

**A PRELIMENERY REPORT ON**

**Machine Learning Driven Anemia**

**Identification and Classification System**

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE

IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE AWARD OF THE DEGREE

OF

**BACHELOR OF ENGINEERING (COMPUTER ENGINEERING)**

##### SUBMITTED BY

ANISH MUKHERJEE 72142644B

TANMAY PARKAR 72142678G

ASHUTOSH SATHE 72142783K

PRATHMESH PATIL 72142698M

## DEPARTMENT OF COMPUTER ENGINEERING

## AISSMS COLLEGE OF ENGINEERING

PUNE 411001

## SAVITRIBAI PHULE PUNE UNIVERSITY

## 2023 -2024



**CERTIFICATE**

This is to certify that the project report entitles

**“MACHINE LEARNING DRIVEN ANEMIA**

**IDENTIFICATION AND CLASSIFICATION SYSTEM”**

Submitted by

ANISH MUKHERJEE B190214276

TANMAY PARKAR B190214286

ASHUTOSH SATHE B190214310

PRATHMESH PATIL B190214291

are bonafide students of this institute and the work has been carried out by them under the supervision of **Prof. S. G. Dhengre** and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune University, for the award of the degree of **Bachelor of Engineering** (Computer Engineering).

**Dr. S. V. ATHAWALE**

Head

Department of Computer Engineering

**Prof. S. G. Dhengre**

Guide

Department of Computer Engineering

**Dr. D. S. BORMANE**

Principal

AISSMS College of Engineering Pune – 41

**External Examiner**

Place: Pune

Date:

**ACKNOWLEDGEMENT**

Deep gratitude is extended to Prof. S. G. Dhengre from the Department of Computer Engineering at AISSMS COE, PUNE, for guidance marked by unsurpassed knowledge and immense encouragement. Appreciation is expressed to Dr. S. V. Athawale, Head of the Department, Computer Engineering, for providing the required facilities for the completion of the project work. Thanks are given to the Principal, Dr. D. S. Bormane of AISSMS COE, PUNE, for the encouragement and cooperation provided in carrying out this work. Acknowledgment is extended to Project Coordinator Prof. V. V. Navale for continuous support and encouragement. The entire teaching faculty of the Department of Computer Engineering is thanked for their suggestions during reviews that facilitated the accomplishment of the project. The Department of Computer Engineering, AISSMS COE, PUNE, is acknowledged for the significant assistance provided in the project's completion. Appreciation is also expressed to Dr. S. D. Patsute and Naidu Hospital for their invaluable support and assistance towards the project's completion. Gratitude is extended to parents, friends, and classmates for their encouragement throughout the project period. Finally, thanks are given to everyone for the direct or indirect support received in the successful completion of this project.

**ANISH MUKHERJEE**

**TANMAY PARKAR**

**ASHUTOSH SATHE**

**PRATHMESH PATIL**

**ABSTRACT**

In India, there is a significant anemia challenge, impacting 57% of women (aged 15-49) and 67% of children (6-59 months). The existing method of anemia detection using Complete Blood Count tests is time-consuming, and rural areas suffer from a lack of access to advanced diagnostic equipment. Existing machine learning models either don't eliminate the need of automated machines or classify only one type of anemia.​ The "Machine Learning Driven Anemia Identification and Classification System" represents an initiative aimed at revolutionizing the diagnostic landscape for anemia within diverse healthcare settings. This project addresses the critical need for an advanced and efficient system that harnesses the power of machine learning to accurately identify and classify anemia, a prevalent hematological condition affecting various demographic groups.

This undertaking commences with an extensive data collection process, gathering diverse blood samples enriched with pertinent hematological parameters from multiple sources. The project's focal point is the development a quick model that ensuring seamless real-time data streaming and integration with existing healthcare infrastructures. Machine learning algorithms, carefully selected for their efficacy, form the cornerstone of the system, undergoing iterative training to achieve heightened accuracy and adaptability.

The system's deployment is not only focused on technological robustness but also considers ethical considerations, including patient data privacy and responsible use of machine learning in healthcare. By shifting from time-consuming and resource-intensive Complete Blood Count tests to a mobile health solution, we eliminate the need for advanced diagnostic equipment in rural areas. The anticipated outcome of this project is a fully operational Machine Learning Driven Anemia Identification and Classification System, poised to transform the efficiency, accuracy, and accessibility of anemia diagnosis.

**TABLE OF CONTENTS**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sr. No.** | | | **Title of Chapter** | | **Page No.** |
| **1** | | | **Introduction** | | 8 |
|  | **1.1** | | Motivation | | 9 |
|  | **1.2** | | Problem Definition | | 10 |
| **2** | | | Literature Survey | | 11 |
| **3** | | | Software Requirements Specification | | 14 |
|  | | **3.1** | **Introduction** | | 15 |
|  | |  | 3.1.1 | Project Scope |  |
|  | |  | 3.1.2 | User Classes and Characteristics |  |
|  | |  | 3.1.3 | Assumptions and Dependencies |  |
|  | | **3.2** | Functional Requirements | | 18 |
|  | |  | 3.2.1 | Image Upload |  |
|  | |  | 3.2.2 | Anemia Detection |  |
|  | |  | 3.2.3 | User Authentication and Authorization |  |
|  | |  | 3.2.4 | Dataset Security |  |
|  | | **3.3** | External Interface Requirements (If Any) | | 20 |
|  | |  | 3.3.1 | User Interfaces |  |
|  | |  | 3.3.2 | Hardware Interfaces |  |
|  | |  | 3.3.3 | Software Interfaces |  |
|  | |  | 3.3.4 | Communication Interfaces |  |
|  | | **3.4** | Nonfunctional Requirements | | 24 |
|  | |  | 3.4.1 | Performance Requirements |  |
|  | |  | 3.4.2 | Safety Requirements |  |
|  | |  | 3.4.3 | Security Requirements |  |
|  | |  | 3.4.4 | Software Quality Attributes |  |
|  | | **3.5** | **System Requirements** | | 32 |
|  | |  | 3.5.1 | Database Requirements |  |
|  | |  | 3.5.2 | Software Requirements |  |
|  | |  | 3.5.3 | Hardware Requirements |  |
|  | | **3.6** | Analysis Models: SDLC Model to be applied | | 34 |
|  | | **3.7** | System Implementation Plan | |  |
| **4** | | | System Design | |  |
|  | | **4.1** | System Architecture | |  |
|  | | **4.2** | Data Flow Diagrams | |  |
|  | | **4.3** | Entity Relationship Diagrams | |  |
|  | | **4.4** | UML Diagrams | |  |
| **5** | | | Other Specification | |  |
|  | | **5.1** | Advantages | |  |
|  | | **5.2** | Limitations | |  |
|  | | **5.3** | Applications | |  |
| **6** | | | Conclusions & Future Work | |  |
|  | | | **Appendix A:** Problem statement feasibility assessment using, satisfiability analysis and NP Hard, NP-Complete or P type using modern algebra and relevant mathematical models.  **Appendix B:** Details of the papers referred in IEEE format (given earlier) Summary of the above paper in not more than 3-4 lines. Here you should write the seed idea of the papers you had referred for preparation of this project report in the following format.  Example:  Thomas Noltey, Hans Hanssony, Lucia Lo Belloz,”Communication Buses for Automotive Applications” In *Proceedings of the* 3rd *Information Survivability Workshop (ISW-2007)*, Boston, Massachusetts, USA, October 2007. IEEE Computer Society.  **Appendix C:** Plagiarism Report | |  |
|  | | | References | |  |

**LIST OF ABBREVATIONS**

1. **AI:** Artificial Intelligence
2. **API:** Application Programming Interface
3. **HIPAA:** Health Insurance Portability and Accountability Act
4. **HTML:** Hypertext Markup Language
5. **HTTPS:** Hypertext Transfer Protocol Secure
6. **JDBC:** Java Database Connectivity
7. **JSON:** JavaScript Object Notation
8. **MTBF:** Mean Time Between Failures
9. **POST:** Power-On Self-Test
10. **REST:** Representational State Transfer
11. **SRS:** Software Requirements Specification
12. **SSL:** Secure Sockets Layer
13. **TLS:** Transport Layer Security
14. **UI:** User Interface
15. **URL:** Uniform Resource Locator
16. **UX:** User Experience
17. **CNN:** Convolutional Neural Network
18. **SVM:** Support Vector Machine
19. **UML:** Unified Modeling Language
20. **RELU:** Rectified Linear Units

**LIST OF FIGURES**

|  |  |
| --- | --- |
| **Illustration** | **Page No.** |
|  |  |
| 1. Waterfall Model | 38 |
| 1. Microscopic Blood Smear Images | 39 |
| 1. Components of Blood | 40 |
| 1. Convolutional Neural Network | 41 |
| 1. System Architecture | 45 |
| 1. Data Flow Diagram | 46 |
| 1. Sequence Diagram 2. Activity Diagram 3. Use case Diagram | 47  48  49 |
| 1. Class Diagram 2. Component Diagram | 50  51 |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **Table** | **Illustration** | **Page No.** |
|  |  |  |
| 4.1 | Project Plan | 29 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**CHAPTER 1**

**INTRODUCTION**

**1.1 OVERVIEW**

The primary goal of our project is to leverage the capabilities of machine learning to expedite and streamline the identification and analysis of anemia in patients through real-time solutions in healthcare. To achieve this, we are actively engaged in the collection and integration of diverse blood sample data, encompassing crucial hematological parameters, sourced from the local hospital. Our focus extends beyond mere data collection to the establishment of a robust real-time data streaming system that efficiently feeds information to our machine learning infrastructure.

In the realm of real-time analysis, we emphasize the development of swift and effective preprocessing techniques tailored to handle incoming data seamlessly, ensuring minimal delays in our analytical pipeline. Feature engineering is a key consideration, with an emphasis on identifying time-sensitive features crucial for rapid anemia identification. Machine learning models are carefully selected or developed to prioritize low latency and high accuracy, and a mechanism for continuous model training is implemented to adapt to evolving data patterns.

Critical to the success of our project is the seamless integration of our solution into healthcare settings. We are mindful of compatibility with existing healthcare information systems and adherence to data privacy regulations., including alerts for potential anemia cases, catering to healthcare practitioners with varying levels of technical expertise.

Monitoring and evaluation mechanisms are integral components, with real-time performance tracking and alert systems in place for anomaly detection during the analysis process. A continuous feedback loop is established to gather insights from medical professionals, enabling iterative improvements based on practical needs and user experiences. Our holistic approach ensures a comprehensive and effective real-time machine learning solution for the timely identification and analysis of anemia in healthcare settings.

**1.2 MOTIVATION**

In rural and underserved areas, the scarcity of blood sample test machines poses a significant challenge to timely and accurate healthcare diagnostics. Our project specifically addresses this issue by leveraging machine learning and computer engineering techniques to introduce efficient and human-error-free blood sample analysis solutions in these regions. The primary objective is to bridge the diagnostic gap in remote areas where traditional testing infrastructure is lacking.

Our focus extends beyond the implementation of machine learning algorithms for blood sample analysis. We are developing and deploying portable, cost-effective diagnostic tools that integrate seamlessly into existing healthcare infrastructures in rural and backward areas. These tools not only provide swift and reliable results but are also designed with user-friendly interfaces to accommodate healthcare professionals with varying levels of technical expertise.

