VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belagavi-560 014, Karnataka

A Mini Project Report On

"BLOOD CANCER DETECTION"

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE COMPUTER GRAPHICS AND IMAGE PROCESSING LABORATORY [21CSL66]

OF

BACHELOR OF ENGINEERING

IN

Department of Computer Science and Engineering

Submitted By

1. NAGAPOOJA M [1SV21CS049]

2. LAXMI SONNAD [1SV21CS040]

Under the Guidance of

Prof. Shanmukaswamy C.V. B.E.,M.E.,MISTE.

Associate Professor, Dept of CSE, SIET, Tumakuru

and

Mrs. Rashmi N B.E., M.Tech.

Assistant Professor, Dept of CSE, SIET, Tumakuru

Department of Computer Science and Engineering

SHRIDEVI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Affiliated to Visvesvaraya Technological University)

Sira Road, Tumakuru–572 106, Karnataka.

2023-2024

Sri Shridevi Charitable Trust (R.)
SHRIDEVI INSTITUTE OF ENGINEERING AND TECHNOLOGY
Sira Road, Tumkur - 572 106, Karnataka, India.
Phone: 0816 - 2212629 | Principal: 0816 - 2212627, 9686114899 | Telefax: 0816 - 2212628
Email: info@shridevienginеering.org, principal@shridevienginеering.org | Website: www.shridevienginеering.org
(Approved by AICTE, New Delhi, Recognised by Govt. of Karnataka and Affiliated to Visvesvaraya Technological University, Belagavi)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the mini-project entitled "**BLOOD CANCER DETECTION**" has been successfully carried out by 1..NAGAPOOJA M [1SV21CS049], 2. LAXMI SONNAD [1SV21CS040] in partial fulfillment for the **Computer Graphics and Image Processing Laboratory [21CSL66]** Mini Project of **Bachelor of Engineering in** the **Department of Computer Science and Engineering** of the **Visvesvaraya Technological University, Belagavi** during the Academic year **2023-24.** It is certified that all the corrections/suggestions indicated for internal assessments have been incorporated into the report. The Mini Project Report has been approved as it satisfies the academic requirements in respect of the Subject of the Bachelor of Engineering Degree.

Signature of the Guide
Mrs. Rashmi N B. E, M. Tech,
Assistant Professor, Dept of CSE,
SIET, Tumakuru

Signature of the Guide
Prof. Shanmukaswamy C V B. E, M. E, MISTE
Associate Professor, Dept of CSE,
SIET, Tumakuru

Signature of the H.O.D
Dr. Basavesha D B.E., M.Tech., Ph. D., MISTE.
Associate Professor and HOD, Dept of CSE, SIET, Tumakuru.

Name of the Examiners	Signature with Date
1._____	_____
2._____	_____

Sri Shridevi Charitable Trust (R.)
SHRIDEVI INSTITUTE OF ENGINEERING AND TECHNOLOGY
Sira Road, Tumkur - 572 106, Karnataka, India.
Phone: 0816 - 2212629 | Principal: 0816 - 2212627, 9686114899 | Telefax: 0816 - 2212628
Email: info@shrideviengineering.org, principal@shrideviengineering.org | Website: www.shrideviengineering.org
(Approved by AICTE, New Delhi, Recognised by Govt. of Karnataka and Affiliated to Visvesvaraya Technological University, Belagavi)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DECLARATION

We **NAGAPOOJA M [1SV21CS049] and LAXMI SONNAD [1SV21CS040]**, students of VI semester **B.E** in Computer Science & Engineering , at Shridevi Institute of Engineering & Technology, Tumakuru, hereby declare that, the Mini Project work entitled **"BLOOD CANCER DETECTION"**, embodies the report of our Mini-Project work carried out under the guidance of Prof. Shanmukaswamy C. V., Associate Professor, Dept. of CSE, and Mrs. Rashmi N, Assistant Professor, Department of CSE, SIET as partial fulfillment of requirements for the **Computer Graphics and Image Processing Laboratory [21CSL66]** mini project of Bachelor of Engineering in Computer Science & Engineering of Visvesvaraya Technological University, Belagavi, during the academic year **2023-24**. The Mini Project has been approved as it satisfies the academic requirements in respect to the Mini Project work.

