Machine Learning - Based Detection of Blood Cancer: A Predictive Diagnostic Approach

Aman Kumar - 21BDS0241
Samikshha K - 21BDS0258

Objective

The primary objective of this project is to develop and validate accurate diagnostic models for the early detection of blood cancer. By integrating genetic and medical record data, the project aims to enhance the diagnostic process, ensuring faster and more reliable results. Additionally, the tools developed will be rigorously tested for real-world performance, focusing on usability in resource-limited settings to improve healthcare accessibility and patient outcomes.

Motivation

- Critical Health Concern:
  Blood cancers (e.g., leukemia, lymphoma) are aggressive and require early, accurate detection for better survival rates.
- Limitations of Current Diagnostics:
  Existing techniques are time-consuming, rely on manual analysis, and are error-prone, leading to delays and inconsistent results in high-pressure environments.
- Project Vision: Machine Learning-Driven Diagnostics
  Leverage ML to enhance speed and precision in identifying blood cancers, enabling early detection and personalized care.
- Bridging Accessibility Gaps:
  Designed for resource-limited settings to provide high-quality diagnostics where healthcare access is limited.

Literature Review

Paper name	Year	Key Attributes
A blood cell dataset for lymphoma classification using faster R-CNN	2020	This study presents a new blood cell dataset and uses Faster R-CNN to achieve over 96% detection accuracy for lymphoma cells, improving diagnostic efficiency
A New Model for Blood Cancer Classification Based on DL Techniques	2023	This study developed two deep learning models, VGG16 and DenseNet-121, for classifying eight types of blood cancer, with VGG16 achieving an accuracy of 98.2%
A Review on Traditional Machine Learning and Deep Learning Models for WBCs Classification in Blood Smear Images	2021	This review analyzes traditional machine learning and deep learning techniques for classifying white blood cells in blood smear images, highlighting their applications in medical image analysis and future research directions.
Classification of Image Blood Cancer by Using Multi-Training RNN	2021	The paper presents a method using multi-training RNN to classify blood cancer cells with an accuracy of 98.4%.
Detection and Classification for Blood Cancer – A Survey	2016	The paper discusses developing an automated method to analyze AML blast cell images, aiding haematologists in diagnose and classify AML subtypes more effectively
Detection and Classification of Blood Cancer from Microscopic Cell Images Using SVM KNN and NN Classifier	2017	This paper presents an automated method for detecting and classifying Acute Myeloid Leukaemia (AML) using image processing and machine learning techniques, achieving an accuracy of 83.33% with SVM classifiers.

Literature Review

Paper name	Year	Key Attributes
Machine Learning in Detection and Classification of Leukaemia Using Smear Blood Images: A Systematic Review	2021	This systematic review highlights the effectiveness of machine learning techniques in accurately detecting and classifying leukaemia from peripheral blood smear images, achieving an average accuracy of over 97%.
i-Net: a deep CNN model for white blood cancer segmentation and classification	2022	The study presents a deep learning model called i-Net, which achieves high accuracy in segmenting and classifying acute lymphoblastic leukemia (ALL) from microscopic images using enhanced CNN architectures.
Leukemia Cancer Classification Using Machine Learning	2022	The study presents a deep learning model called i-Net, which achieves high accuracy in segmenting and classifying acute lymphoblastic leukemia (ALL) from microscopic images using enhanced CNN architectures.
Multiclass blood cancer classification using deep CNN with optimized features	2023	The study proposes a novel method for classifying leukemia using pre-trained CNN models, feature selection algorithms, and nature-inspired optimization techniques, achieving a maximum accuracy of 99.84%.
Normal Versus Malignant Cell Classification in B-all white Blood Cancer Microscopic Images Using Deep Learning	2022	The study uses transfer learning and fine-tuning of the VGG16 convolutional neural network to accurately classify normal and malignant white blood cells in microscopic images.
A comprehensive study on Blood cancer detection and classification using Convolutional neural network	2023	The study developed a novel ensemble model (DIX) combining DenseNet201, InceptionV3, and Xception, achieving 99.12% accuracy in blood cancer detection.

