A Comparative Analysis of Transfer Learning in State-of-the-Art Pre-Trained Models on the C-NMC Leukemia Dataset.

Aniketh Vijesh
School of Computing
Amrita Vishwa Vidyapeetam
Amritapuri,Kerala,India
Email: anikethvij464@gmail.com

Kirti Sikka
School of Computing
Amrita Vishwa Vidyapeetam
Amritapuri,Kerala,India
Email: kirtisikka972@gmail.com

***Abstract*—The early detection of Acute Lymphoblastic Leukemia (ALL) poses a significant challenge in the medical field due to the subtle morphological features of ALL cells, which often resemble healthy cells. This necessitates the expertise of experienced hematologists, a reliance on human interpretation that introduces subjectivity and labor-intensive processes. Consequently, timely diagnosis and treatment initiation can be hindered. By leveraging the capabilities of machine learning, we aim to establish a system that can accurately distinguish between healthy and ALL cells, thereby reducing the reliance on subjective human interpretation and expediting the diagnostic process. This article encapsulates a summary and comparison of diverse automated detection and classification methods for acute leukemia, examining seven distinct models, namely: VGG16, VGG19, Inception,Xception, Efficient NetB0, ResNet50, and ResNet101. Of these models, ResNet101 came as the top performer with a Validation Accuracy of 76.36%, Validation Precision of 75.85%, and Validation Recall of 76.36%. This comparative analysis aims to elucidate the strengths and weaknesses of these models, contributing valuable insights.**

1. Introduction

Blood contains mainly three types of cells, Red Blood Cells (RBC), White Blood Cells (WBC), and platelets. Red Blood Cells are important for the transport of oxygen transport from the heart to all tissues and carry away carbon dioxide. They comprise up to 50% of the total volume of blood. White Blood cells are tasked with the important function of defending the body against infections and diseases. Leukemia is a type of blood cancer originating from the bone marrow. It is characterised by an excessive amount of blood cells, which are called blast or leukemia cells. This causes many issues like bleeding, bone pain, fatigue, fever, and an increased susceptibility to infections due to a deficiency of normal blood cells.

Leukemia is broadly classified into four main types based on the type of white blood cell affected and the rate of progression: Acute Lymphoblastic Leukemia(ALL), Acute Myeloid Leukemia(AML), Chronic Lymphoblastic Leukemia(CLL), and Chronic Myeloid Leukemia(CML)[1].

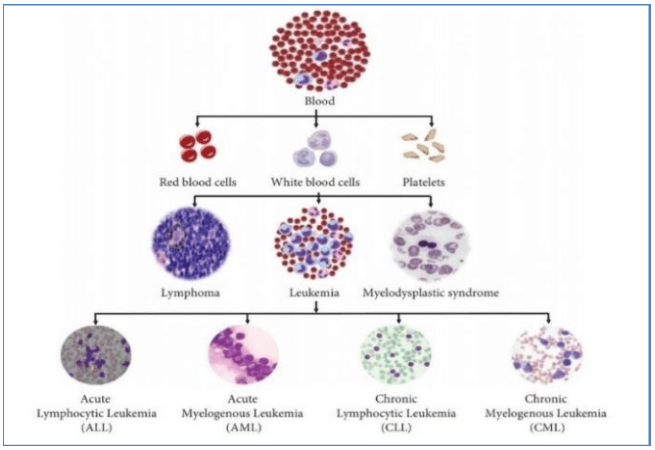

Figure 1. Types of Leukemia

Acute Lymphoblastic Leukemia(ALL) is a common cancer disease that affects mostly children below the age of fifteen. The main reason is the rapid production of immature white blood cells, which decrease the number of good blood cells and platelets, making it harder for the body to fight against infections and diseases. As the name implies, this disease is ”acute”, that is it advances quickly and rapidly. This may result in the disease spreading to different body parts like the lymph nodes, spleen, liver, brain, spinal cord, etc.

Due to the urgency and importance of early-stage detection, it is paramount that the detection is done in a precise and cost-efficient manner. This has proven to be a significant obstacle. Most procedures available in diagnostic laboratories are time-consuming and based on the experience of the hematologists. To address these challenges, our research focuses on harnessing the capabilities of machine learning to develop a system capable of precise and efficient detection and classification of ALL cells.

This paper presents a comprehensive comparative analysis of seven distinct machine learning models, namely VGG16, VGG19, Inception, Xception, Efficient NetB0, ResNet101, and ResNet50. By examining the strengths and

weaknesses of these models, we aim to contribute valuable insights into the development of an automated system for ALL detection. Our objective is to reduce reliance on subjective human interpretation, expedite the diagnostic process, and ultimately improve outcomes for individuals affected by Acute Lymphoblastic Leukemia. The following sections provide an in-depth exploration of each model's performance and its potential applications in the context of leukemia detection.