Place: Tumakuru

Date:

Students' Name and Signature

NAGAPOOJA M [1SV21CS049]

LAXMI SONNAD [1SV21CS040]

ACKNOWLEDGEMENT

This Mini Project would be incomplete without thanking the personalities responsible for this venture, which otherwise would not have become a reality.

We would like to thank our guides **Prof. Shanmukaswamy C. V., Associate Professor** and **Mrs. Rashmi N, Assistant Professor,** Computer Science & Engineering, SIET for their support, sharing their technical expertise timely advice invaluable guidance and assistance throughout this mini-project.

We would like to thank the Head of the Department **Dr. Basavesha D, Associate Professor and Head, Department of Computer Science and Engineering, SIET** for providing all the support and facility.

We express our profound gratitude to **Dr. Narendra Viswanath, Principal**, **SIET**, for his moral support towards completing our mini-project work.

We would like to express our sincere gratitude to all teaching and non-teaching faculty of the Department of CSE for guiding us in this mini-project by giving valuable suggestions and encouragement.

NAGAPOOJA M [1SV21CS049]

LAXMI SONNAD [1SV21CS040]

ABSTRACT

Blood cancer, a life-threatening condition characterized by the abnormal growth of blood cells, requires early and accurate diagnosis for effective treatment. This project explores the application of Convolutional Neural Networks (CNNs) in detecting blood cancer from microscopic images of blood samples. CNNs, known for their proficiency in image recognition tasks, are employed to classify images into categories such as normal, acute lymphoblastic leukemia (ALL), and acute myeloid leukemia (AML).

The dataset comprises labeled images of blood samples, preprocessed to enhance the quality and relevance of the data. The CNN architecture is designed to extract essential features from these images, leveraging layers like convolutional, pooling, and fully connected layers. Techniques such as data augmentation and dropout are applied to improve the model's generalization and prevent overfitting.

The proposed model's performance is evaluated using metrics such as accuracy, precision, recall, and F1-score, demonstrating the CNN's capability to identify blood cancer with high accuracy. This study highlights the potential of deep learning in medical image analysis, offering a non-invasive and efficient method for early detection of blood cancers, thereby contributing to timely diagnosis and treatment planning.

Table of Contents

List of Figures

CHAPTER 1

INTRODUCTION

Cancer is a cluster of cells undergoing unchecked growth in the body, and it can quickly spread to any organ. Cancer comes in various forms; the most common are breast cancer, lung cancer, skin cancer, and blood cancers like leukemia and lymphoma. There have been 9.2 million fatalities from lung cancer, 1.7 million from skin cancer, and 627,000 from breast cancer , according to reports from the World Health Organization (WHO) .When it comes to cancers, leukemia has a remarkably high mortality rate. It's a malignant tumor that forms in the bone marrow when immature white blood cells are cloned in a destructive way. With lung, colon, breast, and prostate cancers, leukemia is among the most frequently diagnosed cancers in the United States. According to projections made by the US government's cancer data collector, the Surveillance, Epidemiology, and End Results (SEER) Program, there were 60,650 newly diagnosed cases of leukemia, and 24,000 death occurred in the US in 2022. According to a review of the cancer database by the WHO, leukemia incidence varies significantly by region and subtype. More than 20,000 cases of pediatric blood cancer are detected annually in India, with approximately 15,000 cases of leukemia only .Around 61,780 instances of leukemia were diagnosed in the United States in 2019, with another 9900 cases being found in the United Kingdom. From 345,000 in 1990 to 518,000 in 2018, the number of newly diagnosed cases of leukemia increased, lowering the Annualized Survival Insusceptibility Rate (ASIR) by 0.43% per year. To achieve the optimum results in multi-class leukemia classification, the proposed research has applied the nature inspired algorithms to find the best features from the extracted features and apply the ML based classifiers and interprets the calculated experimental data accordingly.

