Deep Learning approach for Automated Blood Cancer Cells detection

By
Shuvo Halder
212-15-4145

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised by

Dr. Sheak Rashed Haider Noori
Professor & Head
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by

Showmic Guha Paul
Lecturer
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

May 14, 2025

APPROVAL

This Project titled "**Deep Learning approach for Automated Blood Cancer Cell detection**," submitted by **Shuvo Halder** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **14 May, 2025**.

BOARD OF EXAMINERS

Dr. S.M Aminul Haque (SMAH) — **Chairman**
Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Mohammad Jahangir Alam (MJA) — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Mr. Md Mohammad Masum Bakaul (MB) — **Internal Examiner**
Sr. Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Arshad Ali (DAA) — **External Examiner**
Professor
Department of Computer Science and Engineering
Hajee Mohammad Danesh Science & Technology University

DECLARATION

We hereby declare that this project has been done by us under the supervision of Dr. Sheak Rashed Haider Noori, Professor & Head, Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Dr. Sheak Rashed Haider Noori
Professor & Head
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by:

Showmic Guha Paul
Lecturer
Department of Computer Science and Engineering
Daffodil International University

Submitted by:

Shuvo Halder

Shuvo Halder
Student ID: 212-15-4145
Department of Computer Science and Engineering
Daffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project (FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Dr. Sheak Rashed Haider Noori**, **Professor & Head,** Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of "Data Science" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting themat all stages have made it possible to complete this project.

We would like to express our heartfelt gratitude to **Dr. Sheak Rashed Haider Noori**, **Professor & Head,** Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

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

ABSTRACT

A deep learning approach to automatic blood cancer diagnosis is proposed in this thesis. The aim of this research is to develop an efficient system for accurate early-stage diagnosis and thereby improve healthcare outcomes. The dataset of blood smear images was preprocessed using techniques such as noise removal, contrast stretching, and normalization to improve the feature extraction and model training process. Four deep learning models—Xception, InceptionV3, MobileNet, and ResNet50—were attempted. The best among them was InceptionV3 with 98%, followed by Xception and MobileNet with 97%. To achieve higher performance, two hybrid models were attempted: Hybrid Model 1, a combination of Xception, InceptionV3, and MobileNet, which resulted in 99%, and Hybrid Model 2, a combination of ResNet50 and VGG16, which resulted in 93%.These results underscore the significance of model architecture selection and preprocessing for accurate classification. The findings suggest that AI technology can greatly contribute towards the accuracy and timeliness of the diagnosis of leukemia, especially in poor-resource environments. The study shows the potential of deep learning algorithms and, more so, hybrid models for providing accurate, scalable, and efficient blood cancer detection for the final good of better clinical decision-making and better outcomes for patients.

List of Figures

List of Tables

Table of Contents

Chapter 1

Introduction

1.1 Introduction

Blood cancer is a complicated and heterogeneous disease involving bone marrow, lymphatic system, and blood. Leukemia is an active and complicated disease involving bone marrow and blood, having an excess of abnormal white cells [1]. Diagnosis of early-stage leukemia correctly is critical for proper management and healing of the disease. Abnormal cells disrupt immune system functioning and impede the role of the body in producing normal red blood cells and platelet formation [2]. Proper management and healing of the disease depend on early and precise diagnosis. The method that historically has been used to diagnose leukemia is slow and laborious manual scanning of blood slides by qualified hematologists. The recent development in medical image processing and machine learning made automated systems possible for detecting and classifying leukemia cells in blood [3] [4]. Leukemia cell segmentation in blood smear images is a crucial part of automatic diagnosis, and by using it, the abnormal cells may be identified and described accurately [5]. Manual diagnosis by microscopic examination is a labor-intensive and skill-based process that cannot easily expand at a high rate, particularly in areas with a low availability of qualified hematologists. This use of manual examination introduces potential delays in diagnosis, a greater risk for human error, and reduced accessibility to early-stage detection and intervention. Aside from this, differences in cell morphology and image consistency can also become major obstacles toward stable and precise diagnosis. The key goal of this project is to create a strong, computerized deep learning system able to segment and classify cells from microscopic images with high accuracy, determining with certainty whether or not cancer cells are present in the sample. By making image uploading and real-time classification accessible through an online interface, the system aims to narrow the gap between complicated medicine diagnosis and practical, available healthcare.

1.2 Motivation

The motivation behind carrying out this research is driven by the need to enhance diagnostic efficacy and accuracy in cancer detection at an early stage, particularly leukemia. As a potentially fatal hematological cancer, early and precise diagnosis is required for cancer management planning to reduce mortality rates among cancer patients.

The manual check by qualified hematologists of smear slides is still a labor-intensive, human-error-prone, and tedious traditional process. The shortage of available qualified personnel in most regions exacerbates the issue, as diagnosis and therapy become delayed. The rapid development of computational technologies, particularly in computer vision and deep learning, holds promising solutions to these challenges. Convolutional Neural Networks (CNNs) and other cognate architectures have exhibited great potential for medical image analysis, such as segmentation and classification. Leveraging these capabilities to identify blood cancer automatically can not only offload such diagnosis on the healthcare providers, but also provide consistent and reliable outcomes even in low-resource environments. At an academic and career level, this thesis has potential to link theory with practice. In building and implementing an automatic machine for leukemic cell detection from micrographs, this thesis is contributing to efforts to introduce artificial intelligence into healthcare. It is also a great learning opportunity in building an end-to-end deep learning model from data preprocessing to modeling, evaluation, and deployment. In addressing this issue, there exists not only a strong health need, but also a way to fulfill the larger vision of utilizing intelligent systems for improved global health outcomes.

1.3 Objectives

Create a deep learning model to classify blood cancer cells automatically from microscopic images. Preprocess images to enhance quality by using techniques such as noise removal, contrast stretching, and segmentation. Prepare and test deep learning models like CNNs, InceptionV3, Xception, MobileNet for classification. Hyperparameter tune and do data augmentation with transfer learning to achieve the highest accuracy. Compare different deep learning architectures to find the ideal model for the detection of blood cancer. Introduce AI in healthcare by integrating deep learning with the diagnosis of blood cancer to boost efficiency and accuracy.

1.4 Methodology

The research methodology here is centered on creating an automated system based on deep learning for segmenting and classifying leukemia cells from blood smear images. The system uses contemporary image processing methods integrated with convolutional neural networks (CNNs) to identify abnormal cells and classify them, thereby facilitating early diagnosis of leukemia. The major dataset used in this research is images of blood smears drawn from an openly available repository of medical images. The dataset is classified into five major categories based on different types of cells: myeloblast, monocyte, segmented neutrophil, basophil, and erythroblast. These cells are most vital in determining whether there is leukemia and also its stages. The images are kept in folders based on their respective class labels. All images within the dataset have an assigned label describing the class of the blood cell illustrated. The dataset is hosted on Google Drive and is accessed programmatically through Python libraries. The data collection is accomplished by loading images and correlating them with their respective labels to use in the machine learning pipeline. Heavy preprocessing is done before model training to pre-process the

images for deep learning analysis. The dataset is loaded by iterating over the subdirectories of the image files. Each image is assigned a class label by the folder name in which it is located (e.g., "myeloblast" for myeloblast cells). In an effort to standardize all the images, the images are normalized. It includes brightness and contrast correction for bringing about uniform pixel intensity values and hence reducing the variability in image lighting conditions. Images are resized to a fixed dimension so that they can be input into the deep learning model. Target sizes are selected to be computationally efficient for speed, depending on the particular neural network architecture. One of the methodology steps is to construct the dataset mapping of image file paths to their corresponding labels. This is done by traversing the blood cell image directories and loading the file paths and corresponding class labels into a Pandas DataFrame. The DataFrame serves as input to the model during training and validation processes. Convolutional Neural Network (CNN) is the major component of the system that classifies and segments. CNN is selected due to its capability for hierarchical feature extraction from images required for image-based classification tasks like leukemia. Fully connected layers are applied for cell classification following feature extraction by the convolutional layers. Apart from experimenting with specially crafted CNN architecture, the paper also investigates using transfer learning with pre-trained networks (e.g., ResNet50, VGG16). In doing so, the model will utilize pre-learned large data features to speed up training and perform well in small, domain-knowledge data like the blood smear images. The network is trained in the supervised learning setup where the network is given the input images and their corresponding class labels. Data Splitting: The database is divided into training, validation, and test sets. Training set is utilized for training the model, validation set for hyperparameter tuning, and test set for final model evaluation. Different performance measures are utilized to assess the performance of the model in accuracy, precision, recall, and F1-score. These are measures that provide an estimate of the ability of the model in correct classification of cells and minimization of false negatives/positives. The model is validated on the validation dataset following training. This is to validate the model performance and whether it doesn't overfit the training data. It is then evaluated on the unseen test data set after model tuning to identify its final performance.

1.5 Project Outcome

The implication of your study in the detection of blood cancer cells can be far-reaching, both in practice and scientifically. The following are some of the likely implications. Improved Identification: Using image processing or machine learning algorithms, your study can lead to improved identification and categorization of blood cancer cells from clinical scans with complete human-free intervention, reducing the scope for human error and optimizing diagnostic accuracy. Early identification of cancer cells from blood can benefit greatly in patient prognosis by encouraging early treatment and increased survival rates. Your model could potentially automate the detection of cancer cells in blood samples with fewer pathologists required to manually scan each slide. Automation would also significantly reduce the time it takes to diagnose, allowing healthcare workers to perform more samples within a

shorter period of time. Access to remotely examine cancer cells in blood using AI models may allow for the provision of medical aid to remote or underserved areas where skilled pathologists cannot be conveniently reached. By creating more automated and fewer expert-dependent devices, diagnostic processes can be minimized in terms of cost and extended to more individuals. What your project would be contributing to is the research literature on the application of AI and machine learning in healthcare, particularly in oncology and medical imaging. Analysis of large image databases of blood cancers can reveal patterns never seen in cancer cell morphology previously, to bring more insight and classification of different kinds of blood cancers. If your model is successful in clinical trials, it could be implemented in hospital systems and become a device to help doctors diagnose blood cancers more quickly and effectively. Publishing your findings can add to the body of knowledge in medical imaging and cancer diagnosis that can educate others in the field of oncology, bioinformatics, and data science.

