RETINAL BASED ANAEMIA PREDICTION USING MACHINE LEARNING

Major project report submitted
in partial fulfillment of the requirement for award of the degree of

Bachelor of Technology
in
Computer Science & Engineering

By

K VARSHITHA	(20UECS0424)	**(18312)**
INDLA POOJITHA	(20UECS1059)	**(18123)**
G CHIRANJEEVI	(20UECS0329)	**(17270)**

Under the guidance of
Mrs.D.HEMALATHA M.E.,
ASSISTANT PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
SCHOOL OF COMPUTING

VEL TECH RANGARAJAN DR. SAGUNTHALA R&D INSTITUTE OF SCIENCE & TECHNOLOGY

(Deemed to be University Estd u/s 3 of UGC Act, 1956)
Accredited by NAAC with A++ Grade
CHENNAI 600 062, TAMILNADU, INDIA

May, 2024

RETINAL BASED ANAEMIA PREDICTION USING MACHINE LEARNING

Major project report submitted
in partial fulfillment of the requirement for award of the degree of

Bachelor of Technology
in
Computer Science & Engineering

By

K VARSHITHA	(20UECS0424)	**(18312)**
INDLA POOJITHA	(20UECS1059)	**(18123)**
G CHIRANJEEVI	(20UECS0329)	**(17270)**

Under the guidance of
Mrs.D.HEMALATHA M.E.,
ASSISTANT PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
SCHOOL OF COMPUTING

VEL TECH RANGARAJAN DR. SAGUNTHALA R&D INSTITUTE OF SCIENCE & TECHNOLOGY

(Deemed to be University Estd u/s 3 of UGC Act, 1956)
Accredited by NAAC with A++ Grade
CHENNAI 600 062, TAMILNADU, INDIA

May, 2024

CERTIFICATE

It is certified that the work contained in the project report titled "RETINAL BASED ANAEMIA PREDICTION USING MACHINE LEARNING " by "K VARSHITHA (20UECS0424), INDLA POOJITHA (20UECS1059), G CHIRANJEEVI (20UECS0329)" has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

Signature of Supervisor
Computer Science & Engineering
School of Computing
Vel Tech Rangarajan Dr. Sagunthala R&D
Institute of Science & Technology
May, 2024

Signature of Professor In-charge
Computer Science & Engineering
School of Computing
Vel Tech Rangarajan Dr. Sagunthala R&D
Institute of Science & Technology
May, 2024

DECLARATION

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

(K VARSHITHA)

Date: / /

(INDLA POOJITHA)

Date: / /

(G CHIRANJEEVI)

Date: / /

APPROVAL SHEET

This project report entitled "RETINAL BASED ANEMIA PREDICTION USING MACHINE LEARNING" by K VARSHITHA (20UECS0424), INDLA POOJITHA (20UECS1059), G CHIRANJEEVI (20UECS0329) is approved for the degree of B.Tech in Computer Science & Engineering.

Examiners **Supervisor**

Mrs.D.HEMALATHA M.E.,

Date: / /

Place:

ACKNOWLEDGEMENT

We express our deepest gratitude to our respected **Founder Chancellor and President Col. Prof. Dr. R. RANGARAJAN B.E. (EEE), B.E. (MECH), M.S (AUTO),D.Sc., Foundress President Dr. R. SAGUNTHALA RANGARAJAN M.B.B.S.** Chairperson Managing Trustee and Vice President.

We are very much grateful to our beloved **Vice Chancellor Prof. S. SALIVAHANAN,** for providing us with an environment to complete our project successfully.

We record indebtedness to our **Professor & Dean, Department of Computer Science & Engineering, School of Computing, Dr. V. SRINIVASA RAO, M.Tech., Ph.D.,** for immense care and encouragement towards us throughout the course of this project.

We are thankful to our **Head, Department of Computer Science & Engineering,Dr.M.S. MURALI DHAR, M.E., Ph.D.,** for providing immense support in all our endeavors.

We also take this opportunity to express a deep sense of gratitude to our Internal Supervisor **Internal Supervisor Mrs.D.HEMALATHA M.E.,** for her cordial support, valuable information and guidance,she helped us in completing this project through various stages.

A special thanks to our **Project Coordinators Mr. V.ASHOK KUMAR, M.Tech., Ms.C. SHYAMALA KUMARI, M.E.,** for their valuable guidance and support throughout the course of the project.

We thank our department faculty, supporting staff and friends for their help and guidance to complete this project.

K VARSHITHA (20UECS0424)
INDLA POOJITHA (20UECS1059)
G CHIRANJEEVI (20UECS0329)

ABSTRACT

Anaemia is a prevalent condition that affects a significant portion of the global population. Early detection and accurate prediction of anaemia can greatly contribute to timely medical intervention and improved patient outcomes. It illustrate the study of a predictive model using decision tree, random forest, navie bayes and XGBoost algorithm to determine the presence or absence of anemia based on eye images. To develop the model a comprehensive dataset of eye images from individuals with and without anaemia was collected. The dataset was preprocessed to extract relevant features from the images, including color variations, texture patterns, and structural characteristics. These features were then used as input variables for the decision tree, random forest, navie bayes and XGBoost models. The performance of each algorithm was evaluated using various metrics which including accuracy.Results demonstrated that all three models achieved promising accuracy rates in predicting anaemia based on eye images. Among them the random forest algorithm exhibited the highest predictive performance followed by XGBoost and decision tree. In conclusion to the finding of indicate that eye data can serve as a valuable source for anemia prediction. The developed models utilizing decision tree, random forest and XGBoost algorithms can effectively identify anaemia cases offering a non-invasive and cost-effective approach to early diagnosis and intervention.The proposed methods uses Random Forest,Decision Tree,Naive Bayes to achieve the classification accuracy on the testing data.Among these three methodologies the proposed model suggests the random forest is most suitable for the anaemia prediction.Performance measures show that Random Forest gives 92% accuracy,Decision Tree gives 69% accuracy,Navie Bayes 84% accuracy and XGBoost 62% accuracy.The result also proved that any empirical data on anaemia is compatible with eyelogical theories.The proposed approach also found to be useful for predicting possible anaemia prediction.It suggest that the prediction accuracy for ensemble model is higher than that of the individual classifier.

Keywords:Decision tree,Random forest,XG Boost and Machine learning techniques,Navie bayes.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

CART	Combination Antiretroviral Therapy
HSV	Hue,Saturation,Value
ML	Machine Learning
ORAL	Oral Medicine Radiology
PC	Post Cibum
RBC	Red Blood Cells
SLR	Systematic Literature Review
UML	Unified Modeling Language
WHO	World Health Organisation
XG	XGBoost

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

1.1 Introduction

Anaemia remains a pervasive global health concern, affecting millions of individuals worldwide and significantly impacting their overall well-being. It is characterized by a deficiency in red blood cells or hemoglobin, leading to various health complications if left undetected or untreated. Timely diagnosis plays a pivotal role in mitigating its adverse effects, prompting researchers and healthcare practitioners to explore innovative approaches for efficient screening and intervention. It depicts the study delves into the realm of machine learning (ML) and its predictive potential in diagnosing anaemia based on ocular parameters. Leveraging the capabilities of ML algorithms-specifically Naive Bayes, Random Forest, and Decision Tree-the research aims to discern patterns and relationships between eye conditions and the presence of anaemia. By utilizing a comprehensive dataset encompassing diverse eye-related metrics and corresponding anaemia status these models were trained and rigorously evaluated to classify individuals as affected by anaemia or otherwise.

The primary objective is to investigate the feasibility of utilizing non-invasive eye-related data for accurate anaemia detection. Through rigorous analysis and experimentation this research seeks to showcase the predictive accuracy of these ML algorithms in identifying anaemia based on ocular features. Successful outcomes would signify the potential for integrating these non-invasive metrics into existing diagnostic procedures, enhancing screening efficiency, and enabling healthcare practitioners to intervene promptly. It significance the study lies in its potential contribution to early detection strategies for anaemia. By uncovering correlations between eye conditions and anaemia status, this research aims to pave the way for novel diagnostic tools, facilitating timely interventions and ultimately improving patient outcomes.

1.2 Aim of the Project

The primary objective is to evaluate the predictive capacity of machine learning algorithms-Decision Tree, Random Forest, and Naive Bayes-using diverse eye-related parameters to diagnose anaemia. By training and assessing these models, the study aims to identify patterns and relationships between ocular features and anaemia status,facilitating efficient and non-invasive screening methods.Ultimately,the project seeks to contribute to early detection strategies,augmenting healthcare practitioners diagnostic capabilities for timely intervention. The primary objective is to evaluate the predictive capacity of machinelearning algorithms-Decision Tree, Random Forest,and Naive Bayes-using diverse eye-related parameters to diagnose anaemia.By training and assessing these models,facilitating efficient and non-invasive screening methods. Ultimately,the project seeks to contribute to early detection strategies, augmenting healthcare practitioners diagnostic capabilities timely intervention.

1.3 Project Domain

To predictive capacity of machine learning algorithms- Decision tree,Random forest,Navie bayes.By training and assessing the models and study the aims to identify patterns and relationships between anaemia status and non-invasive screening methods.The primary objective is to evaluate the predictive capacity of machine learning algorithms-Decision Tree, Random Forest, and Naive Bayes-using diverse eye-related parameters to diagnose anaemia.

By training and assessing these models, the study aims to identify patterns and relationships between ocular features and anaemia status, facilitating efficient and non-invasive screening methods.Ultimately, the project seeks to contribute to early detection strategies, aurgmenting healthcare practitioners diagnostic capabilities for timely intervention.

1.4 Scope of the Project

This project aim is to provide a tool assiant professionals and consumers in finding and choosing whether the person is suffering with anaemia or not. To achieve this goal it depict the development of apporach that allow users goals for disease that sat-

ify a condition based on disease properties such as decision tree,random forest,naive bayes. This project encompasses the exploration of machine learning algorithms potential, specifically Decision Tree, Random Forest, and Naive Bayes, in diagnosing anaemia based on eye condition data. Depicts the study which focuses on utilizing a dataset encompassing diverse eye-related parameters and anaemia status for model training and evaluation. The ultimate goal is to contribute to advancing early disease diagnosis and treatment ,fulfilling the unmet global need for effective diagnostic tools.

Chapter 2

LITERATURE REVIEW

[1] A.Porwit,et al,(2020) described that blood and bone marrow pathology.The book is lavishly illustrated with high quality color images that demonstrate the relevant pathological features and immunohistochemical and molecular markers. The text features a well-organized approach that incorporates practical tips and clues to help avoid pitfalls and to ensure optimal diagnosis.Chapters have been totally rewritten and new chapters have been added especially on myeloid malignancies.The chapters on hematological malignancies have been written so that the reader can apply the latest WHO classifications in their routine daily practice (especially the 2008 WHO classification of tumors of hemopoietic and lymphoid tissues).It have been revised to include new aspects of molecular biology and flow cytometry diagnostics.Many new schematic diagrams and color illustrations have been added to illustrate blood and bone marrow pathology.Access the full text online and download images via expert consult.Chapters have been totally rewritten and some new chapters have been added especially on myeloid malignancies in line with the WHO 2008 classification.It have been revised to include new aspects of molecular biology and updated concerning flow cytometry diagnostics

[2] B.Benoist,et al,(2022) recommanded that the worldwide prevalence of anaemia. It is used for anaemia prevalence data from the WHO vitamin and mineral nutrition information system for 1993-2005 to generate anaemia prevalence estimates for countries with data representative at the national level or at the first administrative level that is below the national level. For countries without eligible data are employed with regression-based estimates which used the UN Human Development Index (HDI) and other health indicators. It is a combined country estimates weighted by their population to estimate anaemia prevalence at the global level by UN regions and by category of human development.Anaemia affects one-quarter of the world's population and is concentrated in preschool-aged children and women making it a global public health problem. Data on relative contributions of causal factors are lacking however, which makes it difficult to effectively address the problem.

