SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

Integrating convolutional neural networks for microscopic image analysis in acute lymphoblastic leukemia classification: A deep learning approach for enhanced diagnostic precision

Soumya Madduru[1], Dr. Venkata Raju K[2]

[1]Research Scholar, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh-522502, India
soumya.mahicse@gmail.com ORCID:0000-0001-7567-0357
[2]Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh-522502, India,
kvraju@kluniversity.in

Abstract

Accurate and timely classification of acute lymphoblastic leukemia (ALL) based on microscopic blood smear images is a major priority area in computer-aided diagnosis, but the high intra-class similarity and the uneven or variable stains often make the prior models generally limited in their applicability. A hybrid CNN-ViT model is proposed in the study to combine spatial local representations of Convolutional Neural Networks (CNNs) with global contextual relationships of Vision Transformers (ViTs). To alleviate the scarcity of data and the class imbalance, a lot of data augmentation and transfer learning were done, and model explanations were improved by gradient-weighted Grad-CAM to interpret the discriminative cellular regions. A publicly available leukocyte dataset evaluation shows that the proposed Hybrid CNN-ViT achieved a state-of-the-art validation of 99.60%, beating known benchmark models including DenseNet-121, ResNet-50, and EfficientNet-B0. The comparative analysis proves the strength of the model, and visual interpretations demonstrate that the trained maps of attention reflect the biological salient features of leukocytes. These findings point to the clinical promise of hybrid representation learning in the diagnosis of hematological images, notably, to the support of a diagnostic workflow in ALL. Nevertheless, the framework introduces computational overhead and data set specificities, making it necessary to be further validated within a multi-institutional and real-world clinical setting.

Introduction

Acute lymphoblastic leukemia (ALL) is a severe hematologic malignancy caused by the uncontrolled proliferation of immature lymphoid precursor cells in bone marrow, blood, and lymphoid organs [1]. Although it is less common in adults, where it is frequently linked to worse prognoses and higher relapse rates, it is the most common cancer in children, making up 25–30% of all childhood cases [2]. Early and precise diagnosis, including accurate subtyping, is crucial for risk-adapted therapy and enhanced patient survival [3]. Distinguishing subtypes relies on subtle morphological and cytochemical features, including nuclear size and shape, chromatin patterns, nucleocytoplasmic ratio, and cytoplasmic inclusions, which are challenging to identify even for experienced hematopathologists [4]. Conventional microscopy-based diagnosis is labor-intensive and time-consuming, emphasizing the need for artificial intelligence (AI) and machine learning (ML) approaches to automate cytological analysis, improve accuracy, and deliver interpretable results efficiently.

Over the last few years, deep learning (DL) has transformed the biomedical image analysis field by providing a means to extract hierarchical and complex features from high-dimensional imaging data in an automated fashion. CNNs, specifically, have shown impressive performance in cellular and histopathological structure detection, segmentation, and classification with several studies reporting high classification rates on single-institutional datasets. Nevertheless, these encouraging findings notwithstanding, there are a number of serious limitations that make it difficult to translate these models to the clinic. First, models trained on a single dataset perform poorly in generalization across different populations as well as distinct imaging protocols, where domain shifts caused by differences in the staining technologies, scanners, and sample preparation methods are likely to have a severe impact on model performance [5]. Second, small morphological discrepancies among subtypes of ALL, e.g.,

differences in nuclear-to-cytoplasmic ratio, chromatin granularity, and the presence of cytoplasmic vacuoles, are usually not easily captured by standard CNNs, especially in the presence of little inter-class variance [6]. Third, in most architectures, the decision-making process of DL is opaque and, thus, inhibits the creation of trust and limits clinical application among pathologists [7]. Lastly, a majority of previous research is highly biased towards accuracy without taking into account the efficiency of the models, their interpretability and resistance to noise in the real world where they are needed in the hospital [8]. The combination of these gaps highlights the importance of the desirability of models that can not only reach high accuracy, but also generalize across diverse data sets, be interpretable, and be computable at scale in large-scale clinical settings.

To deal with these drawbacks, this study presents hybrid computational models that integrate CNNs and ViTs in order to exploit local and global feature representations. CNNs are effective in capturing local details like nuclear outline, cellular granularity, and minor chromatin texture, whereas ViTs, through self-attention processes, are able to figure out long-range connections and interdependencies across the whole picture [9]. With the combination of these architectures, one can be able to determine complex cellular features like predicting performance at the same time taking into consideration the wider morphological context, which is especially useful when analysing morphologically similar ALL subtypes [10].

The main goal of this research can be described as having threefold objectives. First, it strives to build an effective, multi-institutional dataset pipeline to record a broad range of morphological variability in ALL. Second, it aims to improve the feature representation capabilities with a hybrid CNN-ViT architecture complemented by sophisticated augmentation techniques, to classify subtle leukocyte subtypes with high accuracy and interpretability by drawing Grad-CAM attention maps. Third, detailed comparisons are done with other reference CNN structures, including ResNet-50, DenseNet-121, EfficientNetB0 and VGG-16, in terms of classification performance, computational resources, robustness and medical interpretability.

This study has important contributions both to the field of computational pathology and clinical diagnostics. First, it proposes a hybrid CNN-ViT that obtains morphological characteristics of the local and global structures, which overcomes the drawbacks of conventional CNNs. Second, it offers an instructive data preprocessing and augmentation pathway that can be used to aid generalizability to substantially different multi-institutional datasets, thus mitigating domain shift vulnerabilities. Third, interpretability is integrated in the form of Grad-CAM that enables understanding the location of the cellular structures that shapes the judgment of the model, which facilitates clinical trust and embracement. The main objectives of this study are structured according to the workflow of the proposed ALL subtype classification framework and can be summarized as follows:

- **Dataset Collection and Integration**: To create a comprehensive microscopic leukocyte image dataset by combining multi-institutional samples that capture variations in staining protocols, imaging devices, and patient demographics, thereby enhancing the generalization capability of the proposed model.
- **Data Preprocessing and Augmentation**: To develop an effective preprocessing and augmentation pipeline that includes normalization, scaling, shuffling, and geometric as well as photometric transformations in order to enhance robustness against clinical imaging variability.
- **Hybrid CNN–ViT Model Development**: To develop and implement a hybrid CNN–Vision Transformer architecture that combines local feature extraction with global contextual modeling, allowing accurate representation of cytomorphological patterns and spatial dependencies for ALL subtype classification.
- **Performance Evaluation and Clinical Interpretability**: In order to provide high diagnostic accuracy and predictions that can be clinically interpreted, the presented framework will be evaluated by benchmarking against cutting-edge deep learning models and integrating explainability mechanisms as Grad-CAM and attention visualization.

The rest of this paper is organized as follows: section 2 reviews related work in image analysis of leukemia and DL methods. The methodology is reported in section 3 with the description of the proposed hybrid CNN-ViT model, design of data augmentation and preprocessing, as well as training regimes. section 4 presents experimental results, the comparison of these results with those of baseline CNN models, and observes how the proposed models can be interpreted clinically. Lastly, section 5 concludes this research implicing the limitations and future scopes for it.

