

A PROJECT REPORT ON

**SICKLE CELL ANEMIA DETECTION USING
CONVOLUTIONAL NEURAL NETWORK**

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE

BACHELOR OF ENGINEERING

In

COMPUTER ENGINEERING

of

SAVITRIBAI PHULE PUNE UNIVERSITY

By

**SUJOT SASANE
SURYANSH SAXENA
OMKAR SURYWANSHI
ACHAL TAJNE**

**SEAT NO: B190234414
SEAT NO: B190234416
SEAT NO: B190234418
SEAT NO: B190234420**

Under the guidance of

DR. M. P. WANKHADE



**DEPARTMENT OF COMPUTER ENGINEERING
SINHGAD COLLEGE OF ENGINEERING, PUNE-411041**

Accredited by NAAC
2023-24

Sinhgad Technical Education Society,
Sinhgad College of Engineering , Pune-41
Department of Computer Engineering



Sinhgad Institutes

Date:

CERTIFICATE

This is to certify that the project report entitled

“Sickle Cell Anemia Detection Using Convolutional Neural Network”

Submitted by

Sujot Sasane

Seat no: B190234414

Suryansh Saxena

Seat no: B190234416

Omkar Surywanshi

Seat no: B190234418

Achal Tajne

Seat no: B190234420

is a bonafide work carried out by him under the supervision of **Dr. M. P. Wankhade** and it is approved for the partial fulfilment of the requirements of Savitribai Phule Pune University, Pune for the award of the degree of Bachelor of Engineering (Computer Engineering) during the year 2023-24.

Dr. M. P. Wankhade

Guide

Department of Computer Engineering

Dr. M. P. Wankhade

Head

Department of Computer Engineering

External Examiner

Dr. S. D. Lokhande

Principal

Sinhgad College of Engineering

ACKNOWLEDGEMENT

Words are often less to reveal once deep regards. With an understanding that worlds like this can never be the outcome of single person. We take this opportunity to express our profound sense of gratitude and respect to all those who directly or indirectly help us through the duration of this work.

It gives us great pleasure to present a dissertation on "Sickle Cell Anemia Detection Using Convolutional Neural Network". Therefore, It become our duty to express our gratitude towards them. We are extremely obliged to our project guide **Dr. M.P. Wankhade**, Head of Department of Computer Engineering for helping and providing us with proper guidance.

At last, we must express our sincere heartfelt gratitude to all the staff members of Computer Engineering Department who helped us directly or indirectly during this course of work .

Sujot Sasane

Suryansh Saxena

Omkar Surywanshi

Achal Tajne

ABSTRACT

Sickle Cell Anemia (SCA) is a prevalent genetic blood disorder characterized by abnormal haemoglobin causing red blood cells to take on a distinct sickle shape. Early and accurate detection of SCA is essential for effective treatment and management. This study presents an innovative approach to SCA detection utilizing Convolutional Neural Networks (CNNs), a class of deep learning algorithms known for their effectiveness in image analysis tasks. The proposed method involves the use of microscopic images of blood smears from patients. These images are pre processed to enhance contrast, normalize intensities, and remove noise. Subsequently, CNN architectures are employed to automatically extract hierarchical features from the blood smear images. The trained CNN model learns discriminative features directly from the images, effectively capturing subtle differences between normal and sickled red blood cells (RBCs). The advantages of using CNNs for SCA detection are manifold.

Firstly, CNNs automate feature extraction, reducing the potential for human error and improving the robustness of the detection process. Secondly, they have demonstrated high accuracy in various image classification tasks, effectively differentiating between normal and sickled RBCs, even in challenging cases with overlapping cells or poor image quality. Thirdly, CNN models are scalable and can process large volumes of data quickly and efficiently, benefiting screening programs in regions with high SCA prevalence. Additionally, CNN models can continuously improve with additional training data, enhancing their accuracy and generalization capabilities over time. Finally, given advancements in mobile computing, CNN-based detection systems can potentially be integrated with mobile devices, enabling point-of-care diagnostics in remote and resource-limited settings.

Keywords: Sickle Cell Anemia, convolutional neural network, deep learning, image analysis, blood smear images, early detection, diagnostic tool.

LIST OF FIGURES

| Figure No. | Title | Page No. |
|-------------------|---------------------------------|-----------------|
| 2.1 | Project estimates..... | 8 |
| 2.2 | Project schedule..... | 11 |
| 4.1 | System Architecture..... | 16 |
| 4.2 | Data Flow(0) diagram | 20 |
| 4.3 | Data Flow(1) diagram..... | 21 |
| 4.4 | Data Flow(2) diagram..... | 21 |
| 4.5 | Class diagram..... | 22 |
| 4.6 | Use case diagram..... | 23 |
| 4.7 | Activity diagram..... | 24 |
| 4.8 | Sequence diagram..... | 25 |
| 4.9 | CNN Layer..... | 26 |
| 4.10 | Watershed Algorithm..... | 27 |
| 4.11 | Watershed Algorithm Result..... | 28 |
| 5.1 | CNN algorithm..... | 30 |
| 7.1 | Starting page..... | 41 |
| 7.2 | Main Page..... | 42 |
| 7.3 | Registration page..... | 42 |
| 7.4 | Login page..... | 43 |
| 7.5 | Output 1 page..... | 44 |
| 7.6 | Output 2 page..... | 44 |

LIST OF TABLES

| Table No. | Title | Page No. |
|------------------|--|-----------------|
| 1.1 | Literature survey conclusion..... | 6 |
| 2.1 | Risk table..... | 9 |
| 2.2 | Risk probability definition..... | 9 |
| 2.3 | Risk impact definition..... | 10 |
| 2.4 | Overview of mitigation, monitoring & management..... | 10 |
| 5.1 | Confusion matrix..... | 30 |
| 6.1 | GUI Testing..... | 39 |
| 6.2 | Registration test case..... | 40 |
| 6.3 | Login test case..... | 40 |

CONTENTS

| | |
|-----------------------------|------------|
| Acknowledgement..... | i |
| Abstract..... | ii |
| List of figures..... | iii |
| List Of Tables..... | iv |

1 Introduction

| | |
|---|---|
| 1.1 Background and basics..... | 1 |
| 1.2 Motivation..... | 2 |
| 1.3 Problem Defination..... | 2 |
| 1.4 Literature Survey..... | 3 |
| 1.4.1 Machine Learning Based Diagnosis And Classification Of Sickle Cell Anaemia In Human RBC..... | 3 |
| 1.4.2 Automated Semantic Segmentation Of Red Blood Cells For Sickle Cell Disease..... | 3 |
| 1.4.3 A Review Of Automated Methods For The Detection Of Sickle Cell Disease..... | 4 |
| 1.4.4 Automatic Sickle Cell Anemia Detection Using Image Processing Technique..... | 5 |
| 1.4.5 Literature Survey Conclusion | 6 |

2 PROJECT PLANNING AND MANAGEMENT

| | |
|---|----|
| 2.1 Project estimates..... | 7 |
| 2.1.1 Project resource..... | 8 |
| 2.2 Risk management with respect to NP hard analysis..... | 8 |
| 2.2.1 Risk identification..... | 9 |
| 2.2.2 Risk analysis..... | 9 |
| 2.2.3 Overview of risk mitigation, monitoring and management..... | 10 |
| 2.3 Project schedule..... | 11 |
| 2.4 Team organization..... | 11 |
| 2.4.1 Team structure..... | 11 |
| 2.4.2 Management reporting and communication..... | 11 |

3 SOFTWARE REQUIREMENT SPECIFICATION

| | |
|--|----|
| 3.1 Functional Requirements..... | 12 |
| 3.1.1 System feature 1..... | 12 |
| 3.1.2 System feature 2..... | 12 |
| 3.2 External Interface Requirement..... | 12 |
| 3.2.1 User Interface..... | 12 |
| 3.2.2 Hardware & software Interfaces..... | 12 |
| 3.3 Non Functional Requirement..... | 13 |
| 3.3.1 Performance requirement..... | 13 |
| 3.3.2 Safety requirement..... | 13 |
| 3.3.3 Software quality attributes..... | 13 |
| 3.3.4 Dataset requirements..... | 14 |
| 3.4 User Classes and Characteristics..... | 14 |
| 3.5 Analysis Model SDLC Model to be Applied..... | 14 |

4 ANALYSIS AND DESIGN

| | |
|---|----|
| 4.1 System Architecture..... | 16 |
| 4.1.1 Elaborated steps of implementation..... | 17 |
| 4.1.2 Module..... | 20 |
| 4.1.3 Data Flow Diagram..... | 20 |
| 4.2 UML Diagrams..... | 22 |
| 4.3 Convolutional Neural Network..... | 26 |
| 4.4 Watershed Algorithm..... | 27 |

5 IMPLEMENTATION AND CODING

| | |
|--------------------------------------|----|
| 5.1 Overview of project module..... | 29 |
| 5.2 Tools and technologies used..... | 29 |

| | |
|---------------------------|----|
| 5.3 Algorithm detail..... | 29 |
| 5.4 Confusion matrix..... | 30 |
| 5.5 Code listing..... | 32 |

6 SOFTWARE TESTING

| | |
|----------------------------------|----|
| 6.1 Introduction..... | 37 |
| 6.2 Test of testing used..... | 37 |
| 6.2.1 Unit testing..... | 37 |
| 6.2.2 Integration testing..... | 37 |
| 6.2.3 System testing..... | 38 |
| 6.3 White box testing..... | 38 |
| 6.4 Black box testing..... | 38 |
| 6.4.1 Test case test result..... | 39 |

7 RESULT

| | |
|-----------------------|----|
| 7.1 Outcome..... | 41 |
| 7.2 Screenshoots..... | 41 |
| 7.3 Output..... | 44 |

8 CONCLUSION

| | |
|---------------------|----|
| 8.1 Conclusion..... | 45 |
|---------------------|----|

9 FUTURE WORK

| | |
|-----------------------|----|
| 9.1 Applications..... | 46 |
|-----------------------|----|

REFERENCE

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND AND BASICS

Sickle Cell Anemia (SCA) stands as a poignant example of the intricate interplay between genetics and health, impacting millions worldwide. This genetic anomaly alters the usual functioning of haemoglobin, the vital molecule responsible for ferrying oxygen within red blood cells. The resulting misshapen form of these cells manifests as the classic 'sickle' shape, setting the stage for a cascade of complications that can severely affect an individual's quality of life. From chronic pain to compromised organ function, the repercussions of SCA are far-reaching, necessitating a precise and efficient means of diagnosis for timely intervention and management.

Traditionally, the identification of sickled red blood cells (RBCs) relied heavily on manual inspection under a microscope, a method fraught with limitations. The inherent subjectivity and time-consuming nature of this process underscore the urgent need for innovative, technology-driven solutions. Convolutional Neural Networks (CNNs), a class of advanced machine learning models uniquely suited to process and interpret visual data. Their exceptional capacity to discern intricate details within images has been harnessed in diverse domains, revolutionizing tasks ranging from image recognition to medical diagnostics.

CNNs operate by systematically extracting and analysing hierarchical features from images, enabling them to discern complex patterns that might elude human observation. In the context of SCA diagnosis, the application of CNNs holds immense promise. By training these networks on vast datasets comprising diverse RBC images, they can learn to discern the subtle variations indicative of sickle cell morphology. This automated approach not only streamlines the diagnostic process but also minimizes the potential for human error and inter-observer variability.

The integration of CNNs into SCA detection signifies a paradigm shift, potentially revolutionizing how this condition is diagnosed and monitored. By leveraging the prowess of machine learning, this technology offers the tantalizing prospect of quicker, more accurate diagnoses, enabling prompt interventions and personalized treatment strategies. Moreover, the scalability and accessibility of CNN-based diagnostic tools hold the potential to bridge gaps in healthcare, particularly in regions where resources are scarce.

1.2 MOTIVATION

The exploration of Convolutional Neural Networks (CNNs) for Sickle Cell Anemia (SCA) detection represents a groundbreaking avenue in medical diagnostics, promising profound implications for healthcare. This proposed research endeavours to revolutionize the current detection methods, potentially offering a non-invasive and more accurate approach. By harnessing the power of CNNs, a deep learning technique adept at pattern recognition, this study aims to enhance the early identification and classification of SCA markers from medical imaging data. Such advancements could significantly expedite diagnosis, enabling timely interventions and personalized treatment strategies. Moreover, the potential integration of CNNs in diagnostic tools could streamline the process, making it more accessible, cost-effective, and scalable, thus addressing critical gaps in healthcare systems worldwide. Consequently, this research not only holds immense promise for advancing SCA detection but also heralds a new era in precision medicine and patient care.

1.3 PROBLEM DEFINATION

This project utilizes convolutional neural network(CNNs) to revolutionize sickle cell anemia (SCA) detection, offering a more accurate and non- invasive approach. By leveraging CNNs' pattern recognition capabilities, we aim to enhance early identification and classification.

1.4 LITERATURE SURVEY

1.4.1 Paper Name: B. Sen, A. Ganesh, A. Bhan, S. Dixit and A. Goyal, "Machine learning based Diagnosis and Classification of Sickle Cell Anemia in Human RBC," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2021, pp. 753-758.

Abstract:

In this research image processing and machine learning techniques is used to automate the process of detection of sickle cells in microscopic images then classify the RBC into three shapes: circular, elongated (sickle cell) and other shape. The microscopic image is pre-processed and Otsu thresholding technique is used for segmentation.

Then, Watershed segmentation is applied to separate the overlapped cells. The geometrical, statistical and textural features are extracted from the images. The machine learning classifier random forest, logistic regression naïve baye sand support vector machine is used. This research describes the comparison among these algorithms.

Techniques used in this paper are very exclusive for the use into our project and very much align according to our requirements, but lacks accuracy which is maintained by the use of CNN altogether with these methods discussed over here.

