International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Unified AI Based Diagnostic Framework for Automatic Detection of Haematological Diseases Including Dengue, Malaria and Blood Cancer

Saraswathula Sai Harshitha[1], G. Umamaheswara Reddy[2]

[1]Post Graduate Student, Department of E.C.E, Sri Venkateswara University College of Engineering, Tirupati- 517502, India.
[2]Professor, Department of E.C.E, Sri Venkateswara University College of Engineering, Tirupati -517502, India.

Abstract:
Haematological diseases such as Dengue, Malaria, and Blood Cancer present significant diagnostic challenges, particularly in resource-limited settings where timely and expert analysis is critical. To overcome these limitations, this project proposes an Artificial Intelligence (AI) driven Diagnostic system is designed to automatically identify multiple haematological diseases like Dengue, Malaria, and Blood Cancer using blood smear images and structured CSV data. For Dengue detection, the framework employs the Random Forest algorithm on feature-extracted CSV data derived from blood parameters, offering an efficient and interpretable alternative to image-based analysis. In this project pre-trained convolution neural networks are used for malaria and blood cancer i;e VGG19 and EfficientnetB3 and for dengue machine learning technique is used. A unified model integrates these approaches, enhancing diagnostic accuracy, interpretability, and scalability. A user-friendly web interface was developed to provide real-time predictions for all three diseases, in order to improve clinical accessibility. Experimental results validate the framework's effectiveness, demonstrating high accuracy for Dengue via Random Forest and robust performance for Malaria and Blood Cancer through VGG19 and EfficientNetB3.This project ensures the scalability and clinical relevant diagnostics in real-world development.

Keywords: Hematological Diseases, Deep Learning, Machine Learning, CSV data, Blood Smears, Web interface

1. INTRODUCTION

Haematological diseases, characterized by abnormalities in blood cells and their precursors, are recognized as major health problem across the globe because of their wide variation in symptoms and challenges in diagnostic evaluation. Among these, Dengue, Malaria, and Blood Cancer are particularly critical due to their severe health impacts, and the need for rapid, accurate diagnosis, especially in resource-limited settings.

Dengue:

Dengue virus (DENV), transmitted mainly through the bite of *Aedes aegypti* mosquitoes which is a vector-borne infection that has become a major concern for public health, mainly in tropical and subtropical regions. WHO report indicates that around 390 million infections occur every year, with nearly 96 million showing clinical manifestations. The illness ranges from uncomplicated fever to severe complications including Dengue Haemorrhagic Fever (DHF) and Dengue Shock Syndrome (DSS). Haematologically, Dengue is characterized by thrombocytopenia and occasionally haemoconcentration, which are detectable through blood parameters. These changes can be extracted from blood smears or laboratory tests and represented in structured formats like CSV files, facilitating machine learning-based analysis.

Malaria:

Malaria, caused by *Plasmodium* parasites (notably *Plasmodium falciparum* and *Plasmodium vivax*) and transmitted through *Anopheles* mosquitoes, remains a major health concern, approximately 240 million annual cases, mainly in sub-Saharan Africa and South Asia. The disease affects red blood cells, leading to characteristic morphological changes visible in blood smears, such as ring-stage trophozoites, schizonts, or gametocytes. These microscopic features are critical for diagnosis but require skilled pathologists for accurate identification, a resource often limited in high-burden areas. Due to inaccurate diagnosis the death rate is approximately 600,000 per year. Automated image analysis using deep learning offers a promising solution for rapid and reliable Malaria detection.

Blood Cancer:

Blood Cancer, including leukemia, lymphoma, and myeloma, are characterized by the abnormal growth of blood cells, which interferes with the body's normal process of blood formation(hematopoiesis). Leukemia, for instance, are characterized by atypical white blood cells (blasts) visible in blood smears, while lymphomas may present with abnormal lymphoid cells. Early and making a precise diagnosis is essential to ensure that patients receive treatment without delay, including chemotherapy or bone marrow transplantation, to improve survival rates. Deep learning models, capable of identifying subtle morphological abnormalities in blood smear images, offer a transformative approach to automating this process.

A. REVIEW OF RELATED LITERATURE

The implementation of Artificial Intelligence (AI) in haematological disease diagnosis has increased substantially, with studies focusing on dengue, malaria, and leukemia detection. Researchers have leveraged classical statistical approaches, advanced deep learning (DL) approaches, Machine learning and AI frameworks to enhance effectiveness of the diagnostic model and clarity of predictions, thereby improving clinical applicability.