Literature Review

The paper will review the approach to using AI-based models, specifically the DL framework through CNNs and transfer learning, to detect and classify leukemia. Even in the light of the advancements, we still have some issues with model generalization, data limitations, and interpretability.

Narayanan et al. [11] proposed a hybrid approach, the fuzzy C-means with random forest to the classification of acute lymphocytic leukemia (ALL) based on microscopic blood smear images. The analysis employed an original data set of 8,637 pictures, comprising infected, normal, and augmented images across several sources. The procedure comprises various main steps: pre-processing of the image (including a conversion of RGB to CMYK as well as an equalization of the histogram), clustering with the fuzzy C-means algorithm, and feature extraction, to be used in a classification procedure. The result of their hybrid model attained a high accuracy of 99.06%, sensitivity of 99.4%, and specificity of 97.8%, surpassing standard techniques and other ML models like support vector machines (SVM) and CNNs. The use of high-quality labeled datasets and computational resources used to train the model is, however, a limitation.

Rai et al. [12] explored the performance of various DL architectures for leukemia classification, comparing models like VGG, EfficientNet, LeNet, AlexNet, CNN, and ResNet on a comprehensive dataset of 12,528 high-resolution images. The study applied advanced data preprocessing techniques to ensure optimal model training. AlexNet achieved the highest performance, with an accuracy of 89.3%, precision of 99.5%, and recall of 86.8%. Nevertheless, there are still issues of model explanations and inconsistency in data to be tackled that indicate further research areas to increase the scale and applicability of the DL models.

Muduli et al. [13] proposed a detailed work including DL models of ALL detection and classification of peripheral blood smear (PBS) images. They tested multiple number of pre-trained models such as VGG16, VGG19, ResNet50, Xception, ResNet152, EfficientNetB0, NASNetMobile, DenseNet169, DenseNet121, and EfficientNetV2B0. The DenseNet121 model was the most successful attaining an impressively high accuracy of 99%, much better than the other models that achieved 72% with EfficientNet-B0 and 81% with NASNetMobile. However, the issue of overfitting, given the small size of the dataset and the complexity of the clinical environment in general, adds the potential of much larger and more diverse biomarker measurement datasets.

Shahbaz et al [14] provided research on early leukemia risk prediction and type classification with the help of Explainable AI (XAI) on the basis of clinical data of Pakistan. This research exploited 364 cases of leukemia and 896 healthy controls based on demographic and lifestyle factors, lab data, and other measures. ML methods were used by the authors, with XGBoost, OR, and CDR determining risk factors and identification of leukemia subtypes. To classify, the research applied the graph-based data as well as tabular structure (structured) data, and the structured data model attained 96% accuracy following a balance in classes using oversampling. Although the method exhibited a high level of prediction, the authors have mentioned that the data imbalances and regional lifestyle differences could affect the generalization.

Park and Miyano [15] developed the GRN-multiClassifier, a network-based classifier that was used to classify acute leukemia cell lines into three categories: B-cell ALL, T-cell ALL, and AML. The model optimizes the GRNs by minimizing the classification errors as well as network estimation errors using multi-class classification algorithms. By using the 50 genes, 100 genes, and 300 genes with the highest variability in gene expression, the model has attained 100% accuracy, indicating that even a small subset of genes (e.g.50 genes) was sufficient in classifying individuals properly. Although accuracy was also slightly enhanced when the number of genes was increased, the utilisation of 50 genes alone showed the best results.

Mustaqim et al. [16] presented FocusAugMix, a sophisticated data augmenting technique that enhances ALL classification, specifically on small data sets. To make the model more robust, the approach combines features extracted with superpixel segmentation, Gradient-weighted Class Activation mapping (Grad-CAM), Multi-Head Attention and SaliencyMix. FocusAugMix was also implemented on different leukemia datasets, such as ALL-IDB1, ALL-IDB2, where the best accuracy of 99.07% was achieved on the multi-cell images, which are much better than the earlier-used augmentation methods. Despite the impressive outcomes, the method faces challenges, especially with overlapping cells and inconsistent staining.

To improve the classification of the ALL based on microscopic blood cell images, Jawahar et al. [17] introduced an attention-based DL model, DDRNet. The architecture combines some stacked sophisticated blocks: Deep Residual Dilated Blocks (DRDB), Global and Local Feature Enhancement Blocks (GLFEB), and Channel and Spatial Attention Blocks (CSAB). It was validated on the blood cell dataset on Kaggle, and its training accuracy stood at 99.86 %, and testing was 91.98 %. The model was far better in performance than existing ones, like ResNet variants, and had an F1-score of 0.96. Although DDRNet exhibits great results, there are still difficulties attached to it, including class imbalance and the fact that its process of decision-making is less obvious.

Bose and Bandyopadhyay [18] discussed an automated leukemia detection system for acute lymphocytic leukemia (ALL) that uses a hybrid learning approach combining traditional image analysis with ML. The initial stages of the system are to capture the cells of leukemia with the help of image preprocessing and segmentation techniques. During the second phase, the identified leukemia cells are further classified into subtypes using the Support Vector Machine (SVM), and the feature extraction with ResNet50 optimised the performance of the model. The accuracy of the proposed system was found to be 99.9%, which is much better when compared to the previous systems of visual inspection.

Asar and Ragab [19] introduced the FOADCNN-LDC method of detecting and classifying leukemia, which used a deep CNN with the Falcon optimization algorithm (FOA) that employs feature extraction with ShuffleNetv2, a Convolutional Denoising Autoencoder (CDAE) classifier, and the FOA to optimise the hyperparameters of the CDAE classification scheme. The model has demonstrated excellent performance with an accuracy of 99.62% in a benchmark medical dataset and has been able to predict and classify leukemia. But the research also identified related difficulties in image noise and data imbalance, and it is suggested that additional studies should be conducted to idealize the computational complexity and advanced generalization abilities of different sets of data. Cheng et al. [20] considered the application of DL to increase the analysis of data on flow cytometry in terms of the detection of acute leukemia and the classification of its cells used in separating different variants of acute leukemia, such as AML, B-ALL, and T-ALL. ResNet-50 and EverFlow were used to train a DL model to analyze flow cytometry data, which showed high sensitivity for AML (94.6%) and B-ALL (98.2%). The model also gave good results in segregation of > 80% physiological cell sensitivities. Nonetheless, AI performance was partially limited by the presence of CD34-negative cell types, which reduced the detection of certain pathological cells.

Although there have been considerable advancements in the use of AI-based leukemia detection, the literature pertaining to these advancements indicates some key limitations. Datasets used in many studies are restricted to a single language (usually English), and thus, the models lack generalizability to use in different multilingual and cross-cultural scenarios. Moreover, low-resourced environments may present difficulties in deploying such computing-intensive models and may also suffer biases in the provided datasets. Notwithstanding these limitations, the reviewed studies form an excellent basis for detecting leukemia. A crystalized report of the approaches, methods, and findings, along with their respective benefits and drawbacks, has been compiled in Table 1.

