IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Leukaemia Detection Using MATLAB

Shreyas Ponnanna T.K.[1], Prof.Adarsha D. [2], Dr. Bhagya H. K. [3], Dr kusumadhara S .[4]

[1]*Mtech, Digital Electronics and Communication,KVGCE, Sullia D.K, Karnataka, India*

[2]*Mtech, ,E&C. Dept,Assistant Professor,KVGCE, Sullia D.K, Karnataka, India*

[3] *M.tech,Ph.D,MISTE,E&C.Dept,ProfessorKVGCE,SulliaD.K,Karnataka,India*

[4]*M.tech,Ph.D,MISTE,E&C.Dept,ProfessorKVGCE,SulliaD.K,Karnataka,India*

Abstract:

Determining cancer's stage (extent) is necessary for most cancer types. The size of the tumor and the scope to which cancer has spread control the location. However, this can be useful for determining a patient's prognosis and treatment. In contrast, Acute Lymphocytic Leukemia (ALL) does not typically form tumors. Instead, it usually affects all of the bone marrow in the body, which can spread to other organs, such as the liver and spleen. So, unlike most other cancers, ALL is not staged.Image segmentation is the most crucial part in image processing techniques. Nu- merous segmentation techniques are used to segment digital images into smaller re- gions called segments, consisting of sets of pixels in order to analyze important infor- mation from the images. Segmentation simplifies the process of information retrieval from the region of interest .It helps in converting the digital image into more relevant information and easier to analyze. This presents a comparative analysis of existing segmentation techniques and its modification to form new segmentation techniques to overcome some of the drawbacks of the existing image segmentation approaches.Leukemia, a type of cancer affecting the blood and bone marrow, requires early detection and accurate classification for effective treatment. Traditional diagnostic methods often involve manual examination of blood smears, which is time-consuming and subjective. In this study, we present a MATLAB-based approach for automated leukemia detection using image processing and machine learning techniques. We preprocess microscopic blood smear images, segment leukemia cells using K-means clustering and Otsu's thresholding method, and extract features to characterize cell morphology and texture. A K-nearest neighbors (KNN) classifier is employed to clas- sify segmented cells into leukemia and non-leukemia categories. The performance of the proposed methodology is evaluated using accuracy, precision, recall, and F1-score metrics. Our results demonstrate the effectiveness of the computational approach in accurately identifying and classifying leukemia cells, with implications for improving diagnostic efficiency and patient outcomes in leukemia management.

Index Terms – Kmeans,Otsu,Matlab,KNN.

I. INTRODUCTION

Introduction to Acute Lymphoblastic Leukemia

Acute lymphoblastic leukemia (ALL) is a rapidly progressing hemato- logic malignancy characterized by the uncontrolled proliferation of im- mature lymphoid cells in the bone marrow and peripheral blood. It is the most common type of cancer in children, although it can also affect adults. Timely and accurate diagnosis of ALL is crucial for initiating appropriate treatment and improving patient outcomes. However, the complexity of ALL diagnosis, which often involves the integration of clinical, laboratory,and molecular data, presents challenges for health- care providers.In recent years, machine learning (ML) techniques have emerged as powerful tools for analyzing complex medical data and im- proving disease diagnosis and management.ML algorithms can extract patterns and relationships from large datasets, enabling the develop- ment of predictive models and decision support systems. In the con- text of ALL, ML holds the potential to enhance diagnostic accuracy, prognostic prediction, and treatment optimization.

Acute Lymphoblastic Leukemia

Leukemia, a malignant neoplasm of the blood-forming tissues, remains a significant health concern worldwide, affecting individuals of all ages. It is characterized by the uncontrolled proliferation of abnormal white blood cells in the bone marrow, leading to compromised immune func- tion and other systemic complications. Timely diagnosis and accurate classification of leukemia subtypes are critical for initiating appropriate treatment strategies and improving patient outcomes.

Conventional methods for leukemia diagnosis often involve manual examination of blood smears by trained hematologists, a labor-intensive and time-consuming process prone to subjective interpretation and hu- man error. In recent years, there has been a growing interest in lever- aging advanced image processing and machine learning techniques to automate the detection and classification of leukemia cells from mi- croscopic blood smear images. These computational approaches offer the potential for faster, more objective, and cost-effective diagnostic solutions, thereby enhancing the efficiency and accuracy of leukemia diagnosis.

