

Advanced Blood Cell Analysis for Cancer Detection Using Deep Learning

Sanjit Kumar Dash ¹, Dibya Ranjan Sahoo ¹, Likuna Pradhan ¹, Mohammed Altaf Ahmed ^{2,*},
Qutubuddin Mohammed ³, and Sultan Alqahtani ²

¹ Department of Information Technology, Odisha University of Technology and Research, Bhubaneswar, Odisha, India

² Department of Computer Engineering, College of Computer Engineering and Sciences,
Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia

³ Department of Electrical Engineering, College of Engineering Wadi Addawasir,
Prince Sattam bin Abdulaziz University, Wadi Addawasir, Saudi Arabia

Email: skdash@outr.ac.in (S.K.D.); sdibya865@gmail.com (D.R.S.); likunapradhan35768@gmail.com (L.P.);
m.alfat@psau.edu.sa (M.A.A.); q.mohammed@psau.edu.sa (Q.M.); su.alqahtani@psau.edu.sa (S.A.)

*Corresponding author

Abstract—Blood cancer, resulting from the uncontrolled growth of white blood cells (leukocytes), is still an important problem in medical diagnosis. In the past, cancerous blood cells were only detected through a manual examination by trained professionals, which took up considerable time and was prone to mistakes. Deep learning algorithms are the new development towards the automation and advancement in the diagnosis of blood cancer. This article talks about how deep learning techniques are making diagnosis of blood cancer more effective and accurate. Deep learning models, such as Convolutional Neural Networks (CNNs), have progressed this further by automatically identifying key patterns in microscopic blood images and thereby making early cancer prediction more precise. The research aims to find the best architectural design which will produce the highest diagnostic accuracy through its assessment of three separate deep learning systems that include Swin Transformer and ConvNeXt and EfficientNetV2. The models achieved high performance through their combination with efficient pre-processing and image enhancement techniques which reduced misclassification rates and improved early cancer detection. The tested models successfully identified the samples according to their actual class distribution which showed slight deviation from the original classification system. The 3 models achieved overall accuracy rates of 99.84% for ConvNeXt, 99.69% for Swin Transformer, and 99.38% for EfficientNetV2, demonstrating that the modernized CNN architecture (ConvNeXt) slightly outperformed the transformer-based model (Swin Transformer) in this specific evaluation. Medical procedures now use deep learning technology to provide quick and accurate patient diagnoses which leads to improved treatment choices and better patient outcomes.

Keywords—blood cancer detection, leukemia classification, deep learning, Convolutional Neural Networks (CNNs), swin transformer, feature extraction, medical image analysis

I. INTRODUCTION

Blood cancer, especially leukemia, serves as an urgent medical emergency which disrupts the body's ability to create and utilize blood cells and needs immediate correct diagnosis for successful treatment [1]. The existing diagnostic methods depend on experts who need to spend substantial time to manually check blood smear images which results in various mistakes [2]. The latest developments in artificial intelligence now allow for machines to analyze medical images automatically which improves both diagnostic precision and operational speed [3]. Researchers have achieved exceptional results in disease classification through their use of deep learning methods which include convolutional neural networks to identify intricate patterns within microscopic blood images [4].

Multiple research studies have developed deep learning-based systems which successfully detected and classified leukemia through testing on various datasets [5]. The hybrid models which combine segmentation and classification methods achieve better accuracy because they focus on key elements of blood smear images [6]. Transfer learning approaches which use pretrained EfficientNet architecture have shown strong results with small medical data collections [7]. The first studies which tested pretrained convolutional neural networks showed that these models could accurately classify different leukemia subtypes [8].

Recent research studies transformer-based architectural development because it enables researchers to investigate how these systems process global contextual information in their medical image processing work [9]. The application of deep learning models to bone marrow and peripheral blood images has increased the accuracy of leukemia diagnosis through their use in medical testing [10]. The common use of transfer learning together

with fine-tuning methods has improved both model performance and classification accuracy [11]. Deep learning diagnostic systems show high accuracy when they detect acute lymphoblastic leukemia [12]. Automatic diagnostic tools which artificial intelligence develops for medical imaging purposes have become critical for oncology work according to recent AI developments in medical imaging research [13]. The development of transformer-based methods enables researchers to create medical imaging systems which handle long-distance relationships while producing better feature representation results for complex medical images [14].

The study aims to create a reliable and efficient deep learning system which will automatically detect blood cancer through the analysis of microscopic blood smear images. The proposed approach aims to leverage advanced deep learning architectures to improve classification accuracy while addressing challenges such as class imbalance and feature extraction. The study develops a dependable system which produces understandable results to assist medical professionals with their early diagnosis tasks and their decision-making needs.

II. LITERATURE SURVEY

Deep learning research has developed methods which achieve better accuracy rates and faster detection times for leukemia through the examination of microscopic blood smear images. The research studies show that Convolutional Neural Network (CNN) systems which enable automated classification achieve better results than conventional classification systems. Ahmed *et al.* [1] developed a deep learning system which detects leukemia through blood smear image analysis and achieved successful performance results. Syed *et al.* [2] developed a hierarchical deep learning system which identifies different types of leukemia through whole-slide image analysis with high diagnostic accuracy. Kizi *et al.* [3] showed that deep learning methods which use CNN systems succeed at classifying leukemia cases.

Researchers have studied Capsule Networks and hybrid systems to develop new design solutions which enhance their ability to extract features and improve their classification system performance. Dhalla *et al.* [4] developed LeukoCapsNet, a modified CapsNet model that efficiently captures spatial relationships in blood smear images. Alzahrani *et al.* [5] proposed a segmentation and classification framework that improves diagnostic accuracy by combining region-based analysis with deep learning. Akter *et al.* [6] introduced a hybrid feature fusion model that uses gated recurrent units and uncertainty quantification to enhance leukemia detection accuracy. A. El-Aziz *et al.* [7] proved that EfficientNet-based models achieved impressive results through their implementation of EfficientNetV2-S cross-validation method which produced trustworthy leukemia prediction results. The research conducted by Shafique and Tehsin [8] demonstrated that pretrained CNNs successfully classified different leukemia subtypes which they studied in their previous research work.

Recent studies investigate transformer-based architectures because these systems can extract complete contextual details from medical imaging data. The TopFormer transformer-based segmentation model developed by Zhang *et al.* [9] has become a foundational technology for multiple medical imaging systems. Elsayed *et al.* [10] used deep learning methods to analyze bone marrow images which resulted in better leukemia diagnosis and classification results. The field of transfer learning has gained popularity after Mollick *et al.* [11] showed that pretrained models could enhance classification results, and Saeed *et al.* [12] developed a dependable deep learning diagnostic system which detects acute lymphoblastic leukemia. The research conducted by Granata *et al.* [13] shows that medical imaging and artificial intelligence technologies have developed to the point where oncologists increasingly use AI-based diagnostic systems. The research conducted by He *et al.* [14] demonstrated that transformer architectures effectively capture global dependencies which exist within medical image analysis, and Konkyana *et al.* [15] studied how transformer-based neural networks function as automated systems for cancer classification.

