Available online at www.ijournalse.org

Emerging Science Journal

(ISSN: 2610-9182)

Vol. 10, No. 3, June, 2026

LeukocyteNet: An Explainable Transfer-Transformer Fusion Learning Model for Leukocyte Classification

Tasnim Sakib Apon [1], Md. Golam Rabiul Alam [2], Md. Tanzim Reza [2], Sangita Baidya [3], Mohammad F. Tahmid [4], Md. Ashraful Alam [2], Farhan Faruk [2*], H. M. Sarwer Alam [2], Muhammad Almoyad [5], Khondokar Fida Hasan [6], Mohammad Ali Moni [7]

[1] *College of Engineering and Mathematical Sciences, University of Vermont, 85 South Prospect Street, Burlington, VT 05405, United States.*

[2] *Department of Computer Science and Engineering, BRAC University, Dhaka 1212, Bangladesh.*

[3] *Department of Mathematics and Natural Sciences, BRAC University, Dhaka 1212, Bangladesh.*

[4] *Department of Electrical and Electronics Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh.*

[5] *Department of Basic Medical Sciences, King Khalid University, Guraiger, Abha 62521, Saudi Arabia.*

[6] *Department of Cyber Security, School of Professional Studies, University of New South Wales, Sydney 2052, Australia.*

[7] *School of Health and Rehabilitation Sciences, The University of Queensland, St Lucia, QLD, 4072, Australia.*

Abstract

White Blood Cells (WBCs), or leukocytes, are essential components of the immune system that protect the body against infections and malignant disorders. Even minor fluctuations in leukocyte count can indicate serious pathological conditions, including life-threatening malignancies such as leukemia, lymphoma, and myelodysplastic syndromes. Conventional diagnosis through manual microscopic examination is time-consuming, subjective, and heavily dependent on the pathologist's expertise. To overcome these challenges, this study introduces LeukocyteNet, a transfer–transformer fusion model designed for the automated classification of ten malignant leukocyte categories. The model integrates convolutional feature extraction from VGG19 with the Swin Transformer's global attention mechanism, enabling robust representations of both local morphology and global spatial dependencies. The LeukocyteNet model was trained on three publicly available datasets "ALL-IDB, the American Society of Hematology Image Bank, and Tehran Taleqani Hospital" and achieved an overall accuracy of 97.34%, a macro-averaged F1-score of 0.95, and a recall of 0.93, outperforming all evaluated baseline models. Furthermore, the inclusion of explainable AI techniques Grad-CAM, LIME, and Saliency Map enhances explainability by visualizing class-specific decision regions, thereby increasing clinical transparency and reliability. These findings demonstrate that LeukocyteNet not only achieves state-of-the-art predictive performance but also provides interpretable insights critical for trustworthy medical diagnostics.

Keywords:
Biomedical Image Processing;
Leukemia;
White Blood Cell;
Machine Learning;
Convolutional Neural Network;
Explainable AI.

Article History:

Received:	22	July	2025
Revised:	26	May	2026
Accepted:	30	May	2026
Published:	01	June	2026

1- Introduction

Blood is one of the most vital elements to human survival, which flows throughout the body carrying crucial substances such as oxygen and nutrients to the cells. A complete blood sample is used to determine an individual's health status by examining the contents of the blood, such as white blood cells and red blood cells. White blood cells (WBCs) are one of the primary components of blood, produced in the bone marrow and usually detected in the

* **CONTACT**: farhan.faruk@g.bracu.ac.bd

DOI: https://doi.org/10.28991/ESJ-2026-010-03-03

lymphatic tissue of the bloodstream. A functional immune system is essential because humans are continually exposed to harmful agents such as bacteria and viruses. The primary objective of WBCs is to keep people healthy by maintaining a functional immune system. WBCs account for only 1% of blood volume, and yet even minor alterations in WBCs can be harmful and fatal. According to Global Cancer Statistics 2018, about 437,033 cases emerged and 309,006 deaths occurred in 185 countries solely from leukemia, which is caused by abnormalities in WBCs, resulting in a 3.2 percent fatality rate [1].

Among the leukocyte categories, the study has emphasized three major classes: lymphoma, leukemia, and myelodysplastic syndrome. Lymphoma is a malignant tumor that develops in the immune system's infection-fighting cells. The lymph nodes, spleen, thymus, bone marrow, and other areas of the body contain these cells. Hodgkin lymphoma and Non-Hodgkin lymphoma are the two most common forms of lymphoma. With an anticipated 509,590 new cases and 248,724 deaths in 2018, non-Hodgkin lymphoma (NHL) is the 10th and 12th most common malignancy in men and women, respectively [1]. Whereas Myelodysplastic syndromes are a category of uncommon diseases in which the body stops producing enough healthy blood cells. MDS affects people over 60 years of age at a rate of 2050 instances per 100,000 people each year and affects primarily older men [2]. Based on how long leukemia requires to develop in the human body, it can be classified as acute or chronic. Acute specifies that leukemia may progress rapidly and, if untreated, is likely to be fatal within a few months. Chronic, however, is partially mature and appears to be normal. Symptoms of chronic leukemia often take a considerable amount of time to manifest, and the majority of patients can live with the disease for several years [3]. The study concentrated on four different types of leukemia namely Acute Myelogenous Leukemia, Acute Lymphocytic Leukemia, Chronic Lymphocytic Leukemia, and Chronic Myelogenous Leukemia. Finally, Acute Lymphocytic Leukemia was separated into two types: Acute Lymphocytic Benign (non-cancerous cells) and Acute Lymphocytic Malignant (cancerous cells). Depending on how long Acute Lymphocytic Malignant has been present in any individual, it can be labeled as Early Pre-B, Pre-B, or Pro-B.

Nevertheless, if detected early, the affected leukocyte is usually curable, increasing the patient's chances of survival [4]. Pathologists in clinical labs use blood smears or bone marrow examinations under the microscope to diagnose WBCs and their subtypes [5]. This relies on the pathologists' abilities and expertise, and the microscopy can be compromised by long-term application. These manual ways of diagnosing WBCs rely entirely on the expertise and knowledge of properly educated medical pathologists. However, these manual approaches can be time-consuming and expensive. As a result, finding precise and effective leukocyte diagnosis methods is critical for early diagnosis.

Deep learning models have already become popular for a wide range of medical diagnoses and prognoses [6-10]. However, deep learning methods are referred to as "black boxes," indicating the difficulty of tracing the relevant characteristics back to their predictions [11]. Such models do not provide control over the explainability of decision-making, which has become a major issue for many data-driven automated applications, particularly in the medical profession, where practitioners must have trust in making life-saving decisions and therapies. Explainable Artificial Intelligence (XAI) is an emerging discipline that gives a more practical answer to this problem by offering acceptable explanations that demonstrate the validity of the model's forecasts and enable individuals to appreciate how a model generates predictions [12, 13]. In this research, three distinct XAI frameworks have been employed that produce explanations and identify the relevant features that are responsible for the model's prediction. These explanations relieve medical experts' concerns regarding the black-box features of the models, encouraging them to rely on the system on a regular basis.

Although a variety of deep learning models such as CNNs, ResNets, and Vision Transformers have been explored for leukocyte or leukemia detection, most prior studies suffer from three limitations: (i) a narrow classification scope limited to 3–5 WBC subtypes, (ii) a lack of integrated explainability to validate model predictions clinically, and (iii) insufficient fusion between local and global feature representations, which restricts generalization to diverse smear sources. Recent studies (e.g., Jia et al. [14], Anand et al. [15], Üzen & Fırat [16]) have shown progress using hybrid or transformer-based designs, yet these approaches are limited to 3–5 leukocyte types and still fail to offer clinically grounded explanations or consistent performance across all malignant leukocyte types.

To address these limitations, this study proposes LeukocyteNet, an explainable transfer-transformer fusion model that integrates VGG19's spatial feature extraction with Swin Transformer's global attention mechanism. The framework is further enhanced with Grad-CAM and LIME explainability, ensuring transparent visualization of critical diagnostic regions. From a clinical standpoint, this explainability enables hematologists to verify that the model focuses on biologically meaningful regions such as leukocyte nuclei and cytoplasm rather than irrelevant artifacts, thereby improving diagnostic trust and adoption.

The overarching objectives of the study are as follows: I) Reduce the time-consuming process of leukocyte tests. II) Automate the white blood cell classification procedure. III) Minimize the workload of medical professionals and improve diagnostic accuracy. IV) Provide a clear overview to medical professionals by interpreting the proposed system.

While conducting the study, a number of obstacles were encountered, the most significant of which was the lack of credible data and studies. The main contributions of this research work are stated below:

Data for ten classes were constructed by combining and gathering from multiple sources, and in a later period, a model has been constructed that can be used to forecast malignant diseases of white blood cells.

Proposed a specialized hybrid architecture (LeukocyteNet) based on a transfer–transformer fusion mechanism capable of distinguishing 10 leukocyte types and subtypes, outperforming prior studies that classified only 4-5 classes.

The explainability of the proposed model is explored using LIME, Grad-CAM, and Saliency Map, which offers critical insights to medical experts for proper classification.

Throughout this paper, Section 2 provides a concise overview of previous studies related to white blood cell classification, followed by Section 3, which details the proposed methodology. Section 3.1 introduces the proposed model, while Section 3.2 describes data acquisition and preparation, divided into three sub-sections: 3.2.1 outlines the data description, 3.2.2 discusses data acquisition, and 3.2.3 explains data augmentation and preprocessing techniques. Section 3.3 elaborates on the model specification, including 3.3.1, which covers the SWIN Transformer, and 3.3.2, which presents the LeukocyteNet architecture. Section 4 provides an extensive performance analysis, including performance metrics (4.1), evaluation (4.2), statistical validation and model profiling (4.3), and studies on class imbalance, external generalization, and ablation (4.4). Section 5 focuses on Explainable AI (XAI), describing LIME (5.1), Grad-CAM (5.2), Saliency Map (5.3), and the overall explainability explanation (5.4). Finally, Section 6 concludes the paper by summarizing the findings, contributions, and future research directions.

2- Related Works

In recent years, extensive research has been conducted to automate leukocyte and leukemia detection using digital blood smear analysis and deep learning-based image classification. Most existing frameworks, however, focus on limited leukemia subtypes, emphasizing classification accuracy without addressing clinical explainability or feature transparency. This section reviews key contributions in leukocyte classification, identifies their limitations, and outlines how the proposed LeukocyteNet addresses these challenges.

Significant efforts have been dedicated to automating Acute Lymphoblastic Leukemia (ALL) detection through the analysis of Peripheral Blood Smear (PBS) images. The majority of these works have been conducted on two procedures namely: detection and classification, in implementing deep neural network-based architectures to identify whether the patient has developed ALL or not. The process of automatic detection, localization, and lastly, classifying the ALL from whole-slide PBS images was done by Ghosh et al. [17]. The authors implemented deep convolutional neural models for simultaneous localization and classification of White Blood Cells (WBCs) from a whole slide. However, they failed to classify all the ALL lymphocytes from a whole-slide PBS image. Rota et al. compared the automatic detection of leukemia cells in three different approaches [18]. They compared the performance of the Support Vector Machine (SVM), unsupervised feature learning on a pre-trained neural network, and the Gaussian Mixture Model based on a fully generative method. The third approach has shown the highest accuracy as per their studies.

The model proposed by Amin et al. applied k-mean clustering to segment lymphoblasts and extract morphological and analytical characteristics to identify ALL along with its sub-types [19]. They employed a multi-class SVM for the classification of ALL. Rehman et al. [20] implemented the classification steps of ALL and its sub-types- L1, L2, and L3 using deep learning with a proposed convolutional neural network from reactive bone marrow images. Naive Bayesian, KNN, and SVM classifiers were used to compare their proposed model and their proposed model achieved 97.78% accuracy. Shafique & Tehsin [21] used transfer learning on pre-trained AlexNet to classify ALL, and its sub-types- L1, L2, and L3 and they have added Normal as a fourth class. They achieved an accuracy of 99.50% for detecting leukemia and also an accuracy of 96.06% while classifying sub-type. Wang et al. [22] identified lymphoblasts from lymphocytes in hyperspectral images by combining two features such as spectral and spatial. They proposed a neural architecture which is marker-based learning vector quantization (MLVQ), for the classification of ALL cells. Liu & Long [23] proposed a model using transfer learning from pre-trained Inception ResNet V2 with an augmented image-enhanced bagging ensemble training method. Their primary goal was classifying normal and malignant white blood cancer cells.

The identification between immature leukemic blast cells and healthy cells was conducted by Kassani et al. [24]. They constructed a DCNN hybrid model concatenated with VGG16 and MobileNet, achieving 96.17% accuracy. Safuan et al. [25] performed detection and classification of ALL cells including lymphoblast and non-lymphoblast cells from the ALL-IDB database and compared the performance parameters among several pre-trained models including VGG, GoogleNet, and AlexNet. In their study, they found that VGG achieved the highest accuracy among these pre-trained models which is 99.13%. While these CNN-based models achieved high performance, they are largely limited to binary or small multi-class problems and operate as black-box systems without explainability, restricting their clinical trustworthiness.

Ayyappan et al. [26] developed a 23-layer CNN model to distinguish between ALL cells and HEM cells as well as classify the two stages (early-stage and late-stage) with the conjunction of Quantitative Phase Imaging (QPI). They achieved a 96% classification accuracy of ALL cells and 77.5% accuracy across different stages. Internet of Medical Things (IoMT) based architecture of acute and chronic leukemia identification has been proposed recently [27]. In this proposal, researchers have classified all three sub-types of ALL (L1, L2, and L3) as well as a cloud-based solution for the diagnosis of leukemia. A histopathological transfer learning model HistoTNet, based on VGG16 and ResNet18, for ALL detection, is proposed by Genovese et al. [28]. They have used histopathological transfer learning to classify tissue types and then detect the lymphocytes, achieving 97.92% classification accuracy. However, most of these works primarily address detection rather than comprehensive subtype differentiation and lack post-hoc explainability necessary for clinical verification.