In this report, we present a comprehensive methodology for leukemia detection using MATLAB, a powerful computational tool widely used in medical image analysis and machine learning applications. Our approach encompasses multiple stages, including preprocessing of mi- croscopic images, segmentation of leukemia cells using state-of-the-art techniques, extraction of relevant features to characterize cell morphol- ogy and texture, and classification of segmented cells into leukemia and non-leukemia categories using machine learning algorithms. By inte- grating these computational techniques, we aim to develop a robust and reliable system for automated leukemia detection, with the ulti- mate goal of assisting healthcare professionals in early diagnosis and treatment planning.

The report begins with a detailed description of the dataset used in our study, followed by an overview of the methodology employed for leukemia detection. We discuss the various image processing techniques utilized for preprocessing and segmentation, as well as the feature ex- traction and classification algorithms implemented for cell classification. Subsequently, we present the results of our experiments, including seg- mentation masks, classification performance metrics, and analysis of the findings. Finally, we offer insights into the implications of our research and potential avenues for future work in this field.

Through our investigation, we seek to contribute to the ongoing ef- forts aimed at advancing computational approaches for leukemia diag- nosis, with the ultimate objective of improving patient care and out- comes in the realm of hematologic malignancies.

II. RELATED WORK

Kyphosis is a spinal disorder that causes an excessive outward curvature of the spine,resulting in a hunchback or rounded posture. [1]Detecting kyphosis early is important for effective treatment and management. Machine learning techniques have shown promise in detecting kyphosis from medical images, such as X-rays and MRI scans.Here are some literature surveys on kyphosis detection for machine learning:

”Special Issue on Using Machine Learning Algorithms in the Pre- diction of Kyphosis Disease by Stephen Dankwa and Wenfeng Zheng ,2019.[2]”Comparative Analysis of Supervised Machine and Deep Learning Algorithms for Kyphosis Disease Detection by Alok Singh Chauhan ,Umesh Kumar Lilhore ,Amit Kumar Gupta,Poongodi Manoharan ,Ruchi Rani Garg, Fahima Hajjej, Ismail Keshta and Kaamran Raahemifar,2023.[2]”Automated Detection of Kyphosis from X-ray Images using Ma- chine Learning Algorithms: A Review” by N. B. Asbahi and H. Y. Al-Nashash (2020): This survey paper examines the use of machine learning algorithms in detecting kyphosis from X-ray images. The au- thors discuss the different machine learning techniques used, such as convolutional neural networks and decision trees, and the challenges in developing accurate and efficient algorithms.[3]”Deep Learning Approaches for the Detection of Kyphosis: A Sys- tematic Review” by S. A. Zayed et al. (2020): This systematic review focuses on the use of deep learning approaches, such as convolutional neural networks and recurrent neural networks, for kyphosis detection.

The authors discuss the different types of medical images used and the challenges in developing accurate and generalizable deep learning models.[4]

The study emphasizes the critical role of blood and its components in diagnosing life-threatening diseases such as leukemia, HIV, anemia, tumors, cancer, and thalassemia. Leukemia, specifically acute lym- phoblastic leukemia (ALL), originates in the bone marrow from im- mature lymphocytes called lymphoblasts and spreads throughout the body via the blood. Timely treatment significantly improves patient outcomes. Lymphoblasts are predominantly found in bone marrow or peripheral blood slides of ALL patients, with ALL subcategorized into L1, L2, and L3 based on morphological structures. Manual morpholog- ical examination of blast cells by hematologists is slow and dependent on expertise, prompting the need for fast and automated diagnostic systems. The paper aims to address this need by focusing on blast cell segmentation, a crucial yet challenging task in automated diagnosis systems. The subsequent sections explore related work, detail materials and methods, present experimental results and discussion, and conclude with insights into the proposed segmentation approach.[5]

Image segmentation is the most crucial part in image processing techniques. Numerous segmentation techniques are used to segment digital images into smaller regions called segments, consisting of sets of pixels in order to analyze important information from the images. Segmentation simplifies the process of information retrieval from the region of interest (ROI).It helps in converting the digital image into more relevant information and easier to analyze. This paper presents a comparative analysis of existing segmentation techniques and its mod- ification to form new segmentation techniques to overcome some of the drawbacks of the existing image segmentation approaches.[6]

III. PROBLEM STATEMENT

Acute lymphoblastic leukemia (ALL) is a type of cancer that affects the blood and bone marrow. It progresses rapidly and requires prompt diagnosis and treatment. Traditional methods of diagnosis and treat- ment planning involve manual examination of blood samples and expert interpretation, which can be time-consuming and subjective. Addi- tionally, predicting treatment outcomes and relapse risk for ALL pa- tients remains a challenge. Machine learning (ML) models can assist in automating diagnosis, predicting patient outcomes, and personaliz- ing treatment strategies and the Solution includes creating a machine learning model that accurately recognizes human actions in videos or photos involves several steps, including data collection, preprocessing, model selection, training, and evaluation.