Researchers developed various hybrid machine learning methods to detect leukemia alongside deep learning models. Dharani *et al.* [16] introduced an improved machine learning algorithm for blood cancer detection, while Devi *et al.* [17] developed a duplet-convolutional neural network to segment and classify white blood cancer cells in their research work. Shehta *et al.* [18] created a deep learning-based blood cancer prediction model which reached high accuracy for blood cancer classification. Awad and Aly [19] conducted comparative studies to test various deep learning models which included You Only Look Once (YOLO) and ResNet architectures for diagnosing leukemia. Ahad *et al.* [20] developed a CNN-based model which used ensemble-based methods to enhance system reliability and operational capabilities. Atteia *et al.* [21] introduced a Bayesian-optimized CNN model which achieved precise leukemia detection, and Sampathila *et al.* [22] created a specialized deep learning classifier for blood smear images. Hybrid approaches which combine deep learning with traditional machine learning methods have been studied through the ResNet-SVM model which Sulaiman *et al.* [23] developed.

The field of medical image analysis has received benefits from the developments in hybrid architectures which extend beyond research studies focused on leukemia. Dhiravidachelvi *et al.* [24] proposed a hybrid CNN-RNN model optimized using artificial hummingbird optimization for improved classification performance. The research of Dese *et al.* [25] established that traditional machine learning methods can successfully detect leukemia through accurate classification results. Awais *et al.* [26] built a decision support system which used hybrid feature extraction methods to assist with leukemia detection tasks. The scientific community has utilized publicly accessible datasets which include the Blood Cell Cancer Classification dataset that Informatica

Mdica released on Roboflow to train and test deep learning models throughout their benchmark evaluations [27].

III. CANCER DETECTION FRAMEWORK

The framework for blood cancer detection and classification, that combines Swin Transformer, Convolutional Next (ConvNeXt), and EfficientNet Version 2 (EfficientNetV2) to maximize segmentation, feature extraction, and classification accuracy. The pipeline starts with input images of blood cells, which are preprocessed and enhanced to guarantee data quality. This process involves resizing, normalizing, and excluding corrupted images to normalize the dataset for training the model. ConvNeXt is employed for hierarchical spatial feature extraction, learning fine-grained information from images of blood cells. At the same time, EfficientNetV2 emphasizes learning high-level abstract patterns to enhance classification ability. Swin Transformer, a vision transformer model, enhances the segmentation performance by capturing long-range dependencies and global contextual information. The illustration indicates

that after the deep learning models extract salient features, Explainable AI (XAI) modules, such as Gradient-Weighted Class Activation Mapping(Grad-CAM), are utilized to localize and highlight the cancerous and non-cancerous areas in the blood cell images.

Fig. 1 Illustrates, instead of fusing these architectures, each of the 3 models is independently trained and evaluated using transfer learning to determine the most effective approach. ConvNeXt is employed for hierarchical spatial feature extraction, learning fine-grained information from images of blood cells. At the same time, EfficientNetV2 emphasizes learning high-level abstract patterns to enhance classification ability. Swin Transformer, a vision transformer model, captures long-range dependencies and global contextual information. During the classification phase, the extracted features from each independent model are fed through their respective classifiers to classify the blood cell smear images. The framework determines which blood cancer detection model performs best as a single model through its comparison of CNN-based and Transformer-based architectural systems.

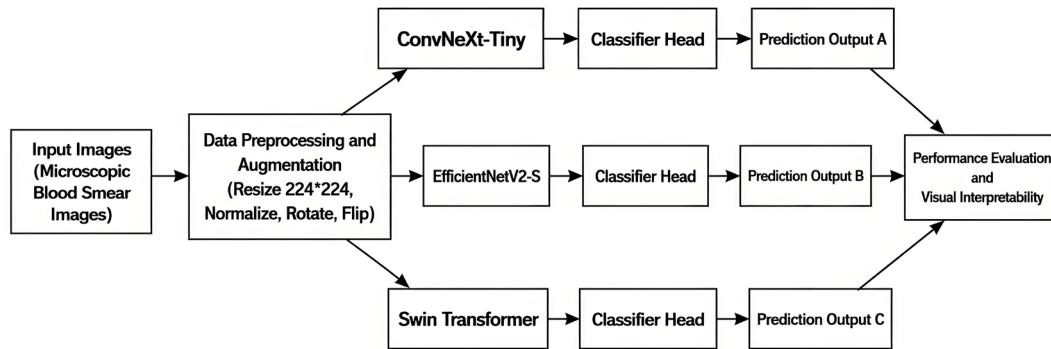


Fig. 1. Block diagram of proposed cancer detection framework.

IV. METHODOLOGY

The method uses a complete process which starts with image preprocessing and moves through deep learning feature extraction and individual model development and ends with classification to achieve maximum accuracy.

A. Dataset Description

The proposed model uses a dataset from Roboflow which was collected through the blood-cell-cancer-classification-fzkr (Version 1) project [27]. The collection contains 3208 high-resolution microscopic images which document Peripheral Blood Smears (PBS) used for blood cancer classification purposes. The dataset includes 5 separate classifications which are Benign, Pre-B [Malignant], Pre-B [Malignant] early, Pro-B [Malignant], and a rare “not cell/unlabeled” class. As we can see in Table I, the dataset was divided into

training and validation and testing sets which contained 2246 and 642 and 320 images for testing purposes. The researchers used advanced microscopy techniques to capture these images which allowed them to see nuclear patterns and cytoplasmic features and complete cell morphology. Each image is stored as a 24-bit RGB file which maintains all color and texture elements that machines need for their automatic assessment. Medical experts verified all dataset annotations through their examination of the complete dataset which they used to establish standardized diagnostic criteria for accurate labeling. The system enables deep learning-based screening of blood smear images which helps with early-stage leukemia detection and disease monitoring and treatment assessment while supporting deep learning-based medical image analysis through its strong foundation.

TABLE I. DATASET CLASS DISTRIBUTION AND SPLITS

Split	Benign	Pre-B [Malignant]	Pre-B [Malignant] early	Pro-B [Malignant]	Not Cell/Unlabeled	Total
Train	351	661	698	529	7	2246
Valid	104	188	184	165	1	642
Test	57	106	97	59	1	320

Furthermore, because this study utilizes a publicly accessible, fully anonymized dataset provided by Roboflow, it does not involve direct human subject research or the handling of sensitive Protected Health Information (PHI). Consequently, institutional ethical clearance and patient consent were waived for this retrospective computational analysis, as the data was collected and anonymized by the original source providers prior to public release.