By promoting the integration of machine learning and computer engineering in these underprivileged healthcare settings, our project aims to revolutionize healthcare delivery. We strive to empower local healthcare practitioners with advanced diagnostic capabilities, ultimately contributing to improved health outcomes in rural communities. This targeted approach aligns with our commitment to addressing specific challenges related to diagnostic accessibility, thereby promoting health equity through the strategic application of technology in resource-constrained environments.

**1.3 PROBLEM DEFINITION**

The existing challenge in anemia detection lies in the limitations of conventional methods, especially in resource-limited healthcare settings. Manual analysis, often prone to human error, can lead to delayed diagnoses and subsequently hinder timely treatment. Recognizing this, our project addresses the need for an advanced solution through the development and implementation of a Machine Learning Driven Anemia Identification and Classification System. By harnessing the capabilities of machine learning algorithms, we aim to revolutionize anemia diagnostics, providing a swift, accurate, and automated system that is particularly crucial in environments where traditional diagnostic infrastructure is scarce.

This proposed system is designed to analyze blood sample data with precision, incorporating a comprehensive set of relevant factors for the accurate identification and classification of anemia. By significantly reducing the time required for diagnosis and mitigating the risk of errors associated with manual assessment, the system not only enhances the efficiency of healthcare delivery in well-equipped facilities but also becomes a vital tool in improving healthcare outcomes, especially in underserved communities where access to advanced diagnostic technologies is limited.

**CHAPTER 2**

**LITERATURE REVIEW**

1. **Paper name:** Diagnostic Approach To Anemia In Adults Using Machine Learning [6]

**Authors:** Dr. V. Geetha, Dr. C. K. Gomathy, Kommuru Keerthi, Nallamsetty Pavithra

The study begins with an introduction highlighting the widespread prevalence of anemia, particularly affecting about a quarter of the world's population. The authors stress the need for an effective machine learning regressor capable of accurately detecting anemia. The primary objective is to identify the individual classifier or combination of classifiers that achieve the highest accuracy in categorizing red blood cells for anemia detection.

Two regression algorithms, Lasso and Ridge, are employed for detecting and estimating anemia. The study finds that the Ridge classifier outperforms Lasso, achieving a higher accuracy in anemia detection. The paper emphasizes the importance of utilizing powerful algorithms to maximize accuracy in medical decision-making.

**Pros:**

* + - Estimates anaemia accurately and quickly
    - Provides method for cross-checking results

**Cons:**

* Blood data is input by user
* Does not remove the need of hematology analyzers.

The proposed method aims to provide a better response to inputs, ultimately determining whether a patient is infected with anemia. The paper highlights the significance of anemia diagnosis, as it can have adverse health consequences and impact economic and social development. The literature survey section reviews related research on anemia, including topics such as iron deficiency's impact on work capacity, maternal anemia's association with adverse outcomes, and the comparison of classification techniques using WEKA for hematological data. The authors also discuss existing machine learning methods for anemia detection and highlight the limitations of traditional methods.

The paper proposes a system architecture for anemia estimation, involving user registration, login, input of blood cell levels, system evaluation, and output display. The implementation and results section outlines the use of logistic regression algorithm for prediction, dataset splitting for training and testing, model training, and result generation.

The user-side implementation involves registration, login, selection of split ratio and model, prediction using tested values, and viewing the output. The paper concludes with the potential benefits of the proposed system, including reduced limitations compared to traditional methods.

Overall, the research paper provides insights into the use of machine learning for anemia diagnosis, highlighting the significance of accurate and efficient medical decision-making in the context of a prevalent global health issue.

1. **Paper name:** Machine learning based Diagnosis and Classification Of Sickle Cell Anemia in Human RBC [3]

**Authors:** B. Sen, A. Ganesh, A. Bhan, S. Dixit and A. Goyal.

Machine learning (ML) techniques can be used to automate the process of diagnosing SCA. In tmhis study, the authors used ML techniques to classify RBCs into three shapes: circular, elongated, and other. They used a variety of ML classifiers, including Random Forest, Logistic Regression, Naive Bayes, and Support Vector Machine (SVM).

The authors also used data augmentation techniques to improve the accuracy of the ML classifiers. Data augmentation involves creating new images from existing images by applying transformations such as flipping, zooming, and rotating. The results showed that data augmentation improved the accuracy of all of the ML classifiers, with the SVM classifier achieving an accuracy of 92%. The authors concluded that ML techniques can be used to accurately classify RBCs into three shapes, which can be used to diagnose SCA. They also concluded that data augmentation can improve the accuracy of ML classifiers. The research addresses the challenges of manual inspection. The comparison of machine learning classifiers reveals that Naive Bayes has lower precision, recall, f-score, and accuracy compared to SVM, random forest, and logistic regression. SVM and logistic regression both have 90% accuracy, while random forest achieves 92%.

**Pros:**

* Uses multiple models, increasing generalization.
* The models have an average accuracy of 90%

**Cons:**

* Detects only one type of anemia.
* Needs more resources to compute results.

1. **Paper name:** Red Blood Cell Classification Using Image Processing and CNN.

**Authors:** Parab, M. A., & Mehendale, N. D. [1]

Convolutional neural networks (CNNs) are a type of deep learning algorithm that has been shown to be effective for image classification tasks. In this study, the authors used a CNN to classify red blood cells (RBCs) into nine different classes: normal, microcyte, elliptocyte, stomatocyte, macrocyte, teardrop RBC, codocyte, spherocyte, sickle cell RBC, and Howell jolly RBC. The authors used a dataset of 5,000 RBC images to train and test the CNN. The results showed that the CNN achieved an accuracy of 98.5%.

The authors also compared the performance of the CNN to a traditional image processing approach based on features such as shape, size, and texture. The results showed that the CNN outperformed the traditional image processing approach.

**Pros:**

* The trained model removes any need for a high-end GPU or processor and can be executed on any low configuration machine
* The system can identify and classify one normal and 9 abnormal diferent types of RBCs

**Cons:**

* Not all the RBCs were classified.
* Features need to be manually extracted.

The authors concluded that CNNs are a promising approach for the classification of RBCs. They also concluded that CNNs can achieve high accuracy even when using a small dataset.

1. **Paper:** Machine Learning Algorithms for Anemia Disease Prediction [4]

**Authors:** Jaiswal, M., Srivastava, A., Siddiqui, T.J.

The remarkable advances in health industry have led to a significant production of data in everyday life. These data require processing to extract useful information, which can be useful for analysis, prediction, recommendations, and decision making. Data mining and machine learning techniques are used to transform the available data into valuable information. In medical science, disease prediction at the right time is the central problem for professionals for prevention and effective treatment plan. Sometimes, in the absence of accuracy this may lead to death. In this study, we investigate supervised machine learning algorithms—Naive Bayes, random forest, and decision tree algorithm—for prediction of anemia using CBC (complete blood count) data collected from pathology centers. The results show that Naive Bayes technique outperforms in terms of accuracy as compared to C4.5 and random forest.

Pros:

* The paper explores the performance of three distinct machine learning algorithms (Naive Bayes, random forest, and decision tree), providing a comprehensive comparative analysis.
* The authors provide information about the dataset, including the number of samples and selected attributes. This transparency enhances the replicability of the study.

Cons:

* The paper does not address ethical considerations associated with disease prediction using machine learning, such as privacy concerns and potential biases in healthcare data.
* The dataset size (200 samples) is relatively small, and the paper does not extensively discuss the potential impact of the dataset's size on the model's performance and generalization.

1. **Paper:** Emerging point-of-care technologies for anemia detection [5]

**Author:** An R, Huang Y, Man Y, Valentine RW, Kucukal E, Goreke U, Sekyonda Z, Piccone C, Owusu-Ansah A, Ahuja S, Little JA, Gurkan UA

Anemia, characterized by low blood hemoglobin level, affects about 25% of the world's population with the heaviest burden borne by women and children. Anemia leads to impaired cognitive development in children, as well as high morbidity and early mortality among sufferers. Anemia can be caused by nutritional deficiencies, oncologic treatments and diseases, and infections such as malaria, as well as inherited hemoglobin or red cell disorders. Effective treatments are available for anemia upon early detection and the treatment method is highly dependent on the cause of anemia. There is a need for point-of-care (POC) screening, early diagnosis, and monitoring of anemia, which is currently not widely accessible due to technical challenges and cost, especially in low- and middle-income countries where anemia is most prevalent. This review first introduces the evolution of anemia detection methods followed by their implementation in current commercially available POC anemia diagnostic devices. Then, emerging POC anemia detection technologies leveraging new methods are reviewed. Finally, we highlight the future trends of integrating anemia detection with the diagnosis of relevant underlying disorders to accurately identify specific root causes and to facilitate personalized treatment and care.

**Pros:**

* Quick results, minimal sample volume.
* Wide accessibility, potential remote monitoring.
* Minimal sample requirements, potential continuous monitoring.
* Comprehensive analysis, rapid results, portability.

**Cons:**

* + Calibration needs, potential cost.
  + Reliability on camera quality, standardization challenges.
  + Fabrication complexity, cost considerations.
  + Limited sensitivity, environmental factors.
  + Higher initial cost, maintenance, consumable costs.

**CHAPTER 3**

**SOFTWARE REQUIREMENT AND SPECIFICATIONS**

**3.1 INTRODUCTION**

**3.1.1 Project Scope:**

Machine Learning Driven Anemia Identification and Classification System

Objective: The system allows registered hospital accounts to upload patients' microscopic blood smear images, utilizing a machine learning model for accurate identification and classification of anemia. The scope also includes the user interfaces, data handling processes, and any other components directly related to the successful operation of the system.The Anemia Detection System is designed to address the identification and classification of anemia based on microscopic blood smear images. It provides a platform for registered hospital accounts to upload patient images, utilizing a machine learning model. The primary purpose is to facilitate accurate and timely detection of anemia, specifically targeting cases caused by iron deficiency or vitamin deficiency.

Goals**:** The key objectives and goals of the Anemia Detection System include:

1. Accurate Anemia Identification**:** Implement a robust machine learning model capable of accurately identifying and classifying anemia in patients based on microscopic blood smear images.
2. Efficient Image Processing**:** Develop efficient image processing algorithms to handle the uploaded blood smear images, ensuring timely and accurate results.
3. User-Friendly Interfaces**:** Provide intuitive and user-friendly interfaces for registered hospital accounts, allowing easy and secure upload of patient images and access to the system's results.
4. Data Security**:** Implement robust data security measures to ensure the confidentiality and integrity of patient information and medical data.

Methodology**:** The project will use a variety of machine learning techniques, including convolutional neural networks (CNNs), to develop the anemia identification and classification system. The system will be trained and evaluated on the dataset of blood images.