1.4.2 Paper Name: M. Zhang, X. Li, M. Xu and Q. Li, "Automated Semantic Segmentation of Red Blood Cells for Sickle Cell Disease," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 11, pp. 3095-3102, Nov. 2020.

Abstract:

In this paper, method adopted is a deep learning based semantic segmentation framework to solve the RBC classification task. A major challenge for robust segmentation and classification is the large variations on the size, shape and viewpoint of the cells, combining with the low image quality caused by noise and artifacts. To address these challenges, technique used is deformable convolution layers to the classic U-Net structure and implement the deformable U-Net (dU-Net). In du-Net technique a combination of offset field and deformed feature map is used, this feature map uses an extra Tan H function in the kernel to perform extraction.

This segmentation method gives an approach for classification of cells using segmentation but it only is insufficient as discrepancies occur due to overlapping of blood cells in pictures of smears, so using segmentation along with some other integrated algorithms this problem can be solved and result can become more precise.

1.4.3 Paper Name: P. K. Das, S. Meher, R. Panda and A. Abraham, "A Review of Automated Methods for the Detection of Sickle Cell Disease," in IEEE Reviews in Biomedical Engineering, vol. 13, pp. 309-324, 2020.

Abstract :

This review emphasizes the state-of-the-art methods and recent advances in detection, segmentation, and classification of sickle cell disease. In here discussed the key challenges encountered during the segmentation of overlapping blood cells. Moreover, standard validation measures that have been employed to yield performance analysis of various methods are also discussed. The content, in terms of methodologies and experiments, of this review paper is useful for people working in this area.

Various segmentation methods which can be used in this project are discussed in this paper. Methods like Region based Segmentation, Threshold Based Segmentation, Clustering Based Segmentation are discussed here which can be used based on explicit study of them. Feature extraction is also explained very extensive in this paper which is very useful technique.

1.4.4 Paper Name: T. S. Chy and M. A. Rahaman, "Automatic Sickle Cell Anemia Detection Using Image Processing Technique," 2018 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE), Gazipur, Bangladesh, 2018, pp. 1-4.

Abstract:

In this paper a new way is introduced to detect and classify sickle cells in RBC using image processing technique to start treatment as early as possible. Firstly, this method collects images of blood. The pre-processing phase is done through gray scale image conversion, image enhancement & median filter. Then, the threshold segmentation is applied to segment the RBCs and morphological operations are used to remove the undesired objects from images. Metric value, aspect ratio, entropy, mean, standard deviation and variance are used as features which are extracted. Finally, the support vector machine classifier is trained to test the images.

Gray scale Image Conversion, Image Enhancement and Noise Filtering techniques which are used in this paper are very useful in this project along with image processing by CNN, in fact mathematical calculation done in this paper helps us give more accuracy to our project.

1.4.5 Literature Survey Conclusion

Table 1.1 Literature Survey Conclusion

| Author(s) | Title | Year | Key Findings |
|---|--|------|---|
| Bheema Sen, Adarsh Ganesh, Anupama Bhan, Shubhra Dixit, Ayush Goyal | Machine Learning Based Diagnosis And Classification Of Sickle Cell Anemia In Human RBC | 2021 | Machine learning based algorithms are used here which lacks accuracy, no CNN is used which can provide more precision in the project. |
| Mo Zhang*, Xiang Li*, Mengjia Xu*, Quanzheng Li. | Automated Semantic Segmentation of Red Blood Cells for Sickle Cell Disease | 2020 | Classification and segmentation techniques are stated in this paper for feature extraction and preprocessing but no further work is done to completely built a project. |
| Pradeep Kumar Das, Rutuparna Panda, Ajith Abraham | A Review Of Automated Methods For The Detection Of Sickle Cell Disease | 2019 | Various segmentation methods which can be used in this project are discussed in this paper. But no discussion of further working algorithms is done.. |
| Tajkia Saima Chy, Mohammad Anisur Rahaman | Automatic Sickle Cell Anemia Detection Using Image Preprocessing Technique | 2018 | Image processing integrated with classification and segmentation is discussed in this paper but not a proper working architecture is made using them, which gives only theoretical idea of project working. |

CHAPTER 2

PROJECT PLANNING AND MANAGEMENT

2.1 PROJECT ESTIMATES

We are using waterfall model for our project estimation.

1. Requirement gathering and analysis: In this step we identified the requirements for the project and arranged them. Requirements were like Spyder IDE downloading installation, various libraries downloading like TensorFlow, NumPy etc, production environment setup, database finalizations and other requirements as well.
2. System Design: In this phase we designed a user friendly front-end interface, containing starting page, then registration, then login and finally main workspace window. Flow of code is also designed here.
3. Implementation: In this phase of implementation we connected all the CNN layers for recurrent working with other used algorithms. And this backend working is intersected with UI for input and result output.
- 4 .Testing: Various test cases are performed on the system to ensure rigorous working. Test cases were run on registration, login and then on main work desk, and expected discrepancies were noted and corrected.
5. Deployment of System: Once the functional and non-functional testing is done, the final software is ready to be used by the end user.
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.

All these phases are cascaded to each other in which progress is seen as flowing steadily downwards like a waterfall through the phases. The next phase is started only after the defined set of goals are achieved for previous phase and it is signed off, so the name "Waterfall Model". In this model phases do not overlap.

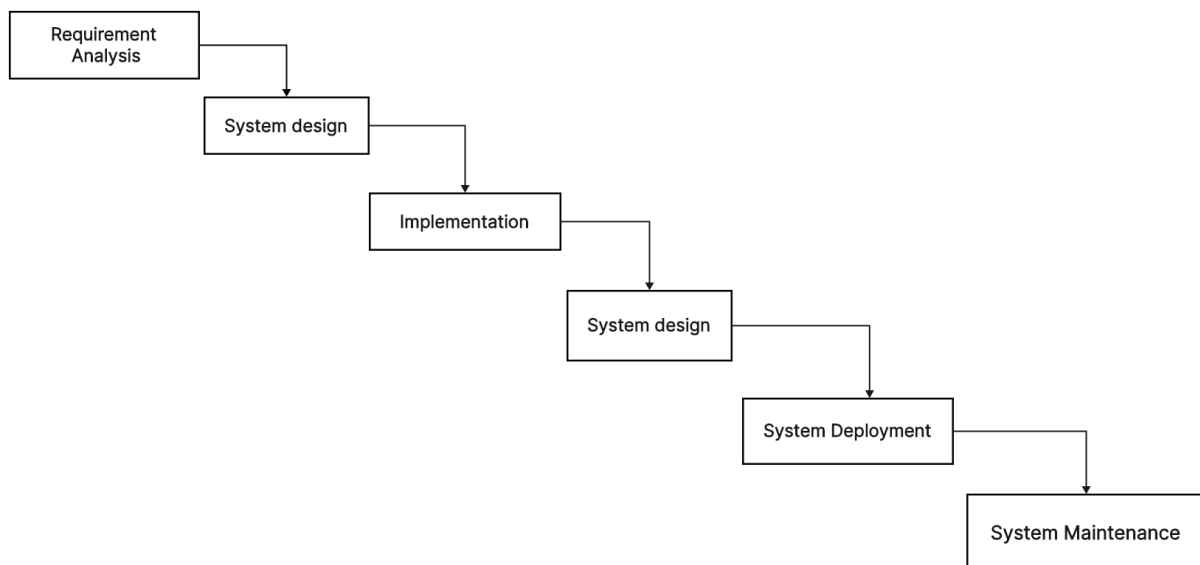


Fig no.2.1 Estimation step flowchart

2.1.1 Project Resources

Well configured Laptop, eclipse IDE, 2 GHZ CPU speed, 8 GB RAM, Internet connection.

2.2 Risk management wrt NP hard analysis

1. In appropriate dataset -To overcome this risk we are trying to use well organized and complete dataset.

2.Security- To overcome and improving security we use multilevel security like access permissions of user.

2.2.1 Risk Identification

Inaccurate shot detection: The system may encounter challenges in accurately identifying and classifying different types of shots, leading to incorrect scoring and analysis. This can occur due to variations in player techniques, occlusions, poor lighting conditions, or noise in the video feed.

False positives or false negatives: The shot detection system may generate false positives, identifying shots where there was none, or false negatives, missing actual shots played by the batsmen. This can occur due to algorithmic limitations, technical issues, or inadequate training data.

2.2.2 Risk Analysis

The risks for the Project can be analysed within the constraints of time and quality.

Table 2.1: Risk Table

| ID | Risk Description | Probability | Impact | | |
|----|------------------|-------------|----------|---------|---------|
| | | | Schedule | Quality | Overall |
| 1 | Description 1 | Low | Low | High | High |
| 2 | Description 2 | Low | Low | High | High |

Table 2.2: Risk Probability Definition

| Probability | Value | Description |
|-------------|------------------------------|-------------|
| High | Probability of occurrence is | >75% |
| Medium | Probability of occurrence is | 26 – 75% |
| Low | Probability of occurrence is | <25% |

Table 2.3 : Risk Impact definitions

| Impact | Value | Description |
|-----------|-------|--|
| Very high | >10% | Schedule impact or unacceptable quality |
| High | 5-10% | Schedule impact or some parts of the project have low quality |
| Medium | <5% | Schedule impact or Barely noticeable degradation in quality Low impact on schedule or Quality can be incorporated. |

2.2.3 Overview of Risk Mitigation, Monitoring, Management

Table 2.4 Overview of mitigation, monitoring and management

| | |
|------------------|--|
| Risk Id | 1 |
| Risk Description | Description 1 |
| Category | Development Environment |
| Source | Software requirements specification document |
| Probability | Low |
| Impact | High |
| Response | Mitigate |
| Strategy | Strategy |
| Risk Status | Occurred |

| | |
|------------------|--|
| Risk Id | 2 |
| Risk Description | Description 2 |
| Category | Requirements |
| Source | Software design specification documentation review |
| Probability | Low |
| Impact | High |
| Response | Mitigate |
| Strategy | Better testing will resolve this issue |
| Risk Status | Identified |

2.3 PROJECT SCHEDULE

| Months | Jun | July | Aug | Sept | Oct | Nov | Dec | Jan | Feb | Mar |
|-----------------------|-----|------|-----|------|-----|-----|-----|-----|-----|-----|
| Finalization of topic | | | | | | | | | | |
| Literature Survey | | | | | | | | | | |
| System Analysis | | | | | | | | | | |
| Mid Sem report | | | | | | | | | | |
| Design, Planning | | | | | | | | | | |
| Implementation | | | | | | | | | | |
| Testing | | | | | | | | | | |
| Final report | | | | | | | | | | |

Figure2.6 Project Schedule

2.4 TEAM ORGANIZATION

Team consists of 3 members and proper planning mechanism are used and roles of each member are defined.

2.4.1 Team structure

The team structure for the project is identified. There are total 3 members in our team and roles are defined. All members are contributing in all the phases of project.

2.4.2 Management reporting and communication

Well planning mechanisms are used for progress reporting and inter/intra team communication are identified as per requirements of the project.

CHAPTER 3

SOFTWARE REQUIREMENT SPECIFICATION

3.1 FUNCTIONAL REQUIREMENTS

3.1.1 System Feature1(Functional Requirement)

Admin: Admin module will first open on web browser. It will provide registration and login option and will perform defined validations for which ever option is selected.

User: User registers into system with personal information. As registrations completes successfully user will be redirected for login, after entering correct login details user can enter the system.

3.1.2 System Feature2 (Functional Requirement)

System: On main work desk, user can select the image to be tested, perform processing and can get the result. Here main system working of detection is performed.

3.2 EXTERNAL INTERFACE REQUIREMENT

3.2.1 User Interface

- Application of Sickle cell anemia detection.

3.2.2 Hardware and Software Interfaces:

Processor: Intel i5 Processor

RAM: 8GB

As we are using Machine Learning Algorithm and Various High Level Libraries Laptop

RAM minimum required is 8 GB.

Hard Disk: 40 GB

Spyder IDE that Integrated Development Environment is to be used and data loading should be fast hence Fast Processor is required

IDE: Spyder

Best Integrated Development Environment as it gives possible suggestions at the time of typing code snippets that makes typing feasible and fast.

Coding Language: Python Version 3.5

Highly specified Programming Language for Machine Learning because of availability of High-Performance Libraries.

Operating System: Windows 10

Latest Operating System that supports all type of installation and development Environment.

3.3 NON-FUNCTIONAL REQUIREMENT

3.3.1 Performance Requirements

The performance of the functions and every module must be well. The overall performance of the software will enable the users to work decently. Performance of encryption of data should be fast. Performance of the providing virtual environment should be fast Safety Requirement The application is designed in modules where errors can be detected and steadily.

3.3.2 Safety Requirement

The application is designed in modules where errors can be detected and fixed easily. This makes it easier to install and update new functionality if required.

3.3.3 Software Quality Attributes

Our software has many quality attribute that are given below: Adaptability: This software is adaptable by all users.

- Availability: This software is freely available to all users. The availability of the software is easy for everyone.
- Maintainability: After the deployment of the project if any error occurs then it can be easily maintained by the software developer.

- **Reliability:** The performance of the software is better which will increase there liability of the Software.
- **User Friendliness:** Since, the software is a GUI application; the output generated is much user friendly in its behaviour.
- **Integrity:** Integrity refers to the extent to which access to software or data by unauthorized persons can be controlled
- **Security:** Users are authenticated using many security phases so reliable security is provided.
- **Testability:** The software will be tested considering all the aspects.

3.3.4 Dataset Requirements

- **Dataset:** Magnified image dataset of blood samples are required.
- **Gathered dataset** from various sources mainly from government medical database.
- **Size and number of images:** There are 2000+ images ranging from 100Kb to 400Kb.
- **Images classes** varies among JPG, JPEG and PNG.

3.4 User Classes and Characteristics

User classes and characteristics are key components of system requirements analysis and design. They help define the different types of users who will interact with a system, as well as the specific attributes or behaviours associated with each user class .Interact with the system to perform specific tasks or access information. May have varying levels of expertise.