B. Image Pre-Processing

The model applies high-resolution microscopic blood cell images to identify both benign and malignant cases while it further differentiates between different leukemia subtypes. The data processing methods resize images to 224 pixel dimensions and use ImageNet mean and standard deviation for normalization and apply extensive data augmentation through random rotation and flipping and contrast and brightness alteration and Gaussian noise introduction to enhance data quality. The system achieves better local contrast through adaptive histogram equalization and Contrast Limited Adaptive Histogram Equalization which enhances the visualization of tiny cellular structures. The model now accurately detects nuclear shapes and cytoplasmic patterns and cell outlines because of the developed enhancements which reduce background noise interference. The proposed method improves feature extraction which leads to better classification results and enhanced generalization and greater system resistance which creates dependable deep learning systems for blood cancer detection support.

C. Model Description

The model employs multi-model deep learning techniques which include 3 recent architectural systems: Swin Transformer, ConvNeXt, and EfficientNetV2 to extract major characteristics from blood cell photographs. The learning process requires these models to fulfill different functions which enable the system to extract all available features thus improving both classification accuracy and system strength.

1) Swin transformer

Swin Transformer is an effective vision transformer that employs a Shifted Window Self-Attention (SW-MSA)

mechanism to examine blood smear images for leukemia diagnosis. The model learns effectively nuclear morphology, cytoplasmic abnormalities, and blast cell features while reducing computational complexity.

The attention mechanism shown in Eq. (1) in Swin Transformer is defined as:

$$Attention(Q, K, V) = \text{soft max} \left(\frac{QK^T}{\sqrt{dk}} \right) V \quad (1)$$

where, Q, K, V are query, key, and value matrices obtained by linearly transforming input feature maps, dk is the dimensionality of the key vector, ensuring numerical stability and T represents the matrix transpose operation

Mathematically, the output of the patch merging layer, which down-samples feature maps while preserving information shown in Eq. (2).

$$Z = \text{Concat}(X_1, X_2, X_3, X_4)W^p \quad (2)$$

where, x_i represents neighboring feature patches, W^p learnable projection matrix that maps concatenated patches into a lower-dimensional space and Z represents the output of the patch merging layer

The model applies hierarchical feature extraction in Eq. (3) using multiple transformer blocks before classification.

$$yC = \text{soft max}(W_{cls} f_{out} + b_{cls}) \quad (3)$$

where, yC represents the predicted class probabilities. Its clinical significance is that it acts as the final classification output, indicating the calculated likelihood that the analyzed blood cell belongs to a specific benign or malignant leukemia subtype. f_{out} is the final feature embedding after transformer layers, W_{cls} , b_{cls} are learnable classification and parameters.

Fig. 2 shows the architecture of swin transformer which is a hierarchical vision transformer that processes input images by dividing them into fixed size patches and applying shifted window attention for efficient feature extraction. It builds a multi-stage representation, gradually merging patches to capture both local and global dependencies in the image.

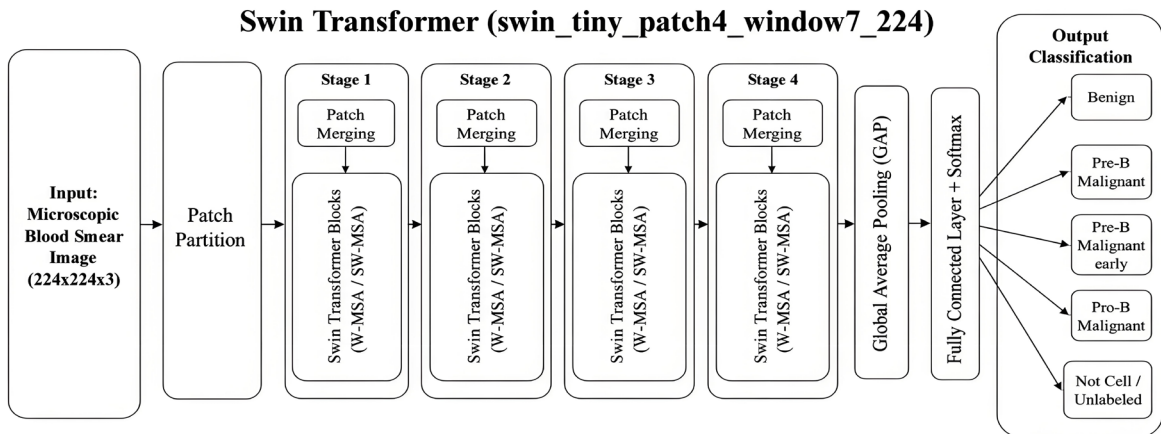


Fig. 2. Swin transformer (swin_tiny_patch4_window7_224) architecture adapted for 224×224 microscopic blood smear image classification.

Swin Transformer applied to blood cell cancer dataset started with preprocessing the images, resizing and normalizing prior to partitioning them into non-overlapping. Hierarchical architecture increasingly fused patches at successive stages, which allowed the simultaneous extraction of both fine-grained cellular information and large-scale structural patterns. Post-feature extraction, the Global Average Pooling (GAP) layer compressed spatial dimensions to obtain a compact representation of features. Lastly, the fully connected layer with a sigmoid activation function recognizes the images as cancerous or non-cancerous. The Swin Transformer performed at high accuracy, and it was very effective at identifying fine morphological variations of blood cells.

2) ConvNeXt

ConvNeXt is a modernized CNN-based model that refines spatial hierarchies using depthwise convolutions, inverted bottlenecks, and large kernels. It enhances blood cell segmentation, capturing the morphological patterns crucial for precise leukemia classification. ConvNeXt applies depth wise separable convolutions presented in Eq. (4), that uphold precision.

$$Y = \sigma(W_{depthwise} \times X) \quad (4)$$

where, $W_{depthwise}$ is a convolutional filter applied separately to each channel, X is the input feature map from the blood smear image, σ is the Gaussian Error Linear Unit (GELU) activation function, improving gradient flow and

Y represents the resulting output feature map after a depthwise convolution operation is performed.

ConvNeXt replaces batch normalization with Layer Normalization (LN) to improve training stability, defined in Eq. (5).

$$x_i^c = \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}} \quad (5)$$

where, c (as a superscript in) represents the channel index. It signifies that the layer normalization calculation is being applied across the specific channels of the extracted feature maps to stabilize the network's training. μ , σ^2 are the mean and variance of the input features, ϵ prevents division by zero. This ensures robust performance even in datasets with varying illumination and staining conditions which is further represented in Eq. (6).

$$Z = W_{expand} X, Y = \sigma(W_{depthwise} Z), Y^c = W_{project} Y \quad (6)$$

where, W_{expand} expands the feature dimension, $W_{depthwise}$ spatial convolution, $W_{project}$ reduces dimensionality back to the input size.