IV. EXISTING SYSTEM

Existing systems for leukemia detection utilize image processing and machine learning techniques to automate the analysis of blood smear images. These systems enhance diagnostic accuracy and efficiency by leveraging segmentation algorithms and classification models. By ex- tracting relevant features and training classifiers on labeled datasets, existing systems can differentiate between normal and leukemic cells. Integration with healthcare information systems streamlines workflow and facilitates seamless clinical deployment. Performance evaluation metrics such as accuracy, sensitivity, and specificity validate the effi- cacy of these systems. Challenges include the need for large and di- verse datasets, standardization of imaging protocols, and robustness to variations in staining techniques. Despite these challenges, exist- ing systems demonstrate promise in augmenting traditional diagnostic methods and improving patient care. Continued research efforts aim to address limitations and further enhance the capabilities of automated leukemia detection systems.

V. OBJECTIVES

For acute lymphoblastic leukemia (ALL) can have several objectives aimed at improving diagnosis, prognosis, treatment planning, and pa- tient outcomes. Here are some potential objectives:

- Early Diagnosis and Detection: Develop ML models that can ac- curately detect and diagnose ALL at an early stage. This can involve analyzing patient data such as blood cell morphology, ge- netic markers, and other clinical parameters to identify patterns indicative of ALL.
- Prognostic Modeling: Develop ML models that can accurately pre- dict patient outcomes, such as relapse-free survival and overall sur- vival. This can involve integrating clinical, molecular, and demo- graphic data to identify prognostic factors associated with disease progression and survival.
- Clinical Decision Support: Develop ML-based decision support systems to assist clinicians in treatment planning and manage- ment. This can involve integrating patient data with evidence- based guidelines and expert knowledge to provide personalized treatment recommendations.
- Patient Outcome Prediction: Develop models that predict long- term outcomes for ALL patients based on various factors such as treatment response, genetic profile, and comorbidities. This infor- mation can help clinicians and patients make informed decisions about treatment options and end-of-life care.
- Patient Monitoring and Surveillance: Develop ML-based monitor- ing systems to track disease progression and detect relapse early. This can involve analyzing longitudinal patient data, such as blood cell counts and minimal residual disease measurements, to identify signs of disease recurrence.

VI. IMPLEMENTATION

The process of implementation is as follows:

- Data Collection: Gather a dataset of microscopic blood smear im- ages, including samples from patients with leukemia and healthy individuals.
- Data Preprocessing:
  - In preprocessing the clamors in the pictures are eliminated.
  - Clamors in the picture addresses the undesirable pixels.
  - The Gaussian channel is applied to eliminate clamors from the picture.
  - This will smooth the picture and make every one of the pixels in the picture all the more clear.
  - Inorder to apply Gaussian channel we use imfilter() capability.
- Image Segmentation: Apply segmentation techniques to distin- guish leukemia cells from the background:
  - Implement K-means clustering to partition the image into clus- ters, with each cluster representing a distinct cell type.
  - Utilize Otsu's thresholding method to find the optimal thresh- old for image binarization, separating leukemia cells from the background.
- Feature Extraction: Extract features from the segmented regions to characterize cell morphology and texture: - Compute statistical

The purpose of the design is to plan the solution of a problem specified by the requirements documents. This phase is the first step moving from problem to the solution domain. In other words, starting with what is needed design takes us to work how to satisfy the needs the design of the system is perhaps the most critical factor affecting the quality of the output and has a major impact on the later phases.

System design aims to identify the modules that should be in the system, their functions and interactions with each other to produce desired results. This chapter presents the High Level Design and a brief description of the modules.

High Level Design

The block diagram shown in Fig. 3.1 is the high level design architecture of the Acute Lymphoblastic Leukemia cell diagnosis model. This design provides an efficient way of understanding the system.

