SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open

Ensemble Deep Learning Approach for Early Detection and Subtype Classification of Blood Cancer Using Transfer Learning and Data Augmentation

Saroja Gangaram Landge[1*], Dr Bokare Madhav Motiram[2]

[1*]Research Scholar, Computer science, SSBES's Institute of technology and Management Nanded. E-mail: sarojalandge@gmail.com
[2]Associate Professor, Computer science, SSBES's Institute of Technology and Management Nanded. E-mail: bokaremadhav@yahoo.com

Abstract
Blood cancer constitutes a significant worldwide health issue that needs precise diagnostic methods to enhance patient treatment results. The latest Deep Learning (DL) methods demonstrate promising results. However, the research studies face difficulties because of their unbalanced datasets and their insufficient feature extraction capabilities together with their missing effective multi-stage classification methods. The study aims to create a hybrid DL system that uses microscopic blood smear images to identify blood cancer at an early stage and determine its different subtypes. The proposed method combines data pre-processing and augmentation with deep feature extraction through Residual Network (ResNet50), followed by dimensionality reduction using Principal Component Analysis (PCA). The system uses Swin Transformer and Vision Transformer (ViT) together with their combined model to classify refined features, which results in better performance outcomes. The outcomes show that the proposed system accomplishes better classification accuracy because the ensemble model reached a 99.54% accuracy rate together with exceptional precision, recall, and F1-score metrics. The research results exhibit that the proposed method delivers an effective automated system for blood cancer diagnosis, which operates with high reliability. The research results demonstrate that advanced feature extraction methods together with dimensionality reduction techniques and hybrid transformer-based models, create better classification performance with increased system stability.

Keywords: Blood Cancer, Deep Learning, Principal Component Analysis, ResNet50, Vision Transformer.

1. Introduction

Blood Cancer is a condition that is affecting more people in society and can include issues with the blood as well as the lymphatic system. It has now become a major issue for healthcare systems worldwide [1-3]. It is a type of cancer where people develop a lot of abnormal cells in the blood, and these cells can take over the functions of the body that defend from infection, deliver oxygen, and control other body functions [4, 5]. Blood Cancer has three major types: leukemia, lymphoma, and myeloma. These types have differing behaviors ranging from varying speeds of progression and can differ in what types of cancer medication can work and respond to [6-8]. Correctly and timely identifying and diagnosing each subtype of blood Cancer in the early stage is critical and can significantly influence how fast a person can heal and the number of cancer deaths [9-12]. Diagnosing blood cancer does not fully depend on advanced and digital medical science. It requires a highly skilled, time-consuming, extremely difficult, and tedious process to create and formulate blood electronic maps using skilled microscopes [13, 14]. There are many things that can greatly influence the results that are obtained from a combination of using traditional microscopy and laboratory work [15].

The limitations of existing diagnostic procedures necessitate the development of automated diagnostic tools that would aid medical professionals in the rapid and accurate interpretation of hematologic images [16, 17]. Figure 1 illustrates the progression of blood cell images, showing how normal blood cells transform into white blood cells and how further changes can lead to lymphoma. The images on the right present representative photo images of each of the four major subtypes of leukemia: Acute Lymphoblastic Leukemia (ALL), Acute Myeloid Leukemia (AML), Chronic Lymphocytic Leukemia (CLL), and Chronic Myelogenous Leukemia (CML). The figure emphasizes the visual differences between healthy cells, lymphoma cells, and blood cancer cells, as well as the distinct dendritic cell populations represented in this study.

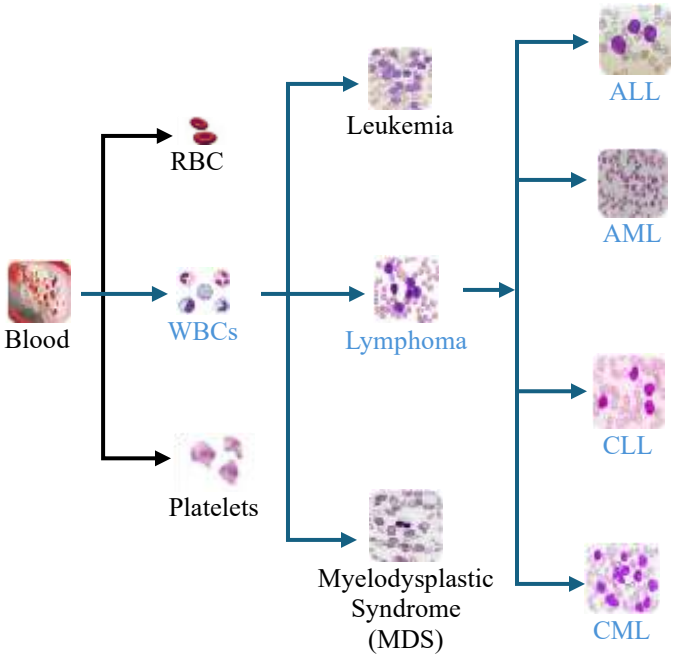


Figure 1: Blood components and main types of blood cancer [18].

