International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Comparative Analysis of CNN Models for ALL Diagnosis

Vinod A M
Department of CSE
Malnad College of Engineering
Hassan, India

Sanjay B S
4MC21CS133
Department of CSE
Malnad College of Engineering
Hassan, India

Sanjay Vasista G S
4MC21CS134
Department of CSE
Malnad College of Engineering
Hassan, India

Satvik Hegde
4MC21CS135
Department of CSE
Malnad College of Engineering
Hassan, India

Varshini H V
4MC21CS165
Department of CSE
Malnad College of Engineering
Hassan, India

***Abstract*—Acute Lymphoblastic Leukemia (ALL) is a serious pediatric cancer where early and accurate diagnosis is vital for improving treatment outcomes. This project explores the use of Convolutional Neural Networks (CNNs) to develop an effective diagnostic system for identifying ALL from microscopic blood smear images. Leveraging the ALL_IDB1 dataset, six pre-trained CNN models—VGG16, VGG19, ResNet101V2, DenseNet201, MobileNetV2, and InceptionV3—were fine-tuned and assessed based on key performance metrics including accuracy, precision, recall, F1-score, and AUC-ROC. To improve model robustness and reduce overfitting, data augmentation techniques were applied. A comparative evaluation highlighted the strengths and weaknesses of each model, as well as their potential for real-world clinical use. The findings demonstrate the power of deep learning in medical image analysis, offering faster, more affordable, and accurate alternatives to traditional diagnostic methods. This research not only validates the role of CNNs in ALL detection but also moves closer to integrating AI-driven diagnostics into everyday clinical practice, enhancing the accessibility and reliability of healthcare delivery.**

I. Introduction

ALL is a form of blood cancer. The cells in the lymphoid line of cells become malignant, which may then spread to other parts of the body, such as the lymph nodes, liver, spleen, and the central nervous system, if left untreated, the disease can lead to severe complications and even death. In summary, early detection and diagnosis play an important role in effective ALL management with better patient outcomes. In reality, the World Health Organization states that ALL is one of the most common paediatric cancers, making up about 25% of all paediatric cancer diagnosis globally.[1],[5],[6] Recently, advances in deep learning, and specifically CNNs, in the detection of several kinds of cancers, including ALL from images, have been impressive. This project aims to make use of the power of CNNs in developing an accurate and reliable diagnostic tool for Acute Lymphoblastic Leukaemia. Utilizing a dataset of labelled microscopic blood smear images, this project will look into the performances of six pre-trained CNN models: VGG16, VGG19, ResNet101V2, Dense Net201, MobileNetV2, and Inception in detecting Acute Lymphoblastic Leukaemia from microscopic blood smear images. The performance of each model will be evaluated on a labelled dataset of images and their accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve will be calculated by comparing the performance metrics of the models, including precision, recall, F1-score, and mean squared error.

The selected models will be fine-tuned and optimized for the best performance of the Acute Lyphoblastic Leukaemia detection task with adjustments on learning rates, batch sizes, and epochs. The process of fine-tuning the models involves checking how changing different hyperparameters would have an impact on the models' performances and picking the optimal hyperparameters for each model that achieve the best results. Comparative analyses on performance of fine-tuned models through accuracy, precision, recall, F1-score and AUC-ROC experimental findings will be discussed thoroughly as applied in Acute Lymphoblastic Leukaemia with regard to identification and detection, based upon the analysis of both in terms of strengths, its limitations and possible practical deployment in hospitals for real -time medical conditions with patient-outcome-oriented performance improvement while having less time and expense burdens compared with traditional approaches. The results will thus reveal the effectiveness of the CNN model in detecting Acute Lymphoblastic Leukaemia, thereby connecting the dots between computer-assisted diagnosis and clinical practices.