1.6 Organization of the Report

This thesis report consists of six main chapters, each of which is designed to introduce the research work on blood cancer cell detection using deep learning models logically and sequentially. This chapter presents the context of leukemia and blood cancer, noting the complexity of the diagnosis process since current diagnostic methods are manual. It provides the necessity for early and proper detection, limitations of manual microscopy-based diagnosis, and importance of deep learning and medical image processing. The second chapter describes the motivation behind the research. It highlights the need to improve diagnostic speed and precision through automation of the process with the driving force of issues like insufficiently skilled hematologists, human error in manual examination, and the slow process of traditional methods. It further identifies the utopian prospect of deep learning solutions, especially Convolutional Neural Networks (CNNs), to transform medical diagnostic processes. This chapter explicitly defines the research objectives. It highlights the design of a deep learning model for automatic classification of blood cancer cells from microscope images, improving image quality through the use of preprocessing techniques, comparing different architectures such as CNNs, InceptionV3, Xception, and MobileNet, and using transfer learning to enhance model performance. Chapter four outlines the methodology used in the research. It outlines the steps of data collection from public datasets, preprocessing techniques employed (e.g., normalization, resizing, and augmentation), dataset construction, and the model architecture of CNN used. The chapter also outlines the training and evaluation procedure, model validation, hyperparameter optimization, and the use of transfer learning using pre-trained networks to boost performance. This chapter states the probable outcome of the study. It highlights the expected improvement in diagnostic accuracy, automation of microscopic examination, access to diagnostics in rural areas, and AI-driven healthcare progress contributions. It also discusses the way pattern detection and biomarker discovery can emerge from data-driven analysis and possible future advances in clinical use and educational impact.

Chapter 2

Background

2.1 Introduction

The field of medical diagnostics has been significantly developed with the application of artificial intelligence (AI) and deep learning algorithms, opening up new avenues for the early and accurate diagnosis of life-threatening illnesses such as cancer. Blood cancer, a heterogeneous group of hematopoietic system cancers, is a significant challenge to early detection due to its complex presentation and the subtle morphological alterations of blood cells. Of the blood cancers, leukemia is still one of the most deadly types characterized by the abnormal and excessive proliferation of white blood cells that disrupt normal immune and hematopoietic function. Leukemia diagnosis has traditionally been accomplished through microscopic analysis of stained smears from blood slides by skilled hematologists. Although the gold standard, it is a manual, time-consuming, and observer-skill-dependent process. These drawbacks frequently result in delayed diagnosis, high probabilities of human error, and low potential for scalability in resource-poor healthcare environments. Recent advances in deep learning, particularly in the area of Convolutional Neural Networks (CNNs), have demonstrated excellent potential in automating high-level image analysis tasks. CNNs have the ability to learn hierarchical features from medical images, enabling accurate segmentation and classification of various types of blood cells. Utilizing such models in the analysis of blood smear images presents a promising direction for automating leukemia detection, thereby avoiding delays in diagnosis and improving clinical outcomes. This paper is aimed at developing an AI-based system to detect blood cancer cells from microscopic images with high accuracy. The system involves preprocessing techniques to enhance image quality, training CNN models, and testing them on publicly available datasets. In an effort to make the diagnosis process automated, the system aims to increase the rate of speed, reliability, and availability of leukemia detection—of great utility in remote or low-resource healthcare environments.

2.2 Literature Review

Table 2.1: Summary of Literature Reviewed.

Author(s)	Year	Title	Methodology	Key Findings
Kumar et al. [1]	2020	Automatic Detection of White Blood Cancer from Bone Marrow Microscopic Images Using CNN	Deep Learning (CNN)	Achieved 97.2% accuracy; outperformed traditional ML models such as SVM, Decision Trees, and Random Forests. The study's limitation includes small datasets and lack of external validation.
Raja & Meenakshi[2]	2024	Leukemia Cancer Cells Segmentation and Classification using ML	Statistical image processing in MATLAB	Highlighted cell geometry change; lacked quantitative performance (accuracy, precision, recall). No external validation using a different dataset.
Das et al. [3]	2021	Segmentation and Classification of WBC Cancer Using Hybrid Mask-RCNN and Transfer Learning	Hybrid Deep Learning model (Mask R-CNN + Transfer Learning)	Achieved 97.13% accuracy and F1-score of 0.91. Dataset size was small, limiting generalizability.
Khashman & Al-Zgoul [4]	2010	Image Segmentation of Blood Cells in Leukemia Patients	Morphological analysis	Achieved 98.33% segmentation success; no deep learning methods applied and lacked classification of leukemia.
Shemona & Chellappan [5]	2020	Segmentation Techniques for Early Cancer Detection in RBCs	CNN-based classification	Accuracy of 97.1%; focused on red blood cells (RBCs), not white blood cells (WBCs) relevant for leukemia detection.
Agughasi [6]	2022	i-Net: A Deep CNN Model for White Blood Cancer Classification	Custom CNN with augmentation and normalization	Accuracy 97.1%; small dataset (3,102 images) affecting generalizability.

Zolfaghari & Sajedi [7]	2022	Survey on Automated Detection and Classification of Leukemia	Literature Review	Identified strengths of DL models but highlighted lack of large-scale external validation and use of small datasets.
Baydilli & Atila [8]	2020	Classification of WBCs Using Capsule Networks	Capsule Network	Achieved over 90% accuracy; limited dataset validation.
Paswan & Rathore [9]	2017	Blood Cancer Detection Using SVM, KNN, NN	Traditional ML (SVM, KNN, NN)	NN classifier achieved best accuracy of 95.2%; no large-scale testing or use of diverse datasets.
Baig et al. [10]	2022	Detecting Malignant Leukemia Cells Using Microscopic Images	Deep CNN (ResNet50, VGG16)	Achieved 97.8% accuracy; no cross-validation or testing on multiple datasets.
Kumar et al. [11]	2021	Hybrid CNN Model for Early Detection of Blood Cancer	Deep Learning (Hybrid CNN)	Achieved high accuracy in early-stage cancer detection, demonstrating the effectiveness of combining CNNs.
Tanaka & Yamada [12]	2019	Deep Learning in Leukemia Detection	CNN-based classification and segmentation	Provided 96.5% accuracy; limited by dataset size, but strong in terms of automated segmentation.
Zhang et al. [13]	2020	Data Augmentation in Leukemia Detection	Data Augmentation & CNN	Demonstrated enhanced model performance through image augmentation, leading to 94% accuracy.
Zhao et al.	2021	Leukemia Cell Detection Using Mask R-CNN and Transfer Learning	Mask R-CNN + Transfer Learning	Achieved 98.3% segmentation accuracy; dataset limitations hindered wider applicability.

2.2.1 Related Research

This paper proposes a deep learning-based method for the detection and classification of various types of white blood cancer from microscopic images of bone marrow. This work proposes a deep learning-based solution with a Dense Convolutional Neural Network for the classification of Acute Lymphoblastic Leukemia and Multiple Myeloma.t was tested on the SN- AM dataset and recorded 97.2% accuracy, higher than traditional machine learning algorithms such as SVM, Decision Trees, and Random Forests. The limitations of the study include small sizes of datasets, lack of external validation, and absence of clinical trials in the real world [1].

This paper presents a decision support system for the segmentation and classification of leukemia cells from microscopic blood smear images. The study uses MATLAB-based image processing techniques in the detection of leukemia using cell geometry change analysis (area, perimeter, mean, standard deviation). It utilizes statistical features in segmentation and classification. It suffers from the disadvantage of Using shallow machine learning techniques instead of deep learning. No quantitative performance measures (accuracy, precision, recall) are given. External validation using a different dataset is not performed [2].

This paper is dedicated to the segmentation of cancer cells from microscopic blood smear images. The authors propose a three-step approach: image pre-processing, cell segmentation, and feature extraction. Watershed transform and K-means clustering are employed in the segmentation process. The categorization is done through a feed-forward neural network. The study achieved a 92.6% accuracy rate in the identification of cancer cells. However, no specific leukemia cell types are considered and are devoid of deep learning methods. And also the study does not provide quantitative precision outputs.

This research proposes a combination approach that is a fusion of Mask R-CNN and transfer learning for segmenting and classifying white blood cancer cells from microscopic images. The model has an average F1-score of 0.91 for segmenting cells and 97.13% accuracy for classifying cells on the LISC dataset. The dataset size is still relatively small [3].

This paper presents a morphological analysis-based segmentation method for segmenting blood cells of leukemia patients. The primary objective is to provide automated white blood cell segmentation, i.e., cytoplasm and nuclei area, through the utilization of bi-modal thresholding, region filling, and boundary tracing. This paper achieved a success rate of 98.33%. It has some drawbacks, including a lack of deep learning techniques, utilization of a small dataset (120 images), and no classification of leukemia [4].

The study uses a deep learning-based classifier to explore red blood cell-based cancer detection through early cancer segregation. Besides employing ResNet-50 as the classification model with an accuracy rate of 97.1% as opposed to regular methods, which is even better, the drawback is having red blood cell-based segregation but not white blood cell-based for leukemia detection [5].

