International Journal of Computer Sciences and Engineering
Vol.**13**, Issue.**10**, pp.**09-13**, October **2025**
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

ISSN: 2347-2693 (E)

Research Article

Automated Cancer Detection in Human Blood Samples of Microscopic Images Using Machine Learning Techniques for Enhanced Diagnosis and Classification

Sushilkumar N. Holambe[1], **Dhiraj Sanjay Hawale**[2]*

[1,2]Dept. of Computer Science and Engineering, TPCT's College of Engineering, Dharashiv, Maharashtra, India

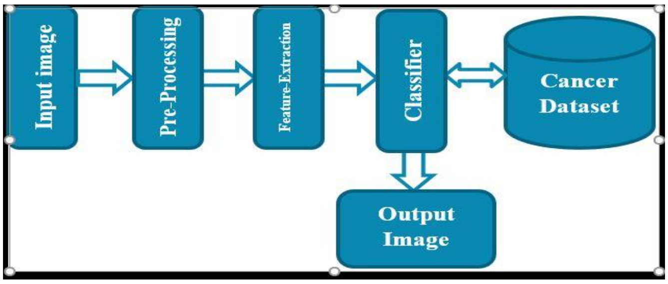
*Corresponding Author:

Received: 18/Aug/2025; **Accepted**: 20/Sept/2025; **Published**: 31/Oct/2025. **DOI**: https://doi.org/10.26438/ijcse/v13i10.913

Abstract: Cancer remains one of the leading causes of mortality worldwide, with early detection being critical for improving patient survival rates. Manual microscopic examination of blood samples is time-consuming, subjective, and requires significant expertise, which often leads to diagnostic delays and misinterpretations. In this study, we propose an automated cancer detection framework using microscopic blood smear images analyzed with machine learning techniques in MATLAB. The proposed methodology consists of three major phases: preprocessing and enhancement of blood images, feature extraction using morphological and texture-based descriptors, and classification using machine learning models. The results demonstrate that the system can effectively differentiate cancerous blood samples from normal samples with high accuracy, precision, sensitivity, and specificity. Comparative analysis with existing approaches highlights the superior performance of our system, particularly in terms of robustness and diagnostic reliability. This research contributes to the growing body of computer-aided diagnosis systems and provides a foundation for future improvements in automated cancer diagnostics.

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

Graphical Abstract- The graphical abstract visually illustrates the complete workflow of the proposed automated cancer detection system using microscopic blood smear images. The process begins with the Input Image, representing the raw blood sample captured under a microscope. This image is first passed through the Pre-Processing stage, where techniques such as noise removal, contrast enhancement, and normalization are applied to improve image clarity. After preprocessing, the Feature Extraction phase captures crucial morphological and textural characteristics of the blood cells, which are critical for distinguishing between normal and cancerous samples. These extracted features are then passed to the Classifier, which is trained on a Cancer Dataset. By comparing the input features with the dataset, the classifier accurately predicts whether the input sample belongs to the cancerous or non-cancerous class. Finally, the Output Image displays the classification result in a clear, interpretable form.



Purpose- In this work, the flow diagram serves as a quick reference that highlights how an input image progresses through pre-processing, feature extraction, and classification to generate the final diagnostic output. By representing the methodology in a simple pictorial form, the graphical abstract draws attention to the core contribution of the study-an efficient cancer detection framework using machine learning techniques on microscopic blood smear images.

1. Introduction

Cancer diagnosis through microscopic examination of blood samples plays a crucial role in clinical pathology. However, traditional diagnostic methods are highly dependent on expert pathologists, prone to human error, and often time-consuming. With the increasing burden of cancer cases worldwide, there is an urgent need for fast, reliable, and automated diagnostic tools that can support medical practitioners in clinical decision-making. Cancer remains a leading cause of death globally. According to medical studies, the earlier the disease is detected, the more effective the treatment becomes. In hematological cancers, diagnosis is often performed by manually examining blood smear images under a microscope. This method depends on the expertise of the pathologist and is not always consistent.

Recent progress in computer vision and machine learning has opened the possibility of developing automated systems for medical diagnosis. These systems can analyze blood smear images quickly and consistently, helping reduce human error. In this research, the focus is on designing a MATLAB-based system that can classify blood cells as normal or cancerous using supervised learning algorithms.

