Deep Feature Extraction-Based Classification of Acute Lymphoblastic Leukemia in Microscopic Blood Images Using MLP Approach

Submitted By

Md. Sazid Khan Shimul

ID: 221-35-870

Department of Software Engineering

Daffodil International University

Supervised By

Dr. S M Hasan Mahmud

Associate Professor

Department of Software Engineering

Daffodil International University

Thesis submitted in fulfillment of requirements for the award of the degree of Bachelor of Science

Department of Software Engineering (Major in Data Science)

DAFFODIL INTERNATIONAL UNIVERSITY

Deep Feature Extraction-Based Classification of Acute Lymphoblastic Leukemia in Microscopic Blood Images Using MLP Approach

Md. Sazid Khan Shimul

Bachelor of Science

DAFFODIL INTERNATIONAL UNIVERSITY

DAFFODIL INTERNATIONAL UNIVERSITY

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name	:	Md. Sazid Khan Shimul
Date of Birth	:	11 May 2003
Title	:	Deep Feature Extraction-Based Classification of Acute Lymphoblastic Leukemia in Microscopic Blood Images Using MLP Approach.
Academic Session	:	2022-2025

I declare that this thesis is classified as:

☐	CONFIDENTIAL	(Contains confidential information under the Official Secret Act 1997)*
☐	RESTRICTED	(Contains restricted information as specified by the organization where research was done)*
☐	OPEN ACCESS	I agree that my thesis to be published as online open access (Full Text)

I acknowledge that Daffodil International University reserves the following rights:

1. The Thesis is the Property of Daffodil International University.
2. The Library of Daffodil International University has the right to make copies of the thesis for the purpose of research only.
3. The Library of Daffodil International University has the right to make copies of the thesis for academic exchange.

Certified by:

Shimul

(Student's Signature)	(Supervisor's Signature)
221-35-870	Dr. S M Hasan Mahmud
Student ID	Name of Supervisor
Date: 23 December 2025	Date: 23 December 2024

NOTE : * If the thesis is CONFIDENTIAL or RESTRICTED, please attach a thesis declaration letter.

APPROVAL

This thesis titled on **"Deep Feature Extraction-Based Classification of Acute Lymphoblastic Leukemia in Microscopic Blood Images Using MLP Approach"**, submitted by **Md. Sazid Khan Shimul (ID: 221-35-870)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Chairman

Dr. S. M. Hasan Mahmud
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 1

A.H.M Shahariar Parvez
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 2

Tapushe Rabaya Toma
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 3

Khalid Been md. Badruzzaman Biplob
Lecturer (Senior Scale)
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

External Examiner

Dr. Md Sazzadur Rahman
Professor
Institute of Information technology
Jahangirnagar University, Bangladesh

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Bachelor of Science.

(Supervisor's Signature)

Full Name : Dr. S M Hasan Mahmud

Position : Associate Professor

Date : 23 December 2025

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Daffodil International University or any other institution.

Shimul

(Student's Signature)

Full Name : Md. Sazid Khan Shimul

ID Number : 221-35-870

Date : 23 December 2025

Deep Feature Extraction-Based Classification of Acute Lymphoblastic Leukemia in Microscopic Blood Images Using MLP Approach

Md. Sazid Khan Shimul

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Bachelor of Science

Department of Software Engineering (Major in Data Science)

DAFFODIL INTERNATIONAL UNIVERSITY

DECEMBER 2025

ACKNOWLEDGEMENTS

I have come this fur by the grace of Almighty. Many people helped me to walk on this journey. Among them, firstly I would like to appreciate my supervisor Associate Professor Dr. S M Hasan Mahmud, who guided, mentored and motivated me all the time.

I would wish to thank every teacher who led to my academic development and made me the person that I am today. I would like to convey my appreciation to Department of Software Engineering and Daffodil International University as the sources.

I want to thank you all so much. This journey would be impossible without you. I am thankful to my family, friends, classmates, seniors, juniors who have directly or indirectly contributed their hands in this odyssey.

DEDICATION

This work is devoted to my loving and hardworking mother, who raised me since my childhood after my father passed away. Her countless sacrifices and nurturing nature made me who I am today.

ABSTRACT

Correct and timely identification of Acute Lymphoblastic Leukemia (ALL) using microscopic images of the blood smear is critical in any attempt to make a clinical decision. This thesis presents a developmental evaluation on the frameworks of deep feature extraction and dimensionality reduction schemes in automated classification of Acute Lymphoblastic Leukemia (ALL) from peripheral blood smear images. The experiments were carried out on a publicly available ALL image dataset (CNMC), in which the natural imbalance in the classes was addressed with the help of ADASYN oversampling to ensure a robust learning process. Featuring seven pretrained architectures- five convolutional neural networks (ResNet50, DenseNet121, InceptionV3, Xception, EfficientNetB0) and two Vision Transformer architectures (ViT-B16 and ViT-L16), high-dimensional feature embeddings were generated and compared. The two dimensionality reduction techniques were Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) with five sizes of features of 1024, 900, 700, 500, and 300, respectively, leading to seventy different possible combinations of features and models.

All configurations were evaluated using a multilayer perceptron classifier based on stratified 5-fold cross-validation and the performance was measured using AUC, accuracy, F1-score, recall, precision, and MCC. The level of performance exhibited by PCA had been found to be higher as compared to that of RFE in most extractors. The CNN-based performance was best with DenseNet121-PCA(700), but the discriminative ability of transformer models were significantly higher. ViT-L16-PCA(1024) had the largest AUC of 96.24, which is the best overall performance. To justify the strength of the classifier, additional comparisons were made with Support Vector Machine (SVM), Random Forest, XGBoost, and Logistic Regression classifiers, in which MLP classifier continued to outperform all alternatives. Explainability of the model was assessed by Vision Transformer attention maps confirmed thate model was able to focus on clinically important elements such as chromatin distribution and nuclear boundaries.

The results associated with the study imply that global attention of transformers along with the dimensionality reduction of variance is a very promising pipeline of hematological image classification. The proposed model has strong potential to be used in automated systems of diagnostic assistance.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

x	Original pixel value
x_norm	Normalized pixel value
F	Extracted feature vector
I	Input image
θ	Pretrained model parameters
h_CLS	CLS-token feature embedding
x_patches	Patch embeddings
x_pos	Positional embeddings
x_CLS	Classification token embedding
X	Input feature matrix
W	PCA loading/eigenvector matrix
Z	PCA-transformed feature matrix
Σ	Covariance matrix
P1, P2, …, Pk	Principal components
W	Feature weight/importance coefficient
G_i	Number of synthetic samples generated for sample i
Δ_i	Imbalance level for sample i
x_i	Minority-class sample
x_nn	Nearest neighbor sample
λ	Random value (0
W1, W2	Weight matrices of layers
b1, b2	Bias vectors
ŷ	Model prediction
y	True label
$\mathscr{L}$	Loss function
N	Number of samples
D	Dataset
D_k	k-th fold of dataset
A_k	Activation feature map (k-th channel)
α_ck	Importance weight for class c and feature map k
L_c	Final Grad-CAM heatmap
y_c	Class score for class c
Q	Query matrix
K	Key matrix
V	Value matrix
d_k	Dimension of key vectors

LIST OF ABBREVIATIONS

ALL	Acute Lymphoblastic Leukemia
CNN	Convolutional Neural Network
ViT	Vision Transformer
MLP	Multi-Layer Perceptron
SVM	Support Vector Machine
MA ViT	Multi-scale Attention Vision Transformer
ResNet	Residual Network (e.g., ResNet50)
DenseNet	Densely Connected Convolutional Networks (e.g., DenseNet121)
VGG	Visual Geometry Group (e.g., VGG16, VGG19)
DDRNet	Deep Dual-resolution Networks
MoCo	Momentum Contrast
AI	Artificial Intelligence
DL	Deep Learning
XAI	Explainable Artificial Intelligence
PCA	Principal Component Analysis
RFE	Recursive Feature Elimination
ADASYN	Adaptive Synthetic Sampling
SMOTE	Synthetic Minority Over-sampling Technique
Grad-CAM	Gradient-weighted Class Activation Mapping
SHAP	SHapley Additive exPlanations
LIME	Local Interpretable Model-agnostic Explanations
GAP	Global Average Pooling
CLS	Classification
ReLU	Rectified Linear Unit
RBF	Radial Basis Function
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
MCC	Matthews Correlation Coefficient
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
TPR	True Positive Rate
FPR	False Positive Rate
GPU	Graphics Processing Unit
CPU	Central Processing Unit
BMP	Bitmap Image File

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background

Acute Lymphoblastic Leukemia (ALL) is a hematological malignant cell disorder that is concerned with uncontrolled multiplication of immature lymphoid cells in the bone marrow and peripheral blood. It is among the most prevalent leukemia type that occur in both children and adults, accompanied by some serious diagnosis difficulties and rapidity of development (Abir et al. [1]). According to American Cancer Society [2], in just 2025 alone, around 6100 new cases of ALL (3,450 in males and 2,650 in females) and about 1400 deaths from ALL (720 in males and 680 in females) are recorded. While it is highly treatable when children are affected from it, but it is totally the opposite scenario when middle aged adults suffer from ALL.

In the planning of treatment and better survival rates, early and correct diagnosis of ALL is essential. Diagnosis is traditionally conducted by microscopic blood smear images, which have to be manually examined by trained hematologists. It is a tedious and time-intensive procedure which is subject to observer bias (Asar et al. [3]).

Deep learning (DL) techniques showed a high potential in automated leukemia classification and detection in recent years. Convolutional Neural Networks (CNNs) and other DL models are able to automatically extract hierarchical feature representations of images and avoids the usage of handcrafted features, and is also very accurate (Tanwar et al. [4]). Moreover, Vision Transformers (ViTs) have become the possible alternative to CNNs, offering better contextual relationships modelling in images at a global level using self-attention mechanisms (Ben-Suliman et al. [5]).

Nevertheless, despite such developments, automated classification of ALL still has a number of challenges. The overfitting and computation inefficiency can be caused by high-dimensional feature representations. Any kind of class imbalance, such as some

subtypes of leukemia being underrepresented in the data, may also bias prediction models. In addition, the majority of DL models are black boxes with little elucidation capacity and clinical dependability (Rodriguez et al. [6]). The limitations of principal components reduction algorithms, like Principal Component Analysis (PCA) and feature selection algorithms, like Recursive Feature Elimination (RFE) along with resampling algorithms such as the ADASYN, have been presented to overcome these shortcomings and enhance the robustness of models (Oybek Kizi et al. [7]).

A viable solution to improve accuracy, generalization, and explainability of leukemia detectors, is through using methods that embrace advance architecture of feature extractors, dimensionality reduction, resampling strategies, and neural classifiers. The strategies can harness the advantages of CNN and ViT backbones to achieve local and global features of blood smear images and combine resampling and feature selection methods to maximize the performance over unbalanced data sets.

1.2 Problem Statement

Despite the potential of the DL-based methods, automated classification of blood smear images is a challenging task because of several factors:

1. **Multi-dimensional feature space:** Deep features extracted from CNN and ViT models have the risk of being redundant and expensive to compute.
2. **Class imbalance:** In most ALL datasets the class variation is uneven between the blast cells which is against the generalization of the model.
3. **Minimal interpretability:** DL models are black-box and less favourable to clinicians, which impacts negatively on practical adoption.
4. **Integration of heterogeneous architectures:** The synergistic combination of CNN and ViT models for improved feature representation remains underexplored.

To overcome these difficulties, a DL framework is essential to combine architectural feature extractor, dimensionality reduction, intelligent resampling, neural classifier, and explainable AI to enhance the precision of classification, generalization, and explanations

1.3 Research Goal

The main objective of this study is to design and test a system of deep learning to classify Acute Lymphoblastic Leukemia in centrifugated blood samples under the microscope. This framework aims to:

- Use several different feature extraction models, such as CNN (Resnet50, DenseNet121, InceptionV3, Xception, EfficientNetB0), and ViT (B-16, L-16), to capture local and global features of an image.
- Optimize the high-dimensional feature representation using the dimensionality reduction methods (PCA) and selecting features (RFE).
- Introduce smart resampling methods (ADASYN) to respond to the issue of class imbalance in the data.
- Apply a Multi-Layer Perceptron (MLP) Classifier to make the use of the extracted and refined features to perform accurate binary classification of Acute Lymphoblastic Leukemia.
- Include explainable AI techniques such as Grad-CAM for CNNs and attention map visualization for ViTs, to explain and understand model decisions.

1.4 Research Question

The key questions to be answered in this research are as follows:

1. What is the best way to achieve an improved accuracy of ALL classification with deep learning framework (based on CNN and ViT) with feature extraction, dimensionality reduction (with PCA), and resampling (with ADASYN) compared to individual architecture?

2. How far can the various feature extraction backbones (CNN vs. ViT) be used to increase the performance, and which configuration is able to achieve maximum generalization across unseen blood smear images?
3. What is the effect of using dimensionality reduction and feature selection method (PCA and RFE) on computation efficiency, convergence of the models, and model's predictive performance?
4. Can the implementation of explainable AI methods, such as Grad-CAM and ViT attention maps, enhance interpretability while maintaining high classification accuracy?