The experiments and models were trained in a Python environment, with the Python version being 3.10.12. The Keras library and TensorFlow libraries were used to aid in training the models. Other libraries used were, numpy, scikit-learn, seaborn, etc.

The subsequent sections of this paper will delve deeper into the existing literature related to leukemia detection and classification in Section 2, followed by a detailed exploration of our proposed methodology in Section 3. Section 4 will present the results obtained from our comparative analysis of machine learning models and draw conclusions based on their performance.

2. Literature Review

Extensive research has been carried out in the field. Different teams of researchers have come up with different algorithms and models to come up with an efficient solution.

Agustin Arif Sukorini et al. [2] have proposed the use of Particle Swarm algorithms and neural networks as a possible solution. The paper first processes the images and then using swarm algorithms, feature extraction is completed. There are two stages for classification, the]first deals with the classification of the cells into lymphoid and non-lymphoid cells, and the second classification step classifies the lymphoid cell into malignant or non-malignant cells. This approach produced an accuracy of 86.92%.

Talaat Gamerl et al [3] used a smaller dataset, hence they used various data augmentation techniques to make a workaround. After data augmentation, feature extraction was carried out using a Convolutional Neural Network(CNN) using the ReLU activation function. The results are then passed to an attention module to extract more features. Finally, the weights received from the attention module are used by another model for the final classification into benign or malignant cells. This approach generated the following results, Precision of 99.97%, Recall of 100%, F1-Score of 99.98%, and an Accuracy of 99.98%.

Sahlol, Kollmannsberger, Ewees et al. [4]This study introduces a robust leukemia detection model. Leveraging transfer learning, it establishes a foundation of features for focused image analysis. A novel Salp Swarm Optimization algorithm is employed, comprising leader and follower components. Correlated features are removed using Chi-square, and recursive feature elimination refines the subset based on a regression model. The model, utilizing tree-based classifiers, achieves an accuracy of 83.2%, comparable to established Convolutional Neural Networks like ResNet, but with the advantage of extracting fewer features.

Hariprasath Dharani Mohammad et al.[5]proposed a streamlined methodology for leukemia cell classification. It begins with image preprocessing, converting RGB images to CMYK format, and applying histogram equalization or contrast stretching. Image segmentation follows by using Zack thresholding or the triangle method to distinguish foreground and background regions. Feature extraction focuses on statistical and texture features from the cell's nucleus. Subsequently, an SVM or KNN classifier categorizes samples into healthy and malignant classes, validated using K-fold cross-validation on the ALL-IDB dataset. In noisy training samples, the KNN classifier achieves 90.33% accuracy, 90% sensitivity, and 90.9% specificity, while the SVM-R classifier attains 91.5% accuracy, 90% sensitivity, and 92% specificity. The dataset employed is ALL-IDB1.

Genovese Hosseini Piuri Plataniotis Scotti et al[6] introduce a multi-step image processing and classification methodology for leukemia detection. Initially, image registration is employed to normalize the radius of each cell image in the database. Subsequently, an adaptive preprocessing technique focuses on improving sharpness quality based on focus quality estimation. The approach performs adaptive unsharpening by tuning a unique parameter, th0 unsharp, using a shallow CNN. The final adaptive unsharpening is applied to all images in the dataset. Classification is carried out using a pre-trained deep CNN, fine-tuned for binary classification (normal or lymphoblast). Qualitative results demonstrate the effectiveness of VGG16 fine-tuned on ALL-IDB2unsharp, achieving the highest classification accuracy (96.84%). Quantitative results using Grad-CAM indicate improved and more accurate evaluations when employing the proposed algorithm on the dataset.

The above papers discuss various pre-processing methods and model architectures. Each paper introduces a novel technique to come up with a solution, some of them use pre-trained models like VGG16 to achieve results. However, it has to be noted that the performances of the plethora of pre-trained models themselves have not been compared and studied. This paper aims to find the objectively best-pre-trained model from a set of popular models, which will aid future researchers to focus on the best models and augment the research process.

3. Methodology

The steps involved in the proposed methodology are described in Figure 2. Each step is explained in the following sections in detail.

3.1. Data Acquisition

The initial step in the process is the collection of the images. The dataset used is the C-NMC dataset.