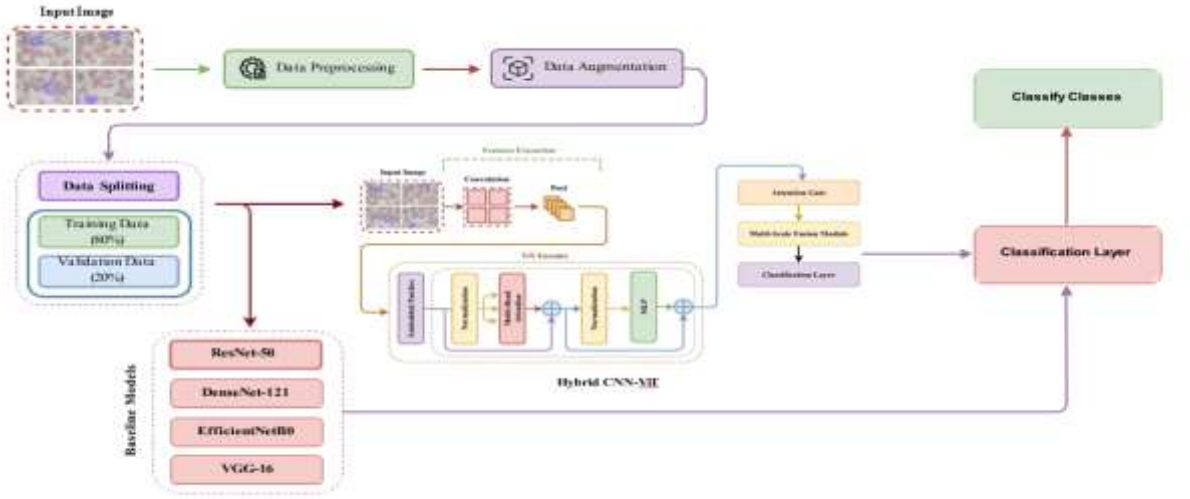
Table 1 : Summary of Leukemia Classification Models

Year	Reference	Model	Results	Limitations
2025	[11]	Fuzzy C-means + Random Forest	Accuracy: 99.06%, Sensitivity: 99.4%, Specificity: 97.8%	Requires high-quality labeled datasets; computational resources
2025	[12]	Various DL Architectures	Accuracy: 89.3%, Precision: 99.5%, Recall: 86.8%	Model explanations and data consistency issues
2025	[13]	DenseNet121	Accuracy: 99%, Precision: 99.3%, Recall: 99%	Overfitting due to small dataset; clinical complexity
2025	[14]	XGBoost & OR with XAI	Accuracy: 96%	Data imbalances; regional lifestyle differences affect generalization
2025	[15]	GRN-multiClassifier	Accuracy: 100%	Limited to small gene set; needs larger datasets for generalization
2025	[16]	FocusAugMix (Data Augmentation)	Accuracy: 99.07%	Issues with overlapping cells and staining variations
2025	[18]	SVM + ResNet50	Accuracy: 99.9%	Needs high-quality blood smear images; preprocessing required
2025	[19]	FOADCNN-LDC (CNN + FOA)	Accuracy: 99.62%	Image noise and data imbalance issues
2025	[20]	ResNet-50 + EverFlow	Sensitivity: 94.6% (AML), 98.2% (B-ALL)	Insufficient performance on CD34-negative cells

Methodology

In this work, a customized CNN with a ViT-based hybrid DL model is proposed as a technique to automatically identify the subtypes of leukocytes in ALL. The workflow in general is shown on Figure 1.

Figure 1: Workflow of the proposed framework



Dataset Description

The data used in this paper takes the form of microscopic images of white blood cells (WBCs) from Kaggle [21], which are intended to facilitate the design of automatised diagnostic processes in hematological diagnosis. The images are classified into four WBC subtypes: Eosinophil, Lymphocyte, Monocyte, and Neutrophil, which are significant in clinical diagnostics and diagnosis of diseases. The source data set had 410 manually tagged cell imageries with respective subtype annotations and annotations of bounding box information. To overcome the issue of class imbalance and provide more samples needed to train DL, we extensively augmented the data, so that now 12,500 high-quality images are available, which are distributed over the four cell classes. The same augmented samples were duly organized in special folders based on their cell type.

To clarify the dataset characteristics in accordance with the workflow requirements, the input data for this study consists of RGB microscopic images of white blood cells obtained from the Kaggle Blood Cell Images dataset, which belongs to one of four leukocyte subtypes (Eosinophil, Lymphocyte, Monocyte, and Neutrophil), which are used as the classification categories in the proposed framework. After augmentation, a total of 12,444 images were available for model development, of which 9,955 images were used for training, and 2,489 images were reserved for validation. Furthermore, this work does not rely on handcrafted tabular attributes; instead, the proposed CNN–ViT model learns feature representations directly from the raw pixel values and spatial morphological patterns of the input images in an end-to-end manner. The explicit breakdown of the number of images that belong to each of the four white blood cell classes into the training and validation sets can be seen in Table 2, which proves that the dataset is not biased towards any of the classes.

Table 2: Class distribution of blood cell images in the training and validation sets.

Class	Training Samples	Validation Samples
Eosinophil	2,452	668
Lymphocyte	2,505	598
Monocyte	2,523	575
Neutrophil	2,475	648
Total	**9,955**	**2,489**

Data Preprocessing

Before the microscopic images of blood cells are fed into the DL models, several preprocessing steps are performed that should improve model generalization and stabilize training.

Data Augmentation: To improve generalization and robustness, we perform both geometric and photometric augmentations.

- Geometric transformations: These modify the spatial domain of the image while preserving semantic content. Examples include:

$$X_g=f_g(X) \quad (1)$$

- where f*g* denotes a transformation such as random resized cropping, horizontal flipping, affine translations, or rotations within ±10◦.

- Photometric transformations: These alter pixel intensities to mimic real-world variability in slide preparation and staining:

$$X\_p=f\_p\ (X\_g\ ) \tag{2}$$

- where f*p* includes random brightness, contrast, and saturation noise. Formally, the augmentation pipeline can be expressed as a composition:

$$X\tilde{}=f\_p\circ f\_g\ (X) \tag{3}$$

Normalization: Deep models are sensitive to input data distributions. Hence, pixel values are normalized channel-wise using ImageNet statistics:

$$X' = \frac{\tilde{X}-\mu}{\sigma}, \qquad \mu = [0.485,0.456,0.406], \quad \sigma = [0.229,0.224,0.225] \tag{4}$$

Normalization rescales each channel such that:

$$\mathbb{E}[X'_c] \approx 0, \qquad \mathrm{Var}[X'_c] \approx 1, \qquad \forall c \in \{R, G, B\} \tag{5}$$

ensuring stable gradient flow during training.