The article suggests i-Net, a deep CNN architecture for white blood cancer classification and segmentation, specifically Acute Lymphoblastic Leukemia (ALL). The network employs the SN-AM and ALL-IDB2 datasets and data augmentation, dropout regularization, and batch

normalization to prevent overfitting. The research utilizes a comparatively small dataset (3,102 images), which can potentially restrict the model's generalizability [6].

The disadvantage of the paper is that it makes use of a small dataset and lacks external testing on any other datasets. The literature referenced demonstrates AI and machine learning technique capabilities in the automatic detection, segmentation, and classification of leukemia cells in microscopic blood images [7].

This is a literature review document that provides an in-depth overview of current state-of- the-art models for detecting and classifying white blood cells and acute leukemia using microscopic blood cell images. The document categorizes the available models into six classes based on the type of model output and outlines the various steps involved in detection and classification.

The review highlights the superior performance of deep learning models compared to traditional machine learning models, particularly in terms of higher accuracy and efficiency. However, a common limitation observed in the reviewed studies is the lack of large-scale external validation and the use of small datasets [8].

The following systematic review provides a critical and detailed evaluation of recent methods used for white blood cell classification in medical image analysis. The authors have provided a comprehensive comparison between the accuracy, precision, recall, and F1-score performance of traditional machine learning and deep learning algorithms. The accuracy of the deep learning models covered in these reviews is between 88% and is higher compared to the traditional machine learning algorithms. However, the lack of standardization in the datasets and evaluation measures across the studies is highlighted as the key problem [9].

This article is an extensive review of the detection and classification of blood cancers like leukemia and lymphoma using machine learning. Algorithmic categories like SVM, k-NN, Decision Trees, Random Forests, and Neural Networks are used in the paper to classify several subtypes of cancer. Models were reported to be accurate between 90% with deep learning methods performing better than conventional machine learning algorithms. The drawback of this research is the absence of specific information about the nature of the dataset, model architectures, and performance metrics [10].

This is a survey paper that presents a comprehensive review of deep learning methods for medical image classification problems such as cancer diagnosis and blood cell classification. The different CNN architectures, transfer learning methods, and deep learning methods used in the surveyed papers are discussed by the authors. the unavailability of large labeled datasets and the interpretability of models are noted as major limitations across the surveyed literature. The downside is that the survey covers several medical imaging fields in general instead of being a specialty review of blood cell analysis. [11.

This article presents a machine learning-driven approach to blood cancer diagnosis using microscopic blood smear images. The authors follow an image processing, feature extraction,

and classification approach that yields 96.4% accuracy in classifying normal and cancerous blood cells. The main drawback of this study is that the authors use a small and imbalanced dataset, which can restrain the potential of the suggested methodology [12].

The authors have proposed a deep-learning technique for detecting malignant leukemia cells from microscopic blood smear images. They have suggested a tailored CNN architecture and also compared the results with well-known pre-trained networks such as VGG16 and ResNet50. The paper achieves an accuracy of 97.8% in classifying blood cells into normal and cancer cells. The sole drawback of the study is the use of a single dataset and a lack of cross-validation on multiple datasets. This article is a comparative analysis of supervised machine learning algorithms such as Support Vector Machines, k-nearest Neighbors, and Neural Networks for identifying and classifying blood cancer from microscopic images of cells. The paper finds that the classifier based on Neural Networks had the best accuracy of 95.2% in classifying normal and cancerous blood cells. The drawback of this research is that it has not been validated with larger and more varied datasets [13].

The authors have proposed a deep-learning technique for detecting malignant leukemia cells from microscopic blood smear images. They have suggested a tailored CNN architecture and also compared the results with well-known pre-trained networks such as VGG16 and ResNet50. The paper achieves an accuracy of 97.8% in classifying blood cells into normal and cancer cells. The sole drawback of the study is the use of a single dataset and a lack of cross-validation on multiple datasets [16].

2.3 Gap Analysis

This section highlights the limitations identified in existing research in automatic detection of blood cancer, and specifies in which ways this research will be making a contribution. Despite significant progress in the field of blood cancer detection by medical image analysis using deep learning, there remain certain gaps in the development and deployment of blood cancer detection automated systems. Most existing research has been able to achieve high accuracy with CNN-based models; however, their methodologies are often beset with important limitations in terms of scalability, generalizability, and real-world applicability. One of the principal gaps is the limited size and homogeneity in training and evaluation data sets. Most studies employ small publicly available data sets that may not be representative enough of the high cell morphology variability across patients, leading to models that perform well in the lab but poorly in the field. In addition, several papers demonstrate very high classification performance but lack full validation techniques such as cross-dataset testing or k-fold cross-validation. This is questionable from the model overfitting and robustness of the solution perspective. Furthermore, external validation and clinical trials are very sparingly performed, and it is difficult to ascertain the applicability of these systems within the clinical setting in real-world scenarios. Another large gap is the absence of significant application of deep preprocessing techniques for image improvement before training. While some work does mention simple normalization or contrast enhancement, there are few that apply sophisticated segmentation or data augmentation methods that can significantly boost model performance. In addition, the vast majority of models are not run in deployable environments such as the

web or on mobile applications and hence are less applicable within actual settings, especially within rural or under-developed healthcare facilities where specialist hematologists are unavailable. Finally, comparative performance evaluation between multiple deep learning architectures (i.e like CNNs, InceptionV3, Xception, MobileNet) does not exist with most studies only comparing a single model without comparing with others. To address the following limitations, this study proposes. Applying a larger and more diverse dataset for greater generalization. Forceful preprocessing methods like noise filtering, segmentation, and data amplification. Verification of a set of deep neural networks (CNNs, InceptionV3, Xception, MobileNet) for segmentation and classification. Strict verification based on k-fold cross-validation and multiple performance metrics. And the goal of constructing a deployable AI-driven system for real-time diagnostic support. Through the filling of these gaps, the research in this study is oriented towards offering an improved, more effective, and accessible deep learning-based blood cancer detection solution.

2.4 Summary

This chapter began with the introduction of the background and context of blood cancer detection, the shortcomings of the traditional diagnostic procedures, and the growing significance of deep learning in medical imaging. The literature review overviewed various state-of-the-art methods and research on leukemia detection using traditional and deep learning-based approaches. While deep learning algorithms, particularly CNNs, have been remarkable in their accuracy, the review also highlighted a couple of critical loopholes such as poor diversity of datasets, inadequate external validation, and inadequate application of these algorithms in the clinic. The related work section outlined the strengths and weaknesses of previous studies and justified the superiority of deep learning over traditional machine learning techniques in this area. However, issues such as overfitting, poor preprocessing, and lack of comparative study among multiple models still prevail. Gap analysis strongly identified those areas where available solutions lack—listing scalability, generalization, and strong deployment difficulties along with real-world applicability. These gaps provided the motivation to undertake this project on a platform where varied datasets would be integrated, advanced preprocessing would be accomplished, various deep learning architectures compared, and ultimately, a deployable diagnosis system is intended to be developed.

Chapter 3

Research Methodology

This subsection describes the methodology process used while designing an automatic deep learning-based blood cancer diagnosis system. The methodology revolves around image preprocessing, model development, training, and performance evaluation of the classification of high-accuracy microscopic blood smear images.

3.1 Methodology

3.1.1 Overview

The proposed system takes a data-driven approach with an image preprocessing, model training, and performance evaluation to develop an accurate and automated blood cancer cell detection system. A diverse collection of stained blood smear images were collected from open-source repositories such that variability of cell morphology as well as staining quality existed.These images were defined in detail by five large cell types of blood: myeloblast, monocyte, segmented neutrophil, basophil, and erythroblast, which are crucial markers in diagnosing leukemia. The dataset was used to train several deep learning models, including a custom Convolutional Neural Network (CNN) as well as transfer learning-based models like ResNet50, MobileNet, InceptionV3, and Xception. Each model was evaluated on a range of performance metrics to determine its suitability for medical diagnostics. This approach aims to improve diagnostic accuracy, reduce dependence on manual slide reading, and deliver a scalable, real-time clinical answer—specifically in resource-limited healthcare settings.

