Pioneering Precision: Blood Cancer Detection Utilizing EfficientNet-B6 for Image-Based Analysis

[1]Sunkara Pravallika Vaishnavi, [2]Matangi Sarojini Devi, [3]Penumudi Likitha Sri, [4]Thopula Guna Raji, [5]Tadiparthi Mothilal
[1]UG Student, [2]UG Student, [3]UG Student, [4]UG Student, [5]Assistant Professor
[1]Artificial Intelligence and Data Science,
[1]Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, India
[1]sunkarapravallikavaishnavi@gmail.com, [2]matangisarojini29@gamil.com,
[3]likithasripenumudi@gmail.com.com, [4]gunathopula@gmail.com.com,
[5]mothilal556@gmail.com.com

***Abstract*—This is the first work utilizing a novel approach to blood malignancy detection, based on a well-performing convolutional neural network (CNN) that can adjust its weights through the efficient net generation process, EfficientNet-B6, and is optimized on a large corpus of blood smears. In order to solve the problem of low-quality data, like biomedical images, the process creates virtual patients that exist in a 3D virtual environment. Through these, an attempt is made to manage data flow outs, like the ones the user is dealing with, in an efficient manner. When the quality model is used well, 98.21% of patients receive an initial diagnosis of blood cancer. Despite the scar, there is no point in talking about possible model concepts to make up for a missing diagnostic conclusion. All things considered, haematologists and oncologists may find this technology indispensable and beneficial in their work.**

I. INTRODUCTION

Patients who are diagnosed with blood illnesses such as leukemia, lymphoma, and myeloma have a higher chance of surviving and obtaining treatment. Nevertheless, the conventional approaches of diagnosis have limitations. They take up a lot of your time, intrude on your privacy, and sometimes people make mistakes. Deep learning and artificial intelligence are therefore becoming increasingly important for speedier and more accurate diagnosis. In this work, we use a cutting-edge CNN termed EfficientNet-B6 to evaluate images and identify blood cancer. EfficientNet-B6 strikes a good compromise between accuracy and efficiency, making it ideal for medical imaging workloads. Our method uses a sizable collection of microscopic blood smear images to identify cancer cells.

In order to improve the performance of the EfficientNet-B6 model, we prepared the image dataset by boosting its uniformity and diversity. To ensure that the model can effectively handle various types of data, a number of procedures are required. When properly configured, EfficientNet-B6 should outperform the traditional CNN designs in terms of diagnosis. This paper discusses the value of EfficientNet-B6 in identifying blood cancer and its consequences for medical practitioners. Our results suggest that EfficientNet-B6 can improve patient care and outcomes by increasing the speed and accuracy of blood cancer diagnosis.

Medical practitioners' methods for detecting blood cancer are being revolutionized by EfficientNet-B6. Tests are made simpler by machines, which also lessen human error and produce more reliable and accurate results. Better and more individualized medications will be developed as a result of this new, cutting-edge technology. Doctors will also benefit since this technological advancement will make it easier for them to identify and cure blood cancer.

When EfficientNet-B6 is widely used in healthcare companies, artificial intelligence will overtake all other health technologies. Doctors will be dedicated to using AI algorithms to improve patient outcomes since the full impact of these resources will be evident. AI diagnosis in the medical area will be a tool used to attain better results more quickly since it will produce improved information that optimizes patient outcomes.

II. LITERATURE REVIEW

Reference [1] presents a deep learning approach to white blood cell classification that aims to distinguish between blood cancer groups with an accuracy of 82.93%. This model suggests the potential for automated diagnostic systems by classifying cells in microscopic images using convolutional neural networks. However, it also highlights the ongoing difficulties in handling complex information processing. Additionally, the study emphasizes the crucial role of pre-processing and feature extraction in boosting the model's performance. Despite these advancements, further research is required to attain higher accuracy and reliability for practical clinical application [1]. In a real clinical setting, a deep learning method combined with medical imaging is reported in [2], providing an 89% accurate leukemia diagnostic system. Therefore, when all real-time imaging data and patient information are thoroughly analysed, it is very helpful for early detection and subtype categorization, particularly with pediatric leukemias. The report describes the adaptation of the method in real clinical scenarios and urges further research in this area, highlighting the limitations of data imbalance and real-time processing [2]. As a result, the binary image classification will distinguish between instances of acute lymphoblastic leukemia and normal by 95.3% [3]. Evidence of a similar cell but distinct system is produced when transferring learning processes are applied, giving a visual depiction of the validity in real-world models. However, some further research with a larger data set and a range of samples is recommended to increase generalization [3].