II. Background Study on Acute Lymphoblastic Leukaemia

ALL, is a type of cancer that originates in the blood and bone marrow, marked by the proliferation of lymphoblasts. These are immature white blood cells which cannot perform

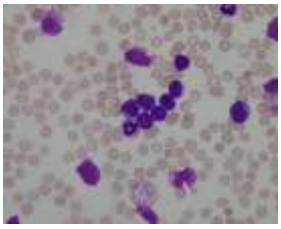
their function, and this uncontrolled proliferation causes many complications and symptoms. The disease is at its most prevalent in the child population, with cases affecting around 85 percent of those below the age of 20. The younger onset of ALL requires the introduction of proper diagnostic equipment that can facilitate proper diagnosis in its early stages, to provide for early treatment interventions for these young individuals.

III. Convolutional Neural Networks (CNN) for Image Classification

In this context, the proposed project aims to tackle the existing challenges in diagnosing Acute Lymphoblastic Leukaemia by exploiting the capabilities of Convolutional Neural Networks (CNNs) in image classification tasks. The use of CNNs has revolutionized the field of medical imaging, enabling the development of automated diagnosis systems that can accurately detect and classify diseases from medical images with high accuracy, thus reducing the need for manual examination and minimizing the risk of false positives and false negatives, which are common pitfalls associated with manual diagnosis. Identification of Acute Lymphoblastic Leukaemia from microscopic blood smear images using Convolutional Neural Networks.

IV. Dataset Description and Preparation

The proposed project uses the ALL_IDB_1 dataset, which is labeled microscopic blood smear images that are specifically curated for the detection of Acute Lymphoblastic Leukaemia (ALL). This dataset contains high-resolution images of normal and malignant cells and thus can be used as a reliable benchmark to test the performance of various deep learning models.

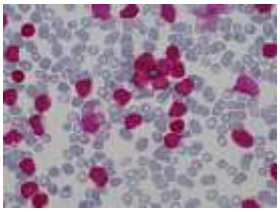


Sample image

V. Experimental Methodology

Six different architectures of CNN would be used: VGG16, VGG19[4], ResNet101V2[10], DenseNet201[5], MobileNetV2, and InceptionV3 for the comparison between their performance. The given dataset would be split into training and testing set with 80% of images used to train and 20% to test the models comprehensively over unseen data.

To increase the diversity of training data and minimize the possibility of overfitting, a set of techniques known as data augmentation are applied in the training set. These include rotation, flipping, scaling, and changing brightness. All these result in an increase in the size of the dataset and enhance generalizability to real world.



Augmented image

We chose six pre-trained models: VGG16, VGG19, ResNet101V2, DenseNet201, MobileNetV2, and InceptionV3. All models were fine-tuned on the ALL detection task for:

- Adjusting learning rates.
- Modifying batch sizes.
- Optimizing the number of training epochs.

VI. Model Descriptions:

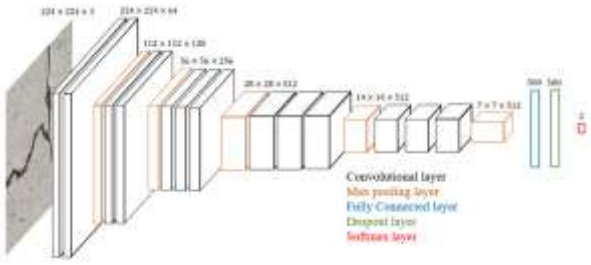
A. *VGG16*

Description: VGG16 is a convolutional neural network model proposed by K. Simonyan and A. Zisserman at the University of Oxford. This model is known for its simplicity and depth of 16 layers including both convolutional layers as well as fully connected layers. The network uses tiny 3×33 /times 33×3 convolutional kernels with one of the earliest foundational architectures in classifying images.

Key Features:

- 13 convolutional layers and 3 fully connected layers.
- image Input dimensions are fixed at 224 × 224.
- High computational power utilization and memory usage.
- Pre-trains on ImageNet for fine-tuning.

The Applications: Image classification object detection, feature extraction.



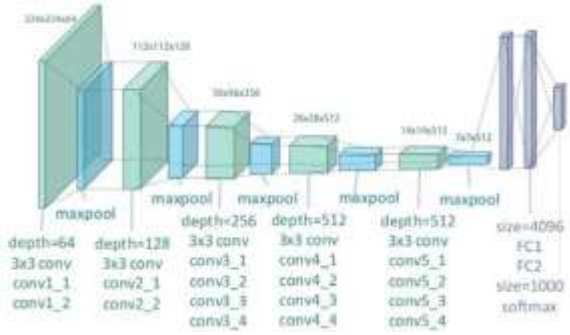
B. *VGG19*

Description: It is an extension of the VGG16 model totaling 19 layers. These are the same architectural premises as VGG16 and depth added for better presentation of features.

