International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 10 | Oct 2023 www.irjet.net p-ISSN: 2395-0072

Deep Learning for Leukemia Detection: A MobileNetV2-Based Approach for Accurate and Efficient Diagnosis

Lagisetty Naga Pavithra[1]

[1]*Student, Dept of Computer Science, Bangalore Institute of Technology, Bengaluru, Karnataka, India*

Abstract – *Acute lymphoblastic leukemia (ALL) is a type of cancer of blood and bone marrow, which is not only fatal but also very expensive to treat. Leukemia detection at early stage would can save lives and money. It is very common in children. Most leukemia in children is treated. Research studies reported that leukemia brings changes in white blood cells count. Currently for initial ALL diagnosis evaluation is done manually. This is time consuming and prone to errors. The proposed model is based on data collected from Kaggle dataset. The MobileNetV2 model is a lightweight model through which I have achieved an accuracy of 98.88% on training data and accuracy of 98.58% on testing data, with precision of 0.986, recall of 0.9858 and F1 score of 0.9857. Experiments were conducted on dataset containing 3256 images from 89 patients suspected of ALL, including 25 healthy individuals. Currently the dataset contains three stages which are Early Pre-B, Pre-B and Pro-B ALL.*

Key Words: *Acute lymphoblastic leukemia (ALL), MobileNetV2, blood cancer, bone marrow, deep learning, leukemia*

1.INTRODUCTION

Leukemia is cancer of blood or bone marrow which produces blood cells. It usually involves white blood cells. White blood cells are infection fighters, where they divide in orderly way to fight as your body needs them. But when it comes to people with leukemia there is abnormal amount of white blood cells. Treatment is very complicated and varies based on the type of leukemia tat the person is facing and involves various other factors. There are many types of leukemia but few of them are very common in children which is ALL. It occurs in children of age 2 to 4. Acute myelogenous leukemia (AML) is second most common in children. ALL can affect different types of lymphocytes called B-cells or T-cells. Blood stem cells originate in the bone marrow, mainly in flat bones in adults (hip, sternum, skull, ribs, vertebrae, scapulae, to name a few.), and can follow two developmental lines. Cells of the myeloid lineage give rise to white blood cells, especially neutrophil monocytes, platelets, and red blood cells; Cells of the lymphoid lining produce white blood cells, also called lymphocytes [1].

Leukemia occurs when there is damage in DNA of developing blood cells, mainly white blood cells which causes the blood cells to divide and grow uncontrollable. Researchers say that leukemia might be genetic and run in the family.

Few of the symptoms of leukemia include fatigue, weakness, pale skin, fever, and chills. It also causes headaches, nausea, vomiting, confusion, seizures. Leukemia can cause petechiae, a rash like collection of pinpoint red spots on the skin.

Currently leukemia is not curable. In few cases it is treatable with chemotherapy, radiation therapy, stem cell transplantation, CART-cell therapy, target therapy and other methods. A risk factor is anything that may increase your chance of having a disease. Some of the risk factors include smoking, exposure to certain chemicals, radiation exposure and blood disorders.

There are several methods that can be used to detect leukemia like CNN, EfficientNet, ResNet, DenseNet and many more. The proposed model is based on MobileNetV2. MobileNetV2 excels due to its efficiency, lightweight design, and fast inference, making it suitable for resource-constrained environments and real-time applications. It offers strong generalization through pre-trained weights, robust feature extraction, scalability, and competitive performance, backed by open-source accessibility and community support.

The rest of the paper is organized as follows: Section II Literature review, Section III presents the proposed architecture, and IV and V explain the experimental settings, results, and conclusion. Finally, Section VI discusses the future work.

2. LITERATURE REVIEW

N. Jiwani et al. [1] introduced a pioneering approach using Pattern Recognition and Computational Deep Learning to enhance acute lymphoblastic leukemia (ALL) diagnosis and management. The ALLDM model achieved impressive accuracy rates, such as 87.92% in chemotherapy management and 94.31% in stem cell transplantation management. This technology holds promise for improving ALL treatment outcomes, especially in children.

A. Batool et al. [2] proposed a comprehensive solution to address the diagnostic complexities of acute lymphoblastic leukemia (ALL). Their work introduces a state-of-the-art DL model based on EfficientNet-B3, achieving remarkable accuracy in leukemia cell classification. This model outperforms existing DL classifiers, offering a robust and reliable tool to enhance clinical leukemia detection and improve patient outcomes.

