Automated Detection Leukemia Subtype Classification from Microscopic Images:

A Convolutional Neural Network Approach

BY

Sheikh Bodrun Nesa Lubna
ID: 201-15-3590

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Fahad Faisal
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By

Amit Chakraborty
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

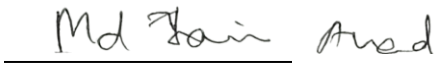
DHAKA, BANGLADESH

JANUARY, 2024

APPROVAL

This Project titled **"Automated Detection Leukemia Subtype Classification from Microscopic Images: A Convolutional Neural Network Approach"**, submitted by Sheikh Bodrun Nesa Lubna, ID: 201-15-3590 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 21/01/2024.

BOARD OF EXAMINERS



Dr. Md. Taimur Ahad (MTA) **Chairman**
Associate Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Md. Sazzadur Ahamed (SA) **Internal Examiner 1**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner 2

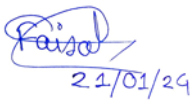
Most. Hasna Hena (HH)
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Abu Sayed Md. Mostafizur Rahaman(ASMR) **External Examiner**
Professor
Department of Computer Science and Engineering
Jahangirnagar University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Fahad Faisal, Assistant Professor, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree.

Supervised by:

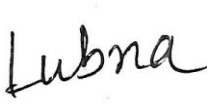


Fahad Faisal
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Amit Chakraborty
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:



Sheikh Bodrun Nesa Lubna
ID: -201-15-3590
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Fahad Faisal, Assistant Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "Machine Learning" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Sheak Rashed Haider Noori, Professor and Head,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Leukemia is a difficult type of blood cancer which can present in a combination of multiple types, that had individual cellular and genetic abnormalities. The prevalent forms of leukemia include Acute lymphocytic leukemia (ALL), Acute myeloid leukemia (AML), Chronic lymphocytic leukemia (CLL) and Chronic myeloid leukemia (CML). The evaluation of diseases, risk detection, and treatment planning all depend on the identification of specific genetic anomalies. This work describes a method for recovering Leukemia blood cancer cells applying blood microscopic images, identifying the subtypes of Leukemia. For detecting leukemia cancer using image processing techniques, that detect classification of leukemia blood cell subtypes. The proposed approach involves utilizing Convolutional Neural Network (CNN) to identify and categorize subtypes to leukemia blood cells based on microscopic images of human blood cells. Aimed to assess the expert Seven independent CNN models such as EfficientNetB7, ResNet-50, VGG19, ResNet101, DenseNet201, MobileNet and also built a CML-1(custom model) detect in the classification of leukemia. After exploit our proposed methodology, ResNet-50 model exhibited superior performance achieving 97.50% accuracy. The remaining models also displayed robust performance, with accuracy rates ranging from 95% to 97%. These finding imply that employing CNN methodologies for the automated identification of leukemia in microscopic blood images holds immense potential and presents substantial benefits in the realm of medical diagnostics.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Introduction

Leukemia is a type of blood cancer where the body produces too many abnormal white blood cells, crowding out healthy cells and affecting the body's ability to fight infections, stands as a formidable adversary in the realm of oncology, posing significant challenges to both medical professionals and patients alike. Defined by the malignant conversion of white blood cells, leukemia disturbs the intricate level of the hematologic system, leading to a diverse range of complex diseases. It’s a like a puzzle where the pieces don’t fit right, creating a mess in our blood and causing problems. The roots of leukemia lie in the intricate interplay of genetic and environmental factors. Although we’ re not sure about the exact causes of leukemia in many cases, some things make it more likely to happen. Being around radiation, certain chemicals, having genes that make it more similar or getting certain viral infections can increase the chances of developing leukemia. It’s a grouped into four main types based on how it grows and the kind of cells involved: Acute Lymphocytic Leukemia (ALL), Acute Myeloid Leukemia (AML), Chronic Lymphocytic Leukemia (CLL) and Chronic Myeloid Leukemia (CML).

Acute Lymphocytic Leukemia (ALL) mainly impacts the growth of lymphoid cells, which are important for the immune system. It interferes with the normal development of these white blood cells. ALL is more prevalent in children, making it the most common childhood leukemia. The rapid proliferation of immature lymphoid cells in the bone marrow and bloodstream leads to symptoms like bone pain, fatigue, and recurrent infections. Due to its aggressive nature, ALL necessitates immediate and intensive treatment, typically involving chemotherapy, radiation therapy, and sometimes stem cell transplantation. There are three categories of ALL known as L1, L2, and L3 [4].

Acute Myeloid Leukemia (AML) is a fast growing type of blood cancer that starts in the bone marrow, impacting myeloid cells responsible for creating red blood cells, platelets, and specific white blood cells. AML is marked by the uncontrollable growth of abnormal myeloid cells, which pushing out the normal blood forming cells. The quick appearance of symptoms like tiredness, anemia, and higher vulnerability to infections how swiftly this type of leukemia develops. Acute myeloid leukemia is more frequently seen in adults than in children and is more prevalent in men than women. It is considered the most dangerous forms of leukemia, with only a 26.9 % survival rate over a five years' period after diagnosis [3].

Chronic Lymphocytic Leukemia (CLL) primarily affects B-lymphocytes, a type of white blood cell involved in the immune response. CLL is characterized by the accumulation of mature but dysfunctional lymphocytes in the blood, bone marrow, and lymph nodes. It is the most common type of leukemia in adults, typically occurring in older individuals. Chronic lymphocytic leukemia is affect in people aged 55 and older, particularly in men, who make up two-thirds of patients. The survival rate for CLL from 2007 to 2013 was 83.2% [3]. CLL often progresses slowly, and some patients may not require immediate treatment. However, in cases where intervention is necessary, treatment options include chemotherapy, immunotherapy, and targeted therapies. The indolent nature of CLL requires careful monitoring and personalized treatment approaches based on the patient's overall health and specific characteristics of the disease.

Chronic Myeloid Leukemia (CML) is a condition where the bone marrow produces too many mature myeloid cells, especially granulocytes, over time. Unlike the acute forms of leukemia, CML progresses more slowly, and patients may not exhibit noticeable symptoms in the early stages. However, if left untreated, CML can transform into a more acute phase known as blast crisis. The hallmark of CML is the presence of the Philadelphia chromosome, a genetic abnormality that plays a crucial role in the disease. Targeted therapies, such as tyrosine kinase inhibitors, have revolutionized the management of CML,

leading to improved outcomes for many patients. Chronic myeloid leukemia mostly affects adults, and the survival rate for this type of leukemia over five years is 66.9%.

Each subtype presents distinct challenges in terms of diagnosis, prognosis, and treatment. The acute forms progress rapidly, necessitating prompt intervention, while the chronic variants often exhibit a more indolent course, requiring long-term management strategies. The identification of leukemia subtypes through blood analysis will be reduced and moved quicker with the use of intelligent diagnostic techniques. Microscopic blood tests are considered the main way to diagnose leukemia due to their detailed features [2].

This research aims to address these challenges by proposing an automated Leukemia Subtype Classification system based on Convolutional Neural Networks (CNNs). CNN, a technology widely used in automation, is gaining popularity for applying artificial intelligence. It has significantly progressed in using advanced methods to diagnose, test, and classify different diseases through medical imaging. [1]. With every convolutional layer, the architecture acquires knowledge about distinct features in the images, generating heightened activations. CNNs function as feature extractors throughout their layers. The assessment of the suggested CNN model's performance involves metrics such as accuracy, precision, recall, sensitivity, and specificity. To offer a comprehensive and measurable evaluation, it is essential to develop an automated framework that incorporates image processing, signal analysis, and pattern recognition, ensuring precise outcomes and efficient processing.

1.2 Inspiration for the Research

Medical professionals need accurate diagnosis quickly to help patients as early as possible. The motivation for this research stems from the need to enhance the efficiency and accuracy of leukemia diagnosis. Manual classification of leukemia subtypes from microscopic images is prone to inter-observer variability and can be labor-intensive. This study uses advanced technology to assist blood doctors in identifying types of leukemia very accurately. This helps avoid mistakes and makes sure patients get better care. Figuring

out leukemia types by hand takes a lot of time and uses up a lot of resources. A computer system that does this automatically can make things easier for doctors and let them concentrate on other important jobs. The idea of using a special computer program, like a smart camera, to decide what type of leukemia someone has is a way to mix modern technology and medical work. If we create more of these smart systems, we can make healthcare better and more dependable. By leveraging the capabilities of deep learning, particularly CNNs, we aim to develop a robust and automated system that can assist medical professionals in accurately categorizing leukemia subtypes, thereby improving diagnostic speed and precision.

1.3 Problem Statement

In automated technology systems, The National Cancer Institute predicts that leukemia killed 24,500 people in 2017 and was responsible for 4.1% of all cancer cases and deaths in the United States. It becomes necessary to address the difficulties involved in manually classifying leukemia subtypes from microscopic images to lower this concerning rate.

There are major challenges with the current manual method of classifying leukemia subtypes in terms of accuracy, speed, and resource usage. In addition to taking a lot of time, this procedure is prone to human error, which increases the possibility of classification errors. An automated system that can effectively and precisely identify leukemia subtypes from microscopic images is desperately needed, as leukemia cases depend critically on quick and accurate diagnoses.

Automation is a workable way to reduce the possibility of a false positive and make the best use of the time and knowledge of medical professionals. Healthcare workers can better organize their time by using an automated Leukemia Subtype Classification system, which can greatly minimize the requirement for manual work.

1.4 Purpose of the Research

The primary purposes of this research are as follows:

1. Develop a Convolutional Neural Network (CNN) model which is good at recognizing patterns in images, to figure out the different types of leukemia from microscopic images.
2. This system can automatically classify leukemia subtypes without human intervention, making the diagnosis process faster and more efficient.
3. Optimize the CNN model to achieve high accuracy, sensitivity, and specificity in distinguishing between different leukemia subtypes.
4. Evaluate the proposed automated system using a diverse dataset, comparing its performance against existing manual classification methods.
5. This project aims to assist healthcare professionals in diagnosing leukemia more effectively, potentially leading to quicker and more targeted treatment for patients.

1.5 Research Question

1. How can Convolutional Neural Networks be effectively utilized for the automated classification of leukemia subtypes from microscopic images, and how does the proposed system compare to manual classification methods in terms of accuracy and efficiency?

1.6 Report Layout

Chapter 1 Introduces the research objectives and primary topics of investigation

Chapter 2 A review of the relevant literature is highlighted.

Chapter 3 Outlines in detail the proposed methodology.

Chapter 4 provides an explanation of the result analysis of existing output.

Chapter 5 Describe the system analysis that built for leukemia classification prediction.

Chapter 6 Discuss about Impact of Social, Environment and Sustainability for our project.

Chapter 7 Summarizes the current research and outlines potential directions for future research.

CHAPTER 2

Literature Review

2.1 Introduction

In this chapter, we delve into existing research relevant to our current work on creating a framework for a medical imaging system that focuses on leukemia blood cancer cell images in the hematology laboratory.

Our research explores the significance of various methods, including manual analysis using a microscope, electronic blood tests, and computer-based approaches for analyzing hematology images. We thoroughly review previous studies to gain insight into the context of our investigation. The focus of this review is on how traditional blood tests and computer-based techniques can improve the analysis of blood cancer tissue in blood tests. Our goal is to compare the accuracy of these methods with the traditional manual analysis approach.

2.2 Related work

We are taking a close look at using a convolutional neural network to automatically classify different subtypes of leukemia from microscopic images.

This study presents an automated method for leukemia diagnosis using bone marrow cell images and convolutional neural networks with transfer learning. The traditional manual microscopy method is subjective and prone to errors. The study collected images from 104 individuals, including healthy and leukemia patients, and used preprocessing techniques. Three CNN frameworks were tested, with DenseNet121 showing the best results. Transfer learning significantly improved prediction accuracy, reaching 95.3%. The method proves to be rapid, accurate, and objective compared to manual microscopy, offering a promising approach for leukemia diagnosis [1].

Leukemia, a severe form of cancer, was studied using microscopic blood cell images. A novel approach employing Convolutional Neural Networks was proposed for accurate diagnosis of the four leukemia subtypes. Data augmentation techniques were used to address limited training samples. The CNN model demonstrated superior performance 88.25% accuracy for binary classification and 81.74% for multi-classification compared to traditional machine learning algorithms. The study emphasizes the importance of CNNs in leukemia diagnosis from blood smears and highlights the need for larger datasets and ongoing research for enhanced accuracy [2].

This paper focuses on early diagnosis of acute myeloid leukemia, a type of blood cancer. The authors propose a Convolutional Neural Network method achieving 96.6% accuracy in distinguishing normal and abnormal blood cell images. Leukemia has different types, and early detection is crucial for effective treatment. The CNN uses advanced features, reducing processing time and improving accuracy compared to traditional methods. Data augmentation methods like histogram equalization and image transformations enhance the dataset. Experimental results show the CNN outperforms conventional statistical features, making it a promising tool for leukemia diagnosis. Future work aims to evaluate the model on a broader leukemia dataset [3].

This study introduces an Internet of Medical Things (IoMT) framework for the quick and accurate diagnosis of leukemia, a white blood cell-related illness. The proposed system utilizes cloud computing to connect clinical devices, enabling real-time coordination among patients and healthcare professionals. The framework employs Dense Convolutional Neural Network is DenseNet-121 and Residual Convolutional Neural Network is ResNet-34 for leukemia subtype identification, using publicly available datasets. Results indicate that the proposed models outperform other machine learning algorithms in identifying leukemia subtypes. The study emphasizes the potential of IoMT in improving healthcare, especially during pandemics like COVID-19, by facilitating remote diagnosis and treatment [4].

The study addresses leukemia detection and classification using three segmentation algorithms: k-means clustering, Marker-controlled Watershed, and HSV color-based segmentation. Utilizing a dataset of 220 blood smear images, the research focuses on automating leukemia detection through image processing techniques. Feature extraction involves parameters like cell size, mean, entropy, and more. The SVM classifier aids in distinguishing leukemic and non-leukemic cells, going further to classify leukemia types (ALL, AML, CML, CLL). The paper emphasizes a comprehensive approach, surpassing previous methods by classifying multiple leukemia subtypes. Experimental results showcase successful segmentation and feature extraction, enabling accurate detection and classification. The study concludes with a call for future exploration, suggesting the potential for subtype detection and the exploration of additional segmentation algorithms for improved outcomes [5].

This paper introduces a new method, LeukemiaNet, a smart computer system that helps doctors identify different types of leukemia from blood cell images. Leukemia is a dangerous cancer, and early detection is crucial. LeukemiaNet uses a special kind of artificial intelligence called a convolutional neural network to analyze images and achieve 97.2% accuracy. Traditional methods are complex, but LeukemiaNet simplifies the process, making it easier for doctors to diagnose leukemia quickly and accurately. The system is trained on a diverse dataset and outperforms previous methods. This research brings a promising solution to improve early clinical leukemia diagnosis [6].

This study introduces a new method to diagnose leukemia using computer technology. Instead of traditional methods, it employs pre-trained Convolutional Neural Networks (CNNs) to extract features from blood images without the need for segmentation. The proposed approach was tested on three diverse image databases and achieved accuracy rates above 99%, outperforming existing methods. Importantly, it doesn't rely on a segmentation process, making it more robust. The results suggest that this CNN-based method is effective in leukemia diagnosis and could be a valuable tool in the future [13].