Fig. 3 shows that the ConvNeXt architecture is a modernized convolutional network inspired by Transformers. It enhances traditional CNNs by removing batch normalization, using Layer Norm (LN), and employing GELU activation for smoother feature transformations.

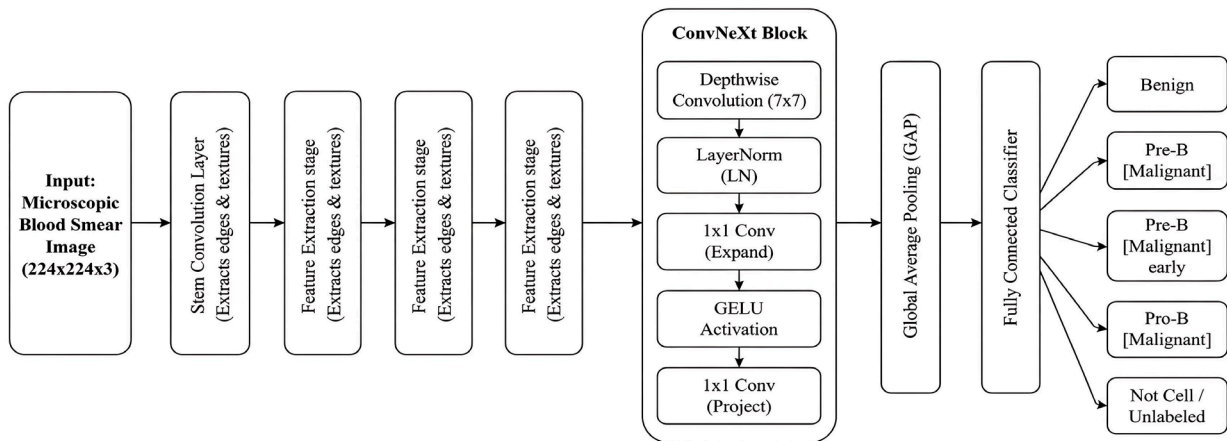


Fig. 3. ConvNeXt-Tiny (convnext_tiny) architecture utilizing transfer learning for the detection of leukemia subtypes from blood cell images.

Fig. 3 shows ConvNeXt architecture where it processes the blood cancer cell dataset through its image resizing and normalization methods which create a standardized input format. The initial stem convolution layer processed the images to extract low-level features which included edges and textures. The model used a hierarchical structure that increased feature depth while preserving spatial resolution to enable learning of intricate cell structures and textures. The extracted features were pooled through Global Average Pooling (GAP) to create a compact representation which then went to the fully connected layer for final classification. ConvNeXt provides a strong balance between efficiency and accuracy, making it an excellent

choice for large-scale image classification. While it was computationally lighter than Swin Transformer, it relied more on local feature extraction, which sometimes limit its ability to capture fine-grained variations in cancerous blood cells.

3) EfficientNetV2

EfficientNetV2 introduces compound scaling, fused convolutions, and Squeeze-and-Excitation (SE) blocks, achieving a balance between speed and accuracy.

Instead of arbitrarily selecting network depth, width, and resolution, EfficientNetV2 scales them uniformly which is presented in Eq. (7).

$$d = \alpha^\phi d_0, w = \beta^\phi w_0, r = \gamma^\phi r_0 \quad (7)$$

where, d, w, r represents the depth, width and resolution of the model, α, β, γ are scaling coefficients ensuring balanced feature extraction and d_0, w_0, r_0 and ϕ represent the baseline network's depth, width, and resolution, is the user-defined compound scaling coefficient,

SE blocks refine channel-wise feature importance presented in Eq. (8):

$$Z_S = \sigma(W_2 \times \text{ReLU}(W_1 \times Z)) \quad (8)$$

where, W_1, W_2 are fully connected layers, σ applies sigmoid activation for feature re-weighting, Z is the input feature map, Z_S is the re-weighted output feature map.

EfficientNetV2 replaces ReLU with Swish activation presented in Eq. (9):

$$\text{Swish}(x) = x \times \text{sigmoid}(\beta x) \quad (9)$$

where, β is a learnable parameter, allowing smoother gradient updates, x is the input feature map.

The final classification step presented in Eq. (10):

$$y^c = \text{softmax}(W_{cls} f_{out} + b_{cls}) \quad (10)$$

where, f_{out} is the final extracted feature embedding, y^c is the predicted class probability, W_{cls} and b_{cls} are the learnable weight and bias parameters of the classification layer.

Fig. 4 shows EfficientNetV2 is an optimized CNN that combines depth wise separable convolutions and mobile inverted residual bottleneck layers for high-speed, accurate image-processing. EfficientNetV2 effectively minimizes misclassifications and improves the overall reliability of leukemia detection. EfficientNetV2 processes the dataset by applying aggressive image resizing and normalization, ensuring input consistency. It used an optimized CNN backbone, starting with a stem convolution layer for initial feature extraction. The architecture depended on mobile inverted residual bottleneck layers together with depth wise separable convolutions to achieve efficient cellular feature extraction while keeping processing costs at minimal levels. EfficientNetV2 developed a new approach which let it modify kernel dimensions and feature depth to match the particular image complexity requirements. The features underwent processing through a 1×1 convolutional head which improved their quality before Global Average Pooling (GAP) used to achieve dimensionality reduction. The blood cell classification process determined cancer status through a final fully connected classification layer which established whether the blood cell was cancerous or non-cancerous. Efficient Network Version 2 (EfficientNetV2) delivered quick inference results together with reliable performance which made it suitable for use in instant classification applications. The system lost essential details because of its strong down sampling methods which caused it to miss small structural changes in blood cells, which decreased its performance.