3.5 ANALYSIS MODEL: SDLC MODEL TO BE APPLIED

The software development cycle is a combination of different phases such as designing, implementing and deploying the project. These different phases of the software development model are described in this section. The SDLC model for the project development can be understood using the following figure The chosen SDLC model is the waterfall model which is easy to follow and fits bests for the implementation of this project.

Requirements Analysis: At this stage, the business requirements, definitions of use cases are studied and respective documentations are generated. **Design:** In this stage, the designs of the data models will be defined and different data preparation and analysis will be carried out

- **Implementation:** The actual development of the model will be carried out in this stage. Based on the data model designs and requirements from previous stages, appropriate algorithms, mathematical models and design patterns will be used to develop the agent's back-end and front-end components.
- **Testing:** The developed model based on the previous stages will be tested in this stage. Various validation tests will be carried out over the trained model. **Deployment:** After the model is validated for its accuracy scores its ready to be deployed or used in simulated scenarios.
- **Maintenance:** During the use of the developed solution various inputs/scenarios will be countered by the model which might affect the models overall accuracy. Or with passing time the model might not fit the new business requirements. Thus, the model must be maintained often to keep its desired state of operation.

Chapter 4

ANALYSIS AND DESIGN

4.1 SYSTEM ARCHITECTURE

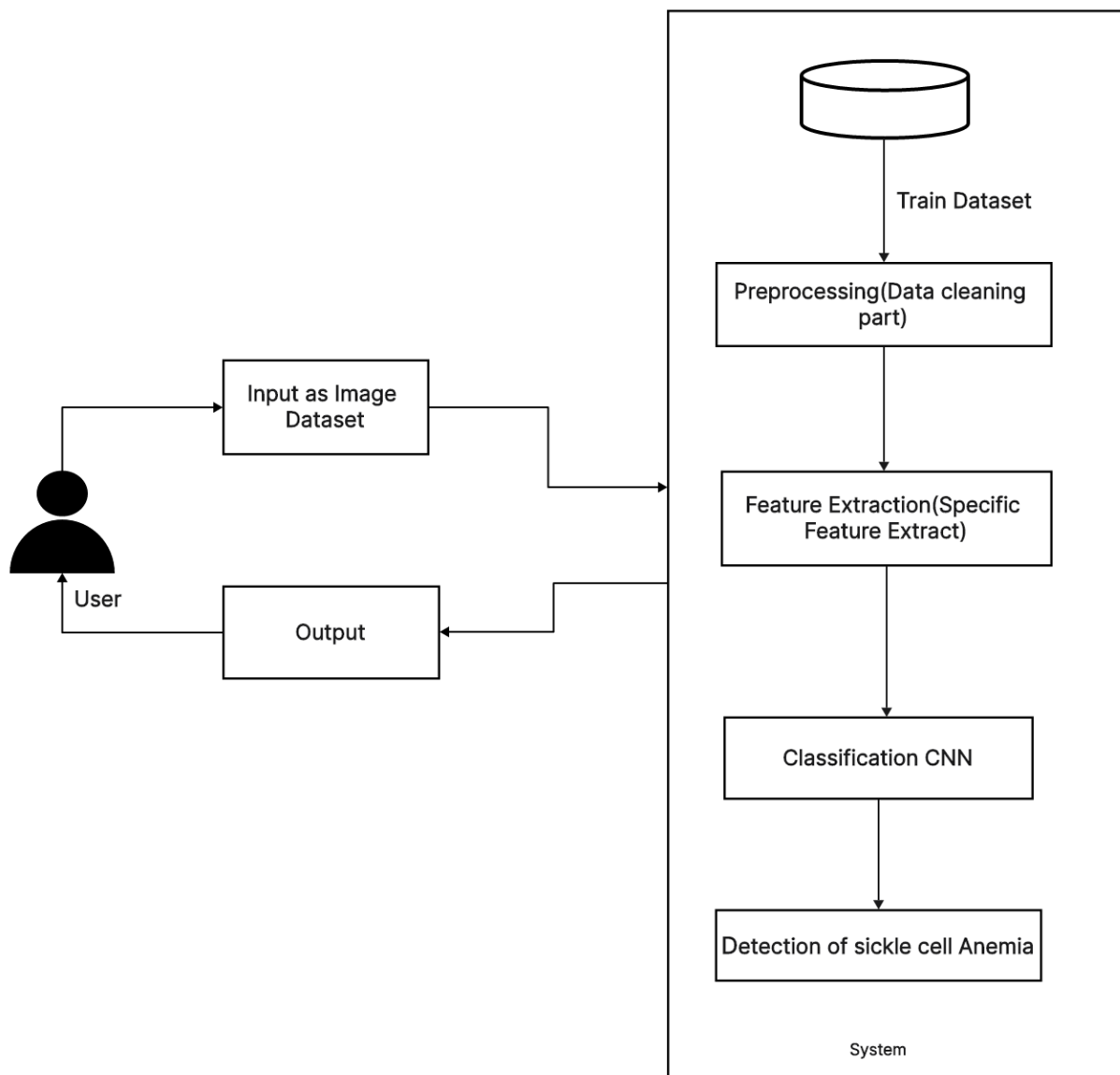


Figure 4.1 System Architecture

4.1.1 Elaborated steps of implementation

1. Preprocessing

a. Normalization

Normalization involves adjusting the intensity values of the images to a common scale, which helps in reducing the variation between images. This step ensures that each pixel value is within a specific range (typically 0 to 1 or -1 to 1), which facilitates better convergence during the training of the CNN.

b. Segmentation

Segmentation is the process of partitioning the blood smear images into meaningful regions, primarily separating the red blood cells (RBCs) from the background and other components like lymphocytes and neutrophils. This step is crucial for isolating the cells of interest for further analysis.

c. Kernel Sizing

Kernel sizing involves determining the size of the filter (kernel) that will be used in convolution operations. The size of the kernel affects the amount of detail captured in the feature maps. Smaller kernels capture finer details, while larger kernels capture more contextual information.

d. Padding and Pooling

Padding adds extra pixels around the borders of the image to ensure that the convolutional layers can process edge pixels as effectively as central pixels. Pooling, such as max pooling or average pooling, reduces the spatial dimensions of the feature maps, helping to control overfitting and reduce computational complexity by down sampling the input representation.

2. Feature Extraction

a. Cell Identification:

This step involves detecting and identifying individual cells within the blood smear images. Techniques such as thresholding and contour detection are often used to locate the cells accurately.

b. Color of Cell:

The color analysis of the cells is performed to identify characteristics specific to sickle cells. Sickle cells may have different staining patterns compared to normal cells, and color features can aid in distinguishing them.

c. Boundary of Cell:

Determining the boundary of each cell helps in accurately defining its shape. Edge detection algorithms such as the Sobel or Canny edge detectors can be used to outline the cells.

d. Outlines of Cell:

Outlining the cells involves tracing the perimeter of the cell boundaries. This step is important for further shape analysis and morphological assessments.

e. Background Removing:

Removing the background noise is crucial for focusing on the cells themselves. Techniques like morphological operations and background subtraction can be used to eliminate irrelevant parts of the image.

f. Shape of Cell:

Shape analysis involves extracting features related to the geometric properties of the cells. Sickle cells have a distinct crescent shape, which can be quantified and used for classification purposes.

3. Convolutional Neural Network (CNN)**a. Image Selection:**

Selecting a diverse and representative set of images from the dataset to ensure that the model is trained on varied examples. This step involves choosing images that include normal cells, sickle cells, lymphocytes, and neutrophils.

b. Image Validation:

Validating the selected images to ensure they are correctly labeled and suitable for training. This step may involve manual inspection or automated checks for quality assurance.

c. Image Transformation:

Applying transformations such as rotation, scaling, and flipping to augment the dataset and improve the robustness of the CNN. These transformations help the model generalize better by exposing it to variations of the same images.

d. Image Feeding into Algorithm:

Feeding the pre processed and transformed images into the CNN algorithm for training. This involves passing the images through multiple convolutional, pooling, and fully connected layers.

e. Image Examination:

During the training process, the CNN examines the images and learns to extract relevant features through backpropagation and optimization techniques. This step is iterative, with the model continually adjusting its weights to minimize the loss function.

f. Result:

After training, the CNN produces results that include class probabilities for each input image, indicating the likelihood of the image belonging to each class (normal, sickle cell, lymphocyte, neutrophil).

4. Detection

a. result from CNN

The output from the CNN is analysed to determine the classification results. Each image is assigned a class label based on the highest probability score from the CNN's output layer.

b. positive/negative showcasing

The final step involves showcasing the detection results, highlighting images that are classified as positive for sickle cell anemia. This step may include generating a report or visual display indicating which images show normal cells and which show sickle cells, along with any relevant metrics (e.g., confidence scores).

4.1.2 Module

- Admin
 - In this module ,the Admin has to log in by using valid user name and password.
After login successful he can do some operations such as View All Users
- View and Authorize Users
 - In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.
- View Charts Results
- View All Products Search Ratio, View All Keyword Search Results, View All Product Review Rank Results.
- End User
 - In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored in the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like Manage Account,

4.1.3 Data Flow Diagram

In Data Flow Diagram, It Shows that flow of data in the system in DFD0 that base DFD in which rectangle present input as well as output and circle shows the system, In DFD1 it shows actual input and actual output of system input of system is text or image and output is rumour detected likewise in DFD2 represent operation of user as well as admin.

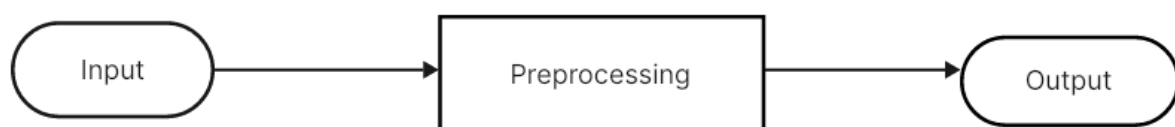


Figure 4.2: Data Flow(0) diagram

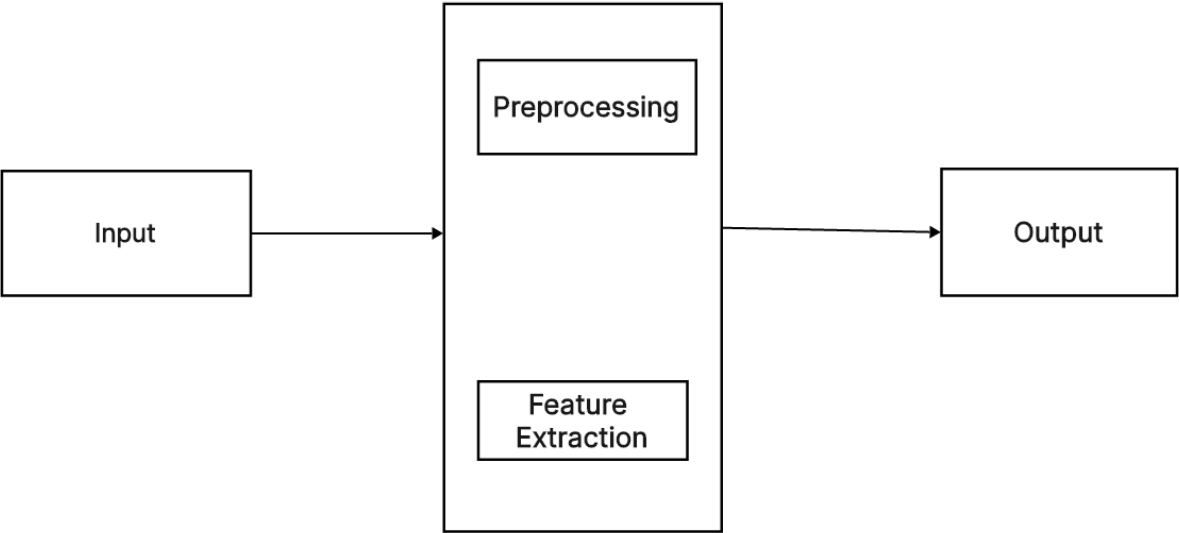


Figure 4.3: Data flow(1) diagram

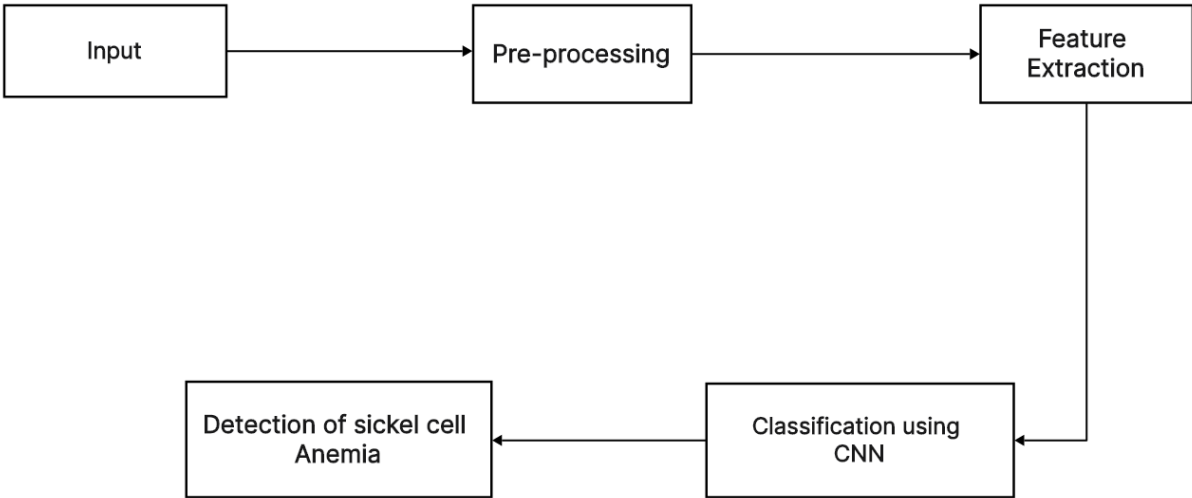


Figure 4.4: Data Flow(2) diagram

4.2 UML DIAGRAMS

Unified Modeling Language is a standard language for writing software blueprints. The UML may be used to visualize, specify, construct and document the artifacts of a software intensive system. UML is process independent, although optimally it should be used in process that is use case driven, architecture-centric, iterative, and incremental. The Number of UML Diagram is available.