1.5 Objectives

In order to find out the answers to the research questions, this study will aim at:

1. To apply the CNN and ViT frameworks for extracting features of ALL blood smears images.
2. To optimize the representations of features through PCA and RFE to reduce the dimensionality in order to enhance the generalization.
3. To balance class distributions using ADASYN resampling techniques to enhance predictive accuracy.
4. To train and evaluate an MLP classifier on extracted and refined features, with the help of performance metrics such as accuracy, precision, recall, specificity, F1-score, MCC, and AUC.
5. To visualize model focus regions using Grad-CAM (CNN) and attention maps (ViT) to gain better interpretability and clinical understanding.

1.6 Significance of The Study

The proposed research contributes significantly to the field of medical image analysis and diagnosis of leukemia:

- **Clinical Impact:** By automating the diagnosis of ALL, the framework decreases the need for manual analysis, accelerate the process, and cut down on the possibility of human error.
- **Technical progress:** The study illustrates how the integration of various deep learning architectures, dimensionality reduction, resampling, and neural classification in a single pipeline can be demonstrated to be successful.
- **Explainability:** The use of Grad-CAM and ViT attention maps present model decision in a way that can be transparent and therefore increase clinician acceptance and uptake.
- **Benchmarking:** The structural framework is utilized as the baseline of further study into automated hematological images analysing and biomedical interventions.

1.7 Scope of The Study

The scope of the research can be described in the following manner:

- This study focuses on microscopic blood smear images of Acute Lymphoblastic Leukemia (ALL) as its subject matter.
- Pre-trained CNNs (ResNet50, DenseNet121, InceptionV3, Xception and EfficientNetB0) and ViTs (B-16 and L-16) are used to extract features.
- Dimensionality reduction (PCA) and feature selection (RFE) are used to optimize model in terms of efficiency and accuracy.
- The issue of class imbalance is solved with the help of ADASYN resampling.
- A Multi-Layer Perceptron (MLP) is employed for binary classification.
- The model predictions are explained using explainable AI methods (Grad-CAM and ViT attention maps).
- The experiment measures the performance of the models based on 5-fold cross-validation and the conventional classification measures.
- The use of augmentation, multi-class leukemia subtypes other than ALL, and other non-DL classification methods do not feature in this research.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

The deep learning technology has been gaining momentum rapidly out of the computer-aided software for diagnosing various medical imaging issues and particularly for identifying the presence of the hematological disorder like leukemia.
Microscopic images from blood smears are complex that has morphological variations which make a manual diagnosis very time consuming and error prone. With the advancement in convolutional neural networks, transfer learning, and transformer-based models, it appears that researchers have come to be remarkably accurate in determining the type of leukemia. However, despite all these improvements, there are still some limitations in the methodology and practicality of the models. Which limits the generalization of the model to new datasets, the efficiency of the data and the interpretability of the results.

2.2 CNN-Based and Transfer Learning Approach

CNNs have been the backbone of the modern medical image classification method because of its automatic learning ability to learn the spatial hierarchy from the pixel data. In the case of diagnosing leukemia, CNNs remove the software's requirement for manual and time-consuming feature extraction by identifying the cells at a morphological level.

Ramesh et al. [8] have used VGG19, ResNet50, and ResNet101 for the classification of acute leukemia and have obtained an accuracy measure of 92.62%. Their results confirmed that deeper CNN layers can detect the fine-grained variations of the structure of blood cells. Similarly, Rai et al. [9] also compared several architectures, including

AlexNet, ResNet, VGG, LeNet and EfficientNet. The reported result is 88.97% accuracy using AlexNet. The research made a point that careful data preprocessing and normalization have a great impact on network stabilization.

Transfer learning further improved the efficiency of the diagnosis by implementing pretrained models using large-scale datasets. Mantri et al. [10] also used EfficientNetB3 with transfer learning and got the testing accuracy of 96.87%. The pretrained weights allowed for better convergence and produced less dependence on labelled data. Haque et al. [11] used a combination of Inception-ResNet with preprocessing and augmentation and achieved F1-scores of better than 96% for binary and multiclass leukemia classification.

Lightweight networks have also attracted attention for their suitability in an environment with real-time and low-resource requirements. Makem et al. [12] published their results of the L1 regularized HSV based augmented model of MobileNet_M, that helped in reducing overfitting. With 95.33% accuracy and a 0.95 F1-score, this proved to be efficient without any sacrifice of diagnostic accuracy. These studies collectively emphasize the adaptability and dominance of CNN-based and transfer learning methods in leukemia detection.

2.3 Ensemble and Optimization Based Learning

In order to further improve the robustness and generalization, ensemble and optimization driven architectures were proposed. These comprise several combinations of CNN or different learning strategies to lower the bias and variance.

Hasanaath et al. [13] proposed the idea of using feature-fusion ensemble combining InceptionV2, ResNet, DenseNet121 and VGG16. A SVM classifier was then applied on the concatenated deep features, achieving accuracy of 91.63%. The approach proved that a fusion of complementary feature representations will result in higher consistency. Abhishek et al. [14] have developed a fuzzy ensemble with Gompertz function weighted ensemble which enhances model adaptivity and provides an accuracy of 88.80% on five-class leukemia data.

Muhammad et al. [15] further advanced this direction by integrating EfficientNet-B7 with optimization-driven strategies and explainability modules, achieving accuracies exceeding 96% across multiple leukemia datasets while reducing computational overhead. Similarly, Huang et al. [16] used a multi-stage ensemble of the most effective CNN models that were improved using Bayesian optimization and achieved an accuracy of 96.26 percent and showed that the optimized ensemble learning has the ability to enhance the diagnostic reliability remarkably.

Optimization techniques have also been applied to optimize the deep learning performance. Abd El-Aziz et al. [17] implemented EfficientNetV2-S with 5-fold cross-validation and the authors added features from geometric symmetry representation of images. Their model achieved 97.34% accuracy, which confirmed that hybrid optimization strategies can improve spatial learning.

Jawahar et al. [18] used dilated residual blocks and channel-spatial attention, and incorporated DDRNet, resulting an accuracy of 91.98% and an F1-score of 0.96. Meanwhile, Alim et al. [19] fine-tuned ResNet-50 using dense and dropout layers. Which led to an accuracy of 99.38%.

These works together emphasize the trend towards the customization and the optimization of architecture for improved performance for diagnostics.

2.4 Transformer Based and Hybrid Architectures

Transformers that were originally created for natural language processing have recently appeared for medical imaging in the form of Vision Transformers (ViTs). Unlike CNNs that utilize local patterns through convolution, ViTs learn long range dependency through the self-attention mechanisms which makes it easy to understand the global features.

Jammal et al. [20] in their paper proposed a multi-scale attention vision transformer (MA-ViT). In which convolutional blocks and attention layers of transformers are integrated together. It achieves accuracy of 98.54% for C-NMC dataset.

Preethika et al. [21] designed a CNN-transformer hybrid model which consists of a convolutional backbone and a transformer encoder, reached an accuracy of 97.40%, outperforming architectures such as EfficientNet and ResNet.

Such models are a hybrid of spatial learning capability of CNNs and the contextual reasoning capability of transformers, which is a new era of balanced feature learning.

2.5 Explainable Artificial Intelligence and Model Interpretability

In the medical field, interpretability plays a very important role in bringing AI to the clinical setting. Even though Black-box models are precise, practices often believe that they need explainable decision-making.

Jia et al. [22] proposed a framework based on Grad-CAM visualization in model interpretation to visualize the impact of each image region in the classification decision. And then allowing the clinicians to judge if the classification results are reasonable. This results in diagnostic trust and accountability.

Meanwhile, Islam et al. [23] proposed explainable deep learning architecture that used visualization-based interpretation techniques to show how particular image regions affect the results of the classification. Their results supported the need to include methods like heatmaps and attribution maps to improve the transparency, trust, and clinical usability - which is all similar to the contemporary XAI tools, such as SHAP, LIME, and Grad-CAM.

2.6 Comparative Analysis

The reviewed studies had accuracies ranging from 88% to 99%, which show that the classification of leukemia has been greatly advanced. CNN-based models are dominant, because they are efficient and easy to understand. While transfer learning and ensemble methods greatly improved the stability and reduced training requirements. Transformer model and hybrid models have introduced new sophisticated attention and global feature awareness, and have achieved near-perfect performance in some cases.

However, despite strong results, these approaches share notable limitations:

- Most are based on the sole use of deep feature extraction without systematic feature selection.
- Few addressed the issue of data imbalance with sophisticated methods of oversampling.
- Many gave priority to architectural tuning rather than end-to-end process optimization.
- Explainability, though emerging, remains secondary to raw accuracy metrics.

2.7 Research Gaps

Based on the reviewed works, there are still several gaps in the current researches about classification of leukemia:

1. **Lack of Incorporation of Feature Selection:**
   None of the studies approach the process of reducing dimensionality of extracted features systematically using dimensionality reducing techniques such as Principal Component Analysis (PCA) or Recursive Feature Elimination (RFE). That results in redundancy and unnecessary complexity of the model.

2. **Inadequate Handling of Data Imbalance:**
   While many studies use augmentation, few use intelligent resampling techniques such as ADASYN to balance minority classes, which is important when it comes to medical dataset that have uneven class distributions.

3. **Limited Exploration Outside of CNN Classifiers:**
   Deep features are seldom combined with typical machine learning classifiers such as MLP, SVM, or Random Forest although they can have a better generalization capability on refined features.

4. **Inactivity regarding Pipeline Integration:**

Most works focus on the model architecture rather than the whole analytical pipeline - from the preprocessing to the feature extraction, selection and classification - which leads to suboptimal utilization of the data and features.

5. **Weak Explainability Integration:**
   In the literature review, very few studies clearly associated the output of their models with an interpretable graphical representation or explanation-generating model, and so there is little confidence in the actual application of the models in clinical practice.

2.8 Proposed Research Solution

With these identified gaps, the following research question is set for the purposes of this study:

- "How can a deep learning pipeline combining multi-architecture feature extraction, dimensionality reduction, intelligent resampling, and neural classification improve the accuracy and generalization of leukemia detection systems?''

To answer this question, the proposed approach is introducing a new integrated pipeline consisting of the following:

- Feature extraction from a number of different pretrained architectures (ResNet50, DenseNet121, InceptionV3, Xception, EfficientNetB0, ViT-B16, ViT-L16).
- Dimension reduction using PCA and RFE to remove the redundant variables.
- Balancing the data through ADASYN as a way of countering class imbalance.
- Classification through a tuned MLP model to optimize decision boundaries.

This framework has a different contribution to the existing literature by considering methodological synergy instead of single-model tuning. Although the achieved accuracy (91.89%) is a little worse than some state-of-the-art transformer models, the approach shows a new direction for researchers that focuses on generalization, interpretability and balanced performance across architectures.

CHAPTER 3

METHODOLOGY

3.1 Overview of The Methodological Framework

The methodological approach taken in this work is based on the automated classification of Acute Lymphoblastic Leukemia, using transfer learning-based deep feature extraction integrated with a Multilayer Perceptron (MLP) classifier. The design of the proposed pipeline is to achieve a good accuracy level, robustness, and generalization by using pretrained Convolutional Neural Network (CNN) models and Vision Transformer (ViT) architectures as the feature extractors, followed by dimensionality reduction based on Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) models. To achieve a better class balance, Adaptive Synthetic Sampling (ADASYN) is applied and the updated refined features are used for training and evaluation of the model using 5-fold cross validation scheme.

The overall workflow contains -

1. Data Acquisition and Preprocessing.
2. Feature Extraction using Transfer Learning.
3. Dimensionality reduction and Feature Selection.
4. Resampling using ADASYN.
5. Training an MLP classifier.
6. Model Testing using Cross validation.
7. Reporting the results.
8. Model Interpretability using XAI.

The system has been coded in Python by using TensorFlow, scikit-learn, PyTorch, and other helper libraries. The general schematic of the methodological pipeline is shown as (Figure 3.1: Overall Research Framework).