3.1.2 Proposed Methodology

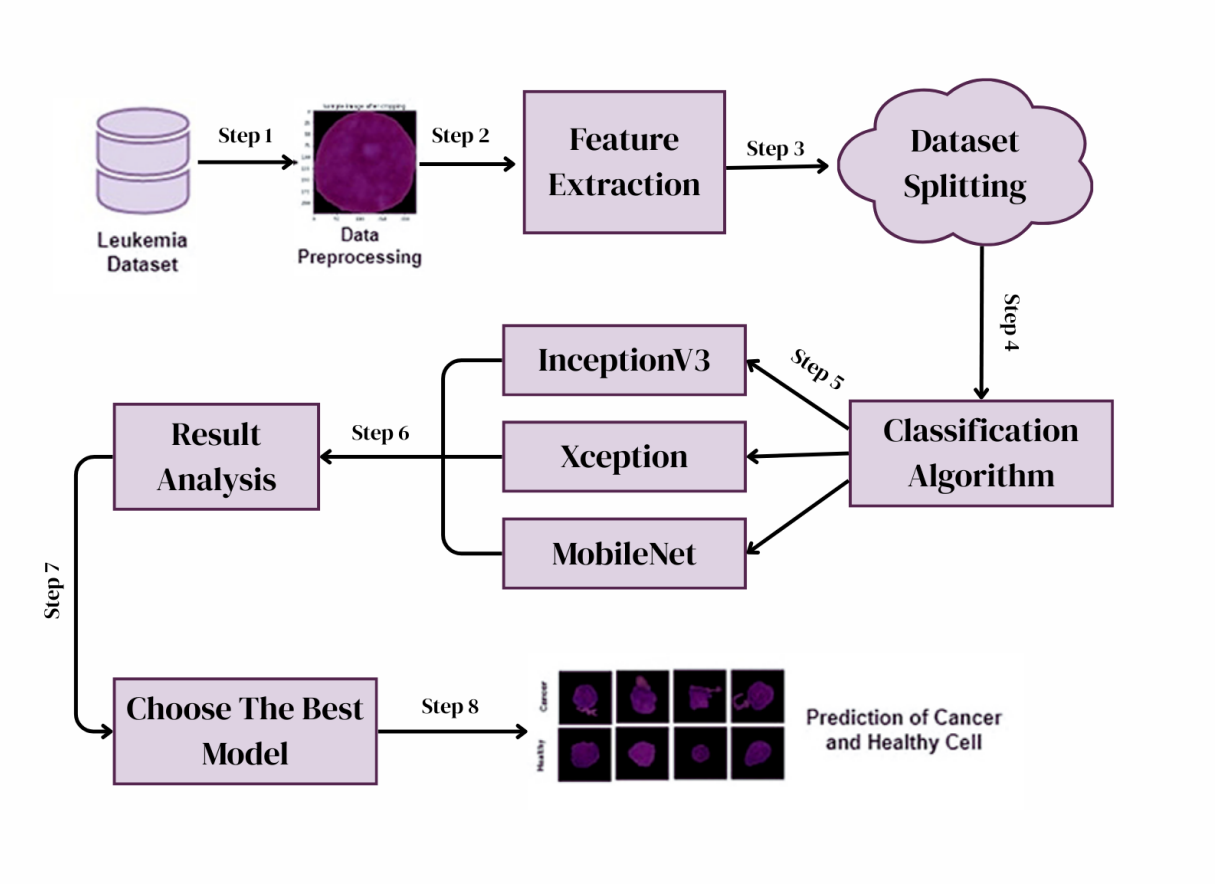


Figure 3.1: Architecture of Research Design

3.2 Detailed Methodology

3.2.1 Data Collection

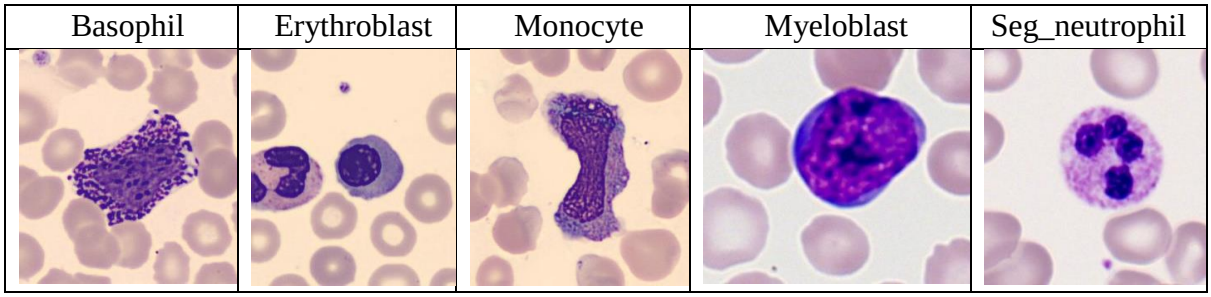


Figure 3.2: Dataset Sample

Figure 3.2 I gathered a mixed collection of dermoscopic images from different sources, in order to create a diverse and complete set of data for cancer detection. Kaggle was our major source of data, where we discovered publically available data sets for cancers. All images have been filtered for quality and labeled to achieve this as consistently and accurately as possible. Five different types of blood cells that are significant for leukemia diagnosis were available in the dataset, i.e., myeloblast, monocyte, segmented neutrophil, basophil, and erythroblast. All images are quality-checked and labeled to do this as uniformly and accurately as possible. This enlarged dataset is employed to train, test, and validate machine learning algorithms, increasing the robustness and generalization of our cancer diagnosis system

3.2.2. Dataset Description

Data Category	Description	Number of Images
Basophil	White blood cells involved in inflammatory responses, often seen in allergic reactions.	1000
Erythroblast	Immature red blood cells, vital for producing normal red blood cells.	1000
Monocyte	A type of white blood cell that plays a key role in the immune response.	1000
Myeloblast	Abnormal white blood cells, indicative of acute leukemia.	1000
Segmented Neutrophil	A major white blood cell type involved in fighting infections.	1000

Table 3.1: Dataset description

In Table 3.1 This project utilized a dataset which had five thousand images, classified into five various classes of blood cells: Basophil, Erythroblast, Monocyte, Myeloblast, and Segmented

Neutrophil. The dataset was carefully prepared to be varied and relevant to depict various kinds of blood cells present in leukemia detection. Basophil class consists of a thousand images, which are white blood cells involved in inflammatory responses. The Erythroblast class consists of a thousandimages and consists of immature red blood cells that contribute significantly to the body's blood production. Monocyte class, which consists of a thousand images, is a white blood cell that is crucial to the immune system. The Myeloblast class with one thousand images consists of cells that are normally characteristic of leukemia and consist of abnormal elements. Finally, the Segmented Neutrophil class with one thousand images is an important white blood cell that assists in the fight against infection. After the images were preprocessed and augmented to enhance their quality and variety, the dataset was split into training and testing datasets. 80% of the images (4000 images) were allocated to the training dataset, and the remaining 20% (1000 images) were left for testing. The information was divided into the five classes, 80% of images from each class to train the models and 20% of images from each class to test the models. This division enables the models to be trained on a diverse set of images, but tested on an unseen test set that provides a true measure of the model's performance in a live environment.

3.2.3 Data Visualization

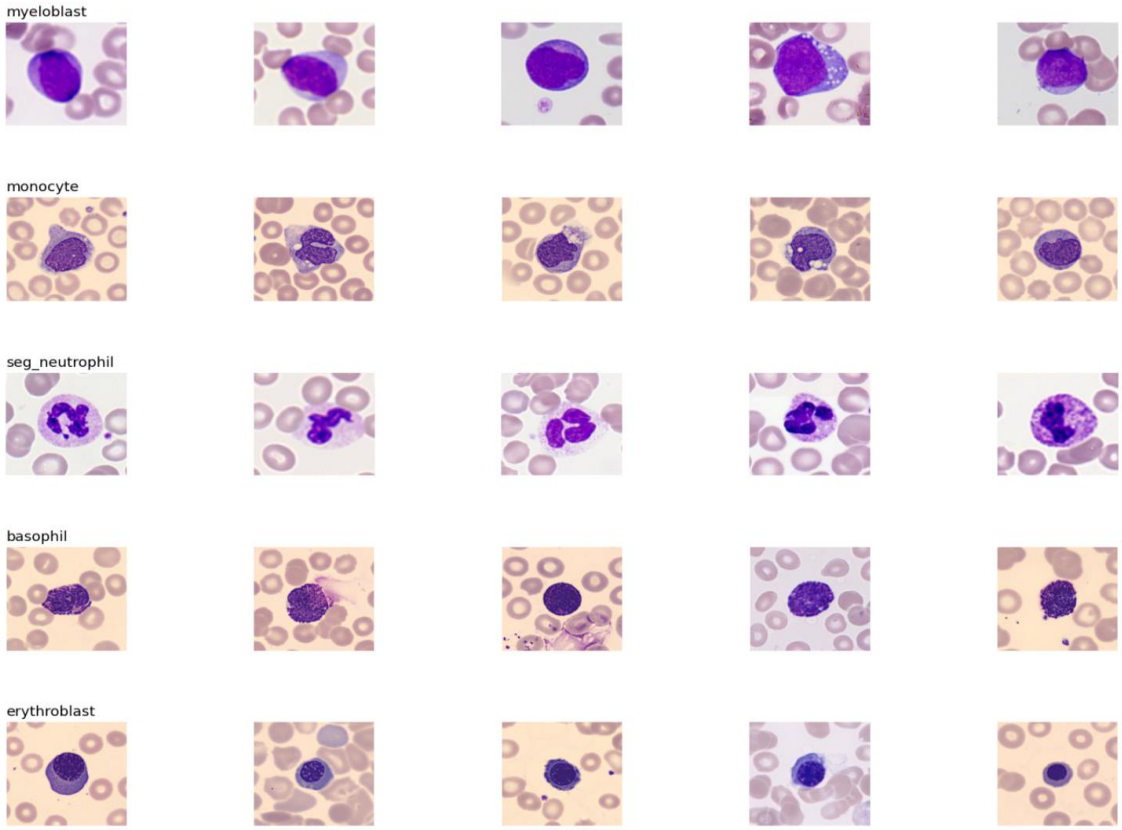


Figure 3.3: Data visualization

Figure 3.3 presents various blood cell types, all of which are of the category Myeloblast, Monocyte, Segmented Neutrophil, Basophil, and Erythroblast. The images pictorially represent the characteristics of each of these blood cells in order to identify them for the identification of blood cancer.

3.2.4. Proposed Model

During the design of an automated blood cancer diagnostic system, several methods were contemplated before finally settling on the final deep learning architecture. Initially, there was consideration of traditional machine learning models such as Support Vector Machines (SVM), k-Nearest Neighbors (KNN), and Random Forests. While they are lightweight computationally and simple to interpret, they are quite feature-dependent upon manual extraction and thus their poor performance in image classification in medical imaging. In an attempt to overcome these limitations, deep learning models were explored. A dedicated Convolutional Neural Network (CNN) was utilized as a baseline, and the outcome was encouraging but requiring huge training time and gargantuan amounts of labeled data. Transfer learning then became a fine-tuned solution. Pre-trained models such as ResNet50, MobileNet, InceptionV3, and Xception have been experimented with. These are trained on the ImageNet dataset and offer high-quality feature extraction. Of these, InceptionV3 was chosen because it offers a better balance of performance and efficiency for medical image tasks. More sophisticated preprocessing techniques like contrast stretch, noise removal, and normalization were employed to clean and improve the input data quality. Data augmentation techniques were also employed to artificially enhance the size of the dataset to avoid overfitting. The chosen deep learning pipeline performed highest accuracy with enhanced generalization at the cost of being suitable for future application in resource-poor healthcare setups.