M. A. Hossain et al. [3] proposed a cost-effective solution to detect early-stage leukemia based on symptoms. Their explainable supervised machine learning model, using decision trees and the Apriori algorithm, outperformed other algorithms with a 97.45% accuracy rate. Sharing the dataset and code enhances resources for future leukemia research.

N. Akram et al. [4] introduced a pioneering solution for leukemia diagnosis, focusing on WBC segmentation. Their multi-scale information fusion network (MIF-Net), with its internal and external spatial information fusion mechanisms, excelled in the accurate segmentation of challenging WBC images. Across four datasets, MIF-Net achieved state-of-the-art segmentation performance, boasting remarkable accuracy, and it maintains computational efficiency with just 2.67 million trainable parameters.

Atteia, G et al. [5] introduced a Bayesian-optimized convolutional neural network (CNN) for acute lymphoblastic leukemia (ALL) detection in blood smear images. The model's hyperparameters were tailored using Bayesian optimization, resulting in enhanced classification performance. This innovative approach yielded superior accuracy, outperforming other optimized deep learning models, promising improved ALL detection.

Chen et al. [6] introduced a Resnet101-9 ensemble model for acute lymphoblastic leukemia (ALL) detection in microscopic images, combining nine trained Resnet-101 models with majority voting. Algorithm hyperparameters were optimized through the Taguchi method. The model achieved an accuracy of 85.11% and an F1-score of 88.94, surpassing individual models and excelling in precision, recall, and specificity.

Houssein EH et al. [7] introduced an end-to-end computer-aided diagnosis (CAD) system for leukocyte classification using deep learning. They combined DenseNet-161 with cyclical learning rate and the one-cycle technique to optimize hyperparameters. The model achieved remarkable accuracy, with 100% on the training set and 99.8% on the testing set, promising significant improvements in white blood cell classification.

Kruse A et al. [8] proposed an advanced model for Minimal Residual Disease (MRD) detection, crucial for predicting leukemia relapse. They leveraged next-generation sequencing (NGS) to enhance MRD diagnostics' sensitivity. The model employed phenotypic markers and differential gene patterns analysed through various techniques like flow cytometry (FCM), PCR, RQ-PCR, RT-PCR, or NGS.

Bibi N et al. [9] presented an IoMT-based framework for quick and safe leukemia identification, aiming to address the shortcomings of existing methods. Leveraging cloud computing, this system facilitates real-time coordination for diagnosis and treatment. Using DenseNet-121 and ResNet-34, the study outperformed other algorithms in identifying leukemia subtypes.

Loey et al. [10] proposed two automated leukemia classification models using transfer learning for early detection. The first model preprocesses images and employs a pre-trained deep convolutional neural network, AlexNet, for feature extraction and classification. In the second model, AlexNet is fine-tuned for improved performance. Experiments on 2820 images demonstrated that the second model achieved a remarkable 100% classification accuracy, surpassing the first model.

3. PROPOSED SYSTEM

Model is designed to automate leukemia classification using microscopic images. This system leverages deep learning and transfer learning techniques to enhance accuracy and efficiency.

3.1 DATASET COLLECTION

The implementation of the proposed model begins by loading the dataset. The dataset used for this model is available publicly on Kaggle. The dataset contains 3256 images classified as healthy, Early Pre-B, Pre-B and Pro-B. There are 504 healthy, 985 Early Pre-B, 963 Pre-B and 804 Pro-B. It is split into training validation and testing in the ratio 80:10:10.

3.2 DATA PREPROCESSING

The preprocessing techniques used in the code involve resizing and rescaling images to a target size of 256x256 pixels and normalizing their pixel values to a range between 0 and 1. Additionally, data augmentation methods such as random flips and rotations are applied to increase the diversity of the training dataset. Caching and prefetching are utilized to improve data loading efficiency during training, ensuring optimal performance of the deep learning model.

3.3 TRANSFER LEARNING

Model utilizes MobileNetV2, a pre-trained deep convolutional neural network, as the feature extractor. This model, trained on a large and diverse dataset, offers a valuable starting point for leukemia classification. By extracting high-level features from images, MobileNetV2 provides valuable insights into image content.