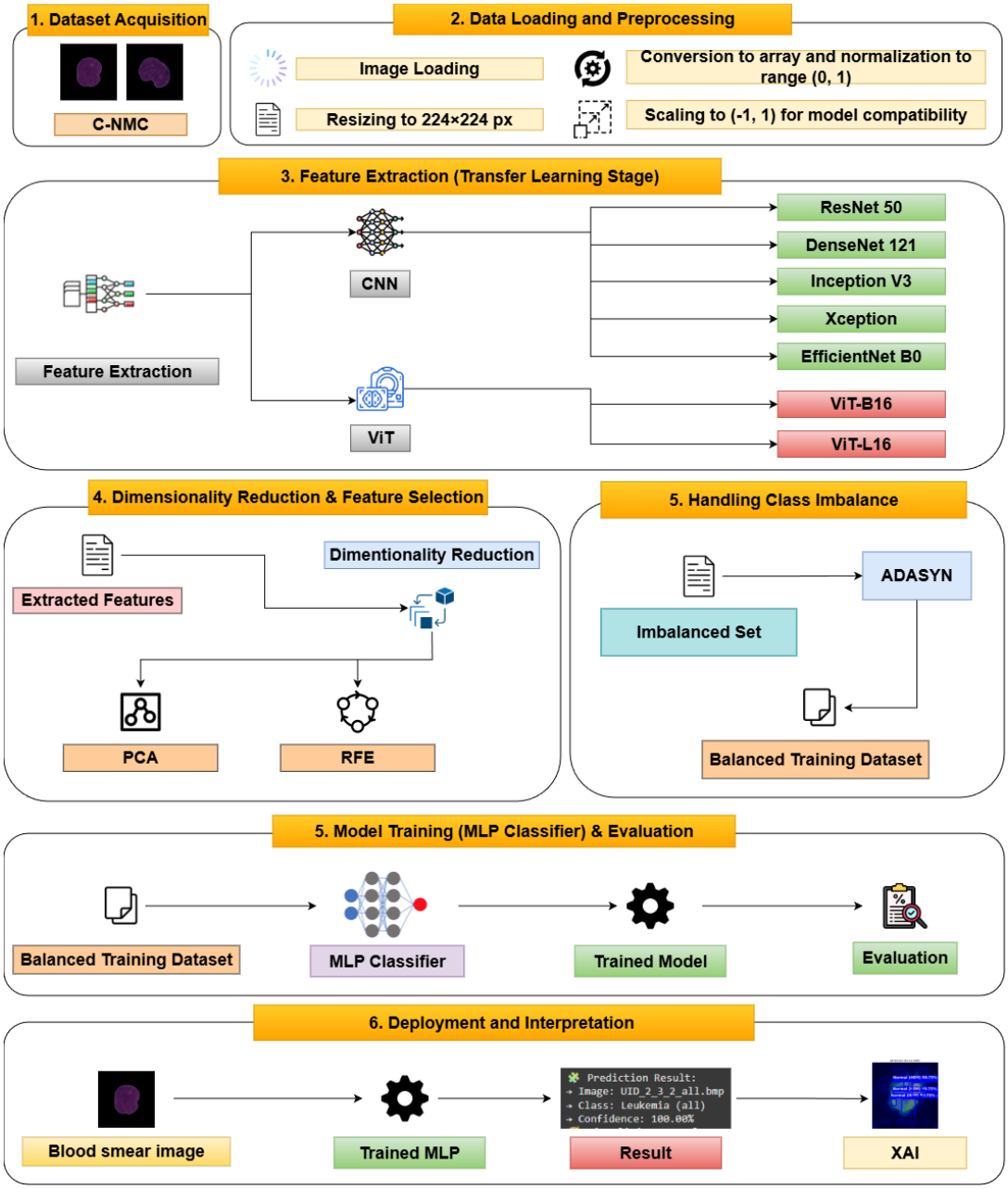


Figure 3.1: Overall Research Framework

3.2 Dataset Description and Preprocessing

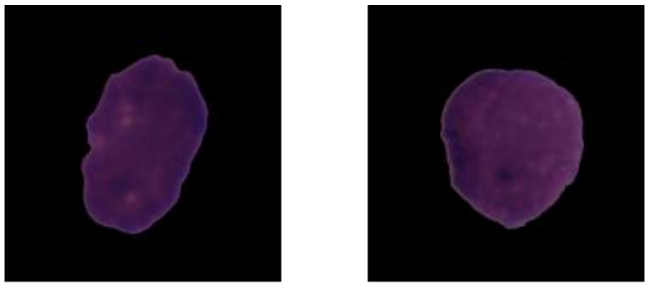
The data that are used in this research are microscopic blood smear images from C-NMC datset, collected from The Cancer Imaging Archive. The dataset of 10661 cell samples are classified into two classes: *acute lymphoblastic leukemia (ALL) and healthy (HEM)* cells. The dataset was divided into two root directories: "ALL" is for leukemia-positive samples, and ''HEM'' is for healthy samples. In each directory, all the image instances were in BMP format.

Each image was resized to spatial dimensions of 224 x 224 pixels, ensuring compatibility with pretrained networks input dimensions. All images were then normalized to the [0, 1] range and standardized to the interval (-1, 1) using the transformation:

$$\mathbf{x}_{\text{norm}} = \frac{x - 0.5}{0.5} \tag{3.1}$$

Image preprocessing and loading was done using the Keras preprocessing module. Each image was converted into an array and saved into memory-efficient batches to bypass the GPU/CPU overload during training. As a result, the standardization processes ensured that different feature extraction models were made comparable so that they could be evaluated comparably.

The figure (Figure 3.2: Sample Leukemia and Healthy Cell Images) depicts the sample images from both classes.



a. *Acute Lymphoblastic Leukemia* b. *Healthy Cell*

Figure 3.2: Sample Leukemia and Healthy Cell Images

3.3 Transfer Learning Based Feature Extraction

Transfer learning was used to extract high-level abstract features from input images using pretrained CNN and Vision Transformer models. Two types of deep models were used:

(i) Convolutional Neural Networks (CNNs).
(ii) Vision Transformer (ViT).

These models were pretrained on massive datasets like ImageNet, and could generalize the visual features very well for medical imaging applications (Mittal et al. [24]).

3.3.1 Convolutional Neural Network Models

The CNNs used in this study are:

1. ResNet50.
2. DenseNet121.
3. InceptionV3.
4. Xception, and
5. EfficientNetB0.

Each of these models offers unique representational capabilities thanks to their architectural innovations:

- **ResNet50** introduces residual connections that reduce the vanishing gradient issues by allowing identity mappings across layers and enhancing gradient flow (Sharma et al. [25]).

- **DenseNet121** is a feed forward structure that links each layer to all the other layers, which reuses the features and has efficient use of parameters (Arulananth et al. [26]).
- **InceptionV3** combines several receptive field filters within inception modules and thus enables the multi-scale feature extraction (Shah et al. [27]).

- **Xception** replaces standard convolutions with depthwise separable convolutions, which has less computational cost while maintaining as much representational capability as possible (Joshi et al. [28]).
- **EfficientNetB0** uses a compound scaling method of width, depth and resolution to maximize the model's performance at low computational budgets (Kanchana et al. [29]).

Each CNN model was initialized using ImageNet weights and and configured with the ***include_top=False***, which removes the classification head. Global Average Pooling (GAP) was used on the final convolutional feature maps to yield compact fixed dimension feature vectors. The extracted features are used as the numerical descriptors describing the contents in the images, which will be used for the dimension reduction and classification later on. The CNN-based feature extraction pipeline is shown in (Figure 3.3: CNN-Based Feature Extraction Pipeline).

As illustrated in the code implementation, for DenseNet121, the extracted features ''F'' have been received using forward propagation:

$$F = f_{CNN}(I; \theta_{ImageNet}) \tag{3.2}$$

Where "I" is the input image and "$\theta_{ImageNet}$" is the pretrained model parameters. Each feature vector was then used for the downstream processing.

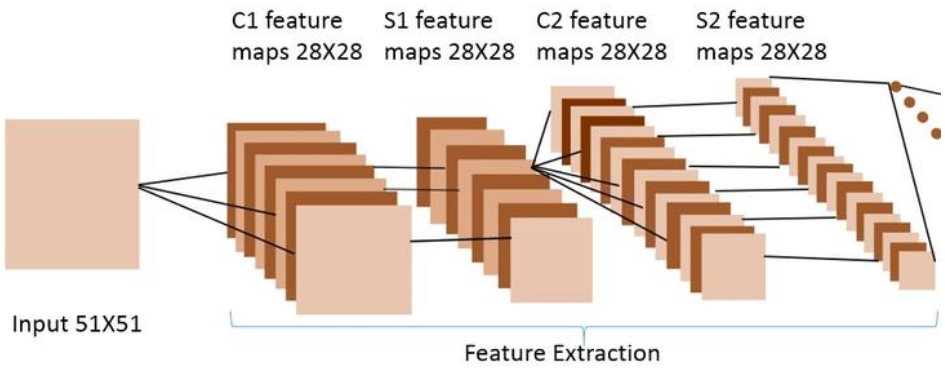


Figure 3.3: CNN-Based Feature Extraction Pipeline

3.3.2 Vision Transformer Models

In addition to using CNNs, models called Vision Transformer or ViTs were used to extract non-local contextual relationships between the image patches. Two versions of the transformer were studied: ViT-B/16 and ViT-L/16, which are pretrained on the ImageNet-21k dataset. The ViT architecture divides each input image into fixed-size patches (16×16) and projects them into a linear embedding sequence. A learnable [CLS] token is added to the sequence, which serves as a global representation of the image after multi-head self-attention processing (Reddy et al. [30]; Dosovitskiy et al. [31]).

Feature extraction from ViT models was done using the transformers library from HuggingFace. The global image embedding was obtained by taking the embedding of the [CLS] token from the last hidden state. Mathematically, the embedding vector ''h_{CLS}'' was calculated in the following way:

$$h_{CLS} = TransformerEncoder([\, x_{patches}, x_{pos}] + x_{CLS}) \tag{3.3}$$

Among all of the transformer variants that were tested, the variant ViT-L/16 with PCA=1024 gave the best classification performance. Dimensionality reduction of the extracted feature matrices was carried out and fed into MLP classifier for training. The working principle of feature extraction mechanism is shown in (Figure 3.4: Vision Transformer Feature Extraction Framework).

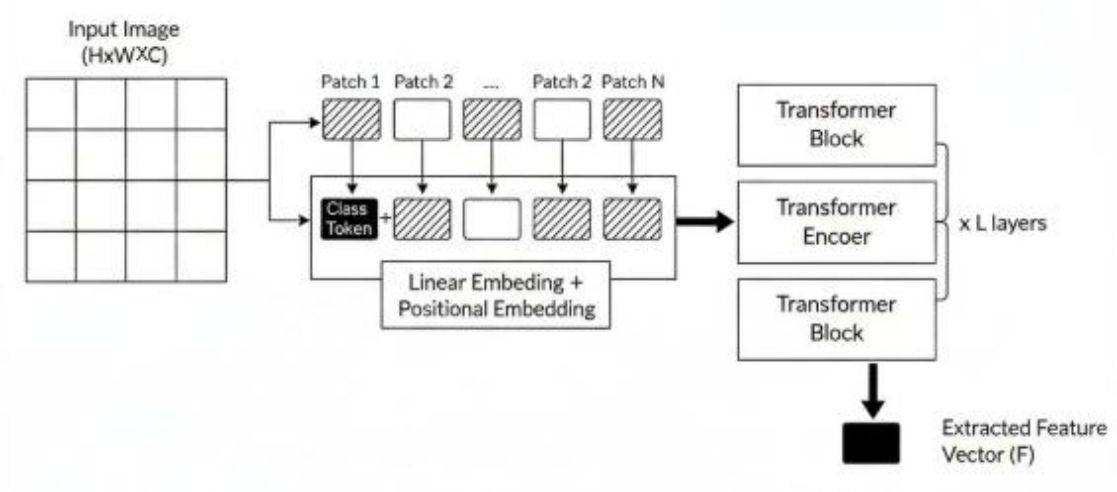


Figure 3.4: Vision Transformer Feature Extraction Framework

3.4 Dimensionality Reduction and Feature Selection

Because the extracted features were always of high dimensionality (usually greater than 10,000), dimensionality reduction was an important step to enhance the computational speed and reduce the overfitting. Two techniques were used:

1. Principal Component Analysis (PCA) and
2. Recursive Feature Elimination (RFE).

3.4.1 Principal Component Analysis (PCA)

PCA is an orthogonal linear transformation that projects the data to a lower dimension of maximum variance (Bharadiya et al. [32]). The transformation identifies principal components (P1, P2, ... P_k) as eigenvectors of the covariance matrix (Σ) of the data matrix (X). The transformation is given as:

$$Z = XW \tag{3.4}$$

Where (W) represents the matrix of principal component loadings. The first few components explain a lot of the variance in the data set, and can be reduced satisfactorily with only a small amount of information loss (Equation 3.4: PCA Transformation).

The dimensionality PCA (k = 1024) has been empirically selected by optimizing its performance in several experimental runs. Slight beyond reducing the calculation cost, PCA also improved the generalization ability by removing redundant features. The PCA system is shown in (Figure 3.5: Dimensionality reduction using PCA).

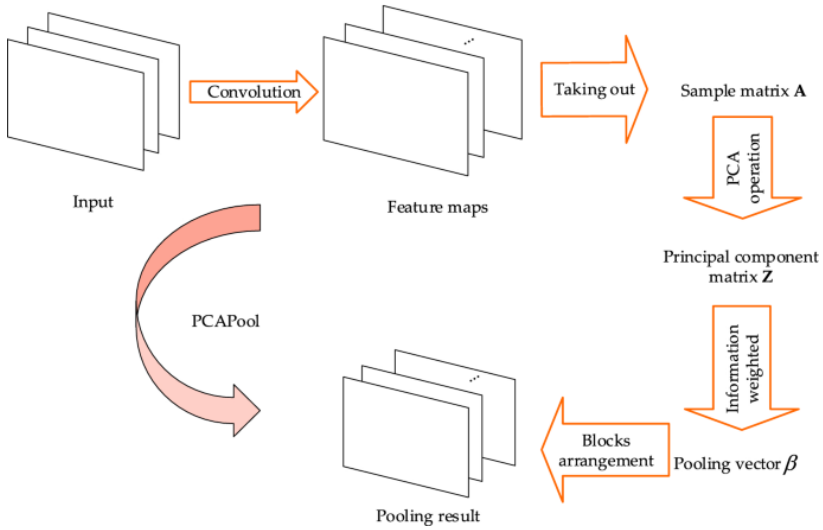


Figure 3.5: Dimensionality Reduction Using PCA

3.4.2 Recursive Feature Elimination (RFE)

RFE was optionally employed to identify the most informative features by recursively training an estimator and removing the least significant features until an optimal subset was obtained (Awad et al. [33]). The ranking of features is determined by the absolute magnitude of the feature weights (coefficients) "W"' from the underlying linear model:

$$Feature\ Rank \propto \frac{1}{|W|} \tag{3.5}$$

This indicates that features with the smallest absolute weights are considered the least important and are recursively eliminated in each iteration.