This research introduces a computer-aided system to enhance the diagnosis of Acute Lymphoblastic Leukemia (ALL), a life-threatening disease. Traditional manual blood testing methods are slow and less accurate. The proposed system uses image processing and deep learning, particularly a Convolutional Neural Network (CNN), to classify ALL subtypes and normal bone marrow in stained images. The method achieves a high accuracy of 97.78%, outperforming other classifiers like Naïve Bayesian, KNN, and SVM. The system proves valuable in automating the detection and classification of leukemia, providing a reliable tool for pathologists [14].

Researchers from the National Institute of Technology Rourkela, India, developed a fuzzy-based image analysis system to automate the detection of leukemia in blood images. Leukemia, a childhood hematological condition, requires careful examination of blood smears for accurate diagnosis. The proposed system uses a two-stage color segmentation method, employing features like nucleus shape and texture for leukemia detection. They introduced novel shape features, Hausdorff Dimension and contour signature, and employed Support Vector Machine (SVM) for classification. The system, tested on 108 blood smear images, showed promising results with 93% accuracy, providing a cost-effective and efficient solution for leukemia screening [15].

2.3 Comparative Analysis and Summary

Table 2.1. Comparison Table of Research Paper

Author	Title	Year	Algorithm	Accuracy
Proposed Work	Automated Detection Leukemia Subtype Classification from Microscopic Images: A Convolutional Neural Network Approach	2024	EfficientNetB7 ResNet-50 VGG19 ResNet101 DenseNet201 MobileNet CML-1	97.50%

Ahmed, N.; Yigit, A.; Isik, Z.; Alpkocak, A.	Identification of Leukemia Subtypes from Microscopic Images Using Convolutional Neural Network.	2014	CNN Naïve Bayes Decision Tree k-NN SVM	88.25% for binary classification and 81.74% for multi-classification
Huang, F., Guang, P., Li, F., Liu, X., Zhang, W., & Huang, W	AML, ALL, and CML classification and diagnosis based on bone marrow cell morphology combined with convolutional neural network: A STARD compliant diagnosis research.	2020	CNN Algorithm-Inception-V3 ResNet50 DenseNet121 Also use Transfer Learning for improved accuracy.	95.3%
Thanh, T. T. P., Vununu, C., Atoev, S., Lee, S. H., & Kwon, K. R.	Leukemia blood cell image classification using convolutional neural network.	2018	CNN model	96.6%
Tran, T., Park, J. H., Kwon, O. H., Moon, K. S., Lee, S. H., & Kwon, K. R.	. Classification of leukemia disease in peripheral blood cell images using convolutional neural network.	2018	CNN architecture-LeukemiaNet	97%

The table presents various studies on automated leukemia subtype classification using convolutional neural networks (CNNs) and other algorithms. The proposed work achieved the highest accuracy of 97.50% by employing multiple CNN architectures. Earlier works in 2014 and 2020 utilized CNNs and additional algorithms, with accuracies ranging from 88.25% to 97.2%. Notably, the 2020 study incorporated Transfer Learning for improved accuracy, achieving 95.3%. Overall, the proposed work achieved the highest accuracy of

97.17%, displaying the effectiveness of combining multiple advanced CNN architectures for automated detection leukemia subtype classification from microscopic images.

2.4 Conclusion

Many different convolutional neural networks are often used in various areas. The research articles used different methods and ways of doing things. To make the results better, researchers use different tricks like changing the data, picking out important features, and removing extra information. This study tried different ways of doing things. To fix the problems mentioned earlier, we used techniques to work with the data and used custom pre-trained convolutional neural networks on our equal dataset.

CHAPTER 3

Materials & Methodology

3.1 Research Methodology

For this study, we utilized EfficientNetB7, ResNet-50, VGG19, ResNet101, DenseNet201, MobileNet and CML-1. Figure 1 show the procedural workflow.

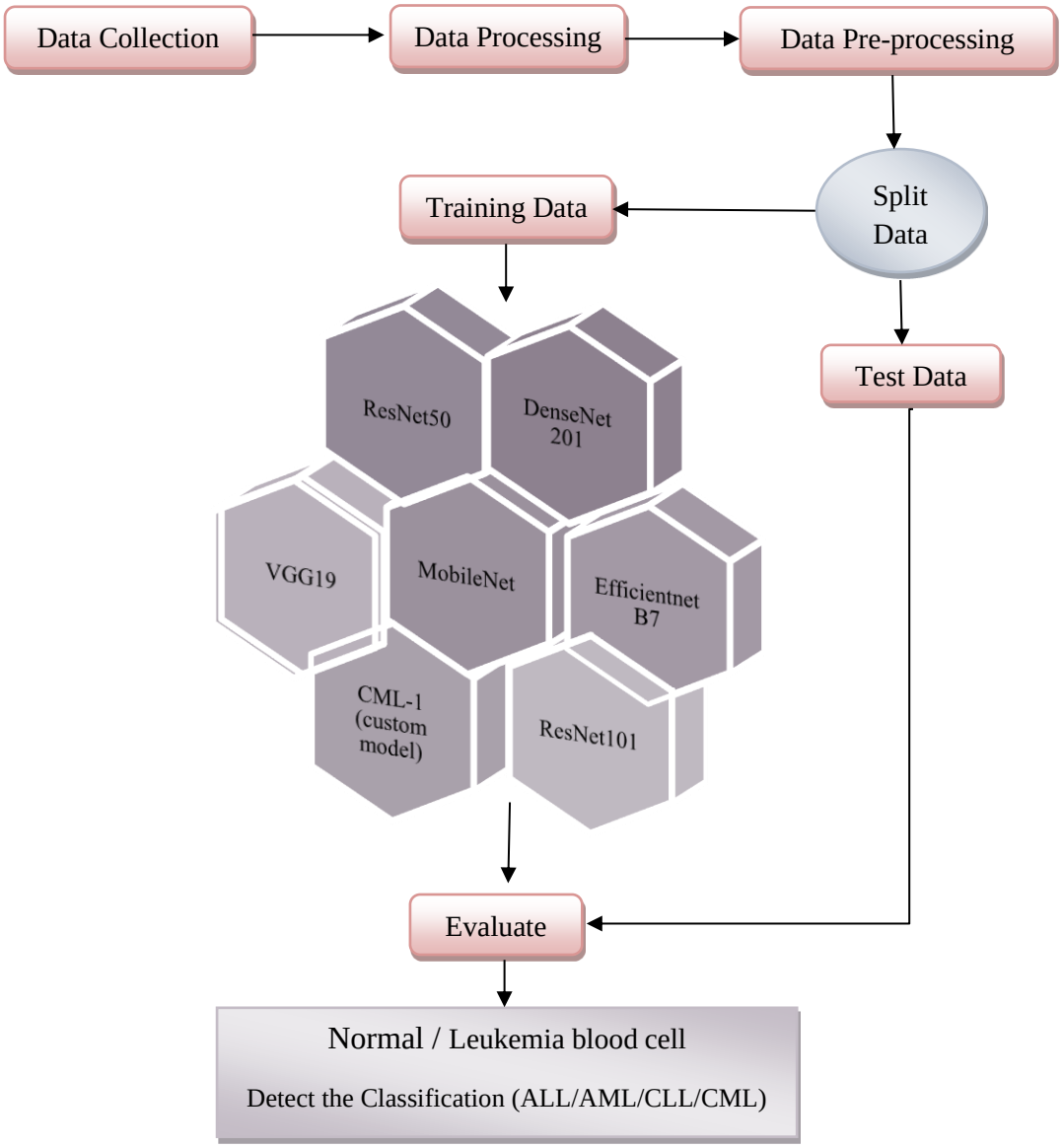


Figure 3.1: Workflow Procedure

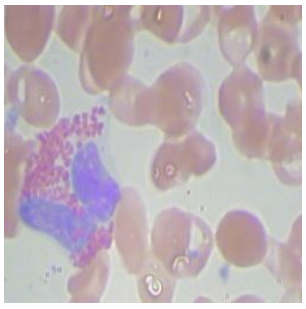
3.2 Data Collection

To learn more about leukemia and find better ways to diagnose and treat it, that's why we gather information from different places. This information helps create a big and effective dataset for research. Effective data collection is important because it helps us understand the different types of leukemia and improves how we can diagnose and treat it. The goal is to build datasets that are complete and of good quality so that we can study the information and learn valuable things from it. This way, we can analyze the data meaningfully and come up with useful insights.

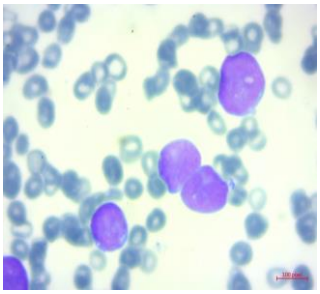
This study utilized data gathered from diverse sources, including LearnHaem (Hematology Made Simple), the American Society of Hematology, and Kaggle. Gathered 366 images of normal blood cells, along with 50 Acute Lymphocytic Leukemia (ALL), 58 Acute Myeloid Leukemia (AML), 68 Chronic Lymphocytic Leukemia (CLL), and 65 Chronic Myeloid Leukemia (CML) images. The images were sourced from Kaggle for normal blood cells and LearnHaem and the American Society of Hematology for the various subtypes of leukemia cancer blood cells.

3.3 Data Processing

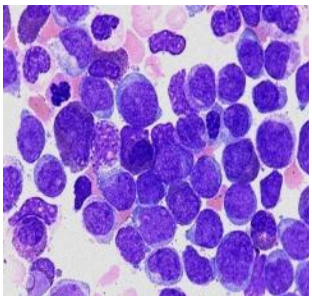
The images categorized into five classes, these classes are represented by healthy_blood_cell, acute_lymphocytic (ALL), acute_myelogenous (AML), chronic_lymphocytic (CLL), and chronic_ myelogenous (CML). Figure 2 show five classes pictures of the dataset.



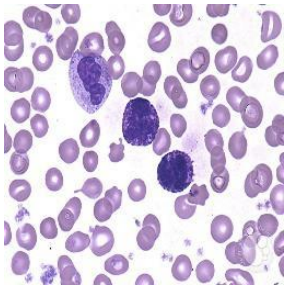
healthy_blood_cell



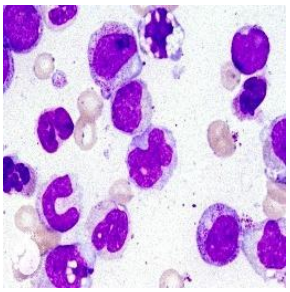
acute_lymphocytic



acute_myelogenous



chronic_ myelogenous



chronic_ lymphocytic

Figure 3.2: Microscopic blood images of each class

3.4 Data Pre-processing

The image format was varied because the images were collected from different places so I converted all the images to the same format. All images are converted to jpg format. After that use Image Augmentation for increasing the dataset to improve model precision and focus on relevant features during training.

3.4.1 Image Augmentation

In this study, we employ image augmentation methods to increase the dataset size, thereby improving the performance of Convolutional Neural Network (CNN) models. Image augmentation involves generating additional data from existing samples, contributing to the diversity of the dataset. For the augmented dataset, we gathered 100 images of normal blood cells, along with 50 ALL, 50 AML, 50 CLL, and 50 CML images. We applied various types of augmentation techniques to these images.

3.4.1.1 Horizontal Flips

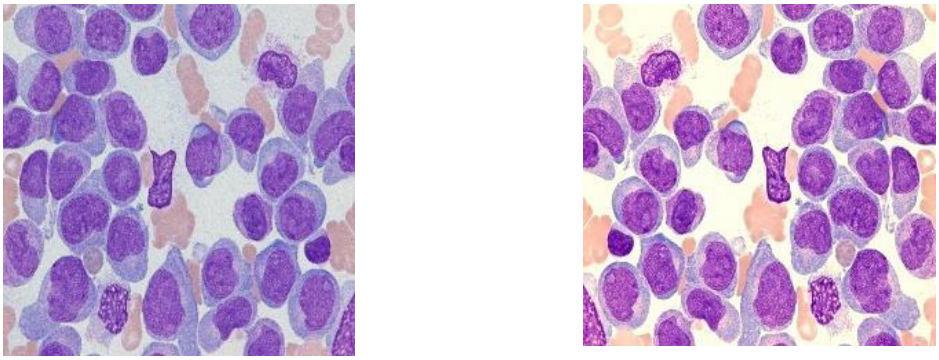
This is a specific augmentation applied in the sequence. It stands for "Flip Left-Right." The augmentation sequence includes the addition of horizontal flips, and each image has a 0.5 chance of undergoing this transformation. This sequence of augmentations can be applied to a set of images, introducing variety to the dataset and aiding in training models to be more robust.

3.4.1.2 Gaussian Blur

There's a 50% chance of applying a Gaussian blur with a random amount of blurring between 0 and 0.5. The blur adds a smoothing effect to the image, and whether it happens or not is determined randomly for each image.

3.4.1.3 Linear Contrast Adjustment

Linear contrast uses to improve the visual quality of an image by optimizing its intensity distribution. This method involves stretching the pixel values in an image to cover the full dynamic range. Adjusts the contrast of the image linearly by a factor randomly chosen between 0.75 and 1.5.



Original Image After Linear Contrast Adjustment

Figure 3.3: Image Contrast Adjustment

3.4.1.4 Additive Gaussian Noise

Gaussian noise to images in a way that for 50% of the images, the noise will be added independently to each color channel of the image. This parameter controls set dynamically between 0.0 and 0.05 times 255. Since pixel values typically range from 0 to 255, this range is proportional to the image's intensity. For the remaining 50%, the noise is added uniformly across all channels, affecting the brightness but not the color.

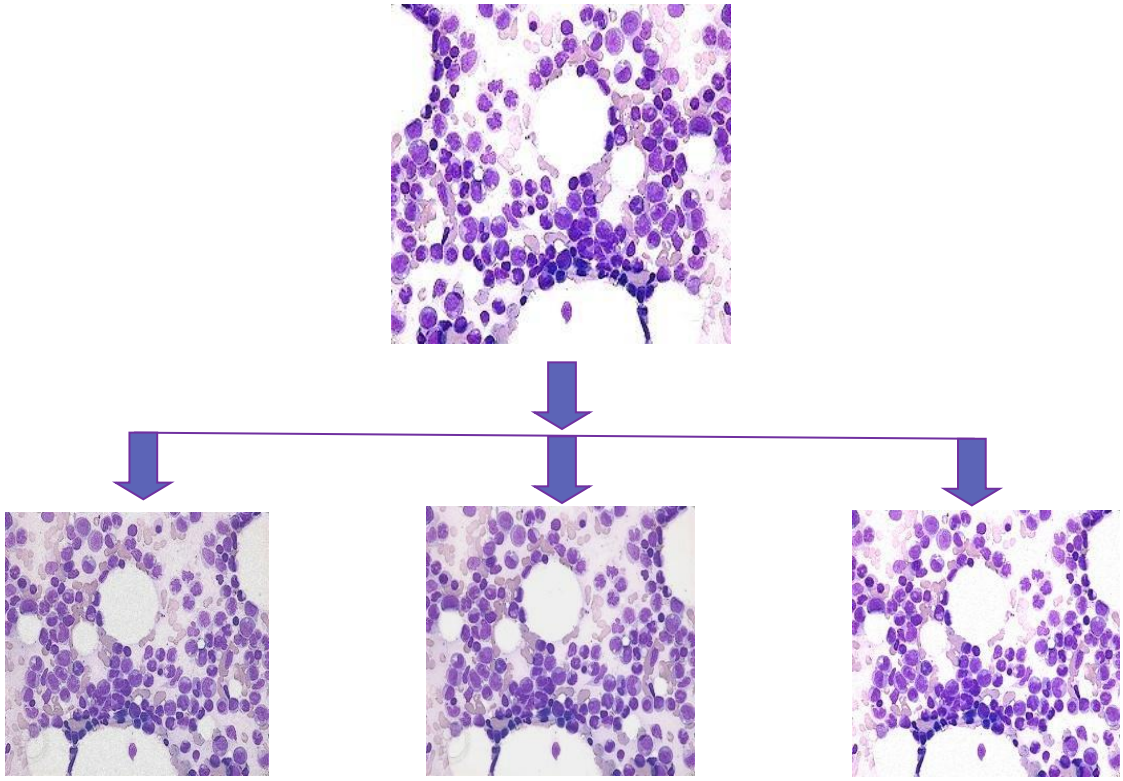
3.4.1.5 Multiplier

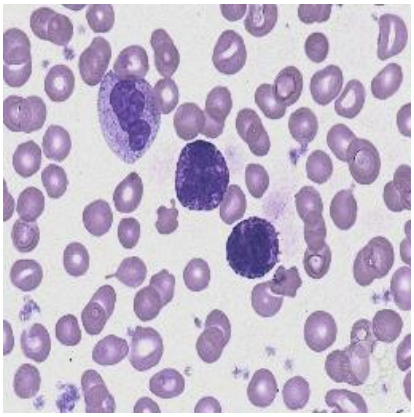
Multiplies the pixel values of the image by a random factor between 0.8 and 1.2. This can make some images darker or brighter. The multiplication is applied independently for each color channel with a 20% probability.