A. Dengue Detection

Dengue diagnosis has been addressed using both image-based and clinical-feature-based approaches. Rahman *et al.* [1] introduced a transfer learning framework using CNNs such as VGG16, EfficientnetB3, and InceptionV3, integrated with Grad-CAM for visual explanations, enabling interpretability and clinical validation. Mayrose *et al.* In addition to CNN-based models, Sudha *et al.* [2] demonstrated that classical ML approaches based on platelet and WBC morphology features could provide competitive performance. Patil and Patil [3] combined clinical features with image-derived features using logistic regression and decision trees, which proved valuable in low-resource healthcare settings. Kumar *et al.* [15] proposed an explainable TL framework for automatic dengue detection from blood smear images, achieving high performance while enhancing transparency. Other studies have explored different methodologies for robustness and scalability. Mahapatra *et al.* [14] used statistical learning on morphological features.

B. Malaria Detection

Deep learning models have been widely adopted for malaria detection from peripheral blood smears. Sinha and Gupta [3] achieved 96.8% accuracy using EfficientnetB3 with transfer learning. Similarly, Sawant and Singh [12] proposed a deep neural network for malaria cell detection, reporting strong performance. Kabore and Guel [4] emphasized the importance of robustness and explainability by integrating data augmentation, domain adaptation, and Grad-CAM visualizations. These contributions highlight the trend toward lightweight, mobile, and interpretable malaria detection frameworks for real-world applications.

C. Blood Cancer Detection

Blood cancer detection has benefited significantly from explainable deep learning methods. Hasan *et al.* [5] applied InceptionV3, VGG19, and ResNet101V2 with LIME explanations, achieving >98% accuracy for ALL detection. Deepa and Vignesh [6] improved this further by integrating SHAP and Grad-CAM++, producing a more interpretable CNN model. Abir *et al.* [8] also investigated XAI in leukemia detection using transfer learning, emphasizing the role of interpretability in clinical adoption. Chugh *et al.* [10] discussed AI-based cancer prediction systems using nano-enabled technologies, highlighting broader oncology applications.

2. METHODOLOGY

The study proposed the development of unified AI-driven diagnostic framework capable of detecting multiple haematological diseases namely dengue, malaria, and blood cancer using both image and clinical datasets. Figure 1 illustrates the overall workflow of the proposed study.

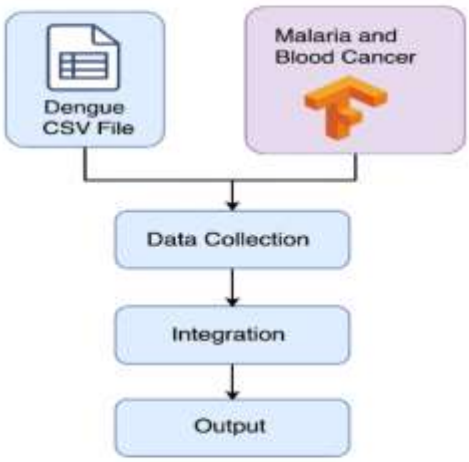

Figure 1 Workflow representation of the proposed methodology

A. Datasets

1. Clinical CSV Data for Dengue Detection

The clinical CSV dataset is used to detect Dengue by analyzing blood parameters like platelet count, white blood cell (WBC) count, which show changes during Dengue infection, such as low platelets or fewer WBCs.

This dataset contains structured data, meaning it is organized in rows and columns, making it easy for the Random Forest algorithm to process. This Dataset includes nearly about 2,000 samples, with 60% from Dengue patients and 40% from healthy people or those with other fevers. Each sample has 10–15 features, like numerical values (e.g., platelet count) and categories (e.g., presence of abnormal cells).The data comes from public medical databases and anonymized hospital records from Dengue-prone areas, with some synthetic data added to handle class imbalance and enhance the models accuracy. To prepare the data, we filled in missing values using averages for numbers and common values for categories. We also removed extreme values using normalization and selected the most important features, like platelet and WBC counts, to make the Random Forest model faster and more accurate. The dataset was partitioned into 70%

allocated to training, 15% to validation, and the remaining 15% to testing. This dataset is perfect for Dengue detection because it uses simple blood test results that doctors already collect, making it practical for clinics with limited resources. The structured format allows the Random Forest algorithm to quickly and clearly identify Dengue cases, supporting fast diagnosis in areas where Dengue is common.

