Leukemia Detection Revolution: AI and Machine Learning Enhance Image-Based Diagnosis

[1]Sheikh Sadi Bandan., [1]MD. Samiul Islam Sabbir., [2]Md Sharuf Hossain., [3]Khadiza Tul Kobra

[1]Dept. of Computer Science & Engineering Daffodil International University Dhaka, Bangladesh

[2]Dept. of Data Science Loyola University Chicago, USA

[3]Dept. of Information Technology and Management Illinois Institute of Technology Chicago, USA

DOI: https://doi.org/10.51584/IJRIAS.2024.908057

Received: 25 July 2024; Revised: 08 August 2024; Accepted: 13 August 2024; Published: 17 September 2024

ABSTRACT

Currently, several diseases have become epidemic in Bangladesh, one of them is leukemia. This disease usually affects people of any age. It is usually found in blood cells, blood plasma, bone marrow. If this disease has been going on for a long time and without specific treatment, it often does not turn into cancer. It is called leukemia when it effects the blood and bone marrow. As a general rule white blood cells are affected by leukemia. Leukemia dataset is composed of both images of blood smears from leukemia patients and non-leukemia patients. While earlier research has simply found leukemia or categorized it into a few varieties, this study has recognized leukemia and defined its types, bringing the categorization process one step further. Artificial intelligence can help us with this. For instance, we can use machine learning and deep learning algorithms to detect leukemia cells and be alerted when they are detected. In this paper I have used the Customize CNN algorithm, through which we have used 7 architectures of CNN. The 5 methods of Customize CNN we used are VGG19, MobileNetV2, MobileNetV3, DenseNet201, Inception V3. The accuracy of VGG19 is 63%, that of MobileNetV2 is 97%, that of MobileNetV3 is 99%, that of DenseNet201 is 99%, that of Inception V3 is 96%. From all the methods, it can be seen that the highest accuracy has been found in MobileNetV3 and DenseNet201 methods. The lowest accuracy was obtained using the VGG19 model.

Keywords: Leukemia Disease, Image Processing, Artificial Intelligence, Machine learning.

INTRODUCTION

blood is an important unique part of all living things on this earth. It is with the help of this blood that animals are alive, all their physiological functions are carried out. A normal human body has about 6 liters of blood. With the help of which humans can complete all their physiological functions. Blood is constantly being produced in the human body and the old blood is being discharged from different parts of the body. A person can change his own blood every 4 months, that is, every 4 months, 1 liter of blood can be removed from the human body. All organs are important in human body. Among them, blood is the most important because without blood, humans could not have chosen. For our survival we have to move from one place to another every day, without the proper circulation of blood, this normal life of human beings will be disturbed. Our body needs good nutritious food to function properly otherwise all the organs of our body will not function properly. All parts of the body should be taken care of otherwise different types of bacteria and fungi cause various diseases in the body. Similarly, we have to take proper care of our blood, otherwise we will go blind. Today's world is the world of information technology. Everything now has a touch of technology due to which people are now used to using technology. We have to look at our mobile screen for at least 5/6 hours every day. Most of which are the current young generation. As a result, blood related diseases in our body are more than before. Careless blood can cause various diseases. Among them are infection by bacteria, fungi, viruses. Lack of vitamins in the blood can cause various diseases. Besides, there are many types of blood diseases such as hereditary genetic diseases, long-term drug use, age-related diseases.

Due to all these diseases, our body easily gets tired, fever and night sweats, sudden weight loss, pain in bones and joints, pain under left rib, bleeding from nose and gums, red blisters etc. to be faced. Many babies need help a few years after birth. Much work is being done on this leukemia disease and scientists are constantly looking for new solutions to these diseases. Thanks to technology, it is now possible to diagnose various blood diseases with the help of machines. This saves more time for our doctors, which results in seeing more patients in less time. Treatment of blood diseases is available in a short period of time. It saves us both time and money.

Addressing leukemia in Bangladesh is crucial due to the rising incidence and limited healthcare resources. Early diagnosis and treatment are vital for improving survival rates, yet many patients face delays due to inadequate medical facilities and awareness. Leveraging advanced technologies like AI and machine learning for leukemia detection can significantly enhance diagnostic accuracy and speed, leading to timely interventions. This approach not only saves lives but also reduces the financial burden on families and the healthcare system. Additionally, improving leukemia care can raise overall healthcare standards in Bangladesh, fostering a healthier population and promoting socioeconomic development.