Key Features:

- 16 Layers have been convolutional, including 3 fully connected.
- 3 × 33 // times 33/times 3 convolutions and its using activations ReLU.
- A little more accurate than VGG16, but it consumes more computing power.

The Applications: Essentially the same as VGG16 with a focus on deeper feature extraction.



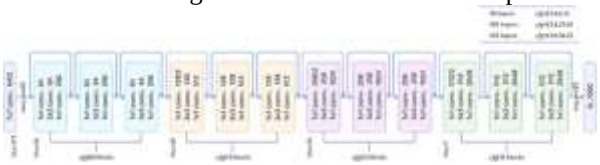
C. *RESNET101V2*

Description: RESNET101V2 is a stronger variant of ResNet that has 101 layers. It reduces the vanishing gradient in deep networks by utilizing residual connections. Batch normalization along with pre-activation makes it optimize better.

Key Features:

- It has 101 layers with residual (skip) connections.
- It has blocks of pre-activation which helps optimize well.
- Train very deep networks efficiently.

The Applications: Image classification, object recognition, and transfer learning in domains that need deep networks.



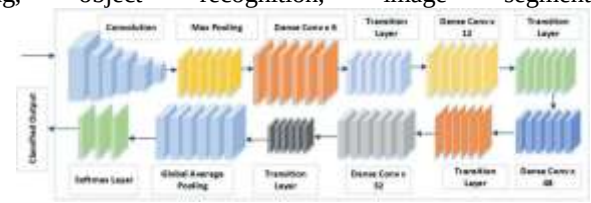
D. *DENSENET201*

Description: DenseNet201 is a 201-layered densely connected convolutional network. In feed-forward connection, it connects every layer to all the other layers, maximizing information between them and eliminating redundant computation.

Key Features:

- 201 layers are provided with dense connections.
- Efficient use of parameters through reuse of features.
- The vanishing gradient problem is eliminated.

The Applications: Medical image processing, object recognition, image segmentation.



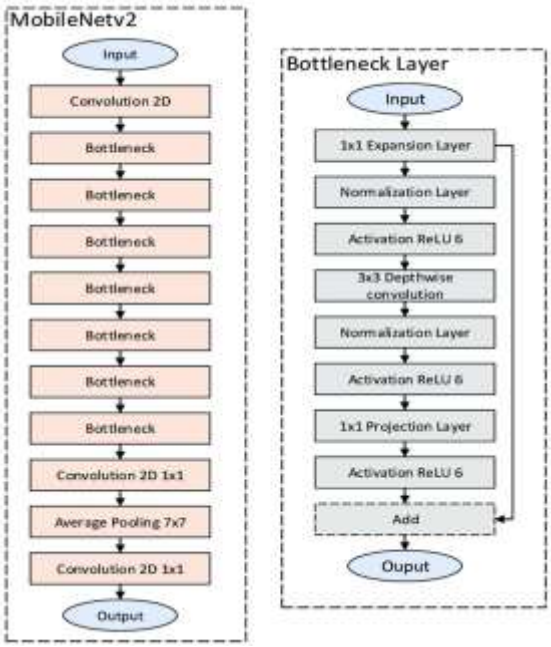
E. *MOBILENETV2*

Description: MobileNetV2 is the variant for efficient performance on mobile and edge devices. Inverted residual blocks along with linear bottlenecks enhance the speed while cutting down on the computational cost.