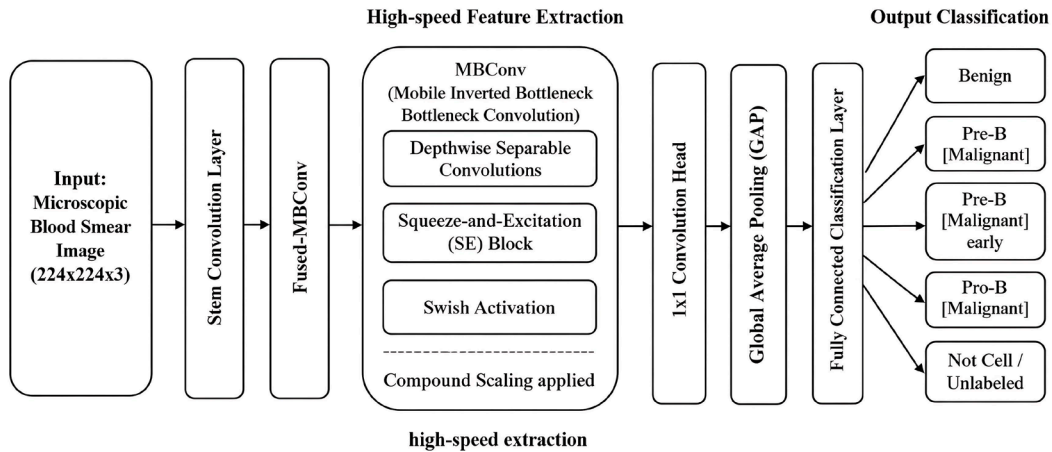


Fig. 4. EfficientNetV2-S (tf_efficientnetv2_s) architecture customized for high-speed feature extraction and classification of benign and malignant blood cells.

D. Localization and Feature Extraction

Fig. 5 suggests the proposed framework independently evaluates ConvNeXt, Swin Transformer, and EfficientNetV2 to improve feature extraction and classification for cancerous blood cell analysis. ConvNeXt improves spatial hierarchies with depth-wise convolutions to separate blood cells from noise and overlaps. Swin Transformer's shifted window self-attention allows accurate localization of nuclear and cytoplasmic abnormalities, while EfficientNetV2 enhances fine-grained discrimination of features to identify cancer affected areas.

Input images are resized to 224×224 and normalized with ImageNet statistics. Data augmentation such as flipping, rotation, contrast change, and addition of Gaussian noise enhance generalization. Explainable AI (XAI) methods such as Gradient-weighted Class Activation Mapping (Grad-CAM) and Eigen Class Activation Mapping (Eigen-CAM) confirm the model's attention on diagnostic areas by calculating gradients of target class scores. Principal Component Analysis (PCA)-based feature map analysis also further refines activation heatmaps by denoising. All models that are used utilize their last

convolutional layers for visualization so that the model pays attention to essential diagnostic features and not artifacts.

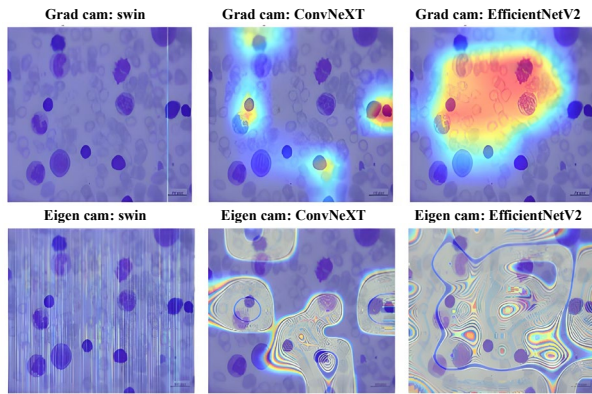


Fig. 5. Eigen CAM and grad CAM view.

V. RESULT AND ANALYSIS

This proposed model evaluates the effectiveness of deep learning models Swin Transformer, ConvNeXt, and EfficientNetV2 in classifying blood cancer cells using microscopic images. The dataset, sourced from Roboflow, consists of five classes of blood cell images. To ensure high accuracy in classification and optimal convergence, the models were trained for 10 epochs using an input image size of 224×224 and a batch size of 32. The models were developed using the AdamW optimizer with an initial learning rate of 1×10^{-4} , a weight decay of 0.01 and an epsilon of 1×10^{-8} , with cross-entropy loss. The learning rate dynamics during training used a ReduceLROnPlateau schedule which had a 2-patience and 0.5 factor. The model performance evaluation used essential assessment metrics which included accuracy and confusion matrix and classification report and Area Under the Receiver Operating Characteristic (AUC-ROC) curve and loss/accuracy graphs.

A. Experimental Setup

The proposed model uses PyTorch as its main deep learning framework while Torchvision handles all its data preprocessing and augmentation needs. The TIMM (PyTorch Image Models) library enabled transfer learning through its loading capabilities which provided access to specific pre-trained models that included Swin Transformer (swin_tiny_patch4_window7_224), ConvNeXt (convnext_tiny), and EfficientNetV2 (tf_efficientnetv2_s). The research team performed fine-tuning on our dataset using pre-trained weights to develop a model that could recognize the unique attributes of microscopic blood smears. The researchers conducted model training and evaluation in a Google Colab environment equipped with an NVIDIA T4

GPU (16 GB VRAM), an Intel Xeon CPU @ 2.20 GHz, and 12.7 GB of system Random Access Memory (RAM) to maximize the performance of GPU resources. The dataset underwent processing through OpenCV which improved image quality while Matplotlib and Seaborn created visualizations that included Grad-CAM and Eigen CAM heatmaps for model explanation. The researchers implemented their deep learning pipeline using a Python 3.9 programming environment. For full reproducibility and transparency of our methodology, the complete source code, Jupyter notebooks, and pretrained model weights utilized in this study are publicly available via Zenodo at <https://doi.org/10.5281/zenodo.20099897> [28].

B. Performance Metrics

Table II shows the performance metrics of the three models SwinTransformer, ConvNeXt, and EfficientNetV2. It provides key metrics including accuracy, weighted precision, weighted recall, and weighted F1-Score for each model. Because the dataset contains a severe class imbalance—specifically a rare “not cell/unlabeled” class with a support of only 1 in the validation set—weighted averages were used to accurately reflect the models’ overall diagnostic capabilities on the primary clinical classes. ConvNeXt achieved the highest weighted F1-Score of 99.77%, with precision and recall exceeding 99%, making it the most reliable model. Swin Transformer and EfficientNetV2 both achieved F1-Scores above 99%, with high recall values, ensuring accurate detection of cancerous cells. However, it is important to note that the macro-average scores were lower, as all models failed to classify the single instance of the extremely rare class, triggering a 0.00% precision and recall for that specific category. To further analyze these performance margins, the minor accuracy differences (0.15%–0.46%) must be contextualized against the dataset size. Given the validation set of 642 images, the 0.15% accuracy advantage of ConvNeXt (99.84%) over Swin Transformer (99.69%) equates to a difference of only a single correctly classified image. Similarly, the 0.46% difference compared to EfficientNetV2 (99.38%) represents a variance of only 3 images. Because this study utilized a single train-validation-test split without formal statistical significance testing (such as McNemar’s test or confidence intervals), this analysis indicates that the observed differences are statistically insignificant. Consequently, these marginal gains are highly likely due to random split variation rather than true architectural superiority. Therefore, while ConvNeXt achieved the highest numerical metric, all three architectures demonstrate comparable, highly effective classification capabilities on this dataset.