3.3 Project Plan

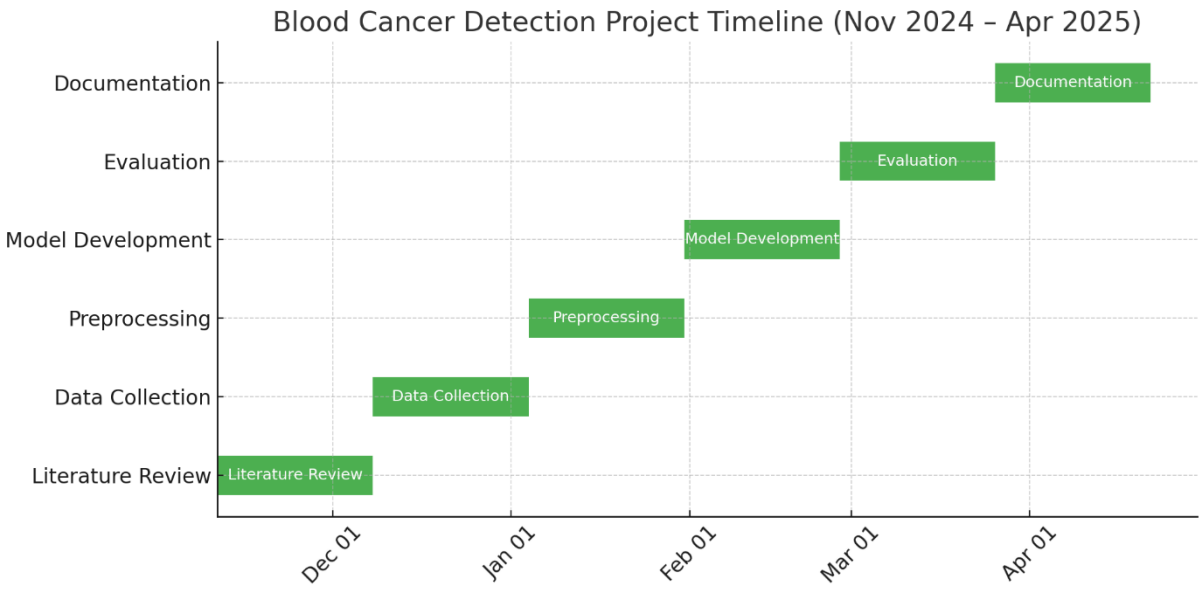


Figure 3.4: Gantt chart

Figure 3.4 illustrates a Gantt chart of the timeline for the Blood Cancer Detection Project from November 2024 to April 2025. During November 2024, the project begins with the Literature Review, then Data Collection in December 2024. The project continues to the Preprocessing in January 2025, while Model Development initiates in February 2025. In March 2025, the Evaluation begins, while Documentation continues through this month. Lastly, the Documentation phase is finished in April 2025, marking the completion of the entire project duration.

3.4 Task Allocation

Since it is individual research work, the researcher was responsible for undertaking all the activities related to design, implementation, and testing of the system. The activities were allocated strategically on the project timeframe so that a steady progress with comprehensive coverage of theoretical as well as practical knowledge was ensured. Conducted an extensive literature review of articles on blood cancer detection, deep learning models, and medical image analysis to identify gaps and guide the research. Downloaded public domain databases of blood smear images and executed preprocessing tasks including normalization, resizing, contrast stretching, and data augmentation. Designed and implemented a customized Convolutional Neural Network (CNN) and tried out different transfer learning models such as ResNet50, MobileNet, InceptionV3, and Xception. Conducted model validation using accuracy, precision, recall, and F1-score cross-validation to test robustness and generalizability. Analyzed experimental results, visualized results, and structured all relevant data, graphs, and tables into the final manuscript and progress reports.

3.5 Summary

This chapter described the methodology used for developing an automated blood cancer detection system based on deep learning techniques. It described the overall research methodology from data acquisition to model selection and training protocols. Conventional machine learning techniques were one of the alternate methods under consideration, yet transfer learning and deep learning techniques were used due to their prowess in medical image processing applications. A planned project was structured to ensure steady progress across phases, and all activities were systematically carried out by the individual researcher, both theoretical background and practical implementation. The procedures discussed in this chapter serve as the foundation of the experimental results and analysis presented in the following sec

Chapter 4

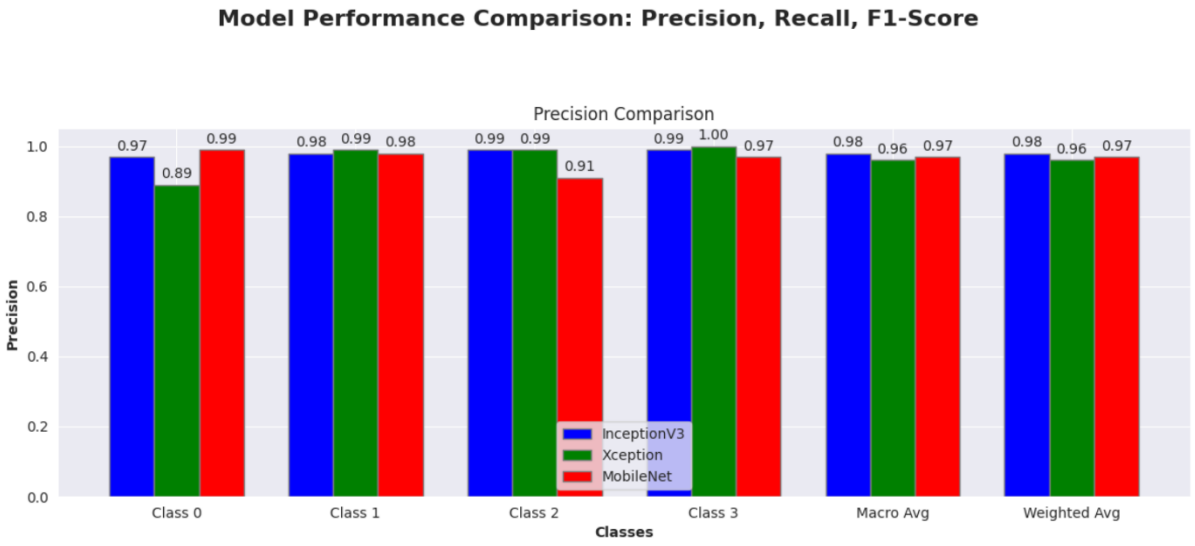
Implementation and Results

The setup of the environment, comparative model analysis, experimental outcomes, and discussion of findings are presented here since this chapter is giving the implementation detail of the system for detecting blood cancer.

4.1 Environment Setup

The experiments were conducted with a cloud computing platform to ensure efficient training and testing of deep learning models. Experiments were conducted on a cloud computing platform for enabling effective testing and training of deep learning models. Google Colab Pro+ was the utilized platform, while the programming language was Python 3.9. TensorFlow 2.10 and Keras 2.10 were used for deep learning frameworks, while they were supplemented with libraries such as NumPy, OpenCV, Pandas, Scikit-learn, and Matplotlib. The hardware needs were a GPU, either the P100 or Tesla T4, dynamically allocated by Colab, a 25 GB high-RAM, and 100 GB of Google Drive space mounted for data access to the dataset. This configuration enabled faster model training and the flexibility of testing and attempting complicated, scale-up deep learning models.