3.4 MODEL ARCHITECTURE

The system builds on MobileNetV2 with additional layers for classification. These layers include a Global Average Pooling 2D layer to reduce dimensionality, a Dropout layer to prevent overfitting, and fully connected Dense layers. The final Dense layer has a softmax activation function to output class probabilities.

3.5 TRAINING AND EVALUATION

Model is trained using the preprocessed training dataset with a batch size of 32 for 30 epochs. The training process is monitored for accuracy and loss. The model's performance is evaluated using a separate test dataset.

3.6 METRICS AND VISUALIZATION:

The system provides insights into the model's performance by calculating metrics such as accuracy and loss. These metrics offer a quantitative measure of the model's classification capabilities. Additionally, the system generates visualizations, including accuracy and loss curves over the training epochs.

4. RESULTS AND ANALYSIS

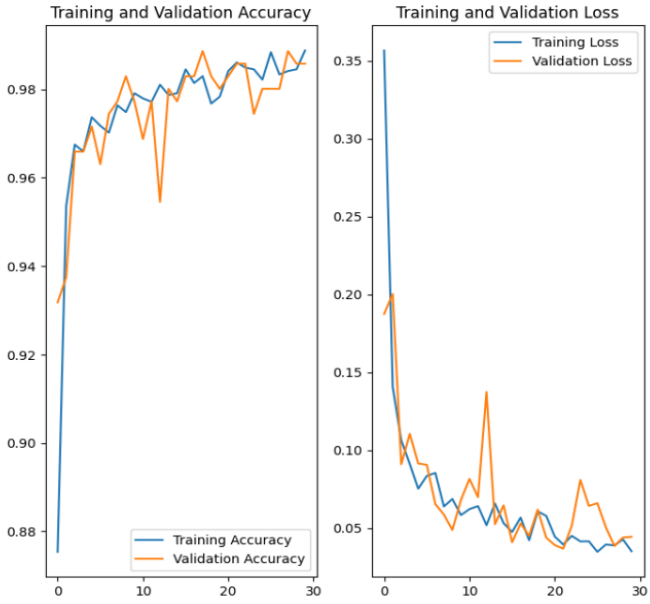

Chart -1: Training and Validation accuracy and loss

It was observed that the proposed neural network model achieved an accuracy of 98.87% on training data and 98.58% on testing data. The loss reduced from 0.1874 to 0.0443. The highest value of accuracy was achieved at 18th epoch.

Table -1: Output for proposed model w.r.t intermediate epochs

Epoch	Loss	Accuracy%
1	0.1874	93.18
15	0.0645	97.73
30	0.0443	98.58

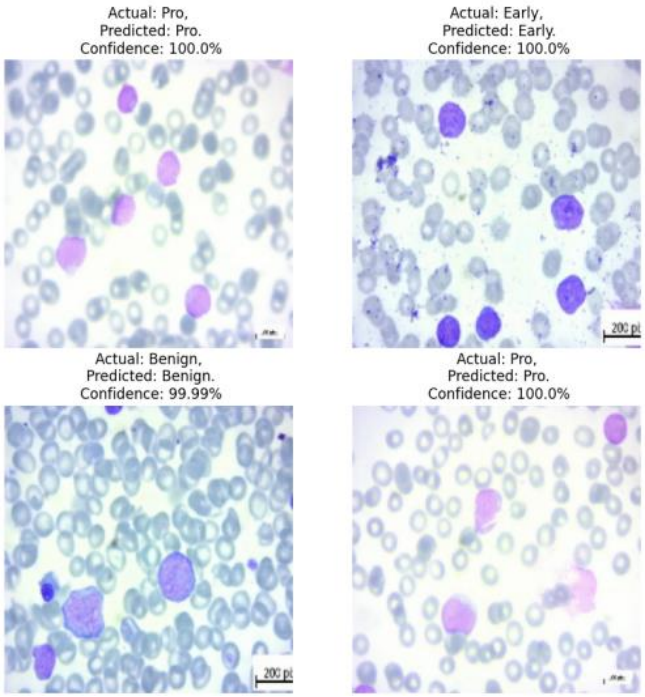

Fig -1: Actual vs Predicted

Figure 1 shows predictions of few of the randomly selected images from dataset.

