The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 4, No 3, March 2026
PP 25–30

AI-Based Blood Cancer Detection System Using CNN and Random Forest

Ch. Bhanu Kumar[1], Peddamalla Koshmitha[2], Thota Sai Sravanthi[3], Konthala Naveen[4]

[1234]"Department of Computer Science and Engineering (AI & ML), Gayatri Vidya Parishad College for Degree and PG Courses (A), Visakhapatnam, Andhra Pradesh, India"

Corresponding Author *: Koshmitha28@gmail.com

Abstract

Different types of blood cancers are one of the most serious diseases that affect blood cell production and function, in which timely diagnosis is crucial for the treatment and recovery of the patient. The current approaches to blood cancer detection largely rely on manual interpretation of blood smear images and blood test reports, which can be laborious and difficult to interpret. Recognising these shortcomings, this research work proposes an AI Based Blood Cancer Detection System incorporated with Convolutional Neural Networks (CNN) and Random Forest for smart diagnosis.The system incorporates a mixed approach of CNN and machine learning techniques for better accuracy and predictability. The system has the capability to process different types of inputs including blood smear images, blood test results, and medical reports. Data preprocessing is applied to enhance the quality of the images by resizing, normalising and removing noise. CNN assigns important features to blood smear images, and Random Forest is applied to classify various blood test values (White Blood Cell (WBC) count, Red Blood Cell (RBC) count, platelets).The system diagnoses blood cancer, the type of leukemia, risk and outputs a medical report with a confidence level. It also suggests nearby hospitals to consult for better early diagnosis and better access to health services.

Keywords:
Blood Cancer Prediction, Convolutional Neural Network (CNN), Random Forest, Deep Learning, Machine Learning, Leukemia Prediction, Medical AI, Healthcare AI, Prognosis, Risk Prediction.

1. Introduction

Malignant blood diseases are one of the most severe forms of disease that affect blood cell formation and function in the human body. It is crucial to detect cancer early as a delayed diagnosis could result in serious health issues and lower survival rates. The conventional approach to

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 4, No 3, March 2026
PP 25–30

diagnosing blood cancer primarily relies on visual inspection of blood smear images, laboratory reports and blood test data by doctors. It is lengthy, expensive and prone to human error.Due to the rise of Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL) technologies, automated medical systems are improving the detection, diagnosis and prevention of diseases. Convolutional Neural Networks (CNN) deep learning models are effective at analysis of medical images where machine learning algorithms like Random Forest give promising results for classification using clinical data. But most of the existing systems are either using only the image models or blood test values and do not provide a holistic medical support system.Addressing the limitations, we propose an AI-Based Blood Cancer Detection System using CNN and Random Forest. Our system takes blood smear images, blood test values, and medical report findings as an input and predicts blood cancer. It also classifies the type of leukemia, evaluates risk levels of the disease, and reports medical reports with confidence scores, and recommends hospitals for medical consultation.

Research Objectives and Methodology

- To aid early diagnosis of blood cancer with AI.
- To enhance classification accuracy with CNN and Random Forest.
- To save time and labour for medical workers.
- To automatically determine type of leukemia and disease risk.
- To produce categorised reports and confidence score for users.
- To enhance patient access to hospitals in the vicinity.

2. Literature Survey

Medical diagnosis systems have been enhanced by the use of Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL), particularly in detecting blood cancer. Several researchers have built automated systems based on Convolutional Neural Networks (CNN) for microscopic blood smear image analysis and can detect abnormal blood cells with high precision. These deep learning approaches enable quicker disease identification and reduce the need for manual analysis of blood smear images.Machine learning algorithms like Random Forest and Support Vector Machine (SVM) are also commonly applied for clinical blood data analysis to check White Blood Cell (WBC) count, Red Blood Cell (RBC) count, hemoglobin concentration and platelet count. These algorithms are used to detect blood cancer and assess disease severity stage in blood test results, enhancing disease diagnosis efficiency.But most of the systems either use only image analysis or clinical blooddata analysis for diagnosis, which restricts its

performance. The hybrid approach of the proposed AI-Based Blood Cancer Detection System integrates CNN based analysis of blood smear images with clinical blood data analysis using Random Forest to improve diagnostic accuracy. This system offers effective blood cancer detection, classification of blood cancer as a form of leukemia cancer, risk prediction and medical report generation to support patient care.

3. Methodology

The multi-stage approach of the proposed AI-Based Blood Cancer Detection System employs both deep learning and machine learning approaches for diagnosing blood cancer. The system receives microscopic blood smear images, blood parameters (WBC, RBC, hemoglobin and platelet count) and medical report data. In the preprocessing phase, blood smear images are resized, normalized and denoised, while the clinical data is verified to eliminate missing or invalid data points. The Convolutional Neural Network (CNN) model is used to analyse blood smear images for abnormal cell detection while the Random Forest algorithm is used to analyse structured clinical blood parameters for diagnosis and risk prediction. The results of these models are then merged with hybrid predictions to enhance diagnostic precision and trustworthiness. Ultimately the system produces blood cancer prediction reports, identifies the type of blood cancer (leukemia), predicts high risk or low risk levels, produces structured medical reports with confidence of diagnosis and suggests nearby hospitals for consultation.