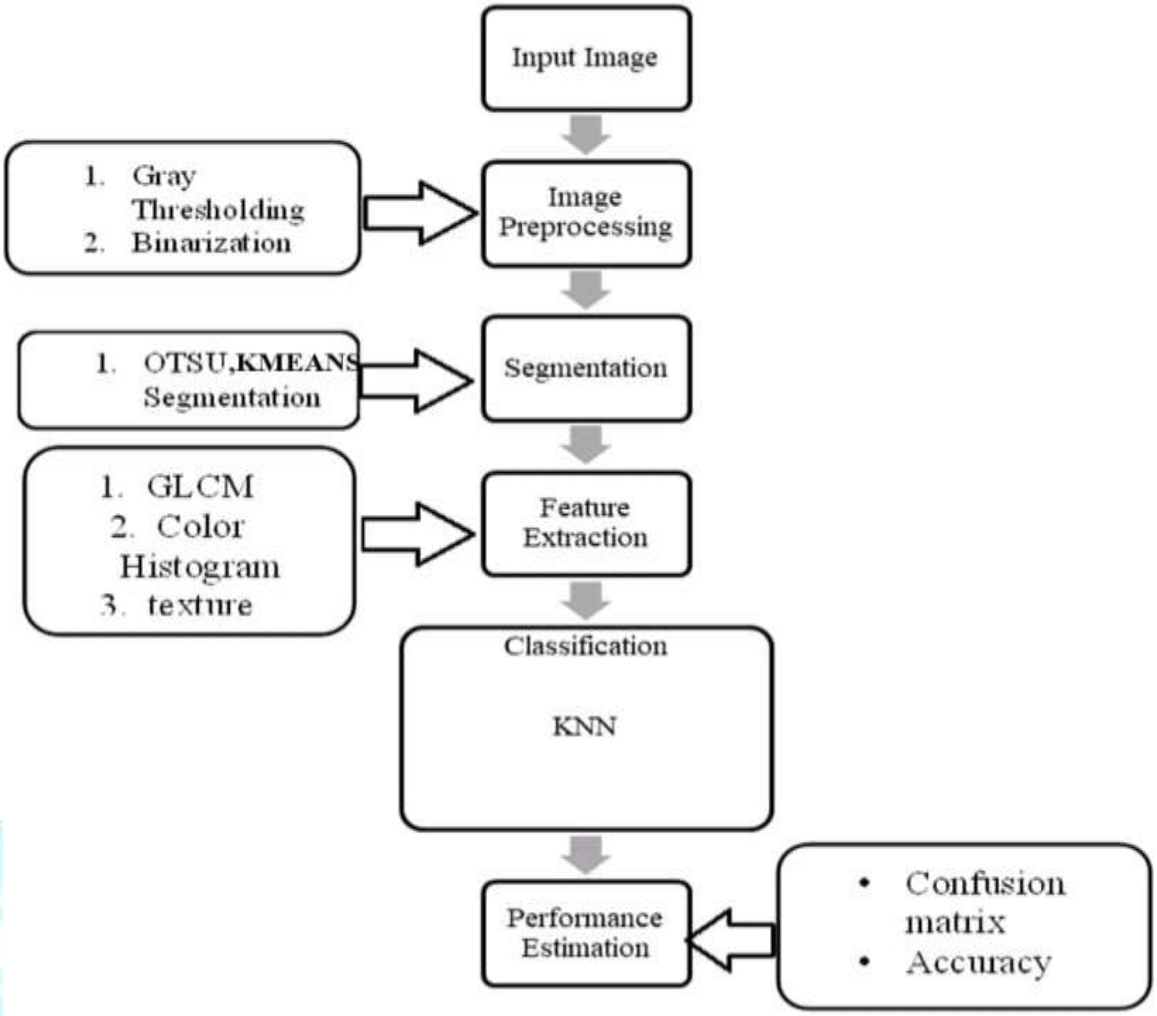

Figure 3.1: High Level Design Analysis

1. Data Collection and Preprocessing:

- Gather a dataset of different natural products with highlights like size, variety, surface, and so on.

- Preprocess the information by taking care of missing qualities, normalizing elements, and encoding straight out factors.

2. Feature Selection/Extraction:

- Conclude which elements are pertinent for arrangement.

- You could have to remove valuable elements from pictures assuming that you're utilizing picture information.

3. Data Splitting:

- Part the dataset into preparing and testing sets. Regularly, you'd utilize 70-80% for preparing and the rest for testing.

4. Implementing Machine Learning Models:

k-Nearest Neighbors (kNN):

- Pick the worth of k (number of neighbors).

- Train the kNN model on the preparation information.

- Test different distance measurements (Euclidean, Manhattan, and so on.).

- Once more, hyperparameter tuning should be possible.

5. Model Evaluation:

- Assess each model utilizing fitting measurements like exactness, accuracy, review, F1-score, and so on, on the testing information.