CHAPTER 2

LITERATURE SURVEY

Many machine learning and deep learning techniques were used to identify or classify the ALL (acute lymphoblastic leukemia) type. Some of the previous papers are described in this section.

Researchers offer a novel Bayesian-based optimized CNN method for identifying ALL in microscopic smeared images in study. A hybrid dataset was formed to be used in this study by combining two subsets of ALL-IDB datasets (ALL-IDB1 and ALL-IDB2). The hybrid dataset consists of 368 blood smear photos. In the test set, the optimized CNN model for ALL identification was found using the Bayesian optimization technique, and it achieved maximum accuracy of 100%.

The article suggested an approach of convolutional neural network called SK U-Net to perform the task of nucleus segmentation for ALL. All 198 input photos come from the publicly available database ALL-IDB2. The SK U-Net achieved a higher Dice score of 0.916 than the traditional U-Net, which only achieved a score of 0.320. A 98% accuracy is achieved with SVM, which is significantly higher than other methods. The proposed method has a higher accuracy of 0.97% than prior methods. Additionally, KNN and SVM achieved an accuracy of 0.85% and 0.98%, respectively.

The latest developments in ALL detection and categorization using deep and machine learning are presented in study through a systematic review. This article thoroughly examines the advantages and disadvantages of many different AI-based ALL detection methods. Lastly, a range of tough topics and potential future scopes are presented, which may inspire readers to develop their own research questions in ALL areas.

CHAPTER 3

SYSTEM ANALYSIS

Detecting blood cancer using image analysis with Convolutional Neural Networks (CNNs) is a cutting-edge approach in medical diagnostics. Here's a detailed analysis of how such a system can be designed and analyzed

3.1 Existing System

3.1.1 Leukemia Detection:

Leukocyte Image Analysis: CNNs are used to analyze images of blood smears to identify and classify leukocytes (white blood cells). Systems like DeepLeuk and LeukemiaNet focus on detecting abnormal cells indicative of leukemia.

Example Study: A 2019 study used a custom CNN architecture to classify leukemia from microscopic images of blood smears with high accuracy. The model achieved an accuracy of around 95%, outperforming traditional methods.

3.1.2 Automated Blood Cell Classification:

Systems: Tools like CellProfiler and ImageJ integrate with machine learning algorithms for automated classification and quantification of blood cells. While these tools often use traditional image processing techniques, they are increasingly incorporating CNNs for enhanced accuracy.

Example System: Blood Cell Classifier uses a CNN model to classify blood cells into normal and abnormal categories. It has shown potential in distinguishing between different types of leukemia.

3.2 Proposed System

3.2.1 Image Acquisition:

Sources: Collect high-resolution images of blood smears from medical institutions or public datasets like the Blood Cell Count and Classification Dataset or the ALL-IDB (Acute Lymphoblastic Leukemia Image Database).

Annotation: Ensure images are labeled by medical professionals indicating the presence or type of cancer (e.g., leukemia).

3.2.2 Diversity and Quality:

Variety: Include images representing different types of blood cancers (e.g., Acute Lymphoblastic Leukemia, Acute Myeloid Leukemia) and non-cancerous conditions.

Quality Control: Maintain high image quality and consistency in preparation.

3.2.3 Image Preparation:

Normalization: Standardize image size and pixel values. Resize images to a consistent dimension (e.g., 256x256 pixels).

Augmentation: Use techniques like rotation, flipping, and color jitter to enhance dataset diversity and robustness.

CHAPTER 4

SYSTEM METHODOLOGY

4.1 System Architecture

In this section, methodology of the research is described. The section is classified into four interconnected subsections such as the research dataset, feature extraction with pre-trained Convolutional Neural Network (CNN) models, the extraction of the feature vectors and classification with the conventional Machine Learning (ML) classifiers. Fig. 4.1 shows the overall proposed system illustration with existing components.