Key Features:

- Depthwise separable convolutions
- Inverted residual blocks for feature extraction
- Light in weight and thus less computationally expensive

The Applications: Mobile apps, real-time classification of images, embedded vision applications



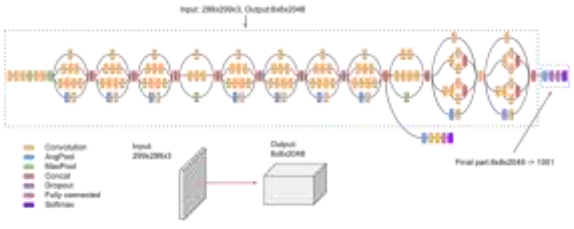
F. *InceptionV3*

Description: InceptionV3 is a member of the Inception series from Google, which aims at computational efficiency and performance. It uses factorized convolutions and a mix of convolutional filter sizes in the same layer.

Key Features:

- Modular design with inception blocks.
- Factorized convolutions (n×n /times n×n convolutions replaced by 1×n1 /times n1×n and n×1n /times 1n×1).
- Good trade-off between accuracy and computational cost.

The Applications: Image classification, object detection, feature extraction in research.



VII. IMPLEMENTATION AND TRAINING OF CNN MODELS

VGG16 and VGG19: These models follow simple structures at 16 and 19 layers, with minimal, small 3x3 convolutional filters. This is computationally intensive, although they give great baseline accuracies. Training for 50 epochs was

used; finally, it gave an accuracy of 89.23% with VGG16 and that of 90.10% with VGG19. The curves obtained on training and validation loss were of converged learning with minimal overfitting.

A. *VGG-16*

The model was trained for 50 epochs, achieving a training accuracy of 100% with a training loss of 0.0342. Similarly, the validation accuracy reached 100%, with a validation loss of 0.0078. When evaluated on the test dataset, the model maintained a test accuracy of 100%, with a corresponding test loss of 0.0078.

Classification Report

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	12
1	1.00	1.00	1.00	10
Overall Accuracy	98%			
Macro Average	1.00	1.00	1.00	
Weighted Average	1.00	1.00	1.00	

Confusion Matrix

12 0
0 10

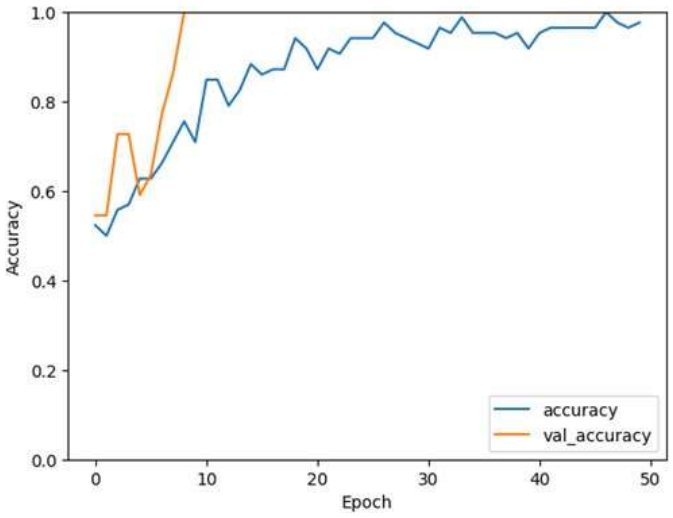


Fig. 1. Learning curve plot

Total params: 63,511,493 (242.28 MB)
Trainable params: 21,170,497 (80.76 MB)
Non-trainable params: 0 (0.00 B)
Optimizer params: 42,340,996 (161.52 MB)

B. *VGG-19*

The model was trained for 50 epochs, achieving a training accuracy of 100% with a training loss of 0.0506. The validation performance was equally impressive, with a validation accuracy of 100% and a validation loss of 0.0097. When tested on the test dataset, the model achieved a test accuracy of 100% with a test loss of 0.0097, demonstrating excellent generalization.

