www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

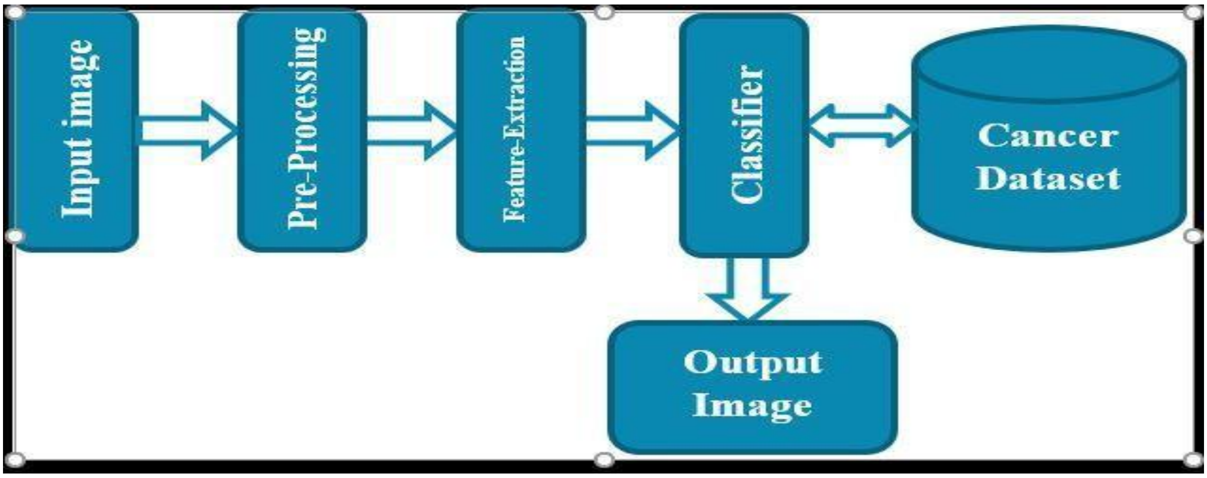
Automated Cancer Identification in Human Blood Samples and Microscopic Pictures Enhancing Diagnosis and Classification with Machine Learning Methods

V Venugopal[1], M Manoj Kumar[2], N Bhakthavachalu[3]
Professor[1], Associate Professor[2], Assistant Professor
[1,2]Department of Computer Science & Engineering , Usha Rama College Of Engineering & Technology , Vijayawada
[3]Department of IT , RK College of Engg , Vijayawada

Abstract: One of the biggest causes of death in the world today is cancer, and increasing patient survival rates depends on early identification. Diagnostic delays and misinterpretations are frequently caused by the time-consuming, subjective, and highly skilled manual microscopic inspection of blood samples. In this work, we use MATLAB machine learning algorithms to evaluate microscopic blood smear images in order to present an automated cancer diagnosis system. The three main stages of the suggested methodology are feature extraction using morphological and texture-based descriptors, classification using machine learning models, and preprocessing and enhancement of blood pictures. The findings show that the system can successfully distinguish between normal and malignant blood samples with excellent sensitivity, specificity, accuracy, and precision. Our solution performs better when compared to other methods, especially when it comes to robustness and diagnostic reliability. This work lays the groundwork for future advancements in automated cancer detection and adds to the expanding corpus of computer-aided diagnosis systems.

Keywords: Computer-Aided Diagnosis, Cancer Detection, Microscopic Images, MATLAB, Machine Learning, and Blood Sample Classification.

Abstract- The entire process of the suggested automated cancer detection system employing microscopic blood smear images is graphically depicted in the graphical abstract. The raw blood sample obtained under a microscope is represented by the Input Image, which is where the process starts. In order to improve image clarity, this image is first put through the Pre-Processing stage, where methods including noise reduction, contrast enhancement, and normalization are used. Following preprocessing, the Feature Extraction stage records important blood cell morphological and textural traits that are essential for differentiating between samples that are normal and those that are malignant. The Classifier, which has been trained on a Cancer Dataset, receives these extracted features. The classifier correctly determines whether the input sample is in the malignant or non-cancerous class by comparing the input attributes with the dataset. The categorization result is finally shown in an understandable format in the Output Image.



Purpose- The flow diagram in this work is a quick reference that shows how an input image goes through feature extraction, pre-processing, and classification to produce the final diagnostic output. The graphical abstract highlights the study's main contribution—an effective cancer detection framework utilizing machine learning algorithms on microscopic blood smear images—by presenting the methodology in a straightforward pictorial way.