DL has emerged as a powerful tool for medical imaging assessment because it enables machines to learn intricate patterns that humans cannot observe [19-22]. The ability of DL models to automatically extract features from data makes them better suited for blood smear image analysis because the disease diagnosis requires the detection of tiny changes in cell shape and texture [23-25]. The Convolutional Neural Network (CNN) and Recurrent Neural Network DL methods demonstrate strong performance in image recognition and pattern analysis, which shows their potential to study blood disorders [26, 27]. However, the training of DL models usually requires large and balanced datasets. In medical research, such datasets are often limited due to privacy restrictions, rare disease occurrence, and challenges in expert-level annotation [28, 29].

To overcome this challenge, this research would create a hybrid DL methodology for early identification of blood cancer and subtype classification based on transfer learning and data augmentation methods, which would permit the system to learn faster, create more stable classifications, and give better diagnostic accuracy from limited quantities of medical data. The developed system would offer accurate, efficient, and reliable diagnostic information while also reducing the number of diagnostic tests each day and decreasing the chance of human error occurring, enabling physicians and other healthcare professionals to make quicker, better decisions in the treatment of patients; thus improving patient care and treatment outcomes. The prime contributions of this research are stated as follows:

- Presents a DL approach for blood cancer detection using enhanced pre-processing to improve ALL smear image quality.
- Utilizes ResNet50 for automated, discriminative feature extraction refined through normalization and PCA.
- Applies advanced data augmentation to increase dataset diversity and boost transformer model generalization.
- Develops a hybrid Swin – ViT model achieving improved classification of Benign, Early, Pre, and Pro stages.

The paper has the following structure, which begins with Section 2, which contains the literature evaluation of this investigation, and ends with the identification of the research gap. The dataset description, proposed methodology, technique, and proposed algorithm are found in Section 3 of the document. The experimental results together with the performance evaluation of the proposed framework, appear in Section 4. The study concludes with essential discoveries in Section 5.

2. Literature Review

Numerous DL and Machine Learning (ML) methods have been investigated recently for the identification and categorization of blood cancer subgroups. A Bayesian DL model was proposed by **Hita et al. (2026) [30]**. It combined transfer learning, data augmentation, and uncertainty quantification to identify leukemic and healthy lymphocytes in blood smear images. Pretrained InceptionV3, VGG16, and ResNet50 models were fine-tuned on

ALL-IDB2 and enhanced with Monte Carlo dropout. VGG16 performed best, achieving 98.65% accuracy. Additionally, **Mollick et al. (2025) [31]** deployed two CNN models: MobileNetV2 and a custom model, employing the ALL-image dataset, then applying the Synthetic Minority Oversampling Technique (SMOTE) to enhance and equilibrate the training dataset. The bespoke model attained a precision of 98.6%. Moreover, **Jain et al. (2025) [32]** built strong automated models to classify all subtypes using peripheral blood smear images. Employing advanced feature extraction using the Swin Transformer framework along with Momentum Contrast (MoCo) for contrastive learning and classifying using a Bidirectional Encoder Transformer, this model attained 92.5% overall classification accuracy. Building upon the integration of advanced feature processing pipelines, **Enas M.F. El Houby (2025) [33]** introduced a Computer-Aided Detection system designed for comprehensive recognition. The system integrated pre-processing, segmentation, feature extraction, and Ant Colony Optimization-based feature selection, subsequently employing Naïve Bayes (NB), Support Vector Machine (SVM), and K-Nearest Neighbors classifiers. The NB classifier attained optimal performance, with an accuracy of 96.15%. Similarly, **Galila et al. (2025) [34]** established a framework for the early diagnosis of ALL in microscopic pictures. The research incorporated data augmentation, pre-processing, and segmentation via the Chan-Vese algorithm, feature extraction using the Gray Level Co-occurrence Matrix, and data normalization employing the Min-Max scaler. Post-normalization, the accuracy of the Acute Lymphoblastic Leukemia-Database (ALL_DB1) dataset rose from 88.42% to 98.42%. Expanding hybrid methodological approaches, **Kasim et al. (2025) [35]** proposed an innovative technique that incorporates pre-trained CNN architectures. The methodology demonstrated strong results, especially for the combination of InceptionV3 and SVM, which achieved 88% accuracy. This result was closely followed by the combination of Visual Geometry Group 16 (VGG16) and EXtreme Gradient Boosting (XGBoost) with 87% accuracy. Additionally, **Mustapha and Ozsahin (2025) [36]** examined the performance of the advanced CNN methodology ConvNeXt when classifying high-quality peripheral blood smear images into subtypes of AML. ConvNeXt achieved 95% accuracy of class assignment compared to 91% accuracy of ResNet50 and 81% accuracy of transformer models, therefore achieving results that outperformed ResNet50. Moreover, an advanced optimized CNN for the early treatment and detection of ALL was proposed by **Anand et al. (2024) [37].** The recommended model was calibrated utilizing hyperparameters including 30 epochs, a batch size of 32, and optimizers including Adam and Adamax. Among both optimizers, the suggested framework demonstrated superior performance with the Adam optimizer, achieving an accuracy of 0.96. Similarly, A novel strategy for leukemia classification was outlined by **Haque et al. (2024) [38]**. This strategy consisted of state-of-the-art image processing, multi-dataset implementations, sophisticated feature extraction, and transfer learning. A thorough comparison with other baseline models showed that Inception-ResNet stood out by achieving the highest F1 score of 96.07% and 95.89% in the binary and multi-class cases, respectively. Furthermore, **Singh et al. (2024) [39]** utilized DL and image processing methodologies to classify lymphoblast and lymphocyte cells through a computer-assisted approach. The proposed ensemble technique attained an accuracy of 98.23%. Subsequently, a novel hybrid deep-learning-based feature engineering methodology was developed by **Atteia et al. (2023) [40]**. The proposed method combined the robust capabilities of PCA and Particle Swarm Optimization (PSO) for picking informative features from BPI images with the efficacy of pre-trained CNNs for feature extraction. The Bayesian-optimized SVM utilizing the suggested hybrid PCA-PSO feature set attained the greatest categorization accuracy of 97.4%. Additionally, **Sheet et al. (2023) [41]** suggested a computer-aided diagnostic method employing Mobilenet V2 and CNNs to differentiate leukemia photos from healthy images. The findings indicated that the proposed model achieved a total accuracy of 96.58%. Moreover, an innovative approach for screening white blood cells was suggested by **Sampathila et al. (2022) [42]**. The suggested DL method utilized tiny pictures of blood samples as input. The model training was conducted on Google Collaboratory utilizing the Nvidia Tesla P-100 Graphics Processing Unit. This precise classifier achieved the highest accuracy of 95.54%.