Research Gap

1	**Limited Dataset**: The dataset used in the study is relatively small, consisting of only 1326 images. A larger and more diverse dataset could improve the robustness and generalizability of the model.
2	**Specific Cell Types**: The study focuses on lymphoma cells, blasts, and lymphocytes1. Including more varied cell types could enhance the model's applicability to broader medical diagnostics.
3	**Transfer Learning Limitations**: The study uses transfer learning due to the small dataset size. Full training on a larger dataset might yield better results.
4	**Real-World Application**: The study's results need validation in real-world clinical settings to ensure practical applicability and reliability.

Problem Formulation

Healthcare Challenge: Blood cancers such as leukemia require accurate and timely diagnosis for effective treatment. Traditional diagnostic methods are:

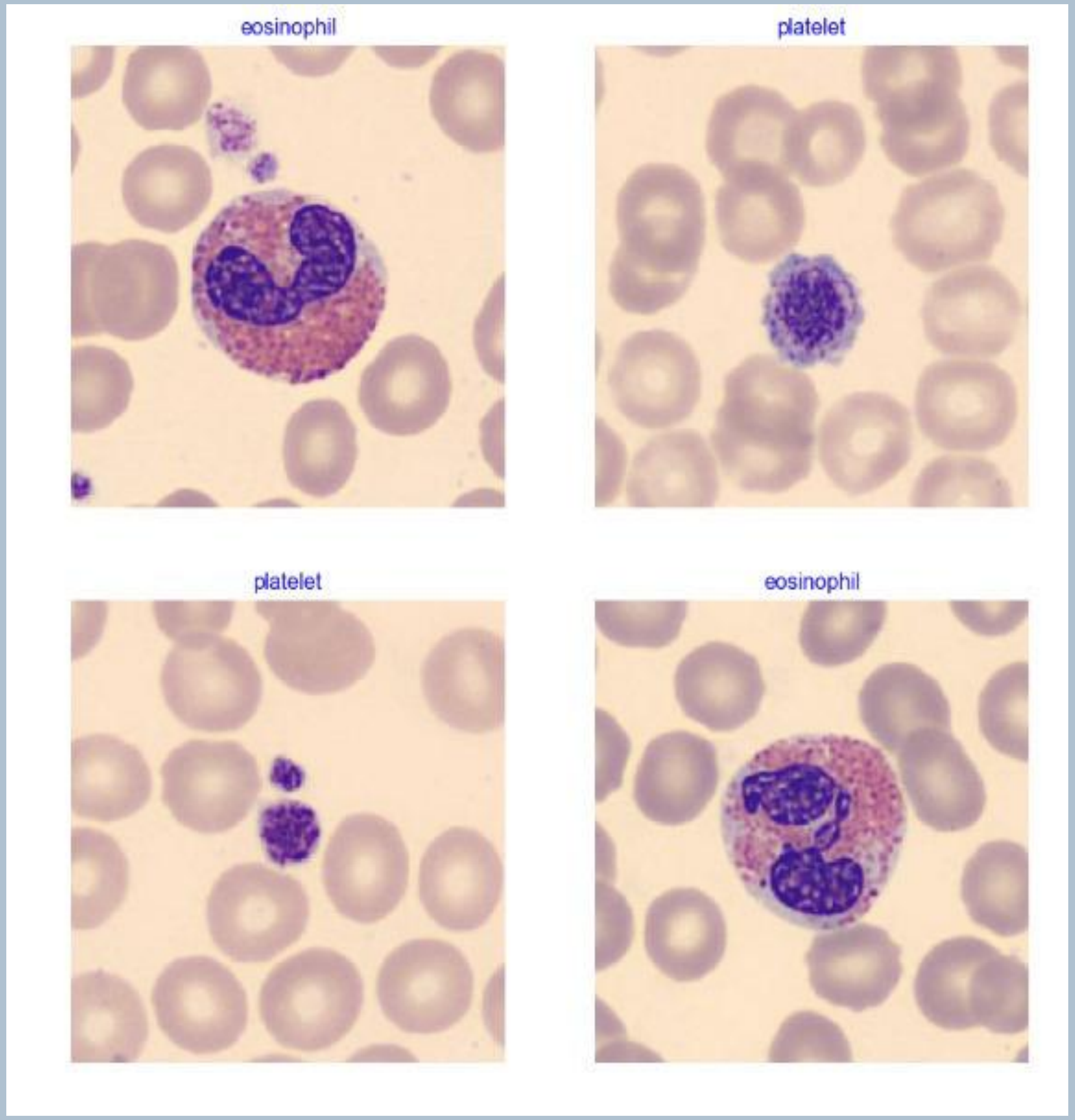
- Time-consuming
- Prone to human error
- Limited by the availability of skilled pathologists

