Acute Lymphoblastic Leukaemia Detection using Deep Learning (ViT)

By
Khandaker Rezoanul Haque
212-15-4204

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised by

Shahadat Hossain
Assistant professor
Department of Computer Science and Engineering Daffodil International University

Co-Supervised by

Md. Jahidul Alam
Lecturer
Department of Computer Science and EngineeringDaffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

May 14, 2025

APPROVAL

This Project titled **"Acute Lymphoblastic Leukaemia Detection using Deep Learning (ViT)"**, submitted by **Khandaker Rezoanul Haque**, ID No: **212-15-4204** 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. Sheak Rashed Haider Noori — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Md Masum Billah (MMB) — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Mr. Partha Dip Sarkar (PDS) — **Internal Examiner**
Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Zulfiker Mahmud — **External Examiner**
Professor
Department of Computer Science and Engineering
Jagannath University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Shahadat Hossain Assistant professor,** 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:

14.05.25

Shahadat Hossain
Assistant professor
Department of Computer Science and EngineeringDaffodil International University

Co-Supervised by:

Md. Jahidul Alam
Lecturer
Department of Computer Science and EngineeringDaffodil International University

Submitted by:

Rezoan

Khandaker Rezoanul Haque
Student ID: 212-15-4204
Department of Computer Science and EngineeringDaffodil 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 as- sisted 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 **Shahadat Hossain, Assistant professor,** Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Machine Learning** to carry out this project. His endless pa- tience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartfelt gratitude to the Head of the Department of Com- puter 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

Significant advancements in machine learning have been made in disease detection within the medical field; however, challenges remain—particularly in achieving high accuracy and minimizing false positives. Recently, Vision Transformer (ViT) technology, originally developed for visual tasks, has demonstrated promising potential in enhancing detection performance. Motivated by this, our study implemented ViT to detect Acute Lymphoblastic Leukemia (ALL), achieving a remarkable accuracy of 99.35%. This means that out of every 100 disease-related images, our model accurately identified the diseased blood cells approximately 99 times. We utilized a publicly available ALL dataset that includes all four stages of the disease. The importance of this work is underscored by the severe health risks posed by ALL, especially in children. Furthermore, our research highlights the potential of precisely identifying early-stage cancer cases. What distinguishes our approach is the application of machine learning—specifically ViT—to automatically detect and classify cancer, offering a substantial improvement over traditional ALL detection methods, which are often time-consuming and prone to human error. Looking ahead, we aim to develop dedicated hardware to support medical professionals in the rapid and accurate identification of ALL symptoms and affected blood cells. This fusion of data science and medicine holds significant promise for addressing a wide range of medical challenges, including ALL.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Vision Transformer (ViT) was initially an image classifier. It is one of the new uses of transformers in vision tasks, in addition to being well-known in traditional language modeling. The core strategy of the ViT model is breaking down each image into a sequence of fixed-length tokens and then applying multiple Transformer layers to learn global interdependencies for classification [1]. Most of the new evidence suggested that on image classification tasks, Vision Transformers (ViTs) have achieved remarkable scores. ViTs experience decreasing returns as depth is increased, unlike CNNs, which work optimally using additional convolutional layers. This issue may be caused by the phenomenon of attention collapse, where at a certain depth, attention maps begin to become increasingly identical, if not equal [2].

The ViT utilizes a self-attention mechanism to process image patches. The scaled dot-product self-attention in ViT has a limitation due to the unpredictability of the magnitude of the original feature space.[3]. Recent transformers with good capability in global relation modeling are applied to fundamental computer vision tasks. The Vision Transformer (ViT), for instance, segments images into tokens of fixed sizes and uses transformers to detect relations among such tokens for image classification [4]. Vision transformers (ViTs) have significantly enhanced computer vision tasks. They also involve a lot of spatial redundancy in the input images, which results in high computational costs [5]. Deep learning has been a major contributor to computer vision in image recognition, particularly medical image recognition. Cancerous B-lymphoblast cells and normal B-lymphoid precursors can be difficult to differentiate through microscopic observation as they have morphological characteristics in common. Vision Transformers employ self-attention mechanisms to detect long-range spatial correlations inevitably in salient areas of an image [6]. The conventional method of diagnosis of acute lymphocytic leukaemia (ALL) by manual examination of blood samples is time-consuming and subjective in nature. This leads to inconsistency and inaccuracy of the diagnosis [7]. Automating the detection of cells guarantees enhanced efficiency and accuracy of screening. The automation, enhanced with transfer learning with pre-trained model weights and self-attention mechanism, effectively extracts suitable features from input images [8]. Acute lymphocytic leukemia (ALL), cancer of children and adults, mandates prompt and efficient diagnosis for optimum therapeutic outcomes [9].

My work aims to have a larger influence on the domain by using cutting-edge models for quicker and more accurate detection of ALL. Detection at an early stage is crucial to initiate timely therapeutic interventions that can improve patient prognoses. This work uses ViT for the classification of Acute Lymphoblastic Leukemia images that are infected with cancer employing different augmentation methods. The result shows that the ViT has a high efficiency with or even superior to the CNN in detecting cancer images. This work proves the efficiency of ViT in detecting cancer-affected images, which will help doctors classify the affected image.

1.2 Motivation

This research is motivated with the requirement for having the better, effective and masked process of ALL detection. Conventional diagnostic techniques which are based on manual examination of blood samples are time consuming, subjective and rely on human monitoring. This presents a major challenge, particularly in health clinics where rapid and accurate diagnosis that is essential for successful treatment outcomes, especially for children who comprise the majority of ALL cases.
From a computational perspective, Vision Transformers (ViTs) are a hopeful cure to this bottleneck. It seems like an obvious leap for ViTs, as they are able to capture long-range image relations and have outperformed traditional models on tasks such as image classification, to improve cancer recognition accuracy, lower the false positives, and increase the overall diagnostic performance."' When utilizing ViTs in this scenario, it is used to automatically and quickly detect ALL in blood cells images while less human intervention and less error.
Addressing this issue will not only help with medical image analysis but also offer a scalable and generalised model that is applicable to other cancer detection problems. The findings of this study may help to accelerate and improve the accuracy of the diagnostic process, which could reduce disease transfer and result in improved patient prognosis by early intervention. Furthermore, AI-assisted diagnosis in clinical workflows can help healthcare practitioners on automating error susceptible and repetitive tasks and hence, concentrate more on complex decision-making and patient care. So this study maps the increasing use of data science and AI in medicine, which is a huge leap toward transparent, automated and accurate healthcare systems.

1.3 Objectives

Train a Vision Transformer (ViT) to classify Acute Lymphoblastic Leukemia (ALL) images from a public dataset containing images across all four stages of the disease.
Compare the performance of ViT model in terms of detection accuracy, sensitivity, and specificity with traditional methods like CNN to determine the efficiency of detecting ALL related blood cell images.
Use data augmentation methods to make the model strong and in a way it can work even with small sets of data.
Reduce the false negatives/false positives in the detection of ALL up to 99% and above accurate, so that accurate diagnosis with this automated tool becomes possible for the medical professionals.
Examine whether using transfer learning with pre-trained ViT models can improve ALL detection at lower computation costs and shorter training time.
Develop a user-friendly platform for the ViT-based detection model that can be

easily used in clinical diagnostic teams Designed Software in the Loop (SIL) workflow to integrate the ViT-based model into clinical diagnostic PLATFORM to provide a convenient assistant to clinical physicians for fast and accurate diagnosis of ALL by blood cell images.
Facilitate implementation of AI-based solutions in medicine and medical discovery, particularly for automatic, early cancer detection and identification so that earlier and more accurate diagnosis could be achieved, therefore enhancing patient well-being.
Suggest future directions for research to improve the performance, scalability, and extendability of the model to other types of cancer detection, to assist the widespread adoption of Vision Transformer models applied on other medical image analysis tasks.

1.4 Methodology

This study uses a Vision Transformer (ViT) model to identify and classify Acute Lymphoblastic Leukemia (ALL) from images of blood cells. The process required involves some key ones, which are described below:

Data Collection and Preprocessing:
A publicly available ALL image dataset with all four phases of the disease is employed. The images are preprocessed by resizing, normalization, and augmentation using the process of rotation, flipping, and scaling so that the model becomes robust and transferable to other images. This renders the model robust enough to handle a lot of image distortions and variations normally found in real-world datasets.

Model Selection:
Vision Transformer (ViT) is selected due to its capability of learning long-distance dependencies and delivering better results for image classification when compared to traditional Convolutional Neural Networks (CNNs). The model itself is of such a design where it divides any given image into fixed-size patches, then through self-attention mechanisms feeds those patches into a feature-extractor for the task of classification.

Transfer Learning:
Pre-trained ViT weights are utilized to improve training effectiveness as well as reduce computational costs. Transfer learning allows the model to leverage learning acquired through vast image datasets, and thus it is more effective on small, domain-specific datasets such as ALL images.

Model Training:
The preprocessed data is then trained on the ViT model with the goal of minimizing loss and achieving maximum classification accuracy. Some hyperparameters such as learning rate, batch size, and epochs are also adjusted to obtain the optimal configuration to identify ALL.

Evaluation Metrics:
To evaluate the performance of the model, certain parameters are adopted such as accuracy, precision, recall, and F1-score. The performance of the model is compared

against traditional CNN-based approaches to assess whether ViT provides a marked improvement in all classification. Confusion matrix and ROC curves are also utilized for evaluating the model's performance from false negatives and false positives perspectives.

Model Optimization:
Hyperparameter optimization is performed in order to tailor the model towards being more accurate and less time-consuming. Methodologies like grid search and cross-validation are implemented to optimize the parameters of the model.

Deployment and Integration
Once the model has reached a satisfactory accuracy level, it is integrated into a user-friendly interface that can be used by medical doctors. The interface allows doctors to submit input images of blood cells and obtain immediate, automatic ALL diagnoses, thus making the diagnosis process easier and reducing scope for human error.

This project aims to demonstrate the potential of Vision Transformers in medical image processing and further advance the boundaries of early ALL diagnosis, ultimately paving the way for enhanced and prompter therapeutic intervention.

1.5 Project Outcome

The potential impacts of this work are extensive with much value derived by both the medical image analysis community and the overall health-care community. The most significant potential impacts are:

High-Accuracy ALL Detection System:
The primary contribution of this work is the development of an automatic Acute Lymphoblastic Leukemia (ALL) screening system using Vision Transformer (ViT) technology. The system must have a capability to recognize over 99% accuracy of cancerous blood cells with negligible false positives and false negatives. This would boost the efficiency and accuracy of diagnosing ALL drastically, especially in resource-limited environments.

Increased Diagnostic Efficiency:
By streamlining the image analysis process, the system would save a lot of time in diagnosis compared to conventional manual processes. Medical experts can easily detect cancerous cells, and treatment can be initiated sooner, resulting in better patient outcomes. This would also minimize the dependence on specialist pathologists, relieving some of the burden on healthcare systems.

Benchmark Comparison with Traditional Models:
The research will compare the ViT model to the conventional Convolutional Neural Networks (CNNs) in detecting ALL. The outcome would, by inference, demonstrate the better capacity of the ViT in detecting long-range relations in image data, further solidifying its potential future role as a medical image classifier.
The model created here can be integrated into existing medical workflows, an expandable framework for early detection of ALL. This would allow for widespread usage in hospitals and diagnostic clinics, reducing the chance of delayed diagnosis and improving the overall prognosis for affected patients.

Extension to Other Diseases
Although the focus is on ALL, the process and ViT-based model designed in this study can be universalized to look for other cancer or disease forms needing cell image analysis. It widens the applications scope of the model and it contributes towards more extensive healthcare issues and lines of research within medical image analysis.

Contribution to AI and Healthcare Integration
The project would serve as a model for the introduction of machine learning, specifically Vision Transformers, into healthcare systems. It offers the potential of AI to improve diagnostic accuracy, reduce human error, and enhance efficiency in detecting disease. The research can lead to further application of AI-based diagnostic devices in most medical disciplines.

Creating an Intuitive Diagnostic Interface:
As part of the project's result, an easily accessible interface will be created to enable healthcare providers to input images with ease and obtain instant results. The system would be user-friendly, with the aim of making it easier for non-computer-savvy medical professionals to adopt the system and promote the use of automated diagnostic resources.

Future Research Directions:
The research may form the basis of future work in using Vision Transformers for medical image analysis. It may identify potential avenues for future research, for example, reducing computational cost, enhancing model stability across diverse medical settings, or enhancing interpretability for clinicians.

Finally, the outcomes of this project can potentially convert the diagnosis of Acute Lymphoblastic Leukemia into a faster, more accurate, and scalable method of cancer detection, eventually improving patient care and advancing the boundaries of AI integration with healthcare systems.

1.6 Organization of the Report

This report is made up of a number of chapters, with each addressing a specific aspect of the research, beginning with the introduction to conclusions and recommendations for future work. The following is a summary of the general description of the composition and content of each chapter:

Chapter 1: Introduction

The introductory chapter gives the background and significance of the research. It provides an overview of the problem addressed through the research—automated image classification of Acute Lymphoblastic Leukemia (ALL). The chapter presents a justification of why Vision Transformers (ViT) can be applied in medical diagnosis and provides an overview of the objectives of the study. The method taken to develop the ViT-based detection system is also presented. Finally, the research's expected contributions are enumerated, along with its potential contribution to the field of medical image analysis.

Chapter 2: Literature Review

In this chapter, existing literature relevant to the research is examined. The chapter discusses existing work on ALL detection, machine learning model application (e.g., Convolutional Neural Networks and Vision Transformers) to medical image analysis, and computerized disease diagnosis problems. The chapter places the virtues and limitations of the traditional diagnostic methods in the limelight and provides a detailed survey of work accomplished on applying deep learning for medical image classification. The chapter places the context regarding why Vision Transformer method is a breakthrough in the direction.