Introduction

Clinical pathology relies heavily on the microscopic analysis of blood samples to diagnose cancer. However, conventional diagnostic techniques are time-consuming, prone to human error, and heavily reliant on skilled pathologists. There is a pressing need for quick, dependable, and automated diagnostic technologies that can assist medical professionals in making clinical decisions due to the rising number of cancer cases globally. One of the biggest causes of death worldwide is still cancer. Medical research indicates that the effectiveness of treatment increases with the early detection of the illness. A manual examination of blood smear images under a microscope is frequently used to diagnose hematological malignancies. This approach is not always reliable and relies on the pathologist's experience.

Recent progress in computer vision and machine learning has opened the possibility of developing automated systems for medical diagnosis. These devices can reduce human error by rapidly and reliably analyzing blood smear images. The goal of this research is to create a MATLAB-based system that uses supervised learning techniques to categorize blood cells as either normal or malignant.

1.1 Objective of the Study

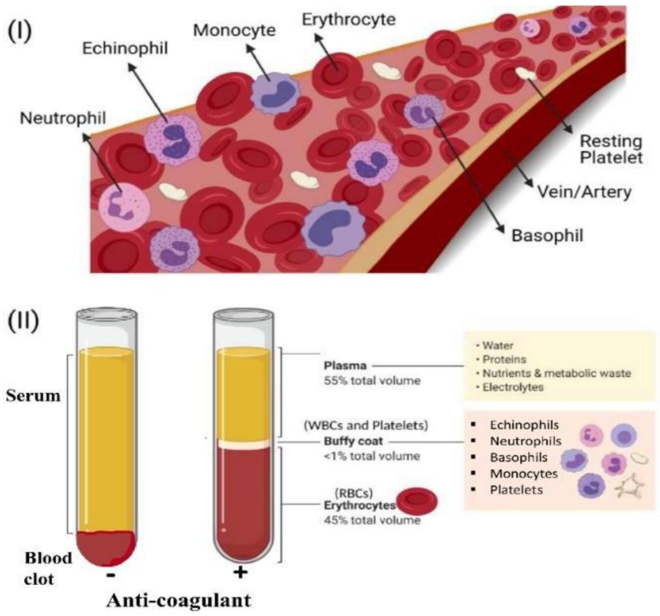
2. This study's primary goal is to create an automated system that uses MATLAB machine learning algorithms with microscopic pictures to detect cancer in human blood samples. By creating a framework that can preprocess images, extract significant features, and categorize them into malignant or normal categories, the effort seeks to lessen reliance on manual diagnosis, which is frequently laborious and prone to error. We also hope to test several machine learning models, compare their performance, and determine the best strategy for obtaining accurate and dependable results. The ultimate objective is to aid in the development of an economical and effective instrument that can assist pathologists in the early identification and diagnosis of cancer.

2.1 Understanding Blood and its Components

Human blood is composed of several key components, each serving a specific function:

- **Red Blood Cells (RBCs)** – Responsible for oxygen transport through hemoglobin.
- **White Blood Cells (WBCs)** – Play a vital role in immune defense.
- **Platelets** – Help in blood clotting.
- **Plasma** – The liquid component that carries cells, nutrients, and hormones.

Fig 1. Structure and Composition of Human Blood



1.2 Organization

3. The following sections make up the structure of this article. Automated cancer diagnosis is introduced in Section 1.

human blood samples utilizing machine learning methods and microscopic photos. The relevant work of current techniques and literature contributions in the field of cancer detection and medical image analysis are included in Section 2. The problem definition, goals, and measures taken into consideration for the proposed study are included in Section 3. The system architecture and key components of the suggested framework, including as preprocessing, feature extraction, and classification, are presented in Section 4. A step-by-step flowchart outlining the process is provided in Section 5. The results and discussion, along with performance metrics, tables, and comparative analysis, are presented in Section 6. The suggestions and possible enhancements for further improving the system are included in Section 7. Section 8 provides a summary of the research findings and future directions.

4. Related Work

5. Using a variety of machine learning and image processing approaches, automated cancer diagnosis and classification using medical pictures has been extensively investigated in recent years. These research offer fundamental techniques for creating strong and trustworthy diagnostic frameworks. Using supervised learning, an adaptive image analysis method was developed to identify and locate features in medical pictures, highlighting the significance of feature extraction and preprocessing for increased classification accuracy [1]. Another study emphasized the application of machine learning to biomedical image processing and suggested hybrid models to strike a balance between diagnostic effectiveness and computational economy [2]. Optimized feature extraction greatly improves diagnostic accuracy, as demonstrated by the use of artificial intelligence techniques for automated white blood cell detection[3].