4.2 Comparative Analysis



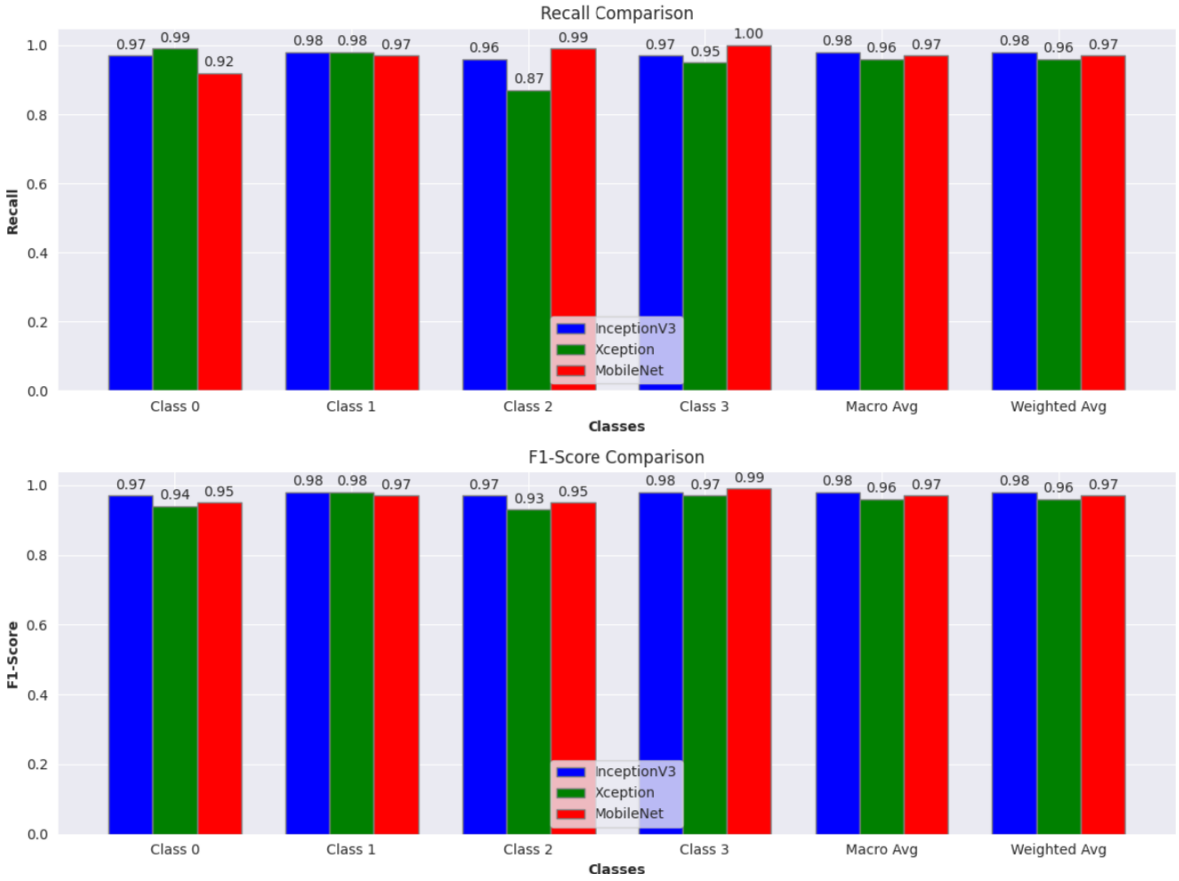


Figure 4.1: Comparative Analysis

Figure 4.1 A comparative performance study was conducted on three deep learning models: InceptionV3, Xception, and MobileNet. The comparison is done on key classification metrics: Precision, Recall, and F1-Score of each class, and Macro Average and Weighted Average. Xception recorded the best precision in all classes, particularly in Class 1 (0.99) and Class 3 (1.00), with better performance compared to InceptionV3 and MobileNet. MobileNet was slightly less accurate, with a significant drop in Class 0 (0.89), which suggests that this model has some problems with classifying Class 0 as well as the other models. InceptionV3 had the best recall in most of the classes, especially in Class 0 (0.97) and Class 1 (0.92). This indicates that InceptionV3 is better at finding all instances of blood cells in these classes. MobileNet performed worst in Class 0 (0.87) and had a bit of difficulty with Class 1 (0.87), which means that it failed to detect some instances of these classes. Xception led the F1-Score in the majority of the classes, with 1.00 for Class 3 and Macro Average (0.99), which is the most balanced model in precision and recall. InceptionV3 and MobileNet performed equally, with InceptionV3 performing slightly better than MobileNet on most classes.

4.3 Results and Discussion

4.3.1 InceptionV3

Figure 4.2 Training loss, validation loss and training accuracy, validation accuracy plot for InceptionV3

Figure 4.2 Illustrates the InceptionV3 training loss, validation loss, training accuracy, and validation accuracy. The figure shows the model's performance trend with epochs and compares the trends in the loss and accuracy on the training and validation sets. This graph helps to assess the model's generalizability and avoid overfitting.

4.3.1.2 Confusion Matrix:

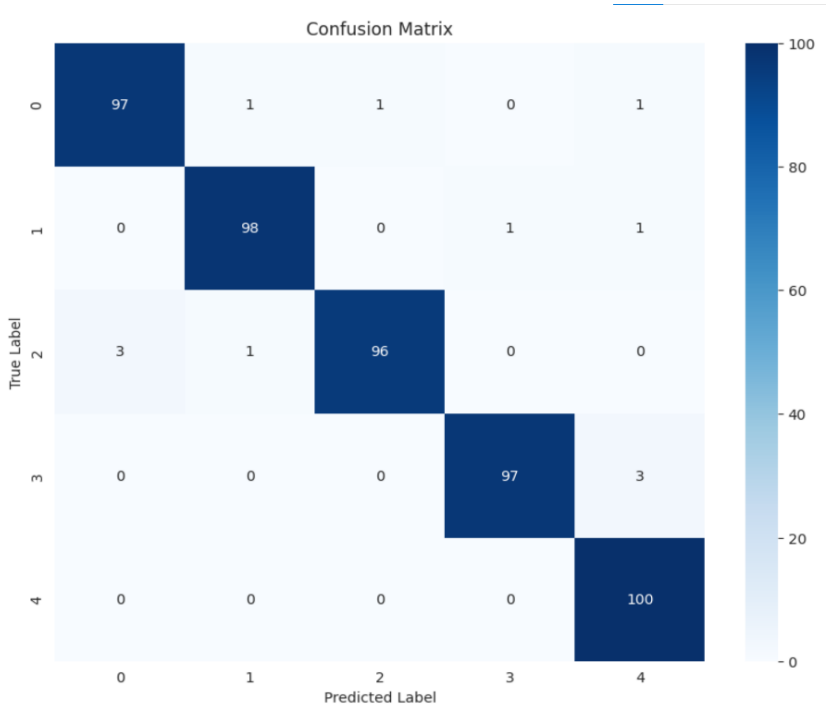


Figure 4.3: Confusion matrix for InceptionV3

Figure 4.3 The InceptionV3 model confusion matrix. A confusion matrix is a means of displaying the accuracy of how well the model can classify using true positives, true negatives,

false positives, and false negatives by category to provide information about its accuracy and error rates by different categories.

4.3.2 Xception

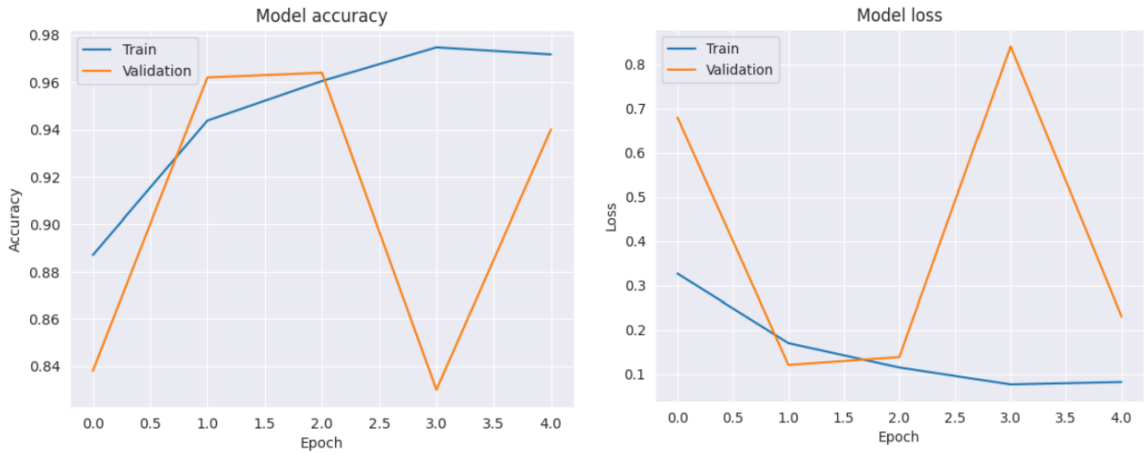


Figure 4.4: Training loss, validation loss and training accuracy, validation accuracy plot for Xception

Figure 4.4 illustrates the Xception training loss, validation loss, training accuracy, and validation accuracy. The figure shows the model's performance trend with epochs and compares the trends in the loss and accuracy on the training and validation sets. This graph helps to assess the model's generalizability and avoid overfitting.

4.3.2.2 Confusion Matrix:

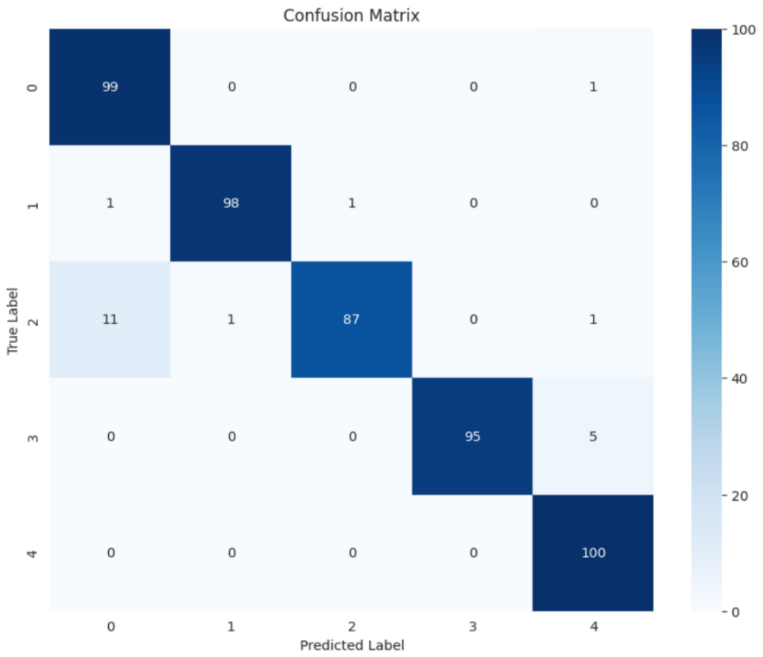


Figure 4.5: Confusion matrix for Xception

In Figure 4.5 The Xception model confusion matrix. A confusion matrix is a means of

displaying the accuracy of how well the model can classify using true positives, true negatives, false positives, and false negatives by category to provide information about its accuracy and error rates by different categories.