The normalized tensor X' serves as the standardized input to the CNN feature extractor:

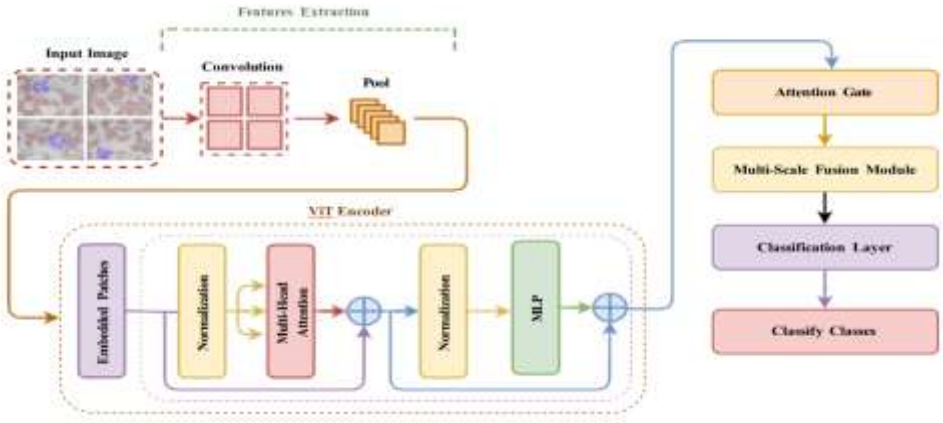
$$X\hat{}'\in R\hat{}(B\times 3\times H\times W), \qquad H=W=224 \tag{6}$$

This guarantees consistency across the dataset, mitigates covariate shift, and aligns the input distribution with that used to pretrain the CNN backbone on ImageNet.

CNN-ViT Model

The proposed hybrid model aims to integrate the inductive biases of CNNs and the global capability of self-attention of ViT. Although CNNs have proven to be very useful in extracting local morphological features (edges, textures, and nuclear boundaries), they form intrinsic limitations in capturing long-range structures. ViTs can be contrasted with tokenising image patches and using self-attention to extract long-range and short-range contextual relationships. In our model, these paradigms are incorporated into a unified framework, which allows for actually learning both local and spatial dependencies simultaneously, at a fine-grained level. To enhance the discriminative ability of the extracted features, we add an Attention Gate (AG) to highlight meaningful features and a Multi-Scale Fusion Module (MSFM) to combine features at different receptive fields. To get a clear understanding of the computing process, the proposed algorithm 1 is outlined to provide a clear picture of the functioning of the hybrid pipeline.

Figure 2: Proposed CNN-ViT Hybrid Model



Algorithm 1 Unified Blood-Cell Classification (CNN-ViT Hyrbid Model)

Require: Roots TRAIN, TEST; classes C; batch B=16; split α=0.2; epochs E=40; patience P =10; lr η=10−4
Ensure: Best model f, checkpoint, validation metrics DATA
1: for $c \in$ C do
2: $Vc \leftarrow$ files in TRAIN/c ∪ TEST/c
3: end foUr
4: Q ← $\{(x, c) \mid x \in Vc\}$; for $(x, \cdot) \in$ Q set $h(x) \leftarrow$ MD5(bytes(x))
5: Q′ ← drop missing $h(x)$; drop duplicates by $h(x)$
6: (Qtr, Qva) ← GroupShuffleSplit(Q′, groups = $h(x)$, test_size = α)
TRANSFORMS/LOADERS
7: ftr : resize 256 → rand-crop 224 → flip/rot/affine/jitter → tensor → normalize
8: fva : resize 224 → tensor → normalize
9: LDtr ← DataLoader((Qtr, ftr), B, shuffle); LDva ← DataLoader((Qva, fva), B)
MODEL/OPTIM
10: f ← EfficientNet-B0 (ImageNet); freeze early backbone
11: Replace head: Dropout → Linear–ReLU–BN (512) → Dropout → Linear–ReLU–BN (256) → Dropout → Linear(|C|)
12: G ← CrossEntropy; pt ← AdamW(η, wd=10−4); Sch ← ReduceLROnPlateau(0.5, 3)
13: $bvl \leftarrow \infty$; $o\_imp \leftarrow 0$
TRAIN
14: for e = 1 to E do
15: Train: for $(X, \ ) \in$ LDtr: $\hat{Y} \leftarrow f(X)$; $\ell \leftarrow$ G($\hat{Y}$, Y); backprop; clip $\|\nabla\| \le 1$; step
16: Validate: compute mean ValLoss, ValAcc on LDva; Sch.step(ValLoss)
17: if ValLoss < bvl then
18: save checkpoint; $bvl \leftarrow$ ValLoss; $no\_imp \leftarrow 0$
19: else
20: $no\_imp \leftarrow no\_imp + 1$
21: end if
22: if $no\_imp \ge P$ then
23: break
24: end if
25: end for
EVALUATE
26: Load best checkpoint → f; on LDva compute Accuracy, macro Precision/Recall/F1, classification report, confusion matrix
27: return f, checkpoint, metrics

Input Representation

Each microscopic blood smear image is represented as a three-channel RGB tensor of fixed resolution:

$$\mathrm{X} \in \mathbb{R}^{\mathrm{B}\times 3\times \mathrm{H}\times \mathrm{W}}, \qquad \mathrm{H} = \mathrm{W} = 224 \tag{7}$$

where B denotes the batch size, 3 corresponds to the color channels (red, green, blue), and (H, W) are the spatial dimensions. Thus, for each mini-batch, the tensor contains B images. The augmented and normalized input tensor, X', can be used as the input to the next CNN feature extractor.

CNN Feature Extraction

The initial path of the hybrid CNN. This network has been trained on ImageNet and cut off its classifier. This CNN transforms the normalized input X' into high-level feature maps through successive convolutional, non-linear, and down-sampling operations.

Feature Representation: The output feature tensor is defined as:

$$\mathrm{F} = \mathrm{CNN}(\mathrm{X}'), \qquad \mathrm{F} \in \mathbb{R}^{\mathrm{B}\times \mathrm{C}\times \mathrm{H}'\times \mathrm{W}'} \tag{8}$$

where B is the batch size, C = 1280 channels, and $H' = W' = 7$ for 224 × 224 input images.

Convolutional Operation: Each spatial element (i, j) in channel c for batch b is computed as:

$$F_{b,c,i,j} = \phi\left(\sum_{c'}\sum_{u,v} K_{c,c',u,v}\, X'_{b,c',i+u,j+v} + b_c\right) \tag{9}$$

where: $K_{c,c',u,v}$ are learnable convolution kernels, b_c is the channel-specific bias term, $\phi(\cdot)$ is a non-linear activation function (e.g., ReLU or Swish). This can be compactly expressed as a discrete convolution:

$$F_c = \phi(X' * K_c + b_c) \tag{10}$$

Convolution acts as a local receptive field operator, enforcing translation invariance while preserving spatial locality.

Down-sampling: Pooling and strided convolutions progressively reduce spatial dimensions:

$$H', W' \ll H, W \tag{11}$$

concentrating representational power into fewer, semantically richer feature maps. The resulting tensor F serves as a compact, hierarchical encoding of cell morphology, highlighting discriminative features such as nuclear boundaries, cytoplasmic granularity, and inter-cellular variations.