TABLE II. CLASSIFICATION PERFORMANCE

Model	Accuracy (%)	Weighted Precision (%)	Weighted Recall (%)	Weighted F1-Score (%)
Swin Transformer	99.69	99.53	99.69	99.61
ConvNeXt	99.84	99.69	99.84	99.77
EfficientNetV2	99.38	99.23	99.38	99.30

C. Confusion Matrix Analysis

Figs. 6–8 represents the confusion matrix of the models Swin Transformer, ConvNeXt, EfficientNetV2 respectively. ConvNeXt exhibited the lowest misclassification rate, with most predictions aligning perfectly with ground truth labels. Swin Transformer showed some misclassifications, mainly in cases where cancerous cells closely resembled non-cancerous ones, indicating minor challenges in distinguishing borderline cases. EfficientNetV2 performed well but had a slightly higher false positive rate than ConvNeXt, suggesting occasional misclassification of healthy cells as cancerous.

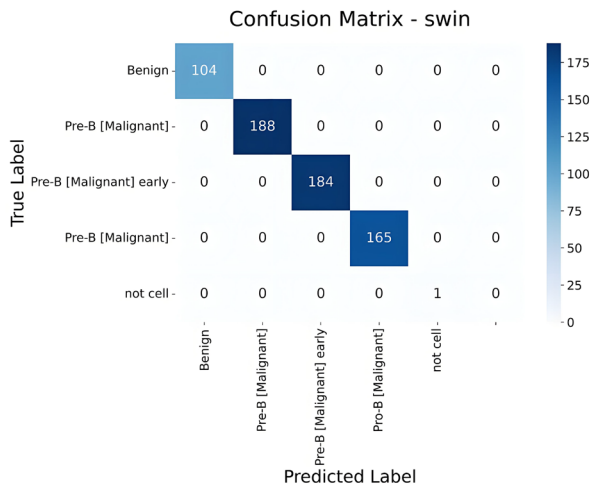


Fig. 6. Confusion matrix-swin transformer.

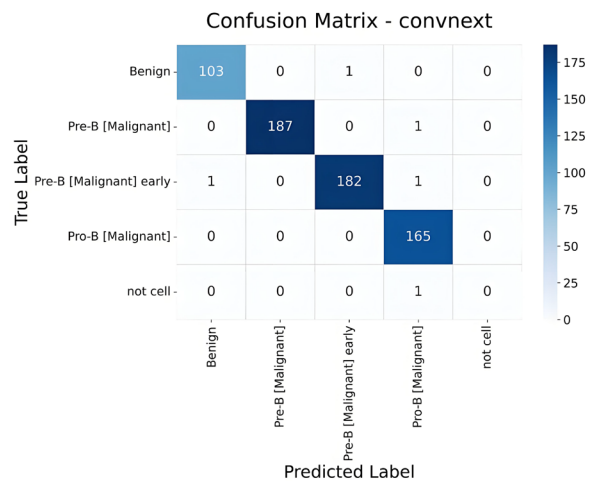


Fig. 7. Confusion matrix-ConvNeXt.

D. Loss and Accuracy Trends

Figs. 9 and 10 shows the training-validation accuracy curve and training-validation loss curve respectively. The figure represents learning and validation behavior of the three models that are used having 10 epochs each. Because these models utilize transfer learning with initialized pre-trained weights, they require significantly fewer epochs to converge and extract robust features. The training-validation loss graph depicts a sharp fall in not only training but also validation loss and then remain stable at input interval, which is an expected behavior when

fine-tuning pre-trained backbones and helps mitigate concerns of data leakage. Thus validating the proper functioning of the model to reduce loss. In the training-validation accuracy graph both training and validation accuracy exhibits a rapid increase during the initial epochs before stabilizing, indicating highly effective feature learning accelerated by the transfer learning approach.

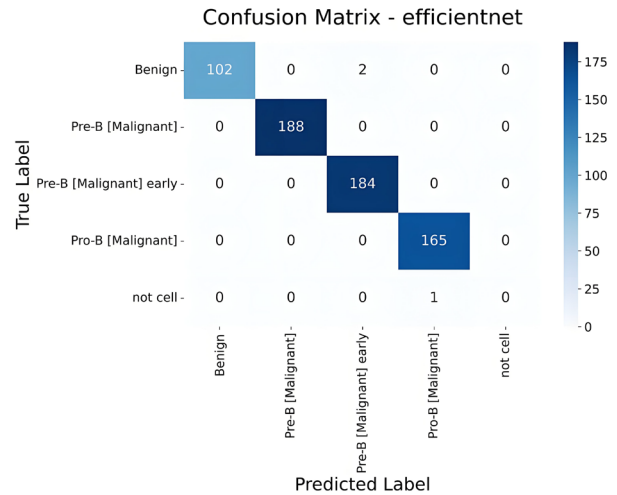


Fig. 8. Confusion matrix-EfficientNetV2.

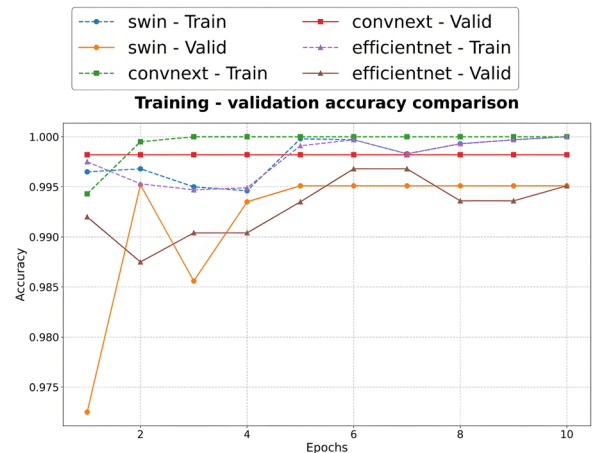


Fig. 9. Training-validation accuracy curve.

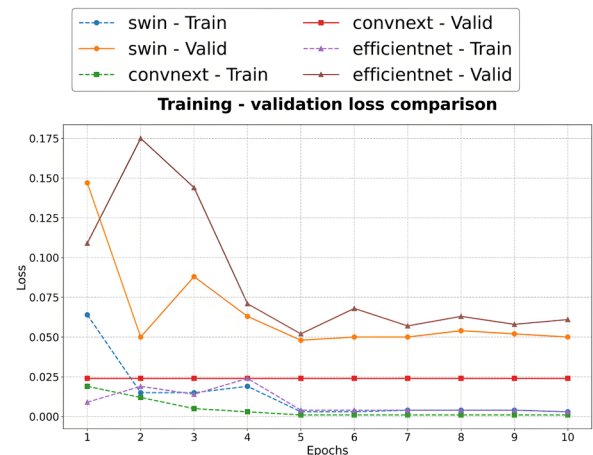


Fig. 10. Training-validation loss curve.