[3] E.Beutler,et al,(2023) summarized that the definition of anaemia: what is the lower limit of normal of the blood hemoglobin concentration.The diagnosis of anaemia is an important aspect of the practice of hematology. The first step is to decide whether the patient is infact of anaemic.Unless earlier blood counts are available and they often are not the physician must make his or her decision on the basis of the population distribution of hemoglobin values.The definition of anaemia has attracted considerable interest recently because of epidemiologic studies that suggest that anaemia may be associated with poorer outcomes in a variety of disorders. In many studies the definition of anemia used is that suggested by a WHO expert committee nearly 40 years ago.Bearing the imprimatur of the WHO apparently carries much weight although it shall point out the numbers so casually presented in that document were based on very few data using methods that were inadequate.A reproduction of the entire section dealing with the establishment of this standard is shown in anaeima recently laminated the near universal acceptance of the WHO criteria of anaemia.

[4] E.J.Wang,et al,(2022) focuesd on the hemaapp: noninvasive blood screening of hemoglobin using smartphone cameras.Depict the primary objective of this study is to accumulate, summarize, and evaluate the state-of-the-art for spatio-temporal crime hotspot detection and prediction techniques by conducting a systematic literature review (SLR).This method utilizes different color spaces, such as RGB (Red, Green, Blue), HSV (Hue, Saturation, Value), or YCbCr (Luma, Blue-difference, Red-difference) The authors were unable to find a comprehensive study on crime hotspot detection and prediction while conducting this SLR. Therefore to the best of author's knowledge this study is the premier attempt to critically analyze the existing literature along with presenting potential challenges faced by current crime hotspot detection and prediction systems.

[5] E.McLean,et al,(2022) proposed that the WHO vitamin and mineral nutrition information system worldwide prevalence of anaemia in 2008. Image processing techniques can be employed to measure parameters such as conjunctival color,vessel tortuosity, and vascular density.Anemia often manifests as a paler conjunctiva due to reduced blood supply and these methods can objectively assess the severity of anemia by analyzing color variations.It is used for anaemia prevalence data from the

WHO vitamin and mineral nutrition information system for 1993-2005 to generate anaemia prevalence estimates for countries with data representative at the national level or at the first administrative level that is below the national level. For countries without eligible data it is employed for regression-based estimates which used the UN Human Development Index (HDI) and other health indicators. It is a combined country estimates the weighted by their population to estimate anaemia prevalence at the global level by UN Regions and by category of human development.

[6] J.D.Cook,et al,(2023) analyzed that the quantitative assessment of body iron. The awareness in recent years of the adverse consequences of iron deficiency has prompted renewed efforts to reduce the prevalence of this common micronutrient insufficiency. One of the main reasons for the limited success of programs to combat iron deficiency is the continuing uncertainty about the optimal epidemiologic approach for identifying it and for measuring its severity. The inadequacy of anaemia surveys is reflected in the wide-ranging estimates by various expert committees of the global prevalence of iron deficiency. In a 1985 World Health Organization (WHO) report it was estimated that 15 to 20 of the world's population had iron deficiency anaemia.Despite the lack of new prevalence data estimates of the global prevalence of iron deficiency anaemia have increased to more than two thirds of the world population.More reliable methods to assess iron status are needed to determine the prevalence of iron deficiency and the impact of iron supplementation and fortification trials.In the present article a new method is described for assessing iron status based on the quantitative measurement of body iron.

[7] J.Zhu,et al,(2021) discussed about the noninvasive hemoglobin measurement using unmodified smartphone camera and white flash.Noninvasive hemoglobin measurement using an unmodified smartphone camera and white flash is an innovative approach that holds great potential for medical diagnostics. By leveraging the advanced capabilities of modern smartphone cameras researchers have developed techniques to estimate hemoglobin levels without the need for invasive blood tests. The method involves capturing a close-up image of the user's fingertip using the smartphone's camera along with a white flash. The flash helps illuminate the tissue and the camera captures the reflected light.Sophisticated algorithms then analyze the color information in the image to estimate hemoglobin levels.This noninvasive approach has several advantages.It eliminates the discomfort and inconvenience associated

with traditional blood tests, making it particularly useful in remote or resource-limited settings. It also allows for frequent and easy monitoring of hemoglobin levels, which is crucial for patients with conditions like anaemia or chronic diseases.Traditional methods of measuring hemoglobin levels involve invasive blood tests, which can be time-consuming and uncomfortable for patients.

[8]K.Hasan,et al,(2020) analyzed that rbc pixel analysis of fingertip video images captured from sickle cell patients with low and high levels of hemoglobin.It can provide valuable insights into the relationship between hemoglobin levels and the appearance of red blood cells (RBCs).In individuals with low hemoglobin levels, such as those with sickle cell anemia the rgb pixel analysis may reveal a higher concentration of darker pixels indicating a higher prevalence of sickled RBCs.Sickle-shaped RBCs can cause vascular occlusion leading to tissue damage and organ dysfunction. Conversely in patients with high hemoglobin levels the RGB pixel analysis may show a higher concentration of lighter pixels suggesting a higher prevalence of normal disc-shaped RBCs.Elevated hemoglobin levels can result from conditions like polycythemia which may lead to increased blood viscosity and impaired circulation

[9]M.K.Hasan,et al,(2020) described that smartphone-based human hemoglobin level measurement analyzing pixel intensity of a fingertip video on different color spaces.Smartphone-based human hemoglobin level measurement is an emerging field that aims to provide a convenient and non-invasive method for analyzing blood parameters.Hemoglobin absorbs light differently depending on its oxygenation state leading to variations in pixel intensity. One approach involves analyzing the pixel intensity of a fingertip video captured by the smartphone's camera. This method utilizes different color spaces such as RGB (Red, Green, Blue), HSV (Hue, Saturation, Value), or YCbCr (Luma, Blue-difference, Red-difference). By converting the fingertip video frames into the desired color space the algorithm can extract relevant information about the hemoglobin levels.Hemoglobin absorbs light differently depending on its oxygenation state leading to variations in pixel intensity. Analyzing these variations can provide an estimation of the hemoglobin concentration. This smartphone-based solution offers several advantages,including accessibility, portability, sustanability and cost-effectiveness.

[10] M.J.Koury,et al,(2023) determined that the red blood cell production and ki-

netics and anaemia.Red blood cell production is termed erythropoiesis and occurs in the bone marrow as a component of general hematopoiesis. The marrow microenvironment and erythropoietin a hormone produced in the kidneys tightly regulate erythropoiesis such that circulating red blood cell numbers vary extremely little despite a turnover rate of one-percent daily. Erythropoietin production is controlled by tissue hypoxia which is function of decreased circulating red blood cells a condition termed anemia. In response to anemia red blood cell production can be expanded several-fold in a rapid but graded manner such that polycythemic overproduction does not occur.Erythropoietin increases survival of erythroid progenitor cells in erythroblastic islands, the marrow niche of terminal erythropoiesis, consisting of a central macrophage surrounded by erythroid progenitor cells and erythroblasts at various stages of differentiation. Normal erythropoiesis requires adequate supplies of iron, vitamin B12, and folate as well as coordinated synthesis of hemoglobin the predominant protein of red blood cells responsible for their principal physiological function the delivery of oxygen from the lungs to the other tissues of the body.Anaemia resulting from deficiencies of these required factors, abnormal marrow microenvironment or intrinsic erythroid cell defects result in distinct patterns of abnormal red cell production.

[11] M.Nelson,et al,(2022) depicted that the anaemia in adolescent girls: effects on cognitive function and activity.The long-term consequences of anaemia in adolescent girls may manifest themselves in later adult life principally in terms of bone health and heart health.Risks of osteopaenia(bone loss) and osteoporotic fracture are inversely proportional to levels of exercisethroughout life (Halioua and Anderson, 1989; Prince et at. 1991).Activity is an important protective risk factor in relation to heart disease.Activity levels in adolescence are likely to be reflected in activity levels at middle age (Fox, 1994).Thus, establishment inadolescence of a regular pattern of physical activity will provide a firm physiological foundation for healthy tissues and lay the foundation for good life-long habits.If anaemia in adolescence militates against the establishment of healthy physical activity patterns it will only be possible to demonstrate the association with poor adult health prospectively. Study of osteoporosis or heart disease in adulthood cannot establish Fe status in adolescence retrospectively.

[12]S.G.Miaou,et al,(2023) described the reducing risks, promoting healthy life,

World Health Organization, Geneva.The method involves capturing a close-up image of the user's fingertip using the smartphone's camera along with a white flash.The flash helps illuminate the tissue and the camera captures the reflected light.Sophist icated algorithms then analyze the color information in the image to estimate hemoglo bin levels.This noninvasive approach has several advantages.It eliminates the discomfort and inconvenience associated with traditional blood tests making it particularly useful in remote or resource-limited settings.It also allows for frequent and easy monitoring of hemoglobin levels which is crucial for patients with conditions like anaemia or chronic diseases.

[13] Y.M.Chen,et al,(2021) proposed that examining palpebral conjunctiva for anaemia assessment with image processing methods.Examining the palpebral conjunctiva for anaemia assessment using image processing methods is an innovative approach in the field of medical diagnostics. The palpebral conjunctiva is the thin transparent membrane that covers the inner surface of the eyelids. By analyzing images of the palpebral conjunctiva various features can be extracted and quantified to provide insights into a patient's anemia status. Image processing techniques can be employed to measure parameters such as conjunctival color, vessel tortuosity, and vascular density. Anaemia often manifests as a paler conjunctiva due to reduced blood supply and these methods can objectively assess the severity of anaemia by analyzing color variations.

Chapter 3

PROJECT DESCRIPTION

3.1 Existing System

To predict anaemia using eye images researcher have utilized machine learning techniques such as XG Boost,Decision Tree. They gather a dataset of eye images from individuals with and without anaemia. Features related to anaemia indicators, such as paleness and blood vessel abnormalities, are extracted. XG Boost, a popular gradient boosting algorithm, is trained on this dataset to classify whether a person has anaemia or not based on the extracted features.

Disadvantages: Dependency on Image Quality: Variations in image quality can impact feature extraction accuracy, leading to inconsistent predictions.

Limited Feature Interpretability: Decision Tree complex ensemble nature may hinder direct interpretation of which features influence predictions.

Sensitivity to Imbalanced Data: XG Boost's performance can be affected if the dataset has an imbalance between anaemia and non-anaemia cases.

Computational Complexity: Training XG Boost's models on extensive datasets demands significant computational resources and time.

Potential Overfitting: XG Boost's flexibility might lead to overfitting, reducing its generalizability to new, unseen eye images.

3.2 Proposed System

The aims to predict anaemia using eye images and machine learning algorithms such as Random Forest.The dataset consists of eye images that will be used to train the models. The output of the system will determine whether an individual has anaemia.This is to ensure that the proposed system is not a burden to the company.By leveraging advanced algorithms this system provides a non-invasive and efficient approach for anaemia prediction potentially aiding in early detection and intervention mention advantages of Proposed system

Advantages: Non Invasive: Eliminates the need for invasive tests, making it comfortable and accessible for widespread screening.
Efficient Screening: Swiftly identifies anaemia cases, streamlining the diagnostic process for timely medical attention.
Cost-Effective: Reduces healthcare costs by utilizing readily available eye imaging technology for predictive analysis.
Early Intervention: Enables timely intervention, improving patient outcomes by addressing anaemia at its early stages.

3.3 Feasibility Study

The feasibility of the project is analysed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential. Three key considerations involved in the feasibility analysis.

3.3.1 Economic Feasibility

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

3.3.2 Technical Feasibility

This study is carried out to check the technical feasibility that is, the technical requirements of the system.Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

3.3.3 Social Feasibility

The aspect of study is to check the level of acceptance of the system by the user.This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

3.4 System Specification

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product.It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner.There are various types of test. Each test type addresses a specific testing requirement.