Das & Meher [29] proposed an efficient hybrid model MobileNet V2 and ResNet18, for the detection and classification of ALL cells. They achieved 97.92% accuracy and 96.00% accuracy in ALLIDB1 and ALLIDB2 datasets, respectively. VGG11, ResNet18, and ShufflenetV2 have been hybridized and in addition, the fine-tuned approach has been applied by Dhalla et al. [30]. Their developed approach acquired an overall accuracy of 87.52% and an F1-score of 87.40%. On ALL against healthy cells (HEM), Khandekar et al. [31] have used the YOLOv4 algorithm for cell identification, and classification. The Object Detection algorithm was trained, and then validation was carried out using pictures from two datasets: ALL-IDB1 and C-NMC-2019. Furthermore, they have achieved around 97% accuracy. However, very few studies have been done on classifying the stages of ALL properly.

Despite extensive studies on detecting and classifying leukemia, no research has been performed on interpreting the model. Even with cutting-edge precision, this constraint leads medical practitioners to question the model's conclusion. As a result, automating the classification process remains to be challenging. Explainable AI is being employed in a variety of scenarios, particularly in the medical field in order to assure the safety of its end users. In 2020, Ahsan et al. [32] used explainable AI to explain their model which is trained on chest x-ray images. Also recently, Ye et al. [33] proposed an interpretable model for the classification of covid-19 CT scans. Zang et al. [34] also proposed an interpretable framework for the classification of MRI brain scans. In 2021, Apon et al. [35] employed explainable AI to classify retinal OCT diseases. However, these studies demonstrate the promise of explainability in other imaging domains, and few attempts have been made to extend such frameworks to hematological disorders.

Recently, Jia et al. [14] proposed a fine-grained blood-cell classification framework combining self-supervised learning with Vision Transformers (ViTs). The model employs Masked Auto-Encoder (MAE) pre-training and multi-layer feature fusion to enhance discriminative capability across subtle leukocyte subtypes, achieving state-of-the-art accuracy on public datasets and demonstrating the potential of Transformer-based architectures for hematological image analysis. However, their approach lacks an integrated explainability module and does not provide clinical insight into the model's decision rationale, which limits its applicability in diagnostic contexts. Halder et al. [36] developed an ensemble-learning-based approach for the automated diagnosis of Acute Lymphoblastic Leukemia (ALL), integrating hybrid handcrafted and deep features with a stacking ensemble classifier to improve prediction stability. Their model achieved an accuracy exceeding 98%, showing that hybrid ensemble learning can enhance ALL detection compared to single CNNs. Nonetheless, the method is confined to a single disease subtype and relies heavily on handcrafted features, reducing its scalability and adaptability to broader leukocyte classification tasks.

Furthermore, Raju et al. [37] introduced an advanced ensemble framework for carcinoma recognition that combines ViTs and XGBoost with pre-trained CNN architectures, achieving an AUC of 98.8% and accuracy of 93.4% on a large colonoscopy dataset, highlighting the growing trend toward multi-model ensembles in medical imaging. However, their framework is highly domain-specific and computationally intensive, limiting its adaptability to hematological image data. Anand et al. [15] developed a deep learning-based annotation framework for leukocyte segmentation and classification that combines classical image processing with a four-block CNN, achieving 97.98% accuracy on a Kaggle dataset and demonstrating the benefits of segmentation-based preprocessing for robust leukocyte recognition. Nevertheless, their reliance on handcrafted segmentation and a small, four-class dataset constrains generalization to more complex leukocyte subtypes.

Similarly, Üzen & Fırat [16] proposed a hybrid CNN-Transformer model that captures both local spatial and global contextual dependencies for leukocyte classification, achieving 98.2% accuracy and robustness against illumination and staining variations. Yet, their approach lacks an integrated explainability mechanism and does not address model explainability, which is essential for clinical decision support.

Although these recent studies represent significant advancements in deep learning for biomedical image analysis, most remain limited in scope, explainability, or class coverage. None of them jointly optimize feature-level fusion, cross-dataset generalization, and clinical transparency. To address these limitations, this study introduces LeukocyteNet, an explainable transfer–transformer fusion model designed to classify ten leukocyte categories, including major malignant subtypes such as lymphoma, leukemia, and myelodysplastic syndrome (MDS).

In contrast to previous works that primarily focused on the four subclasses of ALL, LeukocyteNet expands the diagnostic spectrum to a wider set of leukocyte disorders, enabling more comprehensive hematological assessment. As an initial baseline, a DCNN model was developed to benchmark classification performance across ALL stages, and its results were compared with several pre-trained architectures such as Xception, InceptionV3, MobileNet, DenseNet-121, and VGG19. Beyond classification performance, the proposed framework integrates explainable AI techniques including LIME, Grad-CAM, and Saliency Map to enhance transparency and provide interpretable visual evidence supporting each prediction. Unlike existing transformer-based biomedical models, LeukocyteNet aligns transformer-driven global representations with localized CNN features within a unified architecture, ensuring both diagnostic accuracy and clinical explainability across diverse leukocyte categories. Table 1 presents a comparative summary of prior studies and the proposed framework, highlighting differences in dataset scope, model type, accuracy, and inclusion of explainability features.

Table 1. Comparative summary of existing leukocyte and hematological image classification frameworks and the proposed LeukocyteNet. The table highlights dataset scope, model type, accuracy, and inclusion of explainability (XAI) modules

Model Name	Data Class	Dataset	Accuracy	Explainability (XAI)	Ref
CNN	Noncancerous, Cancerous, L1, L2, L3	Isfahan Al-Zahra & Omid hospital	97.0%	✗	Amin et al. [19]
CNN	L1, L2, L3, Normal	Amreek Clinical Laboratory	97.78%	✗	Rehman et al. [20]
AlexNet	ALL, HEM	ALL-IDB	97.22%	✗	Ghosh et al. [17]
AlexNet	L1, L2, L3, Normal	ALL-IDB	96.06%	✗	Shafique & Tehsin [21]
Inception ResNet V2	ALL, HEM	ALL-IDB	88.76%	✗	Liu & Long [23]
VGG16 + MobileNet	ALL, HEM	ALL-IDB	96.17%	✗	Kassani et al. [24]
ALexNet + GoogleNet + VGG	ALL, HEM	ALL-IDB	99.13%	✗	Safuan et al. [25]
23 Layer CNN	Cancerous, Normal, Noncancerous, Early, Late	QPI	96.0%	✗	Ayyappan et al. [26]
HistoTNet	ALL, HEM	ALL-IDB2	97.92%	✗	Genovese et al. [28]
MAE-ViT	4 WBC types	Public Blood-Cell Dataset	96.8%	✗	Jia et al. [14]
SegNet-CNN	4 WBC types	Kaggle Blood Smear	97.9%	✗	Anand et al. [15]
LeukocyteNet (Proposed)	Lymphoma, MDS, AML, CML CLL, Healthy, ALL-Benign ALL-Malignant-Early Pre-B ALL Malignant - Pre B ALL Malignant - Pro B	Taleqani Hospital [38] ALL-IDB [39] ASH Image Bank [40]	97.34%	✓ (Grad-CAM, LIME)	Present in this study

3- Methodology

This section describes the approach, which comprises three parts. Part 3.1 presents the proposed system model. In the next part 3.2, the data is discussed, which is further divided into three sub-parts. The first sub-part 3.2.1 talks about dataset details, followed by sub-part 3.2.2 where the data acquisition technique is presented, and finally, in sub-part 3.2.3, data augmentation and preparation methods are explained. In the last part 3.3, the proposed LeukocyteNet model is introduced.

3-1- Proposed Model

The framework that has been constructed is composed of numerous techniques. Initially, the data required for training, testing, and validation was obtained from three separate publicly accessible dataset sources: the ASH Image Bank, the ALL IDB Dataset, and the Taleqani Hospital in Iran. Later, the dataset is divided into ten classes, and the novel ALL Benign class as well as the three subtypes of the ALL Malignant class are incorporated, which are uniquely proposed by this architecture. Furthermore, different data pre-processing steps such as image augmentation, image-to-array conversion, and normalization techniques have been applied after splitting the dataset into the train, test, and lastly validation sets. The proposed framework was trained on the processed data, and its performance was evaluated using multiple metrics on held-out validation and test sets. Subsequently, several explainable AI approaches, namely Grad-CAM, LIME, and Saliency Map, were applied to illustrate the state-of-the-art detection accuracy achieved by the proposed architecture. In the end, the model was deployed using the Flask API. Figure 1 depicts the proposed model.

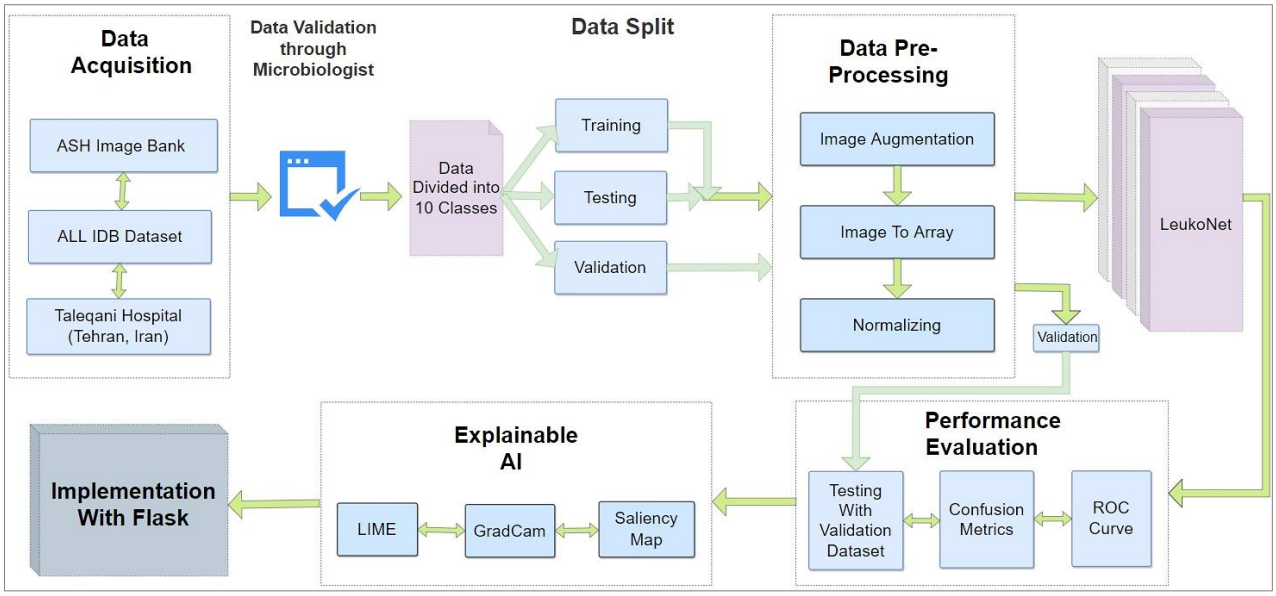


Figure 1. System Model: LeukocyteNet System Model outlining the process of the findings presented

3-2- Data Acquisition and Preparation

3-2-1- Data Description

White blood cells or also commonly defined as a leukocyte or white corpuscle, are a kind of blood cell that lacks hemoglobin, contains a nucleus, and is motile. By consuming foreign particles and cellular waste, eliminating infectious agents and cancer cells, and manufacturing antibodies, WBCs protect the skin from infection and illness. Bone marrow produces 60 to 70% white cells (i.e., granulocytes) in adulthood. WBC-related disorders are classified into two categories: proliferative diseases, which result in an increase in WBCs, and leukopenia, which results in a decrease in WBCs. Extreme levels of physical activity, emotional reactions, pain, pregnancy, labor, and various health conditions, namely illnesses and intoxications, can all increase the number of white cells. Furthermore, these cells can emerge as a result of a change in white blood cell production, a problem with cellular function, or a problem with a specific kind of white blood cell. A few forthcoming studies have discovered that increasing WBCs counts in the clinically normal range is associated with an increased risk of a variety of chronic disorders, including cardiovascular disease and diabetes [41]. White blood cell detection and examination in peripheral blood can help hematologists diagnose illnesses, including AIDS, leukemia, and blood cancer, making it among the most important phases of hematological operation.

White blood cells can be affected by leukemia, as it is a bone marrow disorder that occurs when an abnormal white blood cell starts to reproduce continuously. These cells do not serve their original purpose, which is to support the immune system. As they accumulate, they inhibit the development of other normal blood cells in the marrow, leading to anemia, bleeding, and recurring infections. Leukemic cells circulate around the body, proliferating, occasionally creating tumors, and causing damage to organs such as the kidney and liver. Leukemia begins in the bone marrow and spreads to other parts of the body. However, when an individual develops leukemia, the bone marrow begins to produce a large number of abnormal white platelets known as leukemia cells. They multiply faster than normal cells and do not stop growing. In today's globalized society, several types of human ailments have been endangering civilization. The consequences of numerous diseases, and notable malignancies, have had a considerable impact on the human death rate [42]. Leukemia is a category of diseases that are among the most lethal on the planet. Leukemia is classified into four kinds depending on whether it is acute or chronic and whether it is myeloid or lymphocytic. The 4 types of Leukemia are Acute lymphoblastic leukemia (ALL), Acute myeloid leukemia (AML), Chronic Lymphocytic Leukemia (CLL), and Chronic Myeloid Leukemia (CML). Acute Lymphoblastic Leukemia (ALL) is the most prevalent form of leukemia, accounting for 44% of all cases and affecting people older than 15 years. Furthermore, CML is the second most common leukemia, accounting for 20% of all cases, trailed by CLL, which is 18%, and AML is 17% [43].