3.4.1.6 Affine for transformations

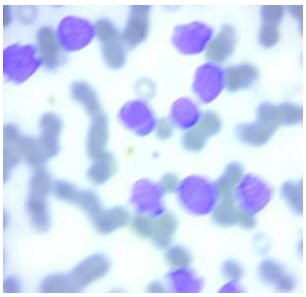
Applies affine transformations to the image, including scaling, translation, rotation, and shearing. These techniques aim to introduce variations and enhance the dataset, ensuring a more robust training process for CNN models.

The augmentation process was applied to the dataset, and the results are visualized in Figure 4.

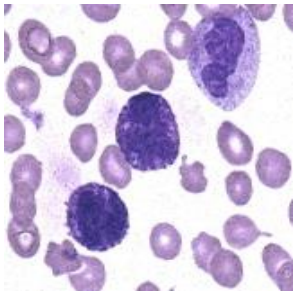




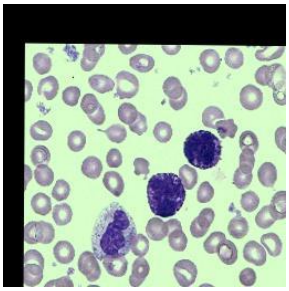
Original Image



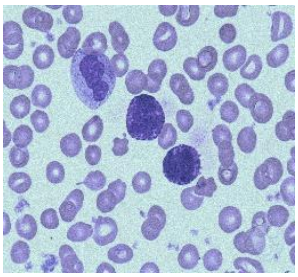
Blurring



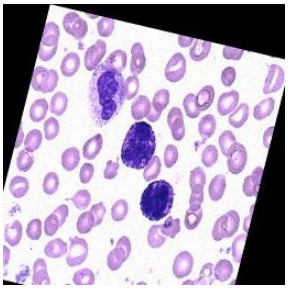
Zooming



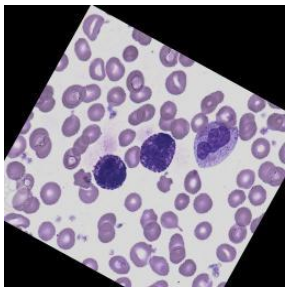
Scaling



Noise



Shearing



Rotation

Figure 3.4: Image Augmentation

After applying the image augmentation technique, the total number of images increased to 3000 images.

Table 3.1. Number of Augmented Images Table

Class Name	Number of Original Images	Number of Augmented Images
acute_lymphocytic	50	600
acute_myelogenous	50	600
chronic_ lymphocytic	50	600
chronic_ myelogenous	50	600
healthy_blood_cell	100	600
Total Number of Image	300	3000

3.5 Data Visualization

For dataset visualization we use EDA analysis. Data visualization with Exploratory Data Analysis (EDA) involves creating charts and graphs to understand and present information from a dataset. By plotting data points, patterns and trends become apparent, helping to identify outlier's correlations. This visual exploration aids in uncovering insights and making data-driven decisions. EDA charts, such as histograms or scatter plots, provide a clear overview of the dataset's distribution, making complex information more accessible to both technical and non-technical.

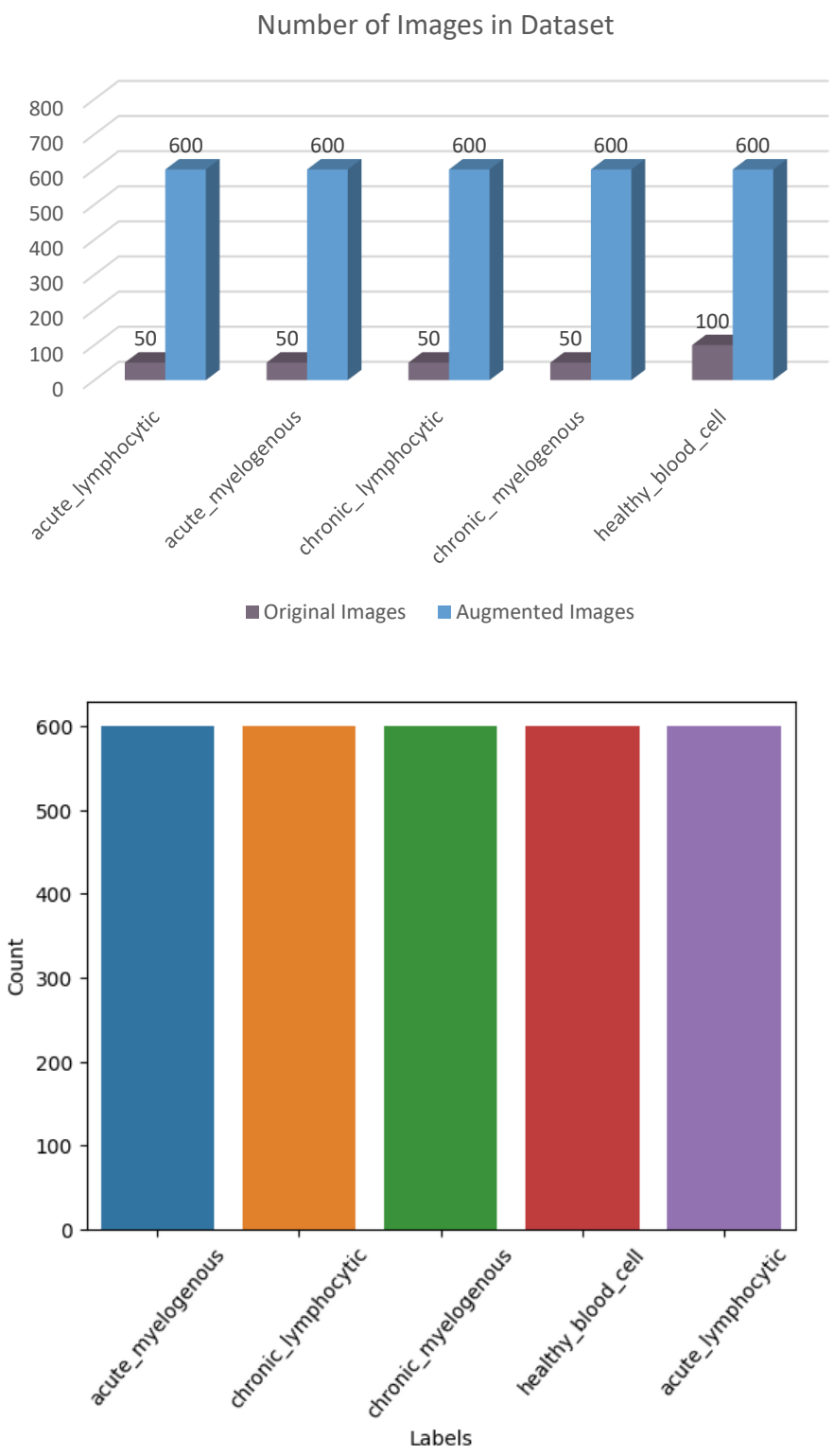


Figure 3.5: Dataset Visualization

3.6 Convolution Neural Network

A Convolutional Neural Network (CNN) is a specialized neural network specifically crafted for tasks related to image processing and pattern recognition. It consists of layers that acquire hierarchical representations of features, making it highly efficient for visual data tasks like the classification of leukemia subtypes based on microscopic images. CNNs used fully connected neural networks to classify the input images after automatically extracting features from them. Layers for max-pooling and convolution were used to extract features. After the application of filters to the image within each layer, the features were extracted, and the process of classification started [2].

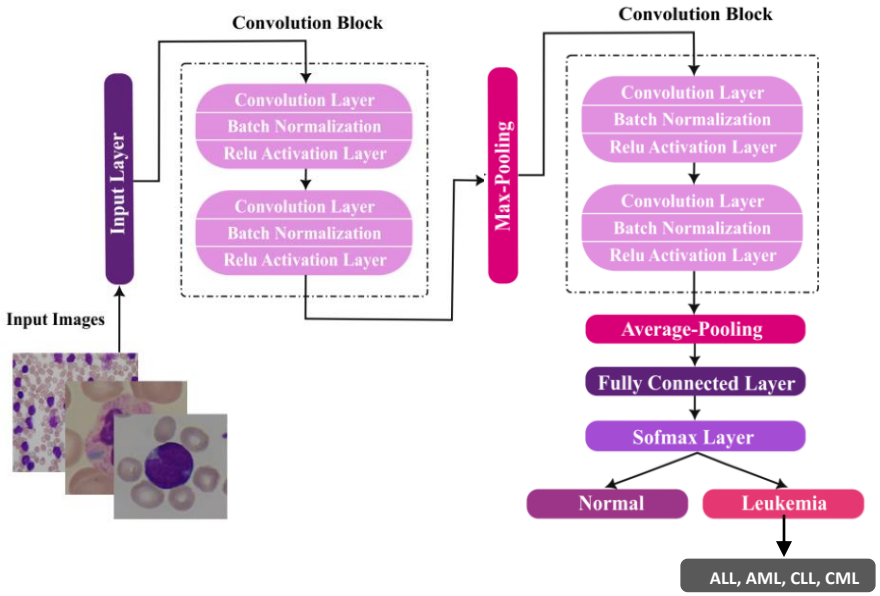


Figure 3.6: Architecture of Convolution Neural Network

3.6.1 Convolutional Layer

The Convolutional Layer in a neural network applies filters to input data, capturing local patterns and features. This enables the network to recognize spatial hierarchies and extract meaningful information from the input. Applying multiple feature detection algorithms for analyzing various filters on the input image was its responsibility. The CNN that we applied featured a 32-feature map that identified 3 × 3. Convolution filters have been placed to the

image. The values of the filters were chosen at random. [2]. Allows the network to capture spatial hierarchies and detect local patterns, crucial for understanding the intricate details in microscopic images.

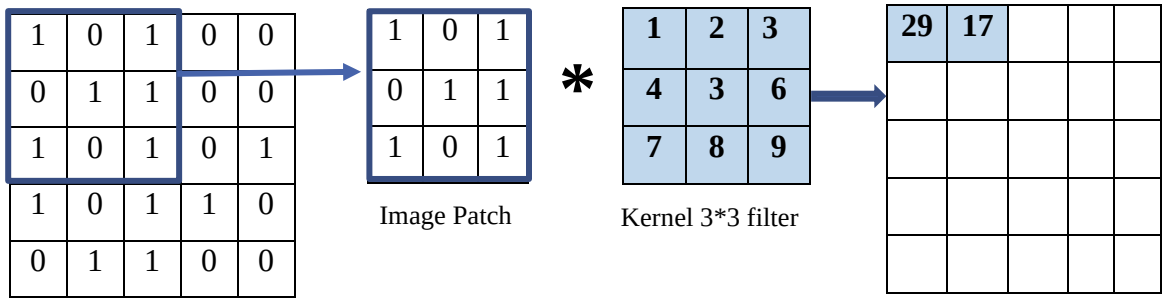


Figure 3.7: Structure of Convolution Layer

3.6.2 Pooling Layer

The Max-Pooling Layer is a crucial component of CNNs that follows convolutional layers. This layer aids in down sampling feature maps, reducing spatial dimensions while retaining essential information. Max pooling involves selecting the maximum value within specific regions of the feature maps, facilitating feature selection and abstraction. This layer was tasked with reducing the dimensions of the filtered image to concentrate on crucial features, areas, or objects within the image. In our network, we employed two max-pooling layers, each with a size of 2 × 2, and also doubled the quantity of this particular layer [2].

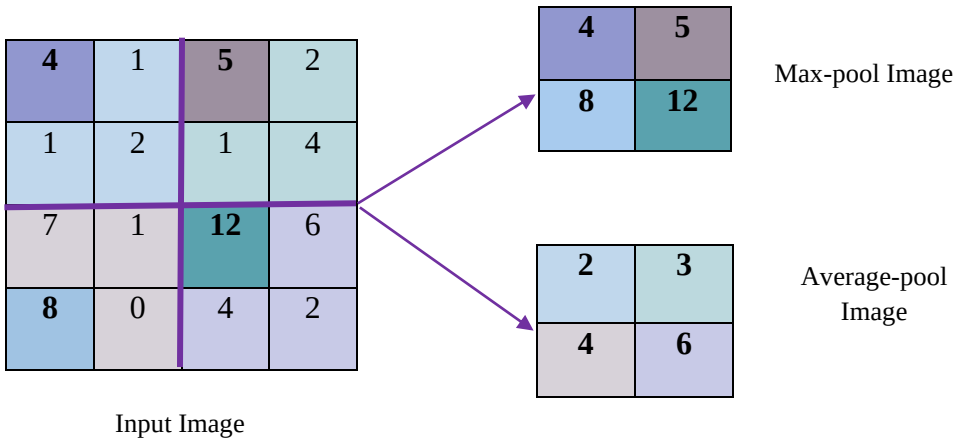
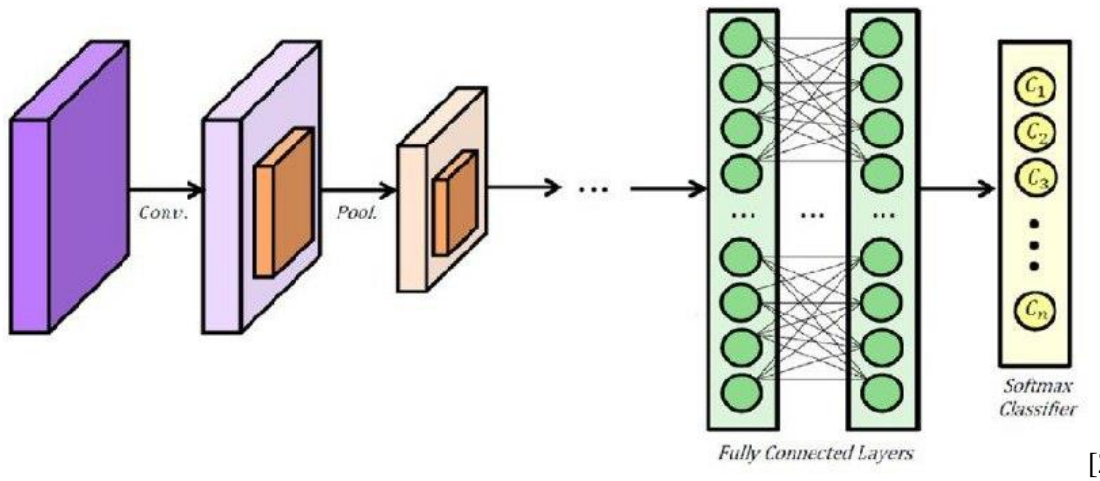