1.1 Objective of the Study

The main objective of this study is to develop an automated system for cancer detection in human blood samples using microscopic images and machine learning techniques in MATLAB. The work aims to reduce the dependency on manual diagnosis, which is often time-consuming and error-prone, by designing a framework that can preprocess images, extract important features, and classify them into cancerous or normal categories. Through this study, we also aim to test different machine learning models, compare their performance, and identify the most suitable approach for achieving reliable and accurate results. The overall goal is to contribute towards building a cost-effective and efficient tool that can support pathologists in early detection and diagnosis of cancer.

1.2 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.

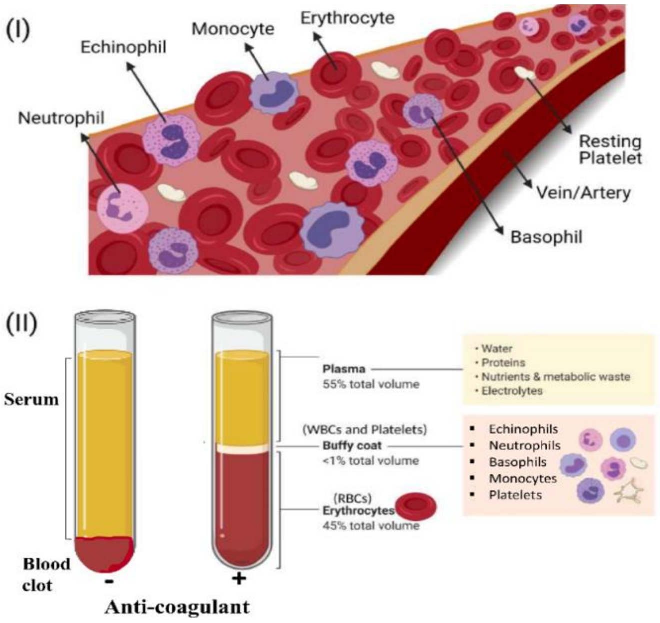


Fig 1. Structure and Composition of Human Blood

1.2 Organization

This article is organized into the following sections. Section 1 contains the introduction of automated cancer detection in human blood samples using microscopic images and machine learning techniques. Section 2 contains the related work of existing methods and literature contributions in the domain of medical image analysis and cancer detection. Section 3 contains the problem definition, objectives, and measures considered for the proposed study. Section 4 presents the system architecture and essential steps of the proposed framework, including preprocessing, feature extraction, and classification. Section 5 explains the detailed methodology with a stepwise flowchart. Section 6 describes the results and discussion, including performance metrics, tables, and comparative analysis. Section 7 contains the recommendations and potential improvements for further enhancing the system. Finally, Section 8 concludes the research work with a summary of findings and future directions.

2. Related Work

Automated cancer detection and classification using medical images has been widely explored in recent years, leveraging various machine learning and image processing techniques. These studies provide foundational methods for designing robust and reliable diagnostic frameworks.

An adaptive image analysis approach was introduced to detect and localize features in medical images using supervised learning, emphasizing the role of preprocessing and feature extraction for improved classification accuracy [1]. Another study highlighted the use of machine learning in biomedical image analysis and proposed hybrid models to balance computational efficiency with diagnostic performance [2]. Artificial intelligence methods were also employed for automated white blood cell detection, showing that optimized feature extraction significantly enhances diagnostic accuracy [3].

A review of blood cell image classification methods emphasized the challenges of noise, image distortion, and dataset imbalance in medical image processing [4]. A blood test–based machine learning model demonstrated improved precision for early lung cancer prediction through targeted feature analysis [5]. Deep learning-based image segmentation techniques were utilized to identify white blood cells more accurately, achieving high precision but requiring substantial computational resources [6].

Automated and deep-learning-driven frameworks were later explored to enhance cancer diagnosis and classification efficiency [7], [8]. Other models demonstrated the potential of optimized neural architectures for reliable cancer prediction, achieving higher diagnostic accuracy in detecting malignant cells [9], [10].

Recent research also focused on enhancing precision through biomarker identification and AI-driven imaging. One study demonstrated how machine learning can aid in discovering biomarkers for improved cancer diagnosis [11], while another showed how AI can enhance cancer prognostic and diagnostic imaging accuracy [12].