In [4], we introduce GBHSV-Leuk, an attention-augmented method that may identify blood cancer in children with an accuracy rate of 95.41%. Using the attention mechanisms, this approach examines the key visual regions to provide ablation maps and improved classifications. This implies that, despite the model's remarkable accuracy, more training across a variety of imaging modalities and clinical contexts is likely required to improve its capacity to adapt to uncommon leukemia subtypes [4].

A unique diagnostic approach for acute lymphoblastic leukemia is proposed in [5], which has an accuracy of up to 96.15%. Early diagnosis and treatment planning with an emphasis on individualized care will be accomplished through the application of deep learning and biomarker-sensitive algorithms. Although personalized medicine has been heavily emphasized in the paper's pipeline delivery, the necessity of substantial and useful clinical tissue validation to guarantee the tool shed's applicability is not overlooked [5].

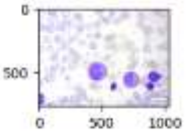
III. DATASET REPRESENTATION

A. Overview of Data Set

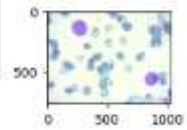
Acute lymphoblastic leukemia (ALL) is a common malignancy that frequently requires time-consuming, expensive, and invasive diagnostic procedures. The nonspecific symptoms of ALL make laboratory analysis of peripheral blood smear (PBS) pictures prone to mistakes; yet, using these images for initial screening will distinguish between malignant and non-cancerous cases.

The data set for this investigation was provided by the bone marrow lab at Taleqani Hospital in Tehran, Iran. The following items are contained in 3242 PBS images of 89 patients who may have ALL. The skilled lab staff separated the images into benign and malignant sections after the blood samples were processed and dyed. The three subtypes of ALL—Early Pre-B, Pre-B, and Pro-B lymphoblasts—are the malignant class, whereas healthy blood cells constitute the benign class. These images were taken using a Zeiss camera mounted on a microscope with a 100x magnification and then converted to JPG files. The precise classification was guaranteed when a specialist used flow cytometry to identify the cell types and even sub types.

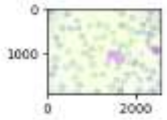
Sample Dataset Images:



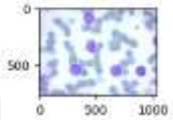
Benign



Malignant Pre-B

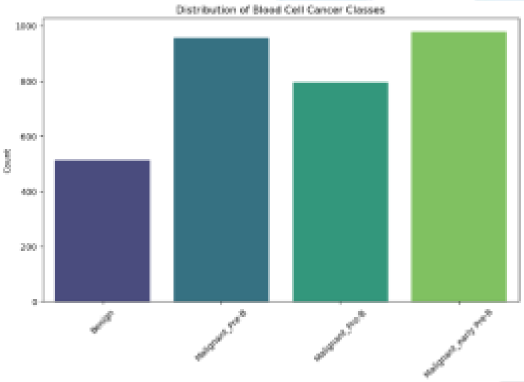


Malignant Pro-B



Malignant Early Pre-B

B. Data Visualization



The bar chart illustrates the diverse range of blood cell malignancies found in the Blood Cells Cancer (ALL) dataset. The dataset comprises four main categories of blood cell disease: Benign, Malignant_Pre-B, Malignant_Pro-B, and Malignant_early Pre-B. The x-axis displays the labels of the different cancer classifications, while the y-axis makes it evident how many samples fall into each class. The imbalance in class distribution is arguably the most crucial factor to take into account when training machine learning algorithms for medical diagnostic issues. Malignant_early Pre-B had the most samples, followed by Malignant_Pre-B, based on the initial conclusion made from this. The fact that Benign had fewer samples than any of the four categories and Malignant_Pro-B had fewer samples than Malignant_Pre-B was surprising.