Fig. 11–13 shows the AUC-ROC Curve of the 3 models Swin Transformer, ConvNeXt, and EfficientNetV2 respectively. A higher Area Under the Curve (AUC) indicates better classification performance. After training, the model's predictions were converted into probability scores using softmax activation, which were then compared against the true labels. The AUC-ROC curves were generated for each blood cancer type, measuring the model's sensitivity and specificity at different decision thresholds. The approach achieved exceptional macro-average AUC scores exceeding 0.99 across the primary clinical classes, which proved its effectiveness to correctly identify leukemia through analysis of images.

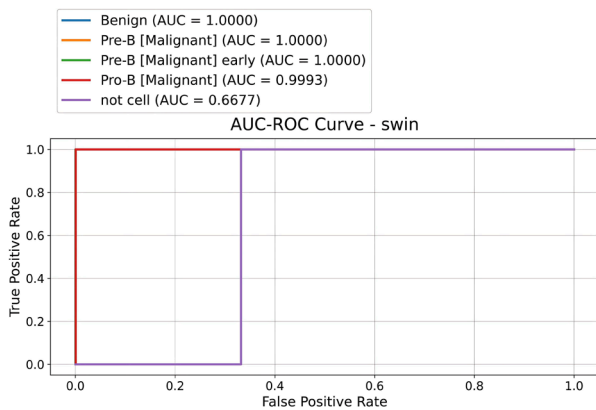


Fig. 11. AUC-ROC curve-swin transformer.

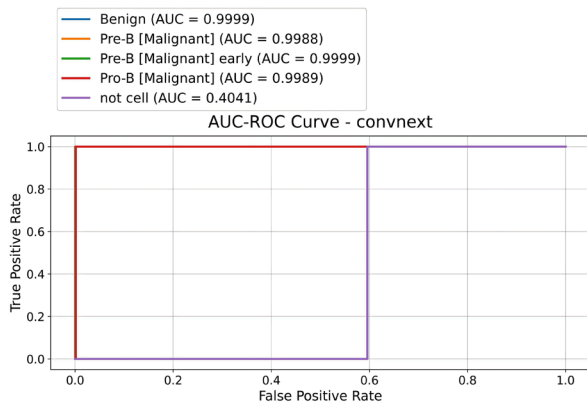


Fig. 12. AUC-ROC curve-ConvNeXt.

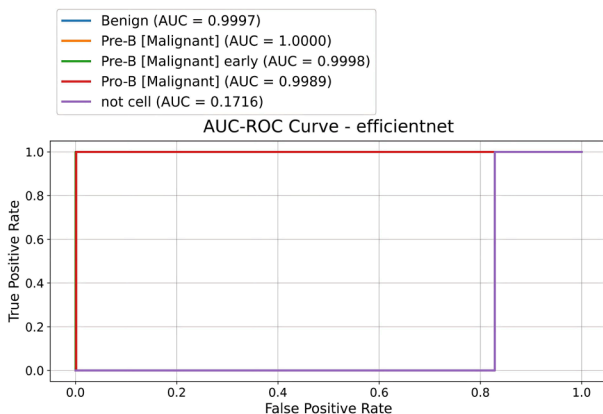


Fig. 13. AUC-ROC curve-Efficientnetv2.

VI. CONCLUSION

The proposed framework uses advanced deep learning techniques to detect blood cancer through an independent comparison of three key technologies which include Swin Transformer and ConvNeXt and EfficientNetV2. The models use transfer learning to classify malignant blood cells with high precision by applying three different architectural designs. The model becomes more transparent and clinically usable because Grad-CAM and Eigen-CAM visualization tools show which parts of the model most affected its decision-making process. The experimental results show that ConvNeXt achieved a peak numerical accuracy of 99.84%, closely followed by Swin Transformer (99.69%) and EfficientNetV2 (99.38%); however, analysis of these marginal differences suggests they are statistically equivalent and likely stem from dataset split variations. The results indicate that modernized CNN-based architecture (ConvNeXt) achieved better results than the transformer-based approach during this specific experiment.

The deep learning framework shows excellent results in leukemia subtype classification but has multiple restrictions which need to be mentioned. First, the dataset exhibits severe class imbalance, particularly regarding the 'not cell/unlabeled' class, which had a validation support of only one image. The existing approach lacks a cross-validation system, which serves as an essential tool to test the system's ability to perform across different data distribution scenarios. The evaluation used only one dataset from Roboflow, which did not include any external validation dataset, and this restriction limits the models' ability to work with various types of real-world diagnostic tools. The upcoming work needs to solve these problems through the implementation of k-fold cross-validation and the acquisition of a wider range of datasets and the application of class-balancing methods that include oversampling and class-weighted loss functions. Beyond methodological restrictions, several clinical deployment constraints must be addressed before this framework can be implemented in real-world diagnostic workflows. Deploying advanced deep learning architectures in active healthcare settings requires careful consideration of hardware limitations, specifically the substantial Graphics Processing Unit (GPU) memory Video Random Access Memory (VRAM) required to run heavy models like ConvNeXt and Swin Transformer effectively. Furthermore, inference time thresholds must be optimized to ensure the system can process microscopic blood smears rapidly enough to provide real-time clinical feedback without causing operational bottlenecks. Finally, the models must be evaluated against real-world image variability; blood smear images in practical clinical environments often exhibit significant differences due to varying lighting conditions, staining inconsistencies, and differing microscopy equipment across hospitals, which can impact the robustness of the automated predictions. Furthermore, due to the single train-validation-test split configuration, statistical significance testing (such as McNemar's test) and confidence intervals for the performance differences between models (e.g., the 0.15%

to 0.46% accuracy differences) were not computed. This limits the ability to conclusively determine the absolute statistical superiority of one architecture over another.

Future developments should concentrate on three main areas which include increasing dataset variety and implementing self-supervised learning techniques and enhancing computational efficiency for real-time operation of healthcare systems. Deep learning techniques which provide explainable results will lead to better model understanding which medical staff will use to build confidence in the system. The proposed method provides major improvements to deep learning-based haematological cancer diagnosis through its development of an accurate and interpretable and scalable diagnostic solution. The research establishes its primary foundation for developing trustworthy medical imaging systems which physicians will use through its advancements in deep learning architectures and its ability to combine multiple data types.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

SKD, DRS, LP, SA, and MAA conceptualized the study; DRS and LP developed the methodology; SKD, DRS, LP, QM and MAA provided the software; SKD, SA, QM, and MAA performed the validation and formal analysis; DRS and LP conducted the investigation and gathered the resources; SKD, DRS, MAA, and LP wrote the original draft; SKD, DRS, and MAA reviewed and edited the manuscript; MAA, acquired the funding. All authors reviewed and approved the final manuscript.