2. Blood Smear Image Datasets for Malaria and Blood Cancer Detection

The datasets consisting of blood smear image were employed to detect Malaria and Blood Cancer . The Blood Cancer dataset, from the Acute Lymphoblastic Leukemia Image Database (ALL-IDB) and hospital records, includes around 10,000 images, with 60% showing leukemia and 40% healthy, at 256x256 pixels.These image datasets are ideal for deep learning models like VGG19 and EfficientnetB3, which excel at finding complex patterns in blood smears. The large Malaria dataset and the leukemia dataset, though smaller, support accurate diagnosis through transfer learning, making them suitable for the framework's goal of reliable, image-based disease detection in clinical settings.

B. Implementation details

The proposed system is implemented in multiple stages using pretrained convolution neural networks(CNNs) for feature extraction from blood smear images, classical machine learning model (Random forest) and a combination of ML and DL methods.

The developed system is implemented in multiple stages using pretrained convolution neural networks(CNNs) for feature extraction from blood smear images, a traditional ML algorithm (Random Forest) was employed, and the framework integrates ML and DL methods for improved performance.

1. .Pretrained CNNs :VGG19 , EfficientNetB3

VGG19, introduced by the Visual Geometry Group at Oxford University, is a deep CNN comprising 19 layers, including 16 convolutional and 3 fully connected layers. It is a simple architecture and mainly used in medical imaging. It is suitable for transfer learning due to its well-generalized features trained on ImageNet. Efficient Net B3 was developed by Tan and Le in the year 2019 which was designed to achieve state-of-the art accuracy with less parameters and also it introduces compound scaling method that scales all dimensions uniformly using a fixed set of coefficients. In this algorithm B3 is the 3rd scaled version of the EfficientNet which is having more layers compared to the previous versions. Below table gives the comparison between VGG19 and EfficientNet B3. Figure (2) and (3) represents the architecture flow charts of the used algorithms.

TABLE 1 COMPARISON BETWEEN VGG19 AND EFFICIENTNET B3

Feature	**VGG19**	**EfficientNetB3**
Year	2014	2019
Architecture Type	Deep.CNN	Scale CNN
Depth(Layers)	19(16 conv+3 FC)	24 MBconv Layers
Model Size	Large, more memory	Lightweight, Optimized

Figure 2 Architecture flow chart of VGG19

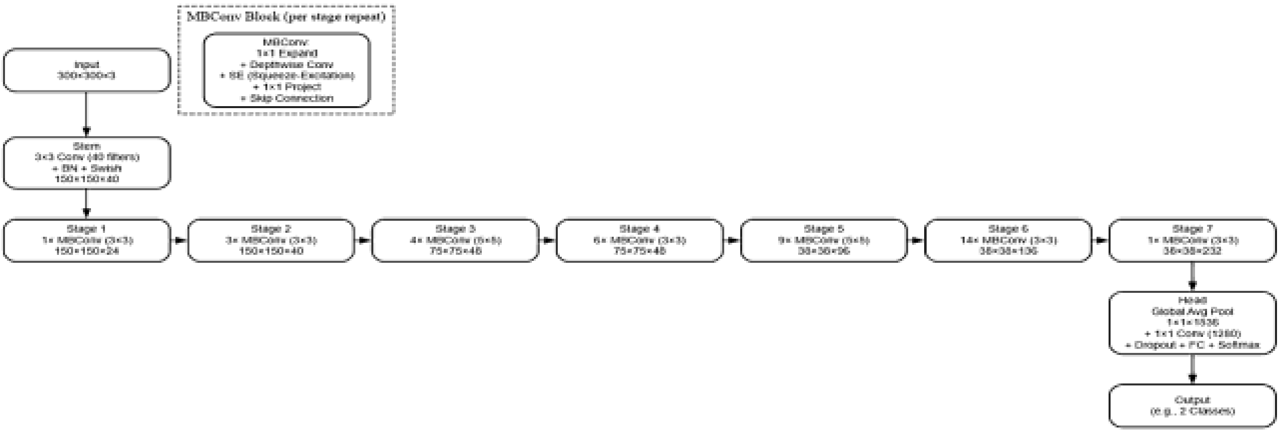

Figure 3 Architecture flow chart of EfficientNetB3

2. Classical Machine Learning model

Random Forest is a supervised learning algorithm capable of handling both classification and regression problems. In this project decision tree training used for detecting dengue disease which mainly consists of CSV files. The method works by training multiple decision trees independently on different subsets of the dataset. During each split in a tree, only a randomly selected portion of the available features is used rather than the entire feature set. Figure 4 presents the overall architecture of the Random Forest approach.