3.4.1 Hardware Specification

Processor : 17/Intel processor
Hard disk : 160GB
Key board : Standard windows keyboard
Monitor : SVGA
RAM : 8GB

3.4.2 Software Specification

Server side script : HTML,CSS,Bootstrap and JS
Operating system : Windows 11
Programming Language : python 3.10
Libraries : Django,pandas,Mysql.connector,Os,Smtplib,Numpy
IDE/Workbench : Pycharm

3.4.3 Standards and Policies

Data set Policy: This policy outlines the data quality requirement for the images and data set used in the training system .It should includes data collection,conjunctival pallor,iris color uniformity,Eyelid swelling to verify the prediction

Privacy and security: This policy should detail the measures taken the protect the privacy and security of the data collected and stored in the system .It should cover data access Transmission.

Health policy: This policy should detail the environmental and health safety requirement for the training system .It should cover any potential hazards associated with the system such as use of eye iris,conjunctival pallor and includes procedures to ensure the safety of the environmental and human health

Standard operating system : Sops should be developed for all activities related to the training system including the data set and management and increase efficiency

Quality assurance : A quality assurance should be developed to ensure that the system data meet the required quality standard .The plan should included for the training of data set.

Anaconda Prompt

Anaconda prompt is a type of command line interface which explicitly deals with the MachineLearning and data modules.A policy should be in place to ensure that only approved packages are installed,and that they are up-to-date and compatible with the system's requirements.This can help to prevent security vulnerabilities and ensure the stability and reliability of the system.

Standard Used: ISO/IEC 27001

Jupyter

It's like an open source web application that allows us to share and create the documents which contains the live code, equations, visualizations and narrative text.To ensure the security of Jupyter,best pratices should followed, such as enabling the authentication and authorization,using HTTP and upcoming data regularly. It can be used for data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning.

Standard Used: ISO/IEC 27001

Chapter 4

METHODOLOGY

4.1 General Architecture

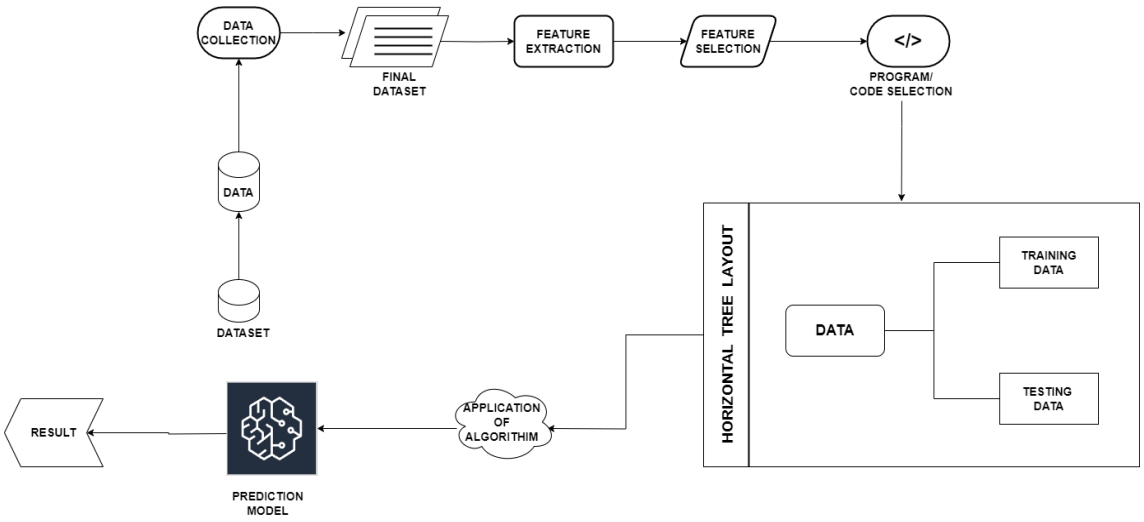

Figure 4.1: **Architecture Diagram**

In the Figure 4.1,analyze the architecture diagram for the Retinal Based Anaeima prediction using machine learning shows a high level view of system's components and their interactions.The diagram illustrates the flow of data from the data set through the data collection and analysis of data to finallise presentation and visualization of anaemia prediction.At the centre of the architecture is the three layout which includes for data training and data testing from the data. These modules work together to extract relevant data set information for the anaeima prediction and transform it into actionable insights.The diagram also shows the different data source and types used in the system such as data extraction, data feature and growth data.Additionally it includes the hardware and software components used in the system such as storage devices and machine learning algorithms.Overall,the architecture diagram provides the clear and comprehensive overview of the system's design highlighting the different components and their interactions.This can help people for the better functionality and capabilities and for improvement or optimization

4.2 Design Phase

4.2.1 Data Flow Diagram

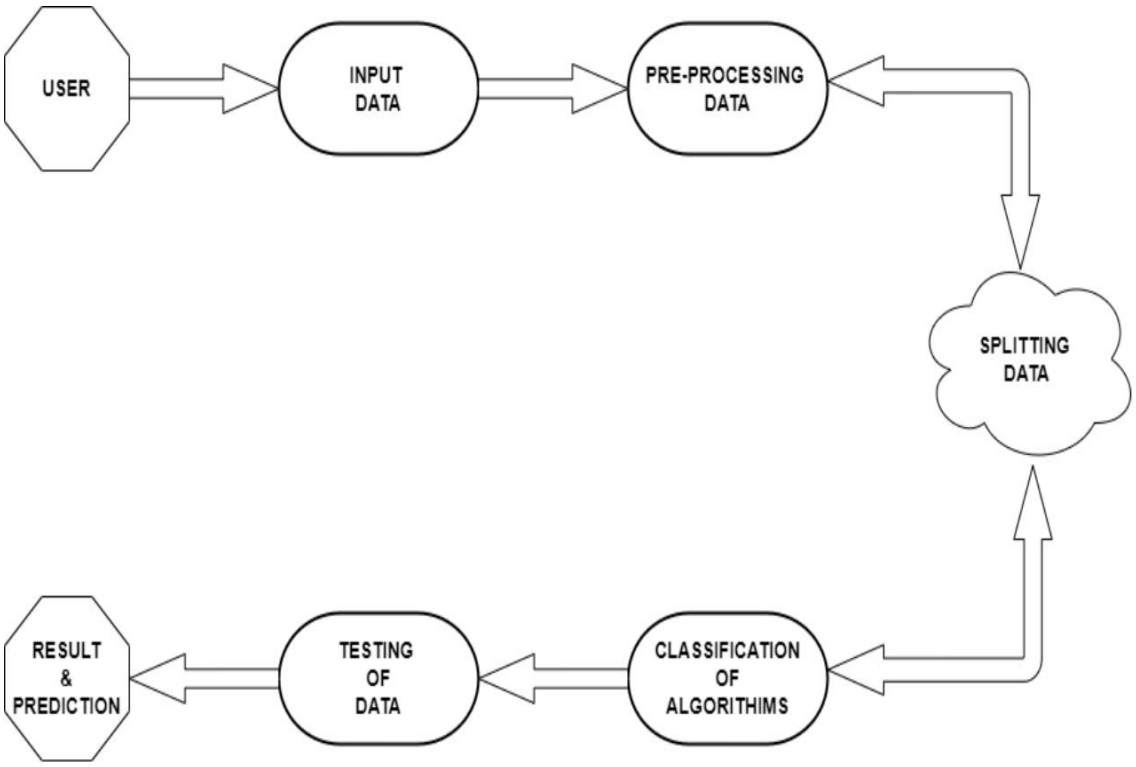

Figure 4.2: **Data Flow Diagram**

In Figure 4.2,describes the information of system, input is the raw data that is processed to produce output. During the input design, the developers must consider the input devices such as PC, MICR, OMR, etc.Therefore, the quality of system input determines the quality of system output. Well designed input forms and screens have following properties It should serve specific purpose effectively such as storing, recording, and retrieving the information.It ensures proper completion with accuracy.It should be easy to fill and straightforward.It should focus on user's attention, consistency, and simplicity.All these objectives are obtained using the knowledge of basic design principles regarding What are the inputs needed for the system.How end users respond to different elements of forms and screens.

4.2.2 Use Case Diagram

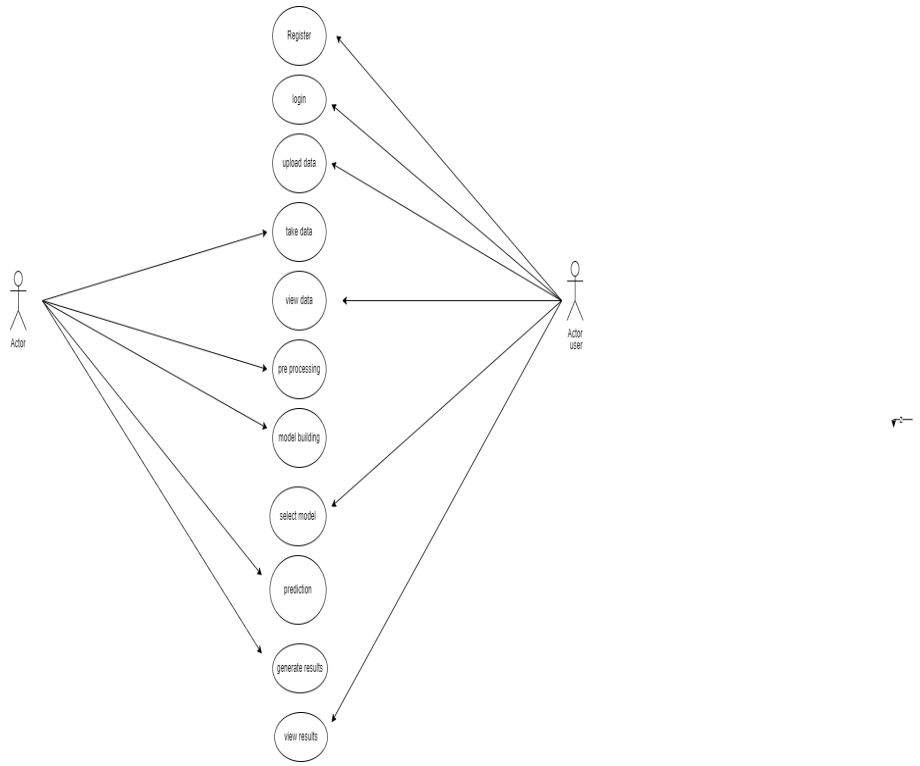

Figure 4.3: **Use Case Diagram**

In Figure 4.3,illustrate the use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted. The model behaviour of a system and help to capture the requirement of the system.It define the context and requirement of either an entire system or the important parts of the system.Use acse describes a function that a system performs to achieve the users goal.An actor represents a role of a user that interacts with the system that you are modeling.

4.2.3 Class Diagram

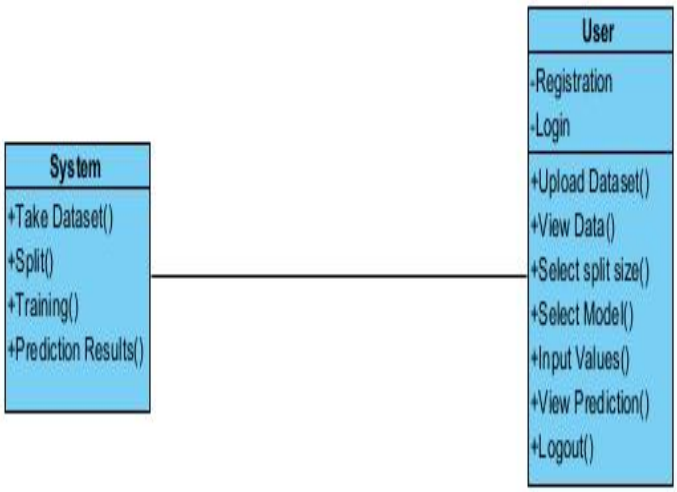

Figure 4.4: **Class Diagram**

In Figure 4.4,shows that the software engineering of a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations or methods and the relationships among the classes. It explains which class contains information.it's a visual representation of class objects in a model system, categorized by class types.Each class type is represented as a rectangle with three compartments for the class name, attributes, and operations.A Class is a blueprint for an object. Objects and classes go hand in hand. We can't talk about one without talking about the other. And the entire point of Object-Oriented Design is not about objects, it's about classes, because we use classes to create objects.A Class is a blueprint for an object. Objects and classes go hand in hand. We can't talk about one without talking about the other. And the entire point of Object-Oriented Design is not about objects, it's about classes, because we use classes to create objects.