**3.1.2 User Classes and Characteristics**

The Anemia Detection System is designed to accommodate various user classes with distinct characteristics, reflecting their roles, responsibilities, and interactions with the system. The identified user classes include:

1. Healthcare Professionals (Doctors):
   * Characteristics:
     + High frequency of use, as doctors may use the system for multiple patients.
     + Require in-depth technical expertise related to medical diagnostics.
     + Need secure access and authorization to ensure patient data confidentiality.
   * Responsibilities:
     + Upload microscopic blood smear images for anemia detection.
     + Access and interpret anemia detection results for patient care decisions.
2. Administrators (Hospital Staff):
   * Characteristics:
     + Moderate frequency of use for system configuration and user management.
     + Technical expertise focused on system administration and configuration.
     + Higher-level privileges for managing user accounts and system settings.
   * Responsibilities:
     + Configure and manage user accounts within the hospital.
     + Ensure the system operates smoothly by handling administrative tasks.
3. System Developers:
   * Characteristics:
     + Infrequent use, primarily during system development, updates, or troubleshooting.
     + High technical expertise in software development and system architecture.
     + May require access to system logs and debugging tools for maintenance.
   * Responsibilities:
     + Develop, update, and maintain the Anemia Detection System.
     + Troubleshoot and resolve any technical issues that may arise.

**3.1.3 Assumptions and Dependencies:**

Assumptions:

1. Availability of High-Quality Images:
   * It is assumed that hospitals will provide high-quality microscopic blood smear images for accurate anemia detection. The effectiveness of the machine learning model relies on the quality of the input data.
2. Stable Internet Connection:
   * The assumption is that users will have a stable internet connection for seamless image uploads and result retrieval. System performance may be impacted if users experience connectivity issues.
3. Compliance with Data Privacy Regulations:
   * The system assumes that hospitals and users will adhere to data privacy and healthcare regulations. Non-compliance may have legal implications and affect the project's viability.
4. Timely Maintenance and Updates:
   * It is assumed that the Anemia Detection System will receive timely maintenance and updates to address evolving technological standards, security vulnerabilities, and user feedback.
5. User Training:
   * The assumption is that healthcare professionals and administrators will undergo sufficient training to effectively use and administer the system. Adequate training is essential for optimal utilization.

Dependencies:

1. Machine Learning Framework:
   * The project is dependent on the availability and compatibility of the chosen machine learning framework (e.g., TensorFlow, PyTorch). Changes in the framework's functionality or support could impact system performance.
2. Database Management System:
   * The Anemia Detection System depends on a compatible and reliable database management system (e.g., MySQL, PostgreSQL) for secure data storage and retrieval.
3. Web Server:
   * The system relies on a stable and well-configured web server (e.g., Apache, Nginx) for hosting and serving the application. The performance and accessibility of the system are influenced by the web server's reliability.

**3.2 FUNCTIONAL REQUIREMENTS**

## Image Upload

* 1. Description and Priority

Enables registered hospital accounts to upload microscopic blood smear images for anemia detection. This feature is of High priority.

* 1. Functional Requirements

REQ-1: The system shall support the upload of image files in formats such as JPEG, PNG, or TIFF.

REQ-2: The system shall validate uploaded images for compatibility and resolution requirements.

REQ-3: The system shall provide real-time feedback on the upload progress.

REQ-4: In case of upload failure, the system shall provide an error message and guidance for resolution.

1. **Anemia Detection**
2. Description and Priority

Utilizes a machine learning model to analyze uploaded images and accurately identify and classify anemia. This feature is of High priority.

1. Functional Requirements

REQ-5: The system shall employ a machine learning model for anemia detection.

REQ-6: The system shall classify anemia based on the analysis results.

REQ-7: The system shall provide detailed insights into the anemia detection, including relevant parameters and confidence levels.

REQ-8: Users shall have the option to download detailed anemia detection reports.

1. **User Authentication and Authorization**
2. Description and Priority

Ensures secure access by implementing user authentication mechanisms, allowing only authorized healthcare professionals to use the system. This feature is of High priority.

1. Functional Requirements

REQ-9: The system shall implement secure user authentication using username and password.

REQ-10: The system shall enforce user roles, providing role-based access to functionalities.

REQ-11: In case of unsuccessful authentication, the system shall display an appropriate error message.

1. **Data Security**
2. Description and Priority

Implements robust data security measures to ensure the confidentiality and integrity of patient information and medical data. This feature is of High priority.

1. Functional Requirements

REQ-12: The system shall encrypt patient data during transmission and storage.

REQ-13: Access to patient data shall be restricted based on user roles and authorization levels.

REQ-14: The system shall implement secure protocols for data exchange with external interfaces.

# 3.3 EXTERNAL INTERFACE REQUIREMENTS

## 3.3.1 User Interfaces

**3.3.1.1 Logical Characteristics**

The user interfaces for the Anemia Detection System are designed to be intuitive, user-friendly, and consistent across various interactions. The logical characteristics include:

1. Main Dashboard:
   1. The main dashboard provides an overview of essential functionalities, such as image upload, result viewing, and user management.
   2. Standard buttons for navigation (e.g., "Upload," "Results," "Logout") are prominently displayed.
2. Image Upload Page:
   1. The image upload page features a clean and straightforward layout.
   2. Users can easily select and upload microscopic blood smear images using the "Browse" button.
   3. Real-time feedback on the upload progress is provided to enhance the user experience.
3. Results Page:
   1. The results page displays a list of uploaded cases.
   2. Users can click on a specific case to access detailed anemia detection results.
   3. Standard buttons for actions such as downloading reports are available.
4. Authentication and Authorization:
   1. The login page captures user credentials (username and password) in a secure manner.
   2. Upon successful authentication, users are directed to role-specific interfaces.
   3. Error messages for unsuccessful login attempts adhere to standardized display conventions.
5. Patient Data Access:
   1. Interfaces for accessing patient data adhere to role-based authorization.
   2. Patient data is presented in a readable format with relevant details and analysis results.

**3.3.1.2 Screen Layout Constraints**

1. Screen layouts follow a responsive design, ensuring usability across various device types and screen sizes.
2. Consistent placement of navigation elements and buttons provides a familiar experience.

**3.3.1.3 Standard Buttons and Functions**

1. Standard buttons for common actions, such as "Upload," "Results," and "Logout," are available across relevant interfaces.
2. Help functionality is accessible for users seeking additional information or guidance.

**3.3.1.4 Error Message Display Standards**

1. Error messages are displayed prominently and include clear, concise information about the nature of the error.
2. Messages follow a standardized format for consistency in user communication.

**3.3.1.5 Keyboard Shortcuts**

1. Keyboard shortcuts are not emphasized due to the nature of the user interactions, which primarily involve button clicks and data entry.

**3.3.1.6 Sample Screen Images**

1. Sample screen images illustrating the main dashboard, image upload page, and results page will be included in the user interface specification document.

**3.3.2 Hardware Interfaces**

**3.3.2.1 Logical Characteristics**

The Anemia Detection System interacts with hardware components to ensure seamless operation. The logical characteristics of hardware interfaces include:

1. Supported Device Types:
   * The system is designed to operate on standard computing hardware commonly used in healthcare settings.
   * Supported devices include desktop computers, laptops, and tablets with compatible web browsers.
2. Nature of Data and Control Interactions:
   * Data interactions involve the transfer of microscopic blood smear images between the user's device and the server for analysis.
   * Control interactions include user commands for actions such as image upload, result viewing, and system navigation.

**3.3.2.2 Physical Characteristics**

1. Server Infrastructure:
   * The system relies on a server infrastructure to host the application, manage user data, and execute the machine learning model.
   * Server specifications include sufficient processing power, memory, and storage for efficient image processing.
2. Client Devices:
   * Client devices, including desktop computers, laptops, and tablets, serve as interfaces for users to interact with the system.
   * The physical characteristics of client devices impact the display and usability of the user interface.

**3.3.2.3 Communication Protocols**

1. Internet Protocol (IP):
   * Communication between client devices and the server is facilitated through standard internet protocols.
   * IP addresses are used to identify and route data between devices.
2. Secure Sockets Layer (SSL) Encryption:
   * All data transmissions, including image uploads and result retrieval, are encrypted using SSL to ensure secure communication.
3. Hypertext Transfer Protocol (HTTP) or HTTPS:
   * The system utilizes HTTP or HTTPS for communication between the web server and client devices, depending on the security requirements.

**3.3.2.5 Device Compatibility**

1. Web Browser Compatibility:
   * The system is compatible with standard web browsers, including Google Chrome, Mozilla Firefox, Microsoft Edge, and Safari.
2. Operating System Compatibility:
   * The system supports multiple operating systems, including Microsoft Windows (Versions: 10 and above), Linux (Ubuntu, CentOS, or similar distributions), and macOS (Versions: 10.12 and above).

Note: The above specifications ensure that the Anemia Detection System is compatible with a range of hardware components and operates with optimal efficiency.

**3.3.3 Software Interfaces**

**3.3.3.1 Database Management System (DBMS)**

1. **Type:** MySQL
2. **Purpose:**
   * It ensures data integrity, supports complex querying for analysis, and manages concurrent access in a multi-user environment.
   * MySQL's scalability, security features, and reliability contribute to the system's ability to handle growing data volumes while protecting sensitive patient information.

**3.3.3.2 Web Server**

1. **Type:** Apache HTTP Server 2.4
2. **Purpose:**
   * Apache serves as the web server hosting the Anemia Detection System.
   * It facilitates communication between client devices and the application, handling HTTP or HTTPS requests.

**3.3.3.3 Machine Learning Framework**

1. **Type:** TensorFlow 2.5
2. **Purpose:**
   * TensorFlow is utilized as the machine learning framework for implementing the anemia detection model.
   * The system communicates with TensorFlow to perform image analysis and classification.

**3.3.3.4 Operating System Compatibility**

1. **Supported Operating Systems:**
   * Microsoft Windows (Versions: 10 and above)
   * Linux (Ubuntu, CentOS, or similar distributions)
   * macOS (Versions: 10.12 and above)
2. **Purpose:**
   * The Anemia Detection System is compatible with multiple operating systems to cater to a diverse user base.
   * Data items include system configurations and user preferences specific to each operating system.

**3.3.3.5 Web Browser Compatibility**

1. **Supported Web Browsers:**
   * Google Chrome
   * Mozilla Firefox
   * Microsoft Edge
   * Safari
2. **Purpose:**
   * The system ensures compatibility with standard web browsers for user accessibility.
   * Data items include user interactions, session information, and preferences stored in browser cookies.

**3.4 NON-FUNCTIONAL REQUIREMENTS**

* + 1. **Performance Requirements**
       1. **Image Upload Performance**
* Requirement: The system shall support concurrent image uploads from multiple users without degradation in performance.
* Rationale: To ensure responsiveness and a positive user experience, especially in hospital settings with multiple healthcare professionals uploading images simultaneously.
  + - 1. **Anemia Detection Performance**
* Requirement: The anemia detection process shall complete within a maximum of 15 seconds per image.
* Rationale: To provide timely results to healthcare professionals, facilitating quick decision-making in patient care.
  + - 1. **System Responsiveness**
    - Requirement: The system shall respond to user interactions (e.g., button clicks, result viewing) within 2 seconds.
    - Rationale: To ensure a smooth and interactive user experience, enhancing usability and efficiency.
      1. **Database Query Performance**
* Requirement: Database queries for patient data retrieval shall have a response time of no more than 3 seconds.
* Rationale: To optimize the efficiency of patient data access and support seamless integration with external systems.
  + - 1. **Error Handling Performance**
* Requirement: Error messages shall be displayed to users within 5 seconds of encountering an issue.
* Rationale: To promptly communicate issues to users and provide guidance for resolution.
  + - 1. **Mobile Responsiveness**
* Requirement: The user interface shall be responsive and maintain acceptable performance on tablets and laptops with screen sizes as small as 7 inches.
* Rationale: To support users accessing the system from various devices while maintaining usability and readability.
  + 1. **Safety Requirements**
       1. **Patient Data Confidentiality**

1. Requirement:
   * The system shall ensure the confidentiality of patient data at all times.
2. Safeguards:
   * Data encryption during transmission and storage.
   * Role-based access controls to restrict data access based on user privileges.
   * Regular security audits to identify and mitigate vulnerabilities.