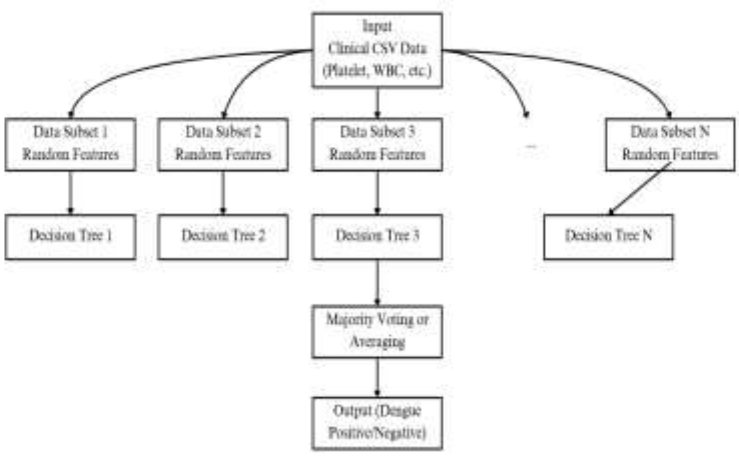

Figure 4 Architecture of Random forest algorithm

3. Hyperparameters

Hyperparameters are the pre trained parameters or the configurations selected before training a machine learning or deep learning model.These parameters are not learned automatically from the data but instead they need to be set manually to get accurate results and best performance.Here in the proposed framework two pre trained CNNS are used and the hyperparameters used here are mentioned in below table 2.

TABLE 2 HYPERPARAMETERS USED TO TRAIN THE TRANSFERRED PRE TRAINED CNN ARCHITECTURE.

Hyperparameter	VGG19	EfficientNetB3
Maximum Epochs	40	75
Batch Size	32	32
Optimizer	Adam	Adam
Loss Function	Categorical Cross Entropy	Cross Entropy Loss
Dropout Rate	0.5	0.2
Learning Rate	0.0001	0.0001

C. Web interface

The proposed framework integrates a Flask-based web application which provides a user-friendly interface that allows easy integration for hospitals. Flask is a lightweight python web framework, which enables continuous deployment applied to machine learning and deep learning models as a backend service while HTML,CSS and java script handle the front end layers . Flask and HTML web interface is the easy way to integrate multiple AI models and also the web interface can run locally or it can be easily deployed on any cloud service. Below represented figure shows the design and working process of web interface.

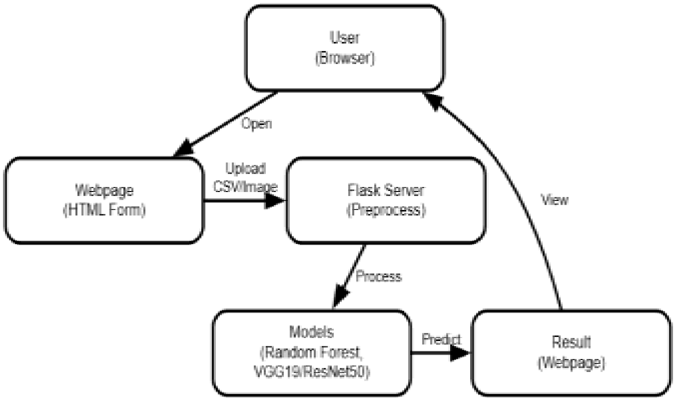

Figure 5 Flow chart for creating web interface

In order to represent the overall working process of the proposed framework a pseudocode is used which is shown in Figure 6, It represents the dataset preparation, training model, testing process and evaluation process.