3.1.1. Dataset. The dataset used in this study is the publicly available C-NMC dataset. It comprises 15,135 images derived from 118 patients. These images have been meticulously segmented from microscopic samples, aiming to

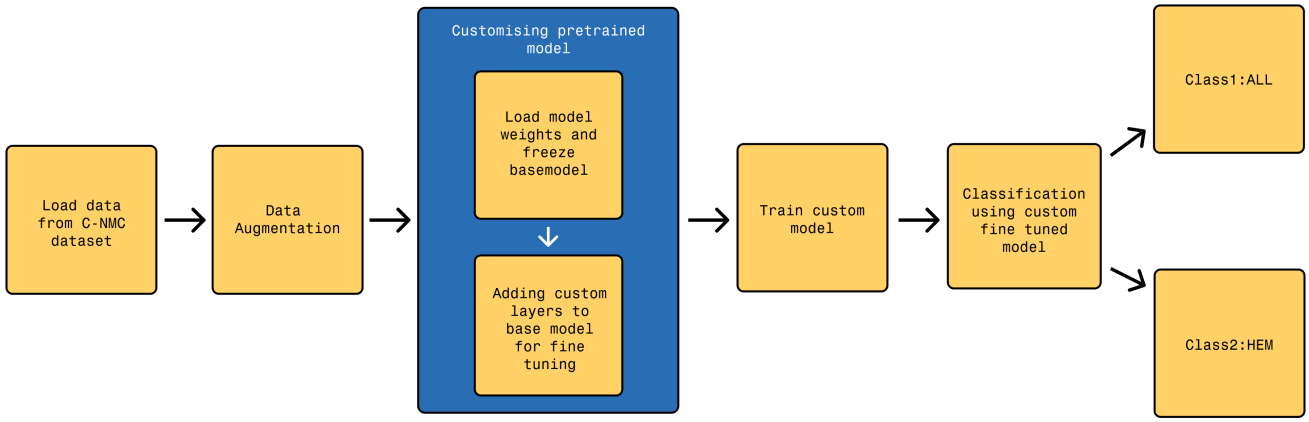

Figure 2. Proposed Methodology

closely replicate real-world conditions by retaining staining noise and illumination errors. The Train Set comprises 73 subjects (47 ALL, 26 Normal) with 10,661 cell images (7,272 ALL, 3,389 Normal). The Preliminary Test Set includes 28 subjects (13 ALL, 15 Normal) with 1,867 cell images (1,219 ALL, 648 Normal). The Final Test Set has 17 subjects (9 ALL, 8 Normal) and 2,586 cell images.[7]. Figure 3 shows some samples of the images in the dataset which is being used to train the model. Some sample images from the dataset with the appropriate labels are given in Figure 3.

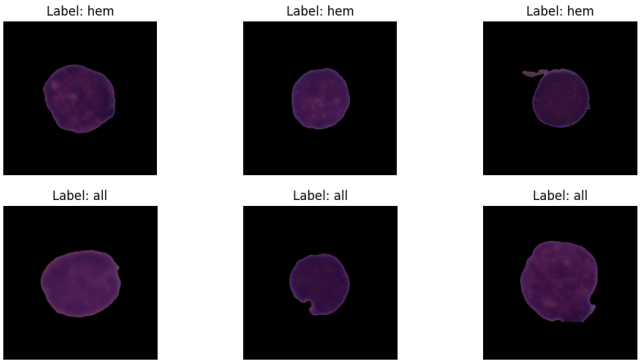

Figure 3. Samples from the C-NMC dataset

3.2. Data Augmentation

Data augmentation is carried out to increase the diversity of the dataset and provide better and more comprehensive results. The ImageDetector function from the TensorFlow library is used for data augmentation. For the training dataset, the images are flipped horizontally, and for the Testing images no changes as made. The batch size is set to 40. The batch size is set to 40 owing to the size of the dataset and, as it offered the best training stability. The results converged more quickly as compared to earlier trials using 32 and 28.

3.3. Customizing the pre-trained models

3.3.1. Loading the model weights and freezing the model. The main focus of this paper is leveraging state-of-the-art pre-trained models, which are established and trained for expansive datasets for general tasks, as the foundational framework for the classification of leukemia cells. This provides us with a robust starting point for this task.

As mentioned earlier, the main point of using a pre-trained model is the ability to use the knowledge acquired from the training. The process of freezing a model is done to preserve the already learned weights and the architecture used in the beginning.