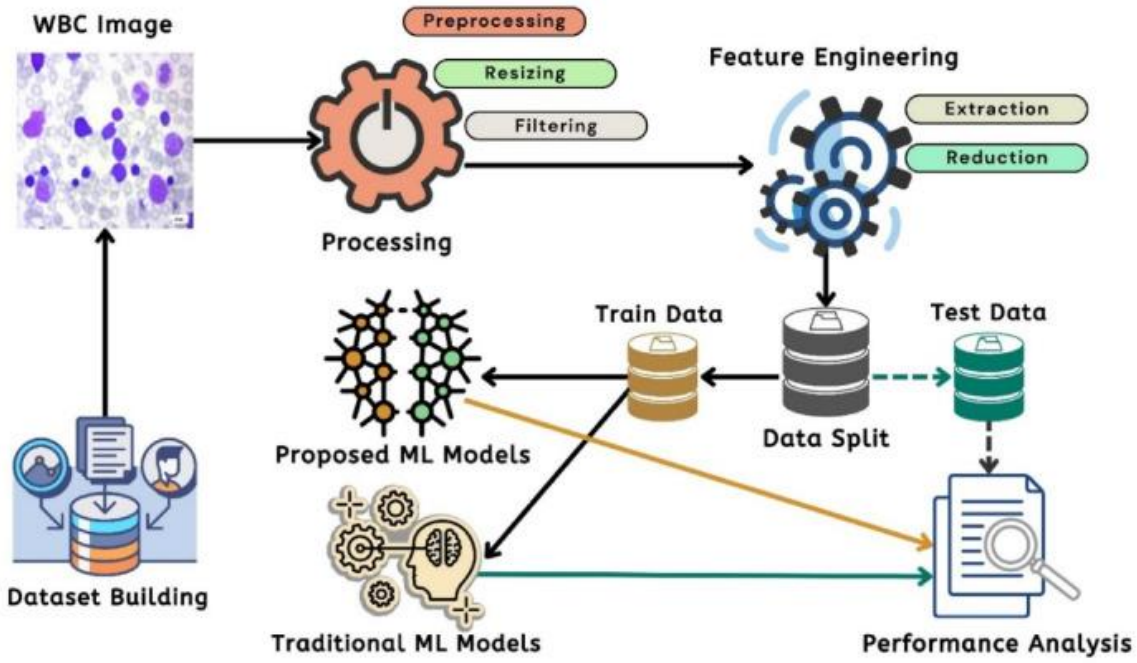


Fig 4.1 : system illustration of the proposed system.

4.1.1. Dataset

This research has been taken the dataset from the secondary sources more specifically from Kaggle .The dataset is comprised with 3262 images of actual peripheral blood smear images. The images were included from the 89 patients where 25 patients were suspected as healthy individuals and rest of the 64 patients were suspected as Acute Lymphoblastic Leukemia (ALL). The dataset is classified into two identical classes such as Benign and malignant categories and further reshaped the dataset into four significant classes with three subtype of malignants namely, Benign, Early Pre-B, Pre-B and Pro-B. All of the images were captured

with a Zeiss camera in a microscope at 100× magnification and stored the images as JPG format in the storage. The types and subtypes of these images were carefully and conclusively determined by a specialist using flow cytometry.fig 4.2 shows the corresponding sample images for this research.