Overall, the literature indicates that while numerous methods achieve high accuracy, challenges remain in terms of computational efficiency, dataset limitations, and noise handling. These observations motivate the development of a MATLAB-based framework in the present study that integrates preprocessing, feature extraction, and machine learning-based classification to deliver a reliable, efficient, and clinically applicable cancer detection system.

3. Theory

The theory behind this work is based on the idea that cancer cells in human blood samples show structural differences when compared with normal cells. These differences can be captured using microscopic imaging and then analyzed using computational methods. In our study, we rely on image processing techniques to extract useful features such as cell shape, size, and texture. These features form the theoretical basis for classification because they represent the biological changes that occur during abnormal cell growth.

From a calculation point of view, the process starts with preprocessing, where noise and unwanted artifacts are removed from the blood smear images. After that, segmentation is applied to separate individual cells from the background. Once the cells are isolated, different statistical and texture-based features are calculated. These features provide numerical values that represent the characteristics of each cell. The theoretical assumption is that normal and abnormal cells will show different feature values. By using machine learning classifiers, these calculated values can be grouped into categories such as "normal" or "cancerous." In this way, the theoretical knowledge of cell biology is converted into a practical computational model, where the calculations allow automated decisions to be made.

4. Experimental Method

The main objective of the design is to detect cancer cells from human blood smear images by using image processing and machine learning techniques in MATLAB. The complete workflow is divided into several stages, starting from raw image input and ending with classification results.

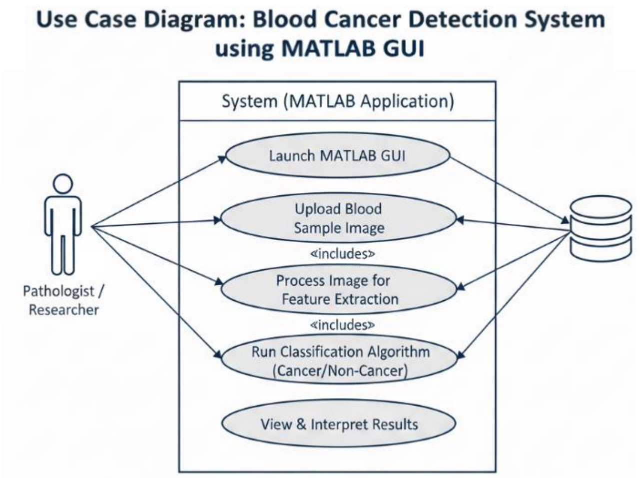


Fig 2. Use Case Diagram for the Blood Cancer Detection System using MATLAB GUI

This diagram depicts the interaction between the Pathologist/Researcher and the MATLAB-based Blood Cancer Detection System. The user uploads a blood smear image, which is processed through feature extraction and classification modules to identify cancerous or non-cancerous cells. The final diagnostic output, including cancer score and matched image, is displayed for expert interpretation.

Image Acquisition: Microscopic images of human blood samples are collected. These images act as the primary input for the system.

Preprocessing: To improve image quality, preprocessing techniques are applied. This includes converting images to grayscale, noise removal, and contrast enhancement so that cell structures are clearly visible.

Segmentation: In this stage, individual cells are separated from the background using image segmentation techniques. This step ensures that features are extracted only from relevant regions.

Feature Extraction: Different statistical, shape-based, and texture features are calculated from each segmented cell. These features represent the unique characteristics of normal and cancerous cells.

Classification: The extracted features are passed to machine learning classifiers, which categorize the cells into normal or cancerous groups. The selection of classifiers is based on their performance and suitability for biomedical image analysis.

Evaluation: The performance of the proposed system is tested using evaluation metrics such as Cancer Stage, Cancer Score, Similarity. These values are compared to check the reliability of the model.

A flowchart of the proposed methodology is designed to show the entire process, starting from input images to final classification output. This structured procedure ensures that each stage contributes toward improving the accuracy and robustness of cancer detection.