3.5 Data Balancing using ADASYN

Moderate class imbalance between the encumbrance of ALL-positive and healthy samples was observed in the dataset. In order to solve this problem, synthetic minority samples were generated in the feature space by a synthetic sampling algorithm- ‘’Adaptive Synthetic Sampling’’ (ADASYN). In contrast to the even sample synthesis in SMOTE, the minority samples in ADASYN are generated adaptively close to the decision boundaries. (Imani et al. [10])

The ADASYN algorithm computes the level of imbalance of class and creates new samples according to the following formula:

$$G_i = \Delta_i \times (x_i - x_{NN}) \times \lambda \tag{3.6}$$

Where (Δ_i) is the imbalance factor, (x_i) represents minority samples, (x_{NN}) denotes nearest neighbors, and (λ) is a random scaling factor between 0 and 1. Using ADASYN enabled greater stability of the model and greater sensitivity in Leukemia classification. The resulting balanced dataset provides a fair learning basis for the MLP classifier. (Figure 3.6: ADASYN Based Data Balancing) shows the effect of ADASYN in class balancing.

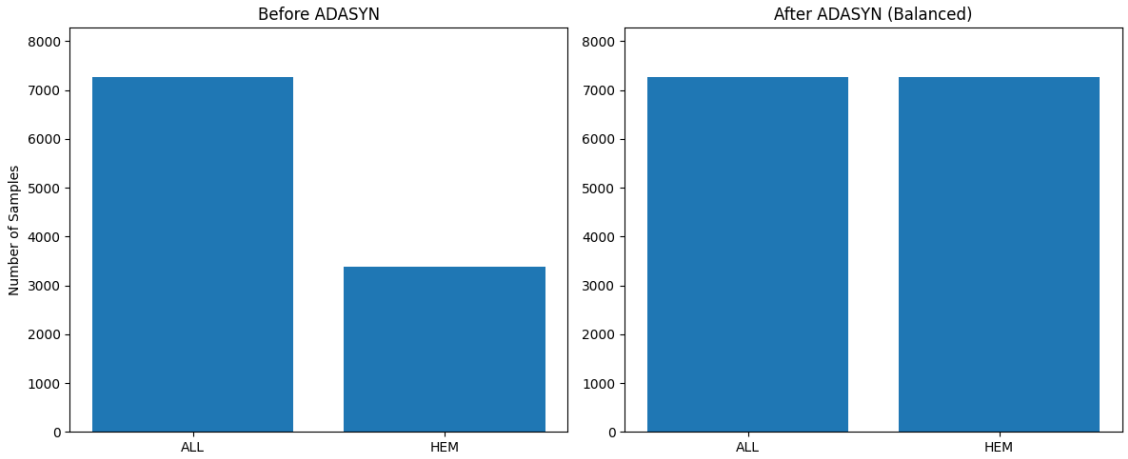


Figure 3.6: ADASYN Based Data Balancing

3.6 MLP Classifier Architecture

Following feature extraction and balancing, classification was performed using a **Multilayer Perceptron (MLP)** network. The MLP was implemented in TensorFlow and had one hidden layer and one output layer. The architecture has been developed as follows:

- Input layer size: depending upon reduced feature dimensions (e.g. 1024 in the case of PCA).
- Hidden layer: 100-1000 neurons dependent on experimental set up.
- Activation: Rectified linear Unit (ReLU).
- Regularization: L2 Weight penalty ($\lambda = 1\times10^{-4}$).
- Optimizer: Adam Optimizer with a learning rate of 1×10^{-3}.
- Output layer: a single neuron with Sigmoid activation for binary classification.

The forward propagation in the MLP may be written as:

$$\hat{y} = \sigma(W_2 \,.\, f(W_1 x + b_1) + b_2) \tag{3.7}$$

Where ($f(.)$) represents the ReLU activation, ($\sigma(.)$) represents the sigmoid function and (W_i, b_i) are learnable parameters. The loss function for the network was Binary Cross Entropy:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}[yi \log(\hat{y}i) + (1 - yi)\log(1 - \hat{y}i)] \tag{3.8}$$

The weights of the model were optimized by stochastic gradient descent using adaptive moments (Safar et al. [35]). A checkpointing mechanism was used to store the model with the best accuracy based on validation. MLP Classification Architecture is shown in (Figure 3.7: MLP Classification Architecture).

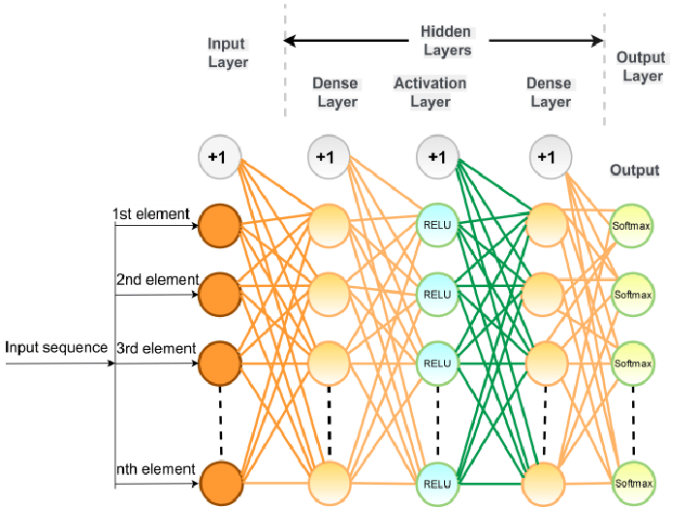


Figure 3.7: MLP Classification Architecture

3.7 Cross-validation Strategy

To guarantee the robustness and generalization of the proposed model, the result was validated using 5-Fold Stratified Cross-Validation. The dataset was partitioned into five equal parts, preserving the class distribution in each fold. In each iteration, the training set consisted of four folds and the testing set consisted of the other fold in a rotating order such that all folds became the validation set once. The average and standard deviation of the performance measure values across all the folds were reported.

Formally, for dataset D = $\{(x_i,y_i)\}_{i=1}^{N}$, it was split in to five subsets- D1, D2, D3, D4 and D5. For each fold (k):

$$Train = D \backslash D_k \quad , \quad Test = D_k \tag{3.9}$$

The 5-fold validation minimized overfitting and gave statistically reliable estimates of model accuracy and model stability (Szeghalmy et al. [36]).

3.8 Evaluation Metrics

The trained models were evaluated with some common classification measures derived from the entries in the confusion matrix: True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). The metrics include:

1. Accuracy = $\frac{(TP + TN)}{(TP + TN + FP + FN)}$
2. Precision = $\frac{TP}{(TP+FP)}$
3. Recall (Sensitivity) = $\frac{TP}{(TP+FN)}$
4. Specificity = $\frac{TN}{(TN+FP)}$
5. F1-Score = $2 \times \frac{(Precision \times Recall)}{(Precision + Recall))}$
6. MCC = $\frac{(TP \times TN)-(FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$

7. Area Under Curve (AUC) = $\int_0^1 TPR(FPR)\, d(FPR)$

These metrics together measure the classification ability of the system, its sensitivity towards the minority class and its robustness. Among them, MCC and AUC were highlighted because of their fair consideration of the true and false predictions in imbalanced consideration cases.

3.9 Explainable Artificial Intelligence (XAI) and Model Interpretability

Explainable Artificial Intelligence (XAI) has become an essential part of deep learning research in the modern research, especially medical imaging, where model transparency and explainability are just as important as the accuracy of the performance. In the context of acute lymphoblastic leukemia classification from microscopic blood smear images, it is very important to understand which parts of the input images have influence on the final decision that the model produces. Such interpretability guarantees that the automated system relies on meaningful hematological features (e.g. morphology of nucleus, cytoplasmic texture, or leukocyte shape) and not spurious correlations for making its prediction. Therefore, in the current study, post-hoc explainability methods were incorporated in the trained models to visualize and interpret its decision-making process.

To do this, two different visualization mechanisms were used depending on the type of model used: Gradient-weighted Class Activation Mapping (Grad-CAM) for Convolution Neural Network (CNN) backbones, and Attention Map Visualization for Vision Transformer (ViT) architectures. Together these methods gave a complete picture of which discriminative patterns were found by the models.

3.9.1 Grad-CAM for CNN Backbones

Grad-CAM is a very common visualization method aimed at providing visual explanations for CNN-based models. It generates a rough localization map which highlights image areas which are more influential for a particular classification decision.

The way that this works is by calculating the gradient of the score of the target class (i.e., the probability to be a particular type of leukemia, referred to as a "leukemia type") with respect to the feature maps of the last convolutional layer. This gradient represents how important each of the neurons in the feature maps will be in making the given decision.

Mathematically, for a given a class (c), the gradient of score (y^c) with respect to the feature map activations (A^k) of the final convolutional layer is given by:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \tag{3.10}$$

Where (α_k^c) is the importance weight of feature map (k), and (Z) is the number of the pixels in feature map. The final Grad-CAM heatmap ($L_{GradCAM}^c$) is then taken as the weighted sum of feature maps followed by a ReLU activation:

$$L_{GradCAM}^c = ReLu\ (\sum_k \alpha_k^c A^k) \tag{3.11}$$

The resulting heatmap is then up-sampled and overlayed to the original input image to create an image that is easy to interpret as an activation map. Areas of greater intensity are regions where the model paid most attention to making its prediction. In the present study, this approach was applied to the CNN-based models - ResNet50, DenseNet121, InceptionV3, Xception and EfficientNetB0, to qualify the focus on hematological features such as leukocyte nuclei and cytoplasm in a qualitative manner.

3.9.2 Attention Map Visualization for Vision Transformers

Unlike CNNs, Vision Transformers (ViTs) by nature learn the global dependencies in an image using self-attention mechanisms. Each transformer layer calculates attention weights which represent how much every image patch contributes to

every other image patch while learning the representation. Using these attention weights, it is possible to visualize the model's internal reasoning process and decision-making process.

In the proposed study, ViT-B16 and ViT-L16 models used the feature extraction module based on attention, and the purpose of their interpretability was realized by extracting the attention matrix corresponding to the classification token (CLS token) in the last transformer block. The attention maps were then reshaped to match the spatial arrangement of the input image patches and up-sampled to the original image resolution to get the ViT Attention Maps.

Given an input image split up into (N) patches, the attention mechanism calculates a score matrix as:

$$Attention\ (Q, K, V) = softmax\ (\frac{QK^T}{\sqrt{d_k}})\ V \tag{3.12}$$

Where (Q), (K), and (V) are the query, key and value matrices, and (d_k) is the dimension of the key vectors. The learned attention weights highlight the interactions between various image patches as well as the global context that influences the classification decision.

By visualizing these attention distributions, the ViT Attention Maps effectively indicate where the image regions are most influential in determining the categories of leukemia. The resulting attention maps proved that the ViT models, especially ViT-L16 with PCA = 1024, concentrated most of their attention mainly on the biologically important features such as nucleus boundary and texture variation. That is coherent with hematological interpretation principles.

3.9.3 Comparative Analysis and Insights of Interpretability

The combination of Grad-CAM and ViT Attention Maps gave complementary insights on model's interpretability. The Grad-CAM visualizations in CNNs tended to yield specific local regions of activation relative to discriminative morphological features, in contrast to the ViT attention maps which presented a more generalized view of the context of inter-patch relationships and spatial orientations of the image. Both visualization methods validated that the models learned medically relevant patterns demonstrating their reliability and interpretability in a clinical decision support application.

In addition, the incorporation of XAI methods helps to increase the trustworthiness and transparency of the proposed deep learning system. In a medical diagnosis pipeline, these types of visual explanations can be used as a form of validation for medical experts, to ensure that the model's predictions are based on biologically relevant evidence, and not an artifact or noise. These interpretability bases reflect the current focus of biomedical research on trustworthy AI adoption, which aims to use not only models with high accuracy but also modes that are explainable and have ethical transparency.

3.10 Implementation Environment

The experiments have been conducted using the platform Google Colab that gave access to GPU acceleration for training deep networks. The implementation environment consisted of:

- Operating System: Windows 11 64-bit.
- Programming Language: Python 3.9.21
- Libraries: TensorFlow, PyTorch, Scikit-learn, Imbalanced-learn, Transformers, NumPy, Pandas, Matplotlib.
- Hardware: NVIDIA Tesla T4 GPU (Colab runtime).
- Reproducibility: random seed value set to 42 in all the frameworks.