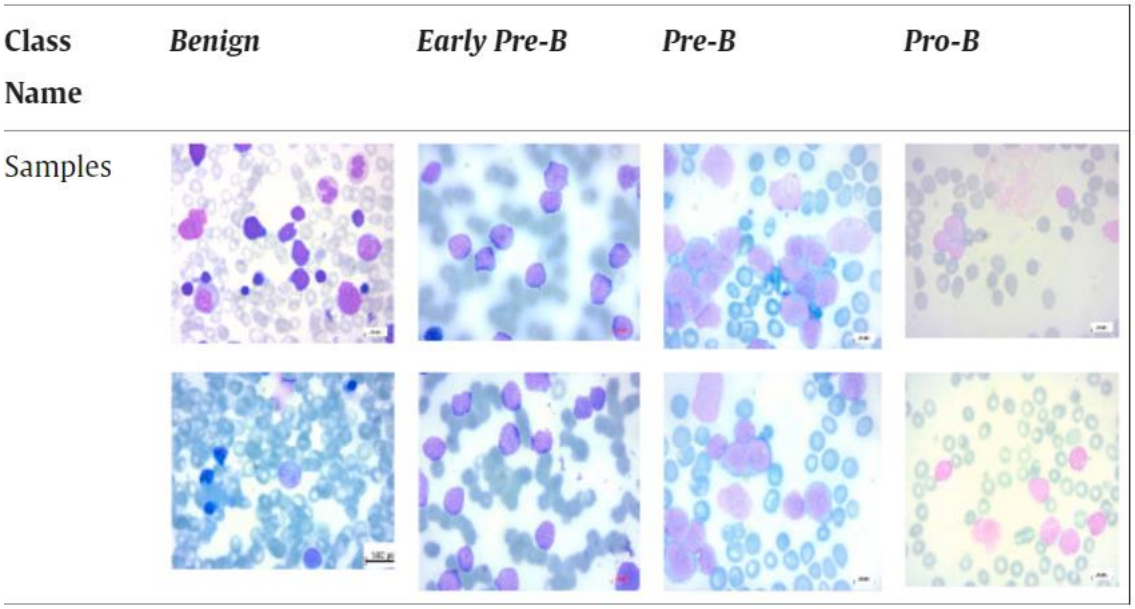


Fig 4.1.1: Sample images of each classes of the dataset.

4.2 System Methodology

4.2.1. Feature extraction

In this sub-section, a mechanism of feature extraction is described. Firstly, the pipeline of feature extraction will be presented with algorithmic annotations. Then, the mechanism of feature selection will be illustrated. After, this subsection will present the working principles of Particle Swarm Optimization (PSO) and Cat Swarm Optimization (CSO) along with the algorithmic annotations and interpretations. This subsection will present the feature vectors and the working mechanism of conventional classifiers.

4.2.2. Feature Extraction with Pre-trained CNN

In this research, four conventional pre-trained Convolutional Neural Network (CNN) models have been applied to extract the features from the single images. shows the corresponding diagram of respective pipeline of feature extraction mechanism of this study. In this figure, initially, the system extracts the images from the dataset to feed the pre-trained models to extract the images. Four traditional pre-trained CNN architecture namely, VGG19, ResNet50, InceptionV3 and Xception have been applied sequentially to extract feature vectors from the images. After extracting the features, seven conventional classifiers have been implemented to classify the images. Due to work with best features, two nature inspired algorithms have been implemented and performed the feature selection method. After that the system measures the performance based on their significant classes. The whole pipeline follows the Algorithm 01 to extract the feature vectors from a particular image.

4.2.3 Algorithm Working mechanism of proposed pipeline to extract feature vectors

The Algorithm shows efficient view of feature extraction from a particular image. In this procedure, the system first initialized the dataset based on the number of images. Where X represent the input image and Y is the output after applying the image filtering and resizing. Then, the pseudocode represents a loop structure to enumerate the feature vectors.

The first part denotes the feature extraction part from the input layer to the last max-pooling layer. The second part represents the residual network of the model, which is mainly responsible for the classification. The proposed solution mainly focuses the VGG19 model on feature extraction; thus, the classification part is declined in this study. The proposed model with VGG19 accepts the Blood cell images of $224 \times 224 \times 3$ and assembles 4096 features from the out of the last layer of feature extraction part for each image In this architecture, the pre-trained model is comprised of a series of Convolutional Layers (CL) and single or multiple Fully Connected (FC) layers. The model is identically classified into two interconnected parts.