Figure 3.8: Structure of Pooling Layer

3.6.3 Connected Layer

In the last part of the Convolutional Neural Network (CNN), there is a crucial layer called the "Fully Connected Layer and Softmax Layer" A fully connected layer in a neural network connects each neuron to every neuron in the adjacent layer, allowing for complex feature learning and transformation. It's essential because it helps the network make predictions using the learned features. The Fully Connected Layer lets these decision-makers share information with each other, discussing what they've learned from the input images. The softmax layer is typically the output layer that converts raw scores or logits into probability distributions, ensuring that the sum of the probabilities is 1, making it suitable for multi-class classification problems. These are allowing the network to make smart predictions about the type of leukemia shown in microscopic images



[25]

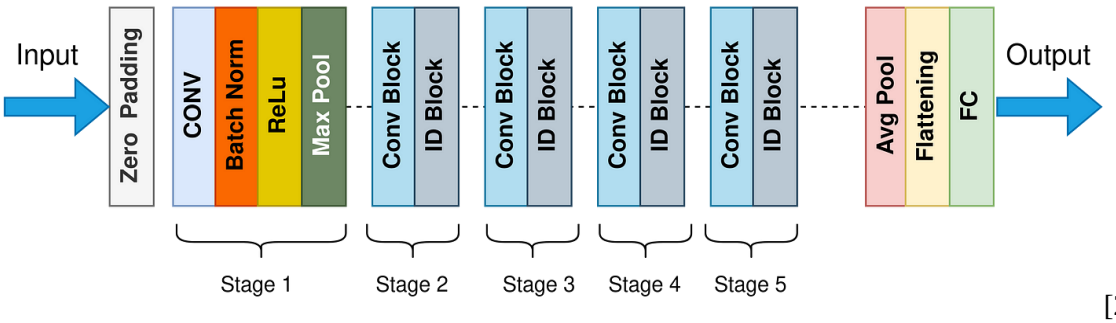
Figure 3.9: Structure of Fully Connected Layer and Softmax Layer

3.7 Convolution Neural Network Models

For leukemia subtype classification, various pre-trained Convolutional Neural Network (CNN) models have demonstrated effectiveness in image classification tasks. These models, which have been previously trained on large and diverse datasets, bring a wealth of knowledge and feature extraction capabilities that can significantly benefit the accurate classification of leukemia subtypes. Leveraging these pre-trained models allows for the transfer of learned features, enhancing the performance and efficiency of the classification process for microscopic images related to leukemia. Some notable custom pre-trained CNN models applied in this context include ResNet-50, DenseNet201, VGG19, ResNet101, EfficientNetB7, MobileNet and built a deep convolutional neural network CML-1 .

3.7.1 ResNet-50

ResNet-50 belongs to the Residual Network (ResNet) family and features 50 layers. ResNet architectures are distinguished by the incorporation of skip connections or residual blocks, utilizing shortcut connections to mitigate vanishing gradient issues. The model's depth enhances its ability to discern intricate patterns, contributing to accurate leukemia subtype classification. The 34-layer network comprises two types of layers, MaxPool and Normal Pool, organized in 2-layer blocks. These blocks undergo a shift using a 3-layer bottleneck block, resulting in a 50-layer ResNet. The image size remains constant at 224 by 224 pixels throughout this process.



[26]

Figure 3.10: Architecture of ResNet-50

3.7.2 DenseNet201

DenseNet201 follows the DenseNet (Densely Connected Convolutional Networks) architecture. DenseNet-201 is a neural network architecture with an impressive depth of 201 layers. Specifically designed for image-related tasks, it operates with an image input size of 224 × 224 pixels. Each architectural design comprises four Dense Blocks, each having a distinct number of layers. Dense connections enhance information flow, aiding the network's understanding of intricate patterns in microscopic images, crucial for accurate and efficient leukemia subtype classification.

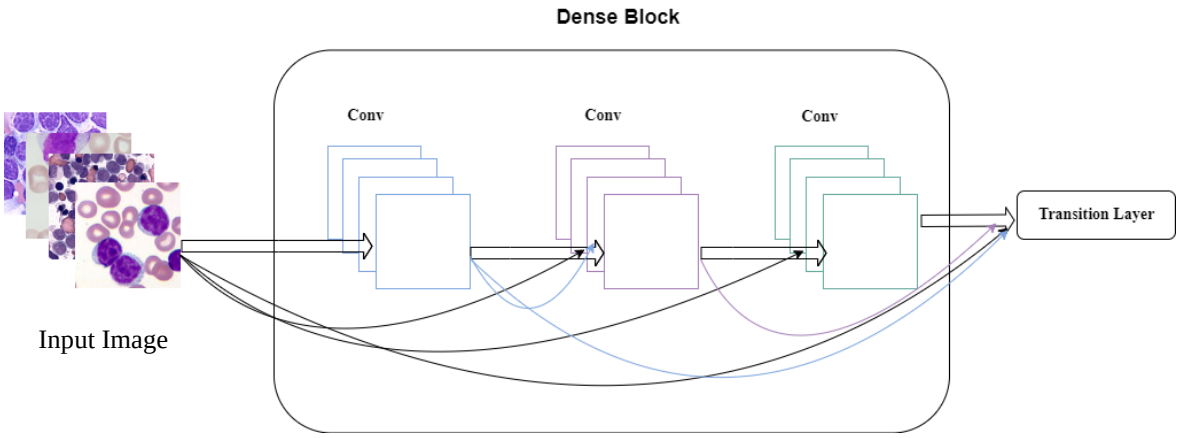


Figure 3.11: Architecture of DenseNet201

3.7.3 MobileNet

MobileNet is a family of lightweight deep learning models designed for mobile and edge devices and also well-suited for tasks like image classification, object detection, and more. The architecture of MobileNet is based on depth wise separable convolutions. There are two fundamental components: 1. Convolutional Unit with a 3x3 kernel 2. Unit consisting of Depth Wise Convolution with a 3x3 kernel, followed by a 1x1 Convolution. The MobileNet architecture consists of 28 layers. Its depthwise separable convolutions enable efficient feature extraction, maintaining accuracy with lower computational requirements. The model learns distinctive patterns in the images, aiding in precise classification of leukemia subtypes.

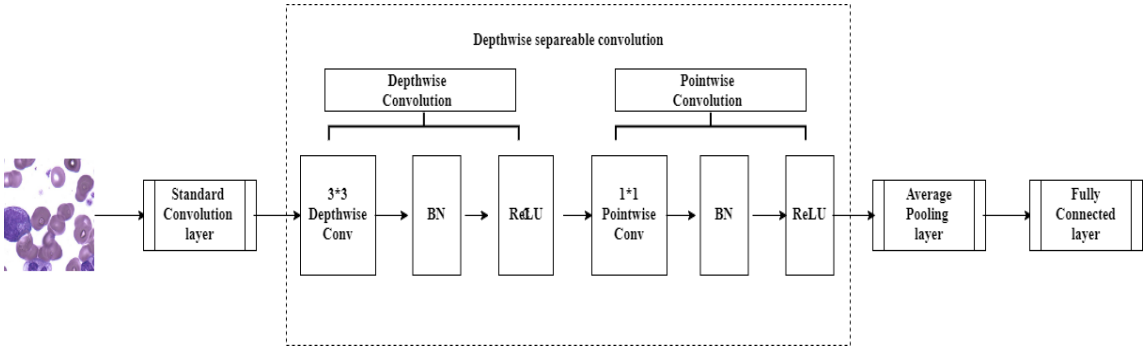


Figure 3.12: Architecture of MobileNet

3.7.4 VGG19

VGG19 is part of CNN architectures. VGG19 is a modified version of the VGG model, characterized by a total of 19 layers with small convolutional filters, effectively capturing hierarchical features. The network is trained to recognize patterns in leukemia images, aiding accurate subtype classification through learned representations. These layers include 16 convolution layers, 3 fully connected layers, 5 MaxPool layers, and culminate in 1 SoftMax layer.

3.7.5 ResNet101

ResNet101 is an extended version of ResNet-50, featuring 101 layers. The increased depth enhances the network's ability to learn and represent complex patterns in images. This makes ResNet101 suitable for tasks where a deeper architecture is advantageous, such as this enables effective feature extraction from microscopic images, aiding accurate classification of leukemia subtypes through learned hierarchical representations, enhancing model performance in complex image recognition tasks.

3.7.6 EfficientNetB7

EfficientNetB7 is a powerful model known for effectively handling various image datasets. It finds a balance between model size and computational efficiency, making it great for accurate classification tasks. EfficientNet-B7 consists of a total of 813 layers. It has more

parameters and is capable of capturing intricate patterns in images. In this study, EfficientNetB7 is likely used as the backbone architecture for the Convolutional Neural Network (CNN) designed for automated classification of leukemia subtypes from microscopic images. The CNN extracts hierarchical features, employs global average pooling, and utilizes fully connected layers with softmax activation for accurate classification. Training optimizes model parameters, enabling automated identification of leukemia subtypes from microscopic images.

3.7.7 CML-1

The model is a deep convolutional neural network designed for image classification tasks. It consists of five convolutional layers to capture hierarchical features, followed by batch normalization and max pooling layers to enhance learning and reduce spatial dimensions. The fully connected layers towards the end of the model contribute to high-level feature representation, ultimately leading to a softmax activation output layer for multi-class classification. The model consists of 23 layers in total. The model uses five convolutional layers, five max pooling layers, six batch normalization layers, one flatten layer, three dense layers and dropout layer.

Table 3. CNN models Layers

Models	**Number of Layers**
ResNet-50	50 layers
DenseNet201	201 layers
MobileNet	28 layers
VGG19	19 layers
ResNet101	101 layers
EfficientNetB7	813 layers
CML-1	23 layers

3.8 Dataset Training and Testing

Dataset consists total 3000 images. Dataset information collected to equip for parted and divided into three parts:

1. Training dataset,
2. Validation and,
3. Test dataset

Training: 80% of total dataset is used for training. Around 1920 images are use foe training. The model learns from the training dataset to understand the patterns and relationship within the data and used to assign the weights and performances that the model relies on.

Validation: From 80% of the training data, 20% data is used for validation. Validation data evaluate trained model in different way. Validation data is used to measure model performance, fundamental accuracy and the parameters of the model. 480 images used for validating. Model is learning and adapting through validation process allowing adjustments and optimization to the model's parameters before final test.

Test: For test dataset, use 20% of data from total dataset. To provide final evaluation of a model and perform similar type of task on different dataset. Use 600 images for testing and evaluate the model performance.

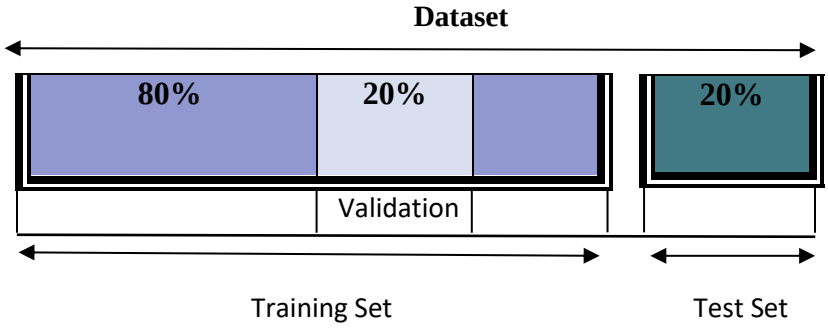


Figure 3.13: Dataset Split

CHAPTER 4

Experimental Result & Discussion

4.1 Result Discussion

Overview of Leukemia blood cancer classification is given in this point that is detected by CNN model. After collecting data, preprocessed data through image enhancement and augmentation process. For detecting the cancer cells and normal cell, create 5 classes such as healthy blood cell, acute lymphocytic cancer cell, acute myelogenous cancer cell, chronic lymphocytic cancer cell, and chronic myelogenous cancer cell. Detect the classification of leukemia, ResNet-50 model performed best. Additionally, EfficientNetB7, DenseNet201, VGG19, ResNet101, MobileNet, CML-1 CNN models are used in this research. ResNet-50, MobileNet and ResNet101 models are achieve the highest accuracy that is 97%, but the ResNet-50 models performed best according to the classification reports. Other models are also performed such as DenseNet201 95%, VGG19 is 95%, EfficientNetB7 is 96% and CML-1 is 96%.

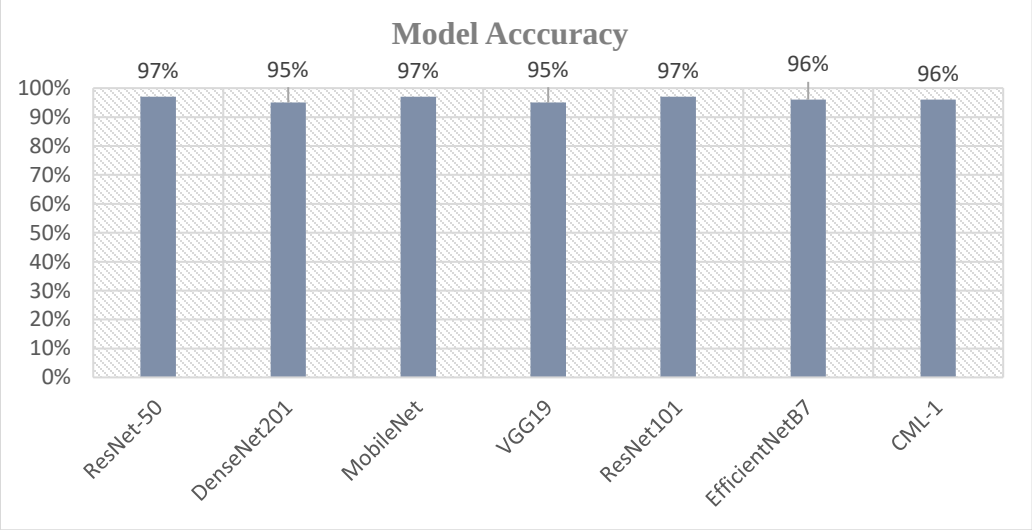


Figure 4.1: Model Accuracy

In this research, we used confusion matrix and many other methods to evaluate the test's validity. These methods analyze and validate the performance of the test on the collected data.