The configurations of each experiment have been repeated in the same way to ensure the reproducibility and comparability of the models. PCA models, metrics and checkpoints were saved for later analysis.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Overview

This chapter will elaborate on the findings of the experiment made with the help of the proposed leukemia (ALL vs. HEM) classification framework. The task of classification was performed with the help of a Multi-Layer Perceive (MLP) model incorporated with several strategies of feature extraction. These extracted features have been further processed through dimensionality-reduction techniques like Principal Component Analysis (PCA) and feature-selection techniques like Recursive Feature Elimination (RFE). A different combination of extractor and subset features created a unique experimental setup and the influence of represented features on model performance was fully evaluated.

In order to establish the best performing configuration, several performance metrics were used such as Accuracy, Precision, Recall, F1-score, Matthews Correlation Coefficient (MCC), and Area Under the ROC Curve (AUC). All these metrics provide a multi-dimensional view of the classifier behaviour, showing not only the general correctness, but the sensitivity of the classification behaviour to each of the classes, the reliability of the prediction, and the strength of the classification boundary. The incorporation of MCC is especially relevant to tasks that involve the medical diagnosis in question since it offers a fair assessment even in the case of the possible imbalances in the class distributions. Although the imbalance in classes is alleviated by the use of ADASYN in the process of data preparation.

Accompanied by the numerically quantitative indicators of performance, the chapter entails an array of visual tools of analysis to further render the performance of the models. Confusion matrices can be used to display how accurately and inaccurately a prediction has been made, while ROC curves provide insight about how well we can classify data

at any learning threshold. All tested configurations and their results are summarized and compared using table, which allows distinguishing which feature extraction and selection pipeline is the most effective.

Overall, this chapter is well organized and covers all the findings of the experiment. These results are arranged in way not only to report numerical results but also to describe performance patterns, point to the improvements made with the help of various feature-processing strategies, and find out the combinations that give the most reliable results.

4.2 Comparative Analysis of Different Feature Extraction Methods

A detailed comparison of seven deep learning models used in feature extraction including five CNN-based architectures (ResNet50, DenseNet121, InceptionV3, Xception, EfficientNetB0) and two ViT-based architectures (ViT-B16, ViT-L16) are provided in this section. The optimal dimensionality reduction configuration is used to evaluate each feature extractor so that it can be fairly compared at optimal performance.

CNNs are also excellent at learning patterns on a localized scale using convolutional operations, and each of them has its unique characteristics. ResNet50 employs residual connections to allow deeper networks, DenseNet121 has dense connectivity to allow feature reuse across layers, InceptionV3 has multi-scale parallel convolutions, Xception makes use of depth-wise separable convolutions, and EfficientNetB0 makes use of compound scaling foe balanced network dimensions. In comparison, ViT models feed images with patch sequences through self-attention mechanisms, enabling modeling of global dependencies across the entire image.

Table 4.1: Performance comparison of feature extractors with optimal configurations

Feature Extractor	Optimal Config	Accuracy	Precision	Recall	Specificity	F1 Score	MCC	AUC
ViT-L16	PCA 1024	91.89	89.50	84.34	95.40	86.84	81.06	96.24
ViT-B16	RFE 768	91.65	88.82	84.37	95.05	86.54	80.55	95.80
DenseNet121	PCA 700	87.58	82.42	77.43	92.30	79.85	70.95	92.09
Xception	PCA 1024	85.98	78.93	76.25	90.52	77.57	67.40	90.16
Inception V3	PCA 700	83.78	76.18	71.24	89.62	73.63	62.01	87.22
ResNet50	PCA 900	82.04	71.04	73.45	86.05	72.23	58.98	85.61
EfficientNetB0	PCA 300	78.20	67.55	60.47	86.46	63.81	48.43	81.60

Table 4.1 summarizes the performance of all the seven feature extractors with their respective optimal configurations identified to after an extensive hyperparameter search among the various techniques of dimensionality reduction and component sizes. The setting with maximum AUC is used to represent the optimal setting of each extractor.

The findings indicate an evident performance of hierarchical order where ViT-based systems achieve significantly higher performance as compared to CNN-based systems. ViT-L16 records the best performance in almost all measures with AUC of 96.24, accuracy of 91.89%, and MCC of 81.06%. The larger ViT-L16 model marginally exceeds best performing ViT-B16 (AUC 95.80), indicating that increased model capacity and deeper attention layers enhance feature representation quality for this task.

Among CNN architectures, DenseNet121 emerges as the strongest performer with AUC of 92.09, surpassing other CNNs by considerable margins. The dense connectivity of DenseNet concatenating features from all previous layers, enables efficient reuse of features and gradient flow, creating deep hierarchical features that are especially useful in the field of medical image analysis. Xception comes in as the second-best CNN with the AUC value of 90.16, proving that depth-wise separable convolutes are capable of finding the important patterns. InceptionV3 and ResNet50 score intermediate results of 87.22 and 85.61 respectively, while EfficientNetB0 demonstrates the weakest performance of 81.60 as the architecture has been designed to focus on efficiency instead of maximizing accuracy.

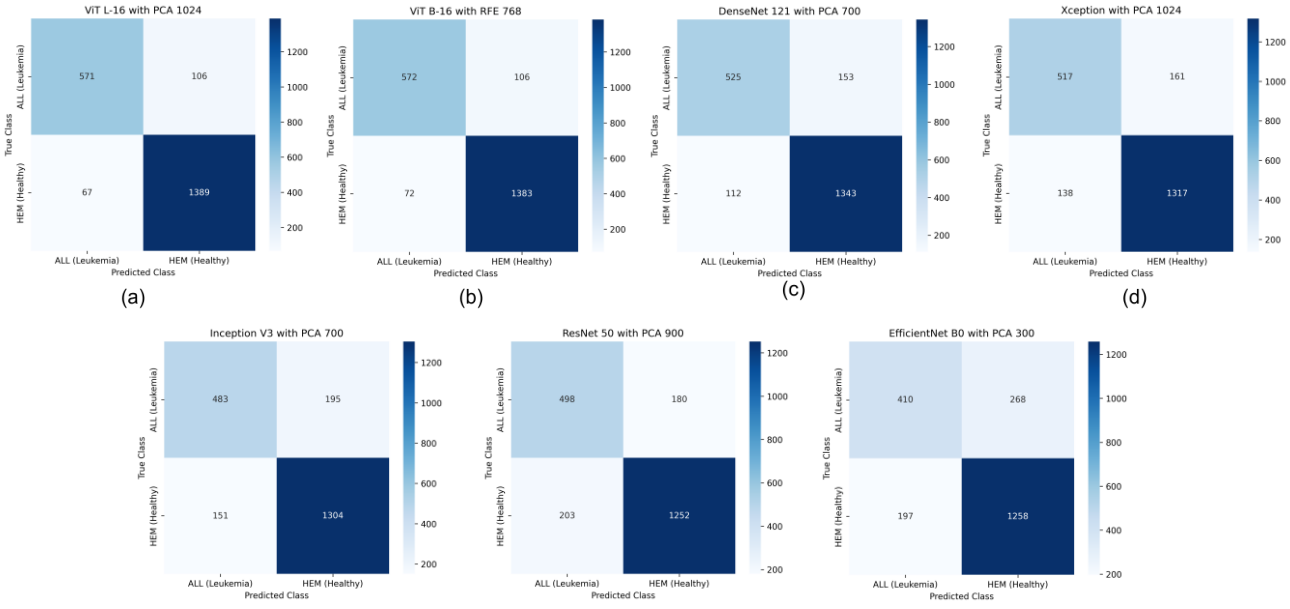


Figure 4.1: Confusion matrix for feature extractors at optimal settings

The substantial performance advantage of ViT architectures over CNNs can be explained by the underlying difference in their feature learning mechanisms. Although CNNs create hierarchical representations using localized receptive fields that gradually expand through network depth, ViTs use self-attention to directly bridge the relationships among the entire image patches regardless of their spatial proximity. In the case of leukemia classification, this global context modeling is especially useful because the indicators of the diagnosis can be in the form of some nuances in cell morphology, chromatin

configurations, and nuclear features distributed over the picture. Self-attention mechanism is capable capturing very long-range dependencies and thus letting ViTs produce more discriminative feature representations.

Moreover, the performance variance among CNN architectures highlights the importance of architectural design for medical image analysis. The high performance of DenseNet121 over ResNet50 indicates that reuse of features via dense connections is more efficient in preserving fine grained details that are important in differentiating subtle morphological variations between ALL and HEM cells. The average result of EfficientNetB0 regardless of its benefit of efficiency indicates that model capacity and representational power remain critical factors for this complex diagnostic task.

The results prove that ViT-L16 is the best feature extractor to classify leukemia, and DenseNet121 is the most suitable CNN model. In The subsequent sections, these high performing architectures are used to explore other components of the pipeline such as dimensionality reduction strategies and classifier configurations.

4.3 Effects of Different Dimensionality Reduction Strategies

Dimensionality reduction is a critical component of classification pipeline, transforming high-dimensional feature spaces into compact representations, increasing computational efficiency and reducing the risk of overfitting and potentially improving generalization by eliminating noisy or redundant features. To identify the most effective dimensionality reduction method to be used in classifying leukemia, this section compares the outcome of two common dimensionality reduction methods- Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE).

PCA performs unsupervised linear transformation of the data, expressing the features in the form of principal components that represent the maximum possible variation in the data. This does not need label information to maintain global data structure and thus, this method is computationally efficient and robust. In contrast, supervised feature selection is utilized in RFE. Here the least significant features are removed iteratively by

comparing their model assigned weights thereby selecting features most relevant for the specific classification task.

Two Representative feature extractors across various levels of performance (ViT-L16 (best overall), DenseNet121 (best CNN) are used in order to extensively compare the two methods. PCA and RFE are encouraged on each extractor with different amounts of retained features: 1024, 900, 700, 500 and 300 (only ViT-B16 used 768, 700, 600, 500 and 300 features due to its native feature dimensionality). This systemic study indicates the role of dimensionality reduction choice on the properties of feature extractors and the properties of the best size of a feature space.

Table 4.2: Performance comparison of best feature extractors with different dimensionality reduction configurations

Extractor	Methods	Components	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1 Score (%)	MCC (%)	AUC (%)
ViT- L16	PCA	1024	91.89	89.50	84.34	95.40	86.84	81.06	96.24
		900	91.28	86.18	86.43	93.54	86.30	79.90	95.40
		700	91.23	89.28	82.30	95.40	86.65	79.49	95.24
		500	90.95	87.02	84.07	94.15	85.52	78.96	95.38
		300	90.62	87.58	82.15	94.57	84.78	78.09	94.95
	RFE	1024	91.79	87.46	86.56	94.23	87.01	81.01	96.20
		900	91.28	87.50	84.66	94.36	86.06	79.73	95.04
		700	91.23	87.25	84.81	94.22	86.01	79.64	95.84
		500	90.67	85.48	85.10	93.26	85.29	78.46	95.48
		300	91.09	87.77	83.63	94.57	85.65	79.25	94.90
DenseNet121	PCA	1024	87.67	80.92	80.09	91.20	80.50	71.49	91.35
		900	88.33	83.89	78.32	92.99	81.01	72.68	91.88
		700	87.58	82.42	77.43	92.30	79.85	70.95	92.09
		500	86.87	80.80	76.99	91.48	78.85	69.39	90.80
		300	87.25	81.33	77.73	91.68	79.49	70.28	91.00
	RFE	1024	86.64	79.28	78.47	90.45	78.84	69.12	84.46
		900	86.87	80.71	77.14	91.41	78.88	69.40	90.73
		700	86.50	83.16	72.12	93.20	77.25	68.07	90.49
		500	86.22	79.63	76.11	90.93	77.83	67.87	90.14
		300	84.56	78.68	71.53	91.07	75.02	64.35	87.94