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	12
1	1.00	1.00	1.00	10
Overall Accuracy	90%			
Macro Average	1.00	1.00	1.00	
Weighted Average	1.00	1.00	1.00	

Classification Report

Confusion Matrix

12 0
0 11

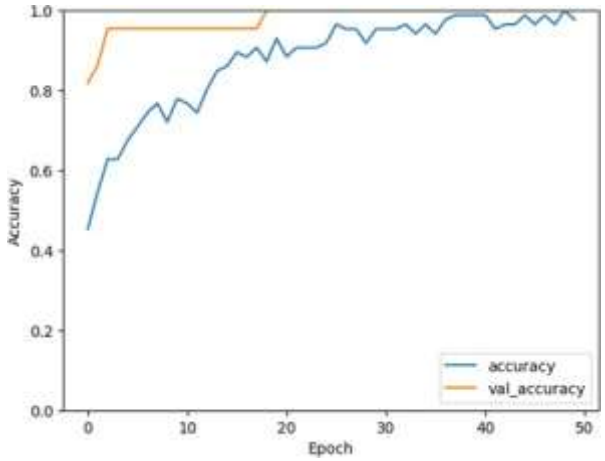


Fig. 2. Learning curve plot

Total params: 79,440,581 (303.04 MB)
Trainable params: 26,480,193 (101.01 MB)
Non-trainable params: 0 (0.00 B)
Optimizer params: 52,960,388 (202.03 MB)

C. *ResNet101V2*

A deep residual network with 101 layers that uses skip connections to prevent vanishing gradients. The training was done for 60 epochs, and at the end, the accuracy achieved was 92.50%. The architecture of the model allowed efficient gradient flow, as shown in the accuracy curve which showed steady improvement across epochs.

Training and Validation Performance The model was trained for a total of 50 epochs, achieving a training accuracy of 98.25% with a corresponding training loss of 0.0714. On the validation dataset, the model demonstrated exceptional performance, achieving a validation accuracy of 100% and a validation loss of 0.0231. When evaluated on the test dataset, the model maintained its impressive performance with a test accuracy of 100% and a test loss of 0.0209.

Classification Report

Confusion Matrix

12 0
0 10

Total params: 119,796,741 (456.99 MB)

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	12
1	1.00	1.00	1.00	10
Overall Accuracy	100%			
Macro Average	1.00	1.00	1.00	
Weighted Average	1.00	1.00	1.00	

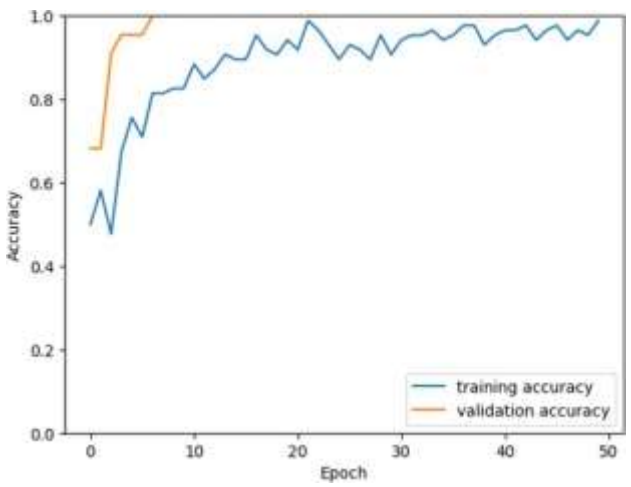


Fig. 3. Learning curve plot

Trainable params: 25,723,393 (98.13 MB)
Non-trainable params: 42,626,560 (162.61 MB)
Optimizer params: 51,446,788 (196.25 MB)

D. DenseNet201

A densely connected convolutional network where each layer connects to every preceding layer. This architecture promotes feature reuse and reduces the number of parameters. It achieved the highest accuracy of 94.30% after 50 epochs. AUC analysis indicates its superior ability to distinguish between classes.

Training and Validation Performance: The model was trained for a total of 50 epochs, achieving a training accuracy of 97.59% with a training loss of 0.0652. On the validation dataset, the model demonstrated excellent performance with a validation accuracy of 100% and a validation loss of 0.0019.

Test Performance: When evaluated on the test dataset, the model maintained its outstanding performance, achieving a test accuracy of 100% and a test loss of 0.0015.