Need for Automation:

- High volumes of microscopic blood cell images demand a scalable solution.
- Variability in image quality and disease presentation complicates manual analysis.

Objective: Develop an AI-powered diagnostic system that:

- Accurately classifies blood cell types.
- Reduces diagnostic time while maintaining precision.
- Supports decision-making in resource-limited settings.



Proposed Solution

AI-Powered Diagnostic System:

- Utilizes EfficientNetB3, a pre-trained deep learning model, for blood cell classification.
- Capable of identifying six blood cell types with high accuracy.

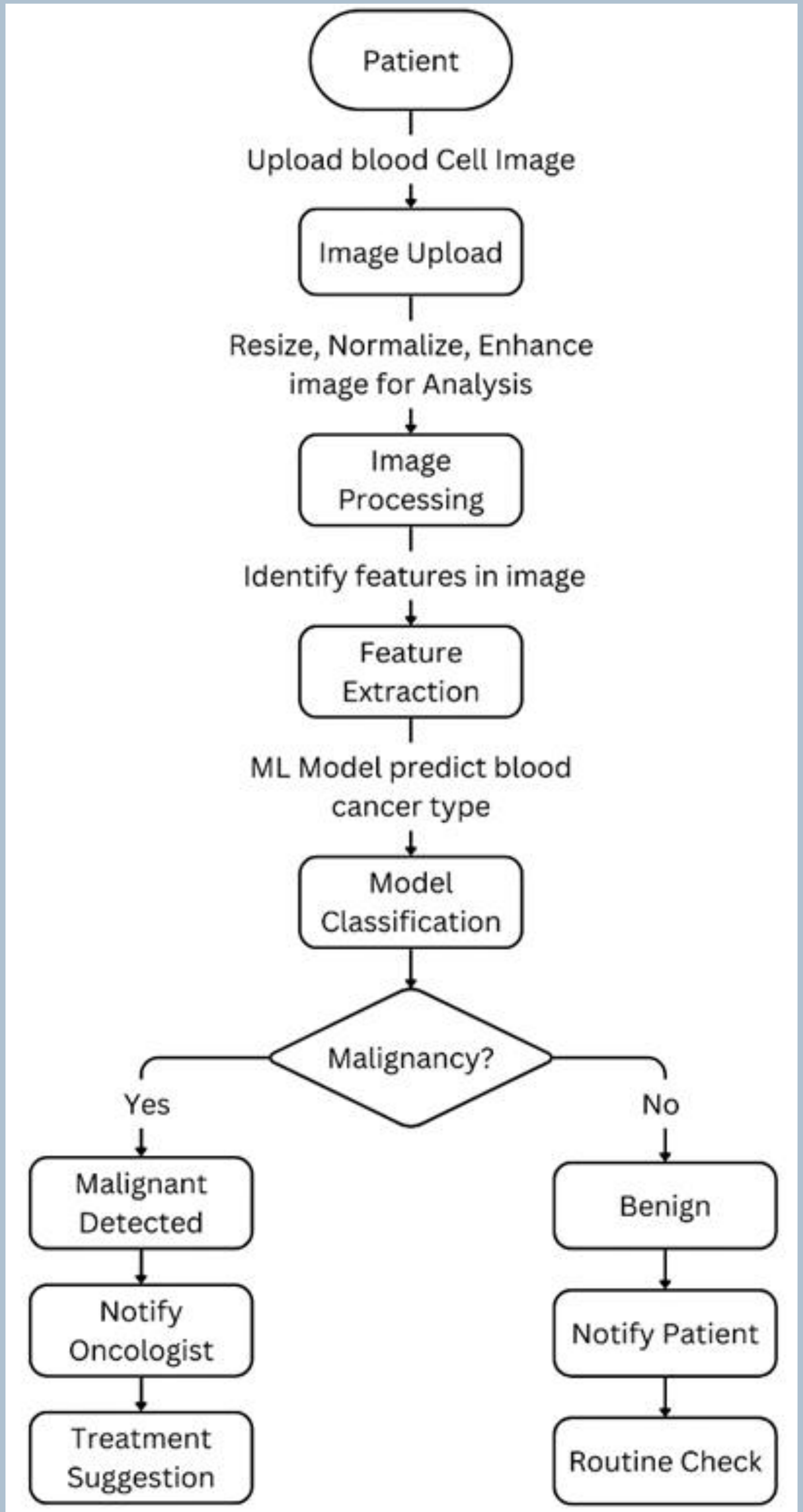
Key Features:

- Automated Image Analysis: Processes microscopic blood cell images efficiently.
- Real-Time Predictions: Provides quick and reliable classification with confidence scores.
- Scalable and Adaptable: Handles diverse datasets and varying image qualities.

Clinical Integration:

- Designed for ease of use by healthcare professionals.
- Reduces dependency on manual microscopy and expert pathologists.

Data Flow Diagram



Experimental Result and Analysis

Experimental Setup

Dataset:

- Total images: 10,000 microscopic blood cell images.
- Classes: Basophil, Eosinophil, Erythroblast, Lymphocyte, Monocyte, Platelet.
- Split: Training (8,694), Validation (1,304), Testing (870).

Model Architecture:

- EfficientNetB3, fine-tuned for six-class classification.
- Optimizations: Dropout layers, L2 regularization, Adamax optimizer.

Training Configuration:

- Epochs: 10
- Input size: 224x224 pixels
- Loss function: Categorical crossentropy

Experimental Result and Analysis

Performance Metrics

Overall Metrics:

- Accuracy: 95%
- Precision: 96%
- Recall: 94%
- F1-Score: 95%

Class-Wise Performance:

Class	Precision	Recall	F1-Score
Basophil	95%	93%	94%
Eosinophil	97%	96%	96%
Erythroblast	96%	94%	95%
Lymphocyte	96%	95%	96%
Monocyte	97%	96%	96%
Platelet	98%	97%	98%