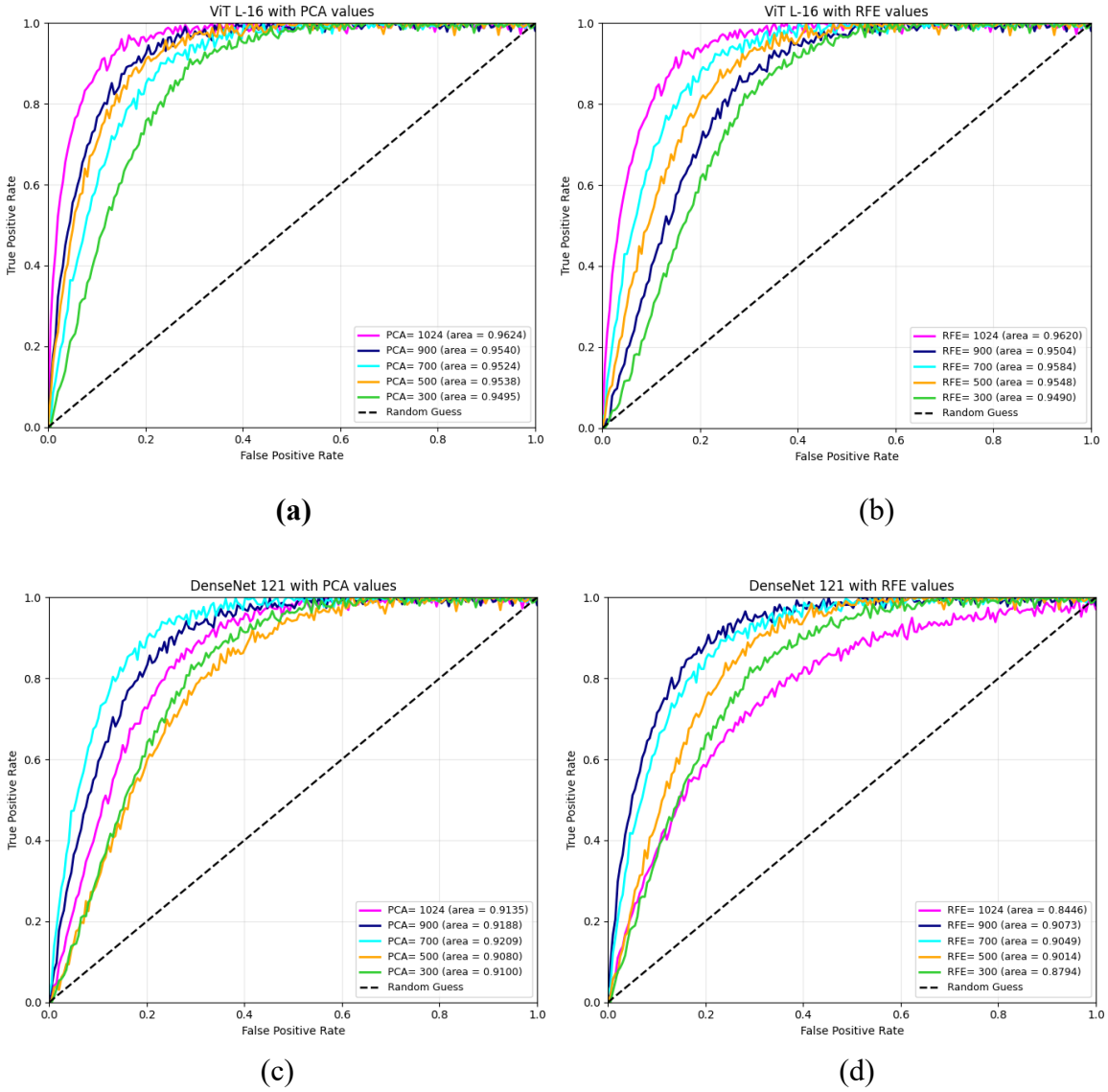


Figure 4.2: AUC-ROC Curve of the Best Performing Feature Extractors On Multiple Dimensionality Reduction Configuration.

The comparative analysis is visualized in several perspectives in the form of figure 4.2. shows line graphs of the AUC performance in relation to the number of components of DenseNet121 and different lines are drawn between the PCA and RFE approaches. The plots show that PCA has maintained similar AUC at most of the dimensionalities with the best performance when AUC is used at 700 components.

These findings all show that PCA is better or equal to RFE in various feature extractors and dimensionality choices. In the case of DenseNet121, PCA has much better results (92.09% vs 90.73% with RFE), which is 1.36% points higher.

It is possible to explain the regular benefit of PCA by its variances conserving traits. PCA preserves the global organization and structure of high-dimensional feature space by plotting data onto major components by their variance explained. The feature is especially useful with deep learning characteristics, which represent fine, spread representations wherein data can be distributed across numerous dimensions. The excessively greedy elimination method at RFE could strip away features that are individually not considered important, but then become significant contributing factors of the overall class separability together with the comprising minor interactions.

Moreover, the best dimensionality also depends on architecture. ViT-L16 performs optimally when the full or near-full dimensionality (1024 components) are preserved implying that the elements of the transformer have limited redundancy and the majority of dimensions contain valuable information in them. Efforts on narrowing Net121 to moderate compression (700 components out of 1024 original) indicate that there is certainly redundancy in CNN features that can indeed be eliminated by PCA to retrieve discriminative information. ResNet50 also enjoys moderate reduction (900 components), which is a compromise between information retention and noise removal.

Interestingly, the aggressive dimensionality reduction (e.g. 300-500 components) continuously deteriorates the performance of all extractors, which means that information loss is high in cases of extreme compression. This observation implies that the deep-networks as learned high-dimensional representations have to be stored in relatively high-dimensional spaces to sustain discriminative capacity in accomplishing this difficult task.

Further practical benefit is given by PCA due to its computational efficiency. Being an unsupervised algorithm requiring only eigen decomposition, PCA executes rapidly without iterative model training. RFE requires training of various models using various feature subsets, which is very expensive, especially when feature space predictors are large and complex predictors are used.

On the basis of these all-inclusive results, it was found that PCA is the strategy of dimensionality reduction that is likely to be chosen to classify leukemia because it demonstrates better or equal effectiveness with reduced computational power and cost.

4.4 Effect of Class Imbalance Handling with ADASYN

Class imbalance is a commonly occurring problem in medical datasets in which disease sample commonly forms the minority group. Models trained using imbalanced data are likely to favor the majority class and still produce large overall accuracy and do poorly on the important minority class. In the case of the leukemia diagnosis, the inability to diagnose the diseased samples (false negatives) has disastrous clinical implications which requires measures to correct imbalance.

The C-NMC Leukemia data has moderate imbalanced classes in the training data, which was used in 5-fold cross-validation. The distribution across folds has different proportions of ALL (majority) and HEM (minority) samples with the imbalance ratios slightly different in each fold as a result of stratified splitting. To overcome such imbalance, Adaptive Synthetic Sampling (ADASYN) algorithm is used.

In order to empirically illustrate the effect of ADASYN, controlled experiments are performed with the optimal configuration of ViT-L16 feature extractor and PCA dimensionality reduction to 1024 components and MLP classifier that have already been developed. There are two training scenarios, one of which involves direct training on unbalanced data without resampling, and the other where the data is balanced with the ADASYN. All other hyperparameters and experimental settings are kept the same in order to isolate the effect of ADASYN.

Table 4.3: Performance comparison of MLP classifier before and after ADASYN

Metrics	With ADASYN (MLP)	Without ADASYN	Difference
Accuracy (%)	91.89	73.42	18.47
Precision (%)	89.50	71.10	18.40
Recall (%)	84.34	69.58	14.76
Specificity (%)	95.40	78.25	17.15
F1 Score (%)	86.84	70.32	16.52
MCC (%)	81.06	55.84	25.22
AUC (%)	96.24	79.10	17.14

The complete visualization of the effects of ADASYN is presented in Figure 4.3. The distribution of the classes is initially presented before the application of ADASYN in a representative fold and indicates the starting highlight of the lack of balance. Implementing ADASYN solved the issues as the model is not showing biasness towards the minority HEM class and the performance of the overall model improved a lot.

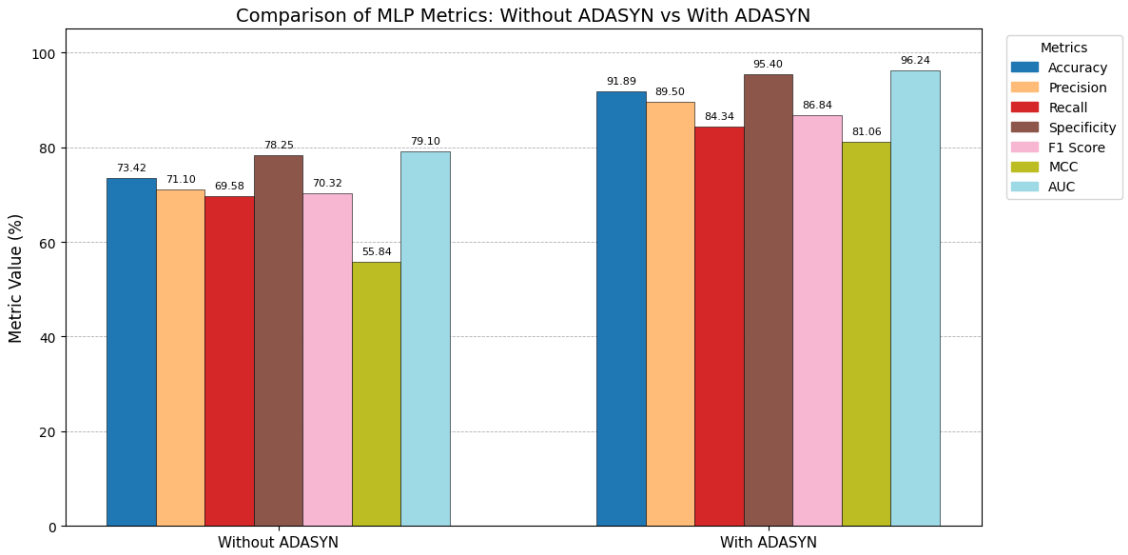


Figure 4.3: Model's performance comparison (before and after ADASYN)

The results demonstrate ADASYN's significant positive impact on model performance, particularly for the minority class (HEM) detection. Specificity improves dramatically from 78.25% to 95.40%, representing a 17.15 percentage point gain. This substantial improvement indicates that the model correctly identifies significantly more HEM cases when trained on balanced data, directly addressing the critical clinical requirement of minimizing false positives (misdiagnosing a healthy cell as ALL).

F1-Score increases from 70.32% to 86.84%, reflecting improved balance between precision and recall. MCC improves by 25.22 percentage points (from 55.84% to 81.06%), indicating enhanced overall classification quality accounting for all confusion matrix elements. AUC increases from 79.10% to 96.24%, demonstrating improved discriminative ability across all classification thresholds.

As every other metrics, Recall increased from 69.58% to 84.34% (14.76 percentage point increment). These minor decreases represent acceptable tradeoffs, as the model shifts its focus to the minority class, generating marginally more false positives while substantially improving the safety of HEM predictions.

These results confirm the use of ADASYN as fundamental element of the classification process. The gains in recall and improved MCC with only the significant cost in other metrics warrants its computational cost. Each further experiment uses ADASYN balancing to be able to detect minority classes.

4.5 Classifier Selection

Despite the fact that the sections above have reached the most appropriate decision concerning the feature extraction and preprocessing components, it is necessary to mention that the adequate classifier architecture is also a key to a good system performance. Although the neural network classifiers are flexible and capable of learning, it must be ensured that the MLP structure applied in them does enhance performance compared to other machine learning techniques. The given section involves the methodical comparison of the given MLP classifier against the four standard

classification algorithms, which include Support Vector Machine (SVM), Random Forest, XGBoost and Logistic Regression.

All of the classifiers are trained on the same conditions with the most successful feature representation identified above: ViT-L16 features with PCA dimensionality reduction to 1024 components, and it is well-balanced with ADASYN. It is a controlled comparison, and so the variations in performance of classifiers are separated by maintaining the same input data and preprocessing the same data on all the methods. Each of the classifiers is initialized using a hand tuned set of hyper parameters which are appropriate to a binary model of the 1024-dimensional feature vectors.

The suggested MLP model is a 1- hidden layer with 1000 neurons and ReLU activation, L2 regularization with $\lambda=1e^{-4}$ to prevent overfitting and sigmoid output activation in binary classification. This model is trained using Adam optimizer (learning rate $1e^{-3}$) for 200-epoches, with early stopping using validation performance. This architecture is able to balance the model capacity to the computational efficiency and also has high representational power.

SVM using Radial Basis Function (RBF) kernel is set using C=1.0 and gamma="scale" which yields nonlinear decision boundaries that are appropriate in complex feature space. Random Forest uses 200 decision trees with default parameters and uses the power of ensemble learning to bring together a number of weak learners. XGBoost is trained on 200 rounds of boosting where the learning rate was 0.1 and gradient boosting was applied to correct the errors. Logistic Regression with L2 regularization (C=1.0) is an example of a linear baseline, which is the simplest reasonable method of high-dimensional binary classification.

Table 4.4: Comparison of metrics to identify the ideal classifier

Classifier	Accuracy	Precision	Recall	Specificity	F1 Score	MCC	AUC
MLP	91.89	89.50	84.34	95.40	86.84	81.06	96.24
XGBoost	86.54	96.72	82.79	93.94	89.32	72.67	95.50
RandomForest	84.81	96.17	80.98	93.06	87.92	69.70	95.00
SMV (Linear)	81.81	94.35	78.02	89.96	85.41	63.81	93.50
Logistic Regression	79.14	93.09	75.00	88.04	83.07	58.98	92.80

Figure 4.4 provides a clear visualization of classifier comparison. It delivers references to a grouped bar chart that compares four major metrics in each of the classifiers (Accuracy, Precision, Recall, F1-Score, Specificity, MCC, AUC) of all classifiers and demonstrates that MLP is always ahead in its measures.