4.2.4 Sequence Diagram

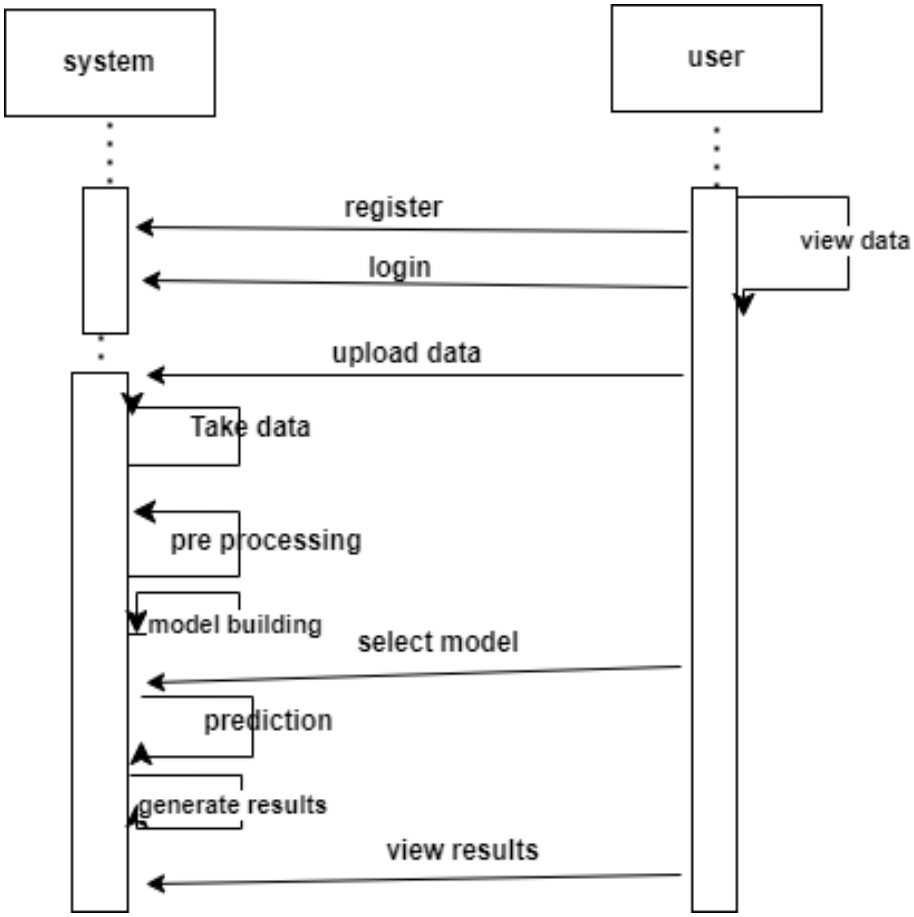

Figure 4.5: **Sequence Diagram**

In Figure 4.5,demonstrates about the sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios and timing diagrams.depict the consists of a group of objects that are represented by lifelines, and the messages that they exchange over time during the interaction.A sequence diagram shows the sequence of messages passed between objects. Sequence diagrams can also show the control structures between objects.Conventionally, the objects involved in the operation are listed from left to right according to when they take part in the message sequence. However, the elements on the horizontal axis may appear in any order.These diagrams are used by software developers and business professionals to understand requirements for a new system or to document an existing process.

4.2.5 Collaboration Diagram

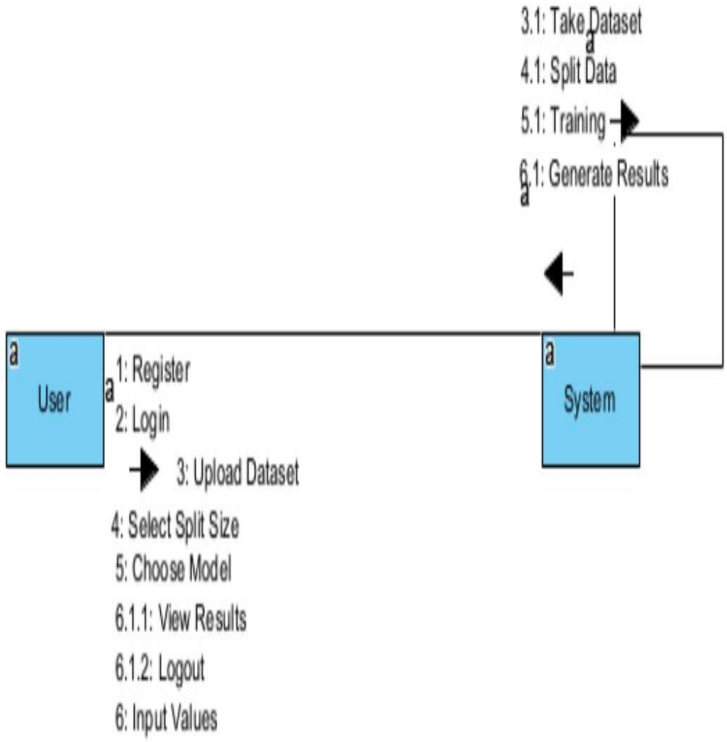

Figure 4.6: **Collaboration Diagram**

In Figure 4.6,depict the collaboration diagram of method call sequence is indicated by some numbering technique as shown below. The number indicates how the methods are called one after another. We have taken the same order management system to describe the collaboration diagram. The method calls are similar to that of a sequence diagram. But the difference is that the sequence diagram does not describe the object organization whereas the collaboration diagram shows the object organization.It is an illustration of the relationships and interactions among software objects in the Unified Modeling Language (UML). Developers can use these diagrams to portray the dynamic behavior of a particular use case and define the role of each object.It is known as communication diagram.It works of projects and tasks.

4.2.6 Activity Diagram

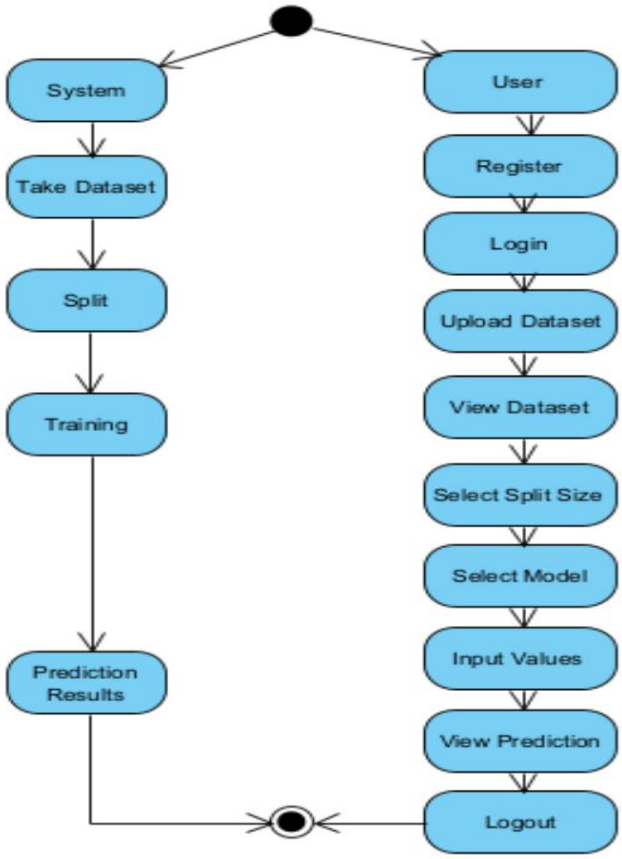

Figure 4.7: **Activity Diagram**

In Figure 4.7,shows the activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.Nodes are nothing but physical hardware's used to deploy the application.The number indicates how the methods are called one after another. We have taken the same order management system to describe the collaboration diagram.A type of Unified Modeling Language (UML) flowchart that shows the flow from one activity to another in a system or process.It is used to describe the different dynamic aspects of a system and is referred to as a 'behavior diagram' because it describes what should happen in the modeled system.

4.3 Algorithm & Pseudo Code

4.3.1 Algorithm

Step 1: User Register with his/her details like Email,password,name.
Step 2: Login with registered credentials.
Step 3: Here, can see a window which user can interact with the machine by giving the input and viewing the result.
Step 4: User can load the dataset he/she want to work on.
Step 5: In this step can able to view dataset uploaded by user and start train the model by the given data set.
Step 6: Preprocess the uploaded data to handle missing values, outliers, and in consistencies. Normalize or standardize the data to ensure uniformity and improve the performance of machine learning models.
Step 7: Split the preprocessed data into Test data and train data. Test data should be maximum 30 percentage only.
Step 8: After that, System Train the machine learning model at backend using algorithms Random Forest,Decision Tree algorithms,Navie bayes to predict whether a individual have obesity or not. Utilize labeled data to train the models on historical patterns of risk factors and their corresponding features.Determines the Labelled part of the system.
Step 9: Model Selection, here user can select one model among others according to model accuracy.
Step 10: User need to provide some parameters like Eye dryness,Eye redness, Pupil size, Tear production,conjunctival,Blink rate,Eyelid swelling,Iris color,Macular health, Uvea health, Corneal activity,Optic nerve health,Ocular motility, Vitreous humor clarity, Lacrimal gland function , Blink rate , Eye pressure ,Visual acuity , Macular health, optical mascular health etc for prediction.
Step 11: Click Predict it gives the data type.shows the type among these accuracy and the prediction of anaemia it show the data type accordingly which the person has given the data set . It show that the person person has the anaemia disease or not
Step 12: Logout

4.3.2 Pseudo Code

```
# Import necessary libraries
from flask import Flask, request, jsonify
import joblib
import pandas as pd
def home():
    return 'Welcome to the data Type Prediction App'


# Initialize Flask app
app = Flask(__name__)

# Load pre-trained machine learning model
model = joblib.load('anaemia_prediction_model.pkl')

# Define route for home page
@app.route('/')
def home():
    return 'Welcome to the data Type Prediction App'

# Define route for prediction
@app.route('/predict', methods=['POST'])
def predict():
    # Receive collection data from user input
    Eye_data = request.get_json()

    # Preprocess collection data
    processed_data = preprocess(Eye_data)

    # Make prediction using pre-trained model
    prediction = model.predict(processed_data)

    # Return predicted crime type to user
    return jsonify({'prediction': prediction})

# Define function for preprocessing data
def preprocess(data):
    # Convert data into appropriate format for model input
    # This might involve encoding categorical variables, scaling numerical features, etc.
    processed_data = pd.DataFrame(data)  # Assuming data is in JSON format
    # Add any preprocessing steps here based on your model requirements
    return processed_data
from flask import Flask, request, jsonify

# Run Flask app
if __name__ == '__main__':
    app.run(debug=True)
```

4.4 Module Description

4.4.1 Data Collection

1.User:
1.1 Register:
User can register for the anaemia web application here
1.2 Login:
After registering the user can access his portal
1.3 View Data:
View data after preprocessing (cleaned dataset)
1.4Input :
User will give the input values
1.5Result History:
After giving the inputs, model will predict the result which it was set according to performance, it will predict that the Anaemia are not.