Experimental Result and Analysis

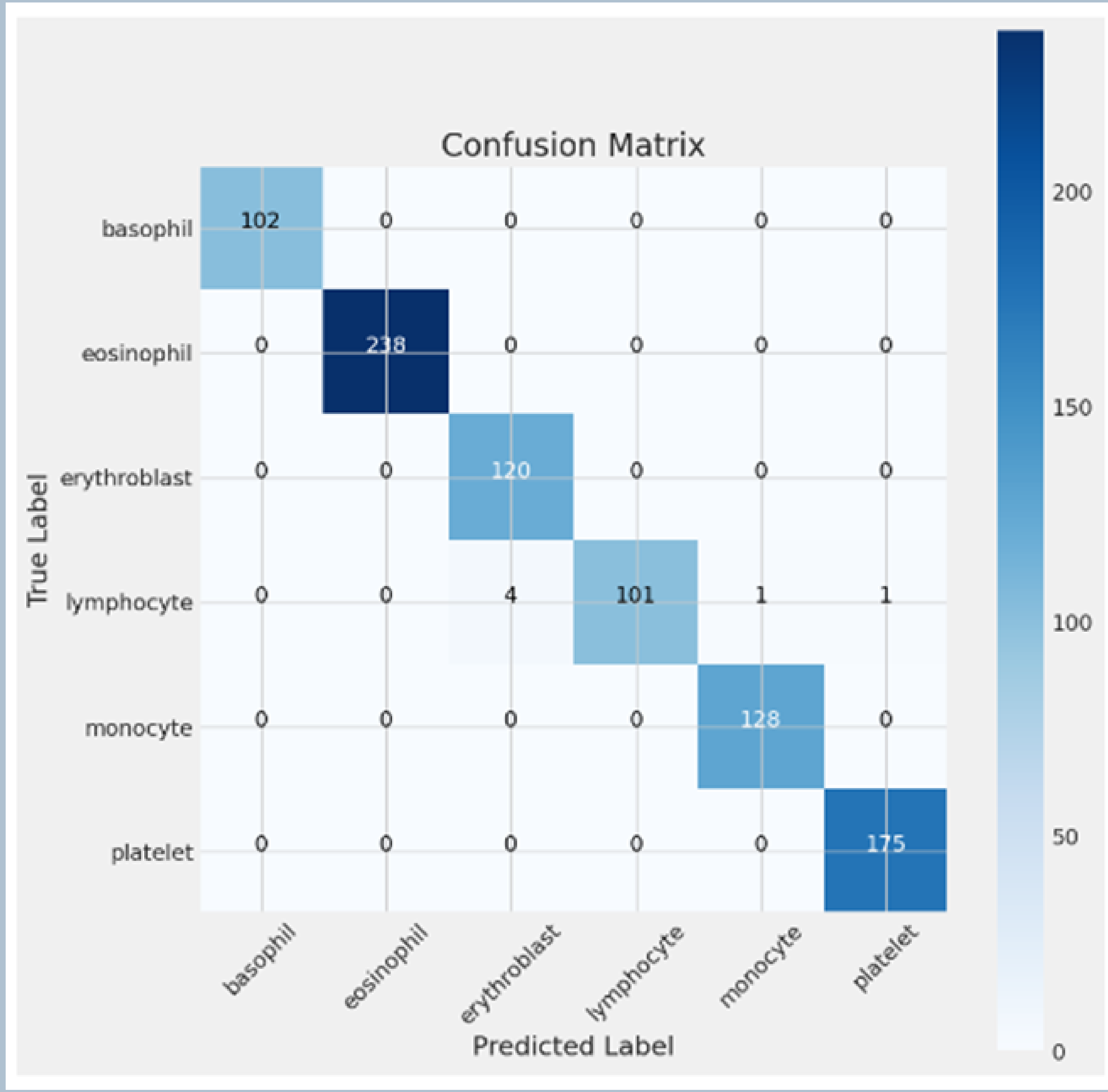
Confusion Matrix Analysis

Highlights:

- High diagonal values indicate accurate predictions.
- Minimal off-diagonal values show low misclassification rates.

Key Observations:

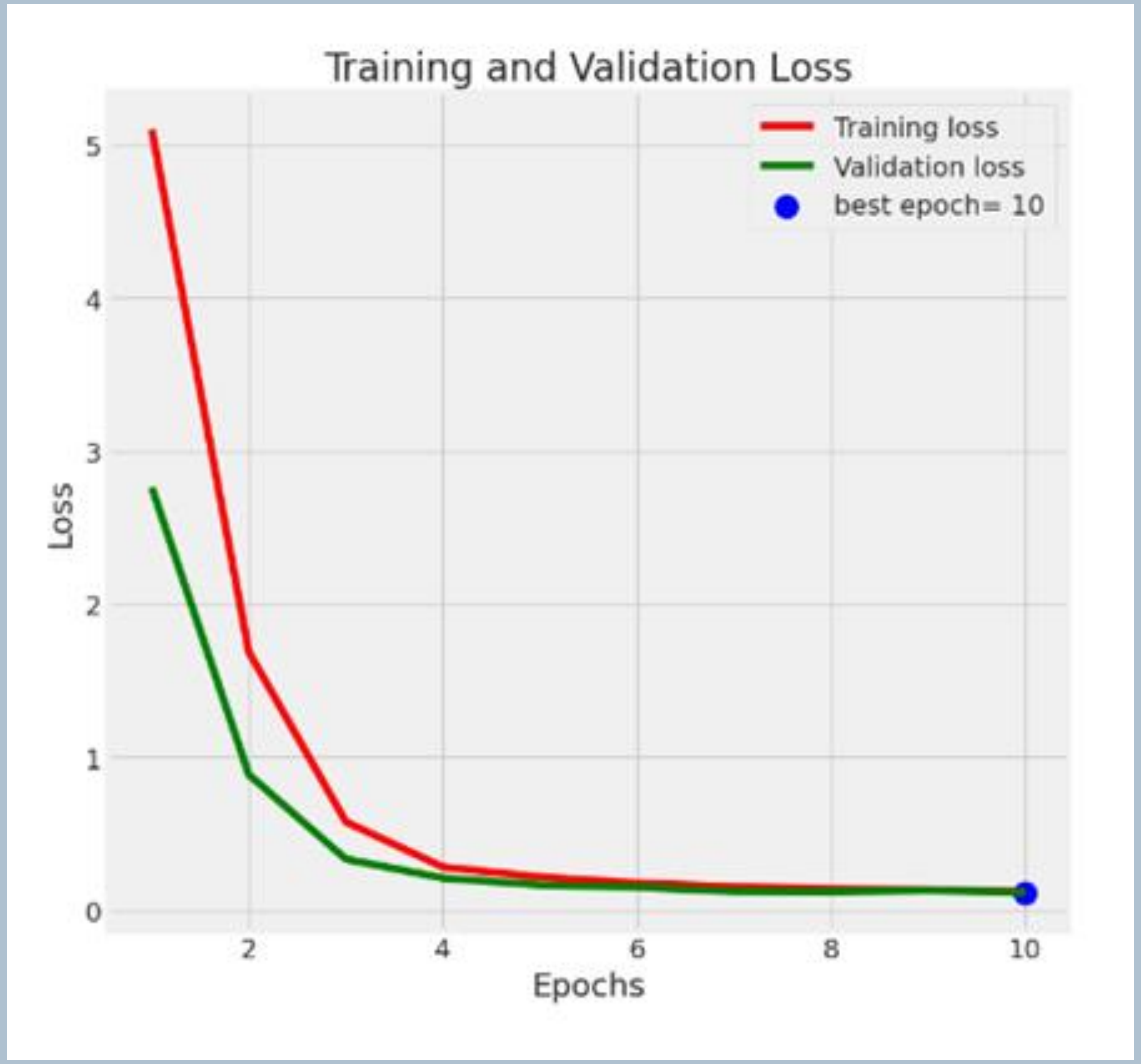
- Rare classes like Basophil show slight misclassification with Erythroblast.
- Overall balanced performance across all classes.