4.2 Performance Metrics

Performance metrics in the context of Convolutional Neural Network and data analysis measure the effectiveness and quality of a model's predictions and classifications. These metrics provide how well a model is performed on a particular task. Enhancing the model's efficacy involves fine-tuning its parameters. In assessing the effectiveness of a classification model, diverse metrics are employed. Some are-

4.2.1 Accuracy

The simplest classification metric to utilize is accuracy, which assesses the overall accuracy of a model's predictions by determining the proportion of correct predictions relative to the total number of predictions made. Accuracy is calculated by using this equation-

$$\frac{\text{(True Positives + True Negatives)}}{\text{(True Positives + False Positive + True Negatives + False Negatives)}}$$

4.2.2 Precision

Accuracy metric constraints are not approved by the accuracy metric. Precision determines the proportion of accurate positive predictions, calculated as the ratio of true positives to the sum of true positives and false positives. It assesses the accuracy of positive predictions by evaluating the relationship between true positives and the combined total of true positives and false positives. Precision is calculated by using this equation-

$$\frac{\text{True Positives}}{\text{True Positives + False Positive}}$$

4.2.3 Recall

Recall, also known as sensitivity or true positive rate, measures the model's effectiveness in correctly identifying all instances of a particular class. It is calculated by dividing the number of true positives by the sum of true positives and false negatives, providing insight into the model's ability to capture all positive observations within a given class. Recall is calculated by using this equation-

$$\frac{\text{True Positives}}{\text{True Positives + False Negatives}}$$

4.2.4 F-1 score

The F1 score, a consolidated metric that balances precision and recall by taking into account both false positives and false negatives, is a weighted average of these two measures. Unlike accuracy, F1 score efficiently considers the trade-off between false positives and false negatives in its calculation. F1 score is calculated by using this equation-

$$\frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision + Recall}}$$

4.3 Classification Report

ResNet-50

	precision	recall	f1-score	support
acute_lymphocytic	1.00	1.00	1.00	122
acute_myelogenous	0.99	0.96	0.97	119
chronic_lymphocytic	0.96	0.93	0.95	105
chronic_myelogenous	0.93	0.98	0.96	130
healthy_blood_cell	1.00	0.99	1.00	124
accuracy			0.97	600
macro avg	0.98	0.97	0.97	600
weighted avg	0.98	0.97	0.98	600

Figure 4.2: Classification Report for ResNet-50

DenseNet201

	precision	recall	f1-score	support
acute_lymphocytic	1.00	1.00	1.00	122
acute_myelogenous	0.98	0.84	0.90	119
chronic_lymphocytic	1.00	0.91	0.96	105
chronic_myelogenous	0.83	0.99	0.90	130
healthy_blood_cell	1.00	1.00	1.00	124
accuracy			0.95	600
macro avg	0.96	0.95	0.95	600
weighted avg	0.96	0.95	0.95	600

Figure 4.3: Classification Report for DenseNet201

MobileNet

	precision	recall	f1-score	support
acute_lymphocytic	0.99	0.99	0.99	122
acute_myelogenous	0.99	0.94	0.97	119
chronic_lymphocytic	0.94	0.97	0.95	105
chronic_myelogenous	0.95	0.94	0.95	130
healthy_blood_cell	0.97	1.00	0.98	124
accuracy			0.97	600
macro avg	0.97	0.97	0.97	600
weighted avg	0.97	0.97	0.97	600

Figure 4.4: Classification Report for MobileNet

VGG19

	precision	recall	f1-score	support
acute_lymphocytic	0.99	0.93	0.96	122
acute_myelogenous	0.93	0.97	0.95	119
chronic_lymphocytic	0.89	0.96	0.93	105
chronic_myelogenous	0.98	0.91	0.94	130
healthy_blood_cell	0.95	0.98	0.97	124
accuracy			0.95	600
macro avg	0.95	0.95	0.95	600
weighted avg	0.95	0.95	0.95	600

Figure 4.5: Classification Report for VGG19

ResNet101

	precision	recall	f1-score	support
acute_lymphocytic	1.00	0.99	1.00	122
acute_myelogenous	0.98	0.95	0.97	119
chronic_lymphocytic	0.96	0.92	0.94	105
chronic_myelogenous	0.91	0.98	0.94	130
healthy_blood_cell	1.00	1.00	1.00	124
accuracy			0.97	600
macro avg	0.97	0.97	0.97	600
weighted avg	0.97	0.97	0.97	600

Figure 4.6: Classification Report for ResNet101

EfficientNetB7

	precision	recall	f1-score	support
acute_lymphocytic	0.98	0.98	0.98	122
acute_myelogenous	0.93	0.92	0.93	119
chronic_lymphocytic	0.97	0.94	0.96	105
chronic_myelogenous	0.92	0.95	0.94	130
healthy_blood_cell	0.98	0.99	0.99	124
accuracy			0.96	600
macro avg	0.96	0.96	0.96	600
weighted avg	0.96	0.96	0.96	600

Figure 4.7: Classification Report for EfficientNetB7

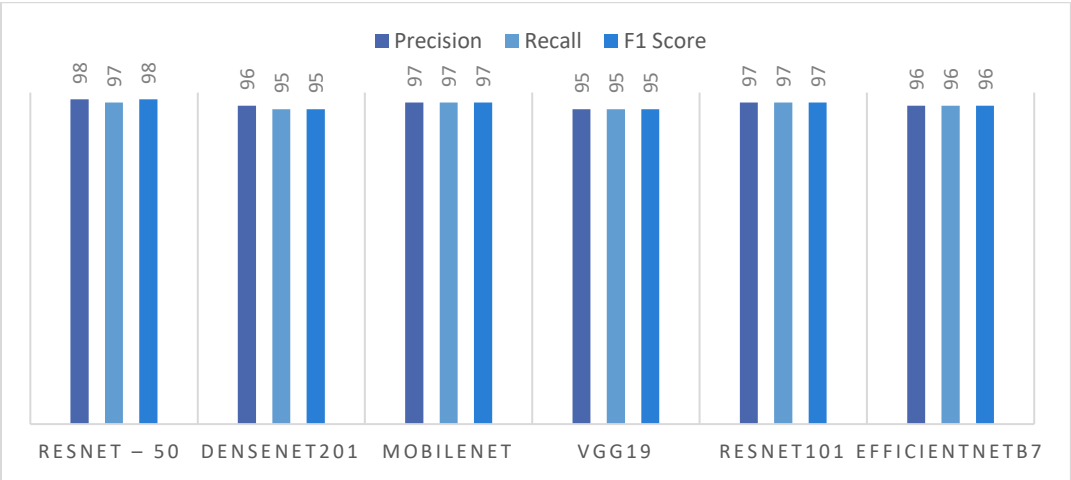


Figure 4.8: Performance Metrics of the models

4.4 Confusion Matrix

In this research, a table used to measure the model's performance is called a confusion matrix. Each class in the dataset's true positive, false positive, true negative, and false negative counts are displayed. A visual representation of the model's performance in accurately predicting each class is provided by a confusion matrix. The predicted classes are shown in columns of a confusion matrix, while the actual classes are represented by rows. The four possible results are as follows:

True Positive (TP): An accurate positive class prediction is made by the model.

False Positive (FP): When a positive class is predicted by the model, the actual class is negative.

True Negative (TN): An accurate negative class prediction is made by the model.

False Negative (FN): While the actual class is positive, the model predicts a negative one.

Multiple metrics can be computed from a confusion matrix to assess the model's performance. Recall, F1 score, and precision are the metrics that are most frequently utilized.

4.4.1 ResNet-50

Table 4.1. Confusion Matrix – ResNet-50

Confusion Matrix – ResNet-50					
acute_lymphoblastic	122	0	0	0	0
acute_myelogenous	0	114	2	3	0
chronic_lymphocytic	0	0	98	7	0
chronic_myelogenous	0	1	1	128	0
healthy_blood_cell	0	0	1	0	123

Confusion matrix summarizes the results of a classification task, showing the number of true positive, true negative, false positive, and false negative predictions. We find from confusion matrix that detected 122 acute lymphoblastic, 114 acute myelogenous, 98 chronic lymphocytic, 128 chronic myelogenous, and 123 healthy_blood_cell after applying the ResNet-50 algorithm on the dataset.

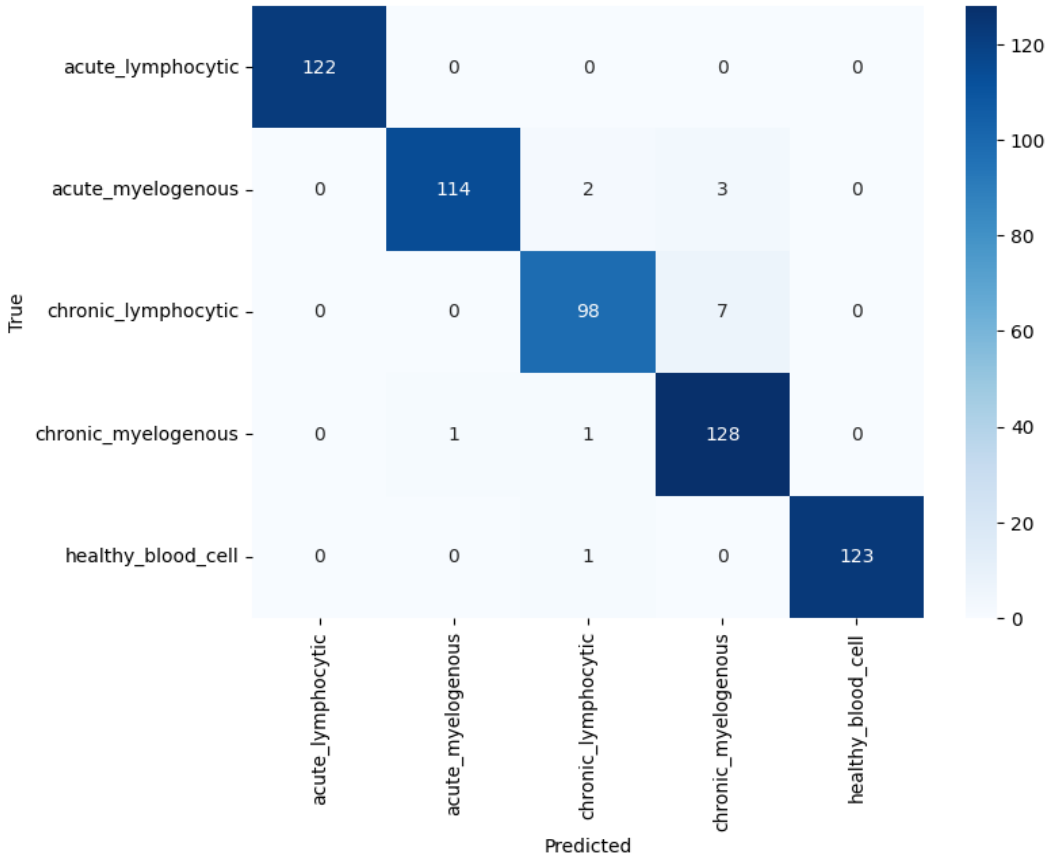


Figure 4.9: Confusion matrix of ResNet-50

4.4.2 DenseNet201

Table 4.2. Confusion Matrix – DenseNet201

Confusion Matrix – DenseNet201					
acute_lymphoblastic	122	0	0	0	0
acute_myelogenous	0	100	0	19	0
chronic_lymphocytic	0	1	96	8	0
chronic_myelogenous	0	1	0	129	0
healthy_blood_cell	0	0	0	0	124

Here is 122 acute lymphoblastic detected, 100 acute myelogenous detected, 96 chronic lymphocytic detected, 129 chronic myelogenous detected, and 124 healthy_blood_cell detected after applying the DenseNet201 algorithm.

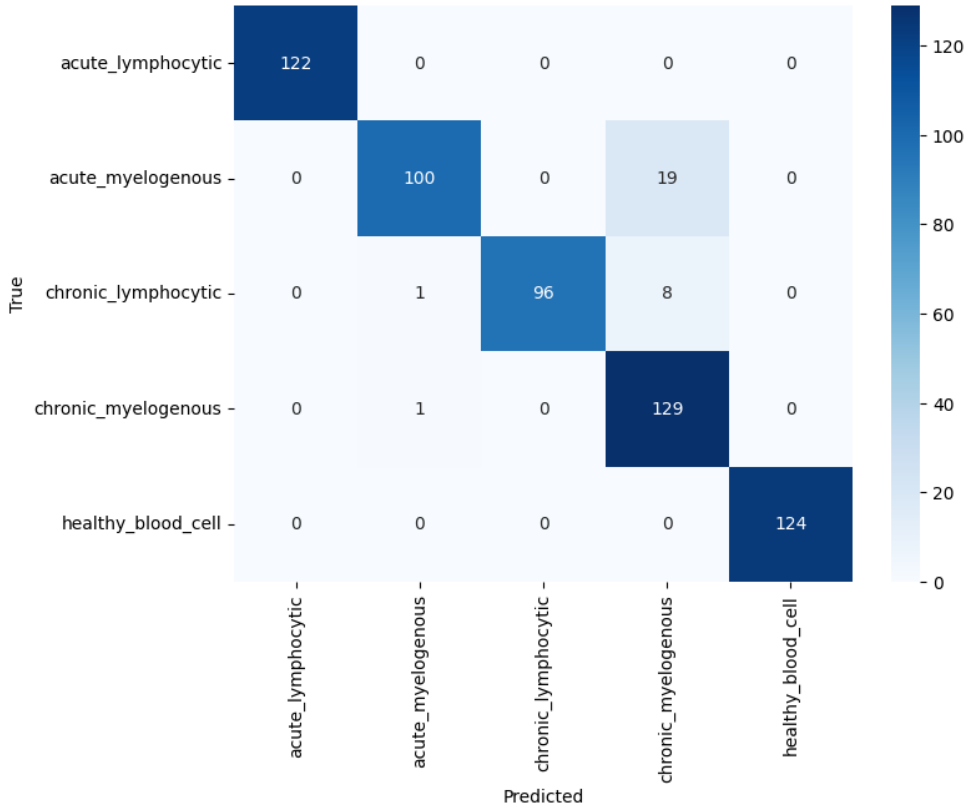


Figure 4.10: Confusion matrix of DenseNet201

4.4.3 MobileNet

Table 4.3. Confusion Matrix – MobileNet

Confusion Matrix – MobileNet					
acute_lymphoblastic	121	0	0	1	0
acute_myelogenous	1	112	2	2	2
chronic_lymphocytic	0	0	102	3	0
chronic_myelogenous	0	1	5	122	2
healthy_blood_cell	0	0	0	0	124

The MobileNet algorithm identified 121 cases of acute lymphoblastic leukemia, 112 cases of acute myelogenous leukemia, 102 cases of chronic lymphocytic leukemia, 122 cases of chronic myelogenous leukemia, and 124 instances of healthy blood cells.

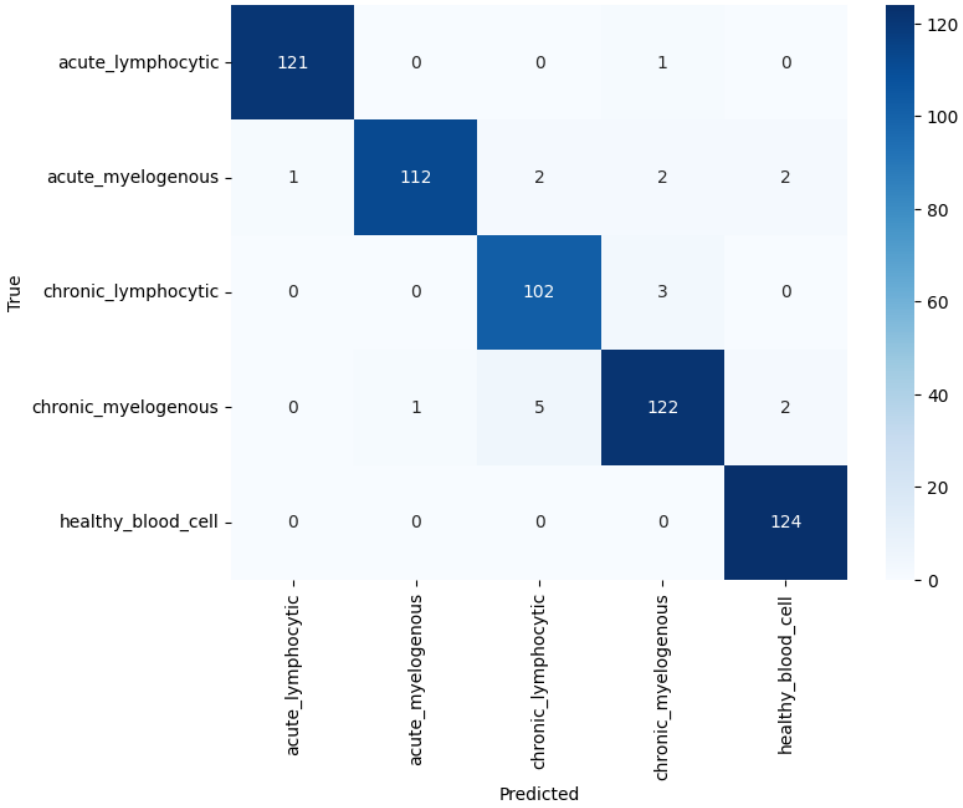


Figure 4.11: Confusion matrix of MobileNet

4.4.4 VGG19

Table 4.4. Confusion Matrix – VGG19

Confusion Matrix – VGG19					
acute_lymphoblastic	114	0	5	1	2
acute_myelogenous	0	115	2	0	2
chronic_lymphocytic	1	3	101	0	0
chronic_myelogenous	0	5	5	118	2
healthy_blood_cell	0	1	0	1	122

Detected 114 acute lymphoblastic, 115 acute myelogenous, 101 chronic lymphocytic, 118 chronic myelogenous, and 122 healthy_blood_cell after applying the VGG19 algorithm on the dataset.