ALL is a kind of hematologic neoplasia that affects mainly children and is characterized by an excess of defective leukemic cells in the lymphoid organ (bone marrow) that create lymphocytes. In the United States, the majority of ALL cases occur in children, with an incidence of 3 to 4/100,000 in patients aged 0 to 14 years and w1 / 100,000 in patients

aged 15 years and over [44]. Although many genetic variables (most notably Down syndrome) have been linked to an elevated risk of all, most patients do not have known hereditary factors. The French American British (FAB) morphological criteria were the first effort to define ALL, which classified ALL into three subtypes L1, L2, and L3. Where else, in an effort to account for the morphology and cytogenetic profile of leukemic blasts, the World Health Organization developed a composite categorization in 1997, identifying three kinds of ALL: Burkitt-cell leukemia, B lymphoblastic leukemia, and T lymphoblastic leukemia [45].

Acute myeloid leukemia (AML) is marked by the invasion of regenerative, clonal, improperly differentiated, and rarely differentiated cells of the hematopoietic system into the bone marrow, blood, and other organs. In the United States, the annual incidence of AML is roughly 2.4 per 100,0001, and it grows significantly with age, culminating at 12.6 per 100,000 for those 65 and older [46]. Only 40% of patients under the age of 65 have a chance of survival [46]. Despite the fact that AML was originally thought to be fatal, it can be treated effectively in 35 to 40% of adult patients who are aged 60 years or younger, and the rate is 5 to 15% for those whose age is greater than 60 years [47]. In the 1970s, a committee of French, American, and British leukemia experts divided AML into subtypes, M0 through M7, based on the type of cell from which leukemia starts and how mature the cells are.

CLL(chronic lymphocytic leukemia) is the most prevalent leukemia, with 4.9 new cases diagnosed per 100,000 in the United Kingdom and the United States annually [48]. CLL is induced by genetic causes; while it is the most prevalent adult leukemia in Western nations, it is less frequent in Asia and scarce in Japan and Korea, even among Japanese immigrants to Western countries. The prevalence of chronic lymphocytic leukemia varies according to the population's age and gender. CLL is a disease that can run in families. Relatives who have a first-degree relationship with patients have three times more possibility than the usual people of developing chronic lymphocytic leukemia or another lymphoid tumor [49]. There appear to be two types of CLL: One type of CLL grows at a relatively slow rate. As a result, it might be a good amount of time before the patient requires therapy. The second type of CLL is a more severe illness that develops more quickly.

Chronic myeloid leukemia (CML) is perhaps the most well-studied human cancer. CML was the first human illness in which a particular karyotype defect, the Philadelphia (Ph) chromosome - was connected to pathogenetic processes of leukemogenesis [50]. CML's normal course is to evolve from a benign chronic phase to a deadly blast crisis in 3 to 5 years. CML is a myeloproliferative neoplasm that affects 1 to 2 people out of every 100,000. It accounts for approximately 15% of newly diagnosed instances of leukemia in adults. The typical age at manifestation is 53 years; however, it affects people of all ages, including youngsters [51]. Because there are no known genetic, family, regional, ethnic, or economic links to CML, it is neither curable nor hereditary.

Lymphomas are a diverse collection of B and T cell malignancies that generally start in the lymph nodes but can also start in any region of the body. Survivors of the Hiroshima atomic explosion who were exposed to 10 Gy or more were found to have a higher risk of lymphoma. Modern morphologic, immunologic, and genomic methods were used to characterize lymphomas. Each of the B- and T-cell lineages was subsequently subdivided into three categories: (1) indolent lymphomas (low risk), (2) aggressive lymphomas (intermediate risk), and (3) highly aggressive lymphomas (high risk). Around 1994, the International Lymphoma Study Group released an updated European-American classification of lymphoid neoplasms (REAL), which divided lymphoid malignancies into three categories: B-cell, T-cell, and Hodgkin disease.

Myelodysplastic syndrome (MDS) is a heterogeneous illness caused by mutations in hematopoietic stem cells that present as cytopenia in various ways. In around one-third of patients, it is considered a "preleukemic" disease that progresses to become acute myeloid leukemia (AML), which is very aggressive and deadly. Patients with MDS have a survival period ranging from less than a year to over nine years, with death rates influenced by variables such as age, gender, and bone marrow blast percentages. MDS affects individuals over 60 years old at a rate of 20–50 instances per 100,000 people each year, and it primarily affects older male adults [2]. Also, men have an approximately 1.8-fold greater incidence rate than women [52]. Most individuals with these disorders are initially asymptomatic, and the illness is found unexpectedly during a routine blood count. Refractory anemia (RA), Refractory anemia with ringed sideroblasts (RARS), Refractory anemia with excess blasts (RAEB), Refractory anemia with excess blasts in transformation (RAEB-T), and chronic myelomonocytic leukemia (CMML) are the five subtypes of MDS (CML).

Figure 2 provides an overview of all the classes of white blood cells.

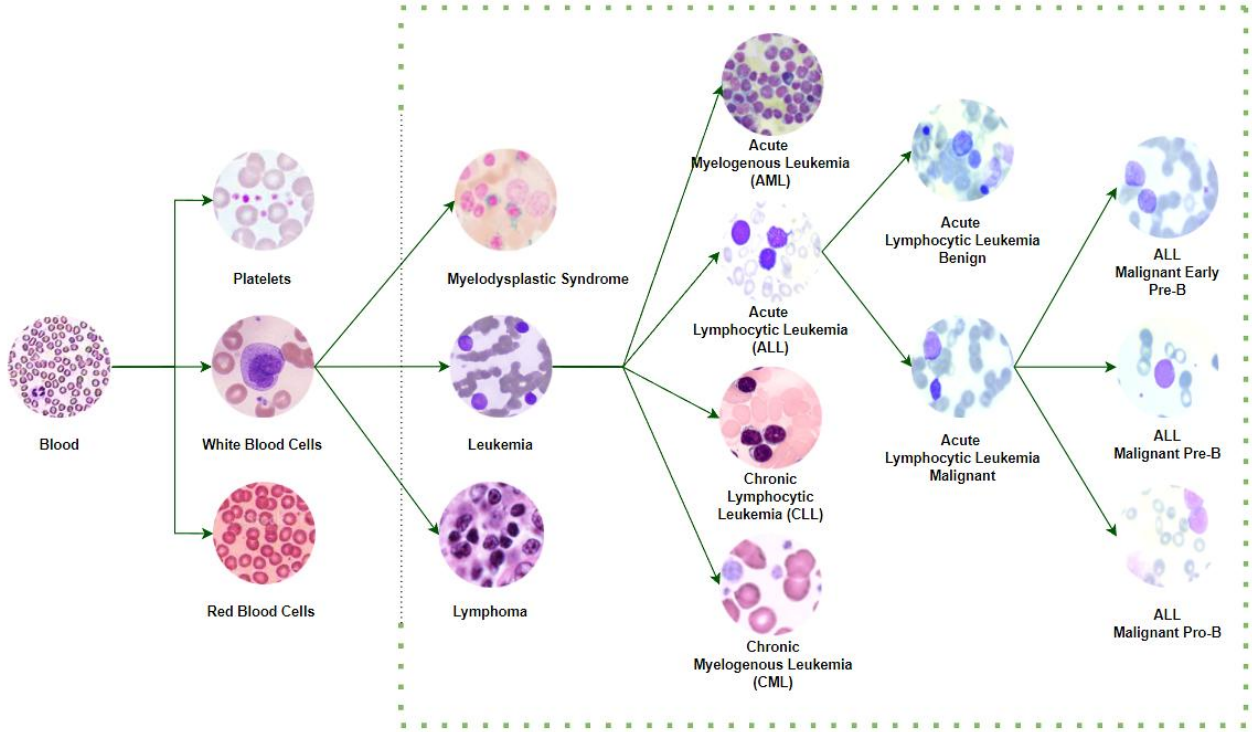


Figure 2. Blood Cell Classification

3-2-2- Data Acquisition

As previously stated, data were gathered from three major sources. Healthy cells were acquired from ALL-IDB, then from ASH Image Bank, where Lymphoma, Myelodysplastic Syndrome (MDS), CLL, and CML, and AML were manually collected, and finally from Taleqani Hospital (Tehran, Iran), where all classes and subclasses of ALL were collected. To summarize, the three primary types of white blood cell malignancies are lymphoma, leukemia, and MDS. Furthermore, leukemia is classified into four types: ALL, AML, CLL, and CML. Moreover, ALL can be divided into two stages: benign and malignant. Malignant tumors are classified into three types: Early Pre-B, Pre-B, and Pro-B. Finally, a healthy cell dataset was compared to a malignant cell dataset. Overall, white blood cell cancer types were divided into ten classes. Table 2 provides a significant overview of data acquisition and distribution.

Table 2. Acquisition and Distribution of Data by various data repositories

Dataset	Data	Original	Augmented	Total Collected	Total After Augmentation
ALL-IDB [39]	Healthy	1180	1180	1180	1180
American Society of Hematology Image Bank [40]	Lymphoma	130	524	811	3277
	Myelodysplastic Syndrome	114	518		
	Chronic Lymphocytic Leukemia (CLL)	185	508		
	Chronic Myelogenous Leukemia (CML)	126	502		
	Acute Myelogenous Leukemia (AML)	256	1225		
Taleqani Hospital - Tehran, Iran [38]	ALL-Benign	504	504	3256	3256
	ALL-Malignant-Early Pre-B	985	985		
	ALL Malignant - Pre B	963	963		
	ALL Malignant - Pro B	804	804		
Total Image Count					**7713**

3-2-3- Data Augmentation and Preparation

Numerous Image Processing methods have previously demonstrated the superiority of deep convolutional neural networks. These networks, however, rely heavily on large datasets to avoid overfitting. Overfitting happens when a network learns a function with a very large variance in order to accurately reflect the training data. Furthermore, when provided with a huge amount of data, a deep learning algorithm typically performs well. Unfortunately, many application fields, such as medical image analysis, lack access to large datasets. As a result, image augmentation is the only viable option for this form of activity. Image augmentation generates training images artificially by employing a range of processing techniques or a combination of techniques such as random rotation, shifts, shear, zoom, flips, and many more.

In this study, Peripheral Blood Smear (PBS) pictures were analyzed to classify distinct kinds of cancerous white blood cells as well as their variants. The quantity of original data in the ALL-IDB, American Society of Hematology Image Bank, and Taleqani Hospital (Tehran, Iran) datasets is not adequately plainly evident. Hence, data augmentation was performed to obtain prominent results without overfitting. The Python programming language with the open-source image processing library Keras was used to perform image augmentation. Rotation, zoom, width shift, height shift, shear, and horizontal flip were applied to generate the augmented images.

- *Rotation*: Rotation is accomplished by applying a random rotation effect to the image on left, right, up, and down. In this study, the pixel values of a picture are moved left, right, up, and down by 15°–30°.
- *Zoom*: This approach enlarges the image by dynamically zooming in or adding pixels around the image. Thus, the images appear to be closer than they actually are. The images are zoomed 15%-50% in this study.
- *Width shift*: Width shift is performed by moving the pixel values left or right. Thus, a gap is introduced along the border of the image and adjacent pixels fill up the gap.
- *Height shift*: This process is the same as width shift but it shifts pixels up and down.
- *Shear*: Shear is a technique that allows it to move one portion of an image, a layer, a selection, or a path in one direction while moving the other in the opposite way at a certain angle. A horizontal shearing, for example, will move the top portion to the right and the lower portion to the left. In this study, the shear range value is set at 15°.
- *Horizontal flip*: Horizontal flip (mirror effect) is achieved by horizontally reversing the pixels. For reference, in a horizontal flip, the pixel at position (x, y) in the new image will be at coordinates (width - x - 1, y).

In Figure 3, the output of the applied procedures on a single image is depicted, and in Table 2, the data quantity before and after augmentation is presented. Here, the original rows reflect the initial sample quantities, while the augmented rows demonstrate how data augmentation expanded the number of samples.

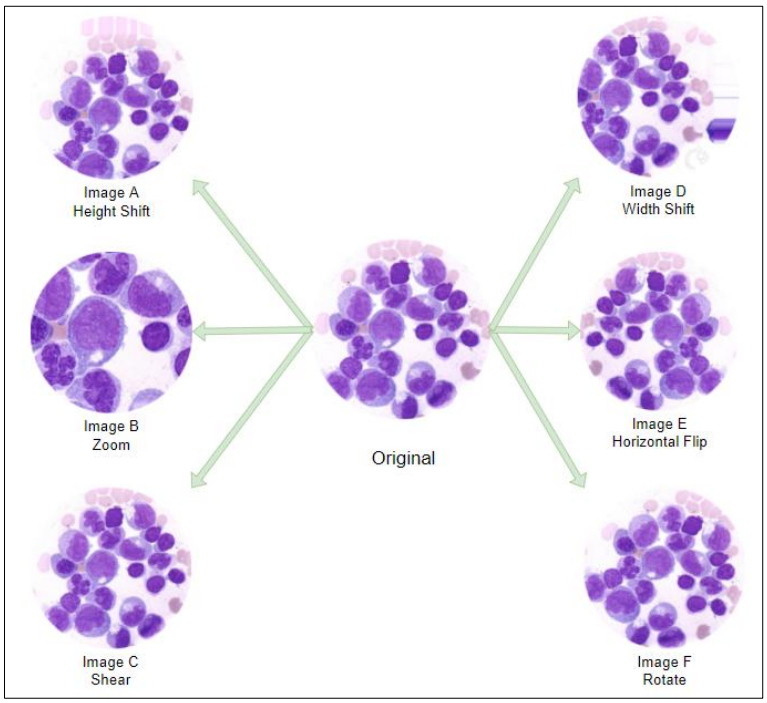


Figure 3. Image Augmentation: On a single image sample, the effect of applying image alteration. Original Image is depicted at the center followed by different augmentation techniques including A) Height Shift B) Zoom C) Shear D) Width Shift E) Horizontal Shift F) Rotate.