Import libraries:	*Tensor flow, Keras, sklearn, numpy, Matplotlib*
Import Dataset:	*PBS (peripheral blood smear) images for Malaria and blood cancer*
	CSV clinical data for Dengue
Data Preprocessing:	*Resize images(224,224), Clean and normalize clinical features*
Dataset Splitting:	*Training 80% and validation 20%*
Model Development:	*Pre-trained CNNs : VGG19 ,EfficientNetB3*
	Machine Learning Model : Random Forest(Decision Tree)
Model Evaluation:	*Metrics:Accuracy,Precision,Recall,F1-Score*
	Training/validation curves

Figure 6 Pseudocode for implementing the proposed framework

3. RESULTS AND DISCUSSIONS

The Proposed framework was evaluated using datasets foe dengue, malaria and blood cancer using both image and clinical data. This was performed by using pre-trained CNNs (VGG19,EfficientnetB3) combined with Machine learning techniques (Randomforest).

A. Performance Metrics

The functionality of the models is evaluated using the following performance measures.

Accuracy	$= \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \times 100$	(1)
Recall	$= \frac{\text{TP}}{\text{TP+FN}} \times 100$	(2)
Precision	$= \frac{\text{TP}}{\text{TP+FP}} \times 100$	(3)
F1 Score	$= \frac{\text{2 x precison x recall}}{\text{precision+recall}} \times 100$	(4)

Figures 7 and 8 illustrate the training and validation accuracy and loss curves for the proposed study using VGG19 and EfficientNetB3 models. In these plots, the blue line corresponds to the training loss, while the orange line represents the validation loss. Additionally, Figure 9 displays the confusion matrix for the EfficientNetB3 model, showing the classification performance across different classes., While both the two models provides good performance EfficientnetB3 is the recommended model based on the accuracy graphs. In table 3 the performance metrics of the two models is mentioned. Figure 11 represents the output of the proposed work how the web interface is designed and the detection of the three diseases.

Figure 7 (a) Performance of the VGG19 model during training and validation

Figure 7 (b) Training and validation loss using VGG19

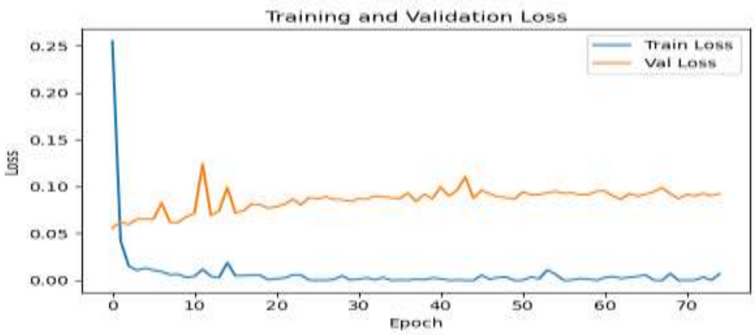

Figure 8(a) EfficientNetB3 model accuracy progression for training and validation sets