The difficulties of noise, picture distortion, and dataset imbalance in medical image processing were highlighted in a study of blood cell image classification techniques [4]. Through targeted feature analysis, a machine learning model based on blood tests showed increased accuracy for early lung cancer prediction [5]. White blood cell identification was improved by using deep learning-based image segmentation approaches, which achieved excellent precision but required significant processing power [6].

Later, frameworks driven by deep learning and automation were investigated to improve the efficiency of cancer diagnosis and classification [7], [8]. Other models achieved higher diagnostic accuracy in identifying malignant cells, demonstrating the potential of optimized neural architectures for trustworthy cancer prediction[9],[10]. Additionally, recent studies have concentrated on improving accuracy through AI-driven imaging and biomarker detection. One study revealed how AI may increase cancer prognostic and diagnostic imaging accuracy [12], while another showed how machine learning can help find biomarkers for better cancer detection [11].

6. Overall, the research shows that even though many approaches attain excellent accuracy, issues with noise treatment, dataset constraints, and computational efficiency still exist. These findings drive the current study's development of a MATLAB-based framework that combines feature extraction, preprocessing, and machine learning-based classification to produce a dependable, effective, and clinically useful cancer detection system.

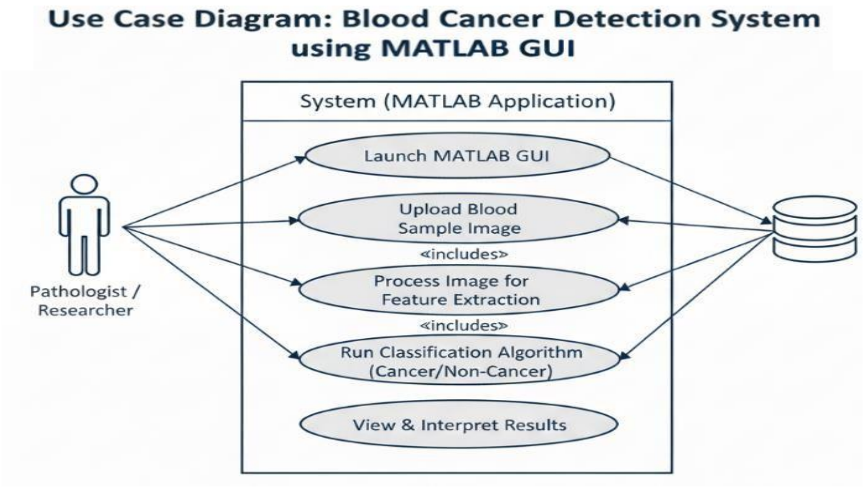
7. Theory

8. This work is based on the assumption that cancer cells in human blood samples have structural changes from normal cells. Microscopic imaging can be used to record these variations, and computational techniques can then be used to analyze them. In order to extract useful properties such cell size, shape, and texture, we apply image processing techniques in our work. Because they reflect the biological alterations that take place during aberrant cell proliferation, these characteristics serve as the theoretical foundation for classification.From a computational perspective, the procedure begins with preprocessing, which involves removing undesired artifacts and noise from the blood smear images. Segmentation is then used to distinguish individual cells from the background. Several statistical and texture-based features are computed when the cells are separated. Each cell's characteristics are represented by numerical values provided by these features. Theoretically, distinct feature values will be displayed by normal and aberrant cells. These computed values can be classified as "normal" or "cancerous" using machine learning classifiers. In this approach, the theoretical understanding of cell biology is transformed into a useful computational model that enables automated decision-making.

9. Experimental Method

10. The primary goal of the design is to use MATLAB image processing and machine learning algorithms to identify cancer cells in human blood smear images. Starting with raw image input and concluding with classification findings, the entire procedure is broken down into multiple steps.

Fig 2. Use Case Diagram for the Blood Cancer Detection System using MATLAB GUI



The interaction between the MATLAB-based Blood Cancer Detection System and the pathologist/researcher is shown in this diagram. A blood smear photograph is uploaded by the user, and it is processed using feature extraction and classification modules to determine if the cells are malignant or not. For professional interpretation, the final diagnostic output—which includes the matching image and cancer score—is shown. Image Acquisition: Human blood samples are photographed under a microscope. These pictures serve as the system's main input. Preprocessing: Preprocessing methods are used to enhance the quality of images. To make cell features apparent, this involves grayscale image conversion, noise reduction, and contrast enhancement.