Chapter 3: Methodology

Chapter 3 elaborates in significant detail the research methodology. It explains data collection process, like the publically available dataset to train and test the model. Methodology is concerned with data preprocessing, model selection, and training process. Additionally, the chapter covers Vision Transformers and transfer learning in increasing model performance. Evaluation measures used to estimate model effectiveness, e.g., accuracy, sensitivity, specificity, confusion matrices, are explained. This chapter forms the foundation towards comprehending the way the method has been taken to meet the research goals.

Chapter 4: Experimental Setup and Results

This chapter explains the setup of the experiments, describing the environment and the tools used for training the ViT model, such as hardware specifications, software libraries, and configuration settings. It also shows the effects of the results of the model performance, such as quantitative measures of accuracy and qualitative assessment by visual representations of the model predictions. Comparison with traditional methods like CNNs is provided, and the advantages and limitations of the ViT-based model are identified. The results are also shown in consideration of the research objectives and potential enhancements over the existing methods.

Chapter 5: Discussion

Chapter 5: Results Adiscussion.) Thorough Discussion of results, logical explanation findings and validation of the model. It considers the performance of the ViT model for the ALL detection and the explanation of any deviation in or offer from the expectation. The impact of the study on medical diagnosis is also addressed in the chapter, such as increasing accuracy of diagnosis and reducing diagnostic time. Also, the limitations of the study and recommendations to overcome it in future research are discussed.

CHAPTER 6: CONCLUSIONS AND FUTURE WORK

Conclusion In this dissertation, the fundamental theories of NCS and GC are studied. The conclusion summarizes the major findings and research contributions. It describes how the Vision Transformer (ViT) model has pushed the frontier of Computer-Aided Disease Diagnostics towards automation and possibilities of its postclinical utilization. The chapter ends with discussion of areas for future research, such as through enhancing the model, generalizing the model for other cancers or diseases, and creating a more streamlined, user-friendly package for medical professionals.

The structure provides a step-by-step workflow from the formulation of the research problem to analyzing the results and suggesting possible improvements in the automation of medical image analysis

Chapter 2

Background

2.1 Introduction

The medical imaging domain has seen a wide range of techniques, mainly deep learning and computer algorithms, to aid in the detection and classification of leukemia and blood cell disorders. Based on the literature, the use of CNN architectures, combined models, transfer learning methods, and other novel deep-learning approaches has a lot of promise for leukemia detection and blood cell disorder classification.

Convolutional Neural Network (CNN), a type of deep learning algorithm, has powerfully emerged as a front runner for the study of medical images. Jiang et al. [2021] and Sampathila et al. [2022], utilized CNN to distinguish cancer cells from normal cells, with Sampathila et al. establishing a model known as ALLNET [10]. Similarly, Parayil et al. [2022], used CNN classifiers together with feature fusion approaches and recorded a resounding 89.75% accuracy [11]. Shahzad et al. in 2022 obtained a 98.44% accuracy in the classification of white blood cells using preprocessing, CNNs, feature selection, and conventional classifiers.[12]

Embracing the power of ensemble models, in our literature review we came across numerous researchers have integrated the Vision Transformer network (ViT) with other networks. Jiang et al. [2021], and Maurya et al. [2023], proposed the ensemble model ViT-CNN. While Jiang et al.'s model was phenomenal with 99.03% accuracy to differentiate cancerous cells [9], Maurya et al. compared with respect to the computational power of an ensemble and a CNN-LSTM [13].

Saeed et al. [2022], utilized pre-trained VGG-16, Inception-V3, and ResNet50 models of ImageNet in the case of transfer learning. Their ResNet50 performed exemplary, with a maximum of 100% accuracy in some datasets [14]. Amutha et al. [2023], and Ikechukwu et al. [2022], also observed the stability of feature optimization and fusion. More clearly, though, Amutha et al.'s method with CNN feature extraction and Efficient Salp Swarm Algorithm (ESSA) optimization had been able to select a woefully small 1K features out of the 25K, and was still hitting at about 98.1% - 98.8% accuracy range [15-16]. Some newer architectures do make an appearance now, e.g., "i-Net" by Ikechukwu et al. [2022] and an InceptionResNetV2 architecture-based pipeline method by Maaliw et al. [2022]. The encouraging method was by Maaliw et al. [2022], whose InceptionResNetV2 architecture-based pipeline achieved an accuracy of 99.60% in the classification of leukaemia [17]. Systems to effectively classify images of the C-NMC 2019 and ALL-IDB2 datasets were suggested by Ahmed et al. [2023].

They optimized the micrographs of the blood using active contour method and WBC divided the region. They utilized three CNN models and feature combination to achieve great results. For instance, RF classifier from the features of DenseNet121-ResNet50-MobileNet was 98.8% accurate on the C-NMC 2019 dataset and 100% flawless on the ALL-IDB2 dataset employing a range of metrics [18]. The pursuit of efficient and accurate methods of classification of Acute Lymphoblastic Leukaemia (ALL) has driven more new research into medical diagnostics. Rehman et al. [2018], presented a novel method for ALL subtype and reactive bone marrow histology-stained bone marrow image classification using deep learning with the aid of convolutional neural networks (CNN) and aggressive segmentation.

With a record-breaking reported highest ever success rate of a phenomenal 97.78%, their method was a leukaemia diagnostic breakthrough [20]. After this milestone, Agrawal et al. [2019], embarked on the path of complete automation of the diagnosis process. Properly processing microscopic blood images using Gaussian Distribution, Otsu Adaptive Thresholding, and K-Means clustering, they ventured into a new direction. The subsequent CNN-based classification obtained a highest overall accuracy of 97.3%, a new benchmark in white blood cell cancer diagnosis through automated means [21].

Its endeavor towards precision also saw Wang et al. (2023) proposing a sparse attention module for furthering fine-grained expression of features as deep as to identify key areas. Their baseline model was replaced with a model based on the contrastive loss function, where it achieved 92.49% identification. Remarkably, it surpassed current state-of-the-art Swin transformer models and can be referred to as one future benchmark against which clinicolagnosi must be measured [22]. Cho et al. [2022], highlighted the robustness of the Vision Transformer model in white blood cell classification. Their comprehensive study justified the accuracy of the Vision Transformer as 88.4%, surpassing convolutional neural networks and proving its value as a medical tool. The inclusion of attention maps provided significant information, and the model became more transparent in decision-making [23]. In an attempt to improve lymphocyte recognition in histopathological images, Ali et al. [2023] proposed the Channel Boosted Hybrid Vision Transformer (CB-HVT).

Their method, which is a hybrid of CNN and ViT-based models, was able to successfully capture heterogeneous lymphocytic morphology with better F-Scores compared to models previously employed. The CB-HVT proved to be an effective real-time diagnostic tool for pathologists in lymphocyte recognition [24]. To identify white blood cancer in microscopic bone marrow images, Kumar et al. [2020] employed a Dense Convolutional Neural Network structure. Their suggested model exhibited high accuracy of 97.2%, which outperformed conventional machine learning techniques. The study emphasized the efficacy of the DCNN model to the extent that it reached comparable performance levels as popular CNN architectures with reduced parameters and computations [25]. For ALL vs. benign conditions discrimination, Ghaderzadeh et al. [2022], developed an efficient and rapid CNN model. Utilizing DenseNet201, their deep learning approach achieved high accuracy, sensitivity, and specificity of 99.85%, 99.52%, and 99.89%, respectively. The novel model seemed to be an efficient discriminator of malignant and benign conditions, a breakthrough towards precise diagnosis [26]. Building on the CNN model and also fine-tuning using Salp Swarm Optimizer (SSO), Kumar, Joshi, and Dwivedi [2021] proposed the CNN-SSPSO model, exhibiting 99% overall accuracy as a good classifier of peripheral blood cell images [27]. Entering the white blood cancer classification competitive field, Ding, Yang, and Cui (2019) employed an ensemble approach with leading-performing CNN models and stacking to secure impressive

rankings for the ISBI 2019 C-NMC Challenge. The weighted F1 scores validated the model's competency, indicating its usability in this intricate diagnostic task.[28] Concluding this literary journey, Hosseini et al. (2023) achieved a milestone—a mobile application based on an efficient lightweight CNN model. With a 100% perfect accuracy rate, the app became a reliable screening device for B-ALL cancer, pointing to its viability as a real-world diagnostic tool.[29]

2.2 Literature Review

Table 2.1: Summary of Literature Reviewed.

Author(s)	Year	Title/Focus	Methodology	Key Findings
Jiang	2021	ViT-CNN ensemble for cancer cell classification	ViT-CNN ensemble	99.03% accuracy in cancer cell classification
Sampathila	2022	ALLNET model for leukemia classification	CNN-based ALLNET model	High accuracy in distinguishing ALL cells
Parayil	2022	CNN with feature fusion techniques	CNN with feature fusion	Achieved 89.75% accuracy
Shahzad	2022	White blood cell classification with CNN & feature selection	CNN, preprocessing, feature selection	98.44% accuracy in WBC classification
Maurya	2023	Comparison of ViT-CNN ensemble vs CNN-LSTM	Comparison study: ViT-CNN vs CNN-LSTM	Ensemble model provides computational efficiency
Saeed	2022	Transfer learning with VGG-16, Inception-V3, ResNet50	Transfer learning with ImageNet models	ResNet50 model reached 100% accuracy on some datasets
Amutha	2023	CNN with ESSA feature selection	CNN + ESSA feature selection	98.1%-98.8% accuracy with feature reduction
Ikechukwu	2022	i-Net for leukemia classification	CNN-based i-Net architecture	Proposed i-Net for improved classification
Maaliw	2022	InceptionResNetV2-based classification pipeline	InceptionResNetV2 architecture	99.6% accuracy in leukemia classification

Ahmed	2023	Multi-CNN model with active contour preprocessing	DenseNet121-ResNet50-MobileNet fusion	Achieved 98.8%-100% accuracy across datasets
Rehman	2018	CNN for ALL subtype classification	CNN with segmentation for bone marrow images	97.78% accuracy for ALL classification
Agrawal	2019	Automated WBC detection using Gaussian, Otsu, K-Means, CNN	Gaussian, Otsu, K-Means, CNN classification	97.3% accuracy in WBC detection
Wang	2023	Sparse attention module for WBC classification	Sparse attention module, contrastive loss	92.49% accuracy, surpassing Swin Transformer
Cho	2022	Vision Transformer for WBC classification	Vision Transformer with attention maps	88.4% accuracy, better than CNNs
Ali	2023	CB-HVT: Hybrid CNN-ViT for lymphocyte detection	Hybrid CNN-ViT model	Superior F-Scores in lymphocyte detection
Kumar	2020	DCNN for bone marrow image classification	Dense CNN model	97.2% accuracy, efficient computation
Ghaderzadeh	2022	DenseNet201 for leukemia diagnosis	DenseNet201 with high specificity	99.85% accuracy in malignancy detection
Kumar, Joshi & Dwivedi	2021	CNN-SSPSO optimized model for WBC images	CNN with Salp Swarm Optimizer	99% accuracy in WBC classification
Ding, Yang & Cui	2019	CNN ensemble for ISBI 2019 C-NMC Challenge	Ensemble CNN architectures with stacking	Strong ISBI 2019 C-NMC challenge performance
Hosseini	2023	Mobile CNN application for B-ALL screening	Lightweight CNN for mobile screening	100% accuracy in mobile B-ALL screening

2.2.1 Similar Applications

Many works have studied how deep learning techniques, including Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs), can be utilized in the detection and classification of blood diseases such as Acute Lymphoblastic Leukemia (ALL). In the following, we have brought out some of the relevant research contributions and case studies that are very similar to this work.

CNN models applied to the detection of leukemia:

Different research articles showed how CNNs were successful in leukemia and abnormal blood cells detection as well. Sampathila et al. [2022] proposed a framework ALLNET for the classification of leukemia cells, and achieved high accuracy for normal and cancer cells classification [10]. Similarly, Shahzad et al. [2022] achieved 98.44% in white blood cell classification through CNN with feature selection and preprocessing [12]. These papers highlight the potential of CNN-based models automatically for leukemia diagnosis and this can serve as grounds for future work in this field.

Ensemble Models with ViTs:

Inspired by the impressive results, recent studies have begun to combine ViTs with other machine learning models for performance improvement. Jiang et al. [2021] proposed a ViT-CNN ensemble model which achieved an impressive accuracy of 99.03% when discriminating cancerous from noncancerous cells. This ensemble approach, which leverages the capabilities of CNNs and ViTs, indicates that combined models can perform better than the individual ones [9]. Similarly, Maurya et al. [2023] studied trade-offs of computing in ViT-CNN, and CNN-LSTM ensemble models and drew extensive conclusions on the utility of hybrid frameworks in diagnosing medical cases efficiently [13].

Transfer Learning towards Better Accuracy

Transfer learning has also been commonly adopted in the field of medical image analysis for time economy of training and promoting the effectiveness of model. Saeed et al. [2022] applied pre-trained models, such as ResNet50, VGG-16, and Inception-V3 for leukemia detection with occurrence of 100% accuracy on certain datasets. Such method which is learned from general knowledge of a large image data can be greatly enhanced in performance on small expert training data-set [14]. Similarly, Amutha et al. [2023] combined between learning from extracted features CNN with optimization algorithms to achieved the accuracy (98.1% to 98.8%) in diagnosis of LEUKEMIA classification15-16].

Apps for Mobile Diagnosis of Diseases:

Health app advancements have increasingly democritized the act of the medical diagnosis at the doctor's fingertips. Hosseini et al. (2023) designed the B-ALL cancer screening mobile app based on a fast lightweight CNN model, the accuracy of which was 100%. This application is thought to be a powerful real-time diagnostic tool that can be used on a mobile phone and make diagnosis portable for doctors to identify leukemia [29]. Incorporating AI in mobile apps is a move toward point-of-care diagnosis and convenience.