Current literature has identified several weaknesses in blood cancer image-based diagnostic research. The problem of data imbalance and low representation diversity continues to exist in blood cancer imaging. Therefore, the use of basic augmentation or SMOTE does not capture the morphological variability necessary for accurate classification of blood cancer images [32, 36]. The majority of studies are still focusing on binary classification instead of subtypes, which is important for clinical decisions [33, 38]. Additionally, there are many advanced ensemble DL models and state-of-the-art architectures, such as ConvNeXt and Hybrid Transformers, that have not been used extensively and therefore have the potential to increase model accuracy and stability [34, 40]. Very few studies utilize Explainable Artificial Intelligence, decreasing the interpretability and credibility of clinicians concerning automatic predictions [35, 39], and most of the existing models do not have external validation with independent clinical datasets, which limits their applicability and generalizability to real-world situations [32, 37, 40].

3. Methodology

The methodology developed for early detection and classification of cancer by microscopic imaging of a blood smear of a child with ALL is described in this section. Pre-processing steps include removing noise from the image, normalizing the intensity of all images, and mitigating all missing and inconsistent data. So, there is one consistent level of quality across all images in the dataset. The dataset was also expanded through augmentation techniques of rotation, reflection (flipping), and rescaling. Semantic segmentation was used to separate cells of interest from other surrounding cells, which would assist in feature extraction during the next stage of the process. The feature extraction process utilized the ResNet50 architecture, allowing for the extraction of deep features from processed images. Features were indexed in a database according to the index of each original photo, and the feature matrices were split into training and validation datasets. Two Transformer DL models (Swin Transformer and ViT) were used for training on the dataset and classifying each image at the benign stage, early stage, pre-stage, and pro-stage of cancer. The models' performance criteria were assessed using various quality metrics. Figure 2 shows the architecture of the suggested analytical approach.