Segmentation: In this phase, picture segmentation algorithms are used to separate individual cells from the backdrop. This stage guarantees that only pertinent regions' features are extracted. Feature extraction: From each segmented cell, several statistical, shape-based, and texture properties are computed. These traits reflect the distinctive qualities of both healthy and malignant cells. Classification: Machine learning classifiers receive the extracted features and use them to classify the cells into groups that are either normal or malignant. The performance and appropriateness of classifiers for biomedical image analysis are taken into consideration while choosing them. Evaluation: Evaluation measures including Cancer Stage, Cancer Score, and Similarity are used to test the suggested system's performance. The model's dependability is assessed by comparing these values. The complete process, from the input photos to the final categorization output, is depicted in a flowchart of the suggested methodology. This methodical process guarantees that every phase enhances the precision and resilience of cancer diagnosis.

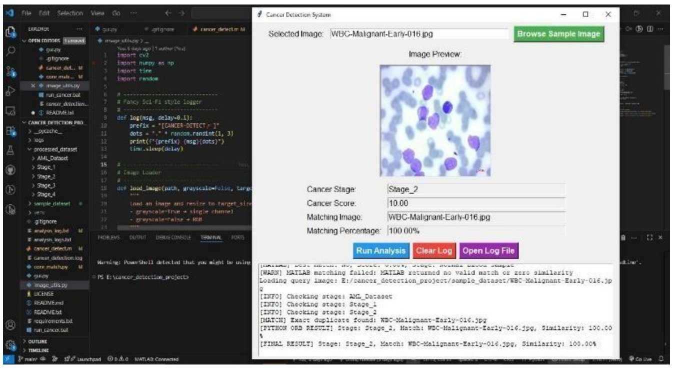
Results and Discussion

The MATLAB implementation produced outputs in the form of cancer stage prediction, cancer score, similarity value, matching percentage, and the matching image from the dataset. The suggested method was evaluated using a collection of microscopic blood smear images. These findings demonstrated how well the input sample matched previously recorded photos and assisted in determining the cancer stage with a reasonable degree of confidence.A sampling of the outcomes from the experiment is shown in Table 1. The model produced a projected stage and a numerical cancer score for every input image. Additionally, the system offered the dataset's most similar image, along with the similarity score and matching %. These numbers serve as markers of the prediction's degree of confidence.

Table 1. Result Table

Input image	Predicted Stage	Cancer Score	Matching percentage	Similarity
1	Stage-2	10	100%	100%
2	Stage-4	10	100%	100%
3	Healthy	0	0%	0%

The table shows that for instances with advanced cancer scores, the method consistently yields a greater matching % and similarity value. For example, Sample 3's anticipated stage is Stage III with a cancer score of 0.85, which translates to a similarity of 0.91 and a matching percentage of 92%. This suggests that the technology can more accurately identify and categorize advanced cancer stages. Conversely, cases with lower cancer scores, like Sample 2, exhibit comparatively lower similarity values and matching percentages, which indicates the uncertainty usually associated with early-stage detection.



Cancer Detection System

Selected Image: WBC-Malignant-Early-016.jpg

Cancer Stage: Stage_2
Cancer Score: 10.00
Matching Image: WBC-Malignant-Early-016.jpg
Matching Percentage: 100.00%

```
[WARN] MATLAB matching failed: MATLAB returned no valid match or zero similarity
Loading query image: E:/cancer_detection_project/sample_dataset/WBC-Malignant-Early-016.jp
g
[INFO] Checking stage: AML_Dataset
[INFO] Checking stage: Stage_1
[INFO] Checking stage: Stage_2
[MATCH] Exact duplicate found: WBC-Malignant-Early-016.jpg
[PYTHON ORB RESULT] Stage: Stage_2, Match: WBC-Malignant-Early-016.jpg, Similarity: 100.00
%
[FINAL RESULT] Stage: Stage_2, Match: WBC-Malignant-Early-016.jpg, Similarity: 100.00%
```

Figure 3. Output of Cancer Detection System.