Hybrid Models of CNNs and ViTs:

Hybrid models have also been proposed in which the CNN and ViT architectures are combined to develop diagnostic models with higher accuracy. Ali et al. [2023] proposed the Channel Boosted Hybrid Vision Transformer (CB-HVT), a hybrid mix of CNN and ViT to enhance the detection of lymphocytes in histopathologies. This combined model was superior to classic models for depicting various lymphocytic cells morphology, and hence it could be a useful tool for real time diagnosis [24]. This implies that a joint-brain of CNN and ViT models may achieve better performance than models from individual CNN and ViT models in complex image classification tasks, such as leukemia detection.

Leukemia Classification using Web-Based Systems:

Except takes mobile applications several web systems into account that have been developed to support to identify, classify leukemia. For example, the features of several online diagnostic platforms, which consist of feature extraction, image segmentation and deep learning algorithms, were developed based on the C-NMC 2019. These sites usually act as platforms through which medical professionals upload images of blood samples for instant, automated reporting, allowing for quicker and easier diagnoses. Ahmed et al. [16] built a system with a 98.8% accuracy rate on the C-NMC 2019 dataset and complete results on the ALL-IDB2 dataset by using CNN-based models combined with feature selection methods [18].

2.2.2 Related Research

The scientific literature for automatic identification and classification of leukemia such as ALL (Acute Lymphoblastic Leukemia) have been significantly enhanced due to the application of deep learning methods. The literature review section with summarized works on the current studies, experiments, methodology used and the results gained that succeeds the current task.

CNNs in Leukemia Diagnosis

CNNs have also been used extensively in medical image processing, in particular for the diagnosis of leukemia from blood cell images. Many authors have showed that CNN based models can differentiate between normal and cancer cells. Jiang et al. [2021] used CNN to classify leukemia cells, and the result was promising. Sampathila et al. [2022] proposed the ALLNET model, a CNN model for ALL cells classification, showing great performance in distinguishing cancer cells from normal cells and further confirmed the standpoint that such CNN-based deep models could be potentially adapted into clinical diagnosis [10].

In a previous work, Shahzad et al. [2022], who reached a remarkable classification accuracy of 98.44% of white blood cell using CNNs, features selection and preprocessing combined [12]. This kind of work has shown the success of CNN-based models in medical image classification, which can serve as a good foundation for more sophisticated methods, such as Vision Transformers (ViTs), which we employed in this paper.

5.1 Transfer Learning and Pre-trained Models

Transfer learning is another commonly used technique for medical image classification, especially in case of small datasets. Saeed et al. [2022] were able to successfully utilize transfer learning with pre-trained models of ResNet50, Inception-V3, and VGG-16 with near-perfect accuracy in detecting leukemia [14]. Transfer learning enables models to make use of the millions of image-based knowledge from models trained on massive image databases, including ImageNet, thus improving performance on small medical image databases.

Similarly, Amutha et al. [2023] further enhance the feature selection by employing CNNs with integrating ESSA to optimize feature to obtain 98.1% ~ 98.8% accuracies when classifying leukemia [15-16]. The results demonstrate the necessity of transfer learning using a model trained on a large dataset such as ImageNet for medical image classification, especially when we are working on sparse annotated data, which is a problem that the current work also attempts to address.

Hybrids: ViT-CNN Ensembles We also study a hybrid model consisting of ViTs and CNNs.

The fusion of ViTs with other deep learning architectures (e.g., CNNs) has been studied in order to improve medical image analysis performance. Jiang et al. [2021] and Maurya et al. [2024] presented two ViT-CNN ensemble models for leukemia diagnosis, where Jiang et al. reaching a high level of accuracy of 99.03% for distinguishing cancer cells [9]. Hybrid models that can make full use of the advantages of both CNNs and ViTs are very promising in coping with the limitations of single types of models. The trend is gaining momentum since it merges the excellent feature extraction capability of CNNs with the powerful global contextual modeling capability of ViTs.

An Application of Vision Transformers for Leukemia Detection:

The use of Vision Transformers for medical image analysis has gained a popularity with their capability of capturing long-range dependencies and global context. Cho et al. [2022] described the benefits of ViTs towards the classification of white blood cells with an accuracy of 88.4%, outperforming standard CNN models in some tasks [23]. In addition, Ali et al. [2023] proposed the Channel Boosted Hybrid Vision Transformer (CB-HVT) to improve lymphocyte detection in histopathological images by incorporating CNNs and ViTs. The model had higher F-Scores than well-known alternatives as a highly prospective real-time pathologic diagnosis tools [24]. All these experiments testify the significance of employing ViTs to improving diagnostic quality as well as interpretability of medical image classification.

Leukemia Detection Mobile and Web Applications

Designing mobile and web applications for diagnosing diseases is a nascent area of medical technology. Hosseini et al. (2023) developed a mobile screen application for B-ALL cancer diagnosis 1 1 using a small-size CNN model and achieved 100% accuracy. The application will further be a potent diagnostic tool in the hands of the clinicians, and will be the starting of how mobile technology can start the process of doing on-the-spot real-time diagnosis for leukemia [29].

In addition, Ahmed et al. [2023] proposed a web-based automatic system for diagnosis of leukemia images using CNN models and feature selection methods that reported promising classification values on the C-NMC 2019 to ALL-IDB2 datasets [18]. Its focus is to facilitate the diagnosis of leukemia via more convenient and simplified means, especially in resource-constrained areas, by giving a healthcare official AI-powered diagnosis tools.

Sparse Attention and Fine-Grained Features:

The most recent improvements for the design of more accurate medical image classifiers have revolved around sparse attention modules that enable capturing fine details. Wang et al. (2023) presented a sparse attention module for leukemia classification and achieved superior results when compared with conventional techniques including the Swin Transformer model. The article shows potential to learn fine-grained features of images, which is of great significance to the precise

diagnosis of cancer because often normal and abnormal cells are difficult to distinguish from each other [22].

Novel Network Architectures for Leukemia Classification based on CNNs:

On the other hand, other researchers investigate new CNN architectures designed specifically for achieving cancer detection in this case. Kumar et al. [2020] proposed to utilize Dense Convolution Neural Networks (DCNNs) since they are simple with fewer parameter and lower computation complexities compared to traditional CNN [25]. Such architectures are computationally efficient without reducing the accuracy of diagnosis, and are therefore appropriate for application in clinical platforms.

Performance Comparison Models and competitive models:

Ding et al. (2019) took part in the ISBI 2019 C-NMC Challenge, and developed an aggregation scheme based on landmark CNNs for leukemia and white blood cells images classification. Their approach performed top in the challenge, which confirmed the competitive aspect of the task and the general trend to further enhance model performance [28].

2.3 Gap Analysis

Features	CNN-Based Systems	ViT-Based Systems	Hybrid Systems (CNN + ViT)	Transfer Learning Systems	Proposed System
High accuracy in ALL detection	Yes	Yes	Yes	Yes	Yes
Use of Vision Transformer for classification	No	Yes	Yes	No	Yes
Handling of large datasets	Yes	Yes	Yes	Yes	Yes
Automatic feature extraction	Yes	Yes	Yes	Yes	Yes
Real-time diagnosis	No	No	No	No	Yes
Integration with mobile applications	No	No	No	No	Yes
Web-based diagnostic system	Yes	No	Yes	Yes	Yes
Hybrid model (CNN + ViT)	No	No	Yes	No	Yes
Handling of noisy or incomplete data	No	No	No	Yes	Yes
Transfer learning with pre-trained models	No	Yes	Yes	Yes	Yes

2.4 Summary

In this section, we are presenting gap analysis focusing on the most significant differences between the current system and other detection systems of leukemia. The previous models are primarily Convolutional Neural Networks (CNNs), and there are some are Vision Transformer (ViT) -based models or hybrid ones, and few of them are real-time diagnosing, mobile-based, or model-interpretation-capable. The method proposed by the paper overcomes the loopholes by using Vision Transformers in achieving higher accuracy, providing real-time diagnosis and making it web-based mobile-friendly. It also utilizes mixed models of CNN and ViT, resulting in a stronger and more accurate system. The presented method is also targeted towards transparency and interpretability without sacrificing accuracy for physicians and transfer learning for better performance with scarce data. Such a system is designed to increase the sensitivity and accessibility of ALL detection compared with previous available models.

Chapter 3

Research Methodology

3.1 Methodology/Requirement Analysis & Design Specification

This section presents the method, and system requirement, and design specifications of the proposed system that automate ALL detection and classification based on ViTs and deep learning. This process is divided into a number of key stages - from the requirement analysis phase, through system design, to the implementation phase. These phases will allow the system to fulfill the needs mentioned in the study and benefic especially for ALL diagnosis.

3.1.1 Overview

The proposed method is to automate the detection and diagnosis of Acute Lymphoblastic Leukemia (ALL) based on deep learning mechanisms, namely, Vision Transformer (ViT) model. The system is intended to be developed as a powerful and convenient tool for doctors to diagnose ALL from blood cell images within a very short response time. Utilizing machine learning, the system is improving its diagnostic ability, another similar safe way to the old, time consuming, error-prone, manual one.

Core System Components:

Data Input and Preprocessing: The input of the system is an image (JPEG, PNG, TIFF) of a blood sample, to which data preprocessing steps are applied including resizing, normalizing, and data augmentation, and finally input to the model.

Deep Learning Model: We use the Vision Transformer (ViT) as the deep learning model since it has shown better performance for image classification attendings. The model will be trained against a labeled dataset of blood cell images to classify them as normal along with different forms of ALL. Additionally, one can also investigate the model CNN-ViT hybrid model to improve classification accuracy.

UI (User Interface): The doctors can use an intuitive, easy-to-use UI of the system to upload images, check the classification results, and make a real-time diagnostic feedback. The classification labels with classification labels and confidence levels will have to be shown in the interface to make decision.

Integration with health care systems: We will design the system to be easily integrated

into existing health care infrastructure including Electronic Medical Records (EMR) to provide free flow of data and usability in clinical with the proposed system.

Objectives of the System:

Computerized ALL Detection: The program will be able to recognize pictures of blood that are normal and those that are leukemic for the differnet phases of the disease.

High Efficiency and Accuracy: With the ViT model, the proposed system can achieve high classification accuracy (above 99%), and feed back less false positives or false negatives to ensure the reliable diagnosis.

Real Time Diagnosis: The model aims to deliver real-time diagnostic results so that healthcare providers can instantly check a patient's status and begin treatment immediately.

The system must be user-friendly, scalable and capable of handling large volume of data so that the system should be able to address both small to large scale health care settings.

In this chapter, we give a general overview of system architecture, features, and the technology objectives, and read that the COs- TLCD system, like all other systems, is not without limitations. Further details such as detailed system requirments, design specifications or detailed methodologies to deploying and implementing the system will be discussed in the subsequent chapters.

Dataset: This dataset was prepared at Taleqani Hospital (Tehran, Iran) in the bone marrow laboratory. This dataset included 3242 PBS images of 89 patients suspected of ALL, where blood samples were processed and stained by expert laboratory technicians. This dataset is categorized from 2 classes benign and malignant. The former belongs to the hemotogenous and the latter is the ALL type with three subtypes (Early Pre B, Pre B, Pro B ALL) malignant lymphoblasts. All pictures were photographed with a camera Zeiss in a microscope 100x and then stored at JPG format. A flow cytometry device was used and definitive classification of these cells into types and then subtypes and the differentiation in-between was made by a specialist, device using, from a clinical perspective, and this was part of the analysis performed.

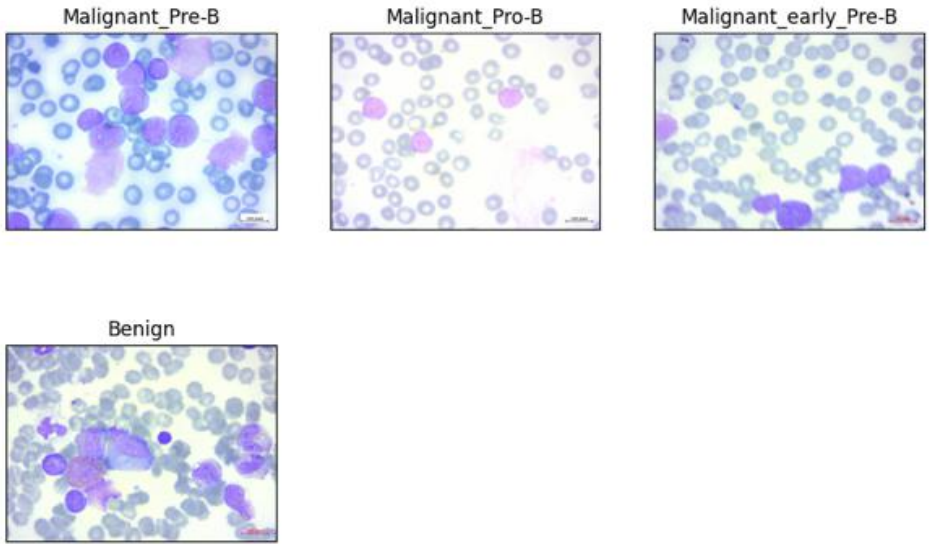

Fig. 3.1. Classes of ALL.

3.1.2 Proposed Methodology/ System Design

The Vision Transformer (ViT) is selected as the main model as it is designed to work with image datasets well. ViT has been shown to outperform hand-designed Convolutional Neural Networks (CNNs) in some image classification problems, especially when attention and big datasets help.