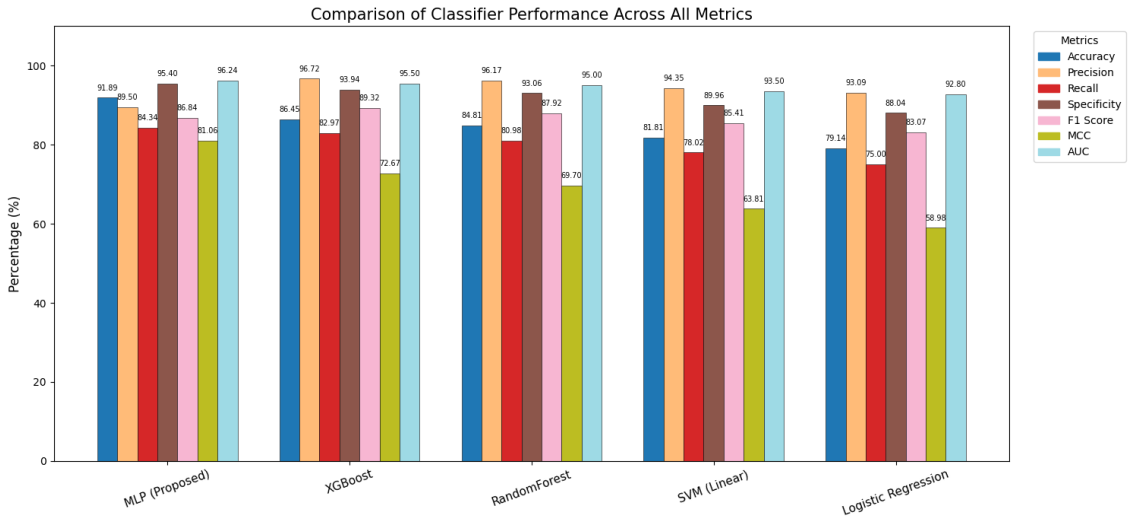


Figure 4.4: Performance comparison of classifiers

The proposed MLP classifier is confirmed to be the best option in the classification of the leukemia through the experiments. MLP is the top performer in all of the critical metrics: it has an AUC of 96.24% which is 0.74 percentage points higher than the second-best XGBoost (95.50%) and an Accuracy of 91.89% which is 5.44 percentage points higher than the second-best XGBoost (86.45%). Crucially, the MCC is 81.06%, which is 8.39 percentage points higher than the second-best XGBoost (72.67%). These may seem small gains but in a high-performance regime, where marginal gains grow progressively harder, they are significant gains.

XGBoost is the second-best classifier with AUC of 95.50% and Accuracy of 86.45% when compared to other classifiers. The ensemble booster method is effective in managing the high-dimensional feature space and nonlinear relationships. SVM using RBF kernel has a similar performance (AUC 93.50%) that is based on the usage of kernel-based techniques that are used to project the features into the space where the linear separation becomes possible. Random Forest comes next (AUC 95.00%), but its result does indicate a possibility of the simple voting of decision trees being not enough to achieve the maximum performance of the complex features representations. The performance of Logistic Regression is not impressive (AUC 92.80%), but it still reaches a respectable level, which means that the features after PCA transformation retain a high level of linear separability.

The excellent performance of the MLP can be explained by a number of architectural strengths. The 1000-neuron hidden layer has enough storage to be able to pick up nonlinear transformations of the 1024-dimensional input features to identify complex patterns that more simple models do not see. Activation functions in ReLU inject nonlinearity that allows representation of more complex decision boundaries whereas L2 regularization deters any overfitting even though the parameter space is enormous. The Adam optimizer using adaptive learning rates can provide efficient convergence to the high quality local minima.

The MLP also has the advantage over SVM of end-to-end gradient-based optimization to jointly learn transformations of the features and decision boundaries and does not use a fixed set of kernel functions. The trained MLP has more continuous differentiable

transformations that allow a smoother generalization and an improved interpolation between the training samples than tree based ensembles such as the Random Forest and the XGBoost. The hidden layer has important nonlinear modeling ability that logistic regression does not have when it comes to complicated medical image characteristics.

All these overall comparisons confirm that the proposed MLP architecture is the best classifier to leukemia diagnosis on ViT-L16 classes. The outstanding performance of the architecture in several metrics, and the moderate level of calculations make the architectural choice explainable. The MLP efficiently converts quality visual descriptors to diagnostic predictions which are accurate, and this completes the end-to-end classification pipeline.

4.6 Visualization and Model Interpretability

The overall final selected model has been subjected to explainable artificial intelligence (XAI) techniques in order to comprehend the areas with the greatest impact on ALL and HEM classification. Transformer-based attention visualization was employed as MLP-ViT-L16-PCA(1024) was found to be the best performing model.

Attention rollout maps were produced to analyse the self-attention distribution across image patches These maps consistently highlighted diagnostically relevant morphological structures, such as nuclear irregularities, chromatin density patterns, and cytoplasmic distortions—regions that are typically associated with leukemic abnormalities. A set of visualization outputs is given that illustrates these interpretative insights.

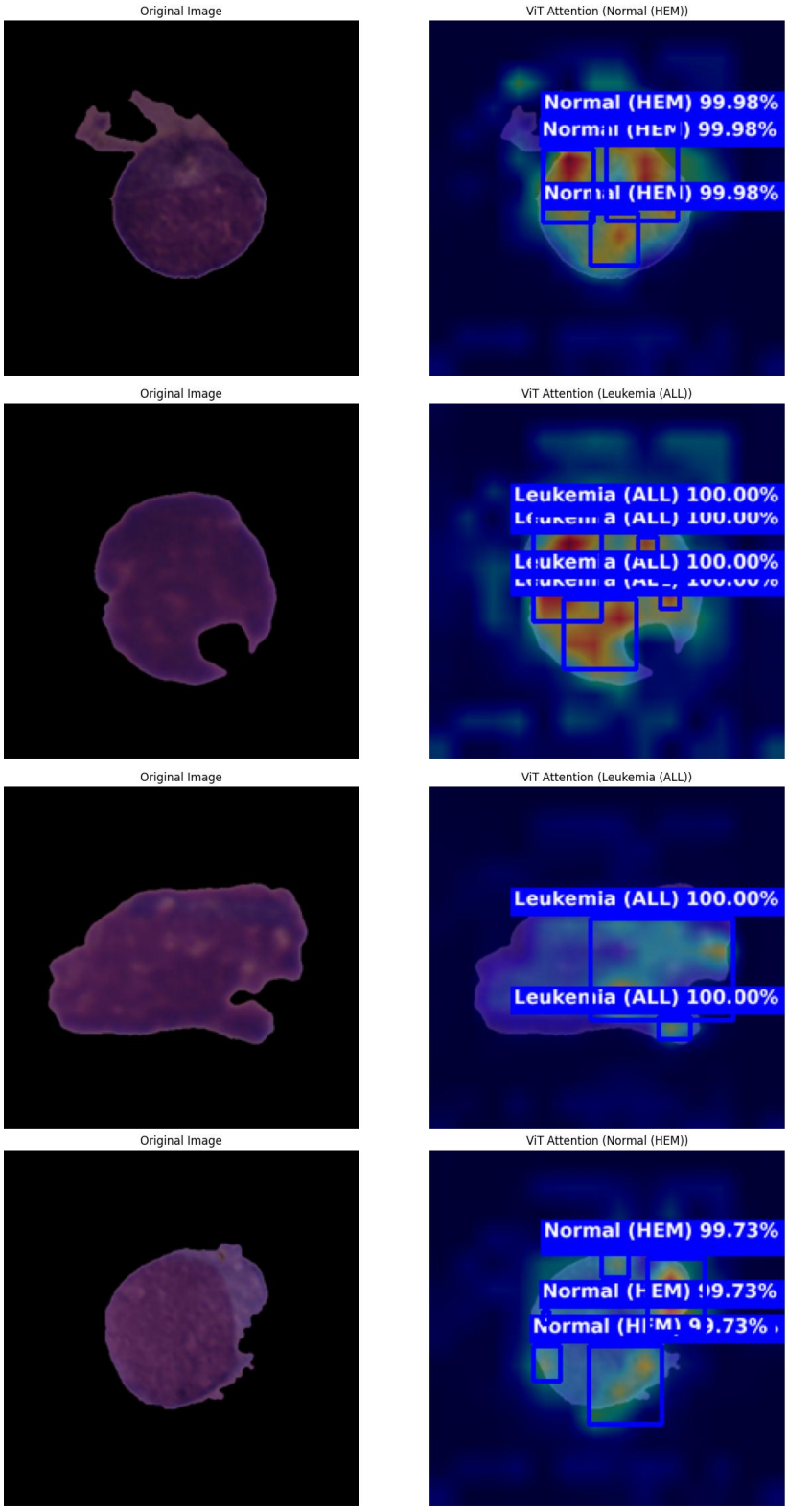


Figure 4.5: Visualising output using ViT attention map

4.7 Result Summary

The chapter was a thorough comparison of 70 experimental settings of CNN-based and ViT-based feature extractors with PCA and RFE dimensionality reduction. PCA always performed better than RFE on the majority of extractors, which demonstrates the significance of the preservation of structures of global variances to downstream classification.

DenseNet121-PCA(700) was the CNN with the best overall performance which corresponds to AUC of 92.09. The models that were based on transformers were however, more generally better at their discriminative ability. The best-performing feature extractor was ViT-L16-PCA(1024) with the best AUC of 96.24, which is much better than any CNN-based model. In addition, the MLP classifier based on these features was superior to XGBoost, Logistic Regression, SVM, and Random Forest in all essential measures, which proved its superiority in the system.

The analysis of interpretability based on ViT attention maps also validated that the model was consistently focused on medically meaningful features, such as nuclear irregularities, chromatin density, and nuclear-to-cytoplasmic ratios. All these findings collectively show that the use of transformer-based feature extraction combined with PCA dimensionality reduction is the most effective framework for MLP-based leukemia classification as witnessed in this study.

CHAPTER 5

CONCLUSION

5.1 Findings and Conclusion

The study has performed a comprehensive comparison of deep learning-based feature extraction methods of automated classification of acute lymphoblastic leukemia (ALL) and health (HEM) cell image. CNN architectures (ResNet50, DenseNet121, InceptionV3, Xception, EfficientNetB0), and two Vision Transformer models (ViT-B16 and ViT-L16) were trained with two dimensionality reduction algorithms (PCA and RFE) on five different feature sizes, which makes the total number of experimental configurations to seventy. It was aimed to determine the optimal combination of feature extractor and dimensionality reduction approach to downstream MLP classification, which was selected after comparing with other regular classifiers.

The results demonstrate that there are a number of definite trends. First, PCA achieved superiority to RFE in majority of the settings. This implies that retaining global variance structures is more beneficial than supervised feature elimination when dealing with high-dimensional deep features. Second, while CNN-based extractors were able to achieve good performance, especially DenseNet121-PCA(700) with Vision Transformers demonstrated a clear performance advantage. ViT-L16-PCA(1024) has received the best AUC score of (96.24) along with strong recall, F1, and MCC values, establishing it as the most reliable and discriminative model in this study. This highlights the effectiveness of transformer-based global attention mechanisms in detecting fine morphological differences that exist in the images of leukocytes.

These findings were also supported by the XAI analysis which consistently indicated medically meaningful regions of ViT attention visualization, which included nuclear contours and chromatin distribution. Overall, the work shows that a combination of ViT-based features extractor with PCA provides a very precise, interpretable and invariant

system of leukemia cell classification. The most promising model to build a set of automated diagnostic tools in the field of hematological imaging is the final one MLP-ViT-L16-PCA(1024).

5.2 Limitations of The Study

Despite the proposed models performing well, there are a number of limitations that need to be acknowledged. To begin with, the used dataset, as common as it is in research, might not fully represent the diversity of actual clinical situations. The differences in staining patterns, radiography equipment and patient-specific factors may introduce variation, which is not reflected in the dataset. Consequently, external generalizability cannot be clear without subsequent confirmation. Second, the fact that larger transformer architectures, more sophisticated self-supervised pretraining and hyperparameter optimization were not explored was due to limited computational resources. Also, dimensionality reduction, although PCA has been used, could have caused certain loss of discriminative data. Lastly, this paper only considered binary classification and in the real clinical environment, it is usually needed to consider multi-class or subtype-level analysis which was outside the scope of this paper.

5.3 Recommendations for Future Work

The research findings of this study can be extended to create a series of research directions in the future which are important in numerous ways. To start with, extending the dataset to a larger variety of diversity and clinically obtained images would help improve model generalizability and validate performance across different imaging conditions. Larger multi-institutional datasets or data on multiple cell types would give a stronger evaluation framework.

Secondly, future work could explore self-supervised learning, contrastive learning approaches, or hybrid CNN–ViT architectures to further enhance feature representation. Additional models such as DINO, MoCo or higher variant forms of ViTs can capture more detailed morphological attributes, especially in low-data settings. Moreover, other

methods of dimensionality reduction, including autoencoders or nonlinear manifold learning, should also be tested as they could achieve better performance than PCA.

Thirdly, the implementation of the model in clinical practice would need the construction of an entire pipeline, including image preprocessing, real-time inference, uncertainty estimation, and more convenient visualization. Lastly, it would improve the clinical applicability of the system enormously by extending the classification task to cover more than just a single subtype of leukemia, or other hematological conditions.