4.4.2 Data Preprocessing

Take Dataset: The dataset for the Anaemia is collected from the kaggle website (kaggle.com).
The size of overall dataset is 73.6 KB (75,404 bytes).
Data preprocessing In preprocessing first of all we will check whether there is any Nan values. If any Nan values is present we will fill the Nan values with different fillna techniques like bfill, ffill, mode, and mean. Here we used the ffill (front fill) technique on our project.
Training the data: Irrespective of the algorithm we select the training is the same for every algorithm.Given a dataset we split the data into two parts training and testing, the reason behind doing this is to test our model/algorithm performance just like the exams for a student the testing is also exam for the model.

4.4.3 Model Selection

Machine learning techniques:
Decision Tree:A tree has many analogies in real life, and turns out that it has influenced a wide area of machine learning, covering both classification and regression. In decision analysis, a decision tree can be used to visually and explicitly represent

decisions and decision making. As the name goes, it uses a tree-like model of decisions. Though a commonly used tool in data mining for deriving a strategy to reach a particular goal.This methodology is more commonly known as learning decision tree from data and above tree is called Classification tree as the target is to classify passenger as survived or died.As the name goes, it uses a tree-like model of decisions.Regression trees are represented in the same manner, just they predict continuous values like price of a house. In general, Decision Tree algorithms are referred to as CART or Classification and Regression Trees.

Random Forest : A random forest is a machine learning technique that's used to solve regression and classification problems. It utilizes ensemble learning, which is a technique that combines many classifiers to provide solutions to complex problems.A random forest algorithm consists of many decision trees. The 'forest' generated by the random forest algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms.The random forest algorithm establishes the outcome based on the predictions of the decision trees. It predicts by taking the average or mean of the output from various trees. Increasing the number of trees increases the precision of the outcome. Features of a Random Forest Algorithm: It provides an effective way of handling missing data. It can produce a reasonable prediction without hyper-parameter tuning. It solves the issue of over fitting in decision trees. In every random forest tree, a subset of features is selected randomly at the node's splitting point.

4.5 Steps to execute/run/implement the project

4.5.1 Installation of Required Dependencies

• Install necessary libraries and tools such as Python,Anaconda,XAMPP Control Panel,Anaconda,sqlyog Enterprise, NumPy, Pandas.
• Use pip to install Anaconda: pip install anaconda.
• Install other required libraries based on project specifications.

4.5.2 Setting up the Development Environment

• Initialize a new Python environment or use an existing one for the project.
• Create a directory structure for the project, including folders for code, data, and models.
• Set up a virtual environment to manage dependencies and ensure project isolation.

4.5.3 Data Collection and Customer Identification

• Gather Eye data from various sources such as private hospital and govt hospital and online plateforms like kaggle .
• Implement data collection mechanisms to Predict Eye disease based on data.
• Develop algorithms for Anaemia identification based on Eye type,Iris collection,Uvea health data,Tear production.

4.5.4 Performance Analysis and Optimization

• Monitor the performance of theload the dataset into work environment and made a check for null values.
• Optimize apply the algorithm and fit the train data and test data.We got the best Accuracy of Randam forest.
• Conduct testing and experimentation after detection, we will get to know the prediction of anaemia model machine learning model detection whether the result Anaemia Type.

4.5.5 Integration and Deployment

• Thoroughly test the integrated system to ensure seamless interaction between components and accurate prediction of anaemia.
• Refactor and optimize code for readability, maintainability, and performance, ensuring efficient processing and analysis of anaemia data.
• Set up the deployment environment by installing necessary dependencies, such as Flask for web application serving, and configuring security measures, including firewalls and encryption protocols.
• Configure monitoring tools to track system performance, detect anomalies, and ensure continuous availability of the anaemia data analysis and prediction application.

Chapter 5

IMPLEMENTATION AND TESTING

5.1 Input and Output

5.1.1 Input Design

Objectives for Input Design: The objectives of input design are To design data entry and input procedures To reduce input volume To design source documents for data capture or devise other data capture methods In Figure 5.1 to design input data records, data entry screens, user interface screens, etc. To use validation checks and develop effective input controls.

	Eye Redne	Eye Dryne	Conjunctiv	Corneal Cl	Pupil Size	Tear Produ	Blink Rate	Visual Acu	Eye Pressu	Retinal He	Iris Color L	Eyelid Swe	Vitreous H	Macular H	Optic Nerv	Ocular Mo	Lacrimal G	Corneal Sc	Uvea Heal	Anemia	Status
2	0.604351	0.625877	0.478576	0.618943	0.561088	0.676392	0.946955	0.435142	0.14665	0.400925	0.627042	0.600381	0.535152	0.695794	0.402496	0.477432	0.654128	0.519117	0.616393		0
3	0.3053	0.70775	0.171636	0.574672	0.576297	0.500416	0.542061	0.166629	0.581399	0.350367	0.302028	0.458866	0.525419	0.673664	0.666972	0.369643	0.561013	0.282678	0.489622		1
4	0.427459	0.572444	0.812538	0.602531	0.651083	0.529676	0.654422	0.504173	0.563186	0.408572	0.584087	0.547697	0.085465	0.473336	0.36607	0.679939	0.408605	0.581402	0.453025		0
5	0.383067	0.395522	0.68964	0.572443	0.36226	0.600682	0.438335	0.634102	0.475251	0.72679	0.469605	0.47381	0.253477	0.381312	0.657833	0.71675	0.248372	0.359432	0.578454		1
6	0.430029	0.524621	0.424506	0.552704	0.58327	0.371061	0.41846	0.489815	0.541876	0.477635	0.614954	0.600051	0.465473	0.745591	0.241634	0.540042	0.511706	0.598109	0.981813		0
7	0.44983	0.58509	0.463958	0.702371	0.500766	0.496773	0.60286	0.509298	0.584973	0.87158	0.389873	0.455268	0.413198	0.75135	0.369177	0.635685	0.565559	0.205777	0.573115		0
8	0.730362	0.396759	0.692133	0.730049	0.529089	0.540322	0.63692	0.432815	0.358708	0.297065	0.536295	0.42385	0.399242	0.553983	0.388904	0.448511	0.657129	0.639371	0.709547		0
9	0.614379	0.725033	0.275113	0.785385	0.473354	0.386299	0.370981	0.576368	0.643552	0.363448	0.561731	0.691103	0.453339	0.4411	0.574782	0.678282	0.620622	0.344763	0.417679		0
10	0.441007	0.495581	0.528735	0.39539	0.343398	0.633357	0.499368	0.448569	0.459548	0.387321	0.596354	0.627858	0.433534	0.346291	0.349564	0.912253	0.488425	0.385092	0.488201		0
11	0.381973	0.362999	0.424048	0.845709	0.484927	0.463104	0.690328	0.273601	0.483988	0.507798	0.518425	0.562079	0.506212	0.280213	0.603729	0.511118	0.590305	0.624362	0.348408		0
12	0.202822	0.381333	0.603152	0.322553	0.50756	0.551689	0.525792	0.579199	0.78111	0.381762	0.551152	0.567702	0.247317	0.610636	0.456351	0.260572	0.525263	0.726317	0.52359		0
13	0.808993	0.344833	0.393476	0.548042	0.482584	0.309388	0.442219	0.534051	0.420544	0.550362	0.256611	0.250811	0.348691	0.408149	0.302549	0.433065	0.521744	0.54823	0.462506		0
14	0.548219	0.774026	0.46971	0.610949	0.392479	0.649284	0.353472	0.774402	0.610876	0.752681	0.485455	0.564305	0.458131	0.356098	0.159154	0.26372	0.419159	0.366733	0.355564		0
15	0.412609	0.258661	0.706858	0.649835	0.435351	0.709091	0.763467	0.500643	0.420306	0.322748	0.468351	0.985359	0.476965	0.374986	0.626284	0.226063	0.357542	0.721204	0.383238		0
16	0.235716	0.527095	0.291866	0.690354	0.401919	0.452823	0.762883	0.51011	0.356782	0.724774	0.410804	0.590546	0.528376	0.39761	0.778933	0.499587	0.810856	0.648742	0.443646		0
17	0.489848	0.501192	0.677365	0.341601	0.381131	0.564932	0.510705	0.812262	0.478639	0.277755	0.682889	0.687538	0.526593	0.316827	0.412595	0.636734	0.598198	0.517354	0.207806		0
18	0.678298	0.386423	0.70268	0.707374	0.252074	0.481779	0.417862	0.376346	0.46128	0.488442	0.663878	0.33422	0.529114	0.412425	0.58182	0.517914	0.804518	0.212074	0.513214		0
19	0.615293	0.541435	0.550844	0.583053	0.370056	0.1543	0.676302	0.561755	0.743819	0.495542	0.330745	0.308529	0.631864	0.327323	0.513861	0.564177	0.51935	0.493795	0.378499		0
20	0.536007	0.339559	0.463602	0.489696	0.320026	0.620174	0.480706	0.666704	0.509705	0.489454	0.549585	0.414143	0.481571	0.465872	0.613414	0.725269	0.483414	0.451107	0.227446		1
21	0.667514	0.694392	0.636872	0.371483	0.514132	0.56256	0.436492	0.863601	0.435658	0.438253	0.71761	0.722657	0.263382	0.410199	0.238159	0.506363	0.422302	0.297091	0.379568		0
22	0.453731	0.28037	0.400917	0.672963	0.300132	0.476505	0.407937	0.600034	0.672112	0.265831	0.672468	0.787484	0.400208	0.474723	0.385338	0.595985	0.582601	0.284905	0.385962		1
23	0.597794	0.839488	0.642591	0.564515	0.505081	0.520105	0.469007	0.351701	0.356397	0.643533	0.415819	0.485962	0.652896	0.455776	0.416186	0.707292	0.640493	0.449575	0.39251		0
24	0.287753	0.534492	0.639765	0.764007	0.494768	0.222351	0.462952	0.710113	0.614532	0.234977	0.363331	0.32656	0.536372	0.462723	0.365597	0.555418	0.536189	0.550226	0.4947		0
25	0.2287	0.261325	0.768764	0.555671	0.358108	0.427289	0.372637	0.392375	0.703477	0.571388	0.346471	0.476213	0.490797	0.661854	0.591611	0.452158	0.475388	0.513765	0.347626		0
26	0.455078	0.417727	0.546851	0.408195	0.265827	0.699891	0.858219	0.280348	0.343558	0.346942	0.71704	0.593852	0.337381	0.478437	0.460364	0.548985	0.626363	0.467387	0.290202		0
27	0.515397	0.610494	0.713734	0.424161	0.720071	0.268929	0.320863	0.720245	0.559649	0.334625	0.663448	0.499144	0.557255	0.455922	0.55439	0.378476	0.579854	0.51917	0.494226		0
28	0.204562	0.669401	0.732427	0.598624	0.429364	0.241959	0.651084	0.325878	0.383219	0.345206	0.681545	0.452444	0.593706	0.526349	0.621297	0.521713	0.465643	0.495848	0.478048		0
29	0.645855	0.587634	0.205906	0.625025	0.806142	0.45399	0.572885	0.83644	0.772983	0.777614	0.404569	0.724549	0.39889	0.481097	0.632788	0.097358	0.541686	0.465743	0.79566		1
30	0.346155	0.371028	0.795583	0.322943	0.366246	0.334772	0.602866	0.619596	0.449297	0.545572	0.70437	0.791389	0.308735	0.539133	0.457416	0.473916	0.834812	0.556085	0.438985		0
31	0.176257	0.44795	0.656215	0.458065	0.595829	0.746047	0.90449	0.443392	0.709585	0.425464	0.664749	0.672174	0.632883	0.438979	0.41631	0.537523	0.302935	0.424582	0.42524		0
32	0.342385	0.778403	0.678766	0.577009	0.569349	0.395524	0.522796	0.598809	0.378215	0.457084	0.525696	0.618888	0.793907	0.525876	0.3366	0.356037	0.319686	0.536044	0.357015		1
33	0.408188	0.470394	0.457473	0.470059	0.452019	0.682257	0.435103	0.747586	0.346761	0.50927	0.406533	0.381761	0.503332	0.391982	0.444944	0.686492	0.433787	0.409905	0.576614		0
34	0.592857	0.811328	0.743313	0.769673	0.512385	0.6507	0.359892	0.469547	0.642129	0.452371	0.306725	0.303959	0.518192	0.228299	0.506988	0.247842	0.452739	0.529166	0.818992		1
35	0.590829	0.301107	0.520475	0.566186	0.199053	0.70991	0.431578	0.230512	0.696772	0.432269	0.7976	0.546044	0.466262	0.619949	0.666732	0.637331	0.398909	0.446789	0.490062		0
36	0.429741	0.541602	0.463779	0.53367	0.641628	0.384334	0.495554	0.280269	0.404496	0.324258	0.54546	0.514571	0.671232	0.640599	0.354982	0.440166	0.482152	0.498369	0.577603		0
37	0.983022	0.550934	0.599859	0.607436	0.584412	0.552329	0.602205	0.646392	0.495907	0.269441	0.652259	0.363595	0.692216	0.374897	0.652636	0.589974	0.786243	0.54719	0.555976		0
38	0.454447	0.357749	0.345223	0.48896	0.48597	0.545862	0.490244	0.452809	0.545396	0.641946	0.373549	0.573325	0.620054	0.745627	0.246049	0.56981	0.5429	0.604907	0.453138		0
39	0.401811	0.63359	0.599665	0.494541	0.474579	0.529606	0.584935	0.5783	0.42554	0.403841	0.620798	0.977215	0.302068	0.346555	0.608026	0.631203	0.395869	0.514897	0.49497		0
40	0.130196	0.780845	0.660067	0.202086	0.536162	0.352883	0.68418	0.631847	0.612445	0.579231	0.374625	0.504799	0.285631	0.528165	0.628345	0.233864	0.517958	0.398855	0.309255		0