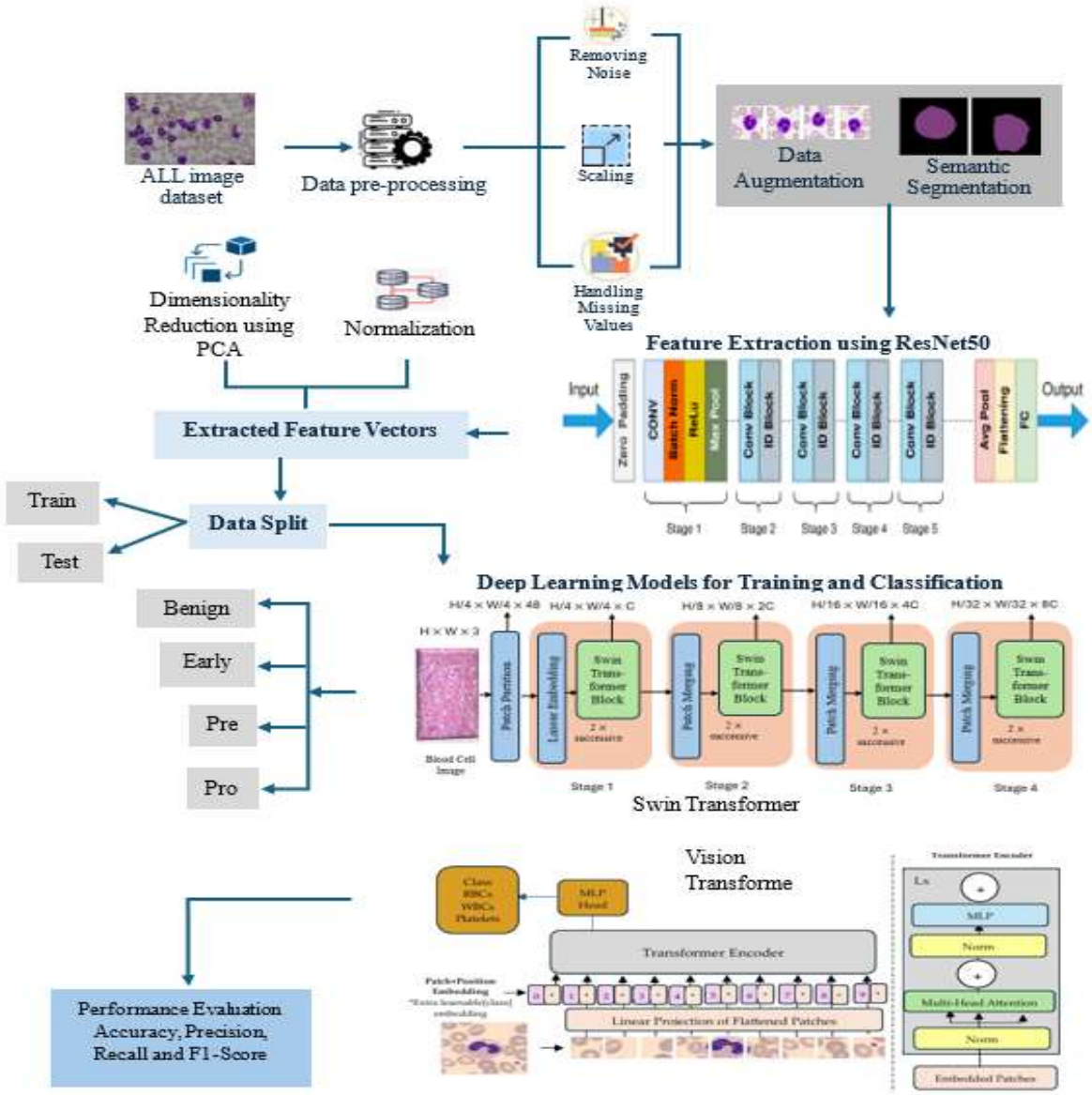


Figure 2: Proposed Methodology

3.1Dataset: ALL-Image Dataset

The ALL-Image Dataset is a large clinical dataset with clinically relevant images to aid in the automated diagnosis of leukemia and other research associated with it. The ALL-Image Dataset comprises 3,256 peripheral blood smear pictures from 89 individuals suspicious of having ALL, and the blood samples are created and

stained at an accredited laboratory using a trained technician [43]. Each image contains a detailed morphology of blood cells under a microscope, providing valuable information on the structure, texture, and differences between leukocytes. As such, it provides good training data for both ML and DL techniques. In addition, the dataset is of sufficient size and is clinically relevant to provide the necessary variation and representational power for constructing robust models for detecting and classifying leukemias.

3.2Data pre-processing

The process of data preparation is defined as all aspects of preparing the ALL-blood smear images through several required methods, such as removing noise, scaling, and managing missing values to produce uniform clean input to the following phases of analysis [44]. The method employed for removing noise is to use Gaussian smoothing to reduce random high-frequency disturbances while preserving significant structures of the cells mathematically delineated:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

where x and y denote pixel coordinates, σ represents the standard deviation of the Gaussian kernel, and G(x, y) is the resulting smoothed intensity value. Resizing of Images is performed using Interpolation (through resampling) using the following equation:

$$I'(x,y) = I(ax,by) \quad (2)$$

where $I'(x,y)$ denotes the resized image value, $I(ax,by)$ refers to the original image sampled at the scaled coordinates, and a and b represent the horizontal and vertical scaling factors, respectively. The process of handling missing pixel data requires both identifying incomplete information/corrupted regions of the image and filling the gaps created by those pixels through imputation methods, for example, mean substitution, as represented by the following:

$$I_{\text{new}} = \frac{1}{N}\sum_{i=1}^{N} I_i \quad (3)$$

Where I_{new} is the imputed pixel value produced from the N available neighboring pixel intensities, and I_{i} represents each contributing pixel value. To ensure that no important area of the images contains blank or undefined values, the dataset is subsequently standardized (particularly after pre-processing) to reduce variability and improve the overall quality of the images before extracting feature points from them.