- Class Diagram.
- Use case Diagram.
- Activity Diagram.
- Sequence Diagram.

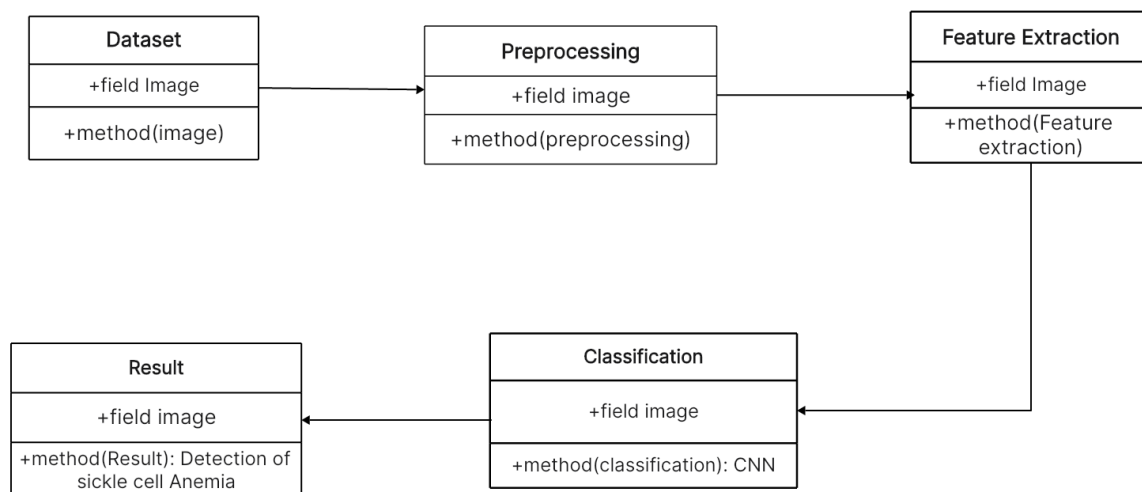


Figure 4.5: Class Diagram

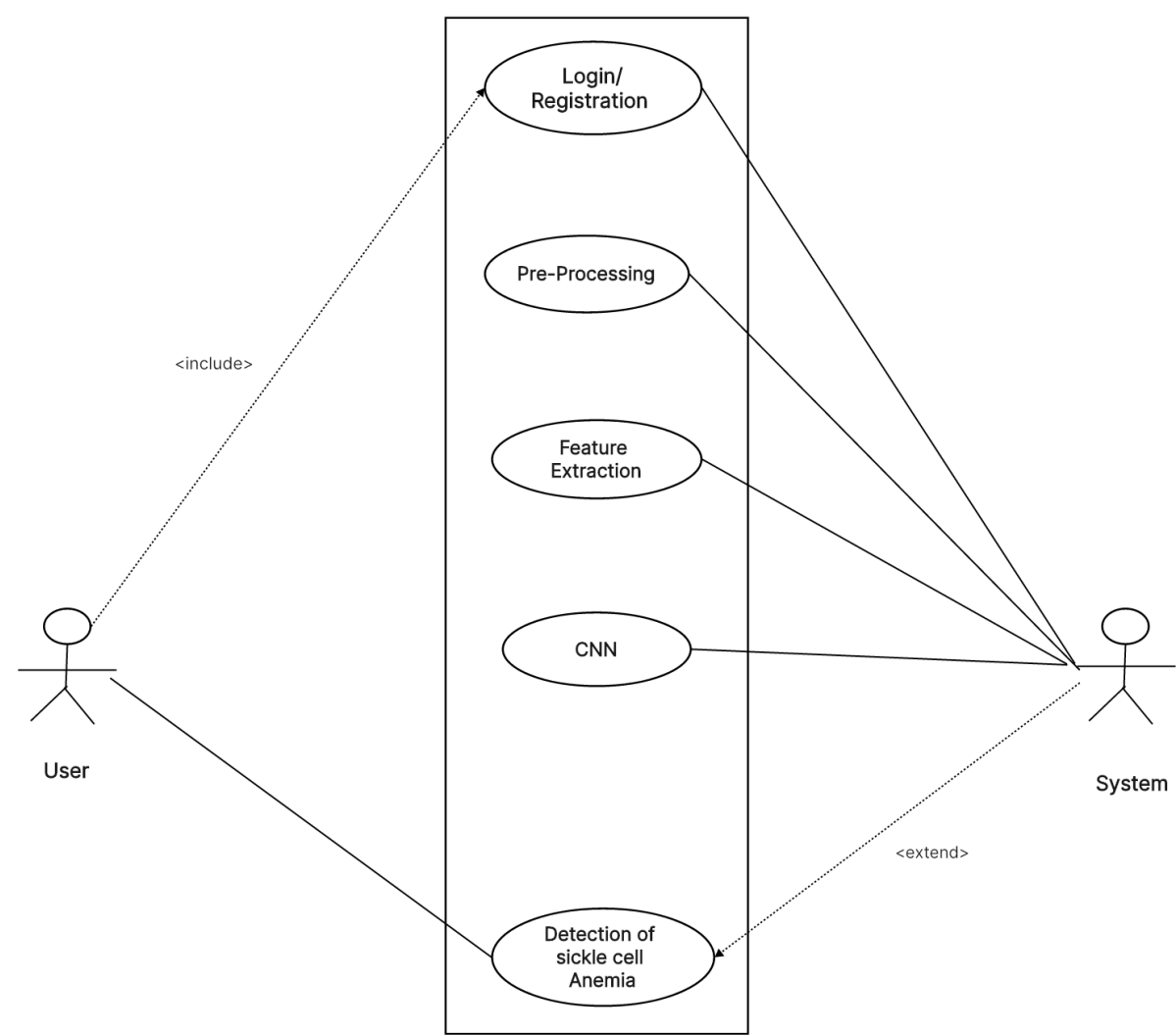


Figure 4.6: Use case Diagram

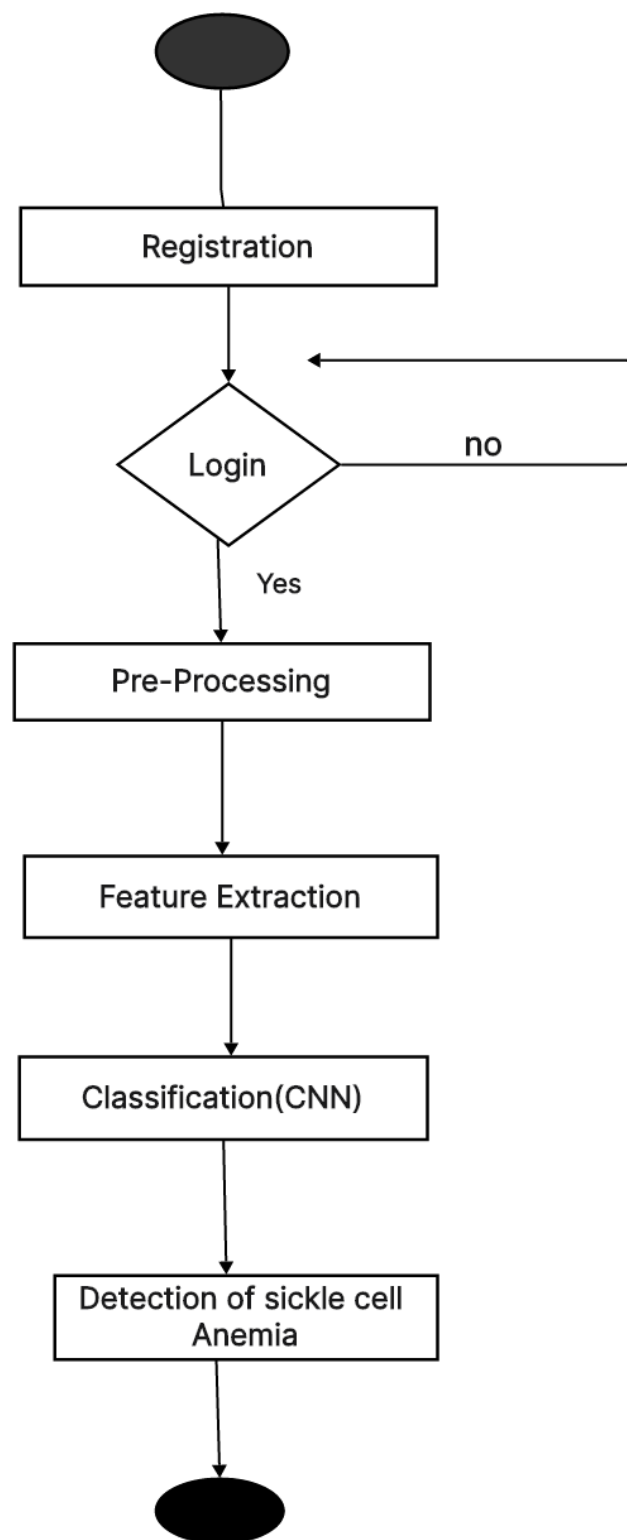


Figure 4.7: Activity Diagram

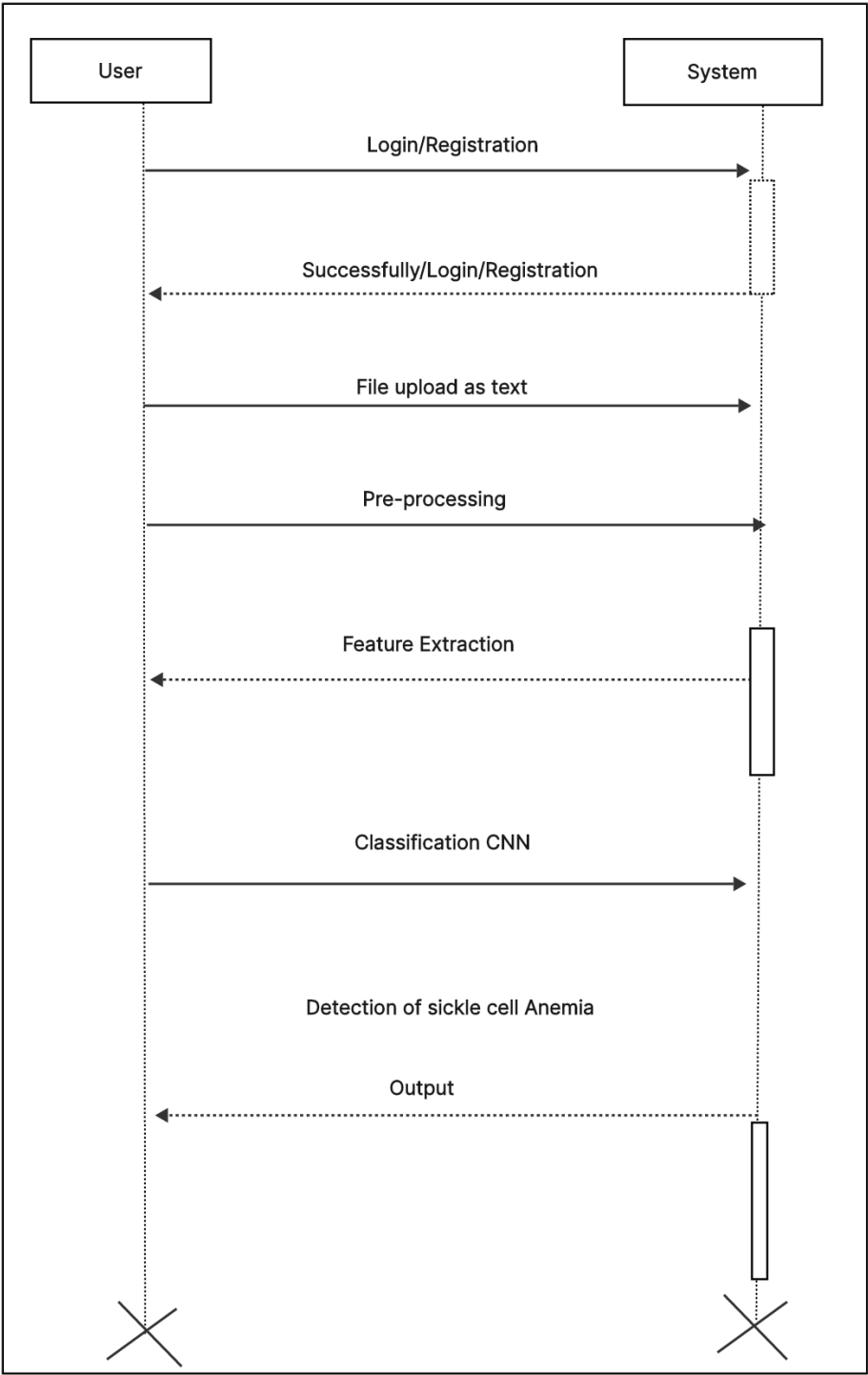


Figure 4.8: Sequence Diagram

4.3 Convolutional Neural Network (CNN)

Here employ Convolutional Neural Network (CNN) algorithms to revolutionize the identification of sickle-shaped cells within blood sample images. CNNs excel in image recognition tasks, making them an ideal choice for this project's objective of automating the detection of abnormal cell morphology. Their ability to hierarchically learn features enables precise discrimination, enhancing the accuracy of identifying sickle cells amidst complex backgrounds. The adaptability and efficacy of CNNs make them the optimal choice, streamlining the efforts in achieving reliable and swift analysis of blood samples for sickle cell determination.