Figure 5.1: **Anaemia Data Set**

5.1.2 Output Design

Objectives of Output Design: To develop output design that serves the intended purpose and eliminates the production of unwanted output. In Figure 5.2 to develop the output design that meets the end user's requirements. To deliver the appropriate quantity of output. To form the output in appropriate format and direct it to the right person. To make the output available on time for making good decisions.

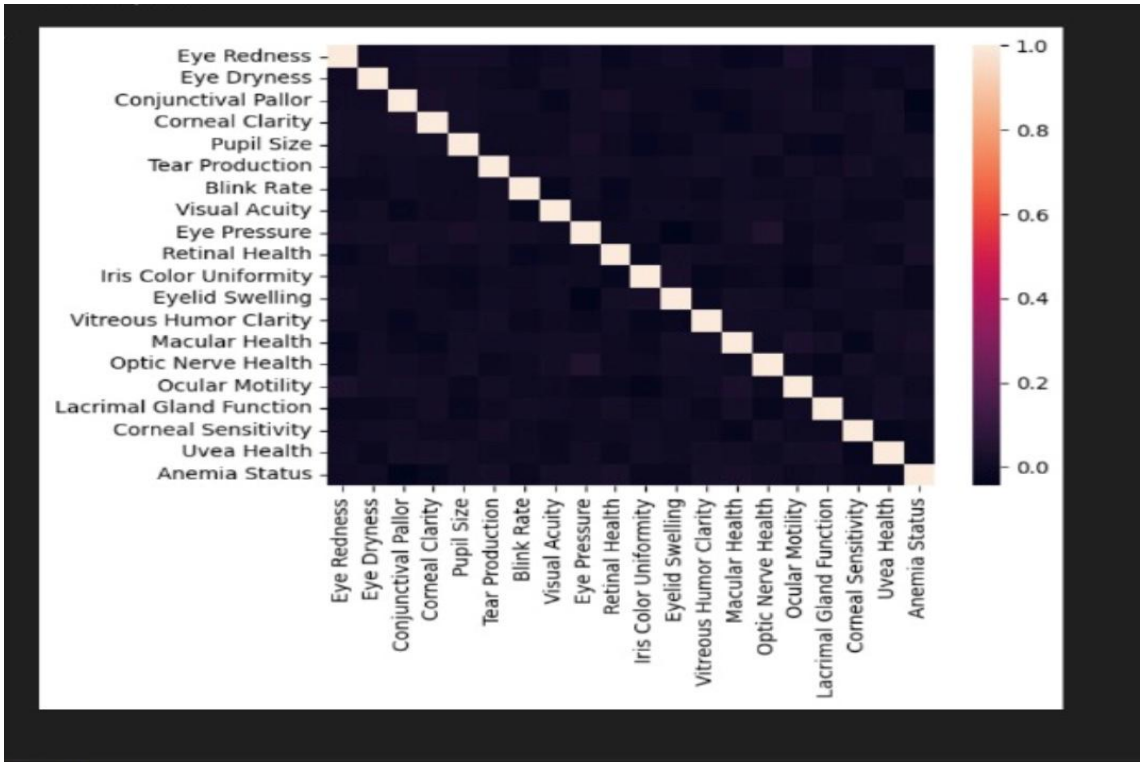

Figure 5.2: **Random Forest Output**

5.2 Testing

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

5.3 Types of Testing

5.3.1 Unit Testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration.

Input

```
from sklearn.ensemble import RandomForestClassifier
ra = RandomForestClassifier()
ra.fit(x_train,y_train)
ra_pred = ra.predict(x_test)
ra_ac = accuracy_score(ra_pred,y_test)
print(f'The accuracy score of RandomForestClassifier {ra_ac}')
ra_clas = classification_report(ra_pred,y_test)
print('----------')
print(ra_clas)
```

Test result

```
    Ran 2 tests in 0.001s
     OK
```

5.3.2 Integration Testing

Integration tests are designed to test integrated software components to determine if they actually run as one program.Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction as shown by successfully unit testing the combination of components is correct and consistent. Integration testing is

specifically aimed at exposing the problems that arise from the combination of components. Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects. The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company.

Input

```
from sklearn.naive_bayes import GaussianNB
nb = GaussianNB()
nb.fit(x_train,y_train)
nb_pred = nb.predict(x_test)
nb_ac = accuracy_score(nb_pred,y_test)
print(f'The accuracy score of naive bayes {nb_ac}')
nb_clas = classification_report(nb_pred,y_test)
print('-----------------------------------')
print(nb_clas)
```

Test result

```
    Ran 4 tests in 0.123s
    OK
```

5.3.3 System Testing

System testing is a crucial part of software development that involves testing the entire system as a whole.It is done after the completion of unit testing and intergration testing and before acceptance testing.The purpose of system testing is to ensure that the system meets all the requirements,functions correctly and is ready to acceptance testing.Retinal based anaemia prediction using machine learning testing involves testing system's performance,functinality and usability.

Input

```
from sklearn.naive_bayes import GaussianNB
nb = GaussianNB()
nb.fit(x_train,y_train)
nb_pred = nb.predict(x_test)
nb_ac = accuracy_score(nb_pred,y_test)
print(f'The accuracy score of naive bayes {nb_ac}')
nb_clas = classification_report(nb_pred,y_test)
print('-----------------------------------')
print(nb_clas)
```

Test Result

```
    Ran 2 tests in 0.234s
    OK
```

Chapter 6

RESULTS AND DISCUSSIONS

6.1 Efficiency of the Proposed System

The proposed system for anaemia detection demonstrates significant efficiency gains through its comprehensive various algorithms such as Random Forest and Navie Bayes on the data set the system harnessess the strengths of each method while identifying their respective limitations.Performance measures show that Random Forest give 92% accuracy,Naive Bayes gives 84% accuracy.

Moreover the system's emphasis on accuracy computation further enhances its efficiency by providing clear metrics for evaluating predictive performance.By combining results from individual algorithm implementation of the system leverages the strengths of ensemble learning which often yields superior predictive accuracy compared to standalone models.This holistic evaluation not only ensures the identification of anaemia with greater precision but also facilitates continuous improvement as new data become available.Ultimately the proposed system stands out for its robustness and effectiveness in leveraging machine learning techincal to address the complex challenge of anaemia detection.

6.2 Comparison of Existing and Proposed System

Existing System: XG Boost Algorithm

To predict anemia using eye data research have utilized machine learning techniques such as XG Boost,Decision Tree.They gather a dataset of eye images from individuals with and without anaemia.Features related to anaemia indicators, such as paleness and blood vessel abnormalities are extracted.XG Boost is trained on this dataset to classify whether a person has anaemia or not based on the extracted features.Another drawback is their inability to provide probabilities directly, requiring additional calibration techniques for probability estimation.Despite these limitations struggle to acheive high accuracy compared to other machine learning algorithms

for eye detection system.XG Boost's flexibility might lead to overfitting, reducing its generalizability to new, unseen eye images.

Proposed System:Random Forest Algorithm

A existing system for retinal based anaemia prediction using machine learning algorithm is random forest.It is equipped with advanced regularization techniques to prevent overfitting a common challenge in machine learning tasks especially when dealing with imbalanced datasets typical of eye detection.Random Forest effectively controls model complexity ensuring generalization to unseen data while maintaining high sensitivity to data indicators.Random forest demonstrates exceptional efficiency in detecting eye data websites by virtue of its robustness, scalability, computational efficiency and adaptability to diverse feature sets.It is a combination of advanced algorithms and optimization techniques.It also empowers machine learning techinques to stay ahead of eye prediction.

Features	**Existing System**	**Proposed System**
Algorithim Used	Limited algorithms like Decision Tree and XG Boost.	Advanced Machine learning models like Random Forest and Navie bayes
Accuracy	Accuracy rate is 69%	Accuracy rate is 92%
Efficiency	Traditional methods may be computationally efficient but may lack the ability to handle large datasets and the efficiency of existing system is 69%.	Machine learning models may require more computational resources for training and prediction, especially if using complex algorithms or dealing with large datasets. However, once trained, they can provide fast predictions the efficiency of proposed system is 92%.
Interpretability	Traditional methods may be more interpretable, as they often rely on straightforward rules or heuristics that are easy to understand.	Some machine learning models, such as decision trees or linear models, offer good interpretability, as they can provide insights into the factors influencing predictions.
Scalability	The scalability of traditional methods may be limited by their design and implementation.	However, scalability may still be a concern depending on the chosen algorithm and computational resources available.