LITERATURE REVIEW

F. Scotti et al. [1] classify leukemia using peripheral blood microscope images, supported by cytogenetics and immunophenotyping as effective diagnostic methods. S. Shafique et al. [2] developed an automated method using SVM to diagnose acute lymphoblastic leukemia with 93.7% accuracy, analyzing images and blood particles in ALL-IDB-1 dataset. S. H. Kassani et al. [3] utilized a hybrid deep learning approach to classify leukemic B-lymphoblast disease, achieving 96.17% accuracy, outperforming CNN models like VGG16 and MobileNet. Mashiat Fatma et al. [4] utilized neural networks for rapid, accurate, and automated acute leukemia classification based on extracted image features, achieving 91% accuracy with a dataset of 50 images. Luis et al. [5] automate leukemia detection through unsupervised segmentation of microscopic blood images, achieving high accuracies of 0.9306, 0.8603, and 0.9119 using a two-color system and K-means clustering on ALL-IDB 2, Kappa index, and leukocyte databases. D. J. Foran et al. [6] developed software using Java for detecting Leukemia from images, aiding in classification with two types of confusion metrics: prototype system and human observation. A. R. J. Begum et al. [7] developed a framework using image processing to detect and analyze leukemia patterns. They used SVM for classifying white blood cells based on various metrics from microscopy images. K. P. Jayavikash et al. [8] developed a machine learning system for leukemia diagnosis using white blood cell characteristics like shape and color. They analyzed cell images, extracting colors and comparing cancerous and normal cells, achieving 93% accuracy with VGG16 architecture. Abdul Gaffar Karim et al. [9] propose automatic leukemia cancer classification using machine learning (LR, DT, SVM) achieving 99.8% accuracy, surpassing other algorithms like RF, GBM, and NB. In their research, Shanbehzadeh et al. [10] automated leukemia classification using machine learning over the past five years. SVM with RBF kernel achieved the highest accuracy of 85.7% among algorithms like XGBBoost and MLP. Y. Zhu et al. [11] employed machine learning to predict Chronic Lymphocytic Leukemia automatically via computer-aided diagnosis, utilizing 6 specific genes and achieving ROC curves exceeding 0.80 for disease classification across 6 datasets.

RESEARCH METHODOLOGY

A. Design Approach

We propose using Convolutional Neural Networks (CNNs) to compare leukemia images systematically, aiming for efficient analysis and classification. Our dataset, sourced from Kaggle, includes 3256 original images categorized into 4 classes for Leukemia disease classification. This approach integrates traditional histopathological analysis with advanced deep learning techniques to enhance early and precise leukemia detection, potentially improving patient outcomes. Our study details each phase, from data collection and preprocessing to model training and evaluation, emphasizing reproducibility, interpretability, and comparative analysis with other methodologies, as illustrated in Figure 1.

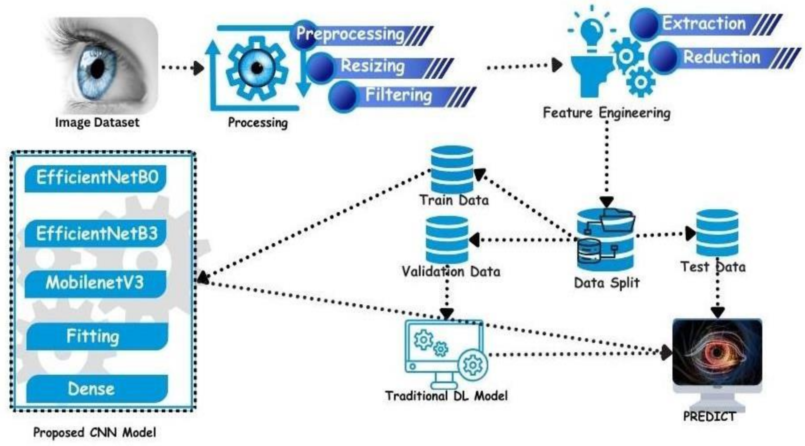

Fig. 1. Architecture of Working Process

B. Dataset Description

This study identified leukemia images in 4 categories. Here the Kaggle [29] dataset is used to diagnose leukemia. A total of 3256 original images will be used in this project and the dataset is collected from Kaggle. The Leukemia Benign category has 504 images, the Leukemia Early category has 985 images, the Leukemia Pre category has 963 images and the Leukemia Pro category has 804 images. The training dataset, testing dataset, and validation dataset are the three sections that make up our dataset. 70% of the data were used for training, 20% for testing, and 10% for validation.