Projection and Tokenization

While CNN features are effective at encoding localized structures, they must be transformed into a sequence representation compatible with the ViT. To this end, we first reduce the channel dimension from C to a smaller embedding width D using a 1 × 1 convolution:

$$Z = W_p * F + b_p, \qquad W_p \in \mathbb{R}^{D\times C\times 1\times 1}, \qquad Z \in \mathbb{R}^{B\times D\times H'\times W'} \tag{12}$$

Flattening the spatial dimensions produces a sequence of tokens of length $N = H'W'$:

$$T = \text{reshape}(Z) \in \mathbb{R}^{B\times N\times D}, \qquad N = 49 \tag{13}$$

A learnable classification token $\mathbf{z}cls$ is prepended to each sequence, and absolute positional embeddings P are added to preserve spatial ordering:

$$X_0 = [z_{cls}; T] + P \tag{14}$$

The resulting sequence X_0 serves as input to the Transformer encoder.

Channel Projection: A 1 × 1 convolution is applied to reduce the feature channel dimension from C to a smaller embedding width D, ensuring computational efficiency while preserving spatial information:

$$Z = W_p * F + b_p, \qquad W_p \in \mathbb{R}^{D\times C\times 1\times 1}, \qquad Z \in \mathbb{R}^{B\times D\times H'\times W'} \tag{15}$$

Tokenization: Flattening the spatial dimensions transforms the feature map into a sequence of tokens of length $N = H'W'$:

$$T = \text{reshape}(Z) \in \mathbb{R}^{B\times N\times D}, \qquad N = 49 \tag{16}$$

Each token corresponds to a spatial patch embedding of the input image, analogous to the patch embedding in standard ViTs.

Sequence Construction: To align with Transformer input requirements:

- A learnable classification token $\mathbf{z}_{cls}$ is prepended to the sequence, designed to aggregate global information.
- Absolute positional embeddings P are added to preserve the spatial relationships among tokens:

$$X_0 = [z_{cls}; T] + P \tag{17}$$

The resulting sequence $X_0 \in \mathbb{R}^{B\times(N+1)\times D}$ is used as input to the ViT encoder.

Vision Transformer Encoder

The ViT encoder is composed of L stacked layers (we use $L = 6$). Each layer integrates two core operations: Multi-Head Self-Attention (MHSA) and a position-wise feed-forward MLP, wrapped with residual connections and normalization.

Self-Attention Mechanism: For input $X \in \mathbb{R}^{B\times(N+1)\times D}$, queries ($Q$), keys ($K$), and values ($V$) are computed as:

$$\mathrm{Q} = \mathrm{XW_Q}, \qquad \mathrm{K} = \mathrm{XW_K}, \qquad \mathrm{V} = \mathrm{XW_V} \tag{18}$$

where $W_Q, W_K, W_V \in \mathbb{R}^{D\times D}$ are learnable matrices.
The scaled dot-product attention is then computed as:

$$\mathrm{Attention(Q,K,V)} = \mathrm{softmax}\left(\frac{\mathrm{QK^T}}{\sqrt{\mathrm{d_h}}}\right)\mathrm{V} \tag{19}$$

where d_h is the per-head embedding size.

Multi-Head Extension: Multiple attention heads capture relationships from different representational subspaces. Their outputs are concatenated and projected to form the layer's attention output.

Residual Learning: To ensure stable training and prevent vanishing gradients, residual connections are applied:

$$\tilde{\mathrm{X}} = \mathrm{X} + \mathrm{MHSA}\big(\mathrm{LN(X)}\big), \qquad \mathrm{X'} = \tilde{\mathrm{X}} + \mathrm{MLP}\left(\mathrm{LN}\big(\tilde{\mathrm{X}}\big)\right) \tag{20}$$

After L layers, the final sequence is:

$$\mathrm{X_L} \in \mathbb{R}^{\mathrm{B\times(N+1)\times D}} \tag{21}$$

Global Representation: The [CLS] token $X[:, 0, :]$ captures the global semantics of the input image and is used for downstream classification.

Attention Gate

Although CNNs capture strong local patterns and Transformers capture global context, combining them effectively is essential. To this end, we introduce an Attention Gate (AG) that adaptively balances contributions from CNN features f_c and Transformer features f_t.

Gating Mechanism: The gating coefficient is computed as:

$$\alpha = \sigma(\mathrm{W_f f_c} + \mathrm{W_t f_t} + \mathrm{b}) \tag{22}$$

where $\sigma(\cdot)$ is the sigmoid activation, and W_f, W_t are learnable weights.
Feature Fusion: The final fused representation is:

$$\mathrm{f'} = \alpha \odot \mathrm{f_c} + (1-\alpha) \odot \mathrm{f_t} \tag{23}$$

where $\odot$ denotes element-wise multiplication.

Interpretation: $\alpha \approx 1$: the model emphasizes CNN-driven local morphology. $\alpha \approx 0$: the model emphasizes Transformer-driven global context. This mechanism enables adaptive weighting depending on the task and input image characteristics.

Multi-Scale Fusion Module

To enhance representation power, we integrate a Multi-Scale Fusion Module (MSFM) that captures multi-resolution features.

Fusion Operation: We define convolutional operators $\phi(\cdot)$ with different kernel sizes or dilation rates, producing features across scales:

$$F_{fusion} = \text{Concat}\left(\phi_1(f'), \phi_2(f'), \ldots, \phi_m(f')\right) \quad (24)$$

Rationale: Small kernels capture fine-grained nuclear structures. Larger kernels/dilated convolutions capture coarse cytoplasmic and contextual features. Concatenation aggregates these features, improving discriminability.

Classification Layer

The classification head leverages the [CLS] token representation from the Transformer output to predict cell categories.

Logit Computation:

$$h = X_L[:,0,:], \qquad o = W_c h + b_c \quad (25)$$

Softmax Probabilities:

$$\hat{y} = \text{softmax}(o) \quad (26)$$

where $\hat{y} \in \mathbb{R}^K$ represents probabilities over $K = 4$ blood cell categories.

Training Objective

The model is trained end-to-end using the categorical cross-entropy loss:

$$\mathcal{L} = -\frac{1}{B}\sum_{i=1}^{B}\sum_{k=1}^{K} \mathbf{1}\,[y_i = k]\log\hat{y}_{i,k} \quad (27)$$

Regularization: - Weight decay prevents overfitting by penalizing large weights. - Dropout is applied in the Transformer and fusion layers to encourage robustness.

We combine explanation techniques based on CNN and Transformer to increase interpretability and confidence. The most discriminative local areas in microscopic images are highlighted by Grad-CAM, while global token-level dependencies across Transformer layers are captured by attention rollout. To ensure highly convincing validation, GroupShuffleSplit was used to divide the dataset into training (80 percent) and validation (20 percent) sets, and duplicate hashes were grouped together to avoid data leakage between splits. During training, data augmentation techniques were employed, such as random cropping, flipping, rotation, affine transformation, and color jitter, as a means to increase the level of generalization capacity. Table Table 3 presents the global training configuration which employed in the experiments.
Baseline Models

This study uses four of the latest CNNs, i.e., ResNet-50, DenseNet-121, EfficientNetB0, and VGG-16, as the backbones of microscopic leukocyte image classification. All of these networks exemplify distinct design principles of DL, including residual connections and dense feature reuse, compound scale, and classical hierarchical convolutions, and would be suitable in modeling the subtle morphological changes that are critical in diagnosing ALL.