3.3.2. Adding custom layers to the model. After freezing the model the main step is to add custom layers to facilitate more efficient binary classification. Three additional layers are added to the frozen model, A GloabalMaxPooling layer, a Dropout layer, and finally a Dense layer with the sigmoid activation function and two nodes, which act as the output layer. The pooling layer is used to flatten the spatial dimensions while retaining the most important features. The dropout layer is used to introduce regularisation of the data and 50% of the input units are dropped out. This helps in preventing overfitting. The output layer is where the final classification happens. The sigmoid function is used as it is specifically optimal for binary classification. The sigmoid function also called the logistic function is a sigmoidal function that maps real values to values between 0 and 1. Mathematically, the formula is defined as follows:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

The graph for the function is shown in Figure 4.

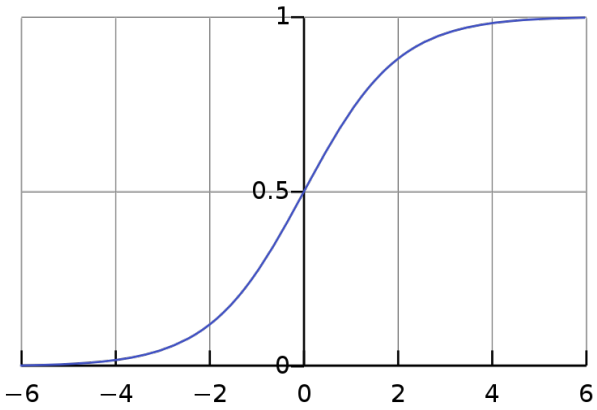

Figure 4. The Sigmoid function

As the real values get mapped to the range between 0 and 1, it is one of the best activation functions to be used in the output layer for a binary classification problem such as this. [11]

Transfer learning, the pivotal strategy employed in this paper involves adapting the pre-trained models to the specific classification task of leukemia cell classification. Instead of starting the training process from scratch, transfer learning allows us to capitalize on the knowledge learned from a different but related domain. As an extension of the transfer learning process, fine-tuning is used to make the pre-trained model into a more specialized model for the new task.

3.4. Training the custom model

The custom model which is the pre-trained model to which the new layers were added, is trained on the augmented images. This process is called fine-tuning as the model is tuned to perform the new classification task. The training is done over 15 epochs. This decision was made as most models reached convergence by the 7th or 10th epoch. The Early Stopping method was employed to come to this conclusion. The patience value was set to 5. Early stopping is a regularization method that is used to reduce overfitting and decrease the training time[8]. The optimization function used is Adamax with a learning rate of 0.001. Adamax was used for its adaptability to varying gradient magnitudes. Also studies related to classification problems have shown that the Adamax function is one of the best optimizer functions[9][10]. The learning rate was set to 0.001 through trial and error. The loss function used is binary cross-entropy, as it is the optimal and standard function to use for binary classification problems. The binary cross-entropy function can be defined as follows:

$$L(y, \hat{y}) = -\left(y \cdot \log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y})\right)$$

3.5. Classification using the custom fine-tuned model

The model classifies the data into classes namely, class 1, ALL; which are cells inflicted to leukemia, and class 2, HEM, which are normal healthy blood cells.

4. Results

The experimental results presented in Table 1 demonstrate the performance evaluation of various deep learning models. The models considered for evaluation include VGG16, VGG19, Inception, Xception, Efficient Net B0, ResNet101, and ResNet50. The evaluation metrics employed for assessing the models' performance are Validation Accuracy, Validation Precision, and Validation Recall.

Precision and recall, fundamental metrics in classification tasks, provide deeper insights into the models' abilities. Precision, denoted as

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives}$$

measures the accuracy of positive predictions. Recall denoted as

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives}$$

gauges the ability of the model to capture all positive instances.

A confusion matrix is a table that summarizes the performance of a classification model. It consists of four components:

- **True Positives (TP):** Instances that are actually positive and were correctly classified as positive.
- **False Positives (FP):** Instances that are actually negative but were incorrectly classified as positive.
- **True Negatives (TN):** Instances that are actually negative and were correctly classified as negative.
- **False Negatives (FN):** Instances that are actually positive but were incorrectly classified as negative.

The confusion matrices associated with each of the models are presented in Figure 5.

ResNet101 emerges as the top-performing model with a Validation Accuracy of 76.36%, Validation Precision of 75.85%, and Validation Recall of 76.36%. Similarly, ResNet50 closely follows, demonstrating a Validation Accuracy of 76.36%, Validation Precision of 75.85%, and Validation Recall of 76.36%. Notably, VGG16, VGG19, Inception, and Xception exhibit competitive performances, with Validation Accuracy values ranging from 70.46% to 73.86%.