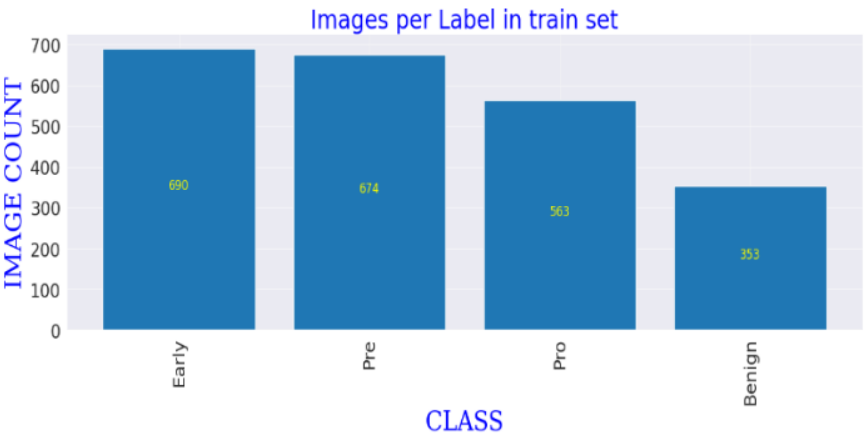

Fig. 2. Splitting Image

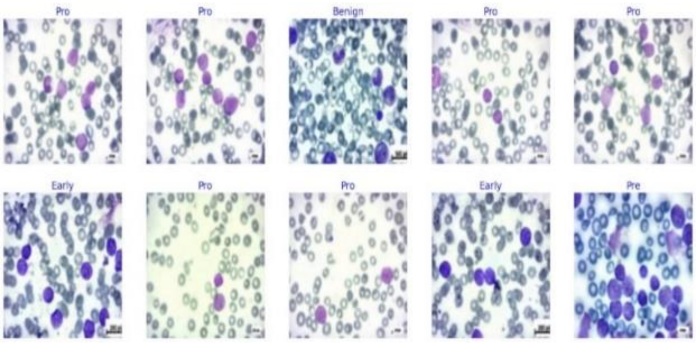

Fig. 3. Class wise Separate Image

C. Data Preprocessing

1. Grayscale preprocessing converts color images to single-channel grayscale, enhancing computational efficiency and focusing on intensity variations for analysis.

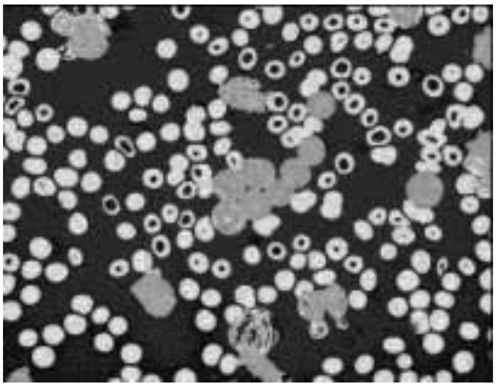

Fig. 4. Grayscale Image

2. Histogram Equalization is a method used in image processing to improve contrast by redistributing pixel intensities, enhancing details in both dark and bright areas of an image.

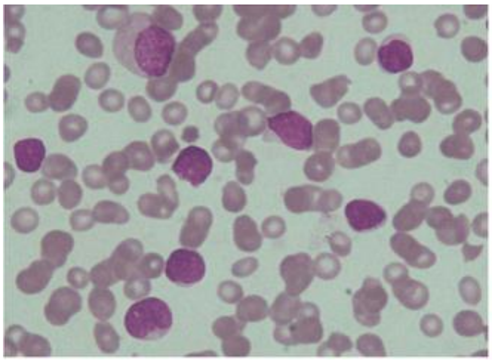

Fig. 5. Histogram Equalization Image

3. Noise reduction preprocessing technique involves filtering out unwanted distortions or random variations from data to enhance signal clarity and improve accuracy in analysis or classification tasks.