Table 6.1: Comparision Between Existing and Proposed System

6.3 Sample Code

```
from django.shortcuts import render,redirect
from django.contrib.auth.models import User
import pandas as pd
from sklearn.preprocessing import LabelEncoder
from imblearn.over_sampling import SMOTE
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
from sklearn.ensemble import RandomForestClassifier
from sklearn.naive_bayes import GaussianNB

# Create your views here.

def index(req):
    return render(req,'index.html')

def about(req):
    return render(req,'about.html')


def login(req):
    if req.method == 'POST':
        uemail = req.POST['uemail']
        upass = req.POST['passw']
        user = User.objects.filter(email = uemail,password =upass ).exists()
        if user:
            return render(req,'userhome.html')
    return render(req,'login.html')

def register(req):
    if req.method == 'POST':
        uname = req.POST['uname']
        uemail = req.POST['uemail']
        upass = req.POST['passw']
        cpass = req.POST['cpassw']
        num =  req.POST['num']
        if upass == cpass:
            User.objects.create(username = uname, email = uemail, password = upass, first_name = num
                )
            return render(req,'login.html')
    return render(req,'register.html')


def userhome(req):
    return render(req,'userhome.html')


```

```
def view(req):
    global df
    df = pd.read_csv('updated_eye_health_data.csv')
    data = df.to_html()
    return render(req,'view.html',{'data':data})


def module(req):
    try:
        global df,x_train, x_test, y_train, y_test
        # Split The Data in Train and Test
        x = df.drop('Anemia Status',axis=1)
        y = df['Anemia Status']
        sm  = SMOTE()
        a,b = sm.fit_resample(x,y)
        x_train, x_test, y_train, y_test = train_test_split(a, b, test_size = 0.3, random_state= 72)
        if req.method == 'POST':
            model = req.POST['algo']
            print("_")
            print(model)
            if model  == "0":
                msg = 'Please select Algorithem.....'
                return render(req,'module.html',{'msg':msg})
            elif model == "1":
                de = DecisionTreeClassifier()
                de.fit(x_train,y_train)
                de_pred = de.predict(x_test)
                de_ac = accuracy_score(y_test,de_pred)
                msg = "The Accuracy score of DecisionTreeClassifier"+" "+str(de_ac)
                return render(req,'module.html',{'msg':msg})
            elif model == "2":
                ra = RandomForestClassifier()
                ra.fit(x_train,y_train)
                ra_pred = ra.predict(x_test)
                ra_ac = accuracy_score(y_test,ra_pred)
                msg = "The Accuracy score of RandomForestClassifier"+" "+str(ra_ac)
                return render(req,'module.html',{'msg':msg})
            elif model == "3":
                nb = GaussianNB()
                nb.fit(x_train,y_train)
                nb_pred = nb.predict(x_test)
                nb_ac = accuracy_score(y_test,nb_pred)
                msg = "The Accuracy score of naive_bayes"+" "+str(nb_ac)
                return render(req,'module.html',{'msg':msg})
    except NameError:
        msg = "View The Data Once"
        return render(req,'view.html',{'msg':msg})
    return render(req,'module.html')


```

```
97
98  def pred(req):
99      col = x_train.columns
100     print(col)
101     if req.method == 'POST':
102         dic = req.POST.dict()
103         print(dic)
104         del dic['csrfmiddlewaretoken']
105
106         inp = []
107         for i in dic.keys():
108             inp.append(float(dic[i]))
109         print(inp)
110         ra = RandomForestClassifier()
111         ra.fit(x_train,y_train)
112         Output = ra.predict([inp])
113         print(Output)
114         if Output == 0:
115             msg = 'NO Anemia '
116             return render(req,'pred.html',{'col':col[:10],'col1':col[10:19],'msg':msg})
117         else:
118             msg = 'Anemia'
119             return render(req,'pred.html',{'col':col[:10],'col1':col[10:19],'msg':msg})
120     return render(req,'pred.html',{'col':col[:10],'col1':col[10:19]})
```

Predicted Anaemia

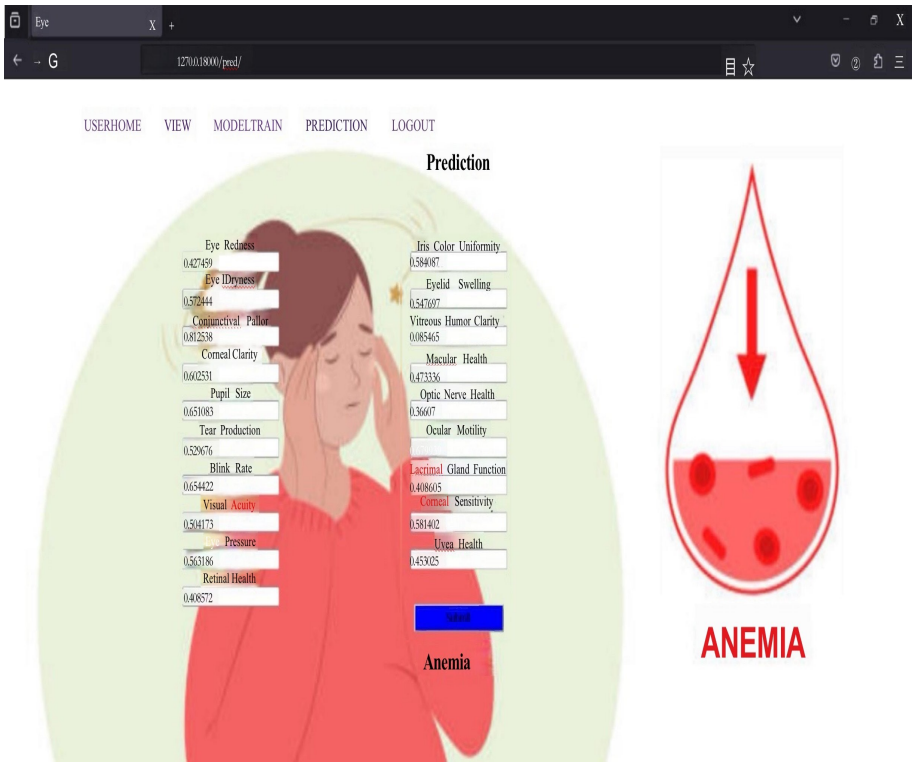

Figure 6.1: **Person With Anaemia**

In Figure 6.1, the idenfied resources for this project include research papers, algorithms, and the anaemia prediction is provided.The reserach papers provide valuable insights into the application of machine learning for anaeima prediction using machine learning. The algorithms such as decision tree,naive bayes are some of the machine learning techniques that can be used to extract features and classify the anaemia data. Together these data provide a comprehensive framework for using machine learning and anaemia prediction in different person bodies.The model handles the imbalanced data and its capacity to adopt varies patterns make it a standout choice for real-world eye detection system.It's strong performance instills confidence in its pratical utility and reinforces its status as a trusted tool in eye detection system.

Predicted without Anaemia

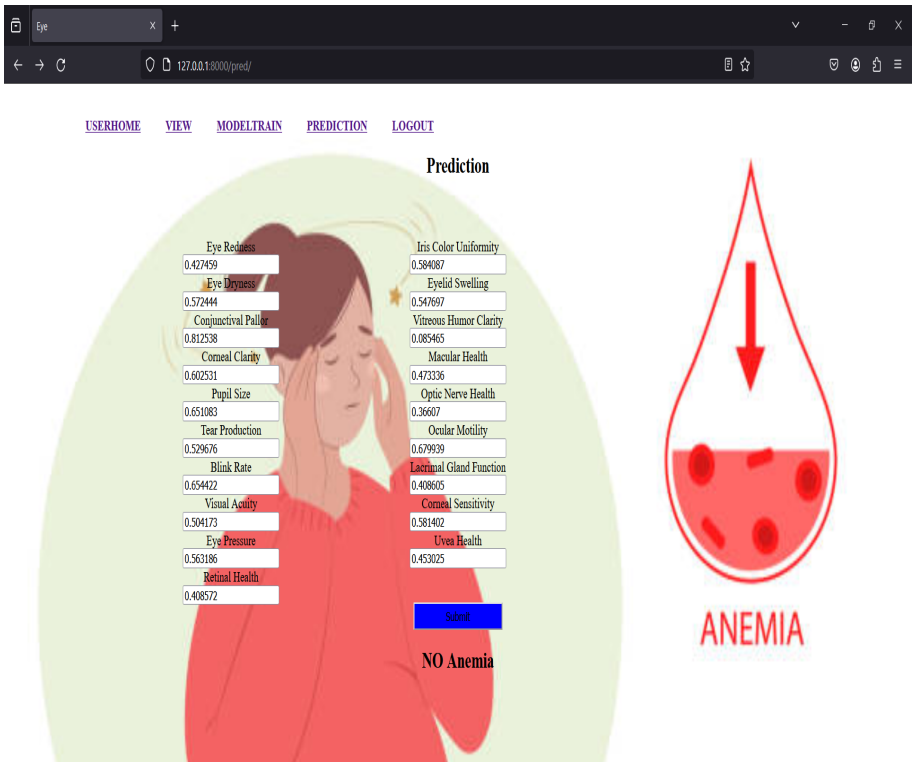

Figure 6.2: **Person Without Anaemia**

In Figure 6.2, Anaemia prediction is the process of determining the physical and biological characteristics of anaemia ,which are used to evaluate its suitability for various uses such as predicting.Anaemia data measures is important for protecting public health and the environment as well as ensuring the sustainability of anameia prediction.In figure 6.2 the data dervies the person is not suffering with anaemia.To develop output design that serves the intended purpose and eliminates the production of unwanted output.To develop the output design that meets the end user's requirements. To deliver the appropriate quantity of output.To form the output in appropriate format and direct it to the right person.To make the output available on time for making good decisions.The design of output is the most important task of any system. During output design, developers identify the type of outputs needed, and consider the necessary output controls and prototype report layouts.

Chapter 7

CONCLUSION AND FUTURE ENHANCEMENTS

7.1 Conclusion

In conclusion, the proposed retianl based anaemia prediction using machine learning offers a more effective approach to anaemia prediction.The system's ability to automate data collection,preprocessing and analysis with the use of advanced technologies such as random forest and data analytics will be real time access to accurate and reliable information on anaemia parameter.This will lead to better decision making and quicker action when there is a anaemia issue ultimately reducing the risk of harm to the environment and human health.

The proposed Retinal based on anaemia prediction using machine learning utilizes an algoritham known as random forest for data processing and analysis.Random Forest is a machine learning that has proven to be highly effective in image processing tasks.The use of random forest is the proposed system enables it to analyze retinal based anaemia using machine learning issues accurately identify eye data isssues accurately. The system's efficiency is reported to be 92% meaning that it can accurately identify eye data quality issues with a high degree of precision.This high level of accuracy is due to the system's ability to analyze vast amounts of data in real time using advanced technologies such as tear production, Uvea health, and data analytics.

7.2 Future Enhancements

Future enhancements in this research could involve expanding the dataset to include a wider range of demographic and health-related variables, allowing for a more comprehensive analysis. Integrating additional machine learning algorithms or employing ensemble methods could further refine predictive accuracy. Exploring real-

time data collection through wearable technology or remote monitoring devices for eye-related parameters might offer a dynamic approach for continuous monitoring and early detection. Moreover, conducting longitudinal studies to validate the predictive capabilities of these models over time could strengthen their reliability in clinical settings. Collaboration with healthcare providers for practical implementation and validation would be a crucial step towards effective deployment in screening programs.

Secondly, the proposed system can be integrated with the environment systems to provide a more comprehensive view of the environment.This will enable the system to provide more accurate and detailed information on eye data quality,Includes the source of data and their impact on the environment. For example the system can integrated with eye systems to identify the impact of eye data events on eye quality.

Thirdly, it depict the system can be enhanced to include machine learning algorithm.This will enable the system to learn from the data collected and improve its accuracy and reliability over time.Additionally,machine learning algorithms can be used to identify patterns in the data that may be missed by human analysts,allowing for early identification oif potential eye data quality issues.

Finally, the system can illustrate that can be enhanced to include a public facing dash board. This will enable the public to access real time information on data quality and make informed decision about recreational activities such uvea health.Additionally the dash board can be used to raise awareness about eye quality issues and encourage responsible behaviour among the protect the environment.

There are several potential future enhancement that can be implemented in the proposed retinal based anaemia prediction using machine learning. These enhancements include integration with other environmental systems,Machine learning algorithms,and a public facing dash board.These enhancements will improve the accuracy,reliability and usefulness of the system,enabling better decision making to protect the environment and public health.