Input: ***2D Images***

Output: **Feature Vectors**

Initialization :

1. $n = 2N-1$, *Where* $N = 1, 2, 3, 4 \ldots \ldots \ldots \ldots n$
2. $X \leftarrow$ *Input Image*
3. $Y_n \leftarrow$ *Apply the median filter on the input image X using the karnel size* $n \times n$
4. $F_v \leftarrow$ *Respective Feature Vector*

Start :

1. ***for*** *each N:*
2. *Find* Y_n
3. *Use* (X, Y_n) *to get* $F_n \mid F_n \{P_0, P_1, \ldots, P_{14}\}$
4. $F_v \leftarrow F_n$
5. ***End for***
6. *Show* F_v

4.3 Implementation

When working with Convolutional Neural Networks (CNNs) for image classification, there are several built-in functions and utilities provided by popular libraries such as TensorFlow/Keras and PyTorch. These functions streamline the process of building, training, evaluating, and deploying models. Below, I'll outline key built-in functions and utilities from TensorFlow/Keras and PyTorch that are useful for blood cancer image detection or similar tasks**.**

TensorFlow/Keras Built-in Functions

Data Preparation and Augmentation

tf.data.Dataset: For handling large datasets efficiently**.**

```
dataset = tf.data.Dataset.from_tensor_slices((images, labels))

dataset = dataset.batch(batch_size).prefetch(tf.data.AUTOTUNE)
```

tf.keras.preprocessing.image.ImageDataGenerator: For data augmentation.

```
from tensorflow.keras.preprocessing.image import ImageDataGenerator
datagen = ImageDataGenerator(
    rotation_range=20,
    width_shift_range=0.2,
    height_shift_range=0.2,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True,
    fill_mode='nearest'
)
```

Model Creation

tf.keras.models.Sequential: For building sequential models.

```
model = tf.keras.models.Sequential([
    tf.keras.layers.Conv2D(32, (3, 3), activation='relu', input_shape=(128, 128, 3)),
    tf.keras.layers.MaxPooling2D((2, 2)),
    tf.keras.layers.Conv2D(64, (3, 3), activation='relu'),
    tf.keras.layers.MaxPooling2D((2, 2)),
    tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(512, activation='relu'),
    tf.keras.layers.Dense(1, activation='sigmoid')
])
```

tf.keras.applications: For pre-trained models

from tensorflow.keras.applications import VGG16

base_model = VGG16(weights='imagenet', include_top=False, input_shape=(128, 128, 3))

Model Compilation

model.compile: Configures the model for training.

model.compile(optimizer='adam',

loss='binary_crossentropy',

metrics=['accuracy'])

Model Training

model.fit: Trains the model.

history = model.fit(train_images, train_labels, epochs=20, validation_data=(val_images, val_labels))

Model Evaluation and Prediction

model.evaluate: Evaluates the model performance.

test_loss, test_acc = model.evaluate(test_images, test_labels)

model.predict: Makes predictions.

predictions = model.predict(test_images)