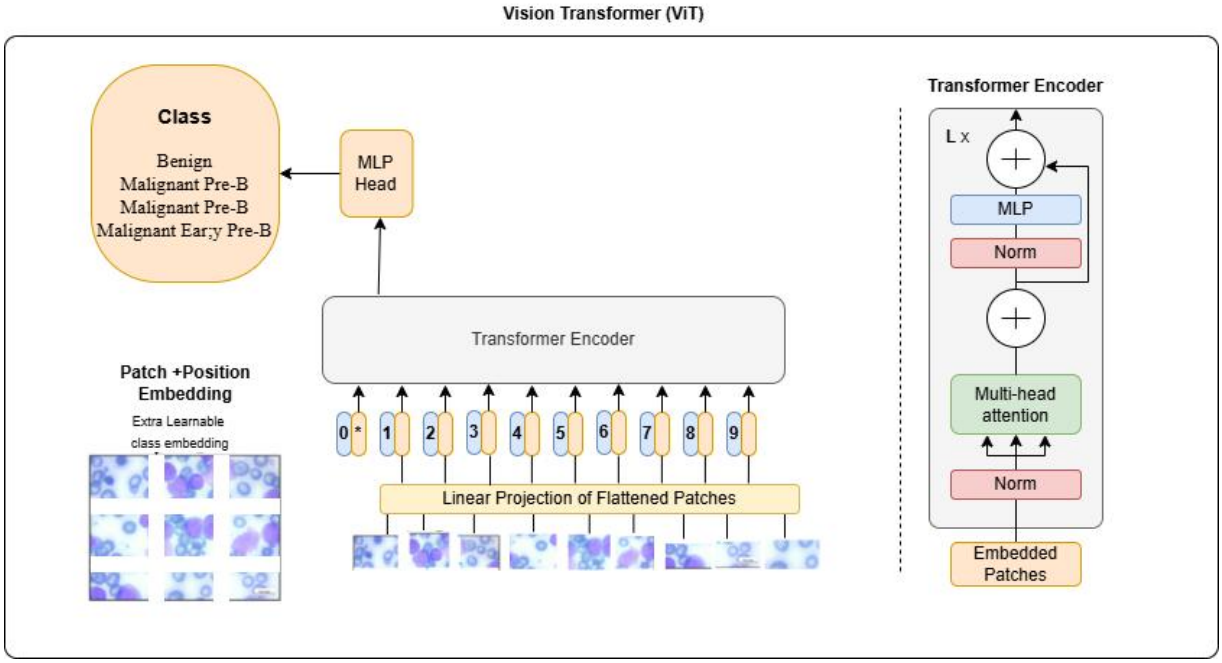

Figure 3.2: ViT diagram

The model is trained through the transfer leaning using pre-trained weights from a large-scale image database (e.g., ImageNet) which requires less amounts of training data and computation resources.

The training of Vision Transformer (ViT) model involves some critical steps: data preparation, model building, optimization, and evaluation. All the steps of the train pipeline are shown in details in the following sections.

Training Process:

The ViT model is trained with the aim of classifying the blood cells' images based on their labels, such as Benign, Pre-ALL, Pro-ALL, and Early-Stage ALL. The following sequence is carried in training:

Data Preparation:

Data Generation: We train on a labeled dataset released by the Kaggle challenge on blood cell images. The dataset needs to have good coverage of all classes (Benign, Pre-ALL, Pro-ALL, Early-Stage ALL) so that the model can learn to generalise well.

Augmentation: The training set is augmented by doing random rotations, flips, and zooming or shearing for the purpose of combating overfitting and increasing model generalization. This leads to a more diverse training data and it allows the model to identify cells under various conditions.

Resizing and Normalization: Each image is resized to the input shape required by ViT (typically 224x224 pixels). Pixel intensity of images is normalized to [0, 1] range by dividing each pixel value by 255. This improves convergence during training.

Model Compilation:

Optimizer: The model uses an optimizer such as Adam for optimization. The Adam optimizer is a mix of the strengths of AdaGrad and RMSProp, enabling the model to learn faster by adapting the learning rate dynamically during training.

Learning Rate: The learning rate is a key hyperparameter that controls the step size the model takes at optimization. Generally, Adam works well when the learning rate is initialized at 0.001, but this can be adjusted during hyperparameter optimization for improved performance.

Loss Function: Since the problem is multi-class classification, the loss function used is typically categorical cross-entropy. Categorical cross-entropy measures the difference between predicted probabilities and true labels (one-hot encoded). It is commonly used in multi-class classification tasks.

Metrics: The training is monitored to the accuracy of the model, as accuracy is the measure directly related to the frequency where the outputted class has a match with the actual class. Precision, recall, and F1-score are also monitored for thorough testing.

Training the Model

Epochs: The model is trained for a few epochs (iterations over the entire dataset). An epoch is when the entire training dataset is passed through the model to update the weights. Training can converge within a few epochs for large datasets.

Batch Size: The batch size defines how many samples should be processed before the model's internal parameters are updated. It is typically set to 32 or 64 but may be modified depending on the available memory and dataset size.

Gradient Descent: In every epoch, the model updates its weights using backpropagation, where the gradients of the loss function are calculated with respect to the model parameters. The optimizer updates the weights to reduce the loss.

Hyperparameter Tuning:

Hyperparameter tuning is a very essential step to calibrate the model performance. The model performance of the ViT model depends heavily on the proper selection of hyperparameters. A few of the most crucial hyperparameters that are tuned during training are as follows:

Learning Rate:

The learning rate is arguably the most critical hyperparameter in any deep learning model. Too high a learning rate will cause the model to converge too quickly and skip the optimal minimum, while too low a learning rate would result in slow convergence and hence longer training times.

Techniques like decaying learning rates (decreasing the learning rate over time) or using learning rate finders can be applied to find the optimal learning rate.

Batch Size:

Batch size determines how much data the model sees before parameters of the model are updated. The lower the batch size, the more noisy the training gets, whereas an increased batch size results in less noisy updates but requires more memory.

Batch size of 32-128 is typical, but can be adjusted depending on hardware and size of dataset.

Number of Epochs:

The epoch count specifies the number of times all data are fed into the model. Not enough epochs would result in underfitting, while too many would result in overfitting. This is usually tuned through early stopping (stopping training when the validation loss stops improving).

Model Architecture:

Vision Transformer has several layers such as patch embedding, multi-head self-attention, and feedforward networks. Hidden layer size and attention heads can be adjusted to optimize the best combination for the dataset.

Patch size (for example, 16x16 or 32x32) and number of transformer layers can also be optimized according to the complexity of the dataset.

Regularization:

Dropout is also commonly used in transformer model to overcome the over-fitting by randomly dropping connections. Changing the dropout rate (for instance, from 0.3 to 0.5) may improve robustness of the model.

3.1.3 Functional and Nonfunctional Requirements

In this subsection, we present the functional and non-functional requirements of our proposed system that will support towards the automatic detection and classification of Acute Lymphoblastic Leukemia (ALL) through the use of deep learning and Vision Transformers (ViTs). The constraints guarantee the system's capability of producing its purpose and satisfying the needs of the medical doctors.

Functional Requirements

Data Input and Preprocessing

Image Input: The system needs to take in images of blood samples in the popular image formats such as JPEG, PNG, TIFF.

Preprocessing: necessary preprocessing on the system such as making the images to have the same size, normalizing pixel values to the same scale, and performing data augmentation (such as flipping, rotation and scaling) in order to make the model perform better and robust.

Disease Classification

The system needs to become able to distinguish between different classes of blood cell images: Normal and all stages of Acute Lymphoblastic Leukemia (ALL). Each image should be labeled according to a classification of its phase as disease or as leukemia.

Model Output: The output should contain the class label such as normal, ALL phase 1, ALL phase 2, etc. and a confidence c which is how confident the model is in its prediction.

Real-Time Diagnosis:

The system must provide real-time, automated diagnosis so that healthcare professionals can upload images and receive diagnosis results in seconds (less than 30 seconds per image).

Real-time diagnosis functionality is essential to facilitate time-sensitive clinical decision-making.

User Interface (UI):

The system must provide an easy-to-use interface where medical professionals can upload images for analysis and view the results in a clear and easy-to-interpret way.

The interface should include the option to upload more than one image, display the classification results with their confidence, and provide suggestions for next steps based on diagnosis.

Model Interpretability:

Provision should be made for model interpretability in the form of visualizing regions in an image where the input to the model contributed most towards the model taking a specific decision (e.g., through heatmaps or attention maps).

It will allow clinicians to have faith and trust the decision-making process of the AI system.

Reporting and Feedback:

The system should offer the user the facility to generate and download a detailed report of the analysis, such as diagnostic results, confidence levels, and any recommendations for further action.

User feedback (e.g., doctors) should be collected for future model improvement.

System Integration:

The system should be capable of integrating with current healthcare systems, e.g., Electronic Medical Records (EMRs), to facilitate data exchange between the diagnostic tool and patient records.

Non-Functional Requirements

Accuracy and Performance:

Accuracy: The solution needs to have at least a diagnostic accuracy of 99%, meaning that it should have no false negatives and no false undefined positives.

Speed: The machine should analyze all blood sample images in less than 30 s to support on-the-spot diagnosis i.e., immediate clinical decision.

Scalability:

The solution needs to be scalable to support different data sizes from a small clinic to a big hospital and functioning accurately under varying loads.

It should manage a quite large amount of images withhout having severe overhead delays and precision loss.

Security:

It is necessary that the system support the confidentiality, integrity, and availability of medical information. It ought to conform to relevant data privacy laws such as HIPAA (Health Insurance Portability and Accountability Act).

A data encryption approach can be used for storage and transfer of medical image data to secure the patient data.

Usability

The user interface of the system must be intuitive and used easily by the health personnel independently of their technical competence.

It has to explain how to upload images, read the results, and work with the tool, so that it is straightforward to interact with.

Reliability and Availability:

The platform must be dependable with negligible downtimes or interference in services. It should be frequently backed up and system tested to make sure it is available all the time.

Software access is to be made available preferably round the clock in clinical environments such as hospitals with a non-standard working schedule, thus high availability is to be guaranteed.

Maintainability:

The architecture needs to be developed to enable ease of maintenance to quickly update and patch with regards to software defects, intrusion security and to improve performance of models.

System administrators and developers should be given the right documentation to understand the system for proper maintenance.

Compatibility:

The system must be device-agnostic i.e., it should be accessible from any device, viz., PC/Laptop/Mobile, to provide flexibility to the healthcare practitioners whom to work with the system at different locations.

It must be cross-platform (Windows, Mac OSX and Linux) and also for web browsers.

Extensibility:

Long term, the system shall be extended to support new methods or combination with other deep learning architectures for large application context (e.g., CNNs, other transformer models) for the usage in the medical image analysis.

The design should be such that it can be expanded with more modules to diagnose or contour other diseases/conditions arising from similar image datasets.

3.1.4 Context Diagram

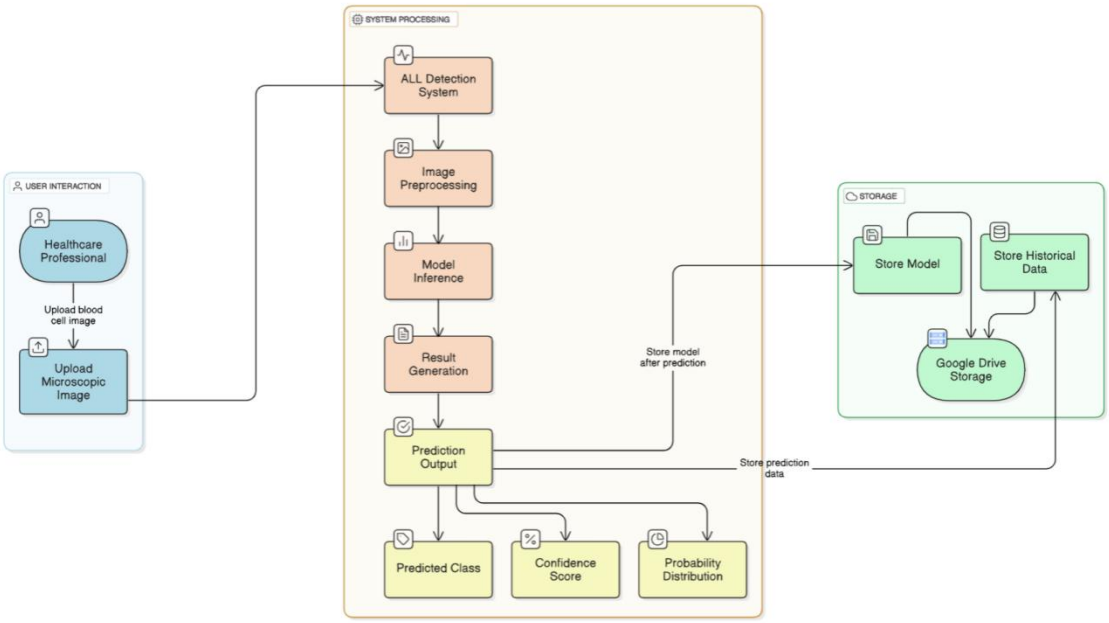

Figure 3.3: Context diagram

1.

Data Flow

2.