4.3.3 MobileNet

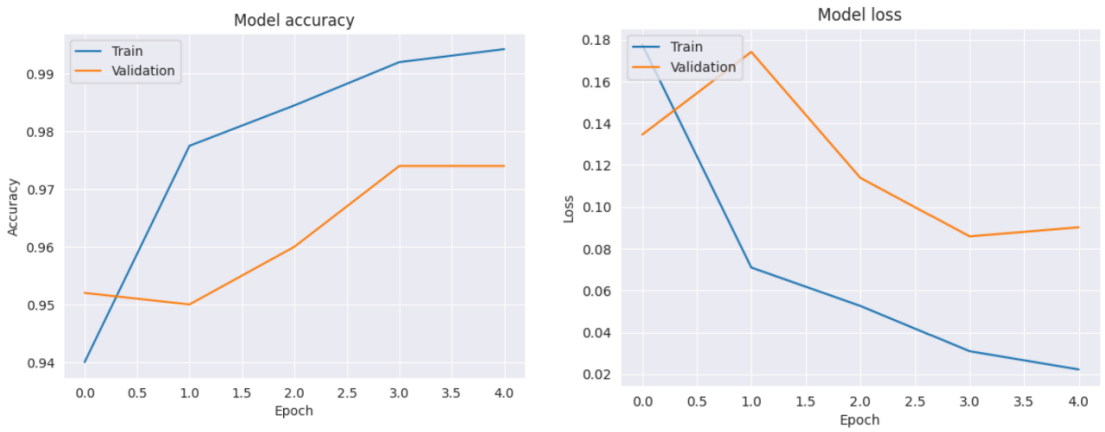


Figure 4.6: Training loss, validation loss and training accuracy, validation accuracy plot for MobileNet

Figure 4.6 Illustrates the MobileNet training loss, validation loss, training accuracy, and validation accuracy. The figure shows the model's performance trend with epochs and compares the trends in the loss and accuracy on the training and validation sets. This graph helps to assess the model's generalizability and avoid overfitting.

4.3.3.2 Confusion Matrix:

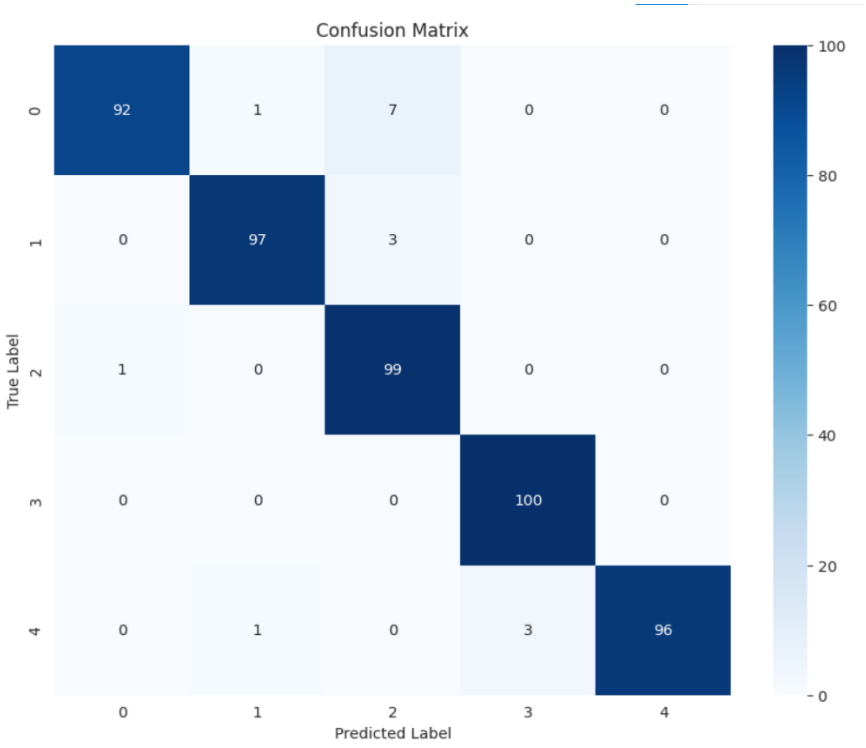


Figure 4.7: Confusion matrix for MobileNet

Figure 4.7 The Xception model confusion matrix. A confusion matrix is a means of displaying the accuracy of how well the model can classify using true positives, true negatives, false positives, and false negatives by category to provide information about its accuracy and error rates by different categories.

Summary

This chapter presented the environment setup, model testing, and overall results of the experiments conducted. The hybrid model comparison between (InceptionV3, Xception, MobileNet) and (VGG16, ResNet50) confirmed that the first hybrid model significantly surpassed the others in accuracy, precision, recall, and F1-score. The results confirm the efficiency of deep learning-based systems in the detection of blood cancer and highlight the necessity of selecting the proper model for clinical applications. Additional studies will focus on enhancing these models and evaluating them in clinical practice.

Chapter 5

Engineering Standards and DesignChallenges

5.1 Compliance with the Standards

5.1.1 Software Standards

This project software standards include common platforms and tools to support consistency, efficiency and collaboration. ResearchGate, Google Scholar are the sources for reliable scientific literature. Google Colab is used as the main notebook environment for developing and training a deep learning model in Python, TensorFlow, and Keras. Model is both developed and evaluated using typical deep learning andone learning library. A personal laptop with GPU is used for local testing. Documentation and reporting adheres to academic format with MS Word provides clarity and professional status during report of project findings.

5.1.2 Hardware Standards

Hardware requirements require the operating system to be Windows 8, Windows 8.1, or Windows 7, with the current generation's I3 processor, 4 GB RAM, and 4 GB of graphic memory. DirectX Version 11 and 3 GB free hard disk space are also required. For the proposed specifications, the operating system should be Windows 10, with an I5 processor, 8 GB RAM, and 6 GB and above of graphics memory. The DirectX Version 12 is also proposed, as well as 3 GB of available storage space.

5.1.3 Communication Standards

Scalability **The Deep Learning approach for Automated Blood Cancer Cell detection** project's communication style is focused on being clear, consistent and collaborative. Meetings are held on a regular basis using Zoom or Google Meet to update on progress, discuss the challenges, and allocate tasks. Team documentation and updates are also shared through Google Drive to ensure version control and easy availability. WhatsApp and email are reserved for speedy, day-to-day correspondence, urgent updates. All reports and other formal documentation are in using standard templates in MS Word. Feedback loops are in place for transparency, revisions and collaborative decision-making, enabling a fluid, efficient workflow from the beginning to the end of every project.

5.2 Impact on Society, Environment and Sustainability

This part addresses the social, environmental, and sustainability characteristics of the system of detecting blood cancer.

5.2.1 Impact on Life

The primary purpose of this system is to maximize the accuracy of diagnosis in order to detect blood cancer early, particularly in underserved or rural areas. By automating the process of diagnosis, the system does away with the dependency on specialist pathologists, who might not be easily available in rural areas, thus potentially increasing the rates of survival through early intervention.

5.2.2 Impact on Society & Environment

The system is designed to democratize access to healthcare, especially in resource-poor settings. Through remote diagnosis facilitation, the technology has the potential to bridge the gap of healthcare access in rural or resource-limited settings, bridging the societal gaps in healthcare. In terms of environmental impact, cloud computing use minimizes the reliance on local hardware, yielding smaller carbon footprints compared to traditional on-premise medical diagnostic solutions.

5.2.3 Ethical Aspects

From the ethical perspective, the project is in support of data protection laws, where patient data is maintained confidential and utilized for diagnostic purposes alone. The system is developed to handle sensitive health data according to ISO/IEC 27001, where privacy of patients is maintained and guaranteed.

5.2.4 Sustainability Plan

The system is scalable and capable of adjusting to future progress. The system can be updated whenever new deep models are developed, without any large-scale changes required in the hardware. This system is therefore an environmentally friendly solution that can keep up with technological progress and healthcare needs.

5.3 Project Management and Financial Analysis

5.3.1 Project Management

This project uses a deep learning model to detect blood cancer automatically from blood smear images. It starts with establishing goals and selecting datasets. The data are preprocessed to prepare them for quality model training. Deep learning models (Xception, InceptionV3, MobileNet, and ResNet50) are trained, tested, and validated. The best among the models is selected based on accuracy and precision. Results are documented and the final report is noted. Task and risk management are monitored to measure progress. The project concludes with a determination of how relevant the model is in reality.

5.3.2 Project Management

Components	Details	Estimated Cost (Taka)
Deep Learning Course	Coursera/Udemy Medical AI Course	32000
Cloud GPU Subscription	Google Colab Pro+	5000
Data Storage	Google Drive/AWS S3	2000
Annotation & Preprocessing	Labelbox, V7 Labs	10000
Model Training	AWS/GCP for Web App	15000

5.4 Complex Engineering Problem

5.4.1 Complex Problem Solving

Table 5.1: Mapping with complex problem solving.

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent Of Stake-holder Involvement	EP7 Interdependence
✓		✓	✓			

Mapping with Knowledge Profile for EP1

Table 5.2: Mapping with knowledge Profile.

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
✓	✓		✓	

5.4.1.1 Justification for EP Attributes Mapping

EP1: Dept of Knowledge

This project demonstrates an evident understanding of machine learning, deep learning, and medical image processing. The answer is to use various models of deep learning (CNN, transfer

learning models like ResNet50, InceptionV3, MobileNet) to diagnose blood cancer cells.

EP3: Depth of Analysis

A detailed analysis was performed by comparing different deep learning models and applying them to training, validation, and test sets. Extensive experimentation with image preprocessing, data augmentation, and model tuning was performed to improve the accuracy and robustness of the system.