Chapter 8

PLAGIARISM REPORT

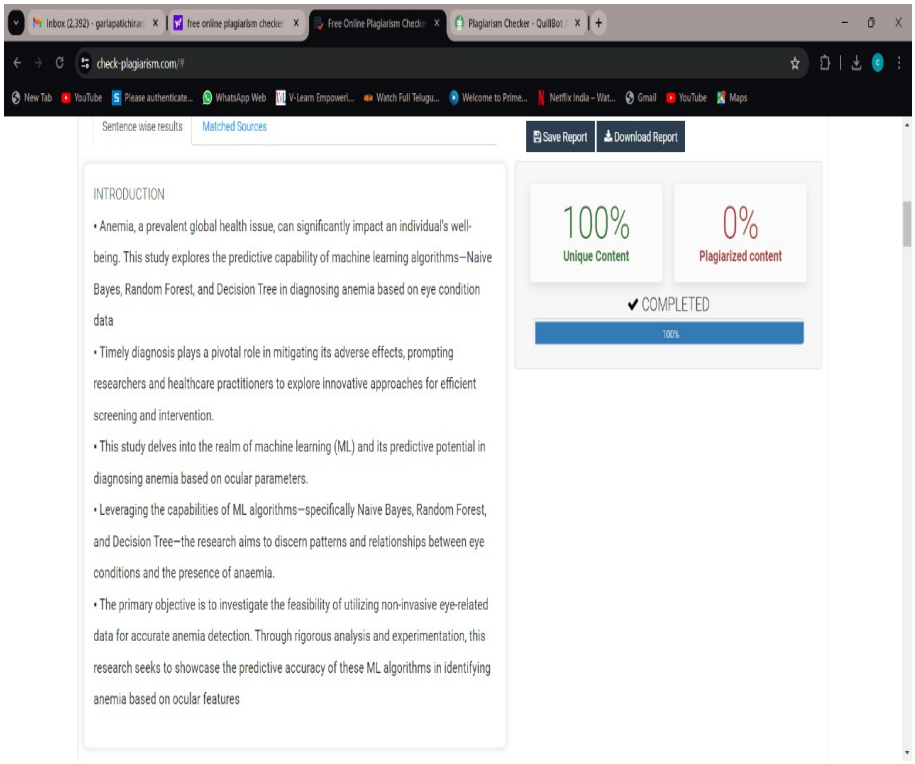

Figure 8.1: **Plagiarism Report**

Chapter 9

SOURCE CODE & POSTER PRESENTATION

9.1 Source Code

```
from django.shortcuts import render,redirect
from django.contrib.auth.models import User
import pandas as pd
from sklearn.preprocessing import LabelEncoder
from imblearn.over_sampling import SMOTE
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
from sklearn.ensemble import RandomForestClassifier
from sklearn.naive_bayes import GaussianNB

# Create your views here.

def index(req):
    return render(req,'index.html')

def about(req):
    return render(req,'about.html')


def login(req):
    if req.method == 'POST':
        uemail = req.POST['uemail']
        upass = req.POST['passw']
        user = User.objects.filter(email = uemail,password =upass ).exists()
        if user:
            return render(req,'userhome.html')
    return render(req,'login.html')

def register(req):
    if req.method == 'POST':
        uname = req.POST['uname']
        uemail = req.POST['uemail']
        upass = req.POST['passw']
        cpass = req.POST['cpassw']
```

```
        num =  req.POST['num']
        if upass == cpass:
            User.objects.create(username = uname, email = uemail, password = upass, first_name = num
                )
            return render(req,'login.html')
    return render(req,'register.html')


def userhome(req):
    return render(req,'userhome.html')


def view(req):
    global df
    df = pd.read_csv('updated_eye_health_data.csv')
    data = df.to_html()
    return render(req,'view.html',{'data':data})


def module(req):
    try:
        global df,x_train, x_test, y_train, y_test
        # Split The Data in Train and Test
        x = df.drop('Anemia Status',axis=1)
        y = df['Anemia Status']
        sm  = SMOTE()
        a,b = sm.fit_resample(x,y)
        x_train, x_test, y_train, y_test = train_test_split(a, b, test_size = 0.3, random_state= 72)
        if req.method == 'POST':
            model = req.POST['algo']
            print("-")
            print(model)
            if model  == "0":
                msg = 'Please select Algorithem.....'
                return render(req,'module.html',{'msg':msg})
            elif model == "1":
                de = DecisionTreeClassifier()
                de.fit(x_train,y_train)
                de_pred = de.predict(x_test)
                de_ac = accuracy_score(y_test,de_pred)
                msg = "The Accuracy score of DecisionTreeClassifier"+" "+str(de_ac)
                return render(req,'module.html',{'msg':msg})
            elif model == "2":
                ra = RandomForestClassifier()
                ra.fit(x_train,y_train)
                ra_pred = ra.predict(x_test)
                ra_ac = accuracy_score(y_test,ra_pred)
                msg = "The Accuracy score of RandomForestClassifier"+" "+str(ra_ac)
                return render(req,'module.html',{'msg':msg})
            elif model == "3":
```

```
                nb = GaussianNB()
                nb.fit(x_train,y_train)
                nb_pred = nb.predict(x_test)
                nb_ac = accuracy_score(y_test,nb_pred)
                msg = "The Accuracy score of naive_bayes"+" "+str(nb_ac)
                return render(req,'module.html',{'msg':msg})
    except NameError:
        msg = "View The Data Once"
        return render(req,'view.html',{'msg':msg})
    return render(req,'module.html')



def pred(req):
    col = x_train.columns
    print(col)
    if req.method == 'POST':
        dic = req.POST.dict()
        print(dic)
        del dic['csrfmiddlewaretoken']

        inp = []
        for i in dic.keys():
            inp.append(float(dic[i]))
        print(inp)
        ra = RandomForestClassifier()
        ra.fit(x_train,y_train)
        Output = ra.predict([inp])
        print(Output)
        if Output == 0:
            msg = 'NO Anemia '
            return render(req,'pred.html',{'col':col[:10],'col1':col[10:19],'msg':msg})
        else:
            msg = 'Anemia'
            return render(req,'pred.html',{'col':col[:10],'col1':col[10:19],'msg':msg})
    return render(req,'pred.html',{'col':col[:10],'col1':col[10:19]})
        if Output == 0:
            msg = 'NO Anemia '
            return render(req,'pred.html',
        else:
        if Output == 0:
            msg = 'Anemia'
            return render(req,'pred.html',{'col':col[:10],'col1':col[10:19],'msg':msg})
    return render(req,'pred.html',{'col':col[:10],'col1':col[10:19]})
```

9.2 Poster Presentation

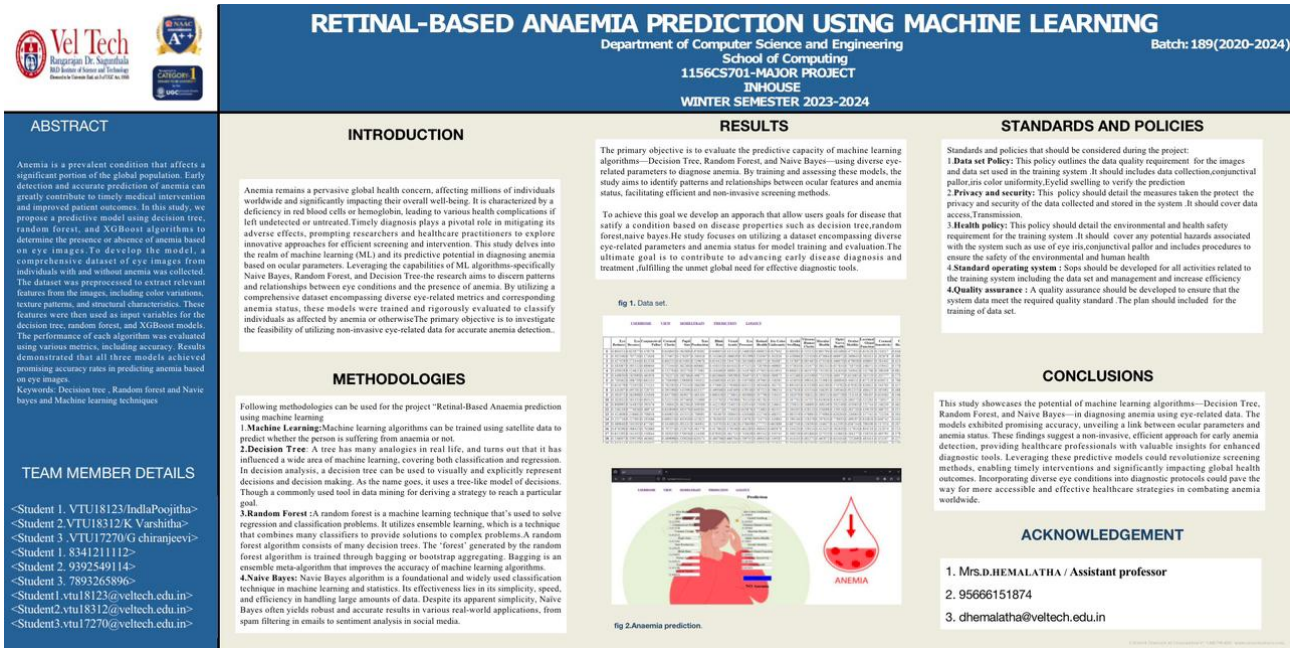

Figure 9.1: **Poster Presentation**

References

[1] A. Porwit,J. McCullough, and W. N. Erber,“Derivation and clinical application of special eye imaging by means of digital cameras and Image for quantification of erythema and pigmentation”, Res.Technol. vol. 14, no. 1, pp. 26–34, Feb. 2021.

[2] B. de Benoist, E. McLean, I. Egli, and M. Cogswell, Eds (2020) “A novel approach to evaluate blood parameters using computer vision techniques”, “A new method and a non-invasive device to estimate anemia based on digital images of the conjunctiva,” vol.107,no.5,6, pp. 46968–46975, 2020.

[3] E. Beutler (2021) ”The definition of anemia”: What is the lower limit of normal of the blood hemoglobin concentration Blood, “A Kalman filtering and nonlinear penalty regression approach for noninvasive anemia detection with palpebral conjunctiva images,”, vol.107,no.5, pp. 1747–1750.

[4] E. J. Wang, W. Li, J. Zhu, R. Rana, and S. N. Patel, “Noninvasive hemoglobin measurement using unmodified smartphone camera and white flash,” in Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Seogwipo, South Korea, 2017,vols pp. 2333–2336. doi: 10.1109/EMBC.2019.8037323.

[5] J.D.Cook,C.H. Flowers, and B.S.Skikne (2020) ‘”The quantitative assessment of body iron”, Blood ‘Toward laboratory blood test-comparable photometric assessments for anemia in veterinary hematology,” J. Biomed. Opt., vol. 21, no. 10, Oct. 2016, Art. no. 107001,, vol.101,no.9, pp.3359–3364.

[6] M. K. Hasan, M. M. Haque, R. Adib, J. F. Tumpa, A. Begum, R. R. Love,Y. L. Kim, and I. A. Sheikh, “SmartHeLP: Smartphone-based hemoglobin level prediction using an artificial neural network,” in Proc. Annu. Symp. (AMIA), vols. 5–6, pp. 26–39, Jan. 2021. pp. 535–544

[7] M.McLean (2022),"Anaemia in adolescent girls: Effects on cognitive function and activity,", "Accuracy of anemia diagnosis by physical examination," Proc. Nutrition Soc., vol. 55, no. 1B, pp. 359–367, 2022.

[8] S.Collings,O.Thompson,E.Hirst, L.Goossens, A.George, and R.Weinkove,"Non-invasive detection of anaemia using digital photographs of the conjunctiva,"vol. 11, no. 4, 2020, Art. no. e0153286. doi: 10.1371/journal.pone.0153286

[9] S.E.McLean (2023)"An Empirical Analysis of Machine Learning Algorithms for Eye Prediction Using images Generalization: An Ensemble Approach" in 2021 IEE Access Doi .1109/ACCESS.2021.3075140

[10] T. Kim, S. H. Choi, N. Lambert-Cheatham, Z. Xu, J. E. Kritchevsky, F.-R. Bertin, and Y. L. Kim, "Toward laboratory blood test-comparable photometric assessments for anemia in veterinary hematology," J. Biomed. Opt., vol. 21, no. 10, Oct. 2016, Art. no. 107001,