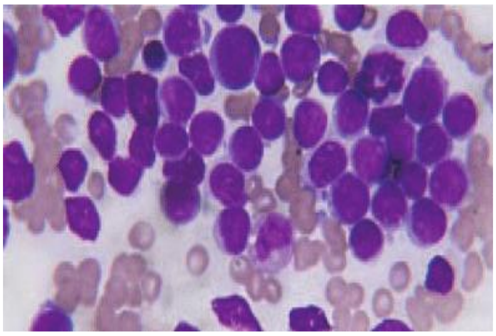

Fig. 6. Noise Reduction Image

D. Proposed Methodology

1. VGG19: VGG19 is a deep CNN architecture consisting of 19 layers, known for its simplicity and effectiveness. It comprises multiple convolutional layers with small 3x3 filters followed by max-pooling layers, enabling it to capture intricate features from input images.
2. Mobile Net V2: MobileNetV2 is a lightweight convolutional neural network (CNN) architecture designed for efficient mobile and embedded vision applications. It employs depth wise separable convolutions and linear bottlenecks to achieve high accuracy with low computational cost.
3. Mobile Net V3: Building upon its predecessor, MobileNetV3 introduces efficient inverted residuals and squeeze-and-excitation blocks to further enhance performance and computational efficiency, making it suitable for resource-constrained environments.

4. Dense Net 201: DenseNet201 is a densely connected CNN architecture characterized by dense block structures, where each layer receives direct input from all preceding layers. This design promotes feature reuse and enhances gradient flow, leading to improved feature learning and model accuracy.
5. Inception V3: Inception V3 is a deep CNN architecture known for its use of inception modules, which allow for efficient multi-scale feature extraction. It utilizes factorization into smaller convolutions to reduce computational complexity and improve performance.

EXPERIMENTAL RESULT AND DISCUSSION

A. VGG19

Table 1: Classification Report of Vgg19

Classes	Precision	Recall	F1_Score	Support
Benign	0.77	0.65	0.71	37
Early	0.96	0.72	0.83	105
Pre	0.48	1.00	0.65	102
Pro	1.00	0.61	0.57	82
Accuracy			0.63	326
Macro Avg	0.80	0.61	0.57	326
Weighted Avg	0.80	0.63	0.58	326

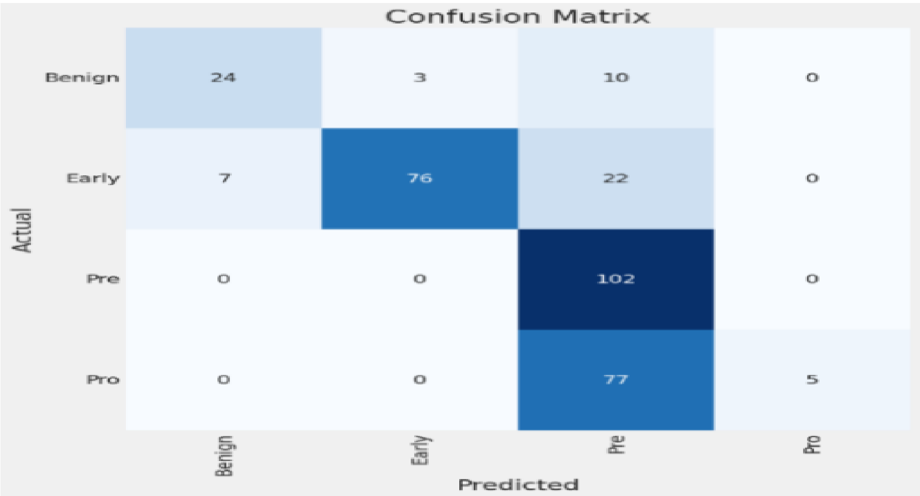

Fig. 7. Confusion Matrix for VGG19

B. MobileNet V2

Table 2: Classification Report of Mobile Net V2

Classes	Precision	Recall	F1_Score	Support
Benign	0.99	0.92	0.95	101
Early	0.96	0.98	0.97	197
Pre	1.00	0.97	0.99	193
Pro	0.96	1.00	0.98	161
Accuracy			0.98	652