3.3Data Augmentation

Data augmentation improves the diversity and variability of the ALL-blood smear dataset by creating copies of certain images and modifying them to help the architecture generalize better and prevent overfitting [45]. This consists of some controlled changes such as rotation, flipping, translation, and zooming, which can all be mathematically modeled using some affine transformations such that $T(x) = Ax + b$, where A represents some transformation matrix that defines which rotation and scaling to apply, and b represents some translation. To perform rotation, one would generally use the rotation matrix.

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad (4)$$

where θ is the rotation angle applied to the image, A is the rotation matrix, and T(x) represents the transformed pixel location. Allowing the image to be rotated by an angle θ. Horizontal or vertical flipping is achieved by applying reflection matrices, such as:

$$A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \quad (5)$$

where A here represents a horizontal reflection matrix that flips the image across the vertical axis. For horizontal flips, while zooming is performed by scaling factors represented as:

$$A = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \quad (6)$$

where s_x ands_y denote scaling factors along the x- and y-directions, respectively. These augmentation operations add variability to orientation, size, and position while keeping the truthfulness of the biological characteristics; thus, these operations increase the models' capacity to identify leukocyte structures across a wide range of visual contexts.

3.4Semantic Segmentation

Semantic segmentation helps to segment the Region of Interest (ROI) from the adjacent backdrop of a slide [46]. This allows for only those structures that have diagnostic importance to be used in subsequent processing and analysis of a sample. Semantic segmentation is initiated through the use of pixel-level comparisons between the cell ROI and the background of the slide using threshold techniques, as described:

$$S(x,y) = 1 \text{ if } I(x,y) > \tau, \text{ else } 0 \quad (7)$$

where $S(x, y)$ is the segmented output, $I(x, y)$ is the pixel intensity at coordinates (x, y), and τ is the threshold separating foreground from background. More refined segmentation can be achieved using morphological operations such as dilation and erosion, represented mathematically as:

$$I \oplus B = \{z \mid (B)_z \cap I \neq \emptyset\} \text{ for dilation} \quad (8)$$

where $I \oplus B$ denotes dilation, B is the structuring element, and z represents pixel locations where the shifted structuring element intersects the image, and

$$I \ominus B = \{z \mid (B)_z \subseteq I\} \text{ for erosion} \quad (9)$$

where $I \ominus B$ denotes erosion, meaning the structuring element B fully fits within the image region I at position z. The structuring element is denoted as B for these operations to assist in removing small areas that are not needed and to also enhance the borders of leukocytes (white blood & red blood cells). This segmented output has clearly identified the cellular area of interest with a clear focus that is appropriate for precise feature extraction and further analysis.

3.5Feature Extraction using ResNet-50

The ResNet50 model extracts features by passing the processed blood smear image through convolutional and residual learning layers that create deep feature representations that differentiate one type of image from another [47]. The remaining learning modules are designed to provide a mapping from original source data to the output from the model.

$$F(x) = H(x) - x \quad (10)$$

where $F(x)$ represents the residual function, $H(x)$ is the desired underlying mapping, and x is the input to the residual block. Allowing the network to produce outputs of the form:

$$H(x) = F(x) + x \quad (11)$$

where this formulation highlights the skip connection, enabling more stable gradient flow during training. Through skip connections, which facilitate stable training and capture complex morphological patterns. Convolutional operations extract spatial and textural characteristics according to:

$$Y(i, j) = \sum_m \sum_n X(i + m, j + n) \cdot K(m, n), \quad (12)$$

where $Y(i, j)$ is the output feature value at location (i, j), X is the input feature map, $K(m, n)$ is the convolution kernel, and m, n index the kernel dimensions, and global average pooling decreases each feature map into a compact feature value using:

$$y_k = \frac{1}{Z} \sum_{i=1}^{H} \sum_{j=1}^{W} X_k(i, j). \quad (13)$$

where y_k is the pooled feature for channel k, $X_k(i, j)$ is the value of the k-th feature map at spatial location (i, j), H and W denote the height and width of the feature map, and $Z = H \backslash times$ W is the total number of spatial positions.

By following the above procedure, a useful representation of leucocyte characteristics is created as shown in Figure 3. The input to the model is processed and turned into a feature vector through a systematic series of steps: zero-padding and convolution, batch normalization, and ReLU activation. Max-pooling continues with further max-pooling layers and the addition of one or more blocks (convolutional and identity) to build incrementally more elaborate feature maps from these feature vector layers until they finally become smoothed and sent to an entirely linked output layer.