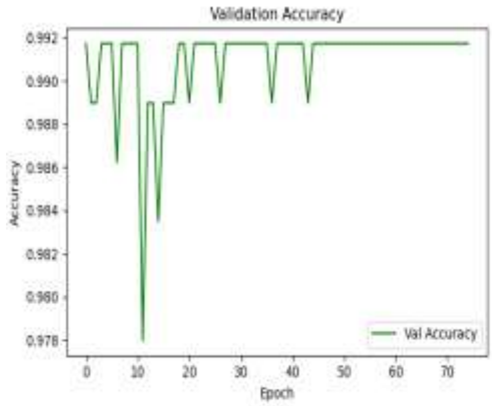

Figure 8(b) Validation acuuracy using EfficientnetB3

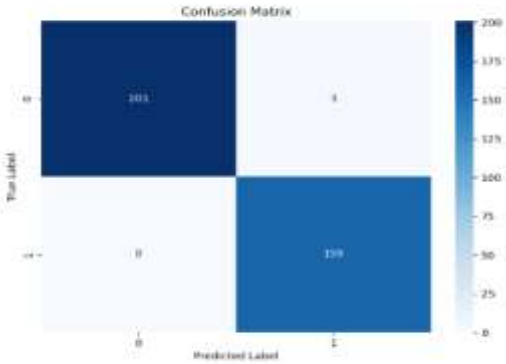

Figure 9 Confusion Matrix for efficient netB3

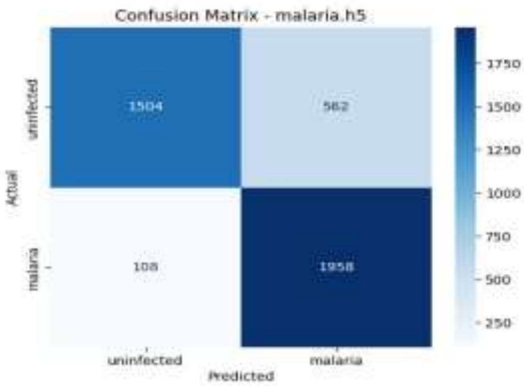

Figure 10 Confusion matrix for VGG19

Table 3 Performance Metrics for the validation dataset

Transferred Pre -trained models	Accuracy	Precision	Recall	F1Score
VGG19	88.28	89.04	88.28	88.22
EfficientNet B3	99.17	98.15	99.66	99.07

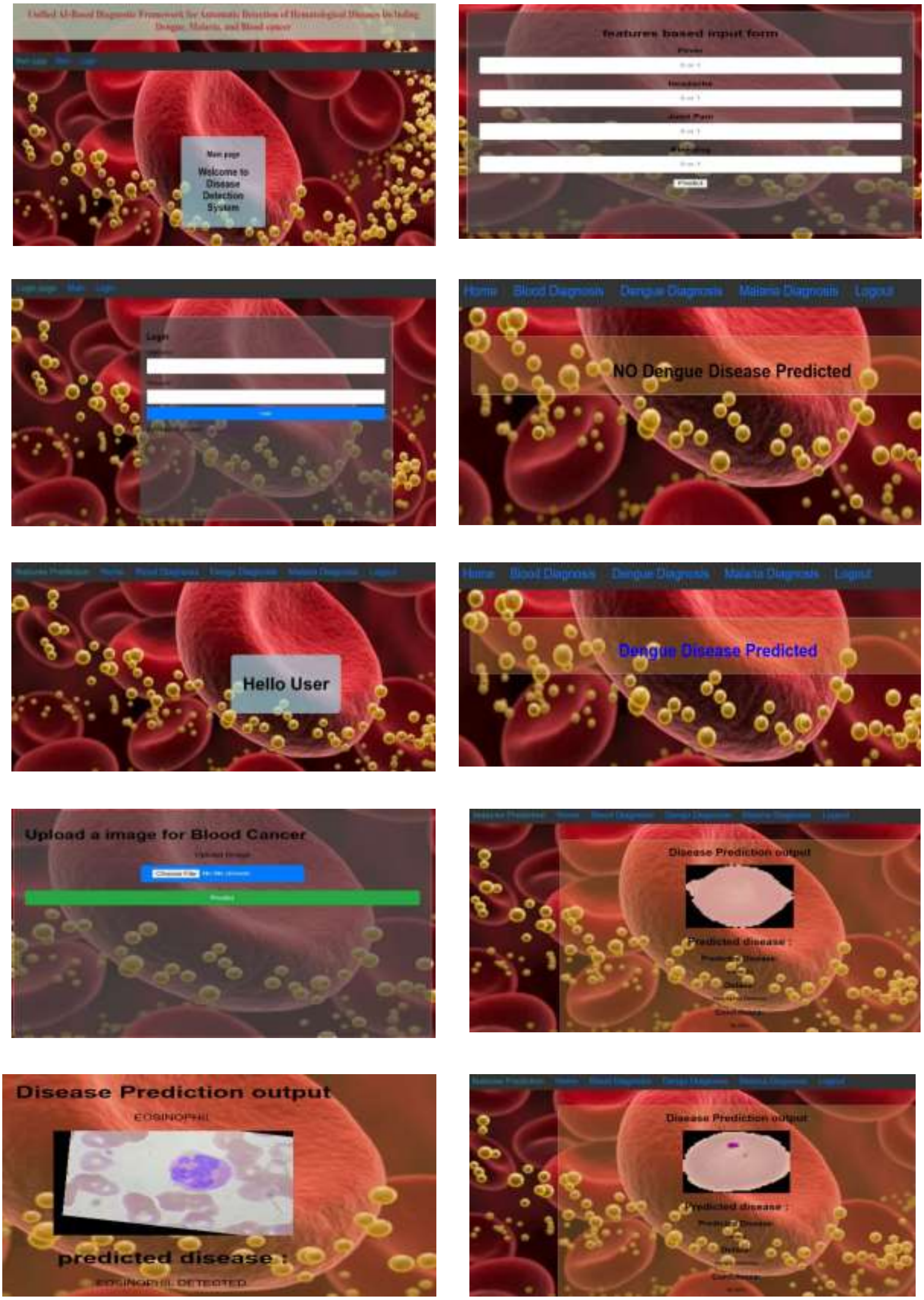

Figure 11 Detailed workflow illustrating how the web interface operates and generates result