Result and Discussion

In this section, the results of the proposed model to recognize ALL on the basis of microscopic images are presented and discussed. To assess the effectiveness of the model, a number of metrics are utilized, such as accuracy, precision, recall, F1 score, and the visual evaluation using Grad-CAM in terms of interpretability.

Hardware and Software Configuration

All the experiments were conducted on a powerful computer to maintain the steady training of the proposed deep learning model. The system configuration was equipped with an NVIDIA GeForce RTX 3080 GPU with 10 GB VRAM, which significantly sped up the training process. A high-performance Intel Core i7-9700K CPU was used for efficient preprocessing and testing. The system was equipped with 32 GB DDR4 RAM to handle the large dataset.

The proposed deep learning model was implemented using the PyTorch framework because of its flexibility and support for efficient neural network training. For the purpose of result visualization and assistance in deployment, TensorFlow was used. PIL was used for image preprocessing, whereas NumPy was used for efficient numerical calculations.

Table 3: Global training configuration used in the experiments

Hyperparameter	Value
Input size / Normalization	224×224 RGB; ImageNet mean/std normalization
Batch size	16
Loss function	Cross-Entropy Loss
Optimizer	AdamW (weight decay = 1e-4)
Learning rate	1×10^{-4}
Learning rate scheduler	ReduceLROnPlateau (val loss, factor=0.5, patience=3)
Epochs	40
Gradient clipping	$\|\nabla\theta\|_2 \leq 1.0$
Dropout	0.4
Data Split	80/20 (GroupShuffleSplit by file hash; duplicates removed)
Augmentations (train)	Resize(256), RandomResizedCrop(224, 0.8–1.0), HorizontalFlip(0.5), Rotation(±10°), Affine(±5% translate), ColorJitter(±0.1)
Validation preprocessing	Resize(224), ToTensor, Normalize(ImageNet stats)
Checkpointing	Save the best model by the lowest validation loss

Training and Validation Accuracy Comparison

Table 4: Training and Validation Accuracy for Each Model

Model	Train Accuracy (%)	Validation Accuracy (%)
Hybrid CNN-ViT (Proposed)	97.35	99.60
ViT	91.88	90.66
Grad-CAM	97.31	88.38
ResNet-50	99.85	99.38
VGG-16	96.56	98.15
DenseNet-121	61.22	85.49
EfficientNetB0	96.03	98.15

This section gives a clear picture of the training and validation accuracy of each of the models and shows details of how each model performed in the diagnosis of ALL. Table 4 reveals the Hybrid CNN-ViT (Proposed) model to be the most ideal, with an improvement of a significant margin over the rest of the baseline models.

As seen in TableTable 4, the CNN-ViT hybrid stands out with the highest overall performance, with 97.35% training accuracy and the highest validation accuracy at 99.60%. This shows the good generalization of the model, effectively utilizing the power of CNNs with the Vision Transformer. On the other hand, the ViT alone trails behind with 91.88% training accuracy and 90.66% validation accuracy.

The Grad-CAM version of the model also shows a similar training accuracy of 97.31%, but the validation accuracy declines to 88.38%, showing the possible negative impacts of the constraints. The ResNet-50 shows a good performance with 99.85% training accuracy and 99.38% validation accuracy, but still trails behind the CNN-ViT hybrid.

The VGG-16 and EfficientNetB0 show a mediocre performance with training accuracy in the mid-90s and validation accuracy at 98.15% and 96.56%, respectively. However, both of these are still outperformed by the CNN-ViT hybrid and the ResNet-50. The DenseNet-121 shows a poor performance with 61.22% training accuracy and 85.49% validation accuracy, showing a volatile performance and the need to further optimize the model.

Validation Performance Metrics

Table 5 : Validation Performance Metrics for Each Model

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Hybrid CNN-ViT (Proposed)	99.60	99.62	99.61	99.62
ViT	90.66	90.61	90.66	90.59
Grad-CAM	88.38	90.84	88.38	88.65
ResNet-50	99.38	99.39	99.38	99.38
VGG-16	98.15	98.16	98.14	98.15
DenseNet-121	84.49	76.50	85.49	80.13
EfficientNetB0	98.15	98.15	98.15	98.16

This section evaluates how all models perform in terms of validity using accuracy, precision, recall, and F1-score. The idea is to show how effective all these models are in reducing both false positives and false negatives, a critical aspect for ALL classification. As shown in the Table 5 for validation metrics, the Hybrid CNN-ViT performs better across all metrics: 99.60% accuracy, 99.62% precision, 99.61% recall, and 99.62% F1-score.

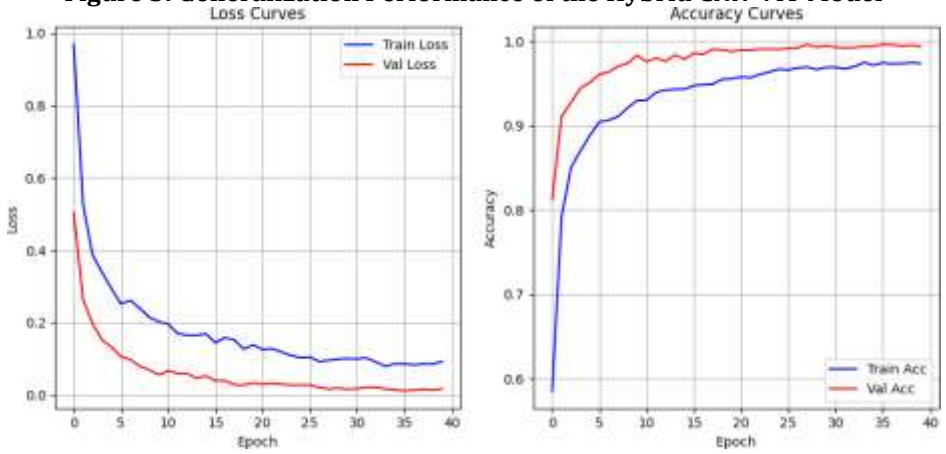
The ResNet-50 is also performing exceptionally well by achieving 99.38% accuracy, 99.39% precision, 99.38% recall, and 99.38% F1-score. However, it is not performing as well as the Hybrid CNN-ViT. The VGG-16 and EfficientNetB0 are also performing reasonably well by achieving 98.15% accuracy, 98.16% precision, 98.14% recall, and 98.15% F1-score.

The Grad-CAM enhanced hybrid model is performing poorly by achieving 88.38% accuracy, 90.84% precision, 88.38% recall, and 88.65% F1-score. This indicates that there is a possibility that incorporating interpretability into a model can be a barrier to a model's ability to generalize. The DenseNet-121 is performing worse by achieving 84.49% accuracy, 76.50% precision, 85.49% recall, and 80.13% F1-score.
Overall, the Hybrid CNN-ViT performs better across all areas than all other models.
Generalization Analysis for the Hybrid CNN-ViT Model

The detailed description of the generalization performance of the hybrid CNN-ViT (Proposed) model is explained in this section, as shown in Figure 3. These graphical representations facilitate analysis of the ability of the model to generalize from training and the training set in training.