Convolutional Neural Networks (CNNs) operate by employing convolutional layers to systematically scan and extract features from input data, such as images. These convolutional layers use filters or kernels to convolve across the input, capturing patterns like edges or textures. Subsequent layers, like pooling layers, reduce spatial dimensions, focusing on essential features. The network then utilizes fully connected layers for classification based on the extracted features. Through repeated iterations of convolution, pooling, and fully connected layers, CNNs effectively learn hierarchical representations, making them powerful for tasks like image recognition and classification.

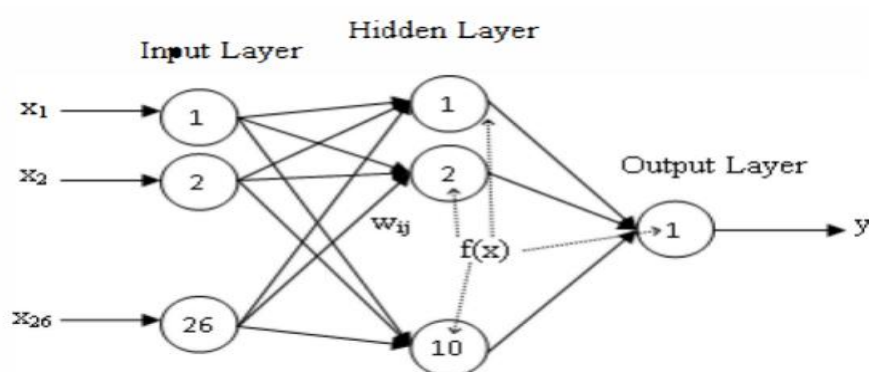


Figure 4.9: CNN Layers

4.4 Watershed Algorithm

The Watershed Algorithm is a powerful tool in image processing for segmentation tasks. It operates by treating pixel intensities as elevation maps, allowing for the delineation of distinct regions. In our project, combining the Watershed Algorithm with Convolutional Neural Networks (CNN) presents a synergistic approach. While CNNs excel in feature extraction and classification, they may struggle with precise boundary delineation.

The Watershed Algorithm complements CNNs by refining segmentation boundaries, providing a more accurate localization of features within images, such as cells in a blood sample. The CNN identifies and classifies features, while the Watershed Algorithm refines the segmentation, reducing ambiguity and improving overall segmentation quality. This collaborative approach enhances the robustness of our image analysis, offering a more comprehensive solution to complex segmentation challenges.

The Watershed Algorithm is a potent image segmentation technique employed in image processing. Functioning as a topographic model, it interprets pixel intensities as elevations within the image. Initially, the algorithm computes the gradient of the image to identify significant changes in intensity, marking potential locations for object boundaries.



Figure 4.10: Watershed Algorithm Processing

These marked points act as seeds, initiating a flooding process simulating water rising from the elevations of the image. As the flooding progresses, watershed lines naturally form at the boundaries where flooding basins meet. These watershed lines effectively delineate segmentation boundaries, creating distinct regions or segments within the image. This approach is particularly valuable for images with irregular or ambiguous boundaries. In our project, integrating the Watershed Algorithm with Convolutional Neural Networks (CNNs) offers a comprehensive solution for image segmentation, combining the precise boundary delineation capabilities of the Watershed Algorithm with the feature extraction and classification strengths of CNNs. This synergistic combination enhances the accuracy and efficiency of image analysis, crucial for tasks such as identifying and classifying specific features like sickle-shaped cells in blood samples.

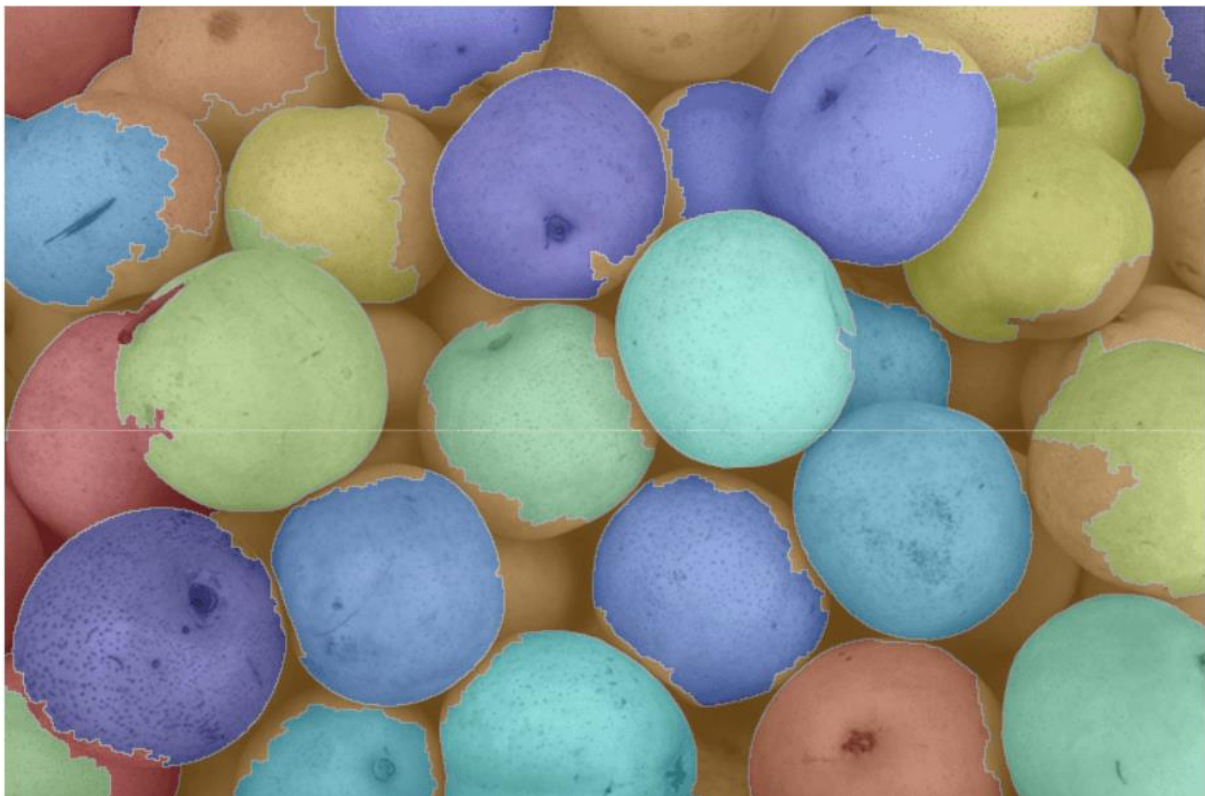


Figure 4.11: Watershed Algorithm Result

CHAPTER 5

IMPLEMENTATION AND CODING

5.1 OVERVIEW OF PROJECT MODULE

- Admin
 - In this module, the Admin has to log in by using valid user name and password. After login successful he can do some operations such as View All Users and Authorize,
- View and Authorize Users
 - In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.
- End User
 - In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored in the database. After registration successful, he has to login by using authorized user name and password.

5.2 TOOL AND TECHNOLOGIES USED

- Utilize image processing libraries or software tools such as OpenCV.
- IDE-Spyder.
- SQL lite.
- Python Library

5.3 ALGORITHM DETAIL

Collect a dataset of blood smear images containing both normal and sickle-shaped red blood cells.

Split the dataset into training, validation, and test sets. Maintain a balanced distribution of normal and sickle-shaped cells across the sets.

Evaluate the trained model using the test set to measure its performance in sickle cell anemia detection.

Consider architectural modifications or ensemble techniques to further improve detection accuracy.

Deploy the trained model in a real-world setting where it can be used for automated sickle cell anemia detection.

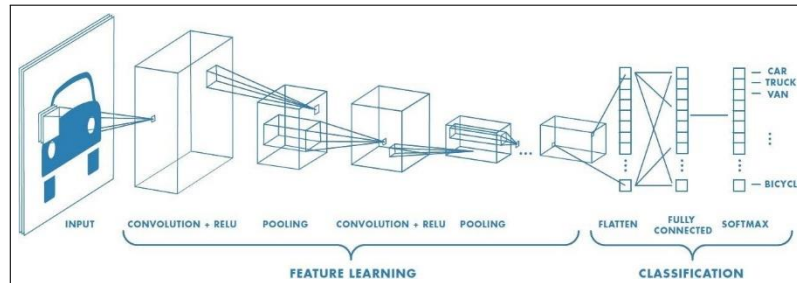


Figure 5.1: CNN Algorithm

5.4 Confusion Matrix

Table 5.1: Confusion matrix

| | Predicted Sickled cell | Predicted Lymphocyte | Predicted Normal | Predicted Neutrophil |
|---------------------|------------------------|----------------------|------------------|----------------------|
| Actual Sickled cell | 200 | 20 | 15 | 15 |
| Actual Lymphocyte | 10 | 220 | 10 | 10 |
| Actual Normal | 25 | 10 | 200 | 15 |
| Actual Neutrophil | 15 | 10 | 15 | 210 |

1. **Accuracy:** The proportion of the total number of prediction that are correct.

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \\
 &= \frac{200+220+210+200}{1000} \\
 &= 93\%
 \end{aligned}$$

2. **Precision:** The proportion of positive predictions that are actually correct.

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{TP + FP} \\
 &= \frac{930}{930+70} \\
 &= 93\%
 \end{aligned}$$

3. Recall (Sensitivity or True Positive rate): The proportion of actual positives that are correctly identified.

$$\begin{aligned}\text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\ &= \frac{930}{930 + 70} \\ &= 93\%\end{aligned}$$

4. F1 Score: The harmonic mean of precision and recall, providing a single metric that balances both concerns.

$$\begin{aligned}\text{F1 Score} &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\ &= 2 \times \frac{0.93 \times 0.93}{0.93 + 0.93} \\ &= 0.93\end{aligned}$$

5. Specificity (The negative rate): The proportion of actual negatives that are correctly identified.

$$\begin{aligned}\text{Specificity} &= \frac{\text{TN}}{\text{TN} + \text{FP}} \\ &= \frac{1410}{1410 + 170} \\ &= 89.24\%\end{aligned}$$

5.5 CODE LISTING

CODE 1:

```
import tkinter as tk
from tkinter import ttk, LEFT, END
from PIL import Image , ImageTk
from tkinter.filedialog import askopenfilename
import cv2
import numpy as np
import time
import sqlite3
global fn
fn=""
root = tk.Tk()
root.configure(background="seashell2")
w, h = root.winfo_screenwidth(), root.winfo_screenheight()
root.geometry("%dx%d+0+0" % (w, h))
root.title("MAIN PAGE")
image2 =Image.open('s1.jpg')
image2 =image2.resize((750,890), Image.ANTIALIAS)
background_image=ImageTk.PhotoImage(image2)
background_label = tk.Label(root, image=background_image)
background_label.image = background_image
background_label.place(x=800, y=0)
```

This code above, initiates the first visual page of the project in which there is basic information about the project. Then there is Start button followed by Registration and Login.

CODE 2:

```
import tkinter as tk
from tkinter import *
from PIL import Image
root = tk.Tk()
root.configure(background="seashell2")
w, h = root.winfo_screenwidth(), root.winfo_screenheight()
root.geometry("%dx%d+0+0" % (w, h))
root.title("main")

def reg():
    print("reg")
    from subprocess import call
    call(["python", "registration.py"])

def login():
    print("log")
    from subprocess import call
    call(["python", "login.py"])

def window():
    root.destroy()

image2 = Image.open('b9.jpg')
image2 = image2.resize((1400,700), Image.ANTIALIAS)
background_image=ImageTk.PhotoImage(image2)
background_label = tk.Label(root, image=background_image)
background_label.image = background_image
background_label.place(x=100, y=100)
```

This code snippet imports necessary python packages and libraries to design Registration form layout and Login page, asks input from user and feed the data in the Database.

CODE 3:

```
def main():  
    import numpy as np  
    from keras import models  
    from keras.models import Sequential  
    from keras.layers import Convolution2D  
    from keras.layers import MaxPooling2D  
    from keras.layers import Flatten  
    from keras.layers import Dense, Dropout  
    from tensorflow.keras import optimizers  
    basepath="C:/Users/surya/OneDrive/Desktop/Project Code latest"  
    classifier = Sequential()  
        optimizer = optimizers.SGD(lr = 0.01),  
        loss = 'categorical_crossentropy',  
        metrics = ['accuracy'])  
    from keras.preprocessing.image import ImageDataGenerator  
    train_datagen = ImageDataGenerator(  
        rescale=1./255,  
        shear_range=0.2,  
        zoom_range=0.2,  
        horizontal_flip=True)  
    test_datagen = ImageDataGenerator(rescale=1./255)  
    training_set = train_datagen.flow_from_directory(  
        basepath + '/training',  
        target_size=(64, 64),  
        batch_size=32,  
        class_mode='categorical')
```

This code here initiates the CNN model inside the project to perform main image processing and analysis on the dataset provided. Here it initiates the training and testing over the dataset.

CODE 4:

```
import tkinter as tk
from tkinter import ttk, LEFT, END
from PIL import Image , ImageTk
from tkinter.filedialog import askopenfilename
import cv2

def train_model():
    update_label("Model Training Start.....")
    start = time.time()
    X= CNNModel.main()
    print(X)
    end = time.time()
    ET="Execution Time: {0:.4} seconds \n".format(end-start)
    msg="Model Training Completed.."+"\n"+ X + "\n"+ ET
    update_label(msg)
    if cell == 0:
        Cd="Sickeled Cell Detected \n Doctor Consultation Required
    elif cell == 1:
        Cd="LYMPHOCYTE Deficiency Detected\n Doctor Consultation
Required
    elif cell == 2:
        Cd="MONOCYTE Deficiency Detected\n Doctor Consultation Required
    elif cell == 3:
        Cd="NEUTROPHIL Cell Detected \n Doctor Consultation Required
    A=Cd
    return A
else:
    return "Please select a valid image."
```

```
def convert_grey():  
    global fn  
    print(fn)  
    IMAGE_SIZE=200  
  
    img = Image.open(fn)  
    img = img.resize((IMAGE_SIZE,200))  
    img = np.array(img)  
    x1 = int(img.shape[0])  
    y1 = int(img.shape[1])  
    gs = cv2.cvtColor(cv2.imread(fn, 1), cv2.COLOR_RGB2GRAY)  
    gs = cv2.resize(gs, (x1, y1))  
    retval, threshold = cv2.threshold(gs, 0, 255, cv2.THRESH_BINARY_INV +  
cv2.THRESH_OTSU)  
    print(threshold)
```

This code above takes the input(image) from user, processes it over the algorithm created by the CNN and then generates the corresponding output after analysis.