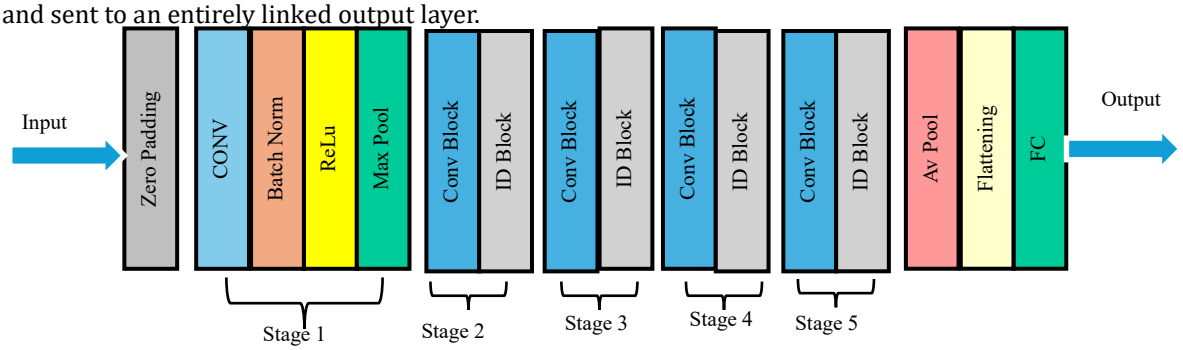


Figure 3: ResNet50 Architecture [48]

3.6Extracted Feature Vectors

Feature vectors produced by ResNet50 are further processed via normalization and dimensionality reduction to improve their quality for future classification. Initially, feature value ranges are normalized on a consistent scale to limit the influence of high magnitude values on model fit through learning. Generally, this is accomplished through the process of min–max normalization as follows:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad (14)$$

where x' is the normalized value, x is the original feature value, and x_{min} and x_{max} represent the minimum and maximum values within that feature range. Z-score standardization, expressed as:

$$x' = \frac{x - \mu}{\sigma}, \quad (15)$$

In which μ and σ are the average and standard variation of each feature. After normalizing the feature vectors, PCA is performed to produce a smaller number of principal component vectors that represent the original feature vectors, and PCA calculates the covariance matrix of the data [49].

$$C = \frac{1}{n-1}(X - \mu)(X - \mu)^T, \quad (16)$$

where C is the covariance matrix, X is the data matrix, μ is the mean vector, and n is the number of samples, and identifies its eigenvalues and eigenvectors using:

$$Cv = \lambda v. \quad (17)$$

where v is an eigenvector of the covariance matrix C, and λ is its corresponding eigenvalue. The eigenvectors corresponding to the largest eigenvalues are selected and used to transform the feature space as:

$$Z = V^T X, \quad (18)$$

where Z is the transformed feature matrix, V contains the selected eigenvectors, and V^{T} is its transpose. Provides a safer, reduced, and richer representation of an image's information. Self-normalization of data gives Z a standardized, smaller, and more compact shape for use with PCA in the training phase to create and classify features more accurately.

3.7Deep Learning Models for Training and Classification

Final refined feature vectors are sent to modern advanced transformers such as Swin Transformers and ViT as represented in schematic diagrams. Swin Transformers and ViT have a structure that captures both local-level (e.g., small-scale) and contextual structure cues together from blood smear images for predicting accurate stages of leukemia classification.

The Swin Transformer operates on non-overlapping image patches and processes them through a hierarchical structure, incorporating shifted-window self-attention [50]. Within each window, the attention function is computed as:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V, \quad (19)$$

The terms $Q, K, and\ V$ represent query, key, and value matrices, while d indicates the dimension of features. The Swin Transformer model processes blood cell images by first partitioning the image into patches, which it combines through a linear process before reaching Stage 1. The model uses four stages, which combine sequential Swin Transformer Blocks with patch merging to reduce spatial resolution while increasing feature dimensions. The hierarchical structure enables the system to obtain both local and global patterns, which create deep feature maps that serve classification purposes.

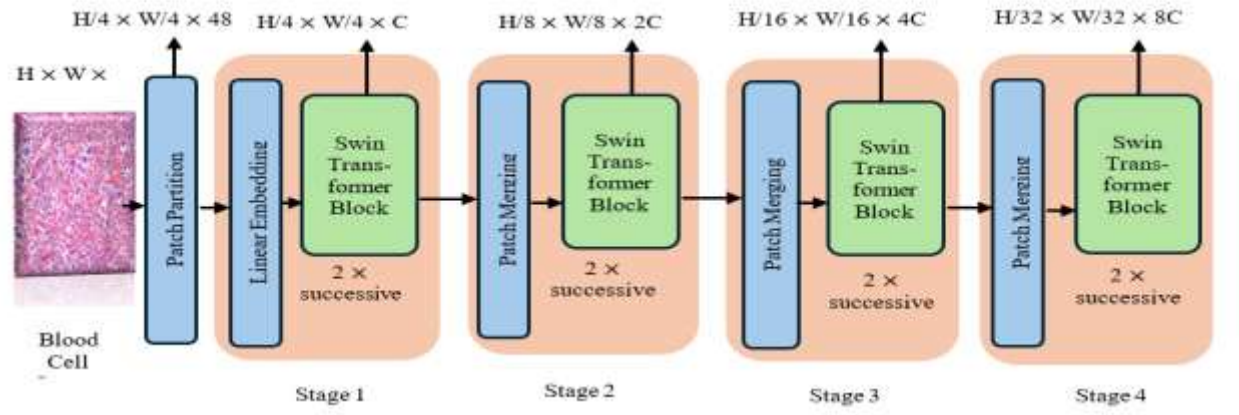


Figure 4: Swin Transformer architecture [51]