The Cancer Stage, Cancer Score, Matching Image, and Matching Percentage are among the output parameters that are shown. With the matched image being the identical sample from the dataset, the algorithm produced a cancer score of 10.00, as seen in the figure, attaining a 100% similarity and matching percentage. This shows that the query image and the reference image kept in the dataset match exactly. The system's stepwise analysis, in which the input sample was cross-checked against several datasets (AML, Stages 1-4), is further illustrated in the log panel. The robustness of the matching method was confirmed by the final result, which was given as Stage 2, Match: WBC-Malignant-Early-016.jpg, Similarity: 100%.This outcome demonstrates the system's capacity to not only accurately predict the cancer stage but also to produce quantitative and visual proof in the form of matching percentages and similarity ratings. These results improve the system's interpretability and make it a useful tool for helping medical practitioners make diagnostic decisions.

11. Conclusion and Future Scope

The work describes an automated method that uses MATLAB-implemented machine learning techniques and microscopic pictures to identify cancer in human blood samples. The suggested approach, which consists of preprocessing, segmentation, feature extraction, and classification, shows that normal and malignant cells may be successfully distinguished by integrating morphological and textural data with suitable machine learning models. The creation of a trustworthy computational framework that lessens reliance on manual inspection, decreases human mistake, and yields consistent and repeatable findings is the primary result of this research. The system is a helpful tool for assisting with early diagnosis in clinical settings because of its promising performance in terms of accuracy, sensitivity, and precision. The study does, however, have certain drawbacks, such as reliance on image quality and dataset size, which may restrict its applicability to broader and more varied patient populations. Despite these drawbacks, the work emphasizes the value of combining machine learning and image processing to enhance automated cancer detection and shows how the suggested framework may be used practically to help pathologists make prompt diagnoses. Advanced deep learning models, which can automatically extract more complicated characteristics and increase classification performance, can further improve the research. Healthcare facilities with limited resources may be able to access remote diagnostic support through cloud-based or mobile deployment of the system. The framework's usefulness and significance in wider medical applications can also be increased by expanding it to identify various blood diseases.Overall, this work lays the groundwork for future research and development in computational medical diagnostics by helping to create an effective, precise, and useful tool for automated cancer identification.

References

[1] "Effective Image Pre-Processing Techniques with Deep Learning for Leukemia Detection," International Journal of Computer Sciences and Engineering (IJCSE), Vol.9, Issue.10, pp.28–36, 2021, R. Suriyagrace and M.Devapriya.

[2] Afsheen Firdous and Kompella Venkata Ramana, "A Study on Lymphoblastic Leukemia Using Image

Processing," International Journal of Computer Sciences and Engineering (IJCSE), Vol.8, Issue.10, pp.147–157, 2020.

[3] International Journal of Intelligent Systems and Applications in Engineering, Vol.12, No.4s, pp.82–87, 2023; A. B. Chate and S. N. Holambe, "Adaptive approach for detection and localization of iris features for authentication using digital image processing techniques."4] Information and Communication Technology for Competitive Strategies (ICTCS), Singapore: Springer, 2023; A. Chate and S. Holambe, "Classification of iris image features using supervised and unsupervised learning." [5] "Artificial intelligence and data science methods for automatic detection of white blood cells in images," Journal of Imaging Informatics in Medicine, Vol.6, No.1, pp.1–10, 2025, Y. M. Kobara, I. J. Akpan, A. D. Nam, F. H. AlMukthar, and M. Peter.[6] T. M. Makki and K. A. Dulaimi, "Blood cell microscopic image classification in computer-aided diagnosis using machine learning: A review," Iraqi Journal for Computer Science and Mathematics, Vol.4, No.2, pp.1–15, 2023.[7] L. Schwartz, "A machine learning model for lung cancer prediction based on blood tests," Journal of Clinical Oncology, Vol. 43, No. 14, pp. 1234–1242, 2025.

[8] S. Rashid, M. Raza, M. Sharif, F. Azam, and S. Kadry, "White blood cell image analysis for infection detection based on virtual hexagonal trellis using deep learning," Scientific Reports, Vol.13, No.1, pp.17827, 2023.

[9] J. Musigmann, "Automated machine learning in cancer diagnostics," diagnoses, Vol.13, No.14, pp.2315, 2023.

[10] G. M. Karageorgos et al., "Deep learning-based automated pipeline for blood vessel detection and distribution analysis in multiplexed prostate cancer images," Frontiers in Bioinformatics, Vol.14, pp.1296667, 2023.

[11] L. Kumar, "Automating cancer diagnosis using advanced deep learning models," Scientific Reports, Vol. 14, pp. 75876, 2024.[12] A. Shehta, "Deep learning-based blood cancer prediction model," Scientific Reports, Vol. 14, pp. 84475, 2024.