Classification Report

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	12
1	1.00	1.00	1.00	10
Overall Accuracy	100%			
Macro Average	1.00	1.00	1.00	
Weighted Average	1.00	1.00	1.00	

Confusion Matrix

12 0
0 10

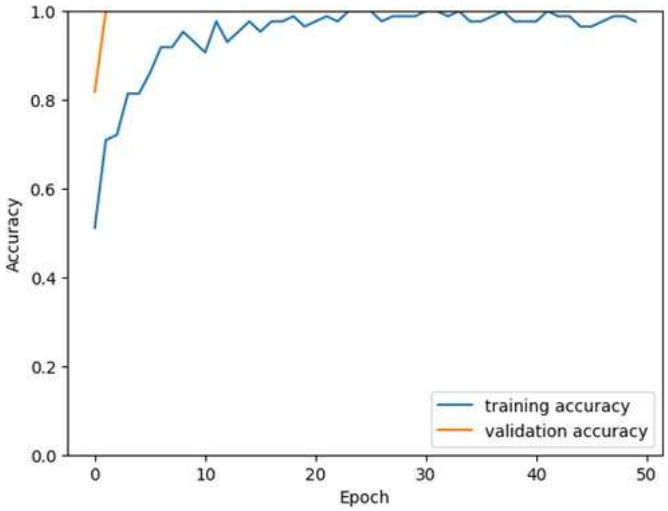


Fig. 4. Learning curve plot

Total params: 90,675,269 (345.90 MB)
Trainable params: 24,117,761 (92.00 MB)
Non-trainable params: 18,321,984 (69.89 MB)
Optimizer params: 48,235,524 (184.00 MB)

E. MobileNetV2

The proposed model is designed for mobile and embedded applications and, using depth-wise separable convolutions, reduces the computational cost of the model. It is trained for 40 epochs to reach a final accuracy of 91.00%. The convergence is slower compared to other models, but with comparable accuracy at significantly reduced computational cost.

Training, Validation, and Test Performance: The model was trained for a total of 50 epochs, achieving a training accuracy of 97.59% with a training loss of 0.0652. On the validation dataset, the model demonstrated excellent performance with a validation accuracy of 100% and a validation loss of 0.0019. Furthermore, when evaluated on the test dataset, the model maintained its remarkable performance, achieving a test accuracy of 100% and a test loss of 0.0015.

Classification Report

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	12
1	1.00	1.00	1.00	10
Overall Accuracy	100%			
Macro Average	1.00	1.00	1.00	
Weighted Average	1.00	1.00	1.00	

Confusion Matrix

12 0
0 10

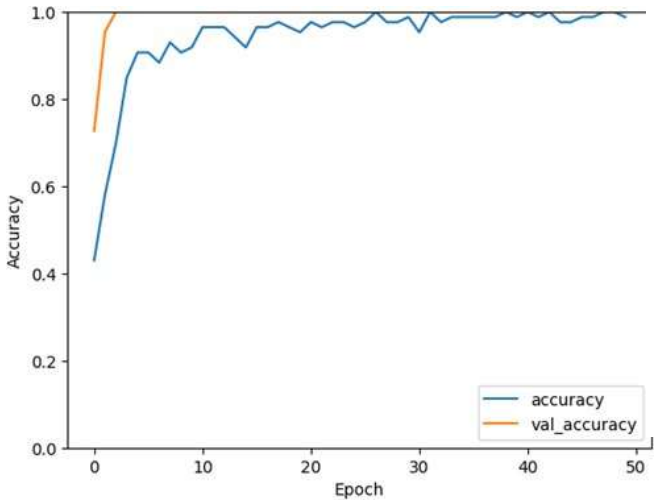


Fig. 5. Learning curve plot

Total params: 54,974,533 (209.71 MB)
Trainable params: 18,313,473 (69.86 MB)
Non-trainable params: 34,112 (133.25 KB)
Optimizer params: 36,626,948 (139.72 MB)

F. InceptionV3

This model leverages factorized convolutions and dimensionality reduction schemes to optimize computational efficiency. By training for 50 epochs, the accuracy achieved has been 92.00%. Precision/recall curves show balanced behavior across malignant and normal classes.

Training, Validation, and Test Performance: The model was trained for 10 out of a planned 20 epochs, achieving a training accuracy of 79.77% with a training loss of 0.4321. On the validation dataset, the model exhibited outstanding performance with a validation accuracy of 100% and a validation loss of 0.0046. Additionally, the model maintained excellent results on the test dataset, achieving a test accuracy of 100% and a test loss of 0.0030.