CHAPTER 5

RESULTS

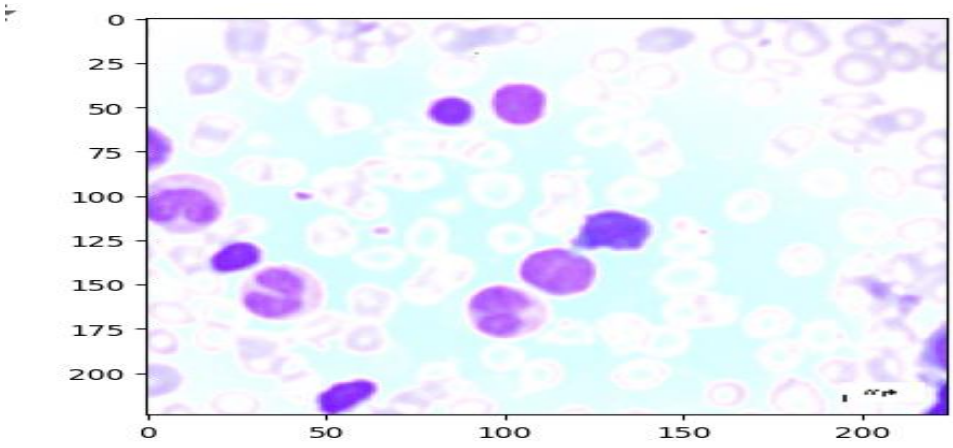


Fig 5.1:Blood Cancer Detected as Benign

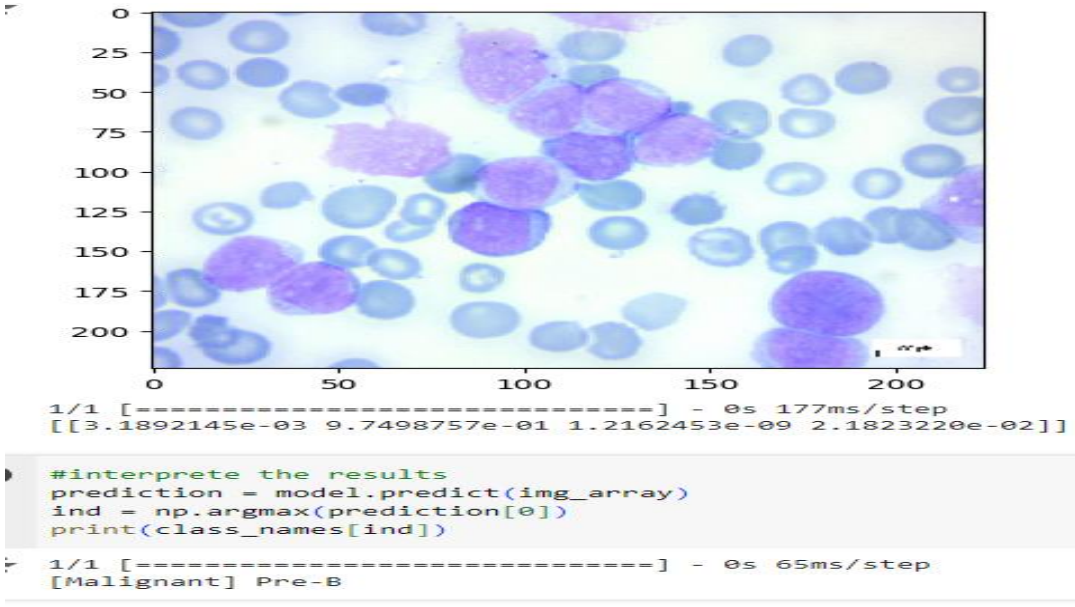
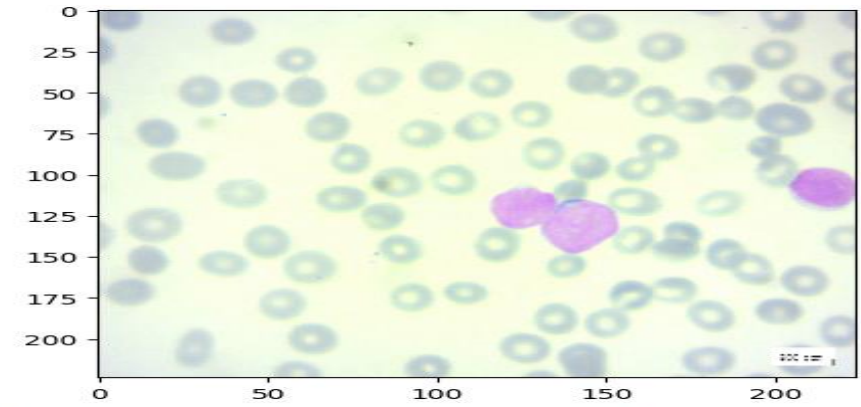