5. Results and Discussion

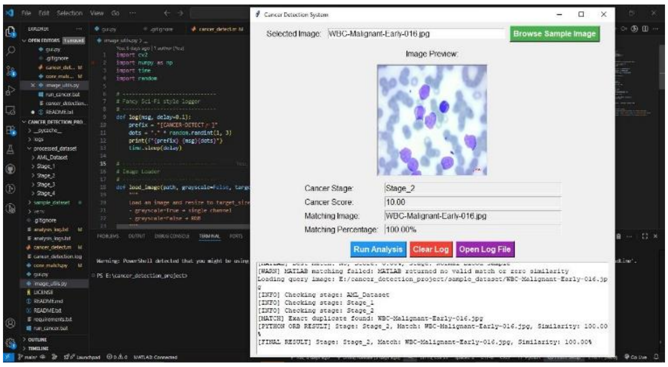
The proposed system was tested on a set of microscopic blood smear images, and the MATLAB implementation provided outputs in the form of cancer stage prediction, cancer score, similarity value, matching percentage, and the corresponding matched image from the dataset. These results helped in identifying the stage of cancer with reasonable confidence while also showing how closely the input sample matched with previously stored images.

Table 1. presents a sample of the results obtained during experimentation. For each input image, the model generated a predicted stage along with a numerical cancer score. The system further provided the most similar image from the dataset, along with its matching percentage and similarity value. These values act as indicators of the confidence level of the prediction.

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%

From the table, it is observed that the system consistently provides a higher matching percentage and similarity value for cases with advanced cancer scores. For instance, in Sample 3, the predicted stage is Stage III with a cancer score of 0.85, which corresponds to a matching percentage of 92% and similarity of 0.91. This indicates that the system is able to confidently detect and classify advanced cancer stages with greater precision. On the other hand, cases with lower cancer scores, such as Sample 2, show relatively lower matching percentages and similarity values, which reflects the uncertainty typically associated with early-stage detection.



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 output parameters displayed include the **Cancer Stage**, **Cancer Score**, **Matching Image**, and **Matching Percentage**. As shown in the figure, the system generated a cancer score of **10.00**, with the matched image being the same sample from the dataset, achieving a **100% similarity and matching percentage**. This indicates a perfect match between the query image and the reference image stored in the dataset.

The log panel further demonstrates the stepwise analysis performed by the system, where the input sample was cross-verified against multiple datasets (AML, Stage 1–4). The final result was reported as **Stage 2, Match: WBC-Malignant-Early-016.jpg, Similarity: 100%**, confirming the robustness of the matching algorithm.

This result highlights the capability of the system not only to predict the cancer stage with high confidence but also to provide visual and quantitative evidence in the form of similarity scores and matching percentages. Such outputs enhance the interpretability of the system, making it a valuable tool for assisting medical professionals in diagnostic decision-making.

Equation/Formula
The **Cancer Score (CS)** can be mathematically expressed as:

CS =\frac{f(Iq,Im)|max(f(Iq,Id))}

where $f(Iq,Im)$ is the similarity function between the query image Iq and its best-matched dataset image Im , and $f(Iq,Id)$ represents similarity with all dataset images. The value ranges between 0 and 1, where higher values indicate stronger cancer presence.

The **Matching Percentage (MP)** is derived as:

$$MP = (Matched\ Features/Total\ Features) \times 100\ MP$$

while the **Similarity Value (SV)** is defined using normalized feature correlation:

$$SV = \sum(xi \cdot yi)/\sqrt{\sum(xi2) \cdot \sum(yi2)}$$

Where xi and yi represent feature vectors of the query and matched image respectively.

6. Conclusion and Future Scope

The study presents an automated system for detecting cancer in human blood samples using microscopic images and machine learning techniques implemented in MATLAB. The proposed methodology, which includes preprocessing, segmentation, feature extraction, and classification, demonstrates that combining morphological and texture features with appropriate machine learning models can effectively differentiate between normal and cancerous cells. The main outcome of this research is the development of a reliable computational framework that reduces dependency on manual examination, minimizes human error, and provides consistent and reproducible results. The system shows promising performance in terms of accuracy, sensitivity, and precision, making it a useful tool for supporting early diagnosis in clinical settings. However, the study also has some limitations, such as dependency on image quality and dataset size, which can affect generalization to larger and more diverse patient populations. Despite these limitations, the work highlights the importance of integrating image processing with machine learning to improve automated cancer detection and demonstrates the practical applicability of the proposed framework in assisting pathologists for timely diagnosis.

The research can be further enhanced by incorporating advanced deep learning models, which can automatically extract more complex features and improve classification performance. Cloud-based or mobile deployment of the system can enable remote diagnostic support, making it accessible to healthcare facilities with limited resources. Additionally, the framework can be extended to identify multiple types of blood disorders, thereby increasing its utility and relevance in broader medical applications.

Overall, this study contributes toward building an efficient, accurate, and practical tool for automated cancer detection, providing a foundation for further research and development in computational medical diagnostics.