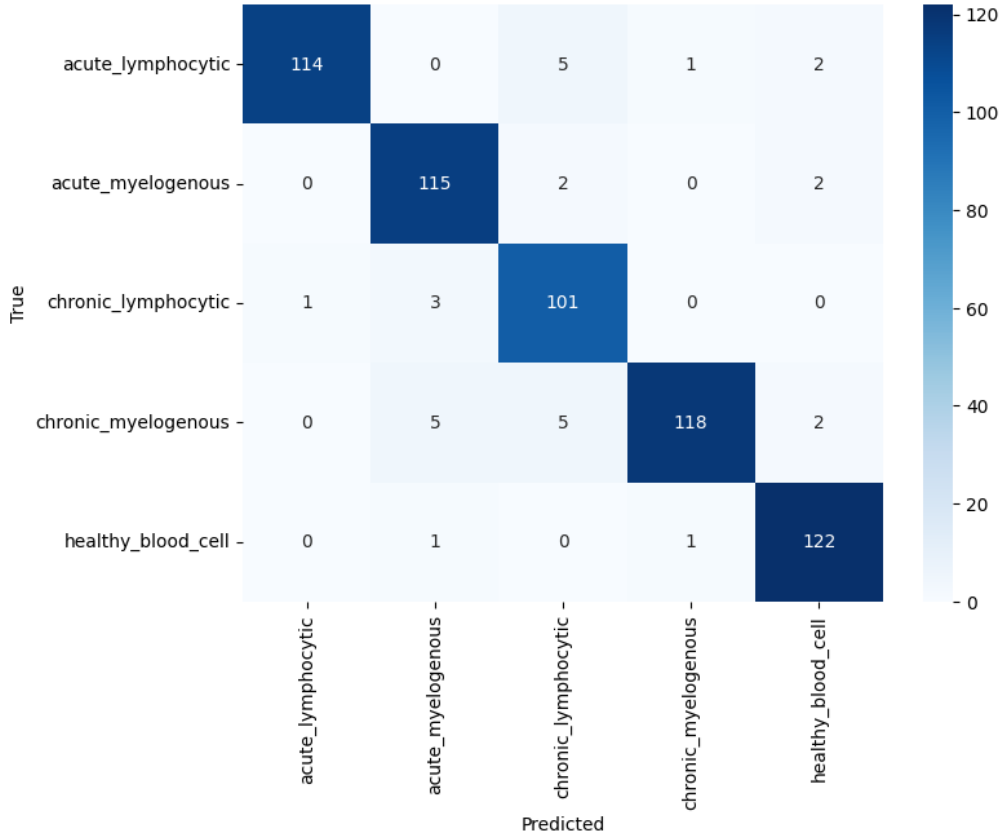


Figure 4.12: Confusion matrix of VGG19

4.4.5 ResNet101

Table 4.5. Confusion Matrix – ResNet-101

Confusion Matrix – ResNet-101					
acute_lymphoblastic	121	0	1	0	0
acute_myelogenous	0	113	2	4	0
chronic_lymphocytic	0	0	97	8	0
chronic_myelogenous	0	2	1	127	0
healthy_blood_cell	0	0	0	0	124

In confusion matrix, we observed that 121 acute lymphoblastic, 113 acute myelogenous, 97 chronic lymphocytic, 127 chronic myelogenous, and 124 healthy_blood_cell have been detected from dataset after applying the ResNet-101 algorithm.

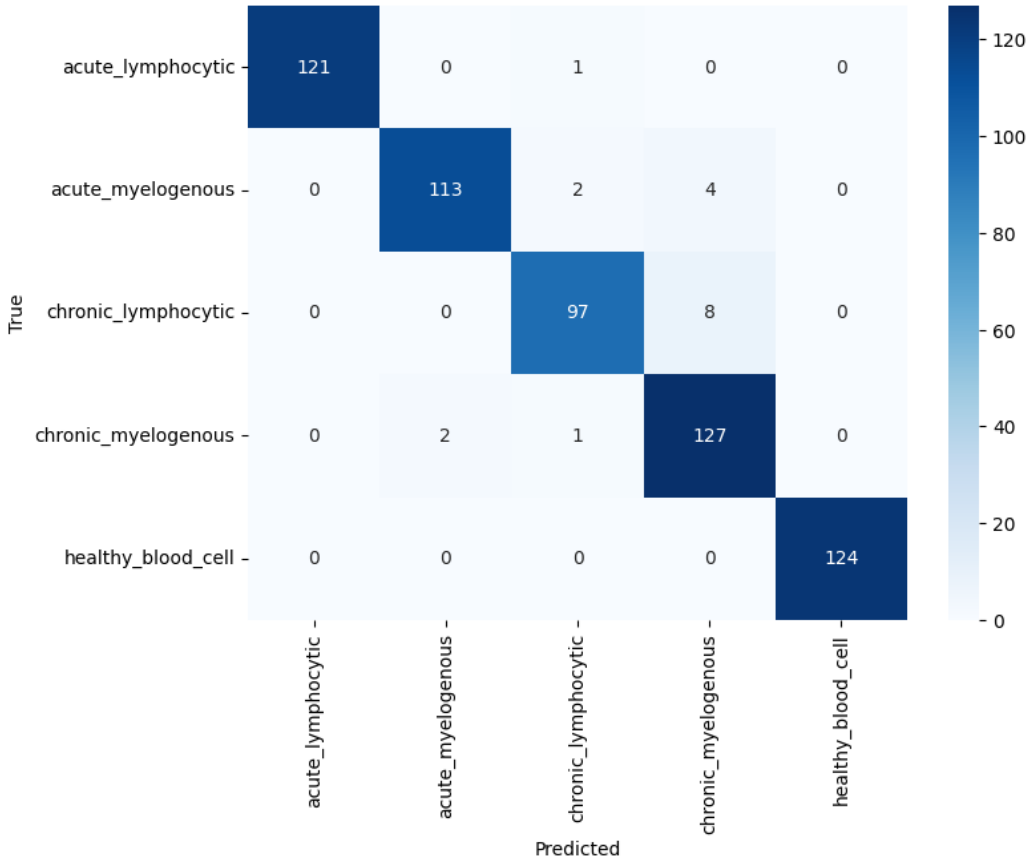


Figure 4.11: Confusion matrix of ResNet-101

4.4.6 EfficientNetB7

Table 4.6. Confusion Matrix – EfficientNetB7

Confusion Matrix – EfficientNetB7					
acute_lymphoblastic	119	3	0	0	0
acute_myelogenous	3	110	1	5	0
chronic_lymphocytic	0	2	99	4	0
chronic_myelogenous	0	3	2	123	2
healthy_blood_cell	0	0	0	1	123

We find from confusion matrix that detected 119 acute lymphoblastic, 110 acute myelogenous, 99 chronic lymphocytic, 123 chronic myelogenous, and 123 healthy_blood_cell after applying the EfficientNetB7 algorithm on the dataset.

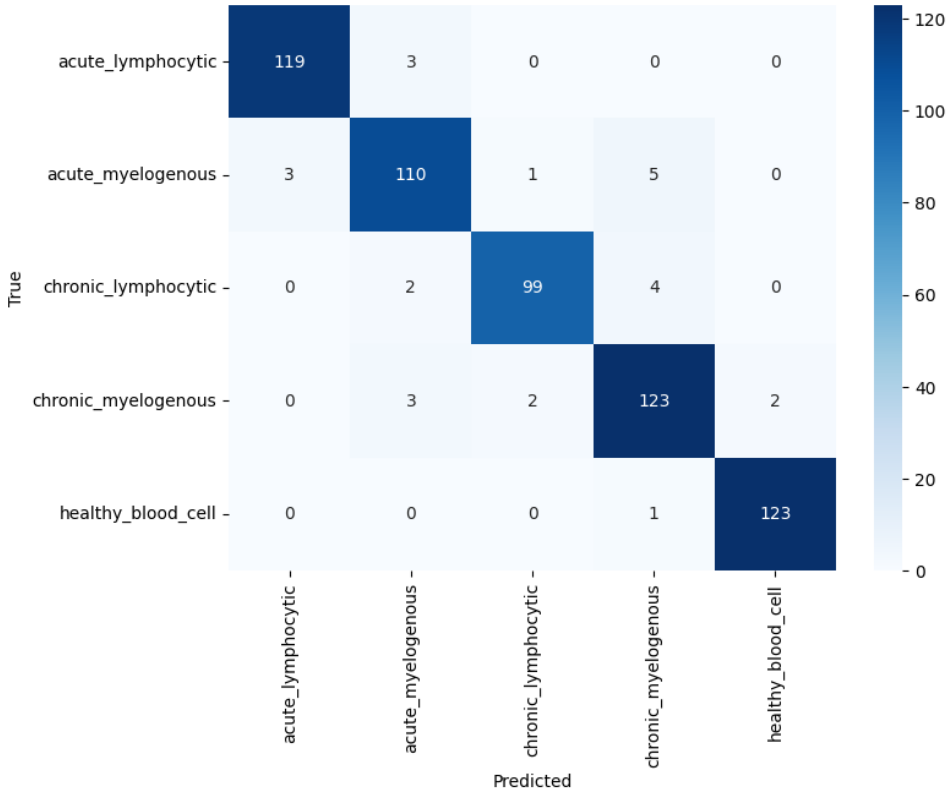


Figure 4.14: Confusion matrix of EfficientNetB7

4.5 Model Accuracy and Model Loss

Model accuracy and model loss are the key metrics used to access the performance of the model during training and evaluation. Model accuracy represents the ratio of correctly classified the classes out of the total classes in the dataset. High accuracy indicates the model is making more correct predictions. Model loss refers to the object function, measures how well the model's prediction matches the actual labels. The focus during training to minimize this loss. Lower loss value indicates the model predictions are closer to the true labels. Minimizing loss is an important aspect of training a CNN to improve its predictive capabilities. During the training process, optimize the model parameter to achieve high accuracy and low loss. To monitor accuracy and loss over epochs to track the model learning process.

4.5.1 ResNet – 50

This graph shows the accuracy and loss over given number of training epochs. Here we took 100 epochs for each model and set patience 15 that help to stop epochs when getting the same value repeatedly. For ResNet-50, 29 epochs were used to measure the accuracy and loss. Achieved 97% model accuracy for accuracy where above 95% valid accuracy. And model validation loss is 0.11%.

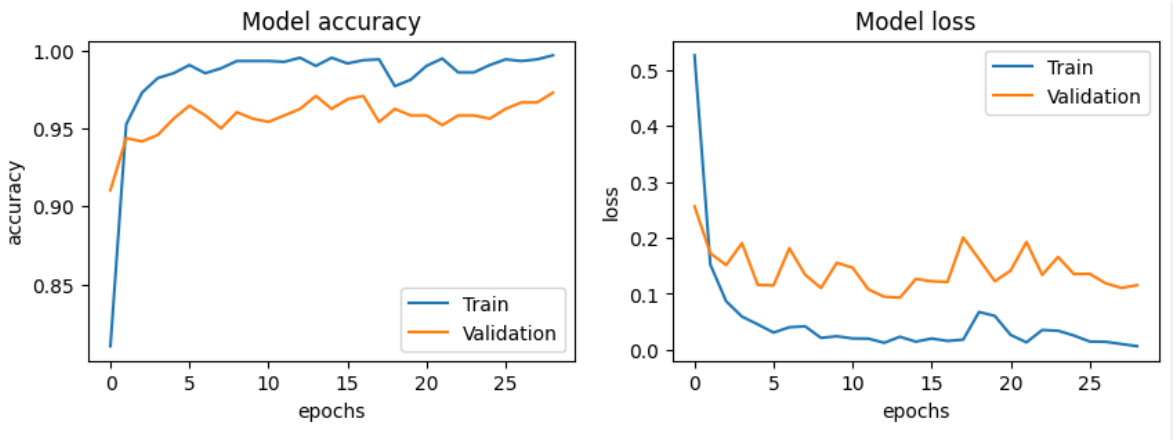


Figure 4.15 Model Accuracy and Model Loss of ResNet–50

4.5.2 DenseNet201

This graph shows the accuracy and loss of DenseNet201 were assessed over 28 epochs, resulting in a 95% accuracy, with more than 93% valid accuracy. And the rest of value is loss data. Here model validation loss is 0.21%.

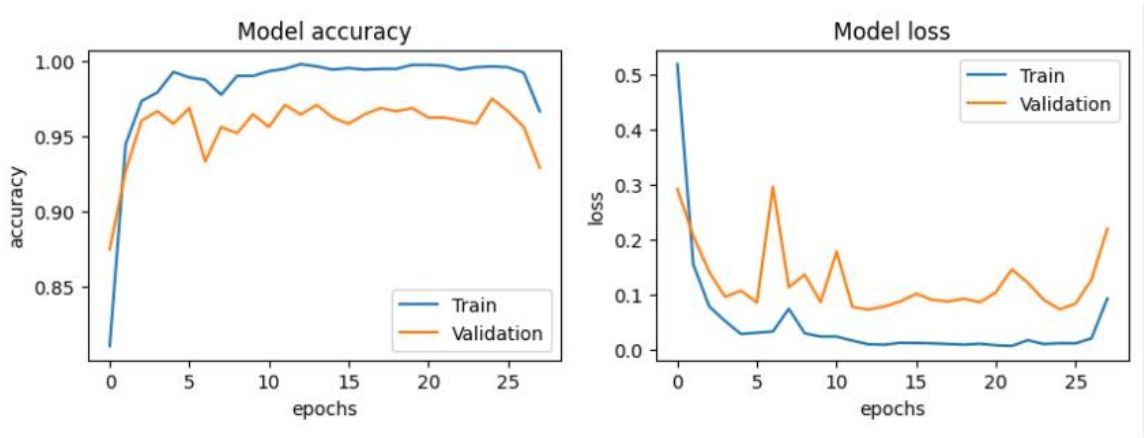


Figure 4.16: Model Accuracy and Model Loss of DenseNet201

4.5.3 MobileNet

MobileNet was trained for 25 epochs, during which its accuracy and loss were evaluated. The model attained a 97% accuracy, with 94% validation accuracy. And model validation loss is 0.14%.