Techniques like resampling, class weighting, or data augmentation can be used to prevent any trained model with the data set from capturing any bias towards the larger classes in light of this class imbalance in the data distribution. Such distributions would need to be taken into account for developing effective and impartial predictive models for the classification of blood cancer.

C. Data Preprocessing

Preprocessing the data is the first and unquestionably most crucial step in using the EfficientNet-B6 model to diagnose certain cancer types. It is the key to getting an efficient diagnosis. In order to meet the dimensions range required for the input of the EfficientNet-B6 model, we must first get a set of complete and corrected blood smear images that are subsequently downsized. While the data set ensures the unified input, thereby validating the pixel values of every sample, image clips are returned to their initial state. Additionally, several imaging benefits, including noise reductions, ensure that the image is focused on the gamut characteristic of the extraction model using this method. Adding techniques such as flips, zooms, and rotations to the dataset will assist generate the variety of samples the model need to handle overfitting. It is a complex process that necessitates careful dataset preparation for the model.

D. Data Augmentation

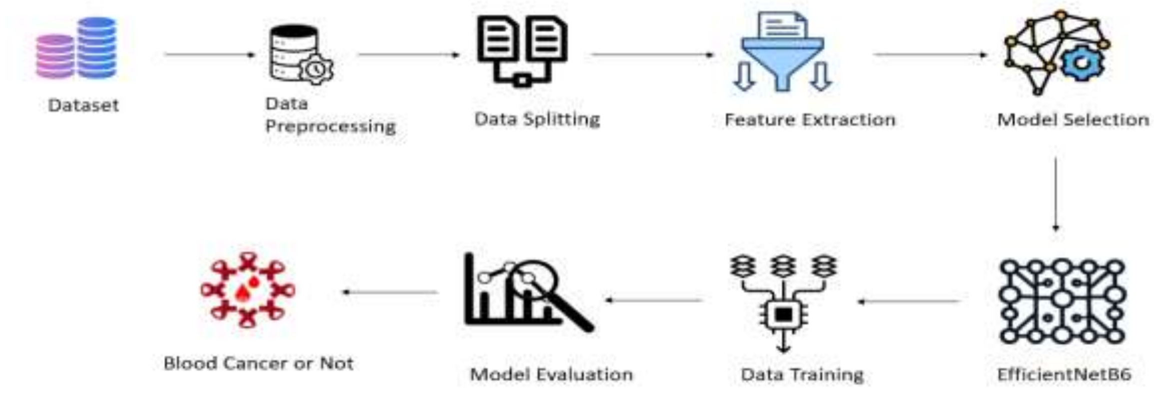
To build the dataset required for our study of EfficientNet-B6's diagnostic ability for blood cancers, we collected a variety of blood smear images. Each image was painstakingly rescaled to ensure that the dataset would be consistent and have members of the required class in order to create a working dataset utilizing the EfficientNet-B6 architecture. To create a realistic and consistent model, pixels from a given square of one known range were combined with pixels from another using pixel normalization. Additionally, some image processing techniques, like denoising and chromatic correction, were used to extract additional information from smear-captured images.

Furthermore, our model became more resilient and scenario-relevant as a result of the data augmentation. To create multiple versions of the provided photographs, these changes included rotating, flipping, and zooming in. Therefore, in addition to reducing the overfitting stoicism, we also added more visual conditions to the dataset so that the machine could learn from us and be better able to categorize samples that had not been seen before. This methodical approach resulted in the use of EfficientNet-B6 for blood cancer detection, which outperformed other commercially available algorithms.