**3.4.2.2 Adherence to Privacy Regulations**

1. Requirement:
   * The system shall comply with relevant privacy regulations, such as HIPAA (Health Insurance Portability and Accountability Act) and GDPR (General Data Protection Regulation).
2. Safeguards:
   * Implementation of privacy-by-design principles in system architecture.
   * Regular reviews and updates to ensure ongoing compliance with evolving regulations.
   * Privacy impact assessments for new features or changes to the system.

**3.4.2.3 Reporting of Safety Incidents**

1. Requirement:
   * Users shall be able to report safety incidents or concerns related to the system.
2. Safeguards:
   * An easily accessible reporting mechanism within the system.
   * Clear procedures for the investigation and resolution of reported incidents.
     1. **Security Requirements**

**3.4.3.1 User Authentication**

1. Requirement:
   * Users shall be required to authenticate their identity before accessing the system.
2. Details:
   * Use of strong password policies, including minimum length and complexity requirements.
   * Implementation of multi-factor authentication for enhanced security.
   * Automatic session timeouts to prevent unauthorized access in case of user inactivity.

**3.4.3.2 Data Encryption in Transit and Storage**

1. Requirement:
   * All data transmitted between client devices and the server, as well as data stored in databases, shall be encrypted.
2. Details:
   * Implementation of TLS (Transport Layer Security) for secure communication.
   * Database encryption mechanisms to protect stored data.

**3.4.3.3 Role-Based Access Control**

1. Requirement:
   * Access to system functionalities and patient data shall be controlled based on user roles.
2. Details:
   * Definition of user roles (e.g., administrator, healthcare professional, lab technician) with specific access privileges.
   * Regular review and adjustment of user roles based on job responsibilities.

**3.4.3.4 Confidentiality and Integrity of Anemia Detection Results**

1. Requirement:
   * Anemia detection results shall be stored and transmitted with confidentiality and integrity.
2. Details:
   * Encryption of anemia detection reports during transmission and storage.
   * Implementation of checksums or digital signatures to verify result integrity.
     1. **Software Quality Attributes**

**3.4.4.1 Usability**

1. Requirement:
   * The system shall achieve a usability score of at least 80% in user satisfaction surveys.
2. Details:
   * Conduct regular usability testing with representative users to gather feedback.
   * Implement user interface improvements based on user feedback.

**3.4.4.2 Reliability**

1. Requirement:
   * The system shall have a mean time between failures (MTBF) of at least 2,000 hours.
2. Details:
   * Continuous monitoring of system performance and reliability metrics.
   * Implementation of automated recovery mechanisms for common failure scenarios.

**3.4.4.3 Maintainability**

1. Requirement:
   * The system shall allow for the implementation of software updates with minimal disruption to user access.
2. Details:
   * Use of modular and well-documented code structures.
   * Automated testing procedures to validate updates before deployment.

**3.4.4.4 Testability**

1. Requirement:
   * The system shall have a test coverage of at least 90% for critical functionalities.
2. Details:
   * Implementation of automated testing suites for unit testing, integration testing, and end-to-end testing.
   * Regular review and update of test cases to reflect changes in system functionalities.

**3.4.4.5 Correctness**

1. Requirement:
   * The system shall achieve a bug-fix resolution time of no more than 48 hours for critical issues.
2. Details:
   * Implementation of continuous monitoring for bug reports and issues.
   * Regular release cycles for bug fixes and updates.
   1. **SYSTEM REQUIREMENTS**

**3.5.1 Database Requirements:**

**MySQL Database Systems:**

MySQL is a popular open-source relational database management system (RDBMS) that offers a robust and reliable platform for storing, managing, and retrieving large volumes of patient data. Its scalability, performance, and open-source nature make it a suitable choice for handling the data used to train and deploy the machine learning model.

MySQL's support for structured data storage and its ability to handle complex queries make it well-suited for organizing and managing the diverse range of patient demographics, clinical parameters, and laboratory test results relevant to anemia diagnosis. Its robust security features and access control mechanisms ensure the confidentiality and integrity of sensitive patient data.

**3.5.2 Software Requirements:**

**Python:**

Python is a versatile and widely used programming language in data science and machine learning due to its ease of use, rich ecosystem of libraries, and extensive community support. Its object-oriented programming paradigm and syntax make it relatively easy to learn and use, allowing developers to focus on the core aspects of machine learning algorithm development without getting bogged down by complex language constructs.

**Machine Learning Libraries:**

Deep learning frameworks like TensorFlow, PyTorch, and Keras have revolutionized the field of machine learning by providing high-level abstractions and efficient implementations of neural network algorithms. These frameworks enable developers to build and train complex machine learning models with relative ease, significantly reducing the time and effort required to develop and deploy state-of-the-art machine learning solutions.

**3.5.3 Hardware Requirements:**

The hardware requirements for training and deploying a machine learning model depend on the complexity of the model and the size of the dataset. In general, more powerful hardware, such as GPUs, is required for training and running complex deep learning models.

* **GPU:** Graphics Processing Units (GPUs) are specialized hardware accelerators designed for parallel computations, making them highly efficient for training and deploying deep learning models. GPUs significantly reduce training time and improve model performance compared to CPUs.
* **Sufficient Storage:** The amount of storage required depends on the size of the dataset and the complexity of the machine learning model. Sufficient storage is crucial to accommodate the data, model files, and intermediate results during training and deployment.
  1. **Analysis Models:**

SDLC Model to be applied is Waterfall Model shown in Figure 1.

* + 1. **Requirement Gathering and analysis** − All possible requirements of the system to be developed are captured in this phase and documented in a requirement specification document.
    2. **System Design** − The requirement specifications from first phase are studied in this phase and the system design is prepared. This system design helps in specifying hardware and system requirements and helps in defining the overall system architecture.
    3. **Implementation** − With inputs from the system design, the system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality, which is referred to as Unit Testing.
    4. **Integration and Testing** − All the units developed in the implementation phase are integrated into a system after testing of each unit. Post integration the entire system is tested for any faults and failures.
    5. **Deployment of system** − Once the functional and non-functional testing is done; the product is deployed in the customer environment or released into the market.
    6. **Maintenance** − There are some issues which come up in the client environment. To fix those issues, patches are released. Also to enhance the product some better versions are released. Maintenance is done to deliver these changes in the customer environment.

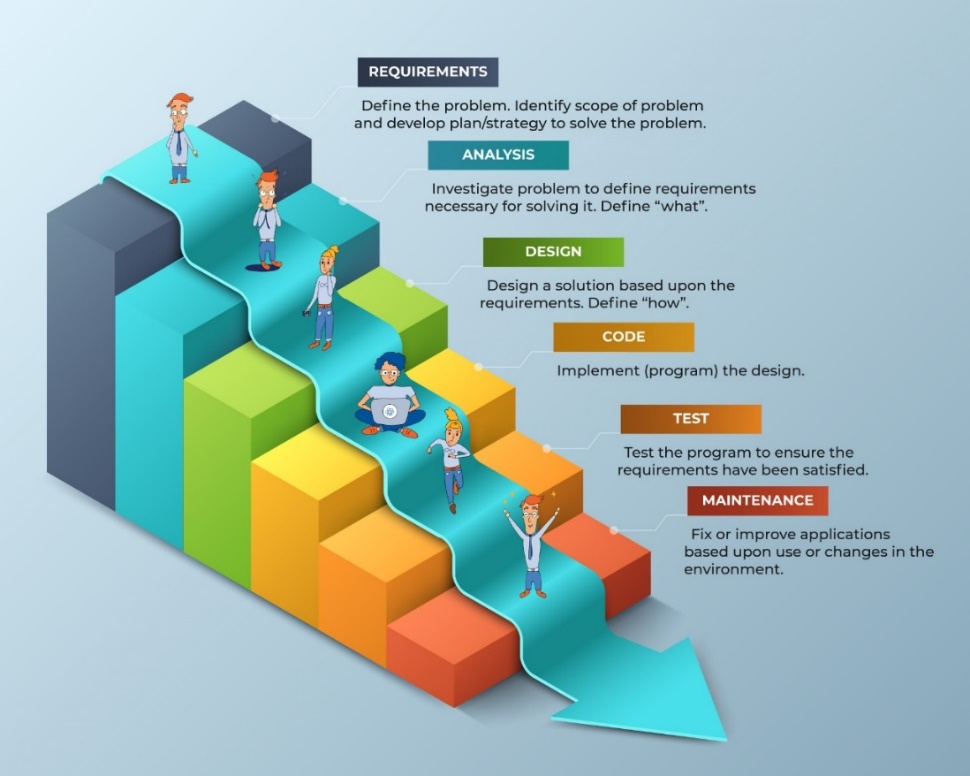


Figure 1. Waterfall Model

* 1. **System Implementation Plan**:

Phases: Divide the implementation into phases, including planning, development, testing, deployment, and maintenance.