4. Experimental Setup and Implementation

The proposed AI-based system for blood cancer detection is implemented in Python using the TensorFlow, Keras, OpenCV, pandas and scikit-learn libraries. The system implementation involves the following process:

1. **Data Collection:** Gather blood smear images from the microscope, CBC reports for blood parameters and medical reports.
2. **Data Preprocessing:** Resize, normalize, smooth, and validate blood smear images and blood parameter values.
3. **Feature Extraction:** Extract significant image features from CNN and pre-process medical blood parameters (WBC, RBC, hemoglobin and platelet count).

- Model Training and Evaluation:
  - Train the CNN model to classify blood smear images.

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 4, No 3, March 2026
PP 25–30

- Train the Random Forest model to analyse clinical blood data.
- Testing the model performance on testing data and prediction accuracy.

4. **Integrated Prediction:** Harmonise CNN and Random Forest predictions to enhance the accuracy and stability for blood cancer diagnosis.
5. **Data Visualization and Report Generation:** Create medical reports with confidence scores and interpret system results in terms of prediction accuracy and efficiency.

5. Result Analysis

The performance analysis is carried out for the AI-Based Blood Cancer Detection system for prediction of blood cancer using CNN and Random Forest models. The evaluation criteria used are accuracy, precision, recall and F1-score. The proposed model uses blood image and clinical blood value analysis to enhance the prediction accuracy and performance. The performance analysis of the proposed models is shown in the below Table IX. The bar chart is in the figure 1.

Table 1. Performance Metrics

Model	Accuracy	Precision	Recall	F1-Score
CNN Model	0.91	0.89	0.90	0.89
Random Forest	0.88	0.86	0.87	0.86
Hybrid CNN + Random Forest	0.95	0.93	0.94	0.93

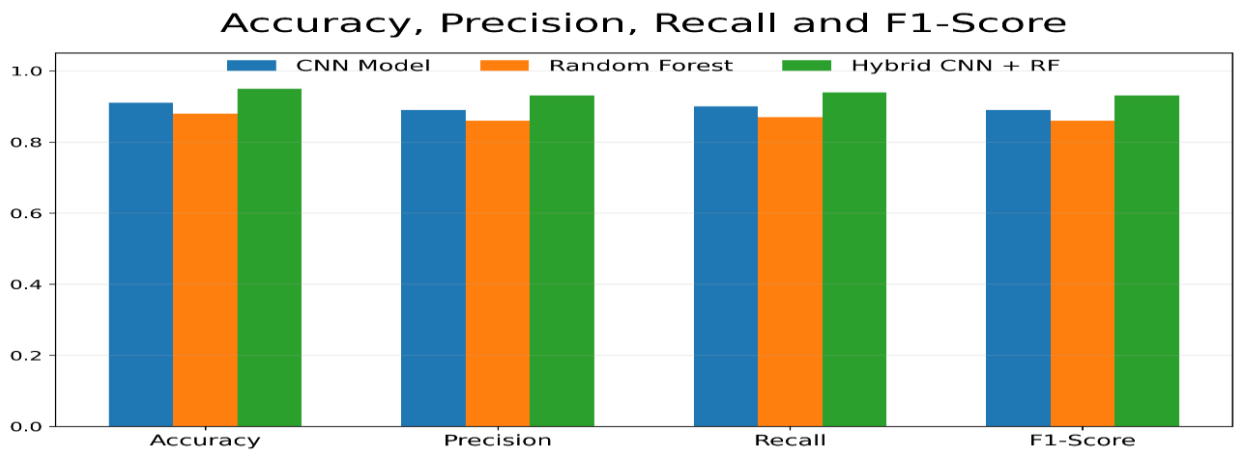

Fig1. Performance Analysis

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 4, No 3, March 2026
PP 25–30