Fig 5.2:Blood Cancer Detected as Malignant pre-B

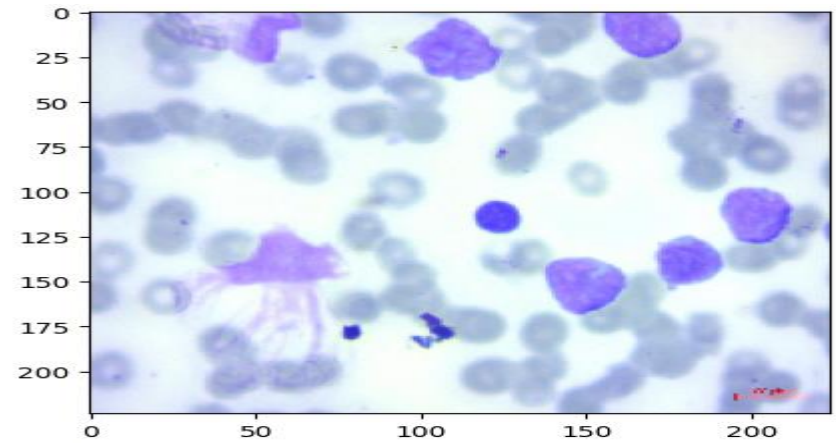



```
1/1 [==============================] - 0s 117ms/step
[[0.03660187 0.00191081 0.94682604 0.01466132]]
```

```
#interprete the results
prediction = model.predict(img_array)
ind = np.argmax(prediction[0])
print(class_names[ind])
```

```
1/1 [==============================] - 0s 74ms/step
[Malignant] Pro-B
```

Fig 5.3 Blood Cancer Detected as Malignant Pro-B




```
1/1 [==============================] - 0s 67ms/step
[[1.10795595e-01 4.77202944e-02 2.93743181e-08 8.41484129e-01]]
```

```
#interprete the results
prediction = model.predict(img_array)
ind = np.argmax(prediction[0])
print(class_names[ind])
```

```
1/1 [==============================] - 0s 87ms/step
[Malignant] early Pre-B
```

Fig 5.4:Blood Cancer Detected as Malignant early Pre-B

CHAPTER 6

CONCLUSION

The many different varieties of cancer, which are the collection of cells that are developing uncontrollably within the body, include breast cancer, lung cancer, skin cancer, and blood cancers like leukemia and lymphoma. One of the most important types of cancer is Acute Lymphoblastic Leukemia (ALL). This study examines the application of a novel technique for categorizing Acute Lymphoblastic Leukemia using cutting-edge technologies like Machine Learning (ML) and Deep Learning (DL). The major components of the proposed research pipeline include dataset construction, feature extraction using Convolutional Neural Network (CNN) architectures that have been pre-trained from each individual image of a blood cell, and classification using traditional ML-based classifiers. The dataset is split into two similar categories—benign and malignant—and then reconfigured into four significant classes, each of which has three subtypes of malignant, namely benign, early pre-B, pre-B, and pro-B. The research first extracts the features from CNN models, and then feeds the extracted features to feature selectors such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and SVC Feature Selectors, along with two nature-inspired algorithms such as Particle Swarm Optimization (PSO) and Cat Swarm Optimization (CSO). The seven ML classifiers have thereafter been used in research. A collection of experimental data has been compiled and analyzed in order to evaluate the effectiveness of the suggested architecture. The research first worked with pre-trained CNN models with conventional ML classifiers and found the highest accuracy of 98.43% accuracy without explicitly using the feature selection algorithms and nature inspired algorithms. The research has executed the proposed model with ResNet50 architecture with feature selection algorithms and PSO & CSO. Then, we have tracked out the highest accuracy up to 99.84%. This is very remarkable improvement in multi-class classification in malignant with the feature fusion and nature inspired algorithms.

CHAPTER 7

FUTURE ENCHANCEMENT

To enhance a blood cancer image detection system using Convolutional Neural Networks (CNNs), several future directions can be considered. Firstly, improving the model architecture can be crucial. Exploring advanced architectures such as Vision Transformers or EfficientNet, and utilizing model ensembling or self-supervised learning techniques, can enhance feature extraction and overall performance. Additionally, sophisticated data augmentation methods like Mixup or CutMix, as well as generating synthetic data through Generative Adversarial Networks (GANs), can enrich the training dataset and boost model robustness.

Fine-tuning pre-trained models on specific datasets and employing domain adaptation techniques can further refine the model's accuracy. Enhancing model interpretability through tools like Grad-CAM or attention mechanisms can provide valuable insights into the decision-making process, which is critical for clinical applications. Addressing robustness and generalization issues by incorporating adversarial training and validating the model on diverse external datasets can ensure the model performs well across various conditions.

CHAPTER 8

REFERENCES

1.Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning.

2. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks.

3. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep Learning.