FUNDING

The authors would like to thank Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2025/01/32835).

ACKNOWLEDGMENTS

The authors thank Prince Sattam bin Abdulaziz University for supporting this research under their research project.

REFERENCES

- [1] A. Ahmed, A. Nagy, A. Kamal *et al.*, "Leukemia detection based on microscopic blood smear images using deep learning," arXiv preprint, arXiv:2301.03367, 2022.
- [2] N. Syed, M. E. S. Saeed, S. Hussain *et al.*, "Novel hierarchical deep learning models predict type of leukemia from whole slide microscopic images of peripheral blood," *Journal of Medical Artificial Intelligence*, vol. 8, 5, 2025.
- [3] R. F. O. Kizi, T. P. T. Armand, and H. C. Kim, "A review of deep learning techniques for leukemia cancer classification based on blood smear images," *Applied Biosciences*, vol. 4, no. 2, 9, 2025.
- [4] S. Dhalla, A. Mittal, and S. Gupta, "LeukoCapsNet: A resource-efficient modified CapsNet model to identify leukemia from blood smear images," *Neural Computing and Applications*, vol. 36, pp. 2507–2524, 2023.
- [5] A. K. Alzahrani, A. A. Alsheikhy, T. Shawly *et al.*, "A novel deep learning segmentation and classification framework for leukemia diagnosis," *Algorithms*, vol. 16, no. 12, 556, 2023.
- [6] M. Akter, R. Khatun, and M. M. Islam, "A hybrid feature fusion deep learning framework for leukemia cancer detection in microscopic blood sample using gated recurrent unit and uncertainty quantification," arXiv preprint, arXiv:2410.14536, 2024.
- [7] A. A. El-Aziz, M. A. Mahmood, S. Abd El-Ghany, "A robust EfficientNetV2-S classifier for predicting acute lymphoblastic leukemia based on cross validation," *Symmetry*, vol. 17, no. 1, 24, 2024.
- [8] S. Shafique and S. Tehsin, "Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks," *Technology in Cancer Research & Treatment*, vol. 17, 2018.
- [9] W. Zhang, Z. Huang, G. Luo *et al.*, "TopFormer: Token pyramid transformer for mobile semantic segmentation," in *Proc. 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 12073–12083.
- [10] B. Elsayed, M. Elhadary, R. M. Elshoeibi *et al.*, "Deep learning enhances acute lymphoblastic leukemia diagnosis and classification using bone marrow images," *Frontiers in Oncology*, vol. 13, 1330977, 2023.
- [11] M. A. A. Mollick, M. M. Rahman, D. M. Asadujjaman *et al.*, "Detection and classification of acute lymphoblastic leukemia utilizing deep transfer learning," arXiv preprint, arXiv:2501.14228, 2025.
- [12] A. Saeed, S. Shoukat, K. Shehzad *et al.*, "A deep learning-based approach for the diagnosis of acute lymphoblastic leukemia," *Electronics*, vol. 11, no. 19, 3168, 2022.
- [13] V. Granata, R. Fusco, S. V. Setola *et al.*, "Immunotherapy assessment: A new paradigm for radiologists," *Diagnostics*, vol. 13, no. 2, 302, 2023.
- [14] K. He, C. Gan, Z. Li *et al.*, "Transformers in medical image analysis," *Intelligent Medicine*, vol. 3, no. 1, pp. 59–78, 2023.
- [15] L. B. Konkyana, J. Rajanikanth, K. C. B. Rao *et al.*, "Transformer-based neural architectures for automated cancer classification in histopathology images," *African Journal of Biomedical Research*, vol. 28, no. 1, pp. 29–39, 2025.
- [16] N. P. Dharani, G. Sujatha, and R. Rani, "Blood cancer detection using improved machine learning algorithm," in *Proc. 2023 International Conf. on Circuit Power and Computing Technologies (ICCPCT)*, 2023, pp. 1136–1141.
- [17] T. G. Devi, N. Patil, S. Rai *et al.*, "Segmentation and classification of white blood cancer cells from bone marrow microscopic images using duplet-convolutional neural network design," *Multimedia Tools and Applications*, vol. 82, pp. 35277–35299, 2023.
- [18] A. I. Shehta, M. Nasr, A. E. D. M. El Ghazali *et al.*, "Blood cancer prediction model based on deep learning technique," *Scientific Reports*, vol. 15, 1889, 2025.
- [19] A. Awad and S. A. Aly, "Acute lymphoblastic leukemia diagnosis employing YOLOv11, YOLOv8, ResNet50, and inception-ResNet-v2 deep learning models," arXiv preprint, arXiv:2502.09804, 2025.
- [20] M. T. Ahad, I. J. Payel, B. Song *et al.*, "Dvs: Blood cancer detection using novel CNN-based ensemble approach," arXiv preprint, arXiv:2410.05272, 2024.
- [21] G. Atteia, A. A. Alhussan, and N. A. Samee, "BO-ALLCNN: Bayesian-based optimized CNN for acute lymphoblastic leukemia detection in microscopic blood smear images," *Sensors*, vol. 22, no. 15, 5520, 2022.
- [22] N. Sampathila, K. Chadaga, N. Goswami *et al.*, "Customized deep learning classifier for detection of acute lymphoblastic leukemia using blood smear images," *Healthcare*, vol. 10, no. 10, 1812, 2022.
- [23] A. Sulaiman, S. Kaur, S. Gupta *et al.*, "ResRandSVM: Hybrid approach for acute lymphocytic leukemia classification in blood smear images," *Diagnostics*, vol. 13, no. 12, 2121, 2023.
- [24] E. Dhiravidachelvi, S. P. S., R. Prabavathi *et al.*, "Artificial humming bird optimization-based hybrid CNN-RNN for accurate exudate classification from fundus images," *Journal of Digital Imaging*, vol. 36, pp. 59–72, 2022.
- [25] K. Dese, H. Raj, G. Ayana *et al.*, "Accurate machine-learning-based classification of leukemia from blood smear images," *Clinical Lymphoma Myeloma and Leukemia*, vol. 21, no. 11, pp. e903–e914, 2021.

- [26] M. Awais, M. N. Abdal, T. Akram *et al.*, “An efficient decision support system for leukemia identification utilizing nature-inspired deep feature optimization,” *Front Oncol.*, vol. 14, 1328200, 2024.
- [27] Blood cell cancer classification classification model. *Roboflow*. [Online]. Available: <https://universe.roboflow.com/informatica-mdica/blood-cell-cancer-classification-fzikt/dataset/1>
- [28] D. R. Sahoo. Blood cancer detection (Model weights). *Zenodo*. [Online]. Availabl. <https://doi.org/10.5281/zenodo.20099897>

Copyright © 2026 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).