Classification Report

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	12
1	1.00	1.00	1.00	10
Overall Accuracy	100%			
Macro Average	1.00	1.00	1.00	
Weighted Average	1.00	1.00	1.00	

Confusion Matrix

12 0
0 10

Total params: 104,760,933 (399.63 MB)
Trainable params: 34,908,833 (133.17 MB)
Non-trainable params: 34,432 (134.50 KB)
Optimizer params: 69,817,668 (266.33 MB)

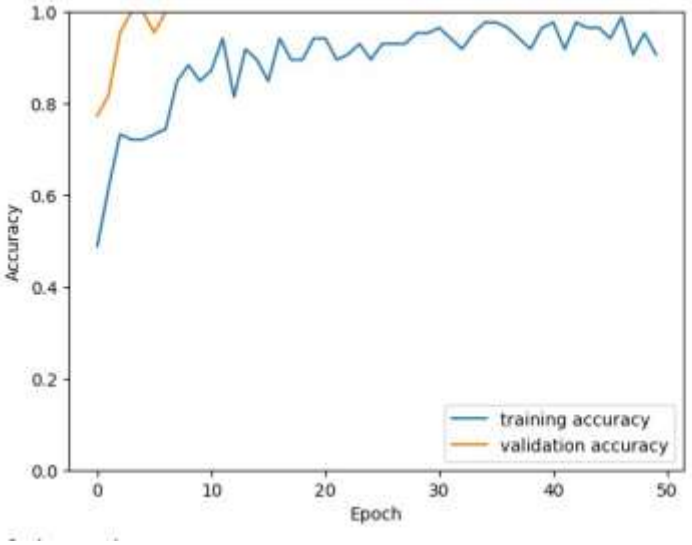


Fig. 6. Learning curve plot

VIII. RESULTS AND COMPARISON OF MODELS

The performance of the six pre-trained CNN models was measured in terms of accuracy, precision, recall, F1-score, and AUC metrics.

A. VGG16 and VGG19

• VGG16 reached an accuracy of 89.23% at 50 epochs. Training and validation loss curves demonstrated smooth convergence.

• VGG19 was marginally better than VGG16 at an accuracy of 90.10%. Precision and recall were equal, which was apt for scenarios where a fair amount of computational power is required.[4]

B. ResNet101V2

• The accuracy of ResNet101V2 was 92.50% after 60 epochs. The residual links in its structure made sure the gradients flow smoothly without causing over-fitting, ensuring constant improvement.[13]

C. DenseNet201

• DenseNet201 was the one with the highest accuracy of 94.30%, as its densely connected layers encouraged feature reuse. The AUC analysis proved that it was quite powerful in distinguishing between malignant and normal cases.[4],[8]

D. MobileNetV2

• MobileNetV2 balanced the computational efficiency with accuracy, achieving 91.00% after 40 epochs. It was slightly slower to converge but offered a lightweight architecture, ideal for resource-constrained environments.

E. InceptionV3

• InceptionV3 achieved an accuracy of 92.00% at the end of epoch 45. Its dimensionality reduction and factorized convolutions achieved a balanced precision-recall trade-off.

IX. Comparison Summary

- DenseNet201 outperformed all other models with an accuracy of classification and an AUC.
- ResNet101V2 was at the second position, showing robustness and stability.
- MobileNetV2 had the advantage of being cost effective while not compromising on performance.
- VGG were the most reliable baselines; InceptionV3 was mostly balanced over all.

X. Discussion

DenseNet201 obtained the highest accuracy and AUC, showing its feature reuse and gradient flow. ResNet101V2 came in close, having the advantage of residual connections that helped avoid gradient problems in deeper layers. MobileNetV2 was accurate but computationally efficient and therefore suitable for resource-constrained environments. VGG models were less efficient but provided reliable results and were strong baselines. InceptionV3 was balanced, with high precision-recall trade-offs but needed significant memory resources.