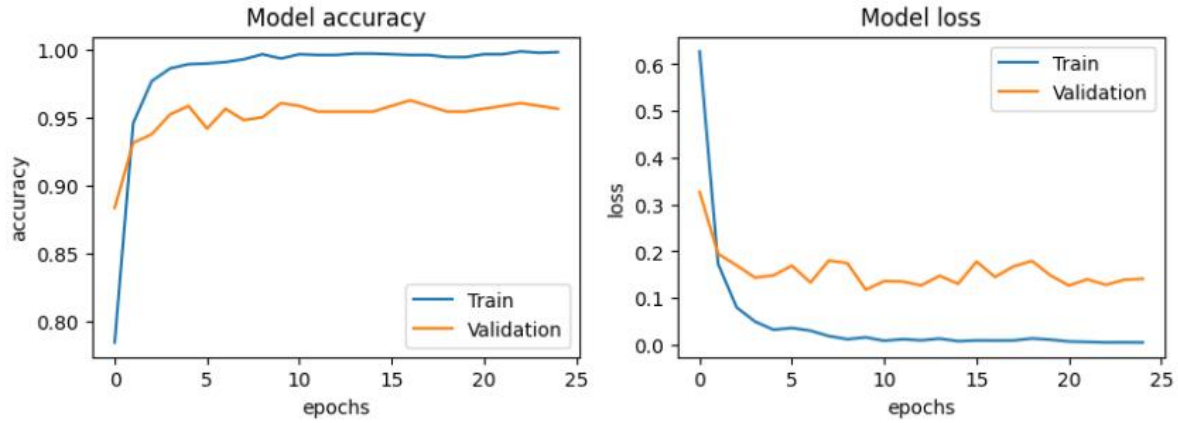


Figure 4.17: Model Accuracy and Model Loss of MobileNet

4.5.4 VGG19

The accuracy and loss of the VGG19 model were evaluated over 24 epochs, resulting in an attained accuracy of 95% on a validation accuracy approximately 92%. Additionally, the model achieved validation loss is 0.2%.

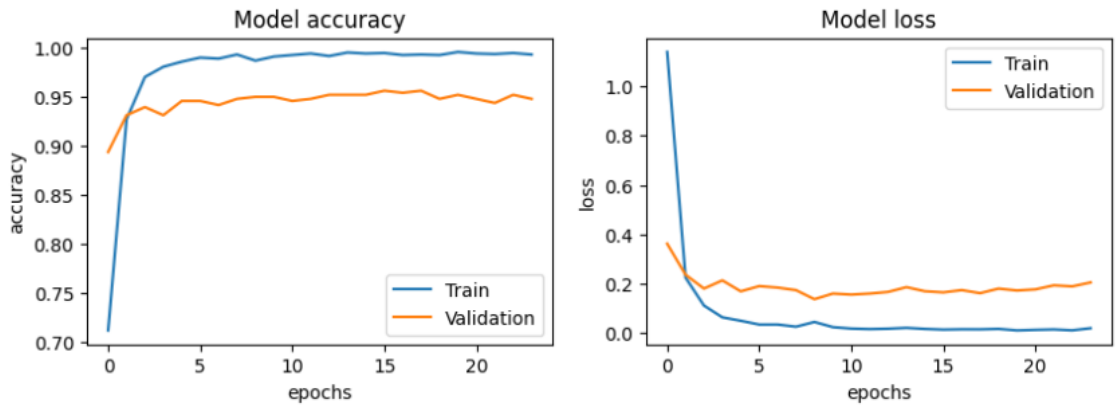


Figure 4.18: Model Accuracy and Model Loss of VGG19

4.5.5 ResNet101

The accuracy and loss of ResNet101 were evaluated over 36 epochs, resulting in a 97% accuracy and validation accuracy, with approximately 92%. Additionally, the model achieved validation loss 0.16%.

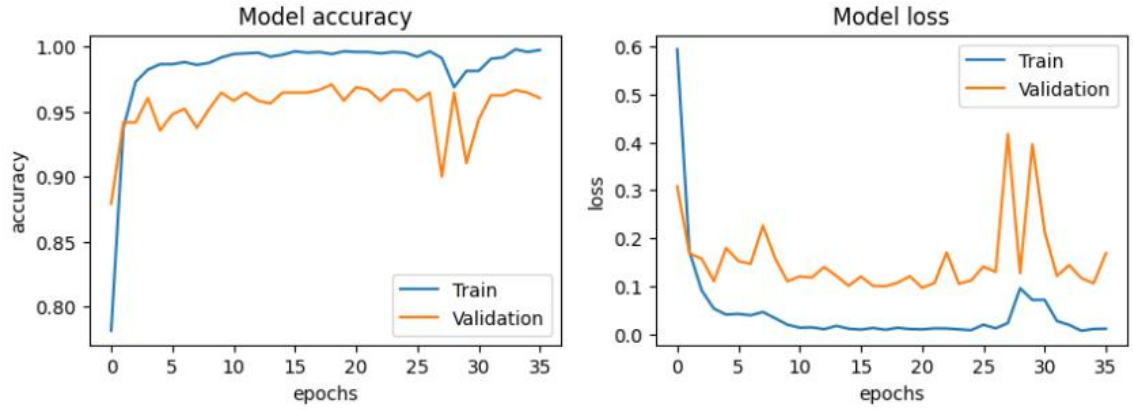


Figure 4.19: Model Accuracy and Model Loss of ResNet101

4.5.6 EfficientNetB7

The EfficientNetB7 model was tested for 50 epochs to see how well it performed. It reached a 96% accuracy rate for the right answers when more than 88% of the data was validated. The model also had a 0.08% loss rate and validation loss is 20%.

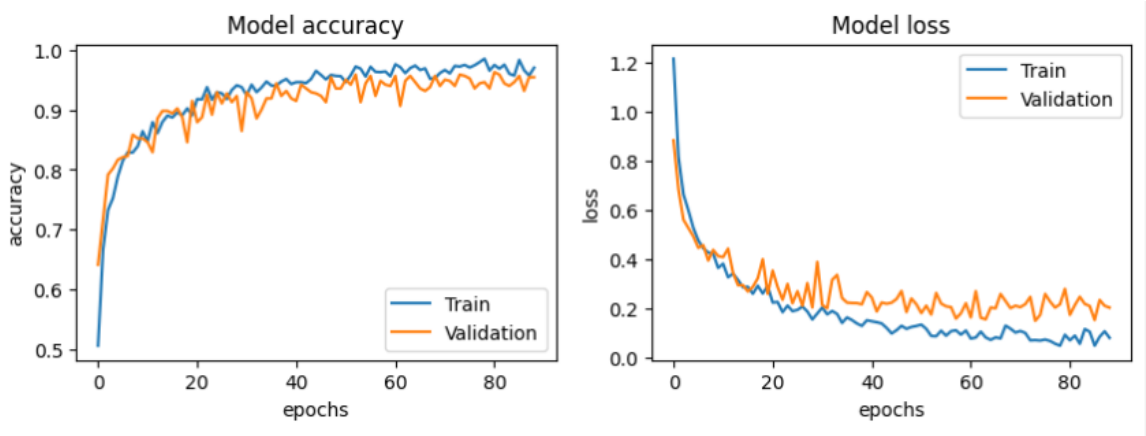


Figure 4.20: Model Accuracy and Model Loss of EfficientNetB7

4.5.7 CML-1

The accuracy and loss of CML-1 were evaluated over 60 epochs and 70 epochs, resulting in a 96% accuracy on the validation data, with approximately 85% of the data considered valid. Additionally, the model validation loss is 16%.

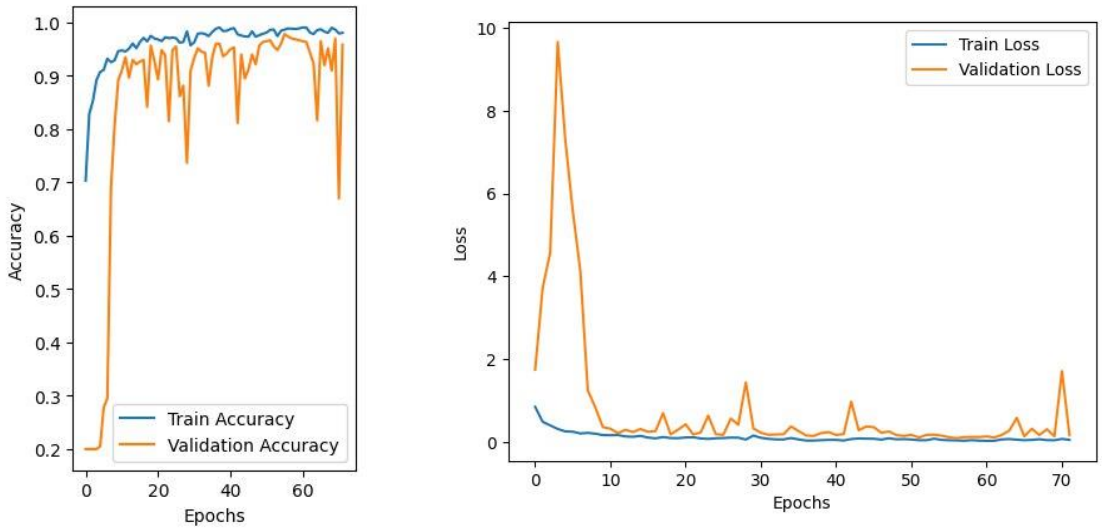


Figure 4.21: Model Accuracy and Model Loss of CML-1

CHAPTER 5

System Design and Development

5.1 Introduction

The project introduced here is an Automated Detection Leukemia Subtype Classification web application developed using Flask, TensorFlow, and Keras. Its primary objective is to provide a user-friendly interface for medical professionals to classify leukemia subtypes based on uploaded images. This innovative approach utilizes CNN models for accurate image classification.

Leukemia, being a critical and time-sensitive medical condition, benefits greatly from automated detection systems. The application allows users to select a specific deep learning model and upload an image, receiving real-time predictions on the leukemia subtype. This technology aids healthcare professionals in making swift and precise decisions, ultimately improving patient outcomes.

5.2 System Design

Flask, serving as the web framework, establishes routes for user interactions, while HTML, CSS, and JavaScript contribute to a user-friendly layout. This combination enhances the overall user experience, providing an intuitive interface for smooth interaction with system.

In the backend, a dictionary stores and loads CNN models, granting users the flexibility to choose a preferred prediction model. Notably, the implementation focuses on incorporating the most accurate convolutional neural network model, ensuring the highest precision in leukemia cancer identification. This strategic integration of design elements and advanced computational capabilities serves as the foundation of the web application. The main aim is to provide users with a dependable identification system, combining a user-friendly design and powerful models for accurate leukemia subtype predictions.

5.3 System Development

System development involves the actual coding and implementation of the proposed design. I chose Python for its flexibility and wide range of tools. The development phase involves implementing the Flask application, integrating CNN models using TensorFlow and Keras, and designing the HTML/CSS interface. The system emphasizes user-friendly interactions, model loading efficiency, and image preprocessing. The development also includes the establishment of routes for rendering the homepage and processing image predictions.

5.4 System Implementation

System implementation involves deploying the application for practical use. The system is implemented by running the Flask application locally or on a server, enabling users to access the system through a web browser. The loaded models are available for predictions, and users can interact with the system through the web interface. The user can select a model, upload an image, and receive prompt predictions for leukemia subtype classification. The implementation phase ensures that the developed system is functional, responsive, and capable of meeting its intended objectives.

5.5 System Testing

System testing is crucial for ensuring the reliability and correctness of the system. The system includes prediction function that processes uploaded images and returns classification results. I assessed the model's functionality within the web framework, examining its capability to accurately identify leukemia cancer types in diverse real-world images. In addition to technical accuracy, I conducted informal tests with individuals representing potential users, gathering valuable insights on the interface's user-friendliness, clarity, and overall usability. This input will inform ongoing enhancements to ensure the application empowers users with varying technical proficiency to identify leukemia cancer types effortlessly and efficiently.

5.6 Conclusion:

Developed Automated Detection Leukemia Subtype Classification system presents a valuable tool for medical professionals. Its utilization of CNN models and a user-friendly web interface empowers healthcare practitioners to make timely and accurate decisions in leukemia subtype classification. The modular design, efficient development, seamless implementation, and rigorous testing collectively contribute to a reliable and impactful solution for medical diagnosis and treatment. Future enhancements may focus on expanding the model repertoire or incorporating advanced image processing techniques to further refine the system's capabilities.

CHAPTER 6

Impact of Social, Environment and Sustainability

6.1 Impact on Social

Understanding social impact is crucial, as decisions, actions, or policies can shape the well-being of people and communities. Positive impact, seen in improved well-being and community development, contrasts with negative consequences like inequality or harm. Research, a powerful catalyst for social change, plays a pivotal role. It addresses pressing issues, informs policies, and advances knowledge, promoting informed decision-making and public awareness.

Leukemia Blood Cancer Detection research has a significant impact on society by improving early detection of leukemia, a form of blood cancer. Early diagnosis and improved treatments not only enhance individual lives but also reducing treatment costs, and easing the emotional burden on patients and their families. Our research's broader influence extends to communities through increased awareness, education, and strengthened support networks, creating a positive ripple effect. It brings hope and positively transforms the lives of those affected by making the diagnostic process more efficient and accessible. By empowering communities, fostering collaboration, and driving innovation, research contributes to shaping a fair and just society. Its multifaceted benefits demonstrate how scientific endeavors, like leukemia research, can serve as a transformative force, offering hope, knowledge, and tangible improvements for individuals and communities alike, ultimately contributing to a better future for society.

6.2 Impact on Environment

Research, like when study things to find new information, can be good and not so good for the environment. On the positive side, it might help us discover better ways to take care of nature and solve environmental problems. But, on the negative side, doing research can use up a lot of resources, create waste, and even harm the environment.

Conducting leukemia blood cancer detection research has environmental impacts primarily through resource utilization, laboratory waste generation, and energy consumption. The production and disposal of lab materials contribute to waste, while energy-intensive equipment and data processing amplify the carbon footprint. Implementing sustainable practices, such as reducing single-use plastics, optimizing energy-efficient technologies, and adopting eco-friendly disposal methods, can mitigate these effects. Striking a balance between medical advancements and environmental responsibility ensures that breakthroughs in leukemia research contribute positively to both human health and the well-being of the planet. We look for eco-friendly ways and think about how to reduce the negative impact on nature while still finding important discoveries. This way, we can keep learning and improving without causing too much harm to our planet.

6.3 Ethical Aspects

Ethical considerations in leukemia blood cancer detection research are paramount, ensuring integrity and respect for participants. We always have to make sure we're doing the right thing and treating people well. Before we start, we need to ask people if they're okay with being part of our study, and we have to keep what they tell us private. We also have to be honest about any possible good or bad things that might happen and stay focused on the facts when we look at our data. For leukemia blood cancer detection research, these rules are even more crucial. We have to keep everything about the people in our study private and make sure they're okay with being part of it. We need to be careful to not cause them any harm and give them the help they need. Talking openly about why we're doing the research and what might happen helps everyone trust us. Treating everyone fairly and being open about any conflicts or where our money comes from keeps things honest. Sharing our results truthfully helps everyone learn more about science. Following these rules makes sure we do our leukemia research in a responsible and kind way, earning the trust of the public and moving medical knowledge forward while respecting everyone.

6.4 Sustainability plan

Developing a sustainability plan for leukemia blood cancer detection research involves comprehensive strategies to minimize environmental impact. First, prioritize the responsible sourcing of materials, opting for suppliers committed to eco-friendly practices. Implement a waste reduction system by promoting the reuse of lab materials and recycling whenever possible. Minimize energy consumption by utilizing energy-efficient equipment and optimizing data processing procedures.

Incorporate green laboratory practices, such as eco-friendly solvents and energy-saving technologies, to reduce the overall carbon footprint. Encourage a culture of sustainability among research personnel through training programs and awareness campaigns. Foster collaboration with institutions that share a commitment to environmental responsibility, promoting the exchange of sustainable research practices.