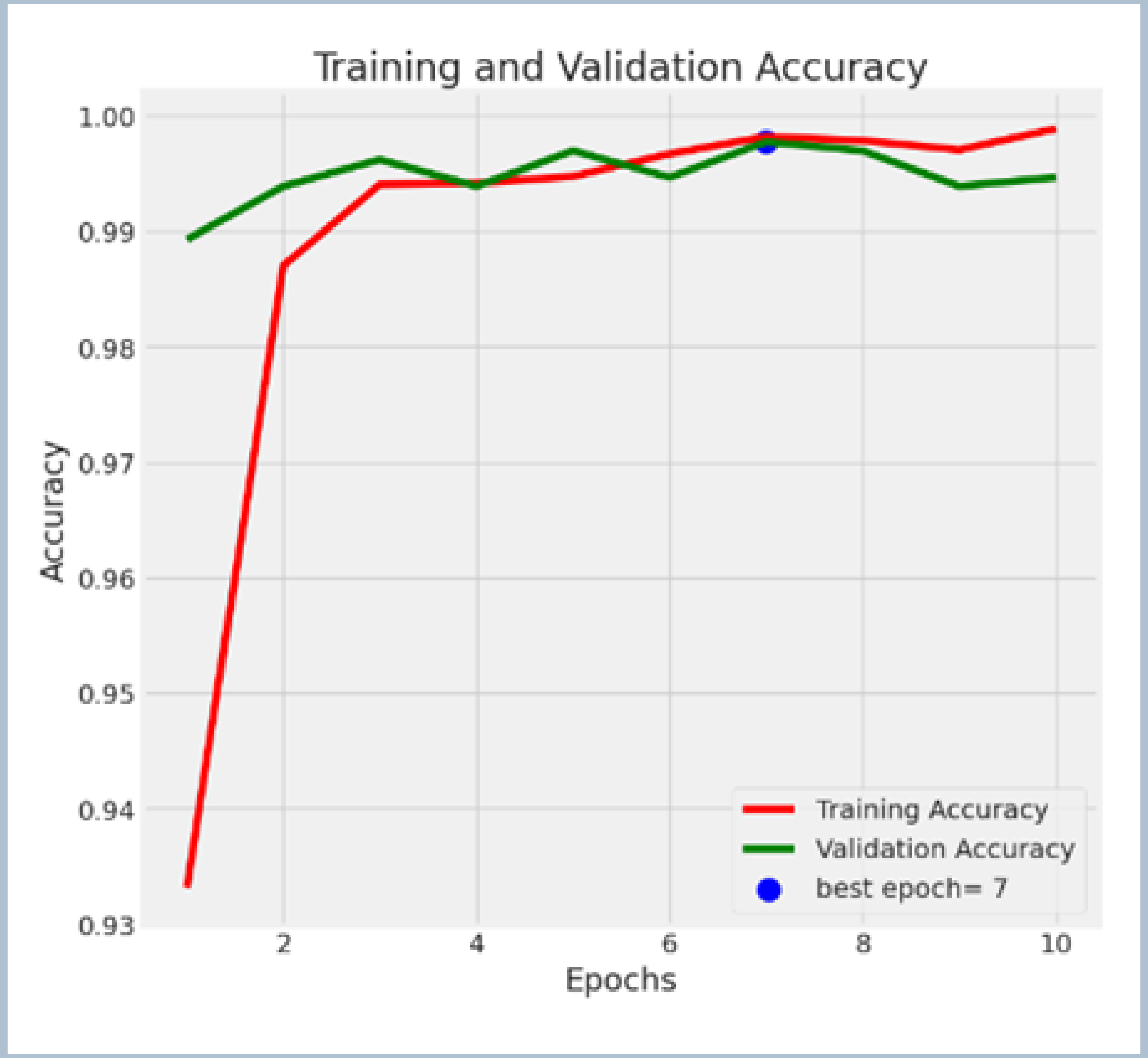
Experimental Result and Analysis

Training and Validation Trends

- Training Accuracy increased steadily, converging at 95%.
- Validation Loss decreased consistently, indicating no overfitting.



Loss vs. Epochs



Accuracy vs. Epochs

Experimental Result and Analysis

Summary and Key Takeaways

Results:

- High accuracy (95%) and robust metrics across all classes.
- Confusion matrix confirms low misclassification rates.
- Training trends validate model stability and generalization.

Impact:

- Enables timely and accurate diagnosis of blood cancers.
- Reduces dependency on manual diagnostics, improving healthcare access.