Macro Avg	0.98	0.97	0.97	652
Weighted Avg	0.98	0.98	0.98	652

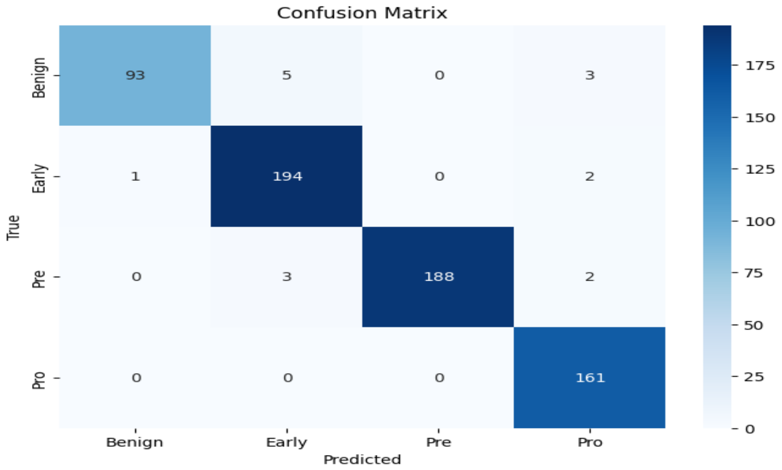

Fig. 8. Confusion Matrix for MobileNet V2

C. Mobile Net V3

Table 3: Classification Report of Mobile Net V3

Classes	Precision	Recall	F1_Score	Support
Benign	1.00	0.97	0.98	101
Early	0.98	1.00	0.99	197
Pre	1.00	0.99	0.99	193
Pro	0.99	1.00	0.99	161
Accuracy			0.99	652
Macro Avg	0.99	0.99	0.99	652
Weighted Avg	0.99	0.99	0.99	652

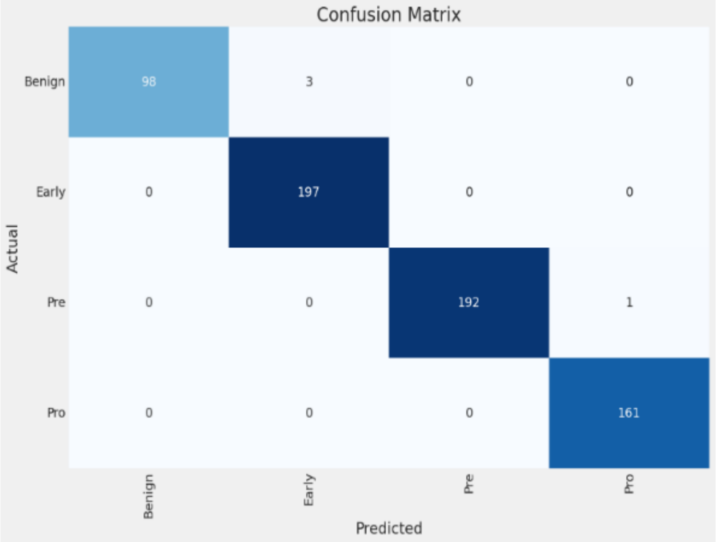

Fig. 9. Confusion Matrix for MobileNet V3

D. DenseNet 201

Table 4: Classification Report of Dense Net 201

Classes	Precision	Recall	F1_Score	Support
Benign	1.00	0.98	0.99	101
Early	0.99	1.00	0.99	197
Pre	1.00	1.00	1.00	193
Pro	1.00	1.00	1.00	161
Accuracy			0.99	652
Macro Avg	1.00	1.00	1.00	652
Weighted Avg	1.00	1.00	1.00	652