Embrace digitalization for data storage and communication, reducing paper usage and streamlining processes. Periodically assess the environmental impact of research activities and adjust strategies accordingly. Consider the life cycle of products and technologies used in the research, opting for those with minimal ecological footprint.

Engage with the community to share insights into sustainable research practices and foster partnerships that promote environmental stewardship. By integrating these sustainability measures into the leukemia blood cancer detection research, not only contribute to scientific advancements but also exemplify a commitment to the long-term well-being of both humanity and the planet.

CHAPTER 7

Conclusion and Future Work

7.1 Conclusion

An important progress in medical diagnostics is the automated leukemia subtype classification using a Convolutional Neural Network (CNN) approach. This study indicates that using CNNs to accurately identify and classify leukemia subtypes from microscopic images is both affordable and effective, due to the combining of modern technology and machine learning.

The CNN models, including ResNet-50, DenseNet20, MobileNet, VGG19, ResNet101, EfficientNetB7 and CML-1, have developed approvable performance that handling the complexities of leukemia classification. The utilization of different augmentation techniques, such as noise removal, zooming, translating, shearing, and rotating, has further enhanced the images quality and generalization capabilities of the models. In this article, seven (ResNet-50, DenseNet20, MobileNet, VGG19, ResNet101, EfficientNetB7, CML-1) potent models has good classification accuracy, but the ResNet-50 model performed best from the others. The model accuracy is 97%.

The achieved results show the promising accuracy rates for the integration of artificial intelligence in leukemia diagnostics. The technology continues to advance, this automated approach holds promise for reducing diagnosis time, minimizing human error, and ultimately improving patient outcomes.

7.2 Future work

In the future we aim to improve our automated leukemia subtype classification system. We'll explore-

- The integration of multi-model data such as the genetic and molecular information to enhance the overall understanding of leukemia subtype.
- To adapt pre-trained models for specific leukemia dataset can be used transfer learning that improve classification performance.

- Flexibility of real-time implementation for immediate clinic applications, accurate leukemia subtype classification can be ensured during routine treatment procedures.
- In this research, we mostly use 2D pictures to find leukemia cancer cells. We can also include ways to detect these cells using 3D images.
- By confirming the generalizability and robustness of the model, large-scale clinical trials can be conducted to validate the proposed CNN method across different patient populations.

Also the CNN models which are designed for our research can be used for even more applications in the future. We will plan to make it work for a variety of diseases and improve the program so that it can be used for a wide range of medical conditions.

References

[1] Huang, F., Guang, P., Li, F., Liu, X., Zhang, W., & Huang, W. (2020). AML, ALL, and CML classification and diagnosis based on bone marrow cell morphology combined with convolutional neural network: A STARD compliant diagnosis research. Medicine, 99(45).

[2] Ahmed, N.; Yigit, A.; Isik, Z.; Alpkocak, A. Identification of Leukemia Subtypes from Microscopic Images Using Convolutional Neural Network. Diagnostics 2019, 9, 104. https://doi.org/10.3390/diagnostics9030104

[3] Thanh, T. T. P., Vununu, C., Atoev, S., Lee, S. H., & Kwon, K. R. (2018). Leukemia blood cell image classification using convolutional neural network. International journal of computer theory and engineering, 10(2), 54-58

[4] Bibi, N., Sikandar, M., Ud Din, I., Almogren, A., & Ali, S. (2020). IoMT-based automated detection and classification of leukemia using deep learning. Journal of healthcare engineering, 2020, 1-12.

[5] Jagadev, P., & Virani, H. G. (2017, May). Detection of leukemia and its types using image processing and machine learning. In 2017 International Conference on Trends in Electronics and Informatics (ICEI) (pp. 522-526). IEEE.

[6] Tran, T., Park, J. H., Kwon, O. H., Moon, K. S., Lee, S. H., & Kwon, K. R. (2018). Classification of leukemia disease in peripheral blood cell images using convolutional neural network. Journal of Korea Multimedia Society, 21(10), 1150-1161.

[7] Sahlol, A. T., Kollmannsberger, P., & Ewees, A. A. (2020). Efficient classification of white blood cell leukemia with improved swarm optimization of deep features. Scientific reports, 10(1), 2536.

[8] Sakthiraj, F. S. K. (2022). Autonomous leukemia detection scheme based on hybrid convolutional neural network model using learning algorithm. Wireless Personal Communications, 126(3), 2191-2206.

[9] Patel, N., & Mishra, A. (2015). Automated leukaemia detection using microscopic images. Procedia Computer Science, 58, 635-642.

[10] S. Mohapatra, D. Patra and S. Satpathi, "Image analysis of blood microscopic images for acute leukemia detection," 2010 International Conference on Industrial Electronics, Control and Robotics, Rourkela, India, 2010, pp. 215-219, doi: 10.1109/IECR.2010.5720171.

[11] Himali P. Vaghela, Hardik Modi, Manoj Pandya and M.B. Potdar Leukaemia Detection using Digital Image Processing Techniques", November 2015 - International Journal of Applied Information Systems (IJAIS)

[12] Gondal, C. H. A. (2020). An Automated Leukemia Screening and Subtypes Classification Using Microscopic Image Analysis.

[13] Vogado, L. H., Veras, R. M., Araujo, F. H., Silva, R. R., & Aires, K. R. (2018). Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification. Engineering Applications of Artificial Intelligence, 72, 415-422.

[14] Rehman, A., Abbas, N., Saba, T., Rahman, S. I. U., Mehmood, Z., & Kolivand, H. (2018). Classification of acute lymphoblastic leukemia using deep learning. Microscopy Research and Technique, 81(11), 1310-1317.

[15] Mohapatra, S., Samanta, S. S., Patra, D., & Satpathi, S. (2011, February). Fuzzy based blood image segmentation for automated leukemia detection. In 2011 International Conference on Devices and Communications (ICDeCom) (pp. 1-5). IEEE.

[16] Mohammed, Z. M., & Mohammed, E. Z. (2023). Leukemia Blood Cell Image Classification Using Machine Learning-A Systematic Literature Review. Journal of Advanced Sciences and Nanotechnology, 2(2), 204-214.

[17] Singhal, V., & Singh, P. (2016). Texture features for the detection of acute lymphoblastic leukemia. In Proceedings of International Conference on ICT for Sustainable Development: ICT4SD 2015 Volume 2 (pp. 535-543). Springer Singapore.

[18] Kumar R, Srivastava R, Srivastava S. Detection and Classification of Cancer from Microscopic Biopsy Images Using Clinically Significant and Biologically Interpretable Features. J Med Eng. 2015;2015:457906. doi: 10.1155/2015/457906. Epub 2015 Aug 23. PMID: 27006938; PMCID: PMC4782618.

[19] Emad A. Mohammed, Mostafa M. A. Mohamed, Christopher Naugler and Behrouz. H.Far, Chronic lymphocytic leukaemia cell segmentation from microscopic blood images using watershed algorithm and optimal thresholding", 2013 26th IEEE Canadian Conference of Electrical and Computer

[20] Panda, A., Pachori, R. B., & Sinnappah-Kang, N. D. (2021). Classification of chronic myeloid leukemia neutrophils by hyperspectral imaging using Euclidean and Mahalanobis distances. Biomedical Signal Processing and Control, 70, 103025.

[21] Ciresan, D.; Meier, U.; Masci, J.; Gambardella, L.; Schmidhuber, J. High-performance neural networks for visual object classification. arXiv 2011, arXiv:1102.0183

[22] Wong, S.C.; Gatt, A.; Stamatescu, V.; McDonnell, M.D. Understanding data augmentation for classification: When to warp? In Proceedings of the 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Gold Coast, Australia, 30 November–2 December 2016.

[23] Perez, L.; Wang, J. The effectiveness of data augmentation in image classification using deep learning. arXiv 2017, arXiv:1712.04621.

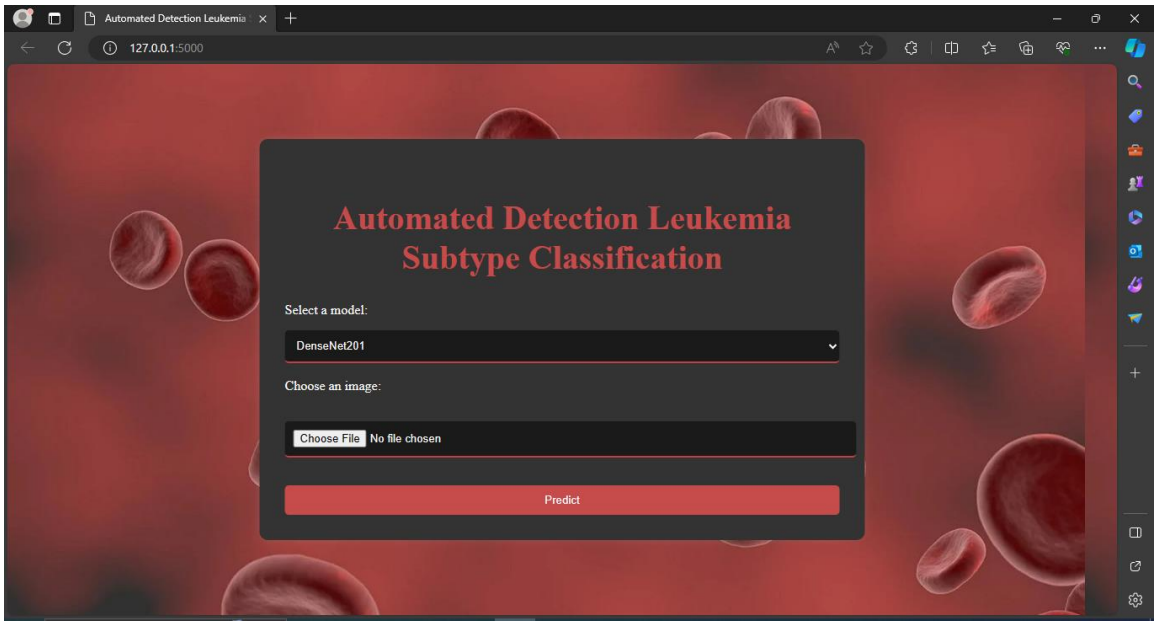
[24] Mikołajczyk, A.; Grochowski, M. Data augmentation for improving deep learning in image classification problem. In Proceedings of the International Interdisciplinary PhD Workshop (IIPhDW), Swinouj´scie, Poland, 9–12 May 2018; pp. 117–122.

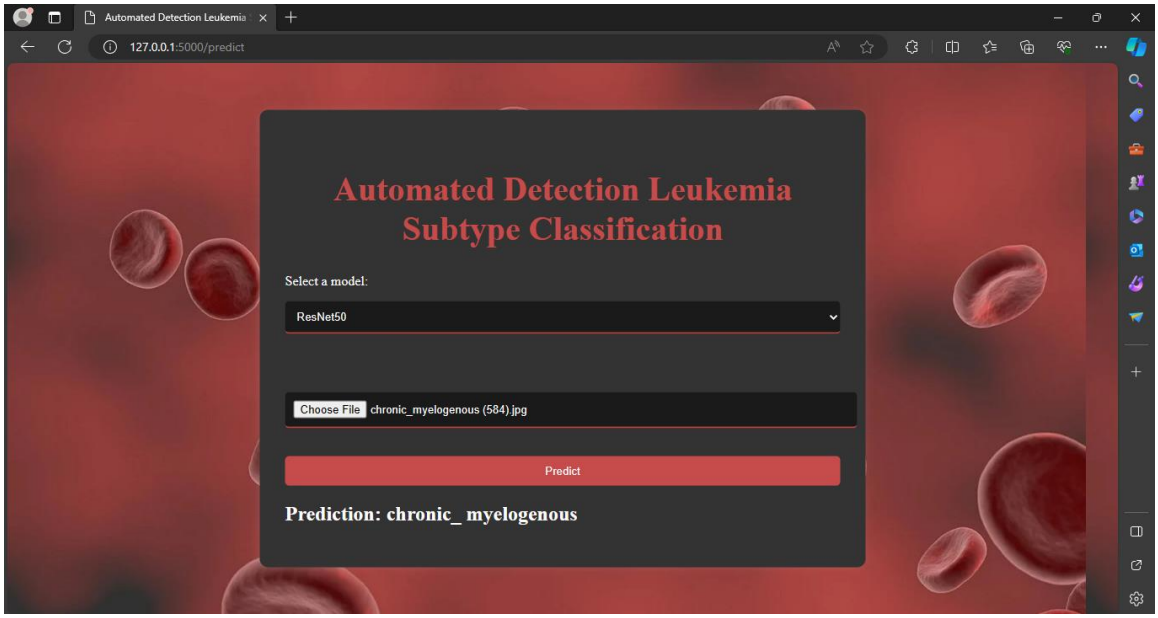
[25] ResearchGate, available at<<Illustration of the structure of fully-connected layers and softmax layer. | Download Scientific Diagram (researchgate.net) >> last accessed on 17-12-23

[26] Medium, available at<<The Annotated ResNet-50. Explaining how ResNet-50 works and why… | by Suvaditya Mukherjee | Towards Data Science >> last accessed on 17-12-23

Appendix

Web application developed for Automated Detection Leukemia Subtype Classification. Here is the screenshot of website-





Turnitin Originality Report

Processed on: 19-Jan-2024 17:26 +06
ID: 2273835251
Word Count: 9456
Submitted: 1

Automated Detection Leukemia Subtype Classification from Microscopic Images: A Convolutional Neural Network Approach By Sheikh Bodrun Nesa Lubna 3590

Similarity Index	Similarity by Source	
15%	Internet Sources:	12%
	Publications:	7%
	Student Papers:	8%

2% match (student papers from 10-Jul-2023)
Submitted to Daffodil International University on 2023-07-10

1% match (Internet from 05-Dec-2022)
https://www.mdpi.com/2075-4418/9/3/104/htm

1% match (Internet from 23-Jul-2023)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/9943/22579.pdf?isAllowed=y&sequence=1

< 1% match (student papers from 10-Jul-2023)
Submitted to Daffodil International University on 2023-07-10

< 1% match (Internet from 11-Jan-2023)
https://www.mdpi.com/1424-8220/22/6/2348/xml

< 1% match (Internet from 22-Dec-2023)
https://www.mdpi.com/2078-2489/14/12/663

< 1% match (Internet from 25-Oct-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5813/171-15-1182%2018%25.pdf?isAllowed=y&sequence=1

< 1% match (Internet from 16-Jan-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/6748/172-15-9716%20%2823%25%29%20clearence.pdf?isAllowed=y&sequence=1

< 1% match (Internet from 22-Sep-2023)
https://www.egekongreleri.org/_files/ugd/614b1f_3d13283380574393978c15d435073d1d.pdf

< 1% match (student papers from 25-May-2023)
Submitted to University of Queensland on 2023-05-25

< 1% match (Internet from 08-Mar-2023)
https://dokumen.pub/deep-learning-with-theano-build-the-artificial-brain-of-the-future-today-9781786465825-1786465825-9781786463050-1786463059.html

< 1% match (Internet from 17-Dec-2023)
https://dokumen.pub/handbook-of-research-on-deep-learning-techniques-for-cloud-based-industrial-iot-team-ira-1668480980-9781668480984.html

< 1% match (Internet from 12-Jan-2023)
https://dokumen.pub/internet-of-things-infrastructures-and-mobile-applications-proceedings-of-the-13th-imcl-conference-1st-ed-9783030499310-9783030499327.html

< 1% match (Internet from 09-Dec-2023)
https://link.springer.com/chapter/10.1007/978-3-031-45214-7_6?code=a0a9c383-8dd9-455c-9001-fc15c3dea0d9&error=cookies_not_supported