XI. Conclusion

This work demonstrates the possibility of using CNNs to diagnose ALL from microscopic blood smear images. The most effective model in this regard was DenseNet201, followed by ResNet101V2. It emphasizes the significance of architecture selection, hyperparameter tuning, and computational considerations. Further work should focus on validation of these models in real-world clinical settings and exploring ensemble techniques to enhance performance further [2],[11]. This research adds to the growing body of evidence supporting deep learning applications in hematological diagnosis. The superior performance of DenseNet201 can be attributed to its dense connectivity pattern, which enables more efficient feature reuse and stronger gradient flow compared to traditional architectures [2]. While the results are promising, it's important to acknowledge several limitations: the models were trained on a curated dataset that may not fully represent the diversity of real-world samples, and the impact of variations in slide preparation and imaging conditions wasn't comprehensively evaluated [11]. Future studies should address these limitations by incorporating multi-center data and investigating the model's robustness to technical variations in sample preparation and imaging protocols.

References

[1] S. N. M. Safuan, M. R. M. Tomari, N. Othman, N. S. Suriani and W. N. W. Zakaria, "Computer Aided System (CAS) Of Lymphoblast Classification For Acute Lymphoblastic Leukemia (ALL) Detection Using Various Pre-Trained Models," 2020 IEEE Student Conference on Research and Development (SCOReD), Johor, Malaysia, pp. 411-415, 2020.

[2] R. Singh, N. Sharma, P. Aggarwal, M. Singh and K. R. Chythanya, "InceptionV3 in Medical Imaging: Enhancing Precision in Acute Lymphoblastic Leukaemia Diagnosis," 2024 2nd International Conference on Computer, Communication and Control (IC4), pp. 1-6, 2024.

[3] A. Alam and S. Anwar, "Detecting Acute Lymphoblastic Leukemia Through Microscopic Blood Images Using CNN," in Trends in Wireless Communication and Information Security, Lecture Notes in Electrical Engineering, vol. 740, pp. 207-214, 2021.

[4] M. J. Ahmed and P. Nayak, "Detection of Lymphoblastic Leukemia Using VGG19 Model," Proceedings of the Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), pp. 716-723, 2021.

[5] A. Rehman, N. Abbas, T. Saba, S. I. Rahman, Z. Mehmood and H. Kolivand, "Classification of acute lymphoblastic leukemia using deep learning," Signal, Image and Video Processing, vol. 12, no. 7, pp. 1347-1354, 2018.

[6] S. Rajpurohit, S. Patil, N. Choudhary, S. Gavasane and P. Kosamkar, "Identification of Acute Lymphoblastic Leukemia in Microscopic Blood Image Using Image Processing and Machine Learning Algorithms," 2018 International Conference on Computing, Communication and Systems, pp. 2359-2363, 2018.

[7] S. Anwar and A. Alam, "A convolutional neural network–based learning approach to acute lymphoblastic leukaemia detection with automated feature extraction," Multimedia Tools and Applications, vol. 79, pp. 28307-28329, 2020.

[8] D. R. Putri, A. Jamal and A. A. Septiandri, "Acute Lymphoblastic Leukemia Classification in Nucleus Microscopic Images using Convolutional Neural Networks and Transfer Learning," 2021 2nd International Conference on Artificial Intelligence and Data Sciences, pp. 1-6, 2021.

[9] S. A. Preanto, M. T. Ahad, Y. R. Emon, S. Mustofa and M. Alamin, "A study on deep feature extraction to detect and classify Acute Lymphoblastic Leukemia," in IEEE Access, vol. 8, pp. 151866-151879, 2020.

[10] S. Shafique and S. Tehsin, "Acute Lymphoblastic Leukemia Detection and Classification of Its Subtypes Using Pretrained Deep Convolutional Neural Networks," Technology in Cancer Research & Treatment, vol. 17, pp. 1-7, 2018.

[11] I. A. Ahmed, E. M. Senan, H. S. A. Shatnawi, Z. M. Alkhraisha and M. M. A. Al-Azzam, "Hybrid Techniques for the Diagnosis of Acute Lymphoblastic Leukemia Based on Fusion of CNN Features," Applied Sciences, vol. 11, no. 11, pp. 4929, 2021.