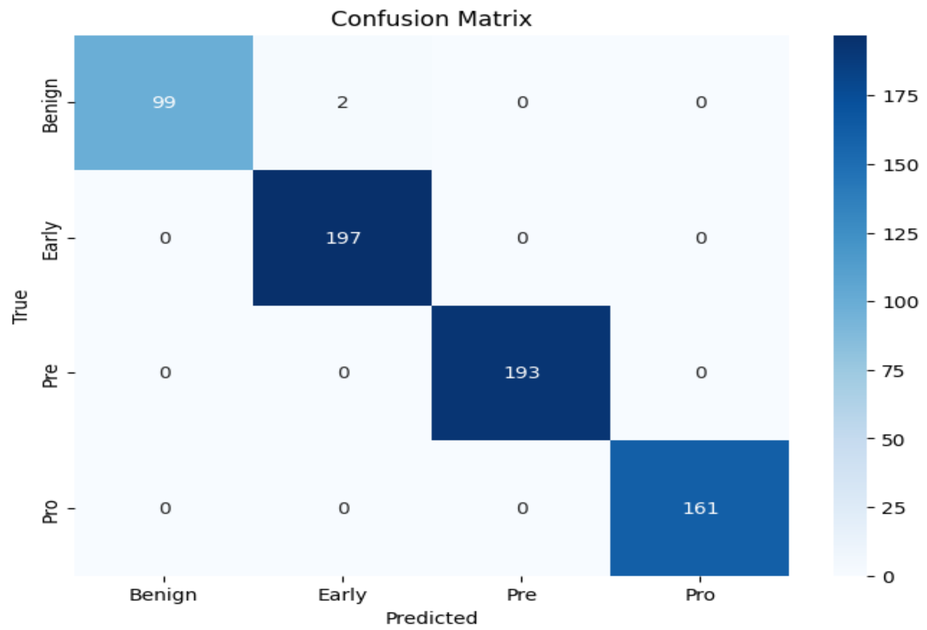

Fig. 10. Confusion Matrix for DenseNet 201

E. Inception V3

Table 5: Classification Report of Inception V3

Classes	Precision	Recall	F1_Score	Support
Benign	0.90	0.96	0.93	101
Early	0.97	0.93	0.95	197
Pre	0.98	0.96	0.97	193
Pro	0.96	0.99	0.98	161
Accuracy			0.96	652
Macro Avg	0.95	0.96	0.96	652
Weighted Avg	0.96	0.96	0.96	652

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue VIII August 2024

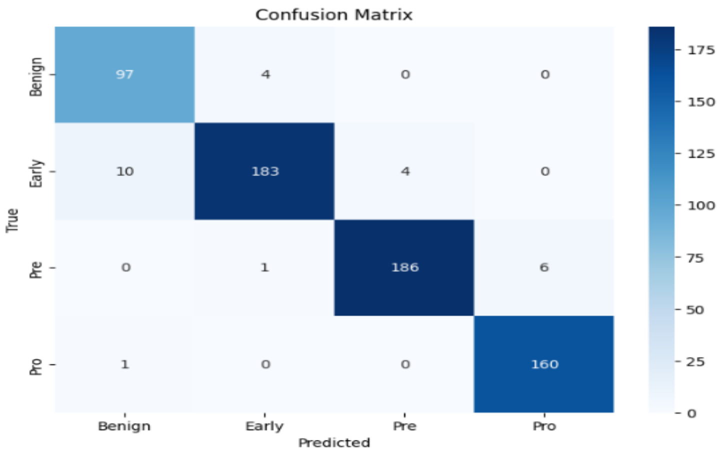

Fig. 11. Confusion Matrix for Inception V3

F. Overall Classification Report

Here the Customize CNN algorithm is used. 7 architectures of the Customize CNN algorithm are used. The 5 methods of Customize CNN we used are VGG19, MobileNetV2, MobileNetV3, DenseNet201, Inception V3. We can see here that the accuracy of VGG19 is 0.63, the accuracy of MobileNetV2 is 0.97, the accuracy of MobileNetV3 is 0.99, the accuracy of DenseNet201 is 0.99, the accuracy of Inception V3 is 0.96. Among these architectures, 2 models of MobileNet namely MobileNetV2 and MobileNetV3 are used, whose accuracy is 0.97 and 0.99 respectively. Among these architectures, 2 model of MobileNetV3 and DenseNet201 have been used, whose accuracy is 0.99 and 0.99 respectively. It can be seen from all the methods that the highest results are obtained in MobileNetV3 and DenseNet201 methods, whose accuracy is 0.99 and 0.99 respectively. The lowest results were obtained using the VGG19 model, with an accuracy of 0.63.