5. Conclusion

The paper advances the field of leukemia detection by presenting a systematic and comparative analysis of machine learning models.In this pursuit, ResNet101 emerges as the

Model	Validation Accuracy	Validation Precision	Validation Recall
VGG16	0.70455	0.5035	0.9886
VGG19	0.7227	0.5497	0.9182
Inception	0.7341	0.7101	0.7682
Xception	0.7386	0.7310	0.7659
Efficient Net B0	0.5000	0.5000	1.0000
ResNet101	0.7636	0.7585	0.7636
ResNet50	0.7591	0.7585	0.7636

TABLE 1. VALIDATION METRIC

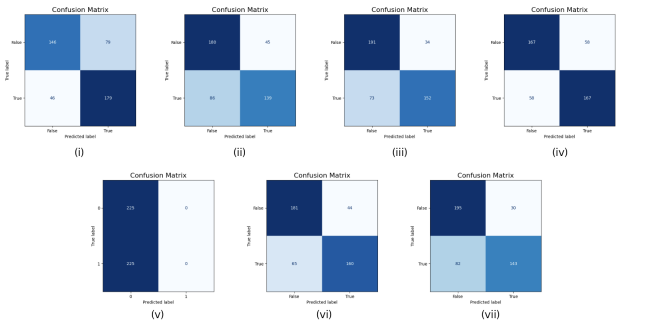

Figure 5. Confusion matrix (i)Vgg16,(ii)vgg19,(iii)Inception,(iv)Xception,(v)Efficient Net BO, (vi)ResNet101,(vii)ResNet50

standout performer, showcasing its exceptional capabilities in accurately distinguishing between healthy and Acute Lymphoblastic Leukemia (ALL) cells. By harnessing the potential of machine learning, the research establishes a robust system capable of reducing reliance on subjective human interpretation and expediting the diagnostic process.

References

[1] *Cancer Center. Leukemia Types.* https://www.cancercenter.com/cancer-types/leukemia/types

[2] Classification of immature white blood cells in acute lymphoblastic leukemia L1 using neural networks particle swarm optimization - Neural Computing and Applications https://link.springer.com/article/10.1007/s00521-021-06245-7

[3] A2M-LEUK: attention-augmented algorithm for blood cancer detection in children - Neural Computing and Applications https://rdcu.be/dxqK3

[4] Efficient Classification of White Blood Cell Leukemia with Improved Swarm Optimization of Deep Features https://rdcu.be/dxqLG

[5] *Automated Detection of Acute Lymphocytic Leukemia Using Blast Cell Morphological Features.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3372907 A. Genovese, M. S. Hosseini, V. Piuri, K. N. Plataniotis and F. Scotti, "Acute Lymphoblastic Leukemia Detection Based on Adaptive Unsharpening and Deep Learning," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 1205-1209, doi: 10.1109/ICASSP39728.2021.9414362. keywords: Deep learning;Training;Databases;Microprocessors;Signal processing algorithms;Computer architecture;Signal processing;Deep Learning;CNN;ALL;XAI,

[6] C-NMC dataset https://www.kaggle.com/datasets/avk256/cnmc-leukemia

[7] Manuel Vilares Ferro, Yerai Doval Mosquera, Francisco J. Ribadas Pena, Víctor M. Darriba Bilbao, Early stopping by correlating online indicators in neural networks, Neural Networks, Volume 159, 2023, Pages 109-124, ISSN 0893-6080,https://doi.org/10.1016/j.neunet.2022.11.035 (https://www.sciencedirect.com/science/article/pii/S0893608022004920)

[8] Comparative Study of First Order Optimizers for Image Classification Using Convolutional Neural Networks on Histopathology Images https://www.researchgate.net/publication/344163092_Comparative_Study_of_First_Order_Optimizers_for_Image_Classification_Using_Convolutional_Neural_Networks_on_Histopathology_Images

[9] The effect of choosing optimizer algorithms to improve computer vision tasks: a comparative study https://rdcu.be/dxqNb Szandała, Tomasz. (2021). Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks. 10.1007/978-981-15-5495-7_11.

[10] H. T. Salah, I. N. Muhsen, M. E. Salama, T. Owaidah and S. K. Hashmi, "Machine learning applications in the diagnosis of leukemia: Current trends and future directions", *Int. J. Lab. Hematol*, vol. 41, no. 6, Dec. 2019.

[11] J. Zhao, M. Zhang, Z. Zhou, J. Chu and F. Cao, "Automatic detection and classification of leukocytes using convolutional neural networks", *Med. Biol. Eng. Comput*, vol. 55, no. 8, Aug. 2017.

[12] M. I. Razzak and S. Naz, "Microscopic blood smear segmentation and classification using deep contour aware CNN and Extreme Machine Learning", *Proc. of CVPRW*, 2017.

[13]