ViT divides the image into fixed-size patches, converts them into embedded tokens, and applies Multi-Head Self-Attention (MHSA) across all tokens [52]. MHSA is formulated as:

$$\text{MHSA}(X) = \sum_{i=1}^{h} \text{Attention}(XW_i^Q, XW_i^K, XW_i^V), \quad (20)$$

where h is the amount of attention heads and W_i^Q, W_i^K, W_i^V Represent the representation matrices for every head. The graphics depict the comprehensive tokenization and transformer-encoder pipeline employed in ViT. Figure 5 demonstrates that the ViT architecture initially divides the input picture into fixed-size segments, which are then linearly projected and integrated with positional and class embeddings. The embedded patches are consecutively fed into a transformer encoder including successive layers of normalization, multi-head attention, and multilayer perceptrons. Upon completion of feature extraction from all patches, the class token is directed via a multilayer perceptron head to provide the final class prediction.

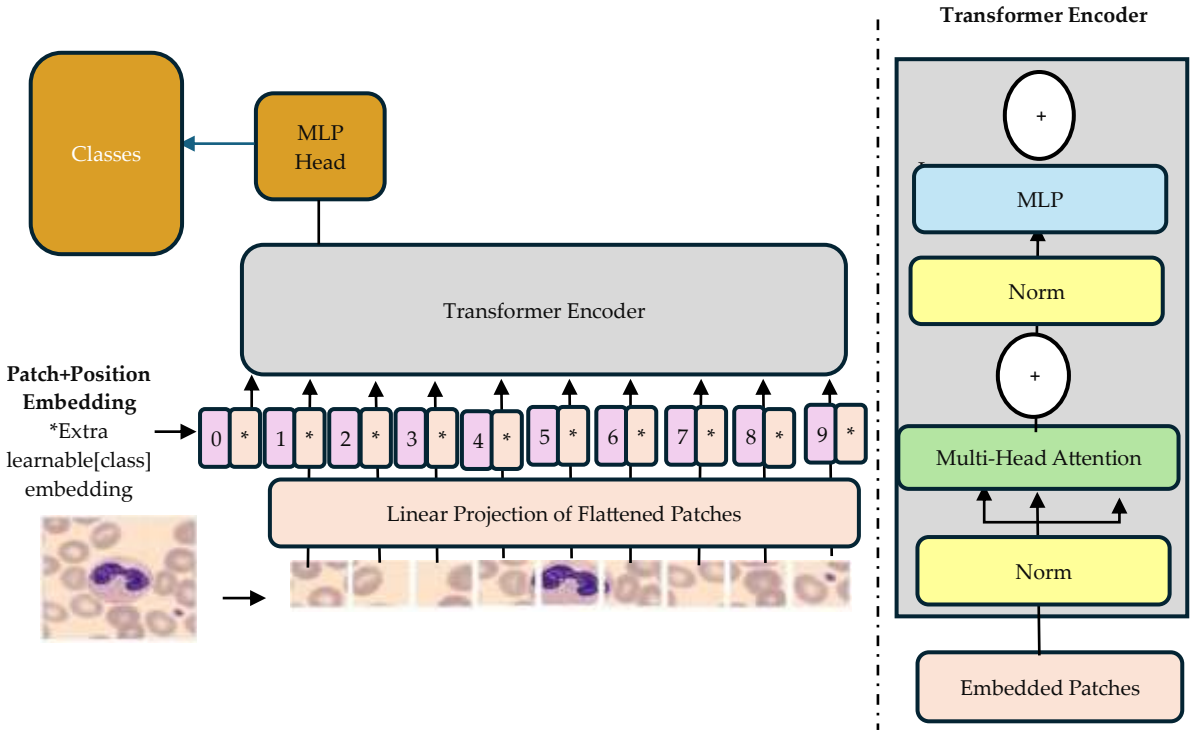


Figure 5: ViT architecture [53]

Upon acquiring discriminative cellular representations, both systems transmit their final encoded features to completely connected layers, subsequently followed by a softmax classifier that generates class probabilities. The softmax function is stated as:

$$P(y = c \mid z) = \frac{e^{z_c}}{\sum_{k=1}^{C} e^{z_k}} \quad (21)$$

In which z_c indicates the logit for class c, and C is the total no. of categories. The label with the highest probability, Benign, Early, Pre, or Pro, is selected as the final prediction, enabling robust stage-wise classification of leukemia.