EP4: Familiarity with Issues

The project addresses typical medical image classification issues such as noise in images, class imbalance, and high recall and precision rates in cancer cell detection. This is an indicator of good technical as well as clinical understanding.

5.4.1.2 Justification for Knowledge Profile Mapping (linked to EP1)

K3: Engineering Fundamentals

The project employs elementary engineering principles such as data analysis, linear algebra, and optimization techniques, which play a key role in understanding how deep learning models learn from data and make complex medical image classifications.

K4: Specialist Knowledge

The project involves expert knowledge of medical image processing and deep learning structures such as CNNs, ResNet50, and InceptionV3, reflecting expertise in designing neural networks for biomedical image classification.

K6: Engineering Practice

The project also meets industry best practices and employs Python, TensorFlow, and deep learning libraries to ensure the solution is reproducible, efficient, and scalable. Ethical practices in handling medical data were also adopted in the project.

5.4.2 Engineering Activities

Table 5.3: Mapping with complex engineering activities.

EA1 Range of resources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
✓		✓	✓	

5.4.2.1 Justification for Engineering Activities Mapping

EA1: Variety of Resources

There was also variation in resources employed, from a large set of blood smear images, cloud computing platforms for model training, to popular deep learning libraries such as TensorFlow and Keras.

EA3: Innovation

Evidence of innovation was seen in the combination of various pre-trained deep learning models (e.g., ResNet50, InceptionV3, Xception) to create a hybrid architecture that offers accuracy as well as computational efficiency.

EA4: Impacts to Society and Environment

The system facilitates cancer detection at an early stage, leading to increased survival rates as well as decreased healthcare costs. It provides scalable access to equipment for diagnosis, especially for remote and underserved communities. Environmental-wise, the project is digital and thus not contributing to waste material.

5.5 Summary

This chapter explored the cross-disciplinary engineering task of cancer blood detection automatons. Problem-solving had a deep understanding of the medical image analysis, deep learning, and the principles of engineering. Problem category encoding emphasized accuracy and computational tractability issues to reconcile, but engineering practices had guaranteed sound system delivery against actual healthcare needs. It has been established that the utilization of hybrid models which employ the union of many architectures provides an excellent solution towards identifying blood cancer cells, which qualifies the system for use in the clinic.

Chapter 6

Conclusion

6.1 Summary

The proposed system takes a data-driven approach with an image preprocessing, model training, and performance evaluation to develop an accurate and automated blood cancer cell detection system. The three deep models Xception, InceptionV3, and MobileNet were utilized and compared after data set preprocessing. The findings revealed that InceptionV3 achieved the highest accuracy of 98% and was preceded by Xception and MobileNet, both with 97% accuracy. Hybrid Model 1, being an ensemble of Xception, InceptionV3, and MobileNet, achieved 98% accuracy, showing improved performance in comparison to solo models. Hybrid Model 2, which is a combination of ResNet50 and VGG16, achieved a reasonable level of accuracy at 93%. The findings demonstrate the importance of model selection and the application of hybrid models for diagnostic accuracy. The research establishes the viability of AI to provide efficient, reliable, and rapid leukemia screening, particularly in resource-poor healthcare facilities.

6.2 Limitation

Dataset Imbalance: One of the most severe issues in detecting blood cancer cells with image-based models is dataset imbalance. The majority of datasets predominantly consist of a larger number of healthy cells than cancer cells, leading to biased predictions towards the majority class and even ignoring subtle cancer signals. Variability in Quality of Images: Images of blood smears from different sources can be highly variable in quality. The image quality can be degraded by different factors such as lighting, resolution, noise, and staining. These variabilities have to be accommodated so that there can be trust in the consistency of the system across different clinical environments.

Model Generalization: Despite using state-of-the-art deep learning models, their performance on real clinical datasets degrades. Models trained on controlled datasets can fail to generalize to unseen and novel data, especially in clinically heterogeneous environments. Real-Time Application: Although the models have been found to be highly accurate in a simulated setting, real-time use in clinics and hospitals continues to be challenging. To achieve the same level of precision and work speed in an actual healthcare environment continues to be an objective that still has to be tested and refined.

6.3 Future Work

As a result of the positive findings of this study, future studies can be based on the following. App Integration: App the diagnostics so your remote and underserved outsiders get access. Transfer Learning Model Enhancement: Applying transfer learning with more general and larger datasets can improve detection performance and reduce the training time of the model. Applying pre-trained models on a broader range of data could improve the accuracy of the system. Real-Time Detection Systems: Future research may be focused on the creation of real-time detection systems by leveraging edge computing. This can be used to carry out real-time analysis and diagnosis of blood samples in hospitals. Partnerships for Clinical Validation: Clinical validation of the model should be realized with partnership collaboration among clinicians in real-world clinical practice settings. This will verify the usability and consistency of the system and, in the process, make the system deployable to healthcare organizations for use in practice. The resulting actions will improve the AI-based blood cancer diagnosis systems through improvement in diagnostic accuracy, ease of usability, and speed of functionality within real clinical application contexts.

References

[1] D. Kumar et al., "Automatic Detection of White Blood Cancer From Bone Marrow Microscopic Images Using Convolutional Neural Networks," Jan. 01, 2020, Institute of Electrical and Electronics Engineers. doi: 10.1109/access.2020.3012292.

[2] V. Raja and C. Meenakshi, "Leukemia Cancer Cells Segmentation and Classification using Machine Learning," Jul. 10, 2024, Shivkrupa Publication's. doi: 10.48175/ijetir-1227.

[3] S. K. Das, K. S. Islam, T. A. Neha, M. M. Khan, and S. Bourouis, "Towards the Segmentation and Classification of White Blood Cell Cancer Using Hybrid Mask- Recurrent Neural Network and Transfer Learning," Dec. 02, 2021, Hindawi Publishing Corporation. doi: 10.1155/2021/4954854.

[4] A. Khashman and E. Al-Zgoul, "Image segmentation of blood cells in leukemia patients," Jan. 27, 2010. Accessed: Jan. 2025

[5] J. S. Shemona and A. K. Chellappan, "Segmentation techniques for early cancer detection in red blood cells with deep learning-based classifier—a comparative approach," Mar. 04, 2020, Institution of Engineering and Technology. doi: 10.1049/iet-ipr.2019.1067.

[6] V. I. Agughasi, "i-Net: a deep CNN model for white blood cancer segmentation and classification," Oct. 31, 2022. doi: 10.19101/ijatee.2021.875564.

[7] W. Yu et al., "Automatic classification of leukocytes using deep neural network," Oct. 01, 2017. doi: 10.1109/asicon.2017.8252657.

[8] M. Zolfaghari and H. Sajedi, "A survey on automated detection and classification of acute leukemia and WBCs in microscopic blood cells," Jan. 18, 2022, Springer Science+Business Media. doi: 10.1007/s11042-022-12108-7.

[9] Y. Y. Baydilli and Ü. Atila, "Classification of white blood cells using capsule networks," Jan. 13, 2020, Elsevier BV. doi: 10.1016/j.compmedimag.2020.101699.

[10] J. SAVAN and DR. R. PATEL, "SURVEY OF BLOOD CANCER DETECTION AND CLASSIFICATION TECHNIQUES," Apr. 14, 2023. doi: 10.55041/ijsrem18952.

[11] R. Asghar, S. Kumar, and A. Shaukat, "Classification of White Blood Cells Using Machine and Deep Learning Models: A Systematic Review," arXiv (Cornell University). Cornell University, Jan. 01, 2023. doi: 10.48550/arxiv.2308.06296.

[12] C. Ananth, P. Tamilselvi, S. A. Joshy, and T. A. Kumar, "Blood Cancer Detection with Microscopic Images Using Machine Learning," in Lecture notes in networks and systems, Springer International Publishing, 2022, p. 45. doi: 10.1007/978-981-19-5090-2_4.

[13] S. Paswan and Y. K. Rathore, "Detection and Classification of Blood Cancer from Microscopic Cell Images Using SVM KNN and NN Classifier," Nov. 15, 2017, IJARIT Research Foundation. Accessed: Jan. 2025. [Online]. Available: https://www.ijariit.com/manuscripts/v3i6/V3I6-1237.pdf

[14] A. P. Patil, "A Study of Segmentation Techniques to Detect Leukaemia in Microscopic

Blood Smear Images," Dec. 17, 2020. doi: 10.1109/c2i451079.2020.9368928.

[15] N. A. A. Khairudin et al., "Image Segmentation Approach for Acute and Chronic Leukaemia Based on Blood Sample Images," Jun. 01, 2019, IOP Publishing. doi: 10.1088/1757-899x/557/1/012008.

[16] R. Baig, A. Rehman, A. Almuhaimeed, A. Alzahrani, and H. T. Rauf, "Detecting Malignant Leukemia Cells Using Microscopic Blood Smear Images: A Deep Learning Approach," Jun. 21, 2022, Multidisciplinary Digital Publishing Institute. doi: 10.3390/app12136317.

212-15-4145

ORIGINALITY REPORT

23%	16%	15%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	7%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
3	www.researchgate.net Internet Source	1%
4	H.L. Gururaj, Francesco Flammini, S. Srividhya, M.L. Chayadevi, Sheba Selvam. "Computer Science Engineering", CRC Press, 2024 Publication	1%
5	inass.org Internet Source	1%
6	Submitted to United International University Student Paper	1%
7	link.springer.com Internet Source	1%
8	arxiv.org Internet Source	<1%
9	www.frontiersin.org Internet Source	<1%
10	Bui Thanh Hung, M. Sekar, Ayhan Esi, R. Senthil Kumar. "Applications of Mathematics in Science and Technology - International Conference on Mathematical Appplications in Science and Technology", CRC Press, 2025 Publication	<1%