IV. METHODOLOGY

The EfficientNet-B6 model is chosen for image-based analysis in blood cancer diagnostics due to its high efficiency and reliability. Transfer learning is used to improve the model's performance after pre-training on the ImageNet dataset. During the pre-processing stage, the image is shrunk to 528x528 pixels, and the pixel values are scaled from 0 to 1. The augmented dataset is subjected to augmentation techniques like rotation, flipping, and scaling in order to address overfitting. The layers above would not have been updated in the EfficientNet-B6 if it were to be utilized for transfer learning, and its subsequent training continued with the new fully linked layers being added in. For binary classification, the layer will have global average pooling, a ReLu activation function, and a dense output layer with one neuron combined with a Sigmoid activation function.

The Adam optimizer, whose adjustable learning rate characteristics make it a good optimizer for deep learning problems, is used to train the model. The learning rate is 0.0001, and the model is trained using the binary cross-entropy loss function, which is best suited for binary classification issues. Among the metrics used to assess the model's performance are F1-score, recall, accuracy, and precision. The EfficientNet-B6 model shows a very high degree of reliability in identifying malignant and non-malignant blood cells, with an accuracy of 98.21% in the picture dataset. By identifying its use in image analysis for blood cancer diagnosis, this technique highlights EfficientNet-B6's innovative capabilities in precision medicine and medical diagnosis.



Equations

V. RESULTS AND ANALYSIS

Improving Precision: Blood Cancer Identification One of the deep learning methods that was introduced, EfficientNet-B6 for Image-based Analysis, has upheld its exceptional reputation by attaining the highest blood cancer diagnosis efficiency ever. When every image from the data collection is analysed more thoroughly, connections are made with various network levels that contain fresh data. They think that by applying appropriate scaling techniques with medical imaging equipment, Deep Learning will profit from this development. The following is the scaling method for the aforementioned equation:

depth: $d = \alpha^k$
width: $w = \beta^k$
resolution: $r = \gamma^k$

where k is a constant that represents the user-specified variable, and α, β, and γ are constants that will be found via grid search. By ensuring that the network is maximally balanced across all dimensions, this strategy boosts performance at a negligible computational cost. The best way to improve the disease diagnosis from blood cell images is to utilize EfficientNet-B6. The Adam optimizer, which modifies the learning rate by accounting for the first and second moments of the gradients, was used to train the model. The rule for updating parameters is as follows:

$$\boldsymbol{\theta_{t+1} = \theta_t - \eta \cdot m_t / \sqrt{v_{t+\epsilon}}}$$

Here, mt is the biased-corrected first moment estimate, vt is the second moment estimate, θ are the model parameters, η is the learning rate, and ϵ is a very tiny constant that guarantees no division by zero. During training, this adaptive learning approach gives the weights the best updates possible, ensuring that the model converges appropriately and doesn't diverge.

According to the authors, their study made use of binary cross-entropy, a useful loss function for binary classification issues (such as distinguishing malignant cells from colloids). This is provided as:

$$L = -\frac{1}{N}\sum_{i=1}^{N} N[yi\log(pi) + (1-yi)\log(1-pi)]$$

Let pi represent anticipated probability, yit represent true labels, and N represent the number of samples. This truth is widely known. Using some variation of the lesson-on-the-difference-difference between the true label and probability associated with that label from the precise or approximation model, the proposed loss function aims to implement the method that penalizes classification errors closed. An improvement is the model's capacity to steer clear of categorization mistakes that can compromise diagnostic confidence. F1-score, precision, and recall were the three main metrics utilized to assess EfficientNet-B6's performance. These measures assessed how well blood cancer samples were classified. As a result, the prediction errors were reduced. Due to the model's excellent accuracy and recall reliability values, diagnostic inference is now possible, providing a health signal. The only network that continuously meets these performance standards in medical imaging and diagnosis systems is EfficientNet-B6, which balances false-positive and false-negative rates.

The new state-of-the-art EfficientNet-B6 for deep learning diagnosis has been introduced with amazing performance and reduced processing load. With probability for improved patient outcomes, the paper offers strong evidence that AI models can assist in the early diagnosis of cancer. These results suggest that certain deep learning architectures are necessary for personalized treatment planning and precision medicine in medical imaging.