Timeline: Establish a detailed timeline for each phase, ensuring realistic deadlines and milestones.

|  |  |  |  |
| --- | --- | --- | --- |
| Sr no. | Name/ Title | Start Date | End Date |
| 1 | Preliminary Survey | 21/07/2023 | 14/08/2023 |
| 2 | Introduction and Problem Statement | 16/08/2023 | 18/08/2023 |
| 3 | Literature Survey | 18/08/2023 | 02/09/2023 |
| 4 | Project Statement | 03/09/2023 | 05/09/2023 |
| 5 | Software Requirement and Specification | 07/09/2023 | 14/09/2023 |
| 6 | System Design | 15/09/2023 | 25/09/2023 |
| 7 | Partial Report Submission | 27/09/2023 | 03/10/2023 |
| 8 | Architecture Design | 18/10/2023 | 27/10/2023 |
| 12 | Paper Publish | 03/11/2023 | 10/11/2023 |
| 13 | Report Submission | 21/11/2023 | 22/11/2023 |

**CHAPTER 4**

**SYSTEM DESIGN**

**4.1 SYSTEM ARCHITECTURE**

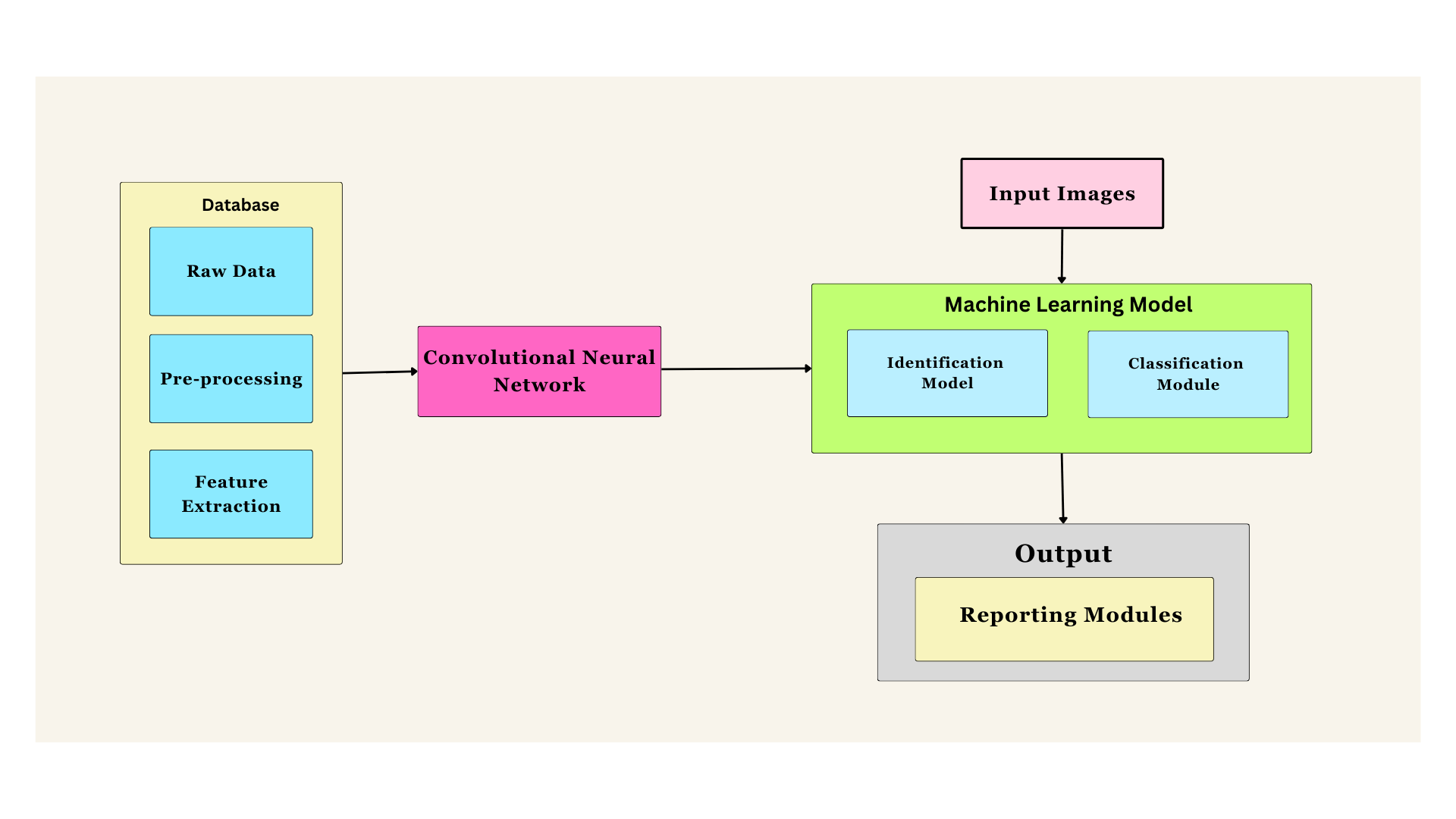
****

Figure 5. System architecture

|  |  |
| --- | --- |
| **Components** | **Description** |
| 1.​Raw Data​ | Microscopic blood smear images and annotations taken from local hospital​ |
| 2.​Pre-Processing​ | Enhance the quality and suitability of the visual data for subsequent analysis.​ |
| 3.​Feature Extraction​ | Strategic selection of relevant features that are indicative of anemia. ​ |
| 4.​Convolutional Neural Network​ | CNNs excel at capturing detailed spatial features in images, making them well-suited for identifying subtle patterns that indicate anemia in blood cells. |
| 5.​Identification Module​ | This module will carry out the identification of anemic cells.​. |
| 6.Classification Module | This module will classify the anemia type |
| 7.​Reporting Module​ | This module will be responsible for reporting the results. |

**4.2 DATA FLOW DIAGRAM**

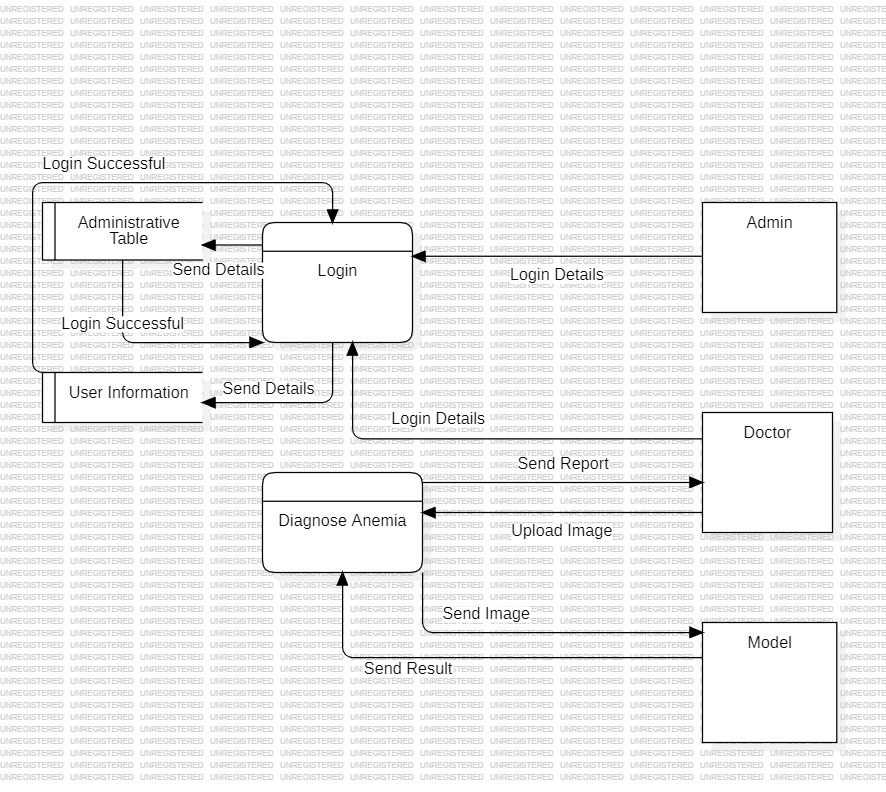


Figure 6. Data Flow Diagram

The machine learning-driven anemia identification and classification system aim to enhance the efficiency of anemia diagnosis. Users submit anemia images, which undergo preprocessing before being analyzed by a Classification Model. The system provides prompt results, including anemia classification and confidence scores, back to the user. The concise data flow diagram illustrates the streamlined process, emphasizing key interactions and data flows. The system's goal is to offer a user-friendly tool for healthcare professionals, contributing to improved diagnostic accuracy and efficiency in anemia identification.

**4.3 UML DIAGRAMS**

**4.3.1 Sequence Diagram**

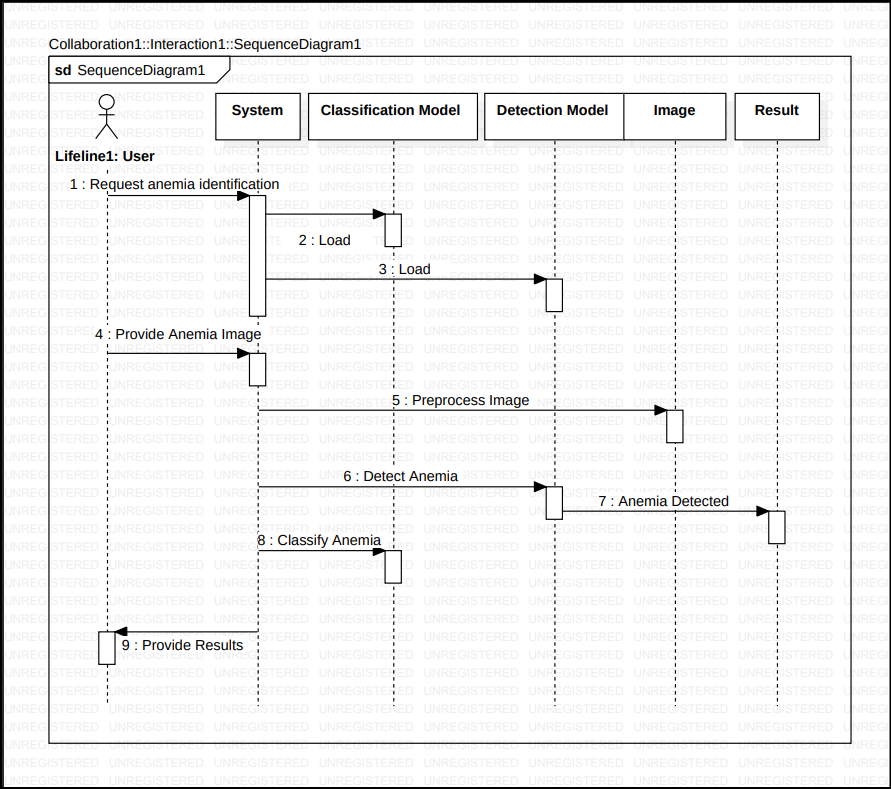


Figure 7. Sequence Diagram

The designed sequence diagram illustrates the series of events that occurs Anemia Identification and Classification System. In this illustration, the actors are represented by a stick man and the transactions or classes are represented by objects. It will give you clear explanation about the behavior of an Anemia Identification and Classification System in terms of processing the flow of instructions.

**4.3.2 Activity diagram**

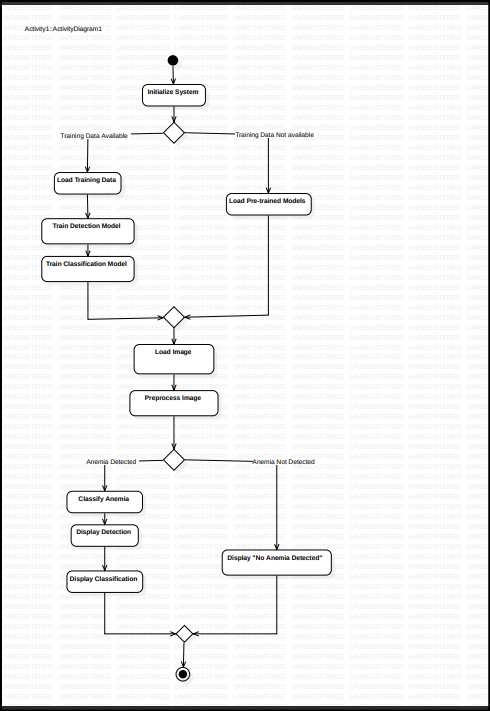


Figure 8. Activity diagram

Here’s the UML activity diagram design of Anemia Identification and Classification System that you can use for your own Final year Project. The UML activity Diagram is used to show the interaction of the user and the system. By creating it, you’ll be able to see the flaws of the system and you may avoid it once you apply it to the project development. So, it is important to have your diagrams designed first before jumping into its development.

**4.3.3 Use case Diagram**

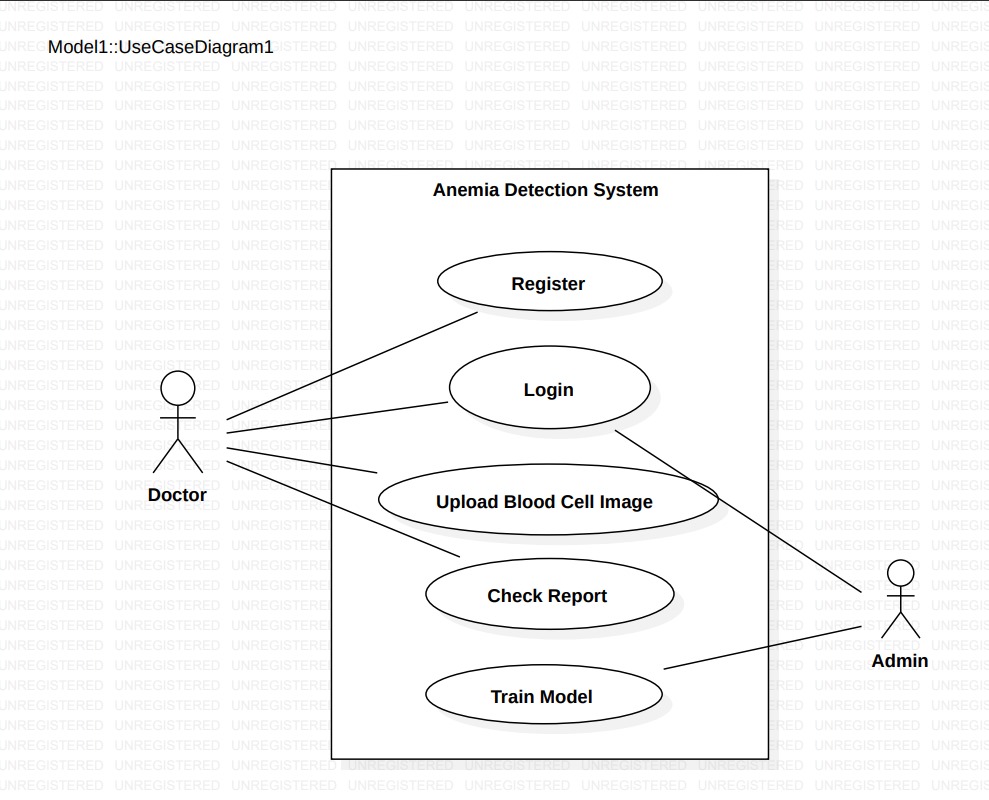


Figure 9. Use case Diagram

The use cases in the diagram represents the main processes in Anemia Identification and Classification System. Then they will be broken down into more specific use cases depending on the included processes of the main use case. Each of these use cases explains how the system handles the actions or scenarios requested by the user.

**4.3.4 Class Diagram**

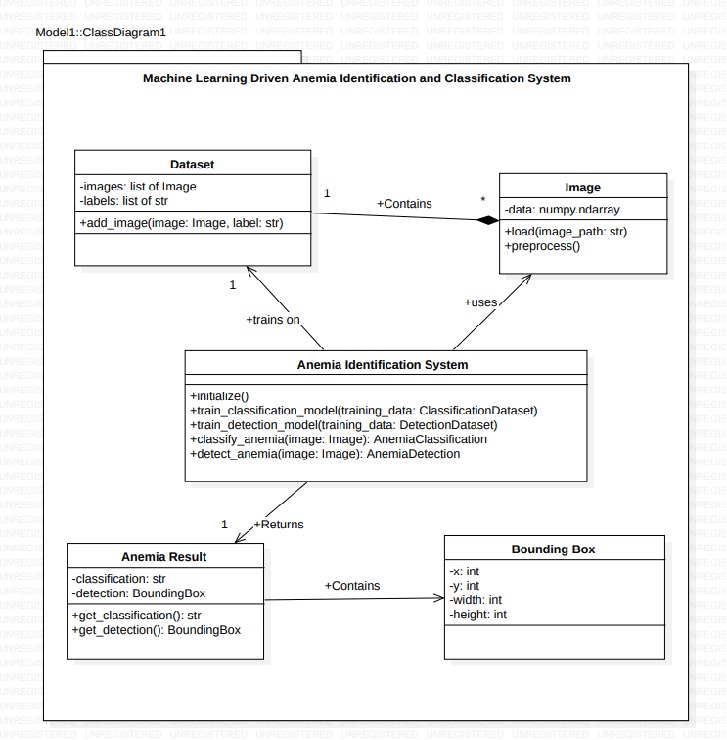


Figure 10. Class Diagram

The Class diagram for Anemia Identification and Classification System shows the structures of information or data that will be handled in the system. These data or information will be represented by classes. Each of the classes will have their attributes in accord to the methods they will use. So, the UML Class diagram was illustrated by a box with 3 partitions and the upper part was the name of the class, the middles are the attributes and the bottom is for the methods. The arrows on them represents their relationships in each other.

**4.3.5 Component Diagram**

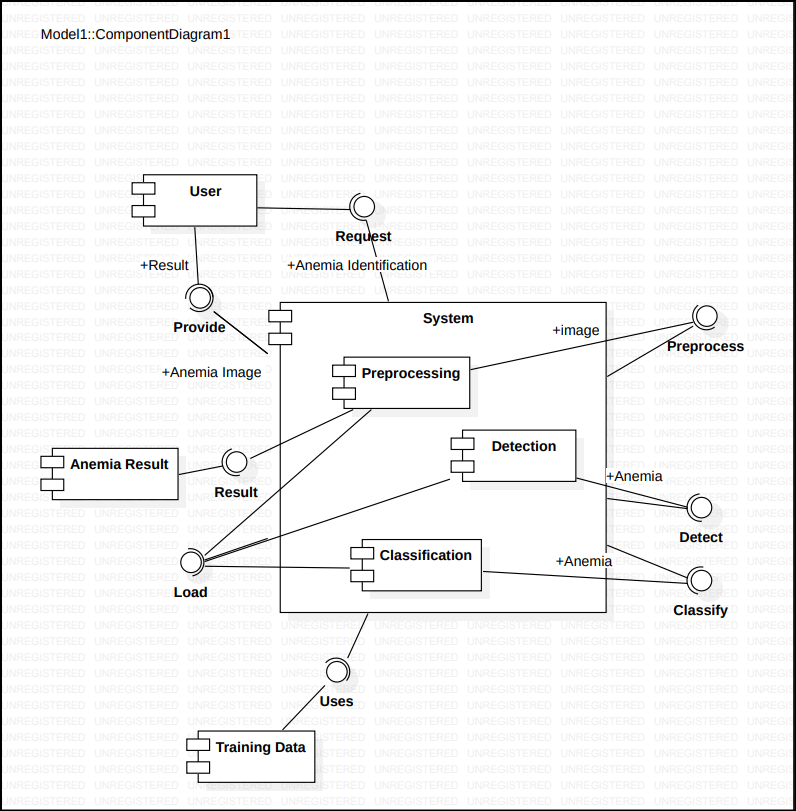


Figure 11. Component Diagram

UML Component diagrams are used in modeling the physical aspects of object-oriented systems that are used for visualizing, specifying, and documenting component-based systems and also for constructing executable systems through forward and reverse engineering. Component diagrams are essentially class diagrams that focus on a system's components that often used to model the static implementation view of a system.

**CHAPTER 5**

**OTHER SPECIFICATIONS**

**5.1 ADVANTAGES**

**5.1.1 Improved Accuracy and Early Detection:**

Machine learning algorithms can analyze vast amounts of patient data, including clinical parameters, laboratory test results, and demographic information, to identify patterns and make predictions with high accuracy. This can lead to earlier detection of anemia, allowing for timely intervention and treatment.

**5.1.2 Enhanced Diagnostic Efficiency:**

The system can automate the analysis of patient data, reducing the time and effort required for manual review and interpretation of laboratory results. This frees up healthcare professionals to focus on providing care and making informed decisions.

**5.1.3 Reduced Human Error and Improved Consistency:**

Machine learning algorithms are less prone to errors and biases that can affect human judgment. They can consistently apply the same decision-making criteria, ensuring consistent and reliable anemia identification across different healthcare settings.

**5.1.4 Continuous Learning and Improvement:**

Machine learning algorithms can continuously learn from new data and improve their performance over time. This allows the system to adapt to changing patterns in anemia diagnosis and classification, ensuring its long-term effectiveness.

**5.2 LIMITATIONS**

**5.2.1 Data Reliance and Potential Biases:**

The system's performance is heavily dependent on the quality and representativeness of the data used to train the machine learning model. Biases or errors in the training data can lead to biased or inaccurate predictions.

**5.2.2 Explainability and Transparency:**

Machine learning models can be complex and difficult to interpret, making it challenging for healthcare providers to understand the rationale behind their predictions. This can raise concerns about transparency and explainability, affecting trust in the system's outputs.

**5.2.3 Overfitting and Generalizability:**

Overfitting occurs when a model performs well on the training data but fails to generalize to new, unseen data. This can lead to inaccurate predictions in real-world scenarios. Regular evaluation and retraining with new data are crucial to maintain generalizability.

**5.2.4 False Positives and False Negatives:**

No machine learning model is perfect, and there is always a risk of false positives (classifying normal patients as anemic) and false negatives (classifying anemic patients as normal). These errors can have significant consequences for patient care and require careful evaluation and management.

**5.2.5 Human Oversight and Clinical Judgment:**

Machine learning systems should not replace clinical judgment or decision-making. Healthcare providers should carefully review the system's outputs and exercise their expertise to make informed decisions based on the totality of patient information.

**5.3 APPLICATIONS**

**5.3.1 Screening and Early Detection of Anemia:**

The system can be used to screen large populations for anemia, particularly in high-risk groups such as pregnant women, children, and individuals with chronic conditions. Early detection allows for timely intervention and treatment to prevent complications and improve patient outcomes.

**5.3.2 Automated Diagnosis and Triage:**

The system can be integrated into electronic health records (EHRs) and laboratory information systems (LIS) to automatically analyze patient data and provide preliminary anemia diagnoses. This can expedite diagnosis, facilitate triage, and prioritize patients for further evaluation and treatment.

**5.3.3 Personalized Risk Assessment and Counseling:**

By analyzing patient-specific data, the system can identify individuals at higher risk of developing anemia. This information can be used to provide personalized counseling and preventive measures, such as dietary modifications or iron supplementation.

**5.3.4 Clinical Decision Support and Treatment Recommendations:**

The system can provide clinical decision support by suggesting appropriate diagnostic tests, treatment options, and follow-up strategies based on patient-specific data and machine learning-derived insights.

**CHAPTER 6**

**CONCLUSION AND FUTURE SCOPE**

**6.1 CONCLUSION**

Machine learning-driven anemia identification and classification systems have the potential to significantly improve healthcare outcomes by providing early, accurate, and personalized diagnosis of anemia. These systems offer several advantages over traditional methods, including improved accuracy, enhanced efficiency, reduced human error, personalized risk assessment, and integration with existing healthcare systems. However, it is crucial to recognize the limitations of these systems and implement appropriate safeguards to mitigate risks. Data reliance, explain ability, overfitting, false positives and negatives, human oversight, regulatory compliance, and continuous monitoring are key areas that require careful consideration and ongoing attention.

**6.2 FUTURE SCOPE**

Future research and development efforts should focus on the following areas to further enhance the effectiveness and applicability of machine learning-driven anemia identification and classification systems:

1. Data Quality and Diversity: Continuously improve the quality, quantity, and diversity of training data to enhance model generalizability and reduce biases.
2. Explainable AI: Develop and incorporate explainable AI techniques to provide insights into model decision-making, fostering trust and transparency among healthcare providers.
3. Adaptive Learning and Continuous Improvement: Implement adaptive learning mechanisms to enable models to continuously learn from new data and improve performance over time.
4. Handling Missing Data and Outliers: Develop robust methods for handling missing data and outliers to ensure model robustness and reliability in real-world scenarios.
5. Integration with Clinical Workflows: Seamlessly integrate the system into clinical workflows to facilitate real-time data exchange, decision support, and personalized care recommendations.

**APPENDIX A**

**1. FEASIBILITY ASSESSMENT:**

**1.1 Technical Feasibility:**

- Machine Learning Model:

- Feasibility: Developing a machine learning model for anemia detection is technically feasible. Existing frameworks such as Scikit-learn or TensorFlow provide robust tools for model development. Utilizing pre-trained models or transfer learning enhances the efficiency and accuracy of the model.

- Image Processing:

- Feasibility: Image processing techniques for feature extraction from microscopic blood smear images are well-established and technically feasible.

- Web Development:

- Feasibility: Web development using frameworks like Django or Flask is technically feasible, ensuring seamless integration with the machine learning model.

**1.2 Economic Feasibility**

-Infrastructure Costs:

- Feasibility: Costs associated with server hosting, storage, and computational resources are economically feasible, especially with the utilization of cost-effective cloud services.

- Development Costs:

- Feasibility: Open-source tools and frameworks reduce development costs, making the project economically viable for a small team or individual developers.

**2. Satisfiability Analysis**

**2.1 User Satisfaction**

-Doctors:

- Satisfiability: Regular feedback sessions and iterative improvements based on user suggestions will contribute to overall satisfaction.

- Patients:

- Satisfiability: Ensuring a user-friendly, secure, and privacy-respecting platform will enhance patient satisfaction.

**2.2 System Performance**

- Response Time:

- Satisfiability: Minimizing response time for image processing and classification is critical for user satisfaction and effective clinical decision-making.

- Reliability:

- Satisfiability: Ensuring reliability, minimal downtime, and constant system accessibility contribute to overall user satisfaction.

**3. Complexity Analysis**

**3.1 NP-Hard, NP-Complete, or P Type**

-Problem Complexity:

-Analysis: The problem of anemia detection using machine learning is classified as a P type problem. The training and prediction tasks are polynomial time solvable.

**3.2 Mathematical Models**

- Machine Learning Models:

- Models: Neural networks, decision trees, and support vector machines are effective mathematical models for classification problems like anemia detection.

- Algorithmic Complexity:

- Complexity: Algorithms used for image processing and feature extraction are chosen based on efficiency and suitability, emphasizing minimizing time complexity.

**APPENDIX B**

[1] P. T. Dalvi and N. Vernekar, "Anemia detection using ensemble learning techniques and statistical models," In *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT).*

It introduces five ensemble learning methods applied to four classifiers. The stacking method achieves the highest accuracy, outperforming Bagging, Voting, Adaboost, and Bayesian Boosting. Individual classifiers' performance is also assessed, with Artificial Neural Network (ANN) leading, and K-Nearest Neighbor (K-NN) performing the worst.

[2] S B. Sen, A. Ganesh, A. Bhan, S. Dixit and A. Goyal, "Machine learning based Diagnosis and Classification Of Sickle Cell Anemia in Human RBC," *in 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*.

The research in the paper addresses the challenges of manual inspection. The comparison of machine learning classifiers reveals that Naive Bayes has lower precision, recall, f-score, and accuracy compared to SVM, random forest, and logistic regression. SVM and logistic regression both have 90% accuracy, while random forest achieves 92%.

[3] Jaiswal, M., Srivastava, A., Siddiqui, T.J. (2019), “Machine Learning Algorithms for Anemia Disease Prediction”, In *Khare, A., Tiwary, U., Sethi, I., Singh, N. (eds) Recent Trends in Communication, Computing, and Electronics*. Lecture Notes in Electrical Engineering, vol 524. Springer.

In this paper, the study assesses Naive Bayes, random forest, and decision tree algorithms using CBC data. The results show Naive Bayes outperforming the others with an accuracy of 96.09%. The numerical values for random forest accuracy (95.32%) and decision tree accuracy (95.46%) provide a quantitative basis for comparison.

[4] An R, Huang Y, Man Y, Valentine RW, Kucukal E, Goreke U, Sekyonda Z, Piccone C, Owusu-Ansah A, Ahuja S, Little JA, Gurkan UA, “Emerging point-of-care technologies for anemia detection”, *Lab Chip. 2021 May 18*.

The research focuses on emerging point-of-care technologies for anaemia detection, emphasizing the need for accessible and cost-effective screening methods, especially in low- and middle-income countries.

[5] R. V. Pellegrino, A. C. Tarrobago and D. L. B. Zulueta, "Development of Anemia Cells Recognition System Using Raspberry Pi", In *2023 15th International Conference on Computer and Automation Engineering (ICCAE).*

The paper focuses on developing an Anaemia Cells Recognition System using Raspberry Pi. The system employs image processing and support vector machine (SVM) for the classification of abnormal red blood cells (RBCs). The automated recognition achieves an average accuracy of 94.31%, contributing to the advancement of anaemia and thalassemia detection.

[6] Shahzad, Muhammad, Arif Iqbal Umar, Syed Hamad Shirazi, Zakir Khan, Asfandyar Khan, Muhammad Assam, Abdullah Mohamed, and El-Awady Attia. 2022, "Identification of Anemia and Its Severity Level in a Peripheral Blood Smear Using 3-Tier Deep Neural Network", *Applied Sciences 12*, no. 10: 5030.

It introduces a 3-Tier Deep Neural Network for identifying anaemia and determining its severity level in peripheral blood smears. The proposed model achieves high accuracies in training (91.37%), validation (88.85%), and testing (86.06%), with recall, F1-Score, and specificity metrics demonstrating its effectiveness in anaemia severity prediction.

[7] Rizal, A.S. et al. (2022), “Detecting anemia based on palm images using convolutional neural network”, In *International Journal of Advanced Engineering Research and Science*, 9(9), pp. 280–287.

The paper explores the use of a Convolutional Neural Network (CNN) for classifying hemoglobin levels based on palm images. Achieving an accuracy of 96.43%, this study offers a non-invasive alternative to traditional blood tests, emphasizing the potential of CNN in anaemia detection.

[8] Zhang A, Lou J, Pan Z, Luo J, Zhang X, Zhang H, Li J, Wang L, Cui X, Ji B, Chen L, “Prediction of anemia using facial images and deep learning technology in the emergency department”, *Front Public Health.* 2022 Nov 9.

It presents a machine learning system utilizing deep learning technology to predict anaemia using facial images. With accuracies ranging from 82.37% to 74.01% for different anaemia severity levels, the system proves valuable in expediting diagnosis and aiding in urgent blood transfusion decisions in emergency departments.

[9] Nithya, R & Nirmala, K. (2022), “Detection of Anaemia using Image Processing Techniques from microscopy blood smear images”, *Journal of Physics: Conference Series.*

The research focuses on the detection of anaemia using image processing techniques from microscopy blood smear images. The proposed algorithm achieves an accuracy of 93.33%, with specificity and sensitivity values of 95.16% and 90.91%, respectively, showcasing its potential for automated RBC counting in anemic conditions.

[10] A. Kovačević, A. Lakota, L. Kuka, E. Bečić, A. Smajović and L. G. Pokvić, "Application of Artificial Intelligence in Diagnosis and Classification of Anemia," In *2022 11th Mediterranean Conference on Embedded Computing (MECO)*.

It introduces the application of artificial intelligence, specifically K-nearest neighbors (KNN), in the diagnosis and classification of anaemia. The study emphasizes the significance of monitoring parameters like age, sex, and various blood parameters in accurate anaemia diagnosis.

[11] S. C, A. M. R, D. M. D and D. M, "Curability Prediction Model for Anemia Using Machine Learning," In *2022 8th International Conference on Smart Structures and Systems (ICSSS)*.

It proposes a curability prediction model for anaemia using supervised machine learning techniques, including Naive Bayes, Logistic Regression, LASSO, and ES algorithms. The study aims to predict the cure or non-cure status of patients after 90 days, with Naive Bayes exhibiting superior accuracy.

[12] Mohammed A. Faris, Fadheelah S. Azeez, and Estabraq A. Hameed, “Estimation of Some Blood Parameters in Anemic Children Patients”, *Tikrit J. Pharm. Sci.*, vol. 16, no. 1, pp. 39–49, Dec. 2022.

The paper estimates blood parameters in anemic children patients, emphasizing gender and age as factors influencing anaemia. The study reveals deteriorated hematological variables in anemic children, with males and children under three years being more affected.

[13] N. B. Noor, M. S. Anwar and M. Dey, "Comparative Study Between Decision Tree, SVM and KNN to Predict Anaemic Condition", In *2019 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON)*.

It compares Decision Tree, SVM, and KNN for predicting anemic conditions. By analyzing conjunctiva color and blood hemoglobin levels, the study achieves an 82.61% accuracy in Decision Tree classification.

[14] M. S. MOHAMMED, A. A. AHMAD and M. SARI, "Analysis of Anemia Using Data Mining Techniques with Risk Factors Specification," In *2020 International Conference for Emerging Technology (INCET)*.

It employs data mining techniques to predict anaemia, comparing Bayesian Network, Naive Bayes, Logistic Regression, and Multilayer Perceptron. Logistic Regression and Multilayer Perceptron demonstrate high performance, with the study highlighting their efficiency in predicting anaemia.

[15] S. J. Mohammed MOHAMMED, A. A. Ahmed, A. A. Ahmad and M. Sami MOHAMMED, "Anemia Prediction Based on Rule Classification", In *2020 13th International Conference on Developments in eSystems Engineering (DeSE)*.

The paper focuses on anaemia prediction based on rule classification, utilizing techniques like ZeroR, OneR, and PART. The PART technique achieves the highest accuracy of 85%, showcasing its potential for creating accurate anaemia prediction systems.

[16] Parab, M.A., Mehendale, N.D, “Red Blood Cell Classification Using Image Processing and CNN” *SN COMPUT. SCI.* 2, 70 (2021).

It combines image processing and CNN to achieve an overall accuracy of 98.5%. The study focuses on classifying RBCs into nine types, showcasing the potential of the proposed method for accurate disease diagnosis.

[17] Dhakal, Prakriti & Khanal, Santosh & Bista, Rabindra. (2023), “Prediction of Anemia using Machine Learning Algorithms”, In *International Journal of Computer Science and Information Technology*.

The paper develops a model for predicting anaemia in children under 5 years of age using machine learning algorithms, with Random Forest proving to be the best performer with an accuracy of 98.4%.

[18] Alzubaidi, Laith, Mohammed A. Fadhel, Omran Al-Shamma, Jinglan Zhang, and Ye Duan. 2020, "Deep Learning Models for Classification of Red Blood Cells in Microscopy Images to Aid in Sickle Cell Anemia Diagnosis", *Electronics 9*, no. 3: 427.

It focused on the classification of red blood cells in microscopy images to aid in the diagnosis of sickle cell anaemia. The researchers proposed three CNN models with different layers and filters. They utilized image processing and machine learning techniques to classify, count, and segment sickle cell anaemia in erythrocytes.

[19] M. Hasan, M. S. Tahosin, A. Farjana, M. A. Sheakh and M. M. Hasan, "A Harmful Disorder: Predictive and Comparative Analysis for fetal Anemia Disease by Using Different Machine Learning Approaches", In *2023 11th International Symposium on Digital Forensics and Security (ISDFS)*.

It investigates machine learning models for predicting anaemia, comparing five different algorithms and highlighting the potential of machine learning in improving disease prevention and treatment.

[20] Justice Williams Asare, Peter Appiahene, Emmanuel Timmy Donkoh, et al, “Iron Deficiency Anemia Detection using Machine Learning Models: A Comparative Study of Fingernails, Palm and Conjunctiva of the Eye Images”, *Authorea*. February 06, 2023.

Reference [26] investigated the detection of iron deficiency anaemia using machine learning models. They compared the performance of machine learning algorithms in analyzing images of fingernails, palms, and conjunctiva of the eye for the detection of iron deficiency anaemia.

[27] Vohra R, Hussain A, Dudyala AK, Pahareeya J, Khan W (2022), “Multi-class classification algorithms for the diagnosis of anemia in an outpatient clinical setting”, In *PLoS ONE* 17(7): e0269685.

It employed various machine learning algorithms, including artificial neural networks, K-nearest neighbors, random forests, naive Bayes, and support vector machines, for the diagnosis of anaemia in an outpatient clinical setting. They compared the performance of these algorithms in classifying anaemia and malaria.

[22] Tuba Karagül Yıldız, Nilüfer Yurtay, Birgül Öneç, “Classifying Anemia Types Using Artificial Learning Methods”, In *Engineering Science and Technology*, an International Journal, Volume 24, Issue 1, 2021, Pages 50-70.

It classifies anaemia types using artificial learning methods, achieving high accuracy with Bagged Decision Trees, Boosted Trees, and Artificial Neural Networks.

[23] Sasikala, N. & Banu, Gulmohamed & Babiker, Thgani & Rajpoot, Pushp. (2021), “A Role of Data Mining Techniques to Predict Anemia Disease”, In *International Journal of Computer Applications*. 174.

The paper discusses the role of data mining techniques in predicting anaemia, using classification techniques like J48, Random Forest tree, and Hoeffding tree, with Random Forest tree exhibiting better accuracy.

[24] Chakraborty, S., Kansara, K., Dinesh Kumar, R. et al, “Non-invasive Estimation of Clinical Severity of Anemia Using Hierarchical Ensemble Classifiers”, In *J. Med. Biol. Eng.* 42, 828–838 (2022).

The research proposed a non-invasive method of anaemia classification using a multi-wavelength photometry sensor and a novel classification algorithm based on hierarchical ensemble classifiers. They leveraged recent advancements in machine learning to develop a non-invasive approach for estimating the clinical severity of anaemia.

**CHAPTER 7**

**REFERENCES**

1. Parab, M.A., Mehendale, N.D, “Red Blood Cell Classification Using Image Processing and CNN*” SN COMPUT. SCI. 2, 70 (2021).*
2. P. T. Dalvi and N. Vernekar, "Anemia detection using ensemble learning techniques and statistical models," *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT).*
3. S B. Sen, A. Ganesh, A. Bhan, S. Dixit and A. Goyal, "Machine learning based Diagnosis and Classification Of Sickle Cell Anemia in Human RBC," *in 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV).*
4. Jaiswal, M., Srivastava, A., Siddiqui, T.J. (2019), “Machine Learning Algorithms for Anemia Disease Prediction*”, In: Khare, A., Tiwary, U., Sethi, I., Singh, N. (eds) Recent Trends in Communication, Computing, and Electronics. Lecture Notes in Electrical Engineering, vol 524. Springer*.
5. An R, Huang Y, Man Y, Valentine RW, Kucukal E, Goreke U, Sekyonda Z, Piccone C, Owusu-Ansah A, Ahuja S, Little JA, Gurkan UA, “Emerging point-of-care technologies for anemia detection”, *Lab Chip. 2021 May 18.*
6. Diagnostic Approach To Anemia In Adults Using Machine Learning”, *Journal of Pharmaceutical Negative Results*, pp. 3713–3717, Dec. 2022, doi: 10.47750/pnr.2022.13.S09.456.
7. R. V. Pellegrino, A. C. Tarrobago and D. L. B. Zulueta, "Development of Anemia Cells Recognition System Using Raspberry Pi", *in* *2023 15th International Conference on Computer and Automation Engineering (ICCAE).*
8. Shahzad, Muhammad, Arif Iqbal Umar, Syed Hamad Shirazi, Zakir Khan, Asfandyar Khan, Muhammad Assam, Abdullah Mohamed, and El-Awady Attia. 2022. "Identification of Anemia and Its Severity Level in a Peripheral Blood Smear Using 3-Tier Deep Neural Network", *Applied Sciences 12, no. 10: 5030*.
9. Rizal, A.S. et al. (2022) “Detecting anemia based on palm images using convolutional neural network”, *International Journal of Advanced Engineering Research and Science, 9(9), pp. 280–287*.
10. Zhang A, Lou J, Pan Z, Luo J, Zhang X, Zhang H, Li J, Wang L, Cui X, Ji B, Chen L, “Prediction of anemia using facial images and deep learning technology in the emergency department”, *Front Public Health. 2022 Nov 9*.
11. Nithya, R & Nirmala, K. (2022), “Detection of Anaemia using Image Processing Techniques from microscopy blood smear images”, *Journal of Physics: Conference Series.*
12. A. Kovačević, A. Lakota, L. Kuka, E. Bečić, A. Smajović and L. G. Pokvić, "Application of Artificial Intelligence in Diagnosis and Classification of Anemia," *in 2022 11th Mediterranean Conference on Embedded Computing (MECO)*.
13. S. C, A. M. R, D. M. D and D. M, "Curability Prediction Model for Anemia Using Machine Learning," *in 2022 8th International Conference on Smart Structures and Systems (ICSSS)*.
14. Mohammed A. Faris, Fadheelah S. Azeez, and Estabraq A. Hameed, “Estimation of Some Blood Parameters in Anemic Children Patients”, *Tikrit J. Pharm. Sci., vol. 16, no. 1, pp. 39–49, Dec. 2022.*
15. N. B. Noor, M. S. Anwar and M. Dey, "Comparative Study Between Decision Tree, SVM and KNN to Predict Anaemic Condition", *in 2019 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON)*.
16. M. S. MOHAMMED, A. A. AHMAD and M. SARI, "Analysis of Anemia Using Data Mining Techniques with Risk Factors Specification," *in 2020 International Conference for Emerging Technology (INCET)*.
17. S. J. Mohammed MOHAMMED, A. A. Ahmed, A. A. Ahmad and M. Sami MOHAMMED, "Anemia Prediction Based on Rule Classification", *in* *2020 13th International Conference on Developments in eSystems Engineering (DeSE).*
18. Dhakal, Prakriti & Khanal, Santosh & Bista, Rabindra. (2023), “Prediction of Anemia using Machine Learning Algorithms”, *International Journal of Computer Science and Information Technology.*
19. Bouwman AC, Hulsegge I, Hawken RJ, Henshall JM, Veerkamp RF, Schokker D, Kamphuis C, “Classifying aneuploidy in genotype intensity data using deep learning”, *J Anim Breed Genet. 2023 May*.
20. M. Hasan, M. S. Tahosin, A. Farjana, M. A. Sheakh and M. M. Hasan, "A Harmful Disorder: Predictive and Comparative Analysis for fetal Anemia Disease by Using Different Machine Learning Approaches", *in* *2023 11th International Symposium on Digital Forensics and Security (ISDFS)*.
21. Saputra DCE, Sunat K, Ratnaningsih T, “A New Artificial Intelligence Approach Using Extreme Learning Machine as the Potentially Effective Model to Predict and Analyze the Diagnosis of Anemia”, *Healthcare (Basel). 2023 Feb 26*.
22. M. S. MOHAMMED, A. A. AHMAD and M. SARI, "Analysis of Anemia Using Data Mining Techniques with Risk Factors Specification”, *in* *2020 International Conference for Emerging Technology (INCET)*.
23. Tuba Karagül Yıldız, Nilüfer Yurtay, Birgül Öneç, “Classifying Anemia Types Using Artificial Learning Methods”, *Engineering Science and Technology, an International Journal, Volume 24, Issue 1, 2021, Pages 50-70*.
24. Sasikala, N. & Banu, Gulmohamed & Babiker, Thgani & Rajpoot, Pushp. (2021), “A Role of Data Mining Techniques to Predict Anemia Disease”, *International Journal of Computer Applications. 174.*
25. Chakraborty, S., Kansara, K., Dinesh Kumar, R. et al, “Non-invasive Estimation of Clinical Severity of Anemia Using Hierarchical Ensemble Classifiers”, *J. Med. Biol. Eng. 42, 828–838 (2022).*
26. Alzubaidi, Laith, Mohammed A. Fadhel, Omran Al-Shamma, Jinglan Zhang, and Ye Duan. 2020, "Deep Learning Models for Classification of Red Blood Cells in Microscopy Images to Aid in Sickle Cell Anemia Diagnosis", *Electronics 9, no. 3: 427.*
27. Justice Williams Asare, PETER APPIAHENE, Emmanuel Timmy Donkoh, et al, “Iron Deficiency Anemia Detection using Machine Learning Models: A Comparative Study of Fingernails, Palm and Conjunctiva of the Eye Images”, *Authorea. February 06, 2023.*
28. Vohra R, Hussain A, Dudyala AK, Pahareeya J, Khan W (2022), “Multi-class classification algorithms for the diagnosis of anemia in an outpatient clinical setting”, *PLoS ONE 17(7): e0269685.*
29. Memmolo, P. & Bianco, Vittorio & Russo, Roberta & Andolfo, Immacolata & Mugnano, Martina & Merola, Francesco & Miccio, Lisa & Iolascon, Achille & Ferraro, Pietro. (2023), “Hierarchical learning approaches for improving label-free single-cell classification in holographic microscopy”.
30. Lancet, “New study reveals global anemia cases remain persistently high among women and children. Anemia rates decline for men”. Available at: https://www.healthdata.org. Accessed on 13 September 2023.

[31] PIB, “ANAEMIA MUKT BHARAT”. Available at: https://pib.gov.in. Accessed on 12 September 2023.