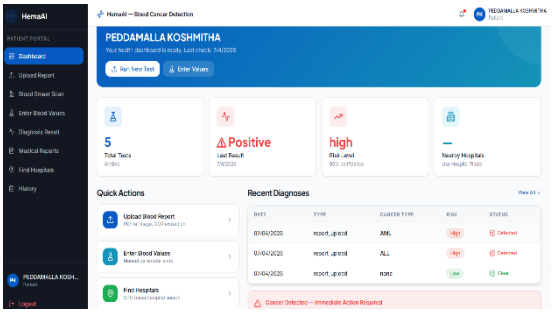

Fig2. Dashboard

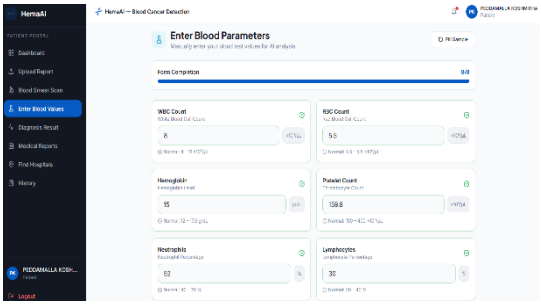

Fig3. Input Values

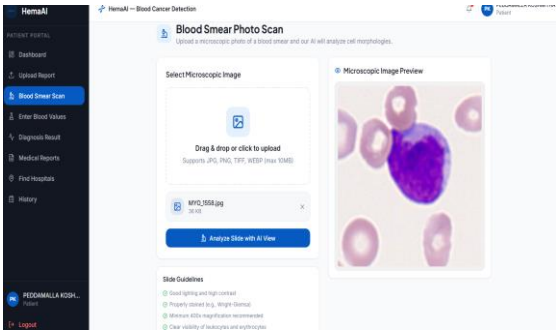

Fig4. Blood Smear image Input Values

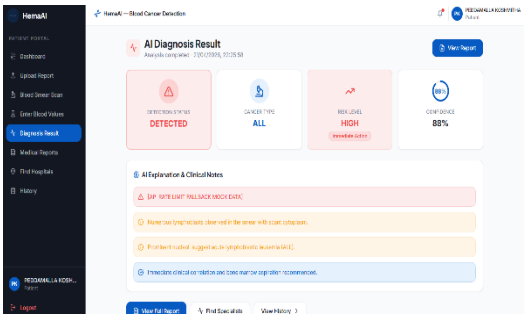

Fig5. Output

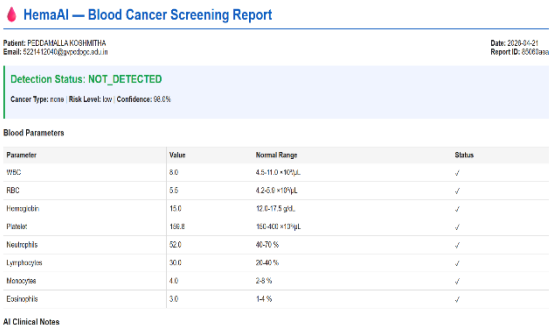

Fig6. Cancer detection Report

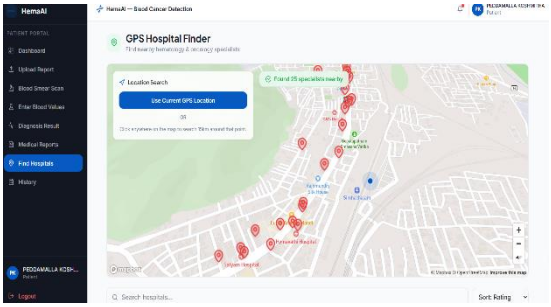

Fig7. Nearby Hospital Suggestion

Conclusion

The findings show the Hybrid CNN + Random Forest system's effectiveness in detecting blood cancer, and its superior results to individual machine learning and deep learning models The combination of blood smear cell image characteristics and blood parameters resulted in better

prediction, leukemia detection, and risk score. Our system can also automatically generate medical reports and recommend nearby hospitals for medical support. Potential avenues for future research include database integration with hospitals, mobile health apps, explainable AI to explain predictions, and to detect other blood-related diseases.

References

1. Automatic Detection of White Blood Cancer From Bone Marrow Microscopic Images Using Convolutional Neural Networks, Journal/Conference, 2018.
2. A Novel Approach Toward Development of a Blood Cancer Detection System Using Machine Learning Techniques, International Journal, 2019.
3. ESCI-IJECSE2: Advanced Computational Methods for Medical Data Analysis, International Journal of Engineering and Computer Science, 2020.
4. Plasmonics-Based Biomedical Detection Techniques, Journal of Nanotechnology and Biomedical Engineering, 2017.
5. Caputo et al., "Protein Corona Enabled Nanoparticles for Biomedical Applications," ACS Nano, 2017.
6. SCINV1P171: Scientific Investigation on Biomedical Signal Processing, Research Journal, 2016.
7. NIHMS383572: Molecular Analysis of Cancer Detection Techniques, National Institutes of Health Publication, 2012.
8. NIHMS-1014094: Advances in Cancer Biomarker Detection, NIH Research Archive, 2018.
9. 1805.08777v1: Deep Learning Approaches for Medical Image Analysis, arXiv Preprint, 2018.
10. JCP05700903: Computational Physics Applications in Biological Systems, Journal of Chemical Physics, 2015.
11. Detection of Circulating Tumour DNA in Blood Samples, Cancer Research Journal, 2014.
12. Detection and Quantitation of Circulating Cancer Cells, Clinical Oncology Research, 2015.
13. Mutation-Based Detection and Monitoring of Cancer, Genomics and Bioinformatics Journal, 2016.
14. Fake News Detection Dataset-Based Machine Learning Approaches (Adapted for Multimodal Analysis Concepts), Kaggle Research Paper, 2020.
15. Multimodal Biomedical Data Analysis Using AI Techniques, IEEE Conference Proceedings, 2021.