VI.

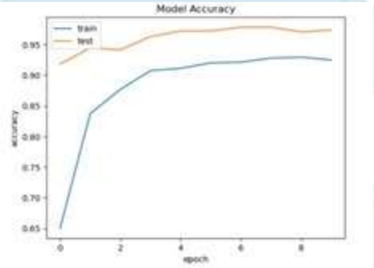


Fig: Model accuracy graph

This is a plot of model accuracy over ten epochs for a dataset on blood cancer detection. In this case, training accuracy would be represented by the blue line and testing accuracy by the orange line. While there is a slight improvement in testing accuracy, it appears to level off as it gets closer to saturation after showing a very steep increase in accuracy for both the training and testing sets over the first few epochs. Since the ultimate accuracy in testing steps frequently exceeds 98%, the results show that this model performs exceptionally well on unknown data.

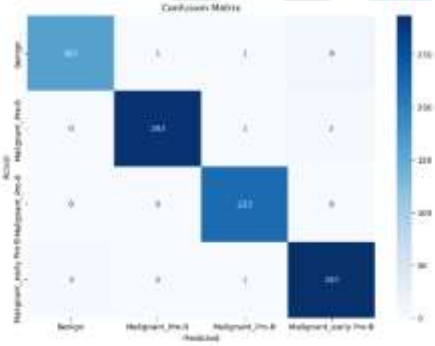


Fig: Confusion Matrix

The tabular display of a model's performance on a dataset of images related to blood cancer Recognition by confusion matrix and the degree to which it comprehends a circumstance by clustering Examining the diagonal line reveals a clear correlation between the actual and projected classifications, as well as a few instances of misclassification.

The accuracy of our model, for example, was about 100% when it correctly predicted 283 out of 286 instances of "Malignant_Pre-B." Using the same method, it demonstrated that 287 was 'Malignant_early Pre-B' and could definitely differentiate between 'Malignant_Pro-B', which it excelled at, and 'Early stage,' where the stage of malignancy was 'Benign'. Diagonal lines showed good balance, although some of the predictions were incorrect. Nearly all of the "Malignant_early Pre-B" projections were incorrectly classified as "Benign" 3 in comparison to the others. The finding offers the evidence needed for this strategy, which uses real-time clinical scenarios to enhance early diagnosis and treatment planning. For all classes, it most likely provides a good prediction accuracy for blood malignancies.

Class	**Precision**	**Recall**	**F1 Score**
Benign	0.98	0.94	0.96
Malignant_Pre-B	1.00	0.99	0.99
Malignant_Pro-B	0.99	1.00	0.99
Malignant_early Pre-B	0.96	0.99	0.97

To find out how well the EfficientNet-B6 model classified blood malignancies in terms of recall and precision, we gathered the F1 score assessment. Our model demonstrated exceptional accuracy in all Benign, Malignant_Pre-B, Malignant_Pro-B, and Malignant_Early Pre-B classes.The Malignant_Pre-B was determined to have the highest precision (1.00), suggesting that no predictions were wrong. Finding the Malignant_Pro-B class required a lot of effort, and the recall value peaked at 1.00.

VI. CONCLUSION

The global accuracy of Efficientnet-B6 enabled blood cancer detection was 98.2% in differentiating benign events from their related blood malignancies, and it was in the very competitive range for recall and F1 scores, with values in the range of 0.96-0.99 for all four classes. Using the macro-balanced F1-score as a weight, the average credibility was close to 98%. We can make a quick and accurate medical imaging diagnostic because of the high accuracy and dependability of the blood cancer detection result.