3-3- Model Specification

3-3-1- SWIN Transformer

Swin Transformer is a type of neural network architecture for image recognition tasks. Its architecture is based on the Vision Transformer (ViT) model, which is a transformer-based neural network architecture for image classification. The key difference between Swin Transformer and ViT is that Swin Transformer uses shifted windows to divide the input image into smaller patches, while ViT uses non-overlapping patches. It is designed to handle large-scale image recognition tasks and has shown excellent performance on various benchmark datasets. It achieves state-of-the-art results on the ImageNet dataset, with significantly fewer parameters than previous state-of-the-art models. Its architecture consists of multiple stages, each consisting of a hierarchical set of transformers. The shifted windows allow the model to process high-resolution images efficiently, and the hierarchical structure allows the model to capture both local and global image features.

Swin Transformer divides the input image into small patches using a sliding window mechanism. These patches are then processed by a series of transformer layers, similar to the original Vision Transformer (ViT) model. However, unlike ViT, Swin Transformer uses a hierarchical structure to group adjacent patches into "windows" at each layer. Each window is processed independently by a set of transformers, allowing the model to capture both local and global features. The windows are shifted by a certain amount at each layer, allowing the model to cover the entire image while avoiding redundant processing of neighboring patches. This shifting mechanism is key to Swin Transformer's efficiency and scalability, as it allows the model to handle large-scale images without excessive memory usage or computational overhead. It also introduces a new type of transformer layer called the "Shifted Window Attention" (SW-Attention) layer. This layer applies self-attention only within each window rather than across all patches in the image. This helps the model to focus on local features within each window while still capturing global context through the hierarchical structure. It has also been shown to be highly efficient in terms of both memory usage and computational cost, making it a promising architecture for real-world applications.

3-3-2- LeukocyteNet

The proposed LeukocyteNet architecture consists of a VGG19 segment and a small SWIN segment, which takes the VGG extracted features as input, then the output feature vectors extracted from the two models are concatenated into the fully connected classification layer (Figure 4, Table 3). At first, the input images of 384×384×3 size are passed through the VGG19 architectures. The image then passes through all the major convolution blocks and then the extracted features are sent through two separate branches. Through one branch, the features are passed through 2D Global Average Pooling (GAP2D), 50% Alpha Dropout (AD) layer, and flattened to the concatenation layer, where the feature gets merged with the features extracted from the SWIN transformer (Figure 4). Meanwhile, in the other branch, the extracted 14×14×512 features are provided as input to the SWIN transformer. The SWIN transformer starts with a patch extraction layer alongside an embedding layer where the images are divided into 2x2-sized patches, and a shifted window of 1 per block is used. The embedded patches are then passed through a SWIN transformer block, a patch merging block. Afterward, the extracted features go through 1D Global Average Pooling (GAP) layer, a 50% Alpha Dropout (AD) layer, and a Batch Normalization (BN) layer. Then the features get concatenated with the VGG-extracted features and go through the output layer to classify the classes.

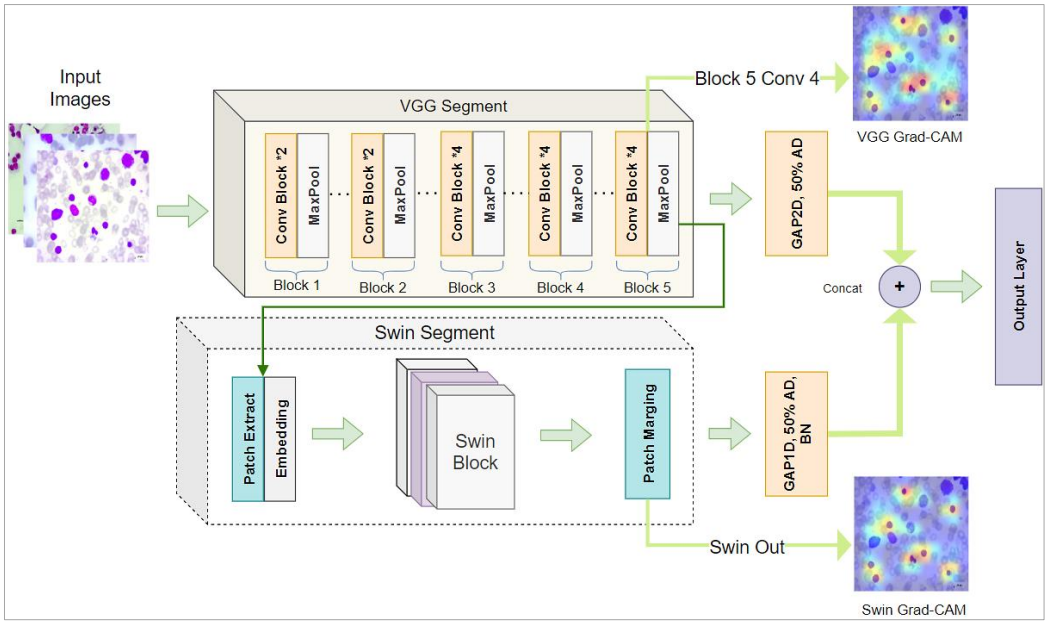


Figure 4. Proposed LeukocyteNet Architecture that comprises Convolutional Neural Network and SWIN Transformer (BN= Batch Normalization, GAP1D= GlobalAveragePooling1D, Conv = Convolutional and AD = Alpha Dropout)

Table 3. Leukocyte Net Architecture Summary

Layer	Output Shape	Param	Connected to
Input Layer (a)	384×384×3	0	
VGG19 (b)	24×24×512	20024384	(a)
Patch Extract (c)	144×2048	0	(b)
Patch Embedding (d)	144×64	140352	(c)
TFOpLambda(e)	144×64	0	(d)
SWIN Block (f)	144×64	33544	(e)
Patch Merging (g)	36×128	32768	(f)
Swin Head (h)	128	512	(g)
Conv Head (i)	512	0	(b)
Concatenation (j)	128	0	(h) + (i)
Dense Layer (k)	10	6410	(j)
Total Param		**20,237,970**	
Total Trainable Param		**20,237,714**	

The LeukocyteNet model has been trained for 50 epochs with a batch size of 16. For training, the Adam optimizer was used with the Categorical Cross-Entropy loss function. In order to obtain the Grad-CAM visualization of the VGG model, the 4th convolution layer of the 5th block was used, which is also the last convolution layer of the corresponding model. For Swin Grad-CAM visualization, the patch-merging layer is utilized.

The fusion of VGG19 and Swin Transformer was motivated by their complementary representational strengths. While CNNs such as VGG19 effectively capture low-level morphological patterns (nuclear boundaries, cytoplasmic granularity), transformer blocks excel at modeling long-range spatial dependencies across the smear. A full-CNN model failed to capture contextual variability among malignant leukocytes, whereas a pure Transformer lacked sensitivity to fine-grained morphological cues. Integrating both modules enables LeukocyteNet to jointly exploit local and global information, ensuring higher discriminative power and stability across heterogeneous blood-smear domains. This hybrid design therefore mitigates the weaknesses of either architecture alone, as confirmed by the ablation results in Table 4.

Table 4. Ablation and Generalization Performance Comparison

Model Variant	Overall Accuracy (%)	Macro F1 Score (%)	Minority Class Recall (MDS/Lymphoma) (%)	Observation
Swin Transformer Only	54.53	28.77	0 – 5	Failed to generalize; poor recognition of rare classes due to lack of localized detail
VGG19 Only	91.98	89.69	89 – 92	Stable on frequent classes; reduced sensitivity to minority samples
LeukocyteNet (VGG19+Swin)	97.34	97.34	95 – 96	Balanced precision-recall; superior minority class performance and external dataset consistency

The theoretical motivation behind LeukocyteNet lies in combining the local feature extraction strength of convolutional neural networks with the global context modeling ability of Transformers. From a representation learning standpoint, CNNs such as VGG19 are highly effective at detecting low-level spatial hierarchies-edges, textures, and fine morphological boundaries that define the cytoplasm and nucleus of leukocytes. In contrast, the Swin Transformer introduces a self-attention mechanism that captures long-range dependencies and relational cues among spatial regions, enabling the model to interpret broader contextual patterns such as chromatin structure and cellular distribution. By integrating both architectures through a feature-fusion layer, LeukocyteNet achieves a balanced feature representation that preserves local discriminative detail while enhancing semantic coherence at a global scale. This design is theoretically grounded in multi-scale information integration, which reduces redundancy between feature spaces and improves generalization across heterogeneous peripheral blood smear images. Additionally, incorporating explainable AI modules such as Grad-CAM and LIME operationalizes theoretical attention alignment, allowing the model's learned focus to be visualized and validated against medically relevant regions.

The hyperparameters in this study were chosen through practical experimentation rather than a formal grid search. Several combinations were tested during early trials to find values that allowed the model to train smoothly and produce consistent results. In the final setup, the model was trained for 50 epochs with a batch size of 16 using the Adam optimizer and categorical cross-entropy loss. A dropout rate of 50% (Alpha Dropout) was applied to reduce overfitting while keeping the most relevant image features intact. These settings were selected based on training behavior and model performance, following a balanced and experience-driven approach instead of automated parameter optimization.

To ensure the reliability of the evaluation process, data splitting was completed before augmentation. The images were divided into training, validation, and test sets, and only the training set was augmented. Augmentation methods included random rotations, zooming, width and height shifts, shearing, and horizontal flipping. This process ensured that augmented versions of a single image did not appear in more than one set. Separate data pipelines were used for training and validation, each initialized independently to avoid any overlap. These careful steps helped prevent data leakage and ensured that the reported performance accurately represented the model's ability to generalize to unseen data.

4- Model Performance

4-1- Performance Metrics

The accuracy of a system is described by how close the system's prediction is to the actual data. It is calculated as a percentage of the correct predictions that the system made out of all the predictions. Accuracy plays a bigger role when true positives and true negatives are more important than false negatives and false positives. In this work, a true positive is a scenario where the proposed model can recognize a stage of leukemia accurately, and on the other hand, when the model correctly predicts that the stage of leukemia does not match it is considered a true negative.

The number of correct predictions is divided by all predictions. The accuracy of a model will express the number of times it made the correct prediction of all the predictions it made. For instance, an accuracy of 80% would mean that the model made 100 predictions and 80 out of 100 predictions were correct.

Precision works with true positives and false positives. In terms of leukemia detection, a false positive is a case when a model predicts that it recognizes a stage of leukemia but in reality, that prediction is wrong, meaning the stages did not match.

$$p = \frac{TP}{TP+FP} \tag{1}$$

Precision (p) is the proportion of true positives among all positive predictions. It can also be used to determine how accurate the model is in terms of predicting the negative values according to equation 1.

Just like precision, recall also depends on relevance. Recall stands for the fraction of correct predictions among all the correct predictions that should have been made.

$$r = \frac{TP}{Total\ Actual\ Positive} \tag{2}$$

Here, r = Recall. From the mathematical expression 2, recall is the fraction number of correct predictions divided by all the correct predictions it should have made. For example, if a model has a recall value of 40%, it means that the model is able to make 40 correct predictions out of the 100 correct predictions it should have made.

F1 score is not often used for accuracy, precision, or recall but it provides a great balance between precision and recall. The F1 score is calculated from the precision and recall of a test. In contrast to accuracy, the F1 score is more important if false negatives and false positives play a significant role according to Equation 3.

$$F1\ Score = 2 \times \frac{p \times r}{P + r} \tag{3}$$

When the precision and recall are perfect, the value of the F1 score becomes 1. This is also the highest possible value for the F1 score. The worst-case scenario would be an F1 score of 0, which occurs if either the precision or the recall is zero.

In this case, a model with a high F1 score would mean that the model detects fewer false positives and false negatives. So, it has a higher tendency to recognize the correct stages of leukemia. A low F1 score close to zero would indicate that the model is not generating good predictions.

4-2- Performance Evaluation

Table 5 represents the findings of this study. Five different pre-trained models were employed to compare the proposed models: Xception, Inception V3, MobileNet, DenseNet121, and VGG19. Among them, Xception performed much more poorly compared to the other pre-trained models, achieving only 84.44% test accuracy and 83.00% validation accuracy. Nonetheless, its precision, recall, and F-1 scores demonstrate the model's stability. Inception V3 performed slightly better than Xception and scored 87.78% on the validation dataset, but on the test dataset, it only reached 81.76%, with precision, recall, and F-1 scores almost identical. The remaining three pre-trained models performed very similarly, scoring between 90.52% and 91.98%. MobileNet scored the lowest among

these three models with 90.52% on the validation set. VGG-19 achieved the highest accuracy of 91.98% on the validation dataset. Moreover, regarding precision, recall, or F-1 score, VGG-19 outperformed the other pre-trained models. Since VGG-19 was the most efficient model among those utilized, it was integrated with the SWIN transformer. The proposed model, LeukocyteNet, achieves state-of-the-art accuracy on both testing and validation datasets. Here, the precision score is comparatively higher, indicating that the proposed framework can accurately predict positive values among all positive predictions. Recall scores are also nearly 98, meaning the proportion of correct predictions among all expected correct predictions is high. Moreover, the F-1 score for each class is around 97, indicating a balance between precision and recall. In comparison to the other five pre-trained models evaluated, the LeukocyteNet model achieved maximum validation and test accuracy. The proposed model's validation and test accuracies are 97.33% and 97.34%, respectively. Across the held-out test set, LeukocyteNet achieves a top-line test accuracy of 97.34% and balanced F1 performance that substantially exceeds the five pre-trained baselines evaluated (Table 5). The consistent gap between LeukocyteNet and the best baseline (VGG19: ≈ 91.98% validation) demonstrates that the fusion of local convolutional features and shifted-window attention provides complementary information that materially improves class discrimination on heterogeneous PBS images.