5.4 Closing Remarks

This thesis has demonstrated that the feature extraction with the help of Vision Transformer, especially in combination with PCA, is a strong and effective method of classifying leukemia cells. The results support the importance of a sensible integration of the modern deep learning architecture strategy with carefully selected dimensionality reduction strategies to unlock the high diagnostic performance. Although the issues of the diversity of datasets, relevance of computational scale, and clinical implementation still persist, the outcomes provide a good basis to continue the improvement of automated hematological diagnosis in the future. The overall outcome of this research highlights both the promise and the responsibility of applying artificial intelligence in medical decision support systems, emphasizing the need for continued refinement, validation, and collaboration with clinical experts.

REFERENCES

1. Hasan Abir, W., Fahim Uddin, M., Rahman Khanam, F., & Monirujjaman Khan, M. (2023). Explainable AI in Diagnosing and Anticipating Leukemia Using Transfer Learning Method. arXiv e-prints, arXiv-2312. https://doi.org/10.48550/arXiv.2312.00487
2. American Cancer Society. (2025, June 30). Key statistics about acute lymphocytic leukemia (ALL). https://www.cancer.org/cancer/types/acute-lymphocytic-leukemia/about/key-statistics.html
3. Asar, T. O., & Ragab, M. (2024). Leukemia detection and classification using computer-aided diagnosis system with falcon optimization algorithm and deep learning. *Scientific Reports*, *14*(1), 21755. https://doi.org/10.1038/s41598-024-72900-3
4. Tanwar, V., Sharma, B., Yadav, D. P., & Dwivedi, A. D. (2025). Enhancing blood cell diagnosis using hybrid residual and dual block transformer network. *Bioengineering*, *12*(2),98. https://doi.org/10.3390/bioengineering12020098
5. Ben-Suliman, K., & Krzyżak, A. (2024, September). Vision Transformer Features-Based Leukemia Classification. In *IAPR Workshop on Artificial Neural Networks in Pattern Recognition* (pp. 111-120). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-71602-7_10
6. Rodriguez, J. R., Fernandez, S., Swartz, N., Alonge, A., Bhullar, F., Betros, T., ... & Jacobs, R. J. (2024). A Chronological Overview of Using Deep Learning for Leukemia Detection: A Scoping Review. *Cureus*, *16*(5). DOI: 10.7759/cureus.61379
7. Oybek Kizi, R. F., Theodore Armand, T. P., & Kim, H. C. (2025). A review of deep learning techniques for leukemia cancer classification based on blood smear images. Applied Biosciences, 4(1),9. https://doi.org/10.3390/applbiosci4010009
8. Ramesh, G., & Thouti, S. (2024). Study of Machine Learning Algorithms on Early Detection of Leukemia. In *E3S Web of Conferences* (Vol. 472, p. 03013). EDP Sciences. DOI: 10.1051/e3sconf/202447203013
9. Rai, H. M., Omkar Lakshmi Jagan, B., Rao, N. T., Mohammed, T. K., Agarwal, N., Abdallah, H. A., & Agarwal, S. (2025). Deep Learning for Leukemia Classification: Performance Analysis and Challenges Across Multiple Architectures. *Fractal and Fractional*, *9*(6), 337. https://doi.org/10.3390/fractalfract9060337
10. Mantri Paswan, R. ., Abdul Hamid Khan, R. ., Jadhav, S. ., Borundiya, V. ., & Dhondge, P. . (2024). Efficient Diagnosis of Acute Lymphoblastic Leukemia using Transfer Learning. International Journal of Intelligent Systems and Applications in Engineering, 12(19s), 667–677. https://ijisae.org/index.php/IJISAE/article/view/5111
11. Haque, R., Al Sakib, A., Hossain, M. F., Islam, F., Ibne Aziz, F., Ahmed, M. R., Kannan, S., Rohan, A., & Hasan, M. J. (2024). Advancing Early Leukemia Diagnostics: A Comprehensive Study Incorporating Image Processing and Transfer Learning. *BioMedInformatics*, *4*(2), 966-991.

https://doi.org/10.3390/biomedinformatics4020054

12. Makem, M., Tamas, L., & Busoniu, L. (2025). A Reliable Approach for Identifying Acute Lymphoblastic Leukaemia in Microscopic Imaging. Frontiers in Artificial Intelligence, 8, 1620252. https://doi.org/10.3389/frai.2025.1620252
13. Hasanaath, A. A., Mohammed, A. S., Latif, G., Abdelhamid, S. E., Alghazo, J., & Hussain, A. A. (2024). Acute lymphoblastic leukemia detection using ensemble features from multiple deep CNN models. Electronic Research Archive, 32(4). DOI: 10.3934/era.2024110
14. Abhishek, A., Deb, S. D., Jha, R. K., Sinha, R., & Jha, K. (2025). Ensemble learning using Gompertz function for leukemia classification. *Biomedical Signal Processing and Control*, *100*, 106925. https://doi.org/10.1016/j.bspc.2024.106925
15. Muhammad, D., Salman, M., Keles, A., & Bendechache, M. (2025). ALL diagnosis: can efficiency and transparency coexist? An explainble deep learning approach. Scientific Reports, 15(1), 12812. DOI: 10.1038/s41598-025-97297-5
16. Huang, M. L., & Huang, Z. B. (2024). An ensemble-acute lymphoblastic leukemia model for acute lymphoblastic leukemia image classification. *Mathematical Biosciences and Engineering*, *21*(2024), 1959-1978. DOI: 10.3934/mbe.2024087
17. Abd El-Aziz, A. A., Mahmood, M. A., & Abd El-Ghany, S. (2024). A Robust EfficientNetV2-S Classifier for Predicting Acute Lymphoblastic Leukemia Based on Cross Validation. *Symmetry*, *17*(1), 24. https://doi.org/10.3390/sym17010024
18. Jawahar, M., Anbarasi, L. J., Narayanan, S., & Gandomi, A. H. (2024). An attention-based deep learning for acute lymphoblastic leukemia classification. *Scientific Reports*, *14*(1), 17447. https://doi.org/10.1038/s41598-024-67826-9
19. Alim, M. S., Bappon, S. D., Sabuj, S. M., Islam, M. J., Tarek, M. M., Azam, M. S., & Islam, M. M. (2024). Integrating convolutional neural networks for microscopic image analysis in acute lymphoblastic leukemia classification: A deep learning approach for enhanced diagnostic precision. *Systems and soft computing*, *6*, 200121. https://doi.org/10.1016/j.sasc.2024.200121
20. Jammal, F., Dahab, M., & Bayahya, A. Y. (2025). Neuro-bridge-X: A Neuro-symbolic vision transformer with meta-XAI for interpretable leukemia diagnosis from peripheral blood smears. Diagnostics, 15(16), 2040. DOI: 10.3390/diagnostics15162040
21. Preethika, P., & Ananthajothi, K. (2025, June). Hybrid Vision Transformer and CNN for Leukemia Detection in Blood Smear Images. In *2025 11th International Conference on Communication and Signal Processing (ICCSP)* (pp. 1344-1349). IEEE. DOI: 10.1109/ICCSP64183.2025.11088481
22. Jia, Y., Pei, H., Xu, D., Pang, J., Chen, Y., Shen, C., & Xiang, M. (2026). Deep learning with explainability: Improving diagnostic accuracy and interpretability in ECG and MCG analysis. Biomedical Signal Processing and Control, 113, 108875. https://doi.org/10.1016/j.bspc.2025.108875

23. Islam, O., Assaduzzaman, M., & Hasan, M. Z. (2024). An explainable AI-based blood cell classification using optimized convolutional neural network. *Journal of Pathology Informatics*, *15*, 100389. https://doi.org/10.1016/j.jpi.2024.100389
24. Mittal, P., Sharma, B., & Yadav, D. P. (2024, December). Comparative Analysis between CNN and ViT using Brain MRI Dataset. In *2024 Eighth International Conference on Parallel, Distributed and Grid Computing (PDGC)* (pp. 290-295). IEEE. DOI: 10.1109/PDGC64653.2024.10984339
25. Sharma, A. K., Nandal, A., Dhaka, A., Zhou, L., Alhudhaif, A., Alenezi, F., & Polat, K. (2023). Brain tumor classification using the modified ResNet50 model based on transfer learning. *Biomedical Signal Processing and Control*, *86*, 105299. https://doi.org/10.1016/j.bspc.2023.105299Get rights and content
26. Arulananth, T. S., Prakash, S. W., Ayyasamy, R. K., Kavitha, V. P., Kuppusamy, P. G., & Chinnasamy, P. (2024). Classification of paediatric pneumonia using modified DenseNet-121 deep-learning model. *IEEE Access*, *12*, 35716-35727. DOI: 10.1109/ACCESS.2024.3371151
27. Shah, S. R., Qadri, S., Bibi, H., Shah, S. M. W., Sharif, M. I., & Marinello, F. (2023). Comparing inception V3, VGG 16, VGG 19, CNN, and ResNet 50: a case study on early detection of a rice disease. *Agronomy*, *13*(6), 1633. https://doi.org/10.3390/agronomy13061633
28. Joshi, S. A., Bongale, A. M., Olsson, P. O., Urolagin, S., Dharrao, D., & Bongale, A. (2023). Enhanced Pre-Trained Xception Model Transfer Learned for Breast Cancer Detection. *Computation*, *11*(3), 59. https://doi.org/10.3390/computation11030059
29. Kanchana, K., Kavitha, S., Anoop, K. J., & Chinthamani, B. (2024). Enhancing skin cancer classification using efficient net b0-b7 through convolutional neural networks and transfer learning with patient-specific data. *Asian Pacific Journal of Cancer Prevention: APJCP*, *25*(5), 1795. DOI: 10.31557/APJCP.2024.25.5.1795
30. Reddy, C. K. K., Reddy, P. A., Janapati, H., Assiri, B., Shuaib, M., Alam, S., & Sheneamer, A. (2024). A fine-tuned vision transformer based enhanced multi-class brain tumor classification using MRI scan imagery. *Frontiers in oncology*, *14*, 1400341. DOI: 10.3389/fonc.2024.1400341
31. Dosovitskiy, A., Beyer, L., & Kolesnikov, A. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. June 3. *arXiv preprint arXiv:2010.11929*. https://doi.org/10.48550/arXiv.2010.11929
32. Bharadiya, J. P. (2023). A tutorial on principal component analysis for dimensionality reduction in machine learning. *International journal of innovative science and research technology*, *8*(5), 2028-2032. DOI: 10.5281/zenodo.8002436
33. Awad, M., & Fraihat, S. (2023). Recursive feature elimination with cross-validation with decision tree: Feature selection method for machine learning-based intrusion detection systems. *Journal of Sensor and Actuator Networks*, *12*(5), 67. https://doi.org/10.3390/jsan12050067
34. Imani, M., Beikmohammadi, A., & Arabnia, H. R. (2025). Comprehensive analysis of random forest and XGBoost performance with SMOTE, ADASYN,

and GNUS under varying imbalance levels. *Technologies*, *13*(3), 8888. https://doi.org/10.3390/technologies13030088

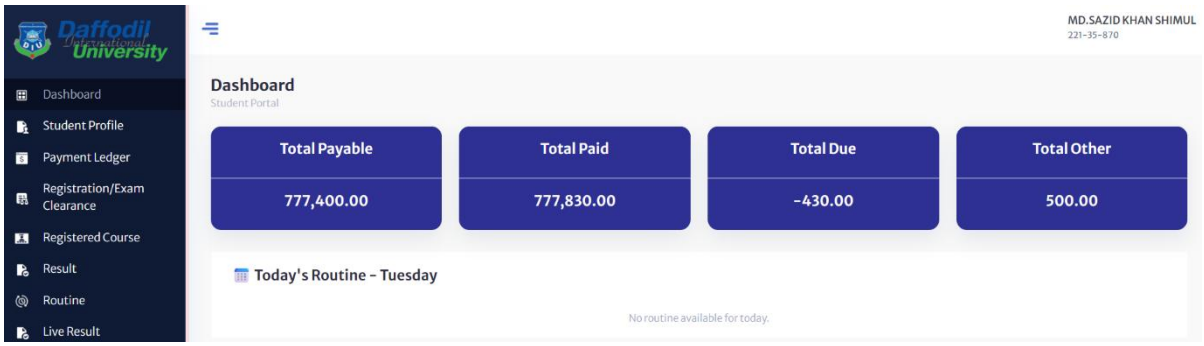
35. Safar, A. A., Salih, D. M., & Murshid, A. M. (2023). Pattern recognition using the multi-layer perceptron (MLP) for medical disease: A survey. *International Journal of Nonlinear Analysis and Applications*, *14*(1), 1989-1998. DOI: 10.22075/ijnaa.2022.7114
36. Szeghalmy, S., & Fazekas, A. (2023). A comparative study of the use of stratified cross-validation and distribution-balanced stratified cross-validation in imbalanced learning. *Sensors*, *23*(4), 2333. https://doi.org/10.3390/s23042333

APPENDICES

Appendix A: Availability of The Used Dataset

Access Link: https://www.cancerimagingarchive.net/collection/c-nmc-2019/

Student Portal Clearance



Library Clearance

Plagiarism Report

221-35-870

ORIGINALITY REPORT

24%	21%	17%	15%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	3%
2	Submitted to Midlands State University Student Paper	2%
3	Submitted to Daffodil International University Student Paper	2%
4	arxiv.org Internet Source	1%
5	journal.yrpipku.com Internet Source	<1%