5. CONCLUSION

In this study, a deep learning model based on MobileNetV2 demonstrated impressive accuracy (98.58%) in classifying acute lymphoblastic leukemia (ALL) from microscopic images. It boasted high precision and recall (both above 98%) and a strong F1 score (0.986). The model was trained on a diverse dataset of 3256 images, encompassing different ALL stages and healthy samples. The utilization of transfer learning with MobileNetV2 enhanced its classification capabilities. This research offers significant potential for early ALL detection, providing a valuable tool for medical professionals and the possibility of improving patient outcomes, while further refinements could advance its clinical utility.

6. FUTURE WORK

In the future, the dataset can be extended by adding new samples and utilizing new augmentation techniques. A variety of deep learning models can be applied to improve accuracy. Various feature extraction techniques could be used.

REFERENCES

[1] N. Jiwani, K. Gupta, G. Pau and M. Alibakhshikenari, "Pattern Recognition of Acute Lymphoblastic Leukemia (ALL) Using Computational Deep Learning," in IEEE

Access, vol. 11, pp. 29541-29553, 2023, doi: 10.1109/ACCESS.2023.3260065.

[2] A. Batool and Y. -C. Byun, "Lightweight EfficientNetB3 Model Based on Depthwise Separable Convolutions for Enhancing Classification of Leukemia White Blood Cell Images," in IEEE Access, vol. 11, pp. 37203-37215, 2023, doi: 10.1109/ACCESS.2023.3266511.

[3] M. A. Hossain, A. K. M. M. Islam, S. Islam, S. Shatabda and A. Ahmed, "Symptom Based Explainable Artificial Intelligence Model for Leukemia Detection," in IEEE Access, vol. 10, pp. 57283-57298, 2022, doi: 10.1109/ACCESS.2022.3176274.

[4] N. Akram et al., "Exploiting the Multiscale Information Fusion Capabilities for Aiding the Leukemia Diagnosis Through White Blood Cells Segmentation," in IEEE Access, vol. 10, pp. 48747-48760, 2022, doi: 10.1109/ACCESS.2022.3171916.

[5] Atteia, G.; Alhussan, A.A.; Samee, N.A. BO-ALLCNN: Bayesian-Based Optimized CNN for Acute Lymphoblastic Leukemia Detection in Microscopic Blood Smear Images. *Sensors* **2022**, *22*, 5520. https://doi.org/10.3390/s22155520.

[6] Chen, YM., Chou, FI., Ho, WH. *et al.* Classifying microscopic images as acute lymphoblastic leukemia by Resnet ensemble model and Taguchi method. *BMC Bioinformatics* **22** (Suppl 5), 615 (2021). https://doi.org/10.1186/s12859-022-04558-5

[7] Houssein EH, Mohamed O, Abdel Samee N, Mahmoud NF, Talaat R, Al-Hejri AM, Al-Tam RM. Using deep DenseNet with cyclical learning rate to classify leukocytes for leukemia identification. Front Oncol. 2023 Sep 12;13:1230434. doi: 10.3389/fonc.2023.1230434. PMID: 37771437; PMCID: PMC10523295.

[8] Kruse A, Abdel-Azim N, Kim HN, Ruan Y, Phan V, Ogana H, Wang W, Lee R, Gang EJ, Khazal S, Kim YM. Minimal Residual Disease Detection in Acute Lymphoblastic Leukemia. Int J Mol Sci. 2020 Feb 5;21(3):1054. doi: 10.3390/ijms21031054. PMID: 32033444; PMCID: PMC7037356.

[9] Bibi N, Sikandar M, Ud Din I, Almogren A, Ali S. IoMT-Based Automated Detection and Classification of Leukemia Using Deep Learning. J Healthc Eng. 2020 Dec 3;2020:6648574. doi: 10.1155/2020/6648574. PMID: 33343851; PMCID: PMC7732373.

[10] Loey, M.; Naman, M.; Zayed, H. Deep Transfer Learning in Diagnosing Leukemia in Blood Cells. *Computers* 2020, *9*, 29. https://doi.org/10.3390/computers9020029

BIOGRAPHIES

Lagisetty Naga Pavithra is a student at Bangalore Institute of Technology, Bengaluru, Karnataka, India pursuing Computer Science and Engineering.