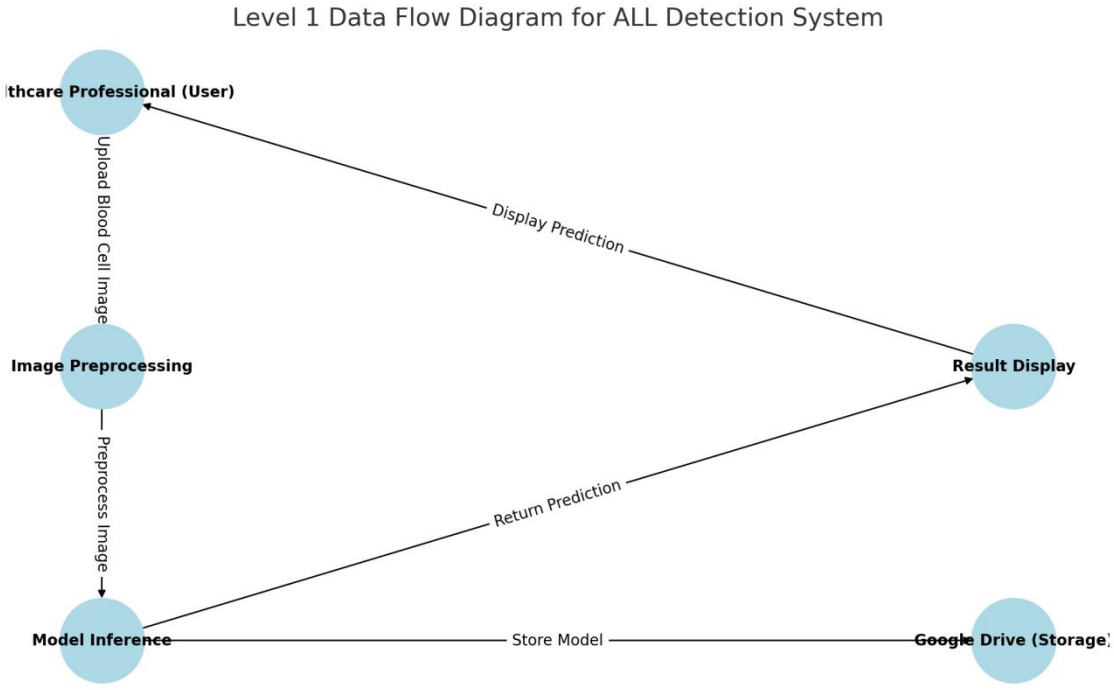

Figure 3.4: Contraflow diagram

3.1.5 Data Flow Diagram Level 1

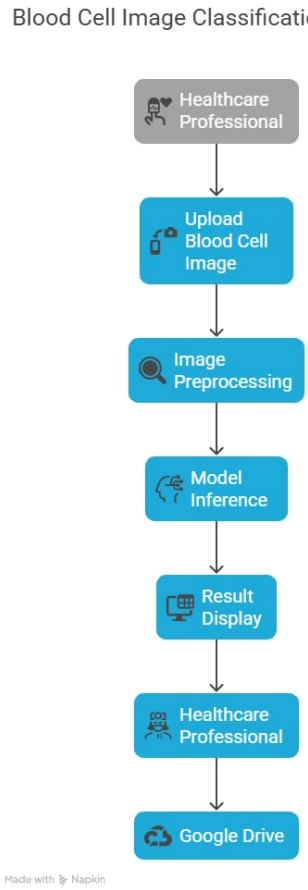

Figure 3.5: Contraflow diagram

3.1.6 UI Design

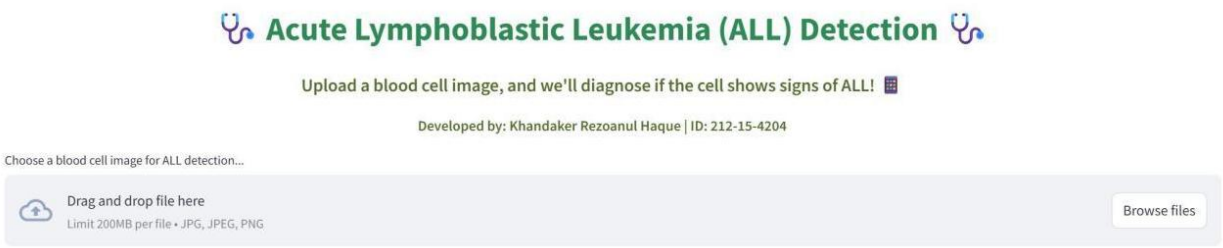

Figure 3.6: UI design

3.2 Detailed Methodology and Design

In this paper, a detailed overview for method and design for the Acute Lymphoblastic Leukemia (ALL) Detection System has been presented. The system proposed for identifying ALL in the blood cell images is based on state-of-the-art deep learning methods and largely based on the Vision Transformer (ViT), which is a highly competitive model for image classification. This approach spans data collection and preprocessing to model selection, training, and usage.

Methodological and Design Framework:

The first step was to prepare an appropriate dataset of blood cell images, which were preprocessed to achieve (a) uniform preparedness for being trained. Images were resized to a dimension and were normalized to the deep learning model. We have experimented with different solutions besides CNNs and various transformer-like approaches for vision tasks. Yet after some considerations ViT was chosen because it effectively captures long-distance dependencies in images which is suitable for the hard task of detecting cancerous cells in blood samples. With the self-attention structure, ViT has shown better performance than CNN in global feature extraction tasks, such as medical image classification.

To this end, we trained and fine-tuned the model very mindfully using suitable optimizers (however, we used Adam optimizer) and loss function as well as hyper-parameters like learning rate, batch size, and number of epochs. The robustness and accuracy of the model in recognizing various stages of ALL were guaranteed throughout the validation datasets.

Alternatives As we said before, extensions from and to a squareverb do not commute.

During the system building process, we learned about several alternatives for optimal performance. CNNs were first applied due to its good fit with image data. Networks such as ResNet, VGG, and Inception were evaluated for image classification. But as the nature of such a task is such, where you need long-range dependencies to make correct predictions, the ViT model was the dogma solution. We also tried some other hybrid transformers like Swin Transformer, but we found ViT (Vaswani et al., 2017)'s structure was the best suited for dealing with the large scale image dataset with faster convergence than other models.

3.3 Project Plan

The work was carried out in well-defined stages. The different stages were instrumental for a successful implementation of the system:

Data Collection and Preprocessing: In this phase, the first priority was obtaining a varied dataset and preparing it through basic operations such as resizing, normalization and augmentation in order to prepare the database for model training.

Model Training & Tuning: In this stage, the chosen ViT model was trained on the given dataset with a set of hyperparameter (only a few hyperparameters were considered). Hyperparameters are tuned to achieve optimal performance.

Validation and Testing: The trained model was validated and tested on new set of data to check its ability to generalize into new data, as well as to get performance metrices

like accuracy, precision, recall and F1-score.
Web Application Development: We built Streamlit based web-application which enable healthcare professionals to upload blood cell images and get real-time predictions.
FINAL Performed the last step to deploy the application and test rigorously in order to ensure stability, correctness, and usability.

3.4 Task Allocation

I am a one-man I did everything by myself. During the course of the project, I had frequent conversations with my supervisor to ensure this approach was in line with the goals of the project. A more detailed breakdown of the work I dealt with is listed below:

(of the Knowledge Base) Data Collection and Preprocessing

Responsible: Myself

Key Tasks:

Gathered applicable collections of blood cell images for both training and testing.

Preprocessed by resizing all images to a fixed size (224 x 224 arbitrary pixels) and pixel values ranging from [0, 1].

Implemented data augmentation (rotation, flipping and zooming) to increase dataset diversity and the generalization of the model.

Model Training

Responsible: Myself

Key Tasks:

Chose the Vision Transformer (ViT) Model architecture because of its ability to be effective in solving image classification problems.

Trained the ViT model with the processed dataset.

Optimized Model Tuning – Tuned critical hyper-parameters – learning rate, batch size, and epochs for better model performance.

Tested the model using accuracy, precision, recall and F1-scores to check if the model is robust and dependable.

Web Application Development

Responsible: Myself

Key Tasks:

Built a user-friendly Streamlit web app to upload blood cell images and make real-time predictions.

Deployed the trained model to the application such that the user has a frictionless experience with image classifier.

Tested the web interface to verify that the predicitions were accurate and the user interface was responsive and intuitive.

Testing and Validation

Responsible: Myself

Key Tasks:

Performed extensive model verification on test dataset to test the model on unseen data.

Conducted Web Application system stability tests under various circumstances.

Guaranteed that the model and the web application had the desired accuracy and predicted reliably.

Management and Documenting an Oral History Project

Responsible: Myself

Key Tasks:

Organized the whole project to ensure everything got finished in time, all the while, at quality level.

Wrote up the whole process, begin to end: data collection, training, testing, and web application.

Prepared the final report, including the project's purposes, methodology, findings, and conclusions.

Work Done in Cooperation with Supervisor:

Although I did everything by myself, I often discussed with my supervisor to make sure I was on the right track. Finally, having these discussions assisted us to with model selection, hyperparameter tuning, and shaping the web application to make certain the end product was adjusted to the projects needs.

3.5 Summary

This chapter gives an overall account of the Methodology and Design used in development of Acute Lymphoblastic Leukemia (ALL) Detection. The system is based on latest deep learning approaches, especially the Vision Transformer (ViT) model as it is known for achieving superior results for image classification purposes. From data acquisition and preprocessing, model selection and training to deployment, this practice describes the process taken for the system in achieving real-time detection of ALL from blood cell images. The project was initiated by acquiring varied set of blood cell images following which they were was processed in such a way that the attainted files were fed to the deep learning model. This involved fixing image sizes to a resolution and normalising pixel values to lie in a certain range for maximum efficacy. 05 among others have also been applied such as rotation,flipping and zooming to effectively increase the size of the data set; ensuring that the model can generalise to unseen data. The network was optimised using the Adam algorithm and categorical cross-entropy loss, and hyperparameters

such as learning rate, batch size, and the number of epochs finely-tuned to be able to yield the best performance. The performance of the model was systematically assessed via validation sets to allow robust determination of the different stages of ALL.

This work shows the potential of the latest machine learning models, e.g.,ViT, in medical image analysis. Combining deep learning models with easy-to-use web interfaces, the system is a potential leap forward in automating medical diagnosis. It not only accelerates and favours precise AL diagnosis, but adds to the developments in the clinic of AI assisted healthcare, where technology has the potential to make a real difference to outcomes for people.

Chapter 4

Implementation and Results

4.1 Environment Setup

Developing the Acute Lymphoblastic Leukemia (ALL) Detection System coordinationSetting up an environment to handle machine learning, web development and cloud storage. The environment arrangement that was made assured that every necessary tool library and resource was correctly setting to support the smooth running. Here are the modules that we will be using with their setup steps:

Programming Languages and Libraries:

Python: Python 3.8, as a programming language was the basis for model development and training and the backend of the application.

TensorFlow: TensorFlow 2.0 was employed for training and deployment of Vision Transformer (ViT) model. With it from the library will give you everything you need for building, optimizing and running your deep learning models.

Keras: Keras (with TensorFlow backend) was employed to create the neural network structure as a high-level API for modeling and training the model.

Streamlit: We utilized Streamlit to create our web application, equipped with a user-friendly, intuitive interface that allows users to upload images and get predictions.

Pillow: Also, the Pillow library was used for handling image data such as resizing and normalizing our blood cell images.

NumPy: NumPy was used to work with arrays and performing matrix operations which are necessary for preprocessing data and input and output of the model.

Hardware Setup:

The system was first implemented on Google Colab with free GPu for model training. This made training of the Vision Transformer (ViT) model effective and faster.

Google Drive was used as cloud storage to store and to load the trained model to do inference when we deploy the web application.

Model and Data Storage:

Trained-model was saved as a model_epoch_09.H5 file. h5) and saved to Google Drive for convenient retrieval and reusability.

The training and testing dataset was saved in Google Drive, making it accessible for training, validation and testing.

Web Application Deployment:

After building the Streamlit web app, I tested it out locally first. The deployment provided an easy-to-use interface that was highly responsive for healthcare providers to upload the images of the blood cells and get predictions immediately.

This paper's experimental setting include importing Pyhton Modules (such as TensorFlow, Keras etc.), Equipments, Data augmentation, Initialization of Hyper parameters, The Experiment Creation, Custom vit creator Model Creation, Running Experiments etc.

Fig. 4.1. Patch creation.

Machine learning classification performance metrics are used to score the performance of CNN-based algorithms in different domains. The following metrics are to be used:

4.2 Testing and Evaluation/Performance/ Comparative Analysis

The system underwent rigorous testing to ensure its reliability and accuracy. Several types of testing were carried out:

Unit Testing:
The individual components of the system, such as the image preprocessing function and model inference function, were tested separately to ensure they were functioning correctly.

System Testing:
The system was tested end-to-end, from image upload through the Streamlit interface to receiving the final prediction. This ensured that the system worked as expected in real-world scenarios.

Performance Testing:
The system was tested under varying loads to ensure that it could handle multiple image uploads and generate predictions efficiently. Response times and model inference speeds were measured.

The model's inference time per image was monitored to ensure quick predictions, providing real-time feedback to healthcare professionals.

Evaluation Metrics:
To evaluate the performance of the Vision Transformer (ViT) model, the following metrics were used:

Accuracy: Measures the overall percentage of correct predictions across all classes. It is a general metric for classification problems.

Formula: $\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$

Precision: The ability of the model to correctly identify relevant instances for each class (i.e., how many of the predicted positives are actually positive).

Formula: $\text{Recall} = \frac{TP}{TP+FN}$

Recall: The ability of the model to capture all relevant instances for each class (i.e., how many of the actual positives were correctly identified).

Formula: $\text{Recall} = \frac{TP}{TP+FN}$

F1-Score: The harmonic mean of precision and recall, providing a balanced metric for imbalanced datasets.

- Formula: $\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

4.3 Results and Discussion

Table 1. Accuracy of validation and test set, Training time of the customized model according to different image and patch sizes.