6. Model Comparison:

- Contemplate the presentation of each model using the appraisal estimations.

- Recognize the characteristics and deficiencies of every procedure for selection of best segmentation.

3. Data Splitting:

- Part the dataset into preparing and testing sets. Regularly, you'd utilize 70-80% for preparing and the rest for testing.

4. Implementing Machine Learning Models:

k-Nearest Neighbors (kNN):

- Pick the worth of k (number of neighbors).

- Train the kNN model on the preparation information.

- Test different distance measurements (Euclidean, Manhattan, and so on.).

- Once more, hyperparameter tuning should be possible.

5. Model Evaluation:

- Assess each model utilizing fitting measurements like exactness, accuracy, review, F1-score, and so on, on the testing information.

6. Model Comparison:

- Contemplate the presentation of each model using the appraisal estimations.

- Recognize the characteristics and deficiencies of every procedure for selection of best segmentation.

USECASE DIAGRAM :

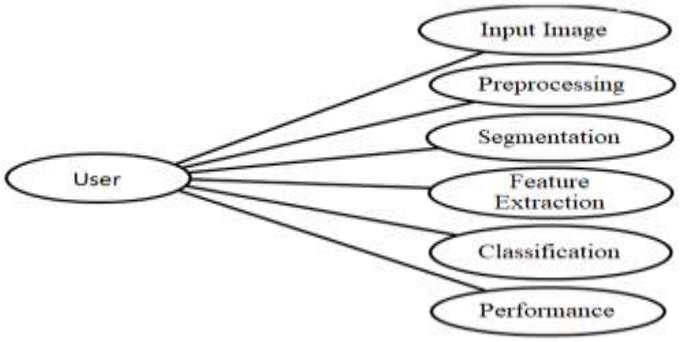

Fig 3.2,Use Case Diagram

SEQUENCE DIAGRAM:

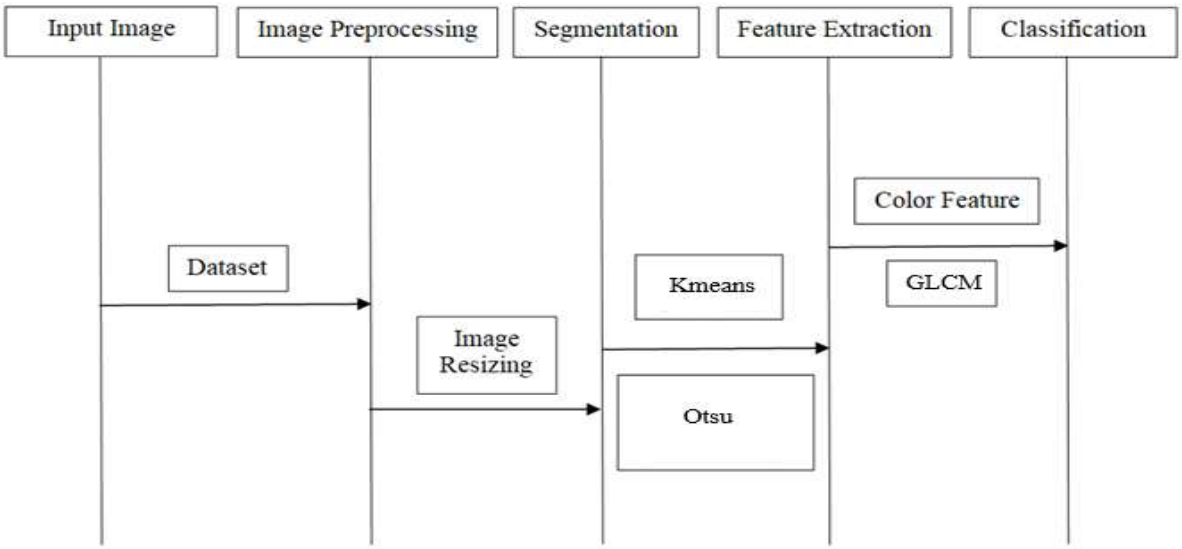

Fig 3.3,Sequence Diagram

CLASS DIAGRAM:

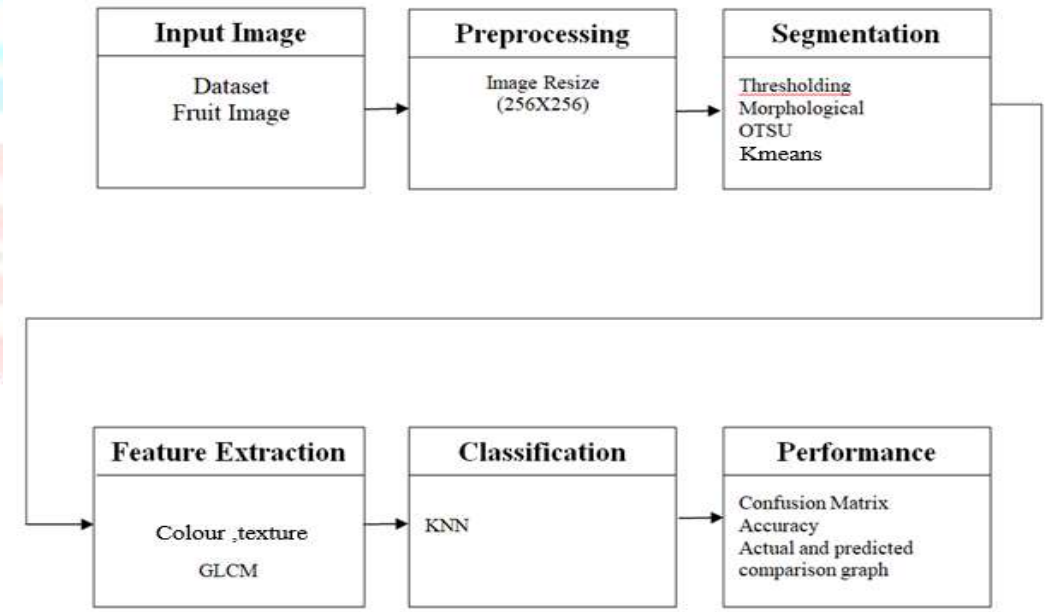

Fig 3.4,Class Diagram

VII.HARDWARE REQUIREMENTS AND SOFTWARE REQUIREMENTS

HARDWARE REQUIREMENTS:

1. Intel i3
2. 4GB DDR RAM
3. 250Gb Hard Disk

SOFTWARE REQUIREMENT:

1. Operating System : Windows 10 above
2. Tool : Matlab R2018a

VII. RESULT AND DISCUSSION

Comparison of Leukemia Detection Algorithms

The table 5.1 gives the summary of the various entities compared for the different leukemia detection algorithms.

Method	Infected Cell	Non-infected cell
OTSU	100.0	100.0
KNN	90.0	92.5

Table 5.1: Comparison of Leukemia Detection Algorithms

As observed in the table 5.1, the OTSU algorithm generates the large number of accuracy.The second largest is given by KNN algorithm.

Fig.5.1 gives the confusion matrix of the two algorithms mainly Ran- dom Forest and Decision tree. Fig. **??** gives the Loan approved or rejected using SVM algorithm.