Figure 3 demonstrates the manner in which the Hybrid CNN-ViT model converges. In the loss curves (left), the loss curves converge rapidly and consistently, with the training loss plummeting early on and continuing to decline smoothly throughout the rest of the iterations. The validation loss follows the same trend, declining smoothly throughout the iterations. By the end, the loss curves are very close to zero, indicating proper optimization and minimal residual error. In the accuracy curves (right), training accuracy rises rapidly and then remains high, exceeding 99% in later iterations. The validation accuracy also increases and then levels off near 100%. The curves demonstrate the excellent generalization and robustness of the Hybrid CNN-ViT framework when faced with new, unseen data.

Figure 3: Generalization Performance of the Hybrid CNN-ViT Model



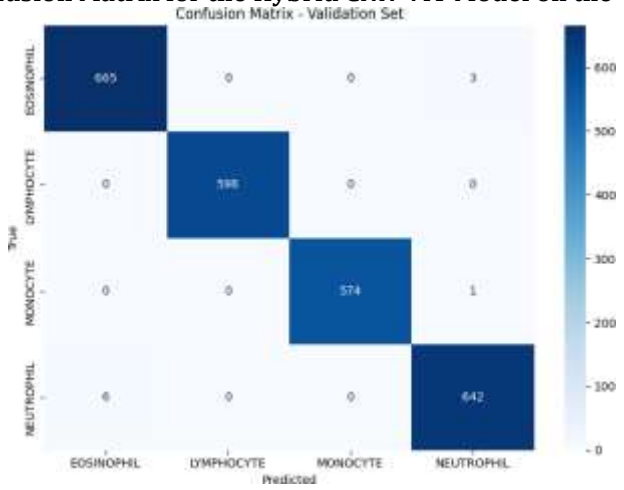
Error Analysis

Figure 4 provides worthwhile information on the distribution of errors in the Hybrid CNN-ViT (proposed) model when used to perform against the validation set.

In the Eosinophil category, the model accurately predicted 665 instances, and only 3 instances were falsely predicted as Neutrophil. This gives a figure of an accuracy of 99.55% for Eosinophil, thus showing a high accuracy and only slight confusion with the Neutrophil class. In the case of Lymphocyte, the model correctly classified all 598 instances without misclassification. This translates to a 100% accuracy, showing that the model has the capability of accurately classifying a Lymphocyte with an accuracy of 100%. Within the Monocyte category, it performed equally well, classifying 574 of the instances correctly and making a single mistake as a Neutrophil. This brings about a precision of 99.83% in the Monocyte, thus gaining more efficiency in the classification reliability of the model. In the case of Neutrophil, the model perfectly predicted it correctly in 642 instances, whereas 6 instances were incorrectly predicted as Eosinophil. This contributed to the highest percentage accuracy of 99.08% with regard to Neutrophil, which, though it is still high, reflects some error on Neutrophil with Eosinophil, even though to a small extent.

On the whole, the Hybrid CNN-ViT (Proposed) model exhibits a high level of generalization and a good performance, making the resulting classification accuracies very high on all four classes.

Figure 4: Confusion Matrix for the Hybrid CNN-ViT Model on the Validation Set



Comparative Analysis and Discussions

Table 6 Comparative Analysis of the Proposed Hybrid Model with Existing Models for ALL Classification

Reference	Dataset	Model	Accuracy (%)
[22]	ALL-IDB2 dataset	Random Threshold Classifier	98.3
[13]	Acute Lymphoblastic Leukemia (ALL) image dataset	DenseNet121	99
[23]	ALL-IDB dataset	Naïve Bayes	96.15
[24]	Microscopic images of blood cells	CNN	97
[17]	Blood Cell Images from Kaggle	DDRNet	91.98
[25]	Microscopic blood cell images	ANFIS	97.14
Ours	**Blood Cell Images from Kaggle**	**CNN-ViT**	**99.60**

In order to ensure the clarity of the comparative study, it should be noted here that the findings reported in Table 6 are collected from the previously published studies and reported here for the purpose of providing a broader performance context for the ALL classification. It should be noted that the proposed model in the study achieves 99.60% accuracy on the Blood Cell Images dataset from Kaggle.

In comparison with Muduli et al. [13], who applied DenseNet121 and reported an accuracy of 99%, our model is 0.60% better. Such a gain indicates the superiority of integrating CNNs and ViT since our method can take advantage of both local and global characteristics in data. Also, Curtidor et al. [22] used Random Threshold Classifier on the ALL-IDB2 dataset with an accuracy of 98.3%. This is bettered by 1.3% in our proposed hybrid model, which proves the effectiveness of DL architecture over the traditional ML classifier, especially hybrid models, like our CNN-ViT model. Similarly, DDRNet used by Jawahar et al. [17] resulted in an accuracy of 91.98%,

which is lower than that of our model by 7.62%, again proving the efficiency of the CNN-ViT in the specified form of dynamics identification.

Our model performed better in terms of accuracy than El et al. [23], who obtained 96.15% through the use of Naive Bayes, by 3.45%, and hence, shows that in medical image analysis, DL models are preferable over more traditional classification methods. What is more, our approach yielded a 2.60% increase in accuracy over the results obtained by Moradiamin et al. [24], who used a standard CNN. Our model performed better than ANFIS proposed by Rejula et al. [25], which reported an accuracy of 97.14%, and our model lifted this accuracy to 99.60%. This again reinstates the benefits of the application of DL models in complex classifications, particularly in medical-related areas.

Finally, the hybrid CNN-ViT (Proposed) model shows the highest accuracy of 99.60% and outperforms existing models significantly, where the performance improvement is between 0.60% and 15.88%. These findings confirm the benefits of hybridizing CNN-ViT networks to classify medical images and present a state-of-the-art solution to the classification of ALL. Combining the advantages of CNNs in local feature extraction and ViTs in modeling global dependencies, the model may not only obtain high accuracy but also improve the capacity of generalization.

Through the efficient use of CNNs for local feature extraction and ViTs for capturing global dependencies, the presented Hybrid CNN-ViT model makes it possible to accurately recognize intricate patterns in medical images. Grad-CAM visualisations improve performance and interpretability by confirming that the model concentrates on biologically significant areas, including cell nuclei. When it comes to tasks where minute distinctions between classes are crucial, this combination offers a richer feature representation than single-architecture models.

Minor overfitting, however, is indicated by a small difference between training and validation loss, which may be reduced by employing strategies like data augmentation, dropout, or early stopping. Furthermore, adoption in resource-constrained or real-time clinical situations may be limited due to the increased processing demands of the ViT component. In spite of these factors, the model shows excellent interpretability, accuracy, and dependability for classifying medical images.