Table 6: Classifiers Description

Model	Accuracy	Precision	Recall	F1-score
VGG19	0.63	0.80	0.61	0.57
MobileNetV2	0.97	0.97	0.97	0.97
MobileNetV3	0.99	0.99	0.99	0.99
DenseNet201	0.99	0.99	0.99	0.99
Inception V3	0.96	0.96	0.96	0.96

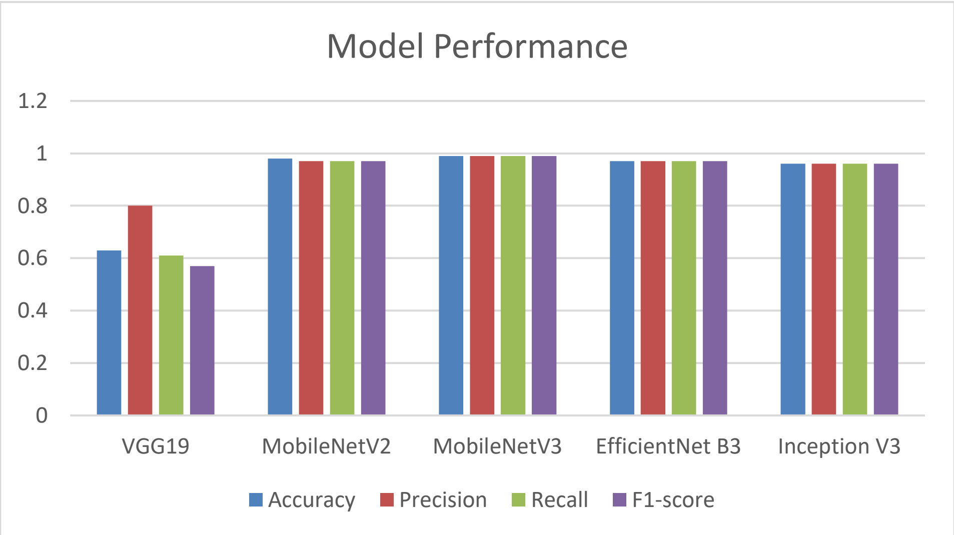

Fig. 12. Overall Model Visualization

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue VIII August 2024

Discussion

We will first collect the image and do preprocessing with that image. We will train the machine with the pre-processed image and then after the machine is trained we will test the machine and see the result. Now the machine is ready to analyze and predict any data. Now, if the machine is given an image of any leukemia disease, it can analyze and predict whether the image has leukemia disease or not. In our project, AI and machine learning, including deep learning algorithms, analyze medical image data to detect diseases like leukemia, enhancing diagnosis speed and patient education on preventive measures. The project employs a Customized CNN algorithm with 7 architectures: VGG19 (accuracy 0.63), MobileNetV2 (0.97), MobileNetV3 (0.99), DenseNet201 (0.99), and Inception V3 (0.96). Notably, MobileNetV3 and DenseNet201 achieved highest accuracies at 0.99. This approach saves time for both patients and doctors, streamlining leukemia diagnosis and treatment recommendations effectively.

CONCLUSION

In conclusion, addressing the rising prevalence of leukemia requires early detection and timely intervention. Adhering to a balanced diet, including moderate consumption of non-vegetarian and sugary foods, plays a crucial role in prevention. Scientific advancements now enable effective treatment of these diseases, supported by artificial intelligence and deep learning algorithms in our project. By leveraging the Customized CNN algorithm with architectures like MobileNetV3 and DenseNet201, both achieving high accuracies of 0.99, we demonstrate significant progress in leukemia detection. This approach not only facilitates accurate diagnosis of various leukemia types such as AML, ALL, CML, and CLL but also enhances patient education and treatment outcomes. Implementing these technologies ensures efficient healthcare delivery, saving time and improving prognosis for leukemia patients.