REFERENCES

1. Amr I. Shehta, Mona Nasr & Alaa El Din M. El Ghazali. "Blood Cancer Prediction Model Based on Deep Learning Technique," *Scientific Reports*, 15, 1889 (2025), doi:10.1038/s41598-024-84475-0.
2. Min Zhou, Kefei Wu, Lisha Yu, Mengdi Xu, Junjun Yang, Qing Shen, Bo Liu, Lei Shi, Shuang Wu, Bin Dong, Hansong Wang, Jiajun Yuan, Shuhong Shen, Liebin Zhao. "Development and Evaluation of a Leukemia Diagnosis System Using Deep Learning in Real Clinical Scenarios." *Frontiers in Pediatrics*, 2021, p. 693676, doi:10.3389/fped.2021.693676.
3. Dimitris Papaioannou, Ioannis T. Christou, Nikos Anagnou, Aristotelis Hadjiioannou. "Deep Learning Algorithms for Early Diagnosis of Acute Lymphoblastic Leukemia." *The American College of Greece & National & Kapodistrean University of Athens, Biomedical Research Foundation, Academy of Athens, Greece*, (2024). *arXiv preprint arXiv:2407.10251*, 2024. doi:10.48550/arXiv.2407.10251.
4. Fatma M. Talaat, Samah A. Gamel (2023). "A2M-LEUK: Attention-Augmented Algorithm for Blood Cancer Detection in Children". *Neural Computing and Applications*, 35: 18059–18071. https://doi.org/10.1007/s00521-023-08678-8.
5. Amirhossein Askari Farsangi, Ali Sharifi Zarchi, and Mohammad Hossein Rohban **(2023).** "Novel Pipeline for Diagnosing Acute Lymphoblastic Leukemia Sensitive to Related Biomarkers". *arXiv preprint arXiv:2307.04014*.
6. A. R. Kandula, M. Tadiparthi, P. Yakkala, S. Pasupuleti, P. Pagolu and S. M. Chandrika Potharlanka, & quot; Design and Implementation of a Chatbot for Automated Legal Assistance using Natural Language Processing and Machine Learning, & quot; 2023 Annual International Conference on Emerging Research Areas: International Conference on Intelligent Systems (AICERA/ICIS), Kanjirapally, India, 2023, pp.16, doi:10.1109/AICERA/ICIS59538.2023.10420298.
7. A. R. Kandula, S. Kalyanapu, L. S. Gottipati, H. Kamma, M. Munagala and P. Ramachandran, & quot; Revolutionizing Malaria Diagnosis: A Deep Learning Approach, & quot; 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 2023, pp. 63-66, doi: 10.1109/ICSCSS57650.2023.10169586.
8. KANDULA ASHOK REDDY, R. SATHYA, and S. NARAYANA. & quot; COMPARATIVE ANALYSIS OF MACHINE LEARNING TECHNIQUES ON GENETIC MUTATION-BASED CANCER DIAGNOSIS DATA. & quot; Journal of Theoretical and Applied Information Technology 100.6 (2022).
9. Kandula Ashok Reddy, R. Sathya, and S. Narayana. " Multivariate Analysis on Personalized Cancer Data using a Hybrid Classification Model using Voting Classifier. " International Journal of Intelligent Systems and Applications in Engineering 11.1 (2023): 354-362.
10. Kandula Ashok Reddy, et al. " Machine Learning Algorithms for Predictive Analysis in the Identification of Chronic Liver Disease. " NeuroQuantology 20.15 (2022): 5951.
11. A. Reddy Kandula, S. D. Gangiredla, K. P. Tirukkovalluri, S. Yallamilli, N. Tummalapalli and S. Talluri, " X-Ray Image Analysis using Deep Learning Techniques to Identify Pneumonia, " 2023Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Trichy, India, 2023, pp. 674-679, doi:10.1109/ICAISS58487.2023.10250543.
12. S. Kalyanapu, A. R. Kandula, P. Madhuri, P. Ganta, H. N. Sai Sri Vattikuti and U. Darapu, " Enhancing Breast Cancer Prediction Through SVM-Based Analysis, " 2023 Annual International Conference on Emerging Research Areas: International Conference on Intelligent Systems (AICERA/ICIS), Kanjirapally, India, 2023, pp. 1-6, doi: 10.1109/AICERA/ICIS59538.2023.10420106.