Algorithm: Hybrid Swin-ViT Model for Blood Cancer Classification

Input:
D_train = { (I_n, y_n) } // preprocessed / segmented blood cell images
D_test = { (I_m, y_m) } // test images and labels
Classes = {Benign, Early, Pre, Pro}

Output:
Trained Swin and ViT models, final predictions for D_test

1. Model Initialization
1: Initialize Swin Transformer parameters θ_s
2: Initialize ViT parameters θ_v
3: Initialize classification heads FC_s and FC_v
4: Choose optimizer and learning rate η
5: Set fusion weight α ∈ [0,1] (e.g., α = 0.5 for equal fusion)

2. Training of Hybrid Model
6: for epoch = 1 to E do

```
7:    for each mini-batch (B_I, B_y) from D_train do
8:       // ----- Swin Transformer branch -----
9:       B_patches_s ← Swin_PatchPartition(B_I)
10:      B_embed_s   ← Swin_LinearEmbedding(B_patches_s)
11:      Z_s         ← Swin_TransformerBlocks(B_embed_s; θ_s)
12:      logits_s    ← FC_s(Z_s)
13:      p_s         ← Softmax(logits_s)     // class probabilities
14:      // ----- ViT branch -----
15:      B_patches_v ← ViT_PatchPartition(B_I)
16:      B_tokens_v  ← ViT_PatchPosEmbedding(B_patches_v)
17:      Z_v         ← ViT_EncoderBlocks(B_tokens_v; θ_v)
18:      logits_v    ← FC_v(Z_v)
19:      p_v         ← Softmax(logits_v)
20:      // ----- Fusion and loss -----
21:      p_hybrid    ← α · p_s + (1 − α) · p_v
22:      L           ← CrossEntropyLoss(p_hybrid, B_y)
23:      // ----- Backpropagation -----
24:      Update θ_s, θ_v, FC_s, FC_v using gradient of L and optimizer
25:   end for
26: end for
```

```
3. Inference and Classification
27: for each image I in D_test do
28:    // Swin prediction
29:    patches_s  ← Swin_PatchPartition(I)
30:    embed_s    ← Swin_LinearEmbedding(patches_s)
31:    Z_s        ← Swin_TransformerBlocks(embed_s; θ_s)
32:    logits_s   ← FC_s(Z_s)
33:    p_s        ← Softmax(logits_s)
34:    // ViT prediction
35:    patches_v  ← ViT_PatchPartition(I)
36:    tokens_v   ← ViT_PatchPosEmbedding(patches_v)
37:    Z_v        ← ViT_EncoderBlocks(tokens_v; θ_v)
38:    logits_v   ← FC_v(Z_v)
39:    p_v        ← Softmax(logits_v)
40:    // Fusion and final label
41:    p_hybrid   ← α · p_s + (1 − α) · p_v
42:    y_hat      ← argmax_c p_hybrid[c]     // c ∈ Classes
43: end for
```

3.8Performance Evaluation

The assessment of classification model effectiveness and reliability requires testing their capacity to identify different stages of leukemia. The assessment uses standard quality metrics, which are derived from confusion matrix components that include True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (22)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (23)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (24)$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (25)$$

With these metrics, the proper sense of model performance and reliability of the leukemia classification system can be obtained, highlighting what it does well and what can still be improved.

4. Results

The section presents experimental results that demonstrate the performance of the proposed DL framework for blood cancer detection and classification through testing on the ALL-image dataset. The dataset is split into an 80:20 ratio for training and testing to ensure reliable evaluation. The model performance assessment occurs

through standard evaluation metrics at various points between the pipeline execution. Table 1 presents the hyperparameter configuration of the Swin Transformer and ViT models, including architectural and training settings. The selected parameters ensure stable convergence and enable a fair comparison between the models.

Table 1: Hyperparameter Configuration

Hyperparameter	Swin Transformer	ViT
Image Size	224 × 224	224 × 224
Patch Size	4 × 4	16 × 16
Embedding Dimension	96	768
Number of Heads	[3, 6, 12, 24]	12
MLP Ratio	4	4
Dropout Rate	0.1	0.1
Optimizer	Adam	Adam
Learning Rate	0.0001	0.0001
Batch Size	32	32
Epochs	10	10

The dataset contains unprocessed blood smear images, which need pre-processing to reach better image quality and uniform image presentation. Figure 6 displays original images that show both visible noise and problems with different brightness levels. The noise removal process results in better visibility of cellular structures because of background improvement, and the scaled images retain their original dimensions while showing improved visual consistency. The complete image quality enhancement through this refinement process results in better outcomes for the next image processing operations.

Figure 6: Data Pre-processing

The pre-processed images become more diverse through data augmentation, which enhances their dataset diversity. Figure 7 demonstrates that horizontal flipping, rotation, and zooming transformations create different looks of identical cellular structures while maintaining their original shapes. The dataset for model training becomes more diverse through these variations, which produce different orientations and scale changes.