REFERENCES

1. F. Scotti, “Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images,” IEEE Xplore, Jul. 01, 2005. https://ieeexplore.ieee.org/document/1522835.
2. S. Shafique, Samabia Tehsin, Syed Mohammed Anas, and F. Masud, “Computer-assisted Acute Lymphoblastic Leukemia detection and diagnosis,” 2019 2nd International Conference on Communication, Computing and Digital systems (C-CODE), Mar. 2019, doi: https://doi.org/10.1109/c-code.2019.8680972.
3. S. H. Kassani, “A Hybrid Deep Learning Architecture for Leukemic B-lymphoblast Classification,” Sep. 2019.
4. Mashiat Fatma and J. Sharma, “Identification and classification of acute leukemia using neural network,” 2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom), Nov. 2014, doi: https://doi.org/10.1109/medcom.2014.7005992.
5. Luis, R. Veras, A. R. Andrade, R. Silva, Flávio H. D. Araújo, and S. de, “Unsupervised Leukemia Cells Segmentation Based on Multi-space Color Channels,” 2016 IEEE International Symposium on Multimedia, pp. 451–456, Dec. 2016, doi: https://doi.org/10.1109/ism.2016.0103.
6. D. J. Foran, D. Comaniciu, P. Meer, and L. A. Goodell, “Computer-assisted discrimination among malignant lymphomas and leukemia using immunophenotyping, intelligent image repositories, and telemicroscopy,” IEEE Transactions on Information Technology in Biomedicine, vol. 4, no. 4, pp. 265–273, 2000, doi: https://doi.org/10.1109/4233.897058.
7. A. R. J. Begum and T. A. Razak, “Diagnosing Leukemia from Microscopic Images Using Image Analysis and Processing Techniques,” IEEE Xplore, Feb. 01, 2017. https://ieeexplore.ieee.org/document/8074532.
8. K. P. Jayavikash, P. Gopinath, U. Nandheeswaran, R. Sivabalan, and T. Divya Bharathi, “Detection of Leukemia Using Machine Learning Algorithms,” Journal of Physics: Conference Series, vol. 1916, no. 1, p. 012220, May 2021, doi: https://doi.org/10.1088/1742-6596/1916/1/012220.
9. Abdul Gaffar Karim, Azhari Azhari, Mobeen Shahroz, Samir Brahim Belhaouri, and Khabib Mustofa, “LDSVM: Leukemia Cancer Classification Using Machine Learning,” Computers, materials & continua, vol. 71, no. 2, pp. 3887–3903, Jan. 2022, doi: https://doi.org/10.32604/cmc.2022.021218.

10. M. Shanbehzadeh, M. R. Afrash, N. Mirani, and H. Kazemi-Arpanahi, “Comparing machine learning algorithms to predict 5-year survival in patients with chronic myeloid leukemia,” BMC Medical Informatics and Decision Making, vol. 22, no. 1, Sep. 2022, doi: https://doi.org/10.1186/s12911-022-01980-w.
11. Y. Zhu, X. Gan, R. Qin, and Z. Lin, “Identification of Six Diagnostic Biomarkers for Chronic Lymphocytic Leukemia Based on Machine Learning Algorithms,” Journal of Oncology, vol. 2022, pp. 1–19, Nov. 2022, doi: https://doi.org/10.1155/2022/3652107..
12. S. S. Bandan, M. Rahman Ajmain, A. R. Rejuan, M. Farhana Khatun and S. A. Khushbu, “State of Survey: Advancement of Knowledge Environmental Sustainability in Practicing Administrative Apps,” 2022 13th International Conference on Computing Communication and Net- working Technologies (ICCCNT), 2022, pp. 1-8, doi: 10.1109/ICC-CNT54827.2022.9984416.
13. M. R. Ajmain, M. F. Khatun, S. S. Bandan, A. R. Rejuan, N. J. Ria and S. R. H. Noori, “Enhancing Sentiment Analysis using Machine Learning Predictive Models to Analyze Social Media Reviews on Junk Food,” 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2022, pp. 1-7, doi: 10.1109/ICC- CNT54827.2022.9984355.
14. M. A. R. Rejuan, S. S. Bandan, M. A. Rakib and M. Assaduzzaman, “A Comparative Study For Measuring The Quality of Dhaka City Transportation System: Survey Based,” 2023 International Conference for Advancement in Technology (ICONAT), Goa, India, 2023, pp. 1-8, doi: 10.1109/ICONAT57137.2023.10080188.
15. Sheikh Sadi Bandan, Sabid Ahmed Sunve, and Shaklian Mostak Romel, “A Deep Learning Approach for Bengali News Headline Categorization,” Jul. 2023, doi: https://doi.org/10.1109/icccnt56998.2023.10307776.