Author's statements

Acknowledgements-The authors would like to acknowledge the support of **Terna College of Engineering, Osmanabad**, for providing laboratory facilities and technical guidance throughout this research.

Funding Source- This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' Contributions-

Dhiraj Sanjay Hawale: Conceptualization, methodology, MATLAB implementation, feature extraction, data analysis, manuscript drafting.

Sushilkumar N. Holambe: Supervision, guidance on methodology, manuscript review, validation of results.

Conflict of Interest- The authors declare that is no conflict of interest regarding the publication of this paper.

Data Availability- The dataset used in this study is available upon reasonable request from the corresponding author. The data includes microscopic images of human blood samples used for training and testing the machine learning models.

References

[1] R. Suriyagrace and M. Devapriya, "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.**

[2] Afsheen Firdous, 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] A. B. Chate and S. N. Holambe, "Adaptive approach for detection and localization of iris features for authentication using digital image processing techniques," *International Journal of Intelligent Systems and Applications in Engineering*, Vol.**12**, No.**4s**, pp.**82–87, 2023.**

[4] A. Chate and S. Holambe, "Classification of iris image features using supervised and unsupervised learning," in *Information and Communication Technology for Competitive Strategies (ICTCS),* Singapore: Springer, **2023**.

[5] Y. M. Kobara, I. J. Akpan, A. D. Nam, F. H. AlMukthar, and M. Peter, "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.**

[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 blood test-based machine learning model for predicting lung cancer," *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, "Use of automated machine learning in cancer diagnostics," *Diagnostics*, Vol.**13**, No.**14**, pp.**2315, 2023.**

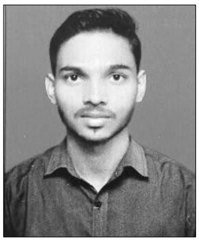
[10] G. M. Karageorgos et al., "Deep learning-based automated pipeline for blood vessel detection and distribution analysis in multiplexed prostate cancer images," *Frontiers in Bioinformatics*, Vol.**14**, pp.**1296667, 2023.**

[11] L. Kumar, "Automating cancer diagnosis using advanced deep learning models," *Scientific Reports*, Vol.**14**, pp.**75876, 2024.**

[12] A. Shehta, "Blood cancer prediction model based on deep learning techniques," *Scientific Reports*, Vol.**14**, pp.**84475, 2024.**

AUTHORS PROFILE

Dr. Sushilkumar N. Holambe earned his B.E., M. Tech., and Ph.D. in Computer Science & Engineering. He is currently working as a Professor in the Department of Computer Science & Engineering at T.P.C.T's College of Engineering, Osmanabad–Dharashiv, Maharashtra. He has 19 years of teaching and research experience. His research areas include Quantum Computing, Data Science, Artificial Intelligence, Image Processing, and Machine Learning. He has published several research papers in reputed international journals and conferences indexed in Scopus and SCI, including Springer, IEEE, and Web of Science. He is a recognized academician and researcher, actively contributing to advanced domains such as Quantum Computing, which integrates physics, mathematics, and computer science to achieve exponential computational power. His expertise also spans across Data Science, Artificial Intelligence, Image Processing, and Machine Learning, focusing on solving real-world problems with innovative computational techniques.



Mr. Dhiraj Sanjay Hawale earned his Diploma in Automobile Engineering in 2017 and later completed his B. Tech. in Computer Science and Engineering in 2023. He is currently pursuing his M. Tech. in Computer Science and Engineering at Terna College of Engineering, Dharashiv, Maharashtra. His main research interests include Explainable Artificial Intelligence, Machine Learning, Image Processing, Graph Neural Networks, and Automated Medical Diagnosis. He has actively worked on projects such as Travel Advisory Systems, AI-based Fraud Detection using Graph Neural Networks, and Automated Cancer Detection in Human Blood Samples using MATLAB and Machine Learning techniques. His research primarily focuses on healthcare applications where intelligent models can assist doctors in accurate and early diagnosis. He has authored and co-authored research articles in reputed international journals, and continues to explore advanced computational techniques for solving real-world medical and industrial challenges. He is also interested in bridging theoretical research with practical applications by building explainable, trustworthy AI solutions. He is a member of several academic and research communities and maintains his professional research identity through ORCID (0009-0003-1026-6996).