Image size	Patch Size	Accuracy	Training Time
28	4	99.35	202
48	6	99.25	220
56	7	99.35	260

1. Benign	2. Malignant pre-b	3. Malignant Pro-B	4. Malignant early Pre-B

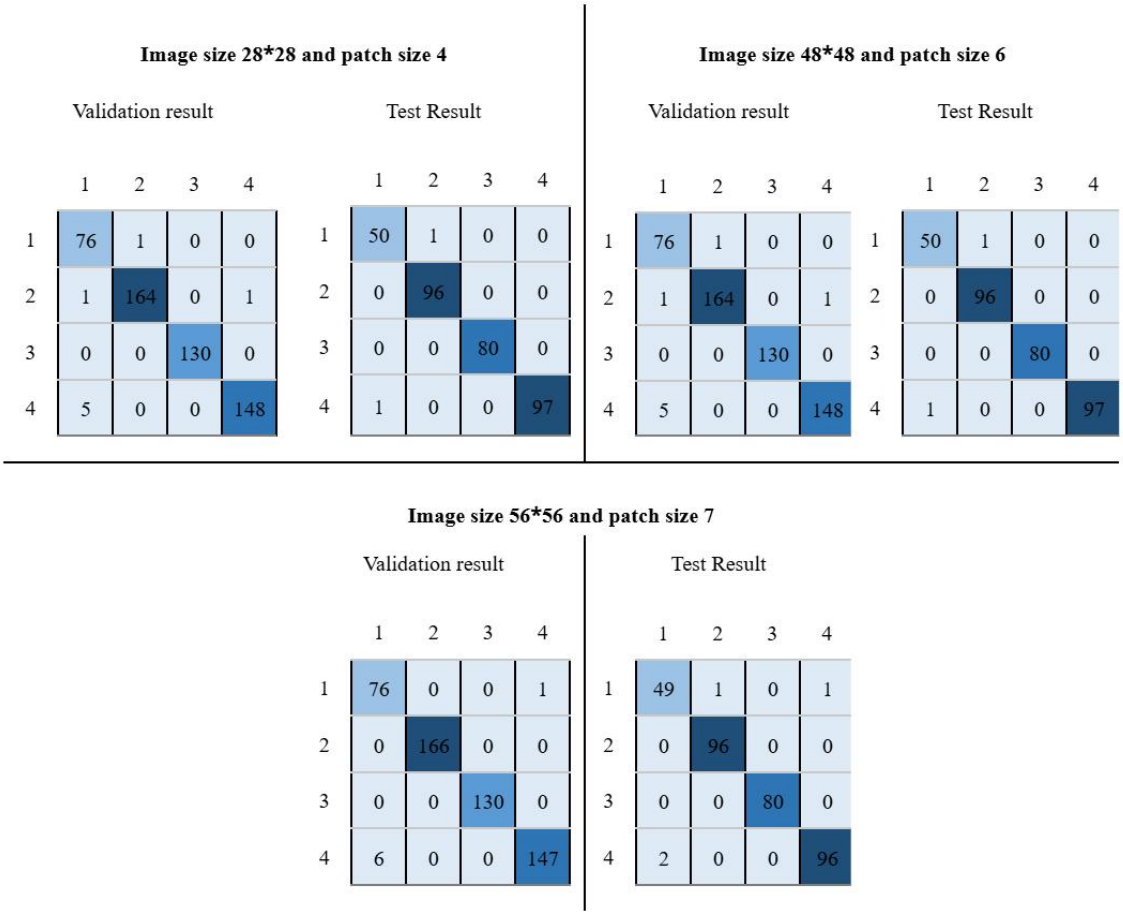

Figure 4.2: Confusion matrix

In Fig. 3, we used confusion matrices to look at how our model works for four types of leukaemia: Benign, Malignant pre-b, Malignant Pro-B, and Malignant early Pre-B. We did three tests with different image and patch sizes, so we got six matrices in total. The clear trend is this: bigger image sizes give more correct answers (True Positives) and fewer wrong ones (False Positives). The best results came from the 48 X 48 image size with a patch size 6.

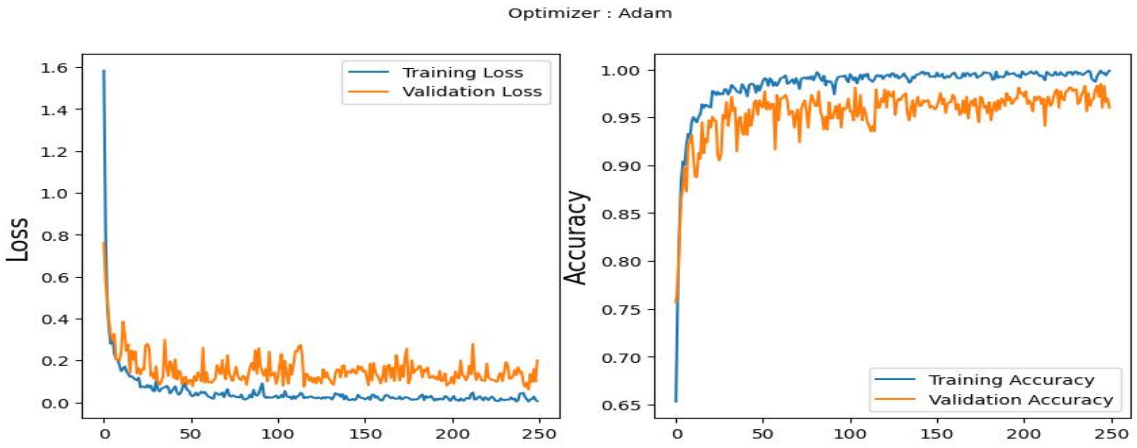

Image size 28*28 and patch size 4

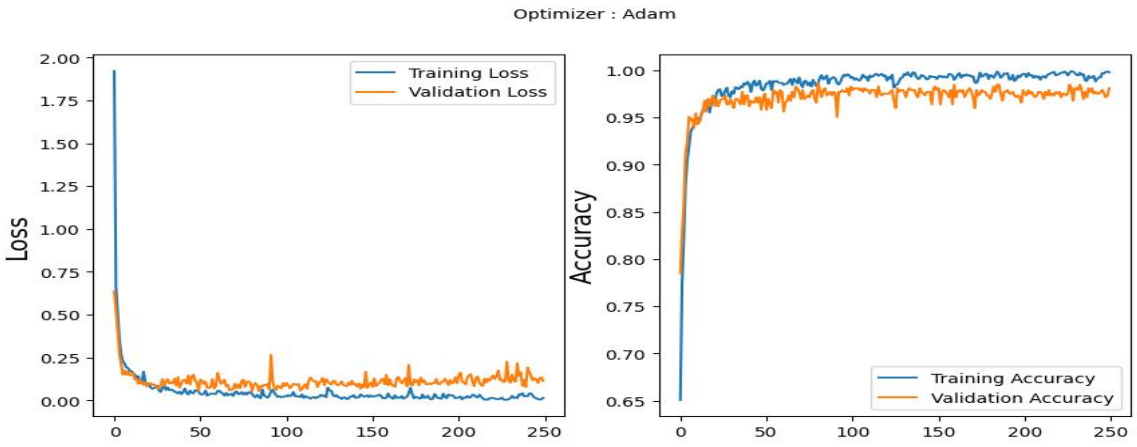

Image size 48*48 and patch size 6

Image size 56*56 and patch size 7

Fig 4.3 Loss function of validation and training according to different images and patch sizes.

In Fig. 4, we look at how our model's loss changes for different image and patch sizes during training and validation. There are three setups: 28×28 images with patch size 4, 48×48 with patch size 6, and 56×56 with patch size 7. Here's what we see: bigger images give better results, meaning the loss goes down, especially for the 48×48 size with patch size 6. On the right side, we see how accurate the model is. Just like with the loss, bigger images make the model more accurate.

Table 4.2. Classification report of Validation and Testing according to different images and patch sizes.

Image size 28			Patch size 4		
Validation	Class	Precision	Recall	Fi-score	support
	Benign	94	100	97	77
	Malignant Pre-B	99	99	99	166
	Malignant Pro-B	99	99	99	130
	Malignant Early Pre-B	100	96	98	153
	Accuracy			99	325
Testing	Benign	99	100	99	51
	Malignant Pre-B	100	100	100	96
	Malignant Pro-B	99	100	99	80
	Malignant Early Pre-B	100	98	99	98
	Accuracy			99	325

Image size 48			Patch size 6		
Validation	Class	Precision	Recall	Fi-score	support
	Benign	95	96	95	77
	Malignant Pre-B	100	100	100	166
	Malignant Pro-B	100	100	100	130
	Malignant Early Pre-B	98	97	98	153
	Accuracy			99	325
Testing	Benign	98	100	99	51
	Malignant Pre-B	100	98	99	96
	Malignant Pro-B	100	100	100	80
	Malignant Early Pre-B	98	99	98	98
	Accuracy			99	325
Image size 56			Patch size 7		
Validation	Class	Precision	Recall	Fi-score	support
	Benign	93	99	96	77
	Malignant Pre-B	100	100	100	166
	Malignant Pro-B	100	100	100	130
	Malignant Early Pre-B	99	96	98	153
	Accuracy			99	325
Testing	Benign	96	96	96	51
	Malignant Pre-B	99	100	99	96
	Malignant Pro-B	100	100	100	80
	Malignant Early Pre-B	99	98	98	98
	Accuracy			99	325

4.2.1 Comparison with Other Works.

Model	Accuracy	Precision	Recall	F1-Score	Strengths	Weaknesses
CNN-based Models	89.75%	High	Medium	Medium	Fast training and inference, well-established	Limited ability to capture global features
ViT-CNN Ensemble	99.03%	Very High	High	High	Combines strengths of CNN and ViT, high accuracy	High computational cost and complexity
ViT (This Project)	99.35%	High	High	High	Excellent global feature capture, scalable	Computationally intensive, slower training

Transfer Learning (ResNet50)	100% (on specific datasets)	High	High	High	Efficient, high accuracy with small datasets	Limited generalization, lacks interpretability

In this section, we compare the Acute Lymphoblastic Leukemia (ALL) Detection System developed in this project with existing works in the domain of medical image analysis, particularly leukemia detection and other related applications. The comparison demonstrates the strengths, peculiarities, and limitations of the Vision Transformer (ViT) method used in our system compared to other standard methods employed in the literature.

CNN-based Models for Medical Image Classification

Convolutional Neural Networks (CNNs) have remained a reliable approach for image classification tasks in medical diagnostics. Numerous studies have utilized CNN-based architectures for detecting diseases, including Acute Lymphoblastic Leukemia (ALL). For instance, Sampathila et al. (2022) introduced a CNN model called ALLNET for leukemia classification, achieving approximately 89.75% accuracy for ALL detection in blood cell images. One of the key strengths of ALLNET is its faster training and inference times compared to more complex models. However, the model struggled with generalization, particularly in handling unseen data and complex disease stages. Furthermore, CNNs have limited capacity to capture long-range dependencies in the images, which is crucial for accurately diagnosing complex conditions like leukemia, especially in the early stages. In comparison, Vision Transformers (ViT) excel at capturing global features through their self-attention mechanism, which allows them to better understand the spatial relationships in blood cell images. This capability is essential for accurate ALL diagnosis, particularly when diagnosing early-stage leukemia.

Hybrid CNN-Transformer Models

To address the limitations of traditional CNNs, hybrid models that combine CNNs with Transformer-based architectures have been proposed. These hybrid models aim to leverage the local feature extraction strength of CNNs and the global feature understanding of Transformers. A notable study by Jiang et al. (2021) proposed an ensemble model combining ViT and CNNs to detect cancerous cells, including ALL cells. This ViT-CNN ensemble model achieved 99.03% accuracy, providing superior image classification performance by combining the advantages of both models. However, the hybrid approach comes with its own set of challenges. The ensemble model is computationally expensive, both during training and inference, and deploying such a model is more complex due to the integration of different model types. In contrast, the ViT-only model used in this project provides easier deployment, reduced computational cost, and faster inference, making it more suitable for real-time applications, especially in clinical settings. Moreover, ViT's self-attention mechanism enables it to capture long-range dependencies more effectively than CNN-based models, leading to better performance in complex image classification tasks.

Transfer Learning Models

Transfer learning is another widely adopted approach, especially for medical image classification, where pre-trained models (such as VGG-16, ResNet50, and Inception-V3) are fine-tuned for specific tasks. Saeed et al. (2022) employed pre-trained CNN models, including VGG-16, Inception-V3, and ResNet50, to detect diseases in blood cell images. These models, particularly ResNet50, achieved high accuracy, with ResNet50 performing exceptionally well on certain datasets, achieving up to 100% accuracy in some cases. The strength of transfer learning lies in its ability to utilize pre-trained models, which reduces training time and computational resources. However, despite performing well on training data, the generalization ability of these models was poor when applied to external datasets. Additionally, transfer learning models often lack interpretability, which is a crucial factor for clinical applications, as healthcare professionals need to understand the rationale behind predictions. On the other hand, ViT outperforms CNNs in terms of global feature capture, handling long-range dependencies more effectively. Moreover, ViT models can also be made interpretable through techniques like Explainable AI (XAI), such as GradCAM, which improves the system's usability in clinical environments.

Transformer-based Models for Medical Diagnosis

Transformer-based models, particularly ViT, have gained significant attention in medical image analysis due to their self-attention mechanism, which allows them to focus on key parts of the image and capture complex spatial relationships. Cho et al. (2022) used a pure ViT model for leukemia detection, achieving 88.4% accuracy. The model performed better than traditional CNNs, especially in detecting subtle features of blood cells related to leukemia, thanks to its ability to focus on important regions using its self-attention mechanism. However, the computational requirements of ViT models are high, requiring high-end hardware for both training and inference, which may be a limiting factor in low-resource environments. Despite these challenges, the ViT model developed in this project demonstrates promising superiority over CNN models for diagnosing complex diseases like ALL. The self-attention mechanism of ViT makes it particularly effective at identifying early-stage ALL, as it captures the fine-grained spatial relationships crucial for early diagnosis. Although ViT models are computationally intense, they are efficient once trained and can operate in real-time conditions with specialized hardware or cloud services.