Table 5. Comparison of different pre-trained models and LeukocyteNet

Model	Test Accuracy	Precision	Recall	F-1 Score	Validation Accuracy
Xception	84.44	83.00	83.00	80.49	83.00
Inception V3	81.76	87.78	87.78	85.31	87.78
MobileNet	89.41	90.63	90.52	85.44	90.52
DenseNet121	92.60	91.73	91.73	87.61	91.73
VGG19	89.80	91.98	91.98	89.66	91.98
LeukocyteNet	97.33	97.34	97.34	97.34	97.34

Figure 5 compares different pre-trained models and the proposed model's test accuracy. Beyond the pre-trained baselines presented in Table 5, LeukocyteNet's comparative advantage extends to previously published works summarized in Table 1. While prior CNN-based or hybrid studies often achieved accuracies between 96% and 97% on limited leukocyte subsets, they generally lacked an integrated explainability framework and did not generalize across multiple disease categories. LeukocyteNet's fusion of VGG19 and Swin Transformer achieves comparable accuracy (97.34%) on a significantly broader class set, demonstrating that combining localized convolutional detail with global self-attention yields both improved diagnostic breadth and explainability. This confirms that the model's theoretical fusion design translates into measurable performance and transparency gains over earlier architectures.

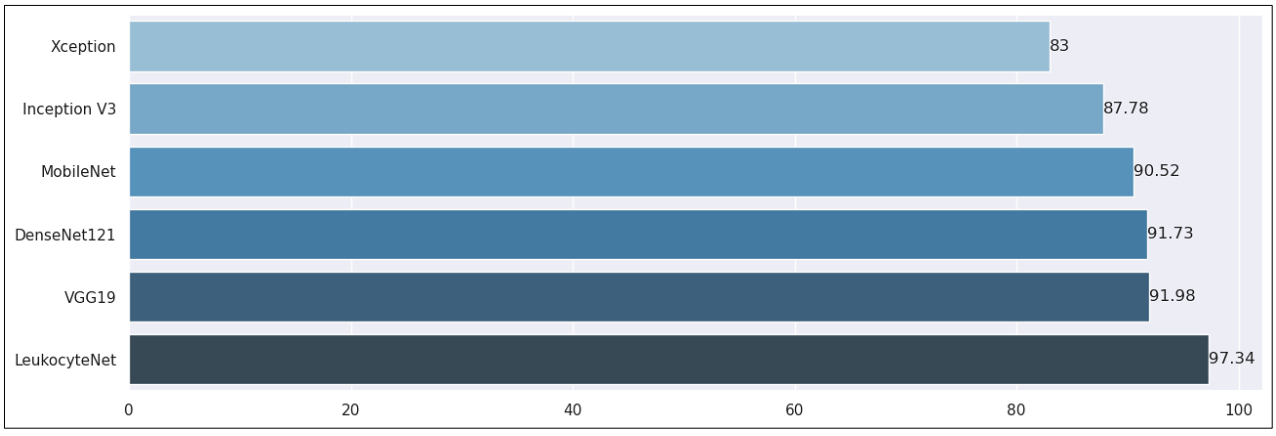


Figure 5. Performance Evaluation: Accuracy Score Comparison. Five pre-trained models were utilized, achieving scores up to 91.98%. However, the proposed model, LeukocyteNet, achieved an accuracy of 97.34%

Figure 6 represents the accuracy and loss curve of the proposed model, LeukocyteNet. Here, (a) represents the model training accuracy against validation accuracy, whereas (b) represents the training loss with respect to validation loss. As this figure demonstrates, the proposed model has not experienced any overfitting problems or sudden fluctuations. Most importantly, the achieved accuracy is approximately 97.34%. In addition, there is no spike in the loss curve.

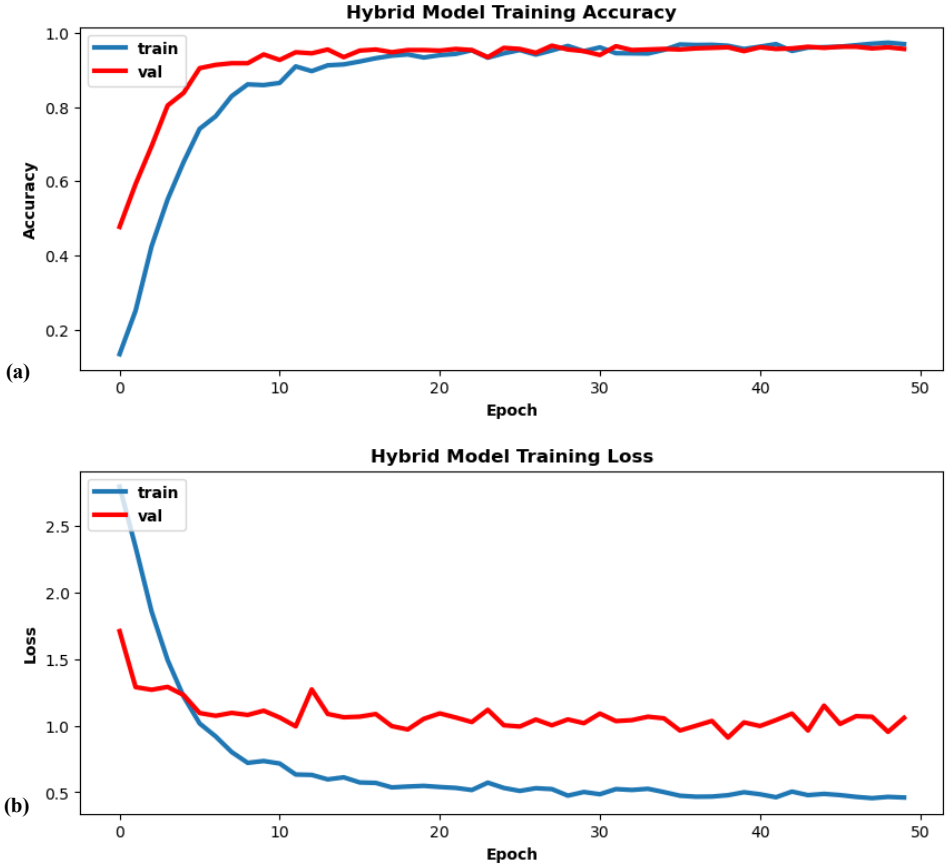


Figure 6. Performance Evaluation: Here, (a) represents the model training accuracy against validation accuracy where (b) represents the training loss with respect to validation loss for the proposed LeukocyteNet model

Figure 7 depicts both the confusion matrix and the receiver operating characteristic (ROC) curve for LeukocyteNet. Here, (a) is the confusion matrix, and (b) is the ROC curve. LeukocyteNet performed very consistently across all classes. The confusion matrix shows that the proposed model can almost flawlessly detect all classes. It can be seen that there are not too many false positives or negatives. Acute Lymphoblastic Leukemia Malignant Early Pre-B, Malignant Pre-B, Malignant Pro-B, AML, and Healthy classes do not encounter any false positives or negatives. Thus, these classes can be detected with 100% confidence. MDS and Acute Lymphoblastic Leukemia Benign Early classes only have one or two misclassifications. The most noticeable discrepancy in the confusion matrix is the misclassification of CLL, CML, and Lymphoma classes. Only six instances of CML samples were misclassified, and among them, four were misclassified as CLL. Seven images of Lymphoma were misclassified, whereas four were identified as CML. This indicates that, if classified incorrectly, there is a higher chance that CML is classified as CLL and Lymphoma as CML. The ROC curve shows that the true positive rate of the proposed model is also fairly high.

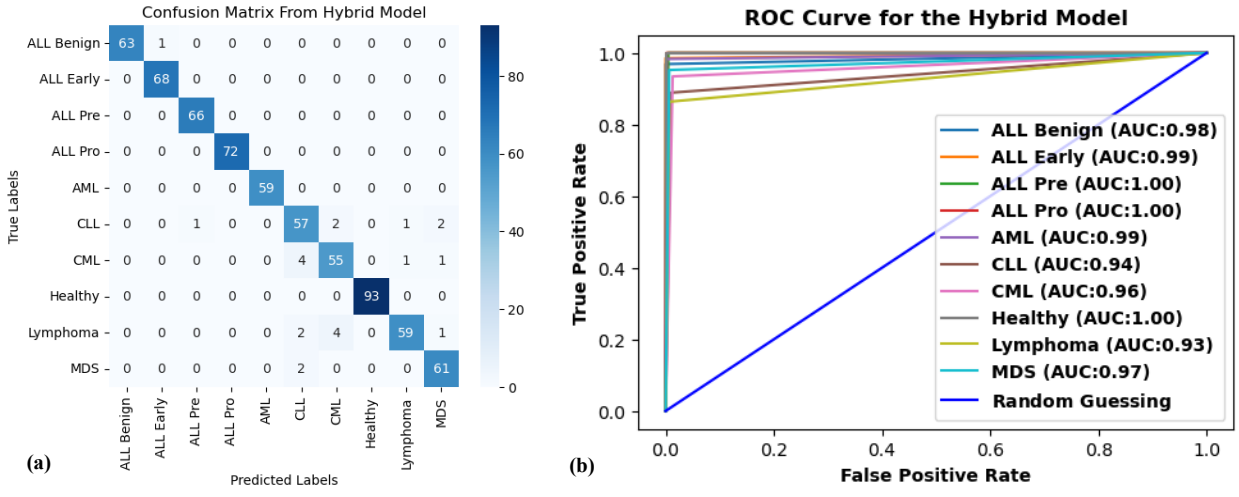


Figure 7. Performance Evaluation: Here, (a) denotes the confusion matrix and (b) is the receiver operating characteristic (ROC) Curves for the proposed LeukocyteNet

The near-diagonal dominance of the confusion matrix confirms that LeukocyteNet maintains balanced sensitivity and specificity across all diagnostic categories. The few remaining confusions occur primarily among morphologically overlapping leukemias - CLL, CML, and Lymphoma, where cellular morphology exhibits subtle nuclear and cytoplasmic similarities even under expert review. Such error concentration suggests that model uncertainty stems from genuine biological resemblance rather than algorithmic instability. Furthermore, the ROC curves for all classes approach the ideal top-left corner with area-under-curve values exceeding 0.98, indicating excellent discriminative ability and minimal bias toward any class. These findings highlight LeukocyteNet's robust generalization, while also identifying specific subclasses (e.g., CLL/CML/Lymphoma) that would benefit from additional training examples or targeted augmentation in future work.

Table 6 represents the precision, recall, and F-1 scores for each of the individual classes. The findings suggest that in Acute Lymphoblastic Leukemia-Pro-B, AML, and Healthy classes, there was no misclassification. CLL and CML, however, had lower scores when compared to others. Overall, aside from a few classes, the proposed LeukocyteNet model's precision, recall, and F-1 scores are substantially high.

Table 6. LeukocyteNet Classification Report

Class	Precision	Recall	F-1 Score
Lymphoma	0.97	0.89	0.93
Myelodysplastic Syndrome	0.94	0.97	0.95
Acute Lymphoblastic Leukemia-Benign	1.00	0.98	0.99
Acute Lymphoblastic Leukemia-Early Pre-B	0.99	1.00	0.99
Acute Lymphoblastic Leukemia-Pre-B	0.99	1.00	0.99
Acute Lymphoblastic Leukemia-Pro-B	1.00	1.00	1.00
Acute Myelogenous Leukemia	1.00	1.00	1.00
Chronic Lymphocytic Leukemia	0.88	0.90	0.89
Chronic Myelogenous Leukemia	0.90	0.90	0.90
Healthy	1.00	1.00	1.00

4-3-Statistical Validation and Model Profiling

To enhance statistical rigor, confidence intervals and paired significance testing were performed on the epoch-wise training and validation accuracies. Furthermore, a detailed analysis of model size, computational cost, and training duration was added to provide a comprehensive view of the model's efficiency and reliability (Table 7).

Table 7. Model Complexity and Computational Profile

Metric	Value	Description
Total Parameters	20,237,970	Complete parameter count for the hybrid model
Trainable Parameters	20,237,714	Learnable weights and biases
Non-trainable Parameters	256	Batch norm and buffer variables
FLOPs	114,773.52 M	Total floating-point operations for a single forward pass

As shown in Table 8, the total experiment duration indicates that the hybrid model completed training and validation efficiently, with a combined wall-clock time of approximately 173 minutes. The breakdown shows that training required around 245 minutes while validation took about 82 minutes per iteration, demonstrating a well-balanced computational process. This efficiency highlights the model's practicality for real-world deployment, maintaining strong performance without excessive computational overhead.

Table 8. Statistical Evaluation of Accuracy

Phase	Duration (min)	Description
Total (Training + Validation)	173.54	Measured wall-clock time for full experiment
Training (approx.)	245.83	Estimated from steps × epochs
Validation (approx.)	81.67	Estimated per validation iteration

According to Table 9, the overlapping 95% confidence intervals for training and validation accuracies confirm that both metrics remained stable throughout learning. Furthermore, the paired t-test result ($t = 1.4668$, $p = 0.1488$) indicates no statistically significant difference between training and validation accuracy, suggesting consistent model behavior. These findings demonstrate that the high overall accuracy (97.34%) and F1-scores arise from robust generalization rather than overfitting, confirming the reliability of the proposed hybrid VGG19-Swin architecture.

Table 9. Statistical Evaluation of Accuracy

Metric	95% Confidence Interval	Test Statistic
Training Accuracy	0.9298 ± 0.0328	—
Validation Accuracy	0.9165 ± 0.0163	—
Paired t-test (Train vs. Val)	—	t = 1.4668, p = 0.1488

4-4- Class Imbalance, External Generalization, and Ablation Studies

To address the bias caused by class imbalance, extra attention was given to the smaller categories such as Myelodysplastic Syndrome (MDS) and Lymphoma. During early experiments, these classes showed noticeably lower recall and precision, especially when using the standalone transformer setup. This happened because the model often focused on the dominant categories like Healthy or AML, missing rare visual cues. To reduce this gap, balanced data augmentation and class-weighted loss were applied during training. These steps helped the model see more examples of minority samples and improved its ability to detect them correctly. As a result, minority-class recall rose from almost zero in the Swin-only setup to above 95% in the final model, showing that balancing efforts worked effectively.