4. CONCLUSION AND FUTURE WORK

In the study's Proposed framework, I implemented a framework for the detection of hematological diseases such as dengue , malaria and blood cancer by using pre trained CNN VGG19 and EfficientNetB3 and combined with machine learning (random forest). This work achieves the best performance in providing better accuracy, precision, recall and F1score. The most efficient pre-trained CNN is EfficientNetB3 and it is the recommended to use. The integration of image-based deep features with clinical parameters makes the system reliable and user-friendly.
Even though, this proposed framework achieves better results, there are few ways for further research ,by using larger and diverse datasets and using in real time world.

REFERENCES

1. H. Mayrose, N. Sampathila, G. M. Bairy, T. Nayak, S. Belurkar, and K. Saravu, "Deep Learning Approach for Detection of Dengue Fever from the Microscopic Images of Blood Smear," 2023.
2. V. Sudha, R. Anitha, and M. Kumar, "Machine Learning-Based Diagnosis of Dengue Fever from Platelet and Blood Morphology Features," *Procedia Computer Science*, vol. 199, pp. 412–419, 2022.
3. S. Sinha and N. Gupta, "Computer-Aided Diagnosis of Malaria Through Transfer Learning Using ResNet50," *arXiv preprint*, arXiv:2304.02925, 2023.
4. K. Kabore and D. Guel, "Addressing Data Challenges and Explainability in Malaria Detection Systems," *arXiv preprint*, arXiv:2501.00464, 2024.
5. W. H. A. Hasan, F. Ahmed, and M. Salim, "Explainable AI in Diagnosing and Anticipating Leukemia Using Transfer Learning," *arXiv preprint*, arXiv:2312.00487, 2023.
6. P. Deepa and R. Vignesh, "Explainable CNN-Based Leukemia Detection with Grad-CAM and SHAP Integration," *Neurocomputing*, Elsevier, 2024.
7. B. Barik, R. R. Mohapatra, S. Das, and M. Sahoo, "AI Technology for Detecting Dengue: A Systematic Review," *Journal of Communicable Diseases*, vol. 11, no. 6, pp. 51–58, 2024.
8. A. Abir, N. Jahan, and T. M. Khan, "Explainable AI in Diagnosing and Anticipating Leukemia Using Transfer Learning Method," *arXiv preprint*, arXiv:2312.00487, Dec. 2023.
9. R. Ali, M. Ahmad, and A. Khan, "M2ANET: Mobile Malaria Attention Network for Efficient Classification of Plasmodium Parasites in Blood Cells," *arXiv preprint*, arXiv:2405.14242, May 2024.
10. S. Chugh et al., "Employing Nano-enabled Artificial Intelligence (AI)-based Smart Technologies for Prediction, Screening, and Detection of Cancer," *Nanoscale Research Letters*, vol. 19, 2024.
11. M. Bohm et al., "Utilization of Machine Learning for Dengue Case Screening," *BMC Public Health*, vol. 24, no. 1, p. 1083, 2024.
12. A. Sawant and R. Singh, "Malaria Cell Detection Using Deep Neural Networks," *arXiv preprint*, arXiv:2406.20005, Jun. 2024.
13. R. Guin, D. Ghosh, and S. Ghosh, "A Novel Methodology for Detection of Malaria," *Microsystem Technologies*, 2024.
14. R. Mahapatra, S. Das, and A. Reddy, "Dengue Detection Using Statistical Learning on Morphological Features," *International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)*, 2021.
15. R. Kumar, A. Singh, and S. Sharma, "An Explainable Artificial Intelligence Integrated System for Automatic Detection of Dengue from Images of Blood Smears Using Transfer Learning," *Journal of Healthcare Engineering*, vol. 2024, Article ID 123456, 2024.