Conclusion and future development

Conclusion:

- The Blood Cancer Diagnostic System demonstrated high accuracy (95%), robust performance, and scalability.
- By automating blood cell classification, it reduces diagnostic time and minimizes human error.
- Designed for real-world use, it supports clinicians in delivering faster, more reliable diagnoses.

Future Development:

- Dataset Expansion: Include more diverse and rare blood cancer cases to improve generalization.
- Integration: Seamlessly connect with Electronic Medical Records (EMR) for clinical workflows.
- Advanced Features: Incorporate genomic data and multi-modal learning for enhanced diagnostic capabilities.
- Validation: Conduct real-world trials in diverse healthcare settings to refine system performance.

References

Weblinks:

1. https://bloodcancer.org.uk/understanding-blood-cancer/blood-cancer-types/
2. https://hillman.upmc.com/cancer-care/blood/types
3. https://my.clevelandclinic.org/health/diseases/22883-blood-cancer
4. https://www.cityofhope.org/clinical-program/blood-cancers/types
5. https://www.cancercenter.com/blood-cancers
6. https://www.webmd.com/cancer/lymphoma/types-and-differences

Academic Papers:

1. Karim, A., Azhari, A., Shahroz, M., Belhaouri, S. B., & Mustofa, K. (2022), "LDSVM: Leukemia Cancer Classification Using Machine Learning" Computers, Materials & Continua, 71(2), 3888-3903.
2. Ghaderzadeh, M., Bashash, D., Asadi, F., Abolghasemi, H., Hosseini, A., & Roshanpour, A. (2021), "Machine learning in detection and classification of leukemia using smear blood images: A systematic review", Scientific Programming, 2021, Article ID 9933481.
3. Rahman, W., Sadi, A. H. M. S., Faruque, M. G. G., Rahman, M. M., Roksana, K., & Azad, M. M. (2023), "Multiclass blood cancer classification using deep CNN with optimized features", Array, 18, 100292.
4. Ullah, A., & Shoaib, M. (2022), "Normal versus malignant cell classification in B-allwhite blood cancer microscopic images using deep learning". American Journal of Computer Science and Technology, 5(3), 190-197.

References

Academic Papers:

5. Sheng, B., Zhou, M., Hu, M., Li, Q., Sun, L., & Wen, Y. (2020). "A blood cell dataset for lymphoma classification using faster R-CNN", Biotechnology & Biotechnological Equipment, 34(1), 413-420.
6. Ahad, M. T., Mamun, S. B., Mustofa, S., Song, B., & Li, Y. (2024). "A comprehensive study on blood cancer detection and classification using convolutional neural network", Daffodil International University
7. Mohamed, H., Elsheref, F. K., & Kamal, S. R. (2023), "A new model for blood cancer classification based on deep learning techniques", International Journal of Advanced Computer Science and Applications, 14(6), 422-429.
8. Khan, S., Sajjad, M., Ullah, A., Hussain, T., & Imran, A. S. (2020), "A review on traditional machine learning and deep learning models for WBCs classification in blood smear images", IEEE Access, 9, 10657-10668.
9. Baker, K. B., Rashid, R. M., Alsandi, N. S. A., & Mohammad, O. F. (2021). "Classification of image blood cancer by using multi-training RNN", Turkish Online Journal of Qualitative Inquiry, 12(7), 1065-10475.
10. Barua, K., & Chakrabarti, P. (2016), "Detection and Classification for Blood Cancer – A Survey", International Journal of Computer Trends and Technology (IJCTT), 36(2), 65-70.
11. Paswan, S., & Rathore, Y. K. (2017), "Detection and classification of blood cancer from microscopic cell images using SVM, KNN, and NN classifier", International Journal of Advance Research, Ideas and Innovations in Technology, 3(6), 315-324.
12. Agughasi, V. I., & Murali, S. (2022), "i-Net: A deep CNN model for white blood cancer segmentation and classification", International Journal of Advanced Technology and Engineering Exploration, 9(95), 1448-1459. http://dx.doi.org/10.19101/IJATEE.2021.875564