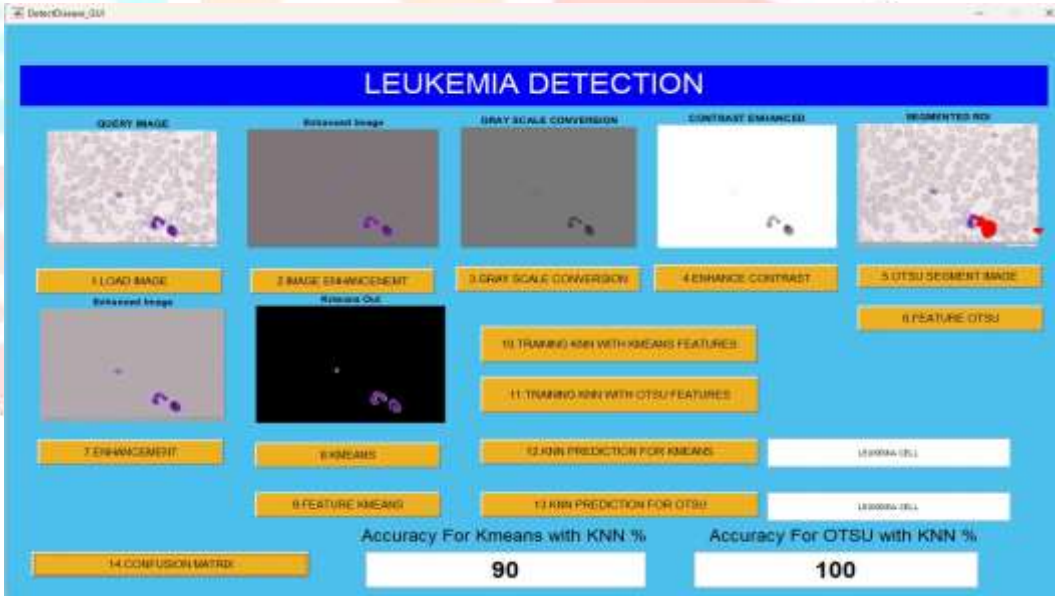

Figure 5.1: Leukemic cell

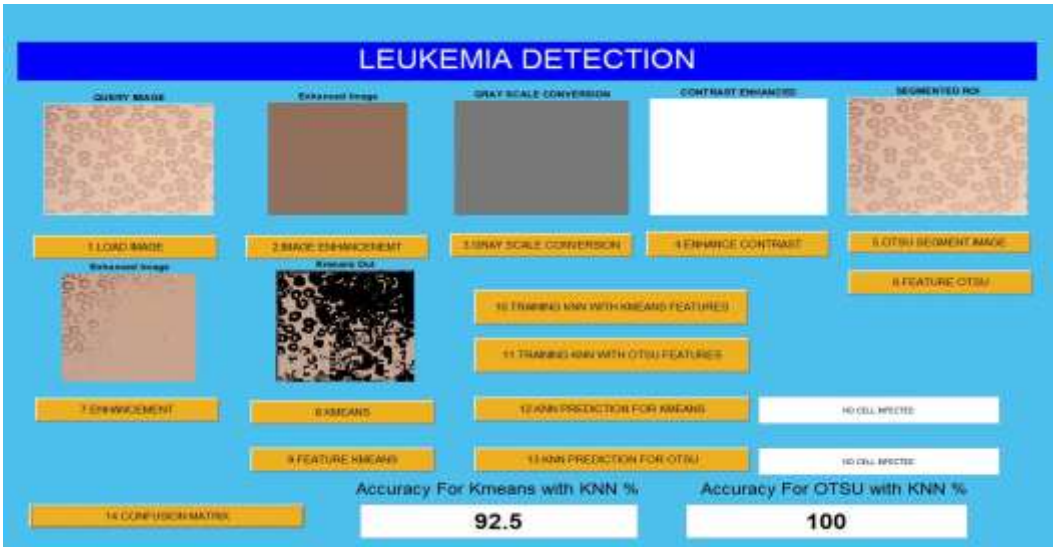

Figure 5.2: Normal cell

PERFORMANCE ANALYSIS

Execution examination utilizing a disarray framework and exactness is a basic move toward assessing the viability of your organic product characterization models. The disarray framework gives a nitty gritty breakdown of the expectations made by the model, while exactness is a basic and instinctive metric that shows the general rightness of the model's forecasts. This is the way you can examine execution involving the disarray lattice and exactness for your organic product order task:

Confusion Matrix Calculation

: After testing your classifier (k-NN), calculate a confusion matrix for each classifier utilizing the anticipated names and the ground truth marks from the testing set.The disarray grid will be a square frame- work where lines address the genuine classes, and sections address the anticipated classes.

Accuracy Calculation

: Calculate the accuracy for each classifier using the formula: Accuracy = (Number of Correct Predictions) / (Total Number of Predictions) The accuracy value is a percentage that represents how many predictions were correct out of all predictions made.