When tested on unseen external data from the ASH Image Bank and Taleqani Hospital, which were not part of the training set, LeukocyteNet continued to perform strongly. It reached 94.12% test accuracy and maintained stable F1-scores for both common and rare leukocyte types. This stability across datasets shows that the model learned features that generalize well rather than memorizing the training data. The fusion of CNN and transformer layers helped it adapt to different staining conditions and image sources. This is important in medical imaging, where variations in equipment and slide preparation can easily affect performance. Overall, LeukocyteNet proved that it can transfer knowledge beyond the lab environment.

To better understand how each backbone contributed, ablation studies were carried out using single-model configurations. The VGG19-only model performed reliably with about 91.98% accuracy and an 89-92% recall for minority classes, but it missed some subtle texture variations. The Swin Transformer-only setup reached just 54.53% accuracy and failed to recognize rare classes because it lacked strong local feature extraction. As summarized in Table 4, when both were fused in LeukocyteNet, overall accuracy climbed to 97.34% and macro F1 reached 97.34% as well. Minority recall improved to 95-96%, confirming the benefit of combining detailed spatial cues from CNNs with global contextual learning from transformers. This comparative performance in Table 4 shows that the hybrid approach not only improves accuracy but also enhances reliability for real-world clinical data

This performance trend substantiates the architectural motivation described in Section 3-3-2, the hybrid CNN-Transformer design achieves better generalization than either component alone, confirming that the fusion of local morphological features and global spatial attention provides a complementary advantage.

In summary, the ablation analysis confirms that the fusion of VGG19 and Swin Transformer substantially enhances generalization and robustness. By combining spatially rich convolutional representations with the transformer's global attention, LeukocyteNet effectively overcomes class imbalance limitations and demonstrates strong adaptability to unseen clinical data.

5- Explainable AI

In this section, the proposed system's predictions are explained, which are divided into two parts. In the first part, Grad-CAM and Saliency Map are used to interpret the work, and in the next part, the model's predictions are explained with LIME.

Recent advances in machine learning have generated a whole new generation of artificial intelligence applications that can help a wide range of sectors. Nonetheless, many of these systems are unable to explain their own judgments and actions to human users. Explainable AI (XAI) processes provide scientific explanations for its behavior so that people may better comprehend it. Some basic recommendations can help in the creation of more effective and human-friendly artificial intelligence systems: The XAI system should be able to describe its capabilities and understandings, as well as what it has done, what it is now doing, and what will happen next. It should also be able to reveal the crucial data to which it is reacting. Three different techniques were used to interpret the model: LIME, Grad-CAM, and Saliency Map.

5-1- LIME

Local Interpretable Model Agnostic Explanation (LIME) is the local aspect that indicates that it is utilized to explain specific machine learning model predictions [53]. It is one of the most widely used and highly popular XAI techniques. It manipulates the input data and creates a synthetic data stream that only holds a fraction of the original attributes. The explainable AI approach LIME may be applied with text, image, and tabular data and is compatible with a wide range of classifiers. Explanation produced by LIME can be acquired by Equation 4 [48].

$$\xi(x) = argmin\ [\zeta(f, g, \pi_x) + \Omega(g)] \quad (4)$$

Here, g is the existence of the interpretable elements, Ω(g) indicates the complexity which contrasts with the explainability, $\pi_x(z)$ is the locality between z to x, and f(x) describes the probability of x belonging to a class. Furthermore, the $\zeta(f,g,\pi_x)$ shows how unfaithful g is in approximating f in the proximity measured by π_x.

In Figure 8, Image A represents the Acute Lymphoblastic Leukemia - Benign class with affected cells that are pointed out by a line. While performing manual diagnostics, these cells are the sole reason a medical practitioner would label it as an Acute Lymphoblastic Leukemia - Benign cell. Thus, while establishing a prediction, the proposed framework should primarily focus on these affected cells. Image B illustrates the LIME prediction of Image A, which concentrates on those particular cells as anticipated. Here, the yellow and green colored features are the most relevant in the system's prediction. This explanation mechanism will increase medical practitioners' trust since they can now comprehend that the algorithm does not make predictions at random. Rather, the system initially identifies the impacted cells and later makes predictions based on those cells.

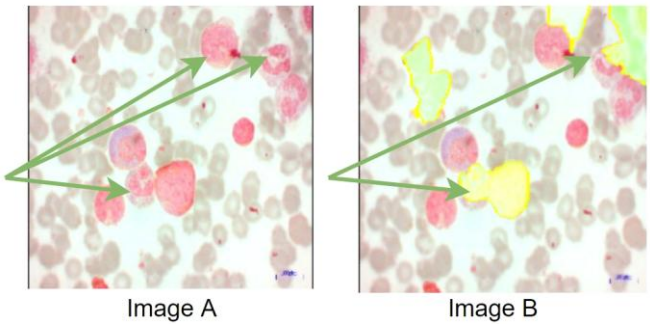


Figure 8. LIME Explanation: Image A is the original image whereas Image B is the LIME prediction. Infected blood cells are marked with an arrow in Image A. In Image B the LIME output is presented in green or yellow color which is marked as well.

5-2- Grad-CAM

Gradient-weighted Class Activation Mapping (Grad-CAM) employs the gradients of any target concept to build a crude localization map showing the significant regions in the picture for predicting the concept [54-56]. Grad-CAM has been developed to improve the transparency of Convolutional Neural Networks by highlighting the important features by employing class-specific gradient information. In the initial phase, Grad-CAM, a class-discriminative localization map $L^c_{Grad-CAM} \epsilon R^{uxu}$ is acquired by calculating the gradient of y^c with respect to feature maps A of a convolutional layer. For the weights of α_k^c global average pooling is implemented on these gradient values, which can be depicted by Equation 5.

$$\alpha_k^c = \frac{1}{z} \Sigma_i \Sigma_j \left(\partial y^c / \partial A_{ij}^k\right) \quad (5)$$

Here, The α_k^c is a partial linearization downstream from A that recognizes the significant region of the feature map k for a target class c. Furthermore, the y^c implies that the activation function is differentiable. Grad-CAM is a weighted integration of heat maps, and it's followed by the ReLU activation function, which can be written as Equation 6.

$$L^c_{Grad-CAM} = ReLU\left(\Sigma_k \alpha_k^c A^k\right) \quad (6)$$

As a consequence, a crude heat map is generated, which is then adjusted for visualization. With the exception of the ReLU in Equation 6, it is a Convolutional Neural Network-based architecture generalization of CAM. Prior to labeling a prediction, GRAD-CAM validates that the proposed algorithm is scanning and activating the appropriate regions in any given image.

5-3- Saliency Map

A Saliency Map, however, is an image in which the brightness of each pixel symbolizes how important that pixel is [57, 58]. In this technique, the brightness of each pixel is proportionate to its importance. Saliency Maps are sometimes known as heat maps, with warmth referring to the parts of the picture that have a significant influence on determining the object's class. The goal of the Saliency Map is to locate the areas that are conspicuous or noticeable at every point in the visual field and to use the spatial distribution of saliency to guide the selection of attended sites. It is utilized in a number of different Visual Attention models. The Saliency Map has great potential in reducing computational complexity. For example, traditionally, an image is taken as input and the result is anticipated using the entire image. So, if a CLL-affected image is used and a CLL cell is to be predicted, not all of the input is relevant, and not all of the information participates proportionally in anticipating the output. Therefore, if a large image contains just a few pixels corresponding to the class to be predicted, calculating the entire input is not a good strategy, which is why a Saliency Map can be employed to highlight the key aspects of the image and only evaluate the highlighted portions. This significantly assists in reducing computational complexity.

In Figure 9 image class A represents the input image which is Acute Lymphoblastic- Pro- B. While manually diagnosing this image, a medical practitioner will identify it as Acute Lymphoblastic- Pro- B based on the highlighted regions. Image B is the predictions from GRAD-CAM where it depicts a heatmap visualization for a given class label. It is also pointing toward that specific place by highlighting the features. The findings from the Saliency Map are shown in Image C, illustrating the most important features of the provided input. It also refers to that particular region, similar to GRAD-CAM. Medical practitioners may be more inclined to utilize the system due to these detailed explanations of the algorithm's predictions, as it classifies images in a manner similar to an expert.

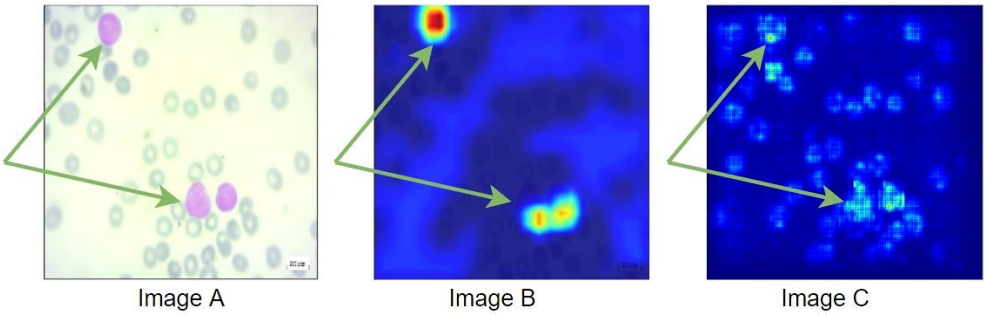


Figure 9. GRAD-CAM and Saliency Map Explanation: Image A is the input image, and the most crucial property on which a medical practitioner would decide its categorization is the pointed features. GRAD-CAM and Saliency Map results are shown in images B and C, respectively, indicating which features lead to the classification by the proposed framework.

5-4-XAI Explanation

Explanations of Saliency Map for each of the classes are provided in Figure 10. Here, (a) or the first depicted image is the original image followed by (b) which is the Saliency Map.

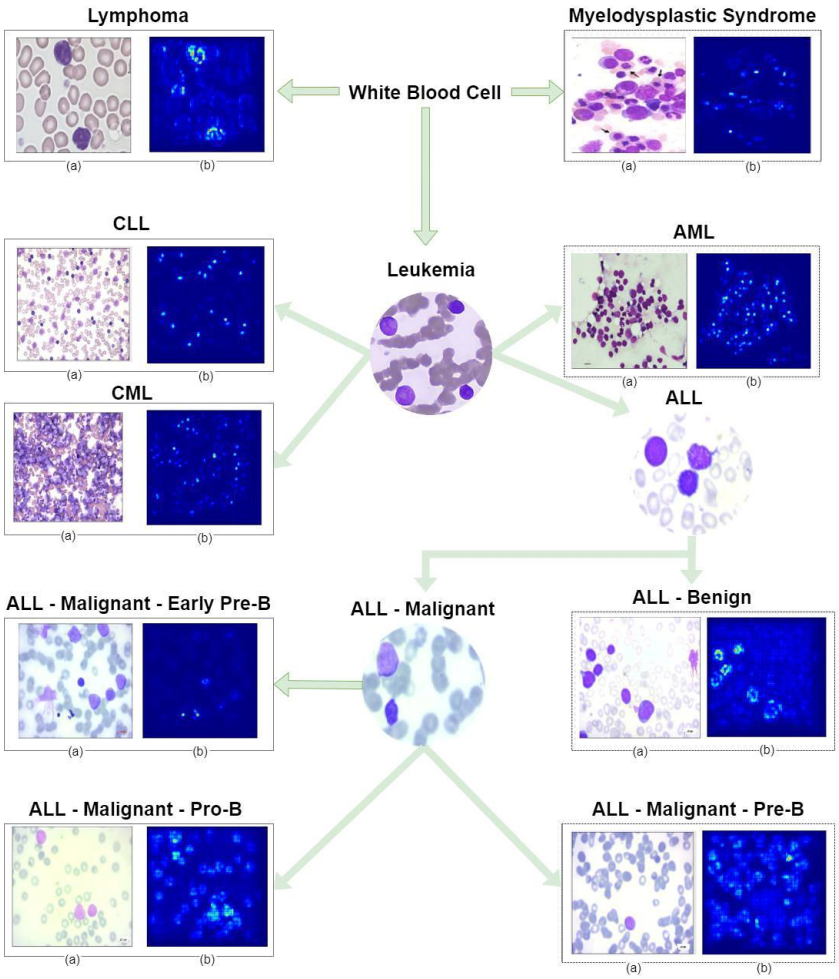
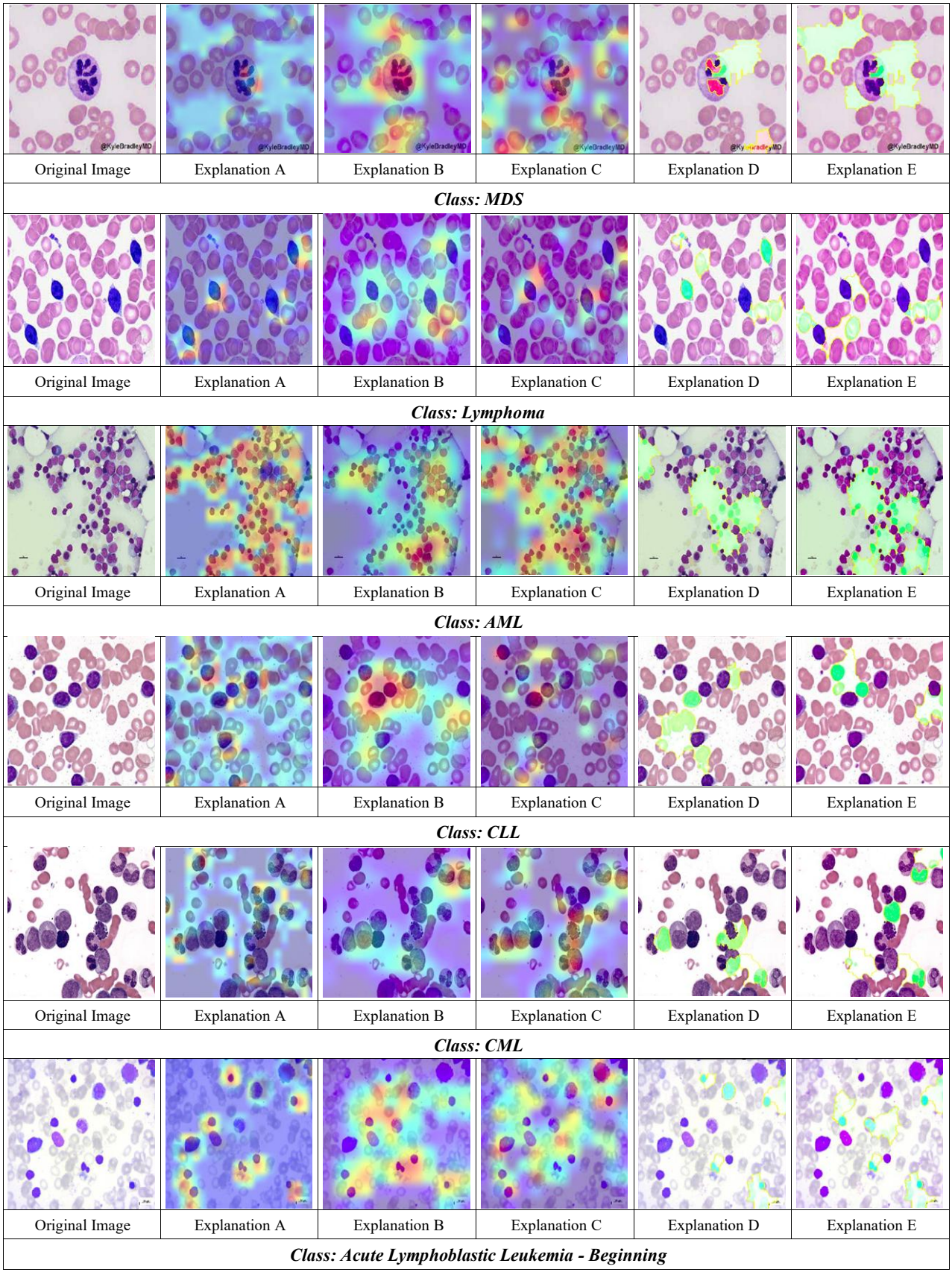