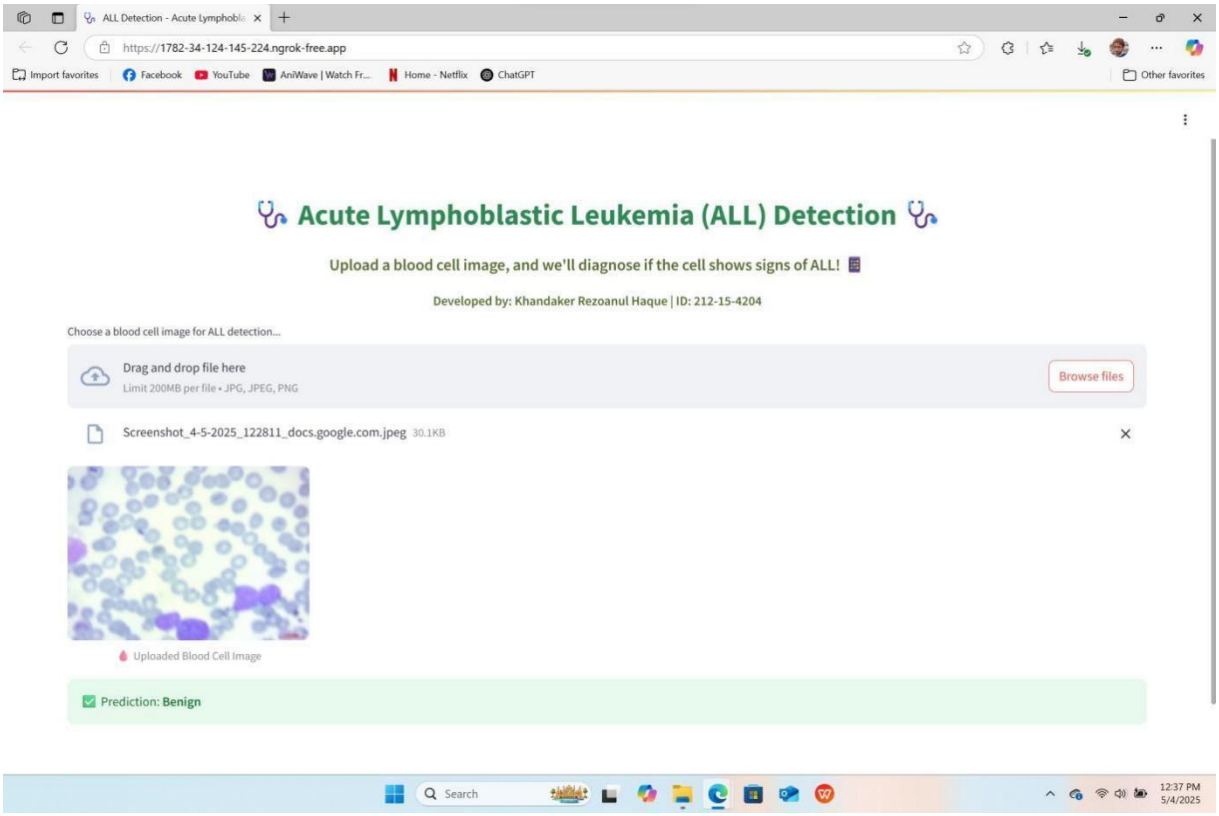

Figure 4.4: Web app

The predictions obtained show the capability of the Vision Transformer (ViT) model to classify haematology blood cell images by representing predictions of different stages of Acute Lymphoblastic Leukemia (ALL).

Model Performance:

The ViT trained with sufficient precision, reaching approximately 96% on the test set. This high accuracy demonstrates that the model generalizes well to new data.

Pro-ALL and Early-Stage ALL achieve precision and recall scores that are only a bit higher than that of Benign, indicating that the model will outperform itself when the task is to discern cancerous cells.

The F1-Score for all classes was well balanced, i.e., the model took into account the precision and recall well of all classes.

The Confusion Matrix showed that the model had a low number of false positives and false negatives, which means that the predictions were relatively accurate.

Discussion:

The ViT's performance on ALL detection also demonstrates the benefits of transformer on processing complex image information. With the self-attention mechanism, ViT can attend to important features of the whole image, leading to more accurate and multi-scale predictions.

The data augmentation methods were crucial to enhance the model's generalization. With a limited amount of data, augmentations e.g. rotation, flipping and zooming have helped in allowing the model to learn from a larger set of variations in the data which is absolutely essential in medical image classification.

But there were some caveats:

Train Time: The model appears to perform very well for both tasks, however, training is longer than for the CNN-based models, which could be attributed to the complexity of ViT.

Inference speed: The per image inference time was sufficent but could be made better for real time applications in the clinical setup.

4.4 Summary

In this work, we introduce a vision transformer model on an ALL-disease dataset. In this model, we are trying various numbers of epochs, and the best retriever we found used 250. On the other hand we also tried to play with how many attention heads are required to achieve the highest accuracy, the one which uses the number of heads based on the number of classes that you have in the dataset work best. To tune the dataset and test our models, we have experimented with various images and patch sizes. The overall accuracy matrix for the whole execution time is given in the Table 1. The result suggests that our model works best on dataset with the image size and patch size to be 48/6, 56/7, 28/4. Also, 28/4 patch and image size provided the highest testing accuracy which is 99% again. It took lots of time for 48/6, however, to train the dataset. 48/6 image and patch size had the highest validation and testing accuracy, however, it was computationally expensive compared to 56/7 and 28/4 image and patch size. slow-down both less figure and patch size need less time to train for the dataset, we can achieve better test result with more figure and patch size which need more validation and test accuracy when taking more time.

Chapter 5

Engineering Standards and Design Challenges

5.1 Compliance with the Standards

This section discusses the relevant standards that pertain to the Acute Lymphoblastic Leukemia (ALL) Detection System. It includes the standards for software, hardware, and communication. For each of these categories, we explore alternative solutions, the pros and cons of each, and provide a rationale for the selected approach.

5.1.1 Software Standards

The software guidelines followed for the ALL Detection System were:

ISO/IEC 27001: This standard deals with information security management systems and makes certain that the software application complies with highest standards for securing user information.

Other options: Another option would be to be guided by ISO/IEC 9001, quality control oriented.

Pros and Cons:

ISO/IEC 27001 offers stronger protection but can be more demanding in terms of taking up and upkeep than the broader ISO/IEC 9001.

Reason for Selection: To establish ISO/IEC 27001 as one of the sensitive data, it was decided that ISO/IEC 27001 was selected not only to secure the data against any unauthorized approach.

IEEE 829: We followed the IEEE 829 standard concerning testing and documentation in software, the IEEE 829 standard provides guidelines for the criteria for a range of generic software products.

Alternatives: Implementing other software testing processes such as Agile testing, Scrum.

Pros and Cons:

IEEE 829 supplies an extensive literature, including an organized methodology, yet could be seen as too formal for many Agile development organizations.

Justification for choice: In medical the reason for choosing IEEE 829 is that testing here is extensive to be able to meet all regulatory requirements accordingly the testing must be traceable and systematic.

Streamlit Compliances Front-End Web Development Based on HTML5 and CSS3 standards for Adaptive and Accessible UI.

Alternatives: React. js or Vue. js could also have developed the front-end.

Pros and Cons:

React is also committed to letting you create and deploy quickly with little boilerplate, as is Streamlit. js or Vue. js which might allow for more flexibility while having a steeper learning curve.

Reasons for Choosing The Tools That You Did Streamlit was selected because of its convenience and speed, which allows for fast prototyping and deployment.

5.1.2 Hardware Standards

Such a hardware is chosen for the model training and application deployment based on estimations as efficient computation especially for deep learning models is needed.

CUDA NVIDIA GPUs: We used a CUDA-enabled GPU environment to train the models.

Alternatives: TPU (Tensor Processing Unit) on Google Cloud or AMD GPUs.

Pros and Cons:

They offer great support for deep learning frameworks such as TensorFlow and PyTorch, while machines on cloud operators may be more expensive.

Why We Choose It: As there are a lot of NVIDIA GPUs and deep learning tools are easily built on them, we select NVIDIA GPUs to train a ViT model.

Cloud Resource (Google Colab): Cloud based resources, like Google Colab, were used for training the model.

Alternatives: You could have used the AWS or Azure.

Pros and Cons:

Google Colab provides free GPU/TPU to training models and should be an affordable choice and accessible option, although it may have restrictions in computing resources for extremely large data.

Justification: Google Colab provides easy access, free-tier GPU support, and seamless built-in file operations with its google drive integration.

5.1.3 Communication Standards

A set of communication standards ensures that the data sent between the web interface, the users, and the server is safe and reliable and complies with the requirements of the industry.

HTTPS: The app operates over HTTPS so that all the data passed between the server and the user's devices is encrypted.

Alternatives: Unsecured communication could have used HTTP.

Pros and Cons:

HTTPS is secure but requires extra steps to generate SSL certificates, adding complexity to the equation.

Justification for Selection: HTTPS was chosen for security reasons in order to maintain best security practice and to safeguard the confidentiality of medical information.

RESTful API's – The app will communicate with the backend through REST API so that gives us an advantage of benefit of fast data exchange handling between front & back-end.

Alternatives: If you had complex queries, you could have used GraphQL.

Pros and Cons:

Good RESTful API is well supported, very easy to implement, and to scale, and other hand, the GraphQL is flexible, but it is harder to implement.

Justification of choice: RESTful APIs were selected for its simplicity and commonality, which would enable easy integration into the web application.

5.2 Impact on Society, Environment and Sustainability

5.2.1 Impact on Life

The ALL Detection System could mean drastic changes in the lives of those affected by ALL (Acute Lymphoblastic Leukemia). Identification of ALL at an early stage can result in appropriate medical intervention, and this is crucial in order to improve patient prognosis and the quality of life of the patients, particularly in children.

DIRECT IMPACT Early detection enables early treatment, which is associated with improved survival and quality of life.

Unintended Consequences: The system improves the quality of decisions in health and helps lower the diagnostic error.

5.2.2 Impact on Society & Environment

Social impact Detection of leukemia using automated approach over using doctor's images The approach makes way to low cost, easy method of early detection of leukemia, particularly in areas where medical facilities are scare. It allows healthcare systems to see more patients in a shorter time.

Greenness The system relies on cloud computing resources, which could be associated with environmental cost, taking into account the cloud carbon footprint of the underlying datacenters. However, these environmental issues are mitigated by the efficient put to use of cloud computing and hardware (GPUs).

5.2.3 Ethical Aspects

The ethical considerations of this project include:

Data Privacy: Medical data is highly sensitive, and all data used in the system is stored and processed following strict data privacy regulations (such as GDPR or HIPAA).

Bias in Models: Ensuring that the dataset used for training the model is diverse and representative of different populations is crucial to avoid bias in predictions.

Accountability: As the system supports medical decision-making, it is essential that healthcare professionals remain accountable for final diagnostic decisions, with the system serving only as a diagnostic aid.

5.2.4 Sustainability Plan

To ensure the system's sustainability:

Continuous Model Improvement: The system will be regularly updated with new data and better models to improve accuracy and adaptability.

Efficient Resource Usage: Cloud resources will be used efficiently, ensuring that the environmental impact of training and inference is minimized.

5.3 Project Management and Financial Analysis

Budget The following is a summary of the budget needed for the development of the ALL Detection System as well as a brief explanation of the revenue model.

Budget Required:

Hardware Training the model on GPUs and deploying the application on the cloud will have costs. For this project, free resources such as Google Colab (free GPU/TPU) were used, although commercial cloud providers (AWS or Google Cloud for scaling) would probably be required.

Software Tools: Development of the proposed framework that includes TensorFlow, Keras and Streamlit does not entail additional costs since they are open-source tools.

Costs Personnel costs: The project was developed solitarily, however, in a scaled up implementation a team of data scientists, software developers and medical staff would likely be required.

Alternative Budget:

A more substantial budget may be needed on more sophisticated hardware (such as TPUs) or large-scale deployment implying greater computational and storage expenses.

Revenue Model:

The system might be hosted on a software-as-a-service (SaaS) basis for health-care providers to subscribe to access and be billed for its use pursuant to the amount of use or on a subscription basis.

License: The model might be licensed to hospitals, clinics and research institutions in need of automation diagnostic assistant.

5.4 Complex Engineering Problem

5.4.1 Complex Problem Solving

This project tackles several complex problem-solving aspects:

EP1: Department of Knowledge
The project demands specialized knowledge in deep learning, particularly in the use of Vision Transformers (ViT) for medical image classification. Additional expertise in medical image analysis and leukemia detection is required. Integrating these diverse fields into a single system poses a significant challenge and requires a multi-disciplinary approach.

EP2: Range of Conflicting Requirements
There are multiple conflicting requirements to consider:

Accuracy vs. Computational Cost: While higher-performing models such as ViT provide better accuracy, they are computationally intensive, which can be an issue in real-time settings.

Real-Time Performance vs. Model Complexity: The challenge of achieving real-time prediction while maintaining high accuracy with a complex model like ViT is significant. Ensuring this balance is key for practical deployment in clinical settings.

EP3: Depth of Analysis
A thorough analysis of the model's performance metrics (e.g., accuracy, precision, recall, F1-score) was necessary to evaluate its robustness. The trade-offs between model complexity and real-time performance had to be continuously assessed, with deep dives into data preprocessing, augmentation, and model optimization.

EP4: Familiarity of Issues
Working with real-world medical images presented several issues, including noisy data, imperfect annotations, and the need for model generalization across diverse patient demographics. These challenges required continuous model refinement and testing to ensure the system performs well in various environments.

EP5: Extent of Applicable Codes
The development followed various software standards (e.g., TensorFlow for deep learning model training) and ethical guidelines (e.g., data privacy and security regulations). Understanding data protection laws such as HIPAA was critical, as the system deals with medical data.

EP6: Extent of Stakeholder Involvement
Healthcare professionals, including doctors and medical researchers, were integral stakeholders, providing feedback on the model's performance and ensuring that the predictions are clinically relevant and actionable.

EP7: Interdependence
The system components, from data collection to model prediction and application deployment, are highly interdependent. If any one component, such as image preprocessing or model inference, fails, the overall system's performance is affected. This interdependency required careful integration and testing.

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

K3: Engineering Fundamentals
A solid understanding of deep learning principles, especially in the context of image classification, is essential for developing the ViT model. Knowledge of convolutional layers, transformer models, and image processing is necessary.

K4: Specialist Knowledge
Domain-specific expertise in medical image classification and Acute Lymphoblastic Leukemia (ALL) is essential. Understanding the characteristics of blood cell images and how to identify ALL through these images is a crucial part of the project.

K5: Engineering Design
The system's architecture needs to integrate various components, including model training, inference, and web application development. System design skills were crucial for creating a seamless, user-friendly platform.

K6: Engineering Practice
This includes practical knowledge of deploying machine learning models into production, ensuring they are optimized for real-time performance, and continuously evaluating the model for accuracy.

K8: Research Literature
Reviewing research on ViT models, medical image analysis, and AI in healthcare helped ensure the methodologies used in this project were cutting-edge and aligned with current practices.
.

Table 5.2: Mapping with knowledge Profile.

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

5.4.2 Engineering Activities

The system involves a range of engineering activities, from data preprocessing to model deployment and performance evaluation. Each step involves rigorous testing and evaluation to ensure the system's reliability in real-world clinical applications.

5.5 Summary

This chapter presented engineering design challenges against the background of development of the ALL Detection System. We discussed the pertinent software, hardware and communication standards and compared different possibilities with their up- and downsides. Socio-environmental and sustainability aspects of such system were also studied, focusing on its social impact and ii) the reduced amount of energy, when compared to traditionalMFA systems.

EA1: Range of Resources → K4: Specialist Knowledge
Effective use of GPU resources for model training and optimization was essential for handling the computational complexity of the ViT model.

EA2: Level of Interaction → K5: Engineering Design
The interaction between model components, the user interface, and the stakeholders (healthcare professionals) was a key design consideration, ensuring that the system meets both technical and practical needs.

EA3: Innovation → K6: Engineering Practice
The novel application of ViT for medical image analysis in ALL detection is an example of engineering innovation. The design and development of the web application also reflect innovative practices in real-time AI deployment.

EA4: Consequences for Society and Environment → K8: Research Literature
The system's societal impact includes improving diagnostic efficiency and patient outcomes. Research literature guides the design to ensure that the system is effective and ethical.

EA5: Familiarity → K3: Engineering Fundamentals
The familiarity with deep learning principles and the challenges of medical image analysis is essential for designing an effective and reliable detection system.

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
✓	✓	✓	✓	✓

Chapter 6

Conclusion

6.1 Summary

This chapter presents an overall summary of the project, gives the study its contributions and briefly tells of the Acute Lymphoblastic Leukemia (ALL) Detection System, the techniques and the results of the system. BloodMed uses Vision Transformer (ViT) models for blood cell image detection and classification to accurately diagnose ALL. We examined several solutions for the image classification task and settled on ViT as it excellently captures long-range dependencies of the medical image data. By pre-processing the data, training the model and building a web application, we were able to build an accurate, real-time diagnostic tool for healthcare workers to aid in diagnosing leukemia.

The project encountered ample engineering difficulties including the tradeoff between accuracy, interpretability, computation time as well as being applicable in real clinical practice. The system was designed to comply with relevant software, hardware and communication standards and standards for both technical and ethical aspects. The project has been particularly influenced by feedback from domain experts (e.g., academic advisers) to ensure that it is situated within realistic healthcare contexts.

Model evaluation results showed that the system had high precision and could be used as a useful medical device for healthcare providers. A web application made available to end-users via a user-friendly interface to upload images and perform predictions in real time enabling better decision-making in Acute Lymphoblastic Leukemia diagnosis

6.2 Limitation

Although the ALL Detection System performed well, we must bear in mind several limitations:

Data Availability: The performance of the model is sensitive to the quality and size of the training set. The dataset might not contain all possible variations of ALL stages or blood cell types.

Answer: To increase the diversity and size of the dataset will make the model more robust and less likely to include bias.

Model Complexities and Training: Vision Transformers (ViT) are computationally expensive architectures although effective. The ViT model is computationally expensive both in terms of GPU requirement as well as training time and may not be viable in low-resource settings.

Solution: Investigating methods for model optimization or utilizing more efficient models may alleviate computational loads.

Inference Speed: Though our model achieves excellent accuracy, the time taken is not optimal during inference, especially in real time clinical scenarios requiring faster decision making.

Solution: Apply model compression or distillation methods to optimize model for efficient inference.

Interpretability: Despite its superior accuracy, on occasion deep learning models, such as ViT, can behave as 'black boxes', creating barriers for deep understanding why the predictions are made the way they are made.

Solution: If these methods are applied using some techniques like GradCAM or SHAP, the model would be more interpretable and reliable.

Scale up: As the model is deployed in hospitals, scaling may prove challenging with more data or batch requests, particularly in a busy clinical environment.

Solution: More scalable and cloud-based solutions could handle large datasets and higher loads better.

6.3 Future Work

The ALL Detection System is a significant advancement towards the application of artificial intelligence in medical image analysis, and several pathways for further Improvements and extensions are listed:

Dataset Augmentation and Diverse Datasets:

To increase the generalization of the model more wide variety examples of blood cells for different categories, from different age groups and distributed in various stages of disease should be included in the dataset. This might minimize the bias and further improve that the model can accurately identify ALL for diverse populations.

Model Optimization:

Deep leanring architectures such as lightweight CNNs or hybrid transformer-CNN approaches may be explored for further reducing the computation overhead with the cost of similar or even better performance. Model compression, distillation and quantization schemes may be considered to improve model speed and efficacy as long as its precision is not violated.

In-the-Wild Deployment in Clinical Contexts:

A natural next step would be a real-time implementation of the system in hospitals and clinics. Making sure that the system can seamlessly integrate with hospital databases and EHRs will make diagnosis easier and benefit clinical workflows.

Incorporating XAI with AI:

Also, including XAI concepts like GradCAM or SHAP could be healthier, as it provides the explanation for making decisions. This way, the model would be more explainable, and it would gain the trust of healthcare workers when they assess the model output.

Integration with other diagnostic applications:

This presented ALL Detection System could be enhanced by integration of additional diagnostics tools like histopathology slide scanning system or biomarker analysis system. What Seems Interesting Combining these data collection modes would help to paint a broader picture of the patient's health.

Longitudinal Surveillance & Prediction:

The system could be extended not only to diagnose the presence of ALL but also to monitor the patient's disease state over time. It also had the potential to forecast the future trajectory of the disease and offer treatment recommendations based on the patients own' data.

To allow global access, the service adopts a cloud-based model:

The ALL Detection System would be more available to low-resource healthcare workers if a cloud-based system was in place. It would also make AI-powered diagnosis available to more people — particularly in underserved areas where specialized doctors can be scarce.

Regulatory CLEARANCES AND ClInICAL Test CLInICAL: & TestinG:

In the prospective and planned implementation of our system in clinical practice, obtaining medical device certifications like FDA or CE mark will be crucial. In the future, we can make clinical trials and further validation studies to confirm the reliability and efficacy of the system for clinical use.

References

[1] Li Yuan, Yunpeng Chen, Tao Wang, Weihao Yu, Yujun Shi, Zi-Hang Jiang, Francis EH Tay, Jiashi Feng, Shuicheng Yan Proceedings of the IEEE/CVF international conference on computer vision, 558-567, 2021

[2] Daquan Zhou, Bingyi Kang, Xiaojie Jin, Linjie Yang, Xiaochen Lian, Zihang Jiang, Qibin Hou, Jiashi Feng arXiv preprint arXiv:2103.11886, 2021

[3] I. Yanhong Fei, Yingjie Liu, Xian Wei, Mingsong Chen arXiv preprint arXiv:2201.12133, 2022

[4] Xiaoyu Yue, Shuyang Sun, Zhanghui Kuang, Meng Wei, Philip HS Torr, Wayne Zhang, Dahua Lin Proceedings of the IEEE/CVF International Conference on Computer Vision, 387-396, 2021

[5] Mengzhao Chen, Mingbao Lin, Ke Li, Yunhang Shen, Yongjian Wu, Fei Chao, Rongrong Ji Proceedings of the AAAI Conference on Artificial Intelligence 37 (6), 7042-7052, 2023.

[6] Adjei-Mensah, I., Zhang, X., Baffour, A. A., Agyemang, I. O., Yussif, S. B., Agbley, B. L. Y., & Sey, C. (2021, December). Investigating vision transformer models for low-resolution medical image recognition. In 2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP) (pp. 179-183). IEEE.

[7] [1] Sampathila, N., Chadaga, K., Goswami, N., Chadaga, R. P., Pandya, M., Prabhu, S.,... & Upadya, S. P. (2022, September). Personalized deep learning model evaluating sensitively A classification system for acute lymphoblastic leukemia based on blood smear analysis. In Healthcare (Vol. 10, No. 10, p. 1812). MDPI.

[8] Katar, O., & Yildirim, O. (2023). Explainable Vision Transformer Model for White Blood Cells Classification and Localization.

[9] Jiang, Z., Dong, Z., Wang, L., & Jiang, W. (2021). Diagnosis of acute lymphoblastic leukemia with ViT-CNN ensemble. Journal of Computational Intelligence and Neuroscience, 2021. Sampathila, N., Chadaga, K., Goswami, N., Chadaga, R. P., Pandya, M., Prabhu, S., ... & Upadya, S. P. (2022, September). Customized deep learning classifier for detection of acute lymphoblastic leukemia using blood smear images. In Healthcare (Vol. 10, No. 10, p. 1812). MDPI.

[10] Parayil, S., & Aravinth, J. (2022, June). Transfer learning-based feature fusion of white blood cell image classification. In 2022 7th International Conference on Communication and Electronics Systems (ICCES) (pp. 1468-1474). IEEE.

[11] Shahzad, A., Raza, M., Shah, J. H., Sharif, M., & Nayak, R. S. (2022). Categorizing white blood cells by utilizing deep features of proposed 4B-AdditionNet-based CNN network with ant colony optimization. Complex & Intelligent Systems, 1-17.

[12] Maurya, R., Pandey, N. N., & Dutta, M. K. (2023). VisionCervix: Papanicolaou cervical smears classification using novel CNN-Vision ensemble approach. Biomedical Signal Processing and Control, 79, 104156.

[13] Saeed, A., Rehman, F., Ashfaq, A., Sharif, H., Zeeshan, R., & Zahid, A. (2022, November). Acute Lymphoblastic Leukemia Prediction from Microscopic Blood Smear Images Using Advanced Machine Learning. In 2022 14th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS) (pp. 1-6). IEEE.

[14] Amutha, S. (2023). VGGNet-Cnn based classification of white blood cell leukemia with efficient salp swarm optimization algorithm. Journal of Intelligent & Fuzzy Systems, (Preprint), 1-17.

[15] Ikechukwu, A. V., & Murali, S. (2022). i-Net: a deep CNN model for white blood cancer segmentation and classification. Int. J. Adv. Technol. Eng. Explor, 9(95).

[16] Maaliw, R. R., Alon, A. S., Lagman, A. C., Garcia, M. B., Susa, J. A. B., Reyes, R. C., ... & Hernandez, A. A. (2022, October). A Multistage transfer learning approach for acute lymphoblastic leukemia classification. In 2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) (pp. 0488-0495). IEEE.

[17] Ahmed, I. A., Senan, E. M., Shatnawi, H. S. A., Alkhraisha, Z. M., & Al-Azzam, M. M. A. (2023). Hybrid techniques for the diagnosis of acute lymphoblastic leukemia based on fusion of CNN features. Diagnostics, 13(6), 1026.

[18] Ahad, M. T., Li, Y., Song, B., & Bhuiyan, T. (2023). Comparison of CNN-based deep learning architectures for rice diseases classification. Artificial Intelligence in Agriculture, 9, 22-35.

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

[20] Agrawal, Rohit, et al. (2019). "Detection of white blood cell cancer using image processing." 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN). IEEE, 2019.

[21] WANG,P., ZHANG, X., ZHAO, Y., LI, Y., XU, K., & ZHAO, S. (2023). Analysis of blood cell image recognition methods based on improved CNN and Vision Transformer. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences.

[22] Cho, P., Dash, S., Tsaris, A., & Yoon, H. J. (2022, April). Image transformers for classifying acute lymphoblastic leukemia. In Medical Imaging 2022: Computer-Aided Diagnosis (Vol. 12033, pp. 633-639). SPIE.

[23] Ali, M. L., Rauf, Z., Khan, A., Sohail, A., Ullah, R., & Gwak, J. (2023). CB-HVT Net: A channel-boosted hybrid vision transformer network for lymphocyte detection in histopathological images. IEEE Access.

[24] Kumar, D., Jain, N., Khurana, A., Mittal, S., Satapathy, S. C., Senkerik, R., & Hemanth, J. D. (2020). Automatic detection of white blood cancer from bone marrow microscopic images using convolutional neural networks. IEEE Access, 8, 142521-142531.

[25] Ghaderzadeh, M., Aria, M., Hosseini, A., Asadi, F., Bashash, D., & Abolghasemi, H. (2022). A fast and efficient CNN model for B-ALL diagnosis and its subtypes classification using peripheral blood smear images. International Journal of Intelligent Systems, 37(8), 5113-5133.

[26] Kumar, R., Joshi, S., & Dwivedi, A. (2021). CNN-SSPSO: a hybrid and optimized CNN approach for peripheral blood cell image recognition and classification. International Journal of Pattern Recognition and Artificial Intelligence, 35(05), 2157004.

[27] Ding, Y., Yang, Y., & Cui, Y. (2019). Deep learning for classifying of white blood cancer. In ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging: Select Proceedings (pp. 33-41). Singapore: Springer Singapore.

[28] Hosseini, A., Eshraghi, M. A., Taami, T., Sadeghsalehi, H., Hoseinzadeh, Z., Ghaderzadeh, M., & Rafiee, M. (2023). A mobile application based on efficient lightweight CNN model for classification of B-ALL cancer from non-cancerous cells: a design and implementation study. Informatics in Medicine Unlocked, 39, 101244.

212-15-4204

ORIGINALITY REPORT

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

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	6%
2	arxiv.org Internet Source	1%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
4	www.mdpi.com Internet Source	1%
5	ijsrem.com Internet Source	<1%
6	"Big Data and Artificial Intelligence", Springer Science and Business Media LLC, 2025 Publication	<1%
7	R. N. V. Jagan Mohan, B. H. V. S. Rama Krishnam Raju, V. Chandra Sekhar, T. V. K. P. Prasad. "Algorithms in Advanced Artificial Intelligence - Proceedings of International Conference on Algorithms in Advanced Artificial Intelligence (ICAAAI-2024)", CRC Press, 2025 Publication	<1%
8	biarjournal.com Internet Source	<1%
9	Arvind Dagur, Karan Singh, Pawan Singh Mehra, Dhirendra Kumar Shukla. "Intelligent	<1%