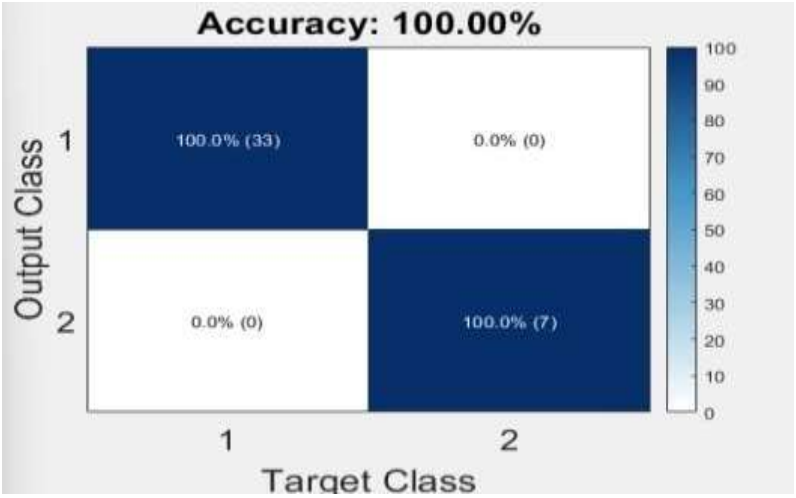

Figure 5.3: OTSU confusion matrix

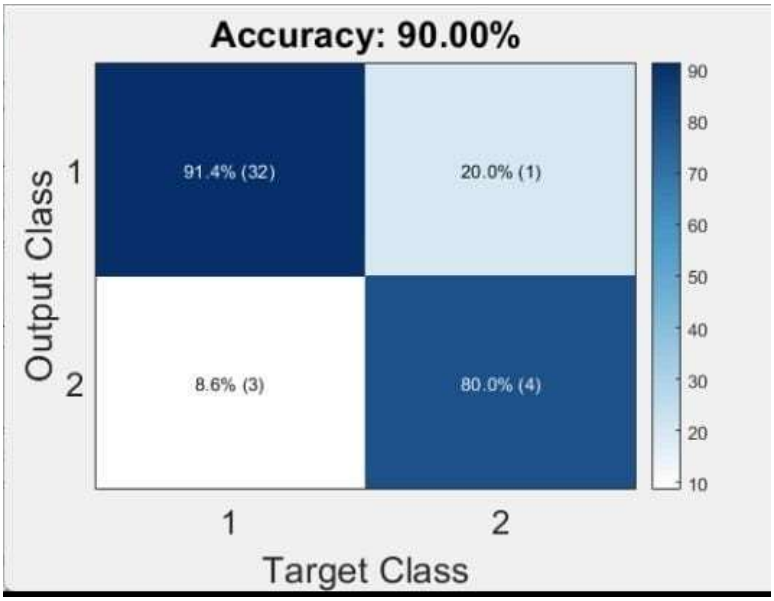

Figure 5.4: K means confusion matrix

CONCLUSION

In conclusion,our study suggests that Otsu's thresholding method out- performs K-means clustering as the preferred method of segmentation for leukemia detection in microscopic blood smear images. Through a comparative analysis of the segmentation results and classification per- formance, it becomes evident that Otsu's method provides superior seg- mentation accuracy and robustness, leading to improved classification outcomes.Otsu's thresholding method effectively separates leukemia cells from the background by automatically determining an optimal thresh- old to maximize inter-class variance, thereby minimizing intra-class variance.This results in clear and precise segmentation masks that accurately delineate leukemia cells, facilitating accurate feature extrac- tion and classification.On the other hand, while K-means clustering also partitions the image into clusters, its performance is contingent upon the initializa- tion of centroids and may be sensitive to noise and variability in cell morphology. Consequently, K-means clustering may yield suboptimal segmentation results, leading to inaccuracies in classification and re- duced diagnostic efficacy.By leveraging Otsu's thresholding method for segmentation, our leukemia detection system demonstrates improved sensitivity and specificity in identifying leukemia cells, leading to en- hanced diagnostic accuracy and clinical decision-making. These find- ings underscore the importance of selecting appropriate segmentation techniques in medical image analysis, with Otsu's method emerging as the preferred choice for leukemia detection in our study.Moving forward, further research may explore hybrid segmentation approaches that combine Otsu's method with other advanced techniques to enhance segmentation accuracy and robustness. Additionally, validation of the proposed methodology on larger and more diverse datasets, as well as integration with clinical workflows, will be essential for its adoption in real-world healthcare settings.In conclusion, our study advocates for the adoption of Otsu's thresholding method as the preferred segmentation technique for leukemia detection in microscopic blood smear images, offering a reliable and efficient solution for automated diagnosis and treatment planning in oncology.

REFERENCES

[1] N. B. Asbahi and H. Y. Al-Nashash. Automated detection of kyphosis from x-ray images using machine learning algorithm. 2020.

[2] Umesh Kumar Lilhore Amit Kumar Gupta Poongodi Manoharan Ruchi Rani Garg Fahima Hajjej Ismail Keshta Alok Singh Chauhan. Comparative analysis of supervised machine and deep learning algorithms for kyphosis disease detection. 2023.

[3] S. A. Zayed. Deep learning approaches for the detection of kyphosis: A systematic review, 2020.

[4] Stephen Dankwa and Wenfeng Zheng. *Special issue on using machine learning algorithms in prediction of kyphosis disease*. 2019.

[5] Sikandar Ali Jiwei Huang Syed Izad, Mishba. Efficient segmentation of lymphoblast in acute lymphocytic leukemia, 2020.

[6] Mohammed Dawood Salman Salva, Khalid Abdul. Comphrehensive study of image segmentation, 2021.