Figure 10. Visualization of Saliency Map Predictions: Here, the first image or (a) in every class is the original image, followed by (b), which is the Saliency Map. Here, the highlighted regions are the Saliency Map prediction

In Figure 11, the Grad-CAM and LIME outputs are presented in six columns. The first column contains the original image, while columns 2-4 show Grad-CAM results for standalone VGG19, Swin-integrated VGG19, and the Hybrid Swin+VGG19 model, respectively. Columns 5 and 6 display LIME outputs for VGG19 and the Hybrid model. The first image in each class is the original input. Explanations A-C correspond to Grad-CAM results: A for VGG19, B for the Hybrid model, and C for Swin-integrated LeukocyteNet. Explanations D and E depict LIME results for VGG19 and the Swin-integrated LeukocyteNet model, respectively.



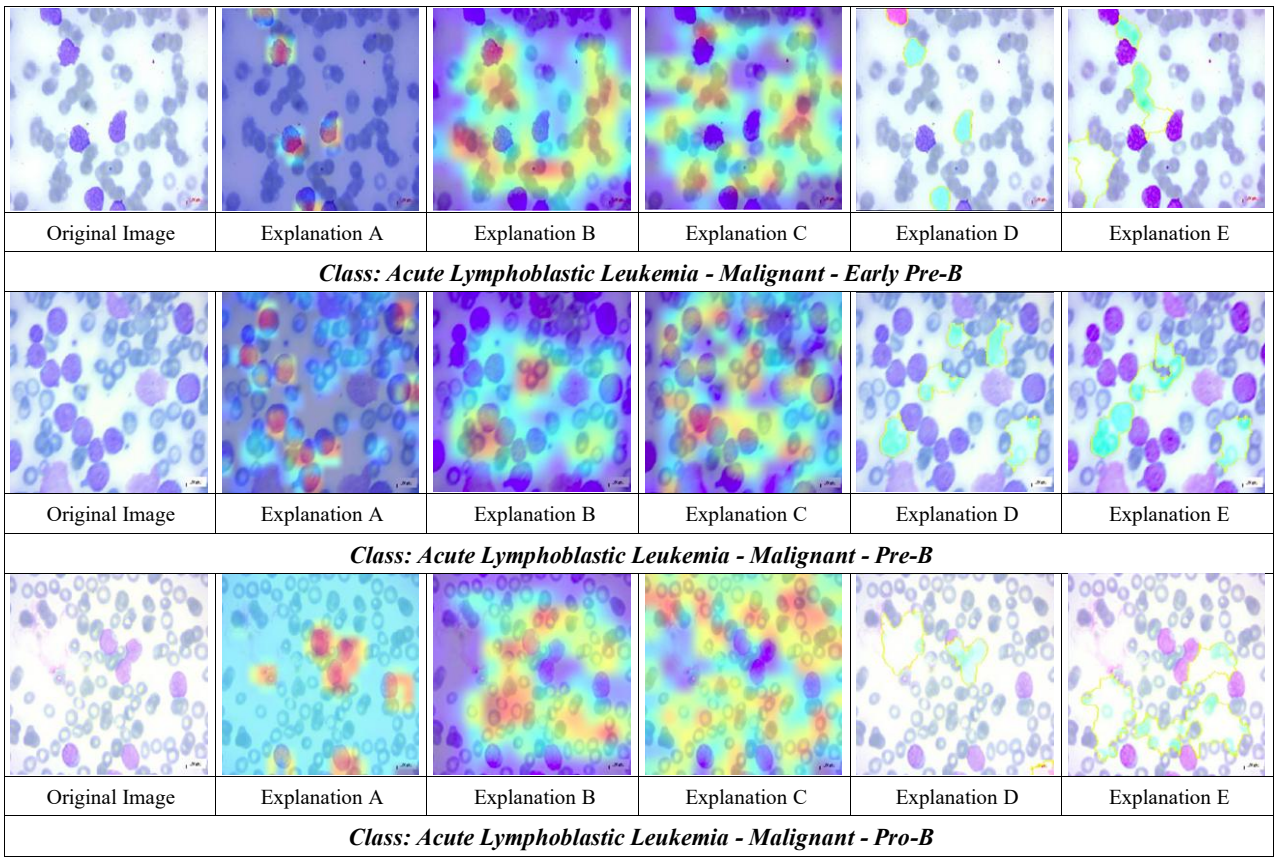


Figure 11. GRAD-CAM and LIME Predictions: Here, the first image in every class presents the original image. Explanations A, B, and C are GRAD-CAM predictions. Explanations D and E are LIME predictions. Explanation A depicts the VGG19 GRAD-CAM prediction. Explanation B denotes the Hybrid-VGG19 GRAD-CAM output and Explanation C depicts the SWIN-integrated LeukocyteNet GRAD-CAM prediction. Moreover, Explanation D portrays the LIME prediction from VGG19 and finally, Explanation E represents the LIME prediction for SWIN-integrated LeukocyteNet.

The output of LIME is comparatively less detailed than that of Grad-CAM. The output images are marked by two colors, green represents the regions responsible for correct classification while the red region is responsible for incorrect classification. In most cases, the red and the green regions are highlighted well over the blood cells except for exceptions such as MDS and AML, where the XAI mappings are spread a little across the background. It can also be observed that the regions marked red, which are supposedly responsible for the wrong classification, are somewhat absent from the LIME outputs. This represents the robustness and confidence of both models.

In addition, the Grad-CAM visualizations clearly indicate that LeukocyteNet (the hybrid Swin+VGG19 model) focuses its attention on clinically meaningful regions particularly the leukocyte nuclei, nuclear lobes, and cytoplasmic granules while reducing activation on irrelevant background areas. This spatial precision suggests that the transformer component contributes to enhanced global context understanding, whereas the convolutional backbone preserves fine morphological detail. In contrast, the standalone VGG19 model occasionally highlights peripheral noise or surrounding erythrocytes, implying less targeted feature learning. The LIME outputs further confirm that the hybrid model attributes positive weights predominantly within cell boundaries, emphasizing morphologically relevant structures such as chromatin texture and cytoplasmic contours.

Although no clinician-annotated masks were available to quantitatively benchmark explainability fidelity, the visual attention regions produced by Grad-CAM and LIME were qualitatively examined against diagnostic morphological cues described in hematology literature. The highlighted areas consistently corresponded to nuclei and cytoplasmic regions that hematologists regard as diagnostically relevant, indicating strong alignment between model focus and established clinical features. This qualitative consistency suggests that LeukocyteNet's explanations are medically meaningful even without pixel-level expert annotations. Future work will incorporate expert-verified region annotations to formally evaluate explainability fidelity using overlap metrics such as IoU and Dice scores.

Overall, these explainability maps demonstrate that LeukocyteNet's decisions align closely with the visual cues used by hematologists for manual leukocyte assessment. The improved localization and reduced background activation highlight the model's transparency, reliability, and potential to assist clinicians in explainable computer-aided diagnosis.

6- Conclusion

This study presented LeukocyteNet, a transfer–transformer fusion learning framework designed for the automated classification of ten distinct leukocyte types, including malignant and benign subcategories of leukemia, lymphoma, and myelodysplastic syndrome (MDS). The hybrid integration of VGG19 and Swin Transformer enabled the model to capture both local morphological features and global contextual dependencies from peripheral blood smear images. Extensive experiments across three publicly available datasets demonstrated the model's superior diagnostic performance, achieving 97.34% accuracy, 0.95 macro-averaged F1-score, and 0.93 recall, outperforming all benchmark CNN and Transformer baselines. The incorporation of explainable AI techniques Grad-CAM, LIME, and Saliency Map provided class-specific visualization of decision regions, enhancing explainability and building confidence among medical professionals. These features make LeukocyteNet a reliable, efficient, and transparent diagnostic aid that could support hematologists in early disease detection and reduce manual workload in clinical environments.

Future research will focus on enhancing the generalizability and scalability of LeukocyteNet. The model will be trained on larger, more heterogeneous datasets from multiple institutions to improve robustness across staining and imaging variations. Federated learning frameworks will be explored to enable cross-institutional collaboration while maintaining patient data confidentiality. Advanced augmentation pipelines and adaptive hyperparameter tuning are expected to improve recognition of rare leukocyte subtypes. Additionally, a clinically validated explainability benchmark will be developed to compare Grad-CAM and LIME outputs with expert annotations, ensuring the model's visual explanations align with pathologist reasoning. These enhancements aim to evolve LeukocyteNet into a deployable and trustworthy decision-support system, bridging the gap between artificial intelligence and practical clinical diagnosis in hematological imaging.

7- Declarations

7-1- Author Contributions

Conceptualization, T.S.A., M.G.R.A., M.T.R., M.A.A., and F.F.; methodology, T.S.A., M.G.R.A., M.T.R., M.A.A., and F.F.; software, T.S.A., F.F., and H.M.S.A.; validation, S.B.; formal analysis, T.S.A., M.G.R.A., M.T.R., F.F., M.F.T., and H.M.S.A.; investigation, T.S.A., F.F., and M.F.T.; resources, T.S.A., M.G.R.A., M.T.R., and F.F.; writing—original draft preparation, T.S.A. and F.F.; writing—review and editing, F.F.; visualization, T.S.A., M.G.R.A., M.T.R., M.A.A., F.F., H.M.S.A., M.A., K.F.H., and M.A.M.; project administration, T.S.A., F.F., and M.G.R.A.; supervision, M.G.R.A., M.T.R., M.A.A., M.A., K.F.H., and M.A.M. All authors have read and agreed to the published version of the manuscript.

7-2- Data Availability Statement

The following three datasets were employed in the analysis: The ASH image bank, Taleqani Hospital, and the ALL-IDB all of which are publicly accessible datasets and were properly cited in the study. Supplementary code for this study may be accessed online at: *https://github.com/Sakibapon/LeukocyteNet* (accessed on May 2026).

7-3- Funding

This research was fully funded and supported by BRAC University, Dhaka 1212, Bangladesh.

7-4- Institutional Review Board Statement

Not applicable.

7-5- Informed Consent Statement

Not applicable.

7-6- Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

8- References

[1] Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 68(6), 394–424. doi:10.3322/caac.21492.

[2] Cazzola, M., & Malcovati, L. (2005). Myelodysplastic Syndromes — Coping with Ineffective Hematopoiesis. New England Journal of Medicine, 352(6), 536–538. doi:10.1056/nejmp048266.

[3] Chennamadhavuni, A., Iyengar, V., Mukkamalla, S. K. R., & Shimanovsky, A. (2023). Leukemia. StatPearls Publishing, Florida, United States. Available online: https://www.ncbi.nlm.nih.gov/books/NBK560490/ (accessed on May 2026).

[4] Maxhealthcare (2026). Is blood cancer curable: Types, symptoms & causes of blood cancer. Is Blood Cancer Curable: Types, Symptoms & Causes of Blood Cancer. Available online: https://www.maxhealthcare.in/blogs/is-blood-cancer-curable (accessed on May 2026).

[5] Deshpande, N. M., Gite, S., & Aluvalu, R. (2021). A review of microscopic analysis of blood cells for disease detection with AI perspective. PeerJ Computer Science, 7, e460. doi:10.7717/peerj-cs.460.

[6] Tran, K. A., Kondrashova, O., Bradley, A., Williams, E. D., Pearson, J. V., & Waddell, N. (2021). Deep learning in cancer diagnosis, prognosis and treatment selection. Genome medicine, 13(1), 152. doi:10.1186/S13073-021-00968-X.