Conclusion

The study moves forward with automated leukemia diagnosis by proposing a hybrid CNN-ViT model that has the potential to reach a satisfactory balance between convolutional feature representations and transformer contextual modeling. The model has been rigorously experimented with and showed better performance on general tasks compared with traditional CNNs and standalone ViTs, reaching 99.60% validation accuracy and showing good interpretability in Grad-CAM heat maps. The paper makes a contribution to the existing knowledge base, which highlights the overall claim that the hybrid structures have the capability to combine local texture sensitivity and global relational reasoning, resulting in more consistent and clinically definite predictions. Despite its strong performance, the model was evaluated on a single public dataset and requires high computational resources, limiting its applicability in diverse clinical settings and real-time use.

Three research directions are recommended as promising avenues of future work: (i) external validation in a multi-center, large-scale clinical data to ensure generalizability; (ii) adding domain adaptation and stain normalization to address distributional shifts; and (iii) assessing model compression, knowledge distillation, and compact versions of transformers as practical deployment options.

References

[1] L. Pagliaro, S.-J. Chen, D. Herranz, C. Mecucci, C. J. Harrison, C. G. Mullighan, M. Zhang, Z. Chen, N. Boissel, S. S. Winter, et al., Acute lymphoblastic leukaemia, Nature Reviews Disease Primers 10 (1) (2024) 41.

[2] S. Walsh, M. Mulraney, M. McCarthy, C. R. De Luca, Fatigue in children who have recently completed treatment for acute lymphoblastic leukemia: a longitudinal study, Health and Quality of Life Outcomes 22 (1) (2024) 27.

[3] Kaczmarska, J. Derebas, M. Pinkosz, M. Niedźwiecki, M. Lejman, The landscape of secondary genetic rearrangements in pediatric patients with b-cell acute lymphoblastic leukemia with t (12; 21), Cells 12 (3) (2023) 357.

[4] L. Ding, J. E. Szymczak, E. Evans, E. Canepa, A. E. Martin, F. Contractor, R. Aplenc, G. Joseph, L. E. Winestone, Factors that contribute to disparities in time to acute leukemia diagnosis in young people: an in depth qualitative interview study, BMC cancer 22 (1) (2022) 531.

[5] T. Tran, T. Zeevi, S. Payabvash, Strategies to improve the robustness and generalizability of deep learning segmentation and classification in neuroimaging, BioMedInformatics 5 (2) (2025) 20.

[6] J. Wong, Neuroendocrine tumors, pathology of the vagina, in: Gynecologic Pathology, Springer, 2024, pp. 344–345.
[7] H. Taherdoost, Deep learning and neural networks: Decision-making implications, Symmetry 15 (9) (2023) 1723.
[8] J. Meng, M. Wu, F. Shi, Y. Xie, H. Wang, Y. Guo, Medical laboratory data-based models: opportunities, obstacles, and solutions, Journal of Translational Medicine 23 (1) (2025) 823.
[9] H. Xu, Q. Xu, F. Cong, J. Kang, C. Han, Z. Liu, A. Madabhushi, C. Lu, Vision transformers for computational histopathology, IEEE Reviews in Biomedical Engineering 17 (2023) 63–79.
[10] Binder, M. Bockmayr, M. Hägele, S. Wienert, D. Heim, K. Hellweg, M. Ishii, A. Stenzinger, A. Hocke, C. Denkert, et al., Morphological and molecular breast cancer profiling through explainable machine learning, Nature Machine Intelligence 3 (4) (2021) 355–366.
[11] K. L. Narayanan, R. S. Krishnan, Y. H. Robinson, S. Vimal, T. A. Rashid, C. Kaushal, M. M. Hassan, Enhancing acute leukemia classification through hybrid fuzzy c means and random forest methods, Measurement: Sensors 39 (2025) 101876.
[12] H. M. Rai, B. Omkar Lakshmi Jagan, N. T. Rao, T. K. Mohammed, N. Agarwal, H. A. Abdallah, S. Agarwal, Deep learning for leukemia classification: Performance analysis and challenges across multiple architectures, Fractal and Fractional 9 (6) (2025) 337.
[13] D. Muduli, S. Parija, S. Kumari, A. Hassan, H. S. Jangwan, A. T. Zamani, S. M. Gouse, B. Majhi, N. Parveen, Deep learning-based detection and classification of acute lymphoblastic leukemia with explainable ai techniques, Array 26 (2025) 100397.
[14] M. Shahbaz, A. Basharat, R. Yasmin, N. Ahmad, R. Abbasi, S. Hussain, Leveraging explainable ai for early risk prediction and type classification for leukemia: Insights using clinical data from pakistan, IEEE Journal of Biomedical and Health Informatics (2025).
[15] H. Park, S. Miyano, Network-based multi-class classifier to identify optimized gene networks for acute leukemia cell line classification, PLoS One 20 (5) (2025) e0321549.
[16] T. Mustaqim, C. Fatichah, N. Suciati, T. Obi, J.-S. Lee, Focusaugmix: A data augmentation method for enhancing acute lymphoblastic leukemia classification, Intelligent Systems with Applications 26 (2025) 200512.
[17] M. Jawahar, L. J. Anbarasi, S. Narayanan, A. H. Gandomi, An attention-based deep learning for acute lymphoblastic leukemia classification, Scientific Reports 14 (1) (2024) 17447.
[18] P. Bose, S. Bandyopadhyay, A comprehensive assessment and classification of acute lymphocytic leukemia, Mathematical and Computational Applications 29 (3) (2024) 45.
[19] T. O. Asar, M. Ragab, Leukemia detection and classification using computer-aided diagnosis system with falcon optimization algorithm and deep learning, Scientific Reports 14 (1) (2024) 21755.
[20] F.-M. Cheng, S.-C. Lo, C.-C. Lin, W.-J. Lo, S.-Y. Chien, T.-H. Sun, K.-C. Hsu, Deep learning assists in acute leukemia detection and cell classification via flow cytometry using the acute leukemia orientation tube, Scientific Reports 14 (1) (2024) 8350.
[21] Blood cell images, [Online; accessed 2025-08-25].
URL https://www.kaggle.com/datasets/paultimothymooney/blood-cells
[22] Curtidor, E. Kussul, T. Baydyk, M. Mammadova, Analysis and automated classification of images of blood cells to diagnose acute lymphoblastic leukemia, EUREKA: Physics and Engineering (5) (2023) 177–190.
[23] E. M. El Houby, Acute lymphoblastic leukemia diagnosis using machine learning techniques based on selected features, Scientific Reports 15 (1) (2025) 28056.
[24] M. MoradiAmin, M. Yousefpour, N. Samadzadehaghdam, L. Ghahari, M. Ghorbani, M. Mafi, Automatic classification of acute lymphoblastic leukemia cells and lymphocyte subtypes based on a novel convolutional neural network, Microscopy Research and Technique 87 (7) (2024) 1615–1626.
[25] M. A. Rejula, S. Amutha, G. Shilpa, Classification of acute lymphoblastic leukemia using improved anfis, Multimedia Tools and Applications 82 (23) (2023) 35475–35491.