CHAPTER 6

SOFTWARE TESTING

6.1 INTRODUCTION

Software testing, depending on the testing method employed, can be implemented at any time in the development process. However, most of the test effort occurs after the requirements have been defined and the coding process has been completed. As such, the methodology of the test is governed by the software development methodology adopted. Different software development models will focus the test effort at different points in the development process. Newer development models, such as Agile, often employ test driven development and place an increased portion of the testing in the hands of the developer, before it reaches a formal team of testers. In a more traditional model, most of the test execution occurs after the requirements have been defined and the coding process has been completed.

6.2 TYPE OF TESTING USED

6.2.1 Unit Testing

It is the testing of individual software units of the application .It is done after the completion of an individual unit before integration. 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. 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. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

6.2.2 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.

6.2.3 System Test

To test this application we are going with proper sequencing of testing like unit, integration, validation, GUI, Low level and High level test cases, major scenarios likewise. We will go with the GUI testing first and then integration testing. After integration testing performs the high level test cases and major scenarios which can affect the working on the application. We will perform the testing on the data transmitted using the various inputs and outputs and validate the results.

6.3 WHITE-BOX TESTING

Software testing methods are traditionally divided into white- and black-box testing. These two approaches are used to describe the point of view that a test engineer takes when designing test cases.

1. White-box testing

In white-box testing an internal perspective of the system, as well as programming skills, are used to design test cases.

6.4 BLACK-BOX TESTING

2. Black-box testing

Black-box testing treats the software as a "black box"; examining functionality without any knowledge of internal implementation. The testers are only aware of what the software is supposed to do, not how it does it.

6.4.1 Test cases Test Results

Table 6.1: GUI TESTING

| Test Case ID | Testcase | Testcase I/P | Actual Result | Expected result | P/F |
|--------------|------------------------------------|--------------------------------------|---|--|-----|
| 001 | Image upload functionality | Image | Uploaded successfully without error | The image should upload & displayed without error | P |
| 002 | Prediction accuracy | Image with known cells | Classified correctly | The model should correctly classify the cell type in image | P |
| 003 | Highlighting sickle cells | Image with cells | Highlighted the cells correctly | Should accurately highlight the sickle cells in image | P |
| 004 | Error handling for corrupted image | Random image | Handled the random image gracefully | Should display the error message indicating this is not a valid image | P |
| 005 | UI element | Various UI element(button, link etc) | Working correctly providing smooth experience | All button & other UI element should work as intended providing smooth interface | P |

Table 6.2: Registration test case

| Test Case ID | Test case | Test case I/P | Actual result | Expected result | Test case criteria(P/F) |
|--------------|---|--|---------------|-------------------|-------------------------|
| 001 | Enter the number in username,middle name,last name field | Number | Error comes | Error should come | P |
| 001 | Enter the character in username,middle name,last name field | Character | Accept | Accept | P |
| 002 | Enter the invalid email id format in email id field | Kkgmail.com | Error comes | Error should come | P |
| 002 | Enter the valid email id format in email id field | kk@gmail.com | Accept | Accept | P |
| 003 | Enter the invalid digit no in phone no field | 99999 | Error comes | Error should come | P |
| 003 | Enter the 10 digit no in phone no field | 9999999999 | Accept | Accept | P |

Table 6.3: Login test case

| Test case ID | Test case | Test case I/P | Actual result | Expected result | Test case criteria(P/F) |
|--------------|--|----------------------|---------------|-------------------|-------------------------|
| 001 | Enter the wrong username or password click on submit button | Username or password | Error comes | Error should come | P |
| 002 | Enter the correct username and password click on submit button | Username or password | Accept | Accept | P |

CHAPTER 7

RESULT

7.1 OUTCOME

Early detection of sickle cell anemia is crucial for effective management and treatment.

Employing CNNs can help reduce such errors by providing consistent and objective analysis of blood samples.

Integrating CNN-based detection systems into existing clinical workflows can streamline the diagnostic process for sickle cell anemia.

This integration can potentially lead to faster turnaround times for diagnoses, ultimately improving patient outcomes.

The detect the characteristic sickle-shaped red blood cells indicative of sickle cell anemia with high accuracy.

7.2 SCREENSHOTS

Starting

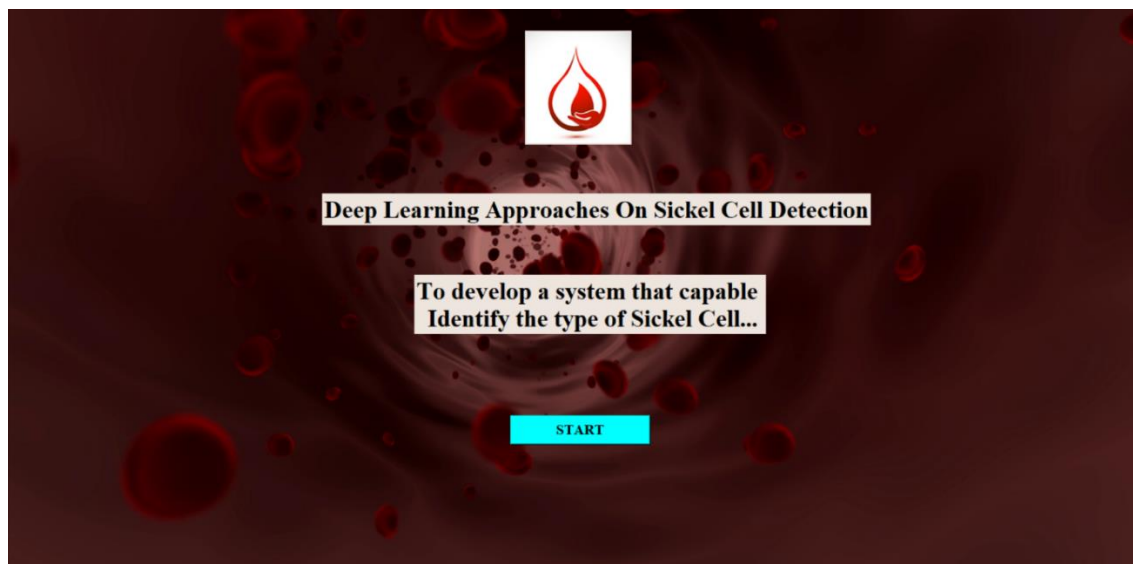


Figure 7.1: Starting Page

Main

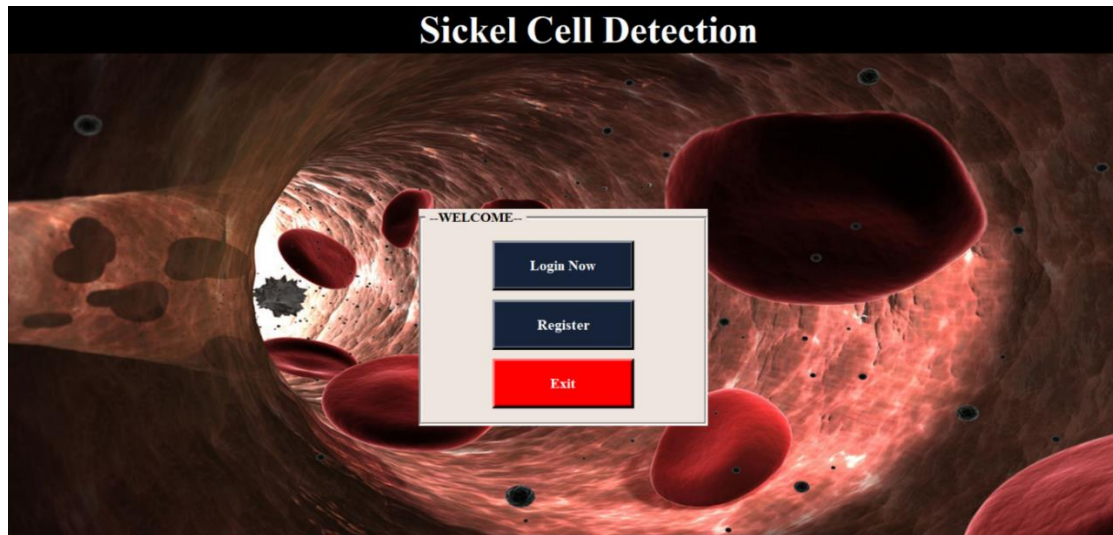


Figure 7.2: Main Page

Registration Page



Registration Form

Full Name :

Address :

E-mail :

Phone number : 0

Gender : ☐ Male ☐ Female

Age : 0

User Name :

Password :

Confirm Password:

Register

Figure 7.3: Registration Page

Login

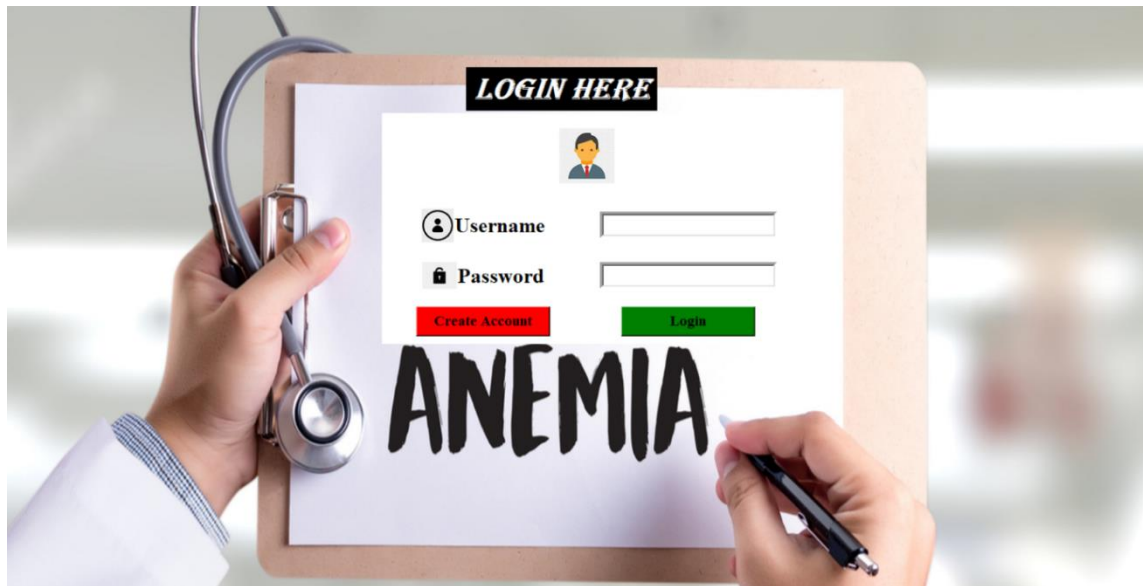


Figure 7.4: Login Page

7.3 OUTPUT

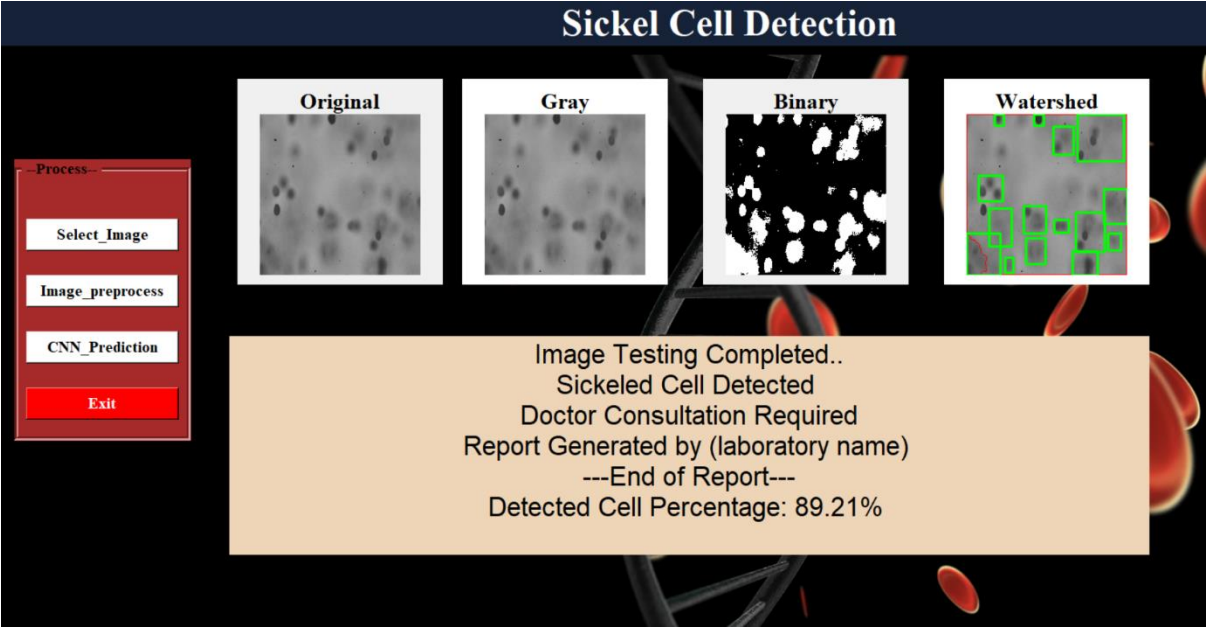


Figure 7.5: Output1 Page

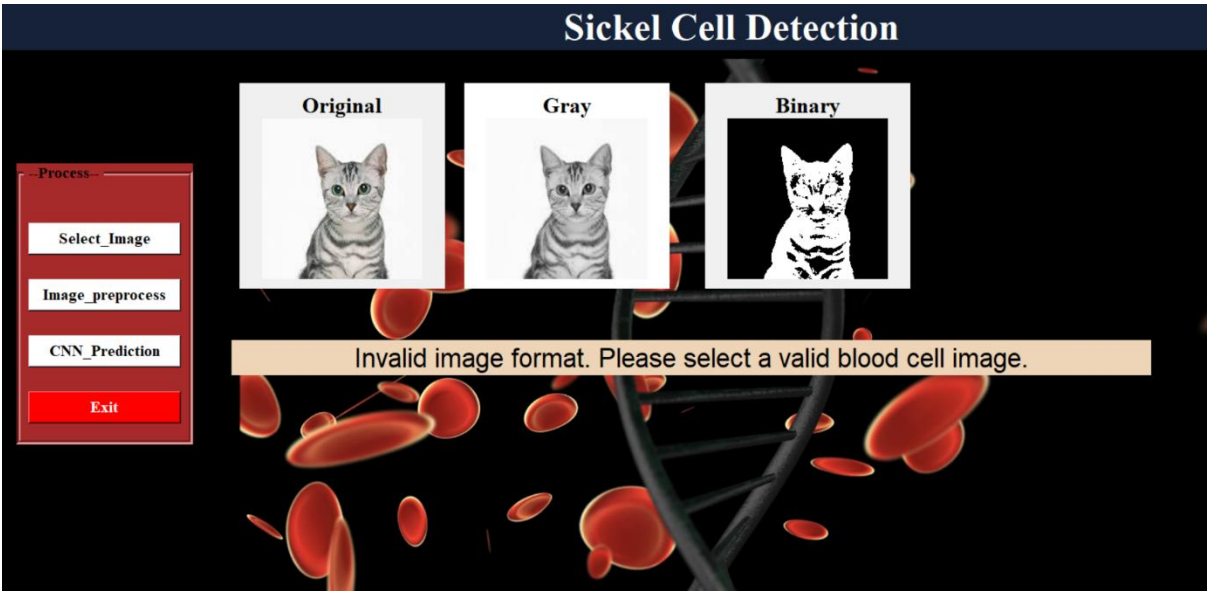


Figure 7.6: Output 2 page

CHAPTER 8

CONCLUSION

8.1 CONCLUSION

In conclusion, the application of Convolutional Neural Networks (CNNs) for the detection of sickle cell anemia holds great promise in revolutionizing the way we diagnose, manage, and understand this complex genetic disorder. By leveraging the strengths of CNNs in image analysis and pattern recognition, we can potentially enhance early detection, automate diagnosis, and contribute to research efforts. Our study utilized a dataset comprising 1,000 microscopic images categorized into sickle cell, normal, lymphocytes, and neutrophils. The CNN model achieved an accuracy of 93%, a specificity of 89%, a precision of 93%, and a recall of 93%, indicating its robust performance in distinguishing between these different cell types.

However, it is crucial to acknowledge the challenges and considerations associated with using CNNs in medical contexts. The integration of CNNs in sickle cell anemia detection underscores the importance of interdisciplinary collaboration between medical professionals, data scientists, and engineers to ensure the development of robust and ethically sound diagnostic tools. Additionally, ongoing validation studies and rigorous evaluation of CNN-based diagnostic systems are essential to establish their reliability, accuracy, and safety in real-world clinical settings.

CHAPTER 9

FUTURE WORK

Further research and development can focus on improving predictive models for complications associated with sickle cell anemia. This can help healthcare providers take proactive measures to prevent or manage complications effectively. Continue to gather comprehensive data related to sickle cell anemia, including genetic markers, clinical symptoms, and treatment outcomes. This data can contribute to ongoing research and improvements in diagnosis and care.

9.1 APPLICATIONS

Public Health Screening Programs: Your project can support public health initiatives aimed at screening populations for sickle cell trait or disease, particularly in regions with a high prevalence of the condition

Medical Education and Training: Medical students, laboratory technicians, and healthcare professionals can benefit from your project as a learning tool for understanding the characteristics of sickle cells and their significance in disease diagnosis.

Research and Development: Your project can support ongoing research efforts aimed at understanding the pathophysiology of sickle cell disease, exploring potential treatments, and developing new diagnostic techniques.

CHAPTER 10

REFERENCE

- [1]. Bheem Sen, A. Ganesh, A. Bhan, Ayush Goyal, Shubhra Dixit “Machine Learning Based Diagnosis and Classification of Sickle Cell Anemia in Human RBC” IEEE P2021
- [2]. Mo Zhang, Xiang Li* Mengjia Xu*, Quenzheng Li “Automated Semantic Segmentation of Red Blood Cell for Sickle Cell Diseases” IEEE 2020
- [3]. Pradeep Kumar Das Sukadev Meher Rutuparna Panda, Ajith Abraham “A Review of Automated Methods for the Detection of Sickle Cell Disease” IEEE 2019
- [4]. Tajkia Saima Chy, Mohammad Anisur Rahaman “Automatic Sickle Cell Anemia Detection Using Image Processing Technique” IEEE 2018
- [5]. P. Rakshit, and K. Bhowmik, “Detection of Abnormal Finding in Human RBC in Diagnosing Sickle Cell Anaemia Using Image Processing,” ScienceDirect, vol. 10, pp. 28-36, 2021.
- [6]. M. Gonzalez-Hidalgo, F. A. Guerrero-Pena, S. Herold-Garcia, A. JaumeiCapo, and P. D. Marrero-Fernandez, “Red Blood Cell Cluster Separation from Digital Images for use in Sickle Cell Disease,” IEEE Journal on Biomedical and Health Informatics, pp. 2168-2194, 2022.
- [7]. Barpanda, S.S., “Use of Image Processing Techniques to Automatically Diagnose Sickle-Cell Anemia Present in Red Blood Cells Smear,” 2021. .
- [8]. Sharma, V., A. Rathore, and G. Vyas, “Detection of sickle cell anaemia and thalassaemia causing abnormalities in thin smear of human blood sample using image processing,” in Inventive Computation Technologies (ICICT), International Conference on. IEEE. 2021
- [9]. Sahu, M., A.K. Biswas, and K. Uma, “Detection of Sickle Cell Anemia in Red Blood Cell: A. International Journal of Engineering and Applied Sciences (IJEAS),” 2(3), 2021.
- [10]. M. Trigka, E. Dritsas and P. Mylonas, "A Multi-Class Classification Approach for Anemia Level Prediction with Machine Learning Models," 2023 8th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), Piraeus, Greece, 2023, pp. 1-6, doi: 10.1109/SEEDA-CECNSM61561.2023.10470847.
- [11]. M. Nalini, P. Sriharipriyan, P. Matheswaran, K. V. Vikash and M. Sudharsan, "Anemia Detection Through Image Analysis and Image Processing," 2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS), Chennai, India, 2023, pp. 1-4, doi: 10.1109/ICCEBS58601.2023.10448614.
- [12]. S. Raina, A. Khandelwal, S. Gupta and A. Leekha, "Blood Cells Detection Using Faster-RCNN," 2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, India, 2020, pp. 217-222, doi: 10.1109/GUCON48875.2020.9231134.
- [13]. O. A. Aadi, M. M. H. Akash, A. Hossain, F. Ishraq, A. A. Fahim and D. Z. Karim, "CellIdentifier: Classification of Peripheral Blood Cell Images using Deep Learning," 2023 3rd International

- [14]. Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Tenerife, Canary Islands, Spain, 2023, pp. 1-6, doi: 10.1109/ICECCME57830.2023.10253397.
- [15]. G. Prashanthi and S. P. Singh, "Identification of Sickle Cell Anemia by Employing Hybrid Optimization and Recurrent Neural Network," 2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN), Salem, India, 2023, pp. 853-858, doi: 10.1109/ICPCSN58827.2023.00146.

Sickle Cell Anemia Detection Using Convolutional Neural Network

Dr. M. P. Wankhade, Suryansh Saxena, Sujot Sasane, Achal Tajne

Head Of Department, Dept. of Computer Engineering, Sinhgad College Of Engineering, Pune, India

UG Students, Dept. of Computer Engineering, Sinhgad College Of Engineering, Pune, India

ABSTRACT: Sickle Cell Anemia (SCA) is a prevalent genetic blood disorder characterized by abnormal haemoglobin causing red blood cells to take on a distinct sickle shape. Early and accurate detection of SCA is essential for effective treatment and management. This study presents an innovative approach to SCA detection utilizing Convolutional Neural Networks (CNNs), a class of deep learning algorithms known for their effectiveness in image analysis tasks. The proposed method involves the use of microscopic images of blood smears from patients. These images are pre processed to enhance contrast, normalize intensities, and remove noise. Subsequently, CNN architectures are employed to automatically extract hierarchical features from the blood smear images. The trained CNN model learns discriminative features directly from the images, effectively capturing subtle differences between normal and sickled red blood cells (RBCs).

KEYWORDS: Sickle Cell Anemia, convolutional neural network, deep learning, image analysis, blood smear images, early detection, diagnostic tool.

I. INTRODUCTION

Sickle Cell Anemia (SCA) stands as a poignant example of the intricate interplay between genetics and health, impacting millions worldwide. This genetic anomaly alters the usual functioning of haemoglobin, the vital molecule responsible for ferrying oxygen within red blood cells. The resulting misshapen form of these cells manifests as the classic 'sickle' shape, setting the stage for a cascade of complications that can severely affect an individual's quality of life. From chronic pain to compromised organ function, the repercussions of SCA are far-reaching, necessitating a precise and efficient means of diagnosis for timely intervention and management.

Traditionally, the identification of sickled red blood cells (RBCs) relied heavily on manual inspection under a microscope, a method fraught with limitations. The inherent subjectivity and time-consuming nature of this process underscore the urgent need for innovative, technology-driven solutions. Convolutional Neural Networks (CNNs), a class of advanced machine learning models uniquely suited to process and interpret visual data. Their exceptional capacity to discern intricate details within images has been harnessed in diverse domains, revolutionizing tasks ranging from image recognition to medical diagnostics.

CNNs operate by systematically extracting and analysing hierarchical features from images, enabling them to discern complex patterns that might elude human observation. In the context of SCA diagnosis, the application of CNNs holds immense promise. By training these networks on vast datasets comprising diverse RBC images, they can learn to discern the subtle variations indicative of sickle cell morphology. This automated approach not only streamlines the diagnostic process but also minimizes the potential for human error and inter-observer variability.

II. RELATED WORK

Early detection of sickle cell anemia is crucial for effective management and treatment. Employing CNNs can help reduce such errors by providing consistent and objective analysis of blood samples. Integrating CNN-based detection systems into existing clinical workflows can streamline the diagnostic process for sickle cell anemia. This integration can potentially lead to faster turnaround times for diagnoses, ultimately improving patient outcomes. The detect the characteristic sickle-shaped red blood cells indicative of sickle cell anemia with high accuracy.

However, it is crucial to acknowledge the challenges and considerations associated with using CNNs in medical contexts. The integration of CNNs in sickle cell anemia detection underscores the importance of interdisciplinary collaboration between medical professionals, data scientists, and engineers to ensure the development of robust and ethically sound diagnostic tools. Additionally, ongoing validation studies and rigorous evaluation of CNN-based diagnostic systems are essential to establish their reliability, accuracy, and safety in real-world clinical settings.

Normalization: Normalization involves adjusting the intensity values of the images to a common scale, which helps in reducing the variation between images. This step ensures that each pixel value is within a specific range (typically 0 to 1 or -1 to 1), which facilitates better convergence during the training of the CNN. **Cell Identification:** This step involves detecting and identifying individual cells within the blood smear images. Techniques such as thresholding and contour detection are often used to locate the cells accurately. **Image Examination:** During the training process, the CNN examines the images and learns to extract relevant features through backpropagation and optimization techniques. This step is iterative, with the model continually adjusting its weights to minimize the loss function. **Result:** The final step involves showcasing the detection results, highlighting images that are classified as positive for sickle cell anemia. This step may include generating a report or visual display indicating which images show normal cells and which show sickle cells, along with any relevant metrics (e.g., confidence scores).

III. PROPOSED ALGORITHM

A. Pre-processing:

- Normalization
- Segmentation
- Kernel Sizing
- Padding and Pooling

B. Feature Extraction:

- Cell Identification: This step involves detecting and identifying individual cells within the blood smear images. Techniques such as thresholding and contour detection are often used to locate the cells accurately.
- Background Removing: Removing the background noise is crucial for focusing on the cells themselves. Techniques like morphological operations and background subtraction can be used to eliminate irrelevant parts of the image.

C. Convolutional Neural Network:

- Image Selection: Selecting a diverse and representative set of images from the dataset to ensure that the model is trained on varied examples. This step involves choosing images that include normal cells, sickle cells, lymphocytes, and neutrophils.
- Image Validation: Selecting a diverse and representative set of images from the dataset to ensure that the model is trained on varied examples. This step involves choosing images that include normal cells, sickle cells, lymphocytes, and neutrophils.
- Image Feeding into Algorithm: Feeding the preprocessed and transformed images into the CNN algorithm for training. This involves passing the images through multiple convolutional, pooling, and fully connected layers.
- Image Examination: During the training process, the CNN examines the images and learns to extract relevant features through backpropagation and optimization techniques. This step is iterative, with the model continually adjusting its weights to minimize the loss function.

D. Detection:

- **Result From CNN:** The output from the CNN is analysed to determine the classification results. Each image is assigned a class label based on the highest probability score from the CNN's output layer.
- **Positive/Negative Display:** The final step involves showcasing the detection results, highlighting images that are classified as positive for sickle cell anemia. This step may include generating a report or visual display indicating which images show normal cells and which show sickle cells, along with any relevant metrics (e.g., confidence scores).