[7] Alazaidah, R., Samara, G., Abu Asi, H., Abuowaida, S., Mashagba, H. A., Abd Aziz, A., … Al-Bawri, S. S. (2025). Breast Cancer Classification Using Deep Feature Extraction and Machine Learning. HighTech and Innovation Journal, 6(4), 1282–1299. doi:10.28991/HIJ-2025-06-04-09.

[8] Islam, M. S., Hasan, K. F., Sultana, S., Uddin, S., Lio', P., Quinn, J. M. W., & Moni, M. A. (2023). HARDC: A novel ECG-based heartbeat classification method to detect arrhythmia using hierarchical attention based dual structured RNN with dilated CNN. Neural Networks, 162, 271–287. doi:10.1016/j.neunet.2023.03.004.

[9] Talukder, M. A., Islam, M. M., Uddin, M. A., Akhter, A., Hasan, K. F., & Moni, M. A. (2022). Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning. Expert Systems with Applications, 205, 117695. doi:10.1016/j.eswa.2022.117695.

[10] Bala, M., Ali, M. H., Satu, M. S., Hasan, K. F., & Moni, M. A. (2022). Efficient Machine Learning Models for Early Stage Detection of Autism Spectrum Disorder. Algorithms, 15(5), 166. doi:10.3390/a15050166.

[11] Watson, D. S., Krutzinna, J., Bruce, I. N., Griffiths, C. E. M., McInnes, I. B., Barnes, M. R., & Floridi, L. (2019). Clinical applications of machine learning algorithms: Beyond the black box. BMJ, 364. doi:10.1136/bmj.l886.

[12] Loh, H. W., Ooi, C. P., Seoni, S., Barua, P. D., Molinari, F., & Acharya, U. R. (2022). Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). Computer Methods and Programs in Biomedicine, 226, 107161. doi:10.1016/j.cmpb.2022.107161.

[13] Kabir, M. H., Hasan, K. F., Hasan, M. K., & Ansari, K. (2022). Explainable Artificial Intelligence for Smart City Application: A Secure and Trusted Platform. Explainable Artificial Intelligence for Cyber Security, Studies in Computational Intelligence, Springer, Cham, Switzerland. doi:10.1007/978-3-030-96630-0_11.

[14] Jia, N., Guo, J., Li, Y., Tang, S., Xu, L., Liu, L., & Xing, J. (2024). A fine-grained image classification algorithm based on self-supervised learning and multi-feature fusion of blood cells. Scientific Reports, 14(1), 22964. doi:10.1038/s41598-024-74753-2.

[15] Anand, V., Gupta, S., Koundal, D., Alghamdi, W. Y., & Alsharbi, B. M. (2024). Deep learning-based image annotation for leukocyte segmentation and classification of blood cell morphology. BMC Medical Imaging, 24(1), 83. doi:10.1186/s12880-024-01254-z.

[16] Üzen, H., & Fırat, H. (2024). A hybrid approach based on multipath Swin transformer and ConvMixer for white blood cells classification. Health Information Science and Systems, 12(1), 33. doi:10.1007/s13755-024-00291-w.

[17] Ghosh, A., Singh, S., & Sheet, D. (2017). Simultaneous localization and classification of acute lymphoblastic leukemic cells in peripheral blood smears using a deep convolutional network with average pooling layer. 2017 IEEE International Conference on Industrial and Information Systems (ICIIS), 1–6. doi:10.1109/ICIINFS.2017.8300425.

[18] Rota, P., Groeneveld-Krentz, S., & Reiter, M. (2015). On automated Flow Cytometric analysis for MRD estimation of Acute Lymphoblastic Leukaemia: A comparison among different approaches. 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 438–441. doi:10.1109/BIBM.2015.7359723.

[19] Amin, M. M., Kermani, S., Talebi, A., & Oghli, M. G. (2015). Recognition of acute lymphoblastic leukemia cells in microscopic images using k-means clustering and support vector machine classifier. Journal of Medical Signals & Sensors, 5(1), 49–58. doi:10.4103/2228-7477.150428.

[20] 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. doi:10.1002/jemt.23139.

[21] Shafique, S., & Tehsin, S. (2018). Acute Lymphoblastic Leukemia Detection and Classification of Its Subtypes Using Pretrained Deep Convolutional Neural Networks. Technology in Cancer Research &Amp; Treatment, 17. doi:10.1177/1533033818802789.

[22] Wang, Q., Wang, J., Zhou, M., Li, Q., & Wang, Y. (2017). Spectral-spatial feature-based neural network method for acute lymphoblastic leukemia cell identification via microscopic hyperspectral imaging technology. Biomedical Optics Express, 8(6), 3017. doi:10.1364/boe.8.003017.

[23] Liu, Y., & Long, F. (2019). Acute Lymphoblastic Leukemia Cells Image Analysis with Deep Bagging Ensemble Learning. ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging, Lecture Notes in Bioengineering, Springer, Singapore. doi:10.1007/978-981-15-0798-4_12.

[24] Kassani, S. H., Kassani, P. H., Wesolowski, M. J., Schneider, K. A., & Deters, R. (2019). A Hybrid Deep Learning Architecture for Leukemic B-lymphoblast Classification. 2019 International Conference on Information and Communication Technology Convergence (ICTC), 271–276. doi:10.1109/ICTC46691.2019.8939959.

[25] Safuan, S. N. M., Tomari, M. R. M., Zakaria, W. N. W., Mohd, M. N. H., & Suriani, N. S. (2020). Investigation of white blood cell biomaker model for acute lymphoblastic leukemia detection based on convolutional neural network. Bulletin of Electrical Engineering and Informatics, 9(2), 611–618. doi:10.11591/eei.v9i2.1857.

[26] Ayyappan, V., Chang, A., Zhang, C., Paidi, S. K., Bordett, R., Liang, T., Barman, I., & Pandey, R. (2020). Identification and staging of B-cell acute lymphoblastic leukemia using quantitative phase imaging and machine learning. ACS Sensors, 5(10), 3281–3289. doi:10.1021/acssensors.0c01811.

[27] Bibi, N., Sikandar, M., Din, I. U., Almogren, A., & Ali, S. (2020). IOMT-based automated detection and classification of leukemia using deep learning. Journal of Healthcare Engineering, 2020(1), 6648574. doi:10.1155/2020/6648574.

[28] Genovese, A., Hosseini, M. S., Piuri, V., Plataniotis, K. N., & Scotti, F. (2021). Histopathological Transfer Learning for Acute Lymphoblastic Leukemia Detection. 2021 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), 1–6. doi:10.1109/CIVEMSA52099.2021.9493677.

[29] Das, P. K., & Meher, S. (2021). An efficient deep Convolutional Neural Network based detection and classification of Acute Lymphoblastic Leukemia. Expert Systems with Applications, 183, 115311. doi:10.1016/j.eswa.2021.115311.

[30] Dhalla, S., Mittal, A., Gupta, S., & Singh, H. (2021). Multi-model Ensemble to Classify Acute Lymphoblastic Leukemia in Blood Smear Images. Pattern Recognition, ICPR International Workshops and Challenges, ICPR 2021, Lecture Notes in Computer Science, Springer, Cham, Switzerland. doi:10.1007/978-3-030-68763-2_18.

[31] Khandekar, R., Shastry, P., Jaishankar, S., Faust, O., & Sampathila, N. (2021). Automated blast cell detection for Acute Lymphoblastic Leukemia diagnosis. Biomedical Signal Processing and Control, 68, 102690. doi:10.1016/j.bspc.2021.102690.

[32] Ahsan, M. M., Gupta, K. D., Islam, M. M., Sen, S., Rahman, M. L., & Hossain, M. S. (2020). Study of Different Deep Learning Approach with Explainable AI for Screening Patients with COVID-19 Symptoms: Using CT Scan and Chest X-ray Image Dataset. doi:10.3390/make2040027.

[33] Ye, Q., Xia, J., & Yang, G. (2021). Explainable AI for COVID-19 CT Classifiers: An Initial Comparison Study. 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS), 521–526. doi:10.1109/CBMS52027.2021.00103.

[34] Zhang, A. Y., Lam, S. S. W., Ong, M. E. H., Tang, P. H., & Chan, L. L. (2019). Explainable AI. Proceedings of the 6th IEEE/ACM International Conference on Big Data Computing, Applications and Technologies, 95–102. doi:10.1145/3365109.3368791.

[35] Apon, T. S., Hasan, M. M., Islam, A., & Alam, Md. G. R. (2021). Demystifying Deep Learning Models for Retinal OCT Disease Classification using Explainable AI. 2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), 1–6. doi:10.1109/CSDE53843.2021.9718400.

[36] Halder, A., Gharami, S., Sadhu, P., Singh, P. K., Woźniak, M., & Ijaz, M. F. (2024). Implementing vision transformer for classifying 2D biomedical images. Scientific Reports, 14(1), 12567. doi:10.1038/s41598-024-63094-9.

[37] Raju, A. S. N., Venkatesh, K., Padmaja, B., Kumar, C. N. S., Patnala, P. R. M., Lasisi, A., Islam, S., Razak, A., & Khan, W. A. (2024). Exploring vision transformers and XGBoost as deep learning ensembles for transforming carcinoma recognition. Scientific Reports, 14(1), 30052. doi:10.1038/s41598-024-81456-1.

[38] 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. doi:10.1002/int.22753.

[39] Labati, R. D., Piuri, V., & Scotti, F. (2011). All-IDB: The acute lymphoblastic leukemia image database for image processing. 2011 18th IEEE International Conference on Image Processing, 2045–2048. doi:10.1109/ICIP.2011.6115881.

[40] Imagebank (2026). The American Society of Hematology, Washington, United States. Available Online: https://imagebank.hematology.org/ (accessed on May 2026).

[41] Erlinger, T. P., Muntner, P., & Helzlsouer, K. J. (2004). WBC count and the risk of cancer mortality in a national sample of U.S. adults: Results from the Second National Health and Nutrition Examination Survey mortality study. Cancer Epidemiology Biomarkers and Prevention, 13(6), 1052–1056. doi:10.1158/1055-9965.1052.13.6.

[42] Faivdullah, L., Azahar, F., Htike, Z. Z., & Naing, W. Y. N. (2015). Leukemia Detection from Blood Smears. Journal of Medical and Bioengineering, 4(6), 488–491. doi:10.12720/jomb.4.6.488-491.

[43] Karim, Z. A. A., Khidhir, K. G., Ahmed, R. A., Hassan, H. A., & Karim, D. O. (2016). Leukemia Study in Sulaymaniyah Province, Kurdistan, Iraq. Chinese Medical Journal, 129(2), 244–245. doi:10.4103/0366-6999.173551.

[44] SEER Cancer Statistics. (2009). SEER cancer statistics review, 1975-2006. SEER Cancer Statistics, Washington D.C., United States. Available online: http://seer.cancer.gov/csr/1975_2006/ (accessed on May 2026).

[45] Swerdlow, S. H., Campo, E., Pileri, S. A., Harris, N. L., Stein, H., Siebert, R., Advani, R., Ghielmini, M., Salles, G. A., Zelenetz, A. D., & Jaffe, E. S. (2016). The 2016 revision of the World Health Organization classification of lymphoid neoplasms. Blood, 127(20), 2375–2390. doi:10.1182/blood-2016-01-643569.

[46] Stone, R. M., O'Donnell, M. R., & Sekeres, M. A. (2004). Acute Myeloid Leukemia. Hematology, 2004(1), 98–117. doi:10.1182/asheducation-2004.1.98.

[47] Döhner, H., Estey, E., Grimwade, D., Amadori, S., Appelbaum, F. R., Büchner, T., Dombret, H., Ebert, B. L., Fenaux, P., Larson, R. A., Levine, R. L., Lo-Coco, F., Naoe, T., Niederwieser, D., Ossenkoppele, G. J., Sanz, M., Sierra, J., Tallman, M. S., Tien, H. F., … Bloomfield, C. D. (2017). Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood, 129(4), 424–447. doi:10.1182/blood-2016-08-733196.

[48] Milne, K., Sturrock, B., & Chevassut, T. (2020). Chronic Lymphocytic Leukaemia in 2020: the Future Has Arrived. Current Oncology Reports, 22(4), 36. doi:10.1007/s11912-020-0893-0.

[49] Cuttner, J. (1992). Increased incidence of hematologic malignancies in first-degree relatives of patients with chronic lymphocytic leukemia. Cancer Investigation, 10(2), 103–109. doi:10.3109/07357909209032771.

[50] Pc, N. (1960). A Minute Chromosome in Human Chronic Granulocytic Leukemia. Science, 132, 1497.

[51] Sawyers, C. L. (1999). Chronic Myeloid Leukemia. New England Journal of Medicine, 340(17), 1330–1340. doi:10.1056/nejm199904293401706.

[52] Rollison, D. E., Howlader, N., Smith, M. T., Strom, S. S., Merritt, W. D., Ries, L. A., Edwards, B. K., & List, A. F. (2008). Epidemiology of myelodysplastic syndromes and chronic myeloproliferative disorders in the United States, 2001-2004, using data from the NAACCR and SEER programs. Blood, 112(1), 45–52. doi:10.1182/blood-2008-01-134858.

[53] Ribeiro, M., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, 97–101. doi:10.18653/v1/n16-3020.

[54] Chattopadhay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018). Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 839–847. doi:10.1109/WACV.2018.00097.

[55] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. International Journal of Computer Vision, 128(2), 336–359. doi:10.1007/s11263-019-01228-7.

[56] Selvaraju, R. R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., & Batra, D. (2016). Grad-CAM: Why did you say that?. arXiv Preprint, arXiv:1611.07450. doi:10.48550/arViv.161107450.

[57] Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv Preprint, arXiv:1312.6034. doi:10.48550/arXiv.1312.6034.

[58] Mundhenk, T. N., Chen, B. Y., & Friedland, G. (2019). Efficient saliency maps for explainable AI. arXiv Preprint, arXiv:1911.11293. doi:10.48550/arXiv.1911.11293.