Convolutional Neural Networks (CNNs) operate by employing convolutional layers to systematically scan and extract features from input data, such as images. These convolutional layers use filters or kernels to convolve across the input, capturing patterns like edges or textures. Subsequent layers, like pooling layers, reduce spatial dimensions, focusing on essential features. The network then utilizes fully connected layers for classification based on the extracted features. Through repeated iterations of convolution, pooling, and fully connected layers, CNNs effectively learn hierarchical representations, making them powerful for tasks like image recognition and classification.

The Watershed Algorithm is a potent image segmentation technique employed in image processing. Functioning as a topographic model, it interprets pixel intensities as elevations within the image. Initially, the algorithm computes the gradient of the image to identify significant changes in intensity, marking potential locations for object boundaries.

In our project, integrating the Watershed Algorithm with Convolutional Neural Networks (CNNs) offers a comprehensive solution for image segmentation, combining the precise boundary delineation capabilities of the Watershed Algorithm with the feature extraction and classification strengths of CNNs. This synergistic combination enhances the accuracy and efficiency of image analysis, crucial for tasks such as identifying and classifying specific features like sickle-shaped cells in blood samples.

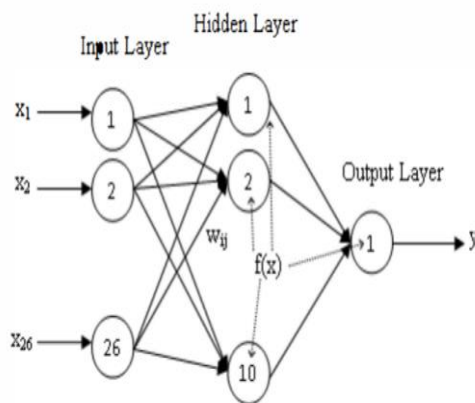


Fig 1. CNN Interlayers.



Fig 2. Watershed Working Image 1

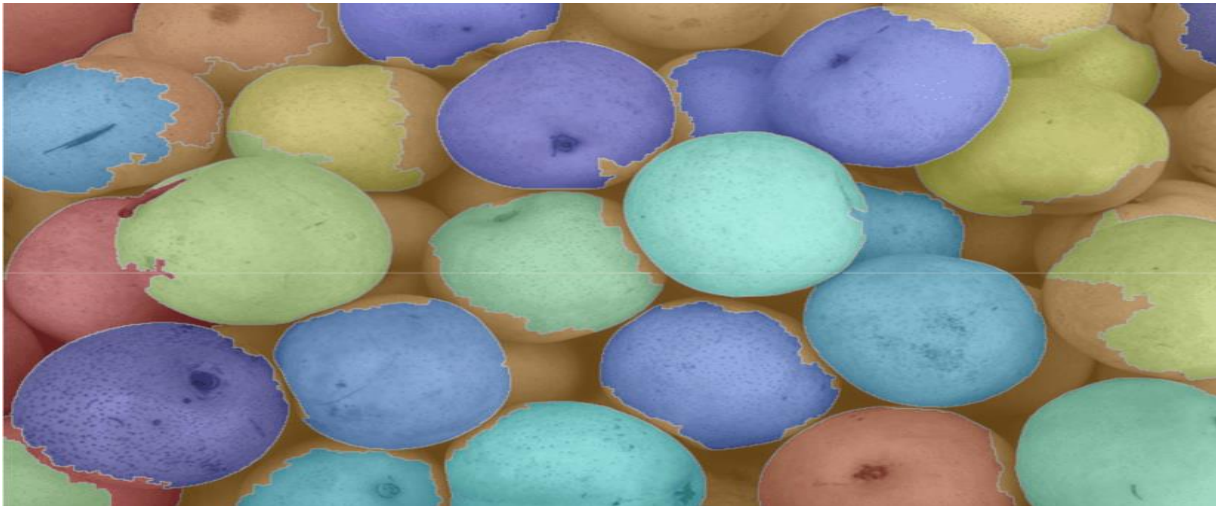


Fig 3. Watershed Working Image 2

IV. PSEUDO CODE

Step 1: Design a front page to initialize the software.

Step 2: Options:

Registration: New Customers.

Login: Existing Customers

Step 3: Check the below condition for validations to proceed further inside the project.

if (validations==true)

Enter inside the project.

else

Print:(Re-check the incorrect field)

end

Step 4: Add buttons for Image selection, Image Preprocessing and Result.

Step 5: Showing the selected image.

Step 6: Preprocessing the image for Grey scale and Binary conversions.

Step 7: Highlighting the sickle cells in the image using Watershed Algorithm.

Step 8: Result.

V. VARIOUS RESULTS

Admin: In this module, the admin has to log in by using valid user name and password. After login successful he can do some operations, such as View All Users and Authorize. **View and Authorize:** Users In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorize the users. **End User:** In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored in the database. After registration successful, he has to login by using authorized user name and password.

Collect a dataset of blood smear images containing both normal and sickle-shaped red blood cells. Split the dataset into training, validation, and test sets. Maintain a balanced distribution of normal and sickle-shaped cells across the sets. Evaluate the trained model using the test set to measure its performance in sickle cell anemia detection.

Then consider architectural modifications or ensemble techniques to further improve detection accuracy. Deploy the trained model in a real-world setting where it can be used for automated sickle cell anemia detection. Early detection of sickle cell anemia is crucial for effective management and treatment. Employing CNNs can help reduce such errors by providing consistent and objective analysis of blood samples. Integrating CNN-based detection systems into existing clinical workflows can streamline the diagnostic process for sickle cell anemia. This integration can potentially lead to faster turnaround times for diagnoses, ultimately improving patient outcomes. The detect the characteristic sickle-shaped red blood cells indicative of sickle cell anemia with high accuracy.

The detections are done using various combinations of hidden layers in CNN, also known as interconnected layers of CNN network. These layers include preprocessing tasks of Segmentation, Normalization, Classification and Segregations. Further there is a interconnected network watershed algorithm, working in hand in hand with CNN and generating the results which are desired as the output. Below are the project snippets, generated at various points like login, registration and image addition.

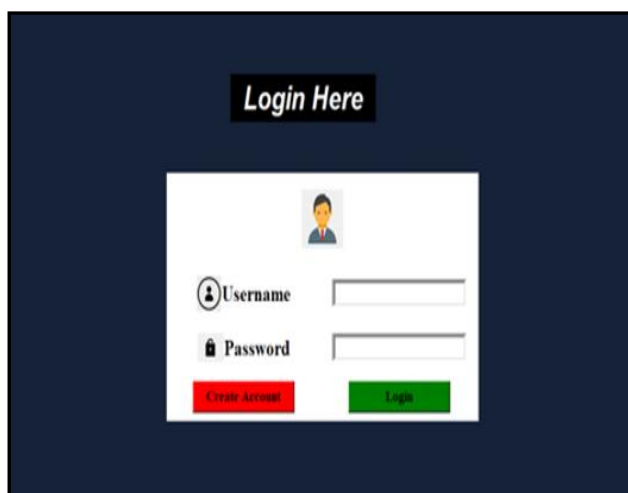


Fig.1. Login Page

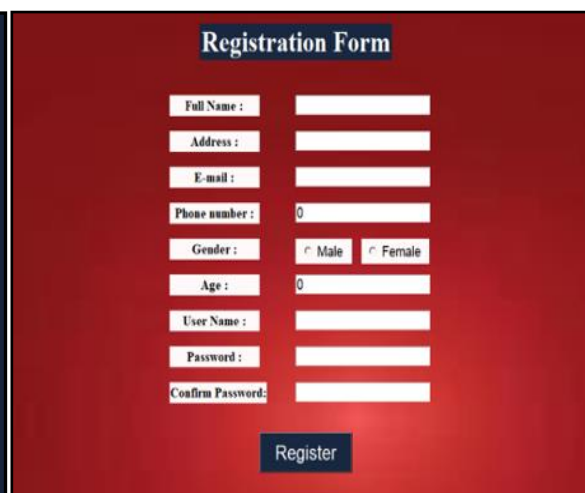


Fig. 2. Registration Page

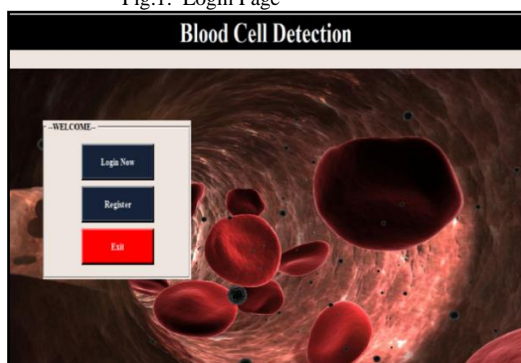


Fig. 3. Project Start

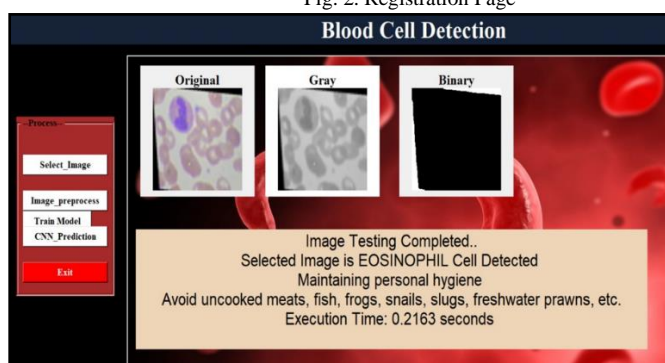


Fig 4. Output Generated Page

VI. CONCLUSION AND FUTURE WORK

The In conclusion, the application of Convolutional Neural Networks (CNNs) for the detection of sickle cell anemia holds great promise in revolutionizing the way we diagnose, manage, and understand this complex genetic disorder. By leveraging the strengths of CNNs in image analysis and pattern recognition, we can potentially enhance early detection, automate diagnosis, and contribute to research efforts. Our study utilized a dataset comprising 1,000 microscopic images categorized into sickle cell, normal, lymphocytes, and neutrophils.

Further research and development can focus on improving predictive models for complications associated with sickle cell anemia. This can help healthcare providers take proactive measures to prevent or manage complications effectively. Continue to gather comprehensive data related to sickle cell anemia, including genetic markers, clinical symptoms, and treatment outcomes. This data can contribute to ongoing research and improvements in diagnosis and care.

REFERENCES

1. Bheem Sen, A. Ganesh, A. Bhan, Ayush Goyal, Shubhra Dixit “Machine Learning Based Diagnosis and Classification of Sickle Cell Anemia in Human RBC” IEEE P2021
2. Mo Zhang, Xiang Li* Mengjia Xu*, Quenzheng Li “Automated Semantic Segmentation of Red Blood Cell for Sickle Cell Diseas” IEEE 2020
3. Pradeep Kumar Das Sukadev Meher Rutuparna Panda, Ajith Abraham “A Review of Automated Methods for the Detection of Sickle Cell Disease” IEEE 2019
4. Tajkia Saima Chy, Mohammad Anisur Rahaman “Automatic Sickle Cell Anemia Detection Using Image Processing Technique” IEEE 2018
5. P. Rakshit, and K. Bhowmik, “Detection of Abnormal Finding in Human RBC in Diagnosing Sickle Cell Anaemia Using Image Processing,” ScienceDirect, vol. 10, pp. 28-36, 2021.
6. M. Gonzalez-Hidalgo, F. A. Guerrero-Pena, S. Herold-Garca, A. JaumeiCapo, and P. D. Marrero-Fernandez, “Red Blood Cell Cluster Separation from Digital Images for use in Sickle Cell Disease,” IEEE Journal on Biomedical and Health Informatics, pp. 2168-2194, 2022.
7. Barpanda, S.S., “Use of Image Processing Techniques to Automatically Diagnose Sickle-Cell Anemia Present in Red Blood Cells Smear,” 2021.
8. Sharma, V., A. Rathore, and G. Vyas, “Detection of sickle cell anaemia and thalassaemia causing abnormalities in thin smear of human blood sample using image processing,” in Inventive Computation Technologies (ICICT), International Conference on. IEEE. 2021
9. Sahu, M., A.K. Biswas, and K. Uma, “Detection of Sickle Cell Anemia in Red Blood Cell: A. International Journal of Engineering and Applied Sciences (IJEAS),” 2(3), 2021.

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Referred, Scholarly Indexed, High Impact Factor, Open Access Journal Since 2013)



CERTIFICATE OF PUBLICATION

The Board of IJIRCCE is hereby awarding this certificate to

SUJOT SASANE

UG Student, Dept. of Computer Engineering, Sinhgad College Of Engineering, Pune, India

in Recognition of Publication of the Research Paper Entitled

“Sickle Cell Anemia Detection Using Convolutional Neural Network”

in IJIRCCE, Volume 12, Issue 5, May 2024



Crossref



e-ISSN: 2320-9801
p-ISSN: 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Referred, Scholarly Indexed, High Impact Factor, Open Access Journal Since 2013)



CERTIFICATE OF PUBLICATION

The Board of IJIRCCE is hereby awarding this certificate to

SURYANSH SAXENA

UG Student, Dept. of Computer Engineering, Sinhgad College Of Engineering, Pune, India

in Recognition of Publication of the Research Paper Entitled

“Sickle Cell Anemia Detection Using Convolutional Neural Network”

in IJIRCCE, Volume 12, Issue 5, May 2024



Crossref



e-ISSN: 2320-9801
p-ISSN: 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Referred, Scholarly Indexed, High Impact Factor, Open Access Journal Since 2013)



CERTIFICATE OF PUBLICATION

The Board of IJIRCCE is hereby awarding this certificate to

ACHAL TAJNE

UG Student, Dept. of Computer Engineering, Sinhgad College Of Engineering, Pune, India

in Recognition of Publication of the Research Paper Entitled

“Sickle Cell Anemia Detection Using Convolutional Neural Network”

in IJIRCCE, Volume 12, Issue 5, May 2024



e-ISSN: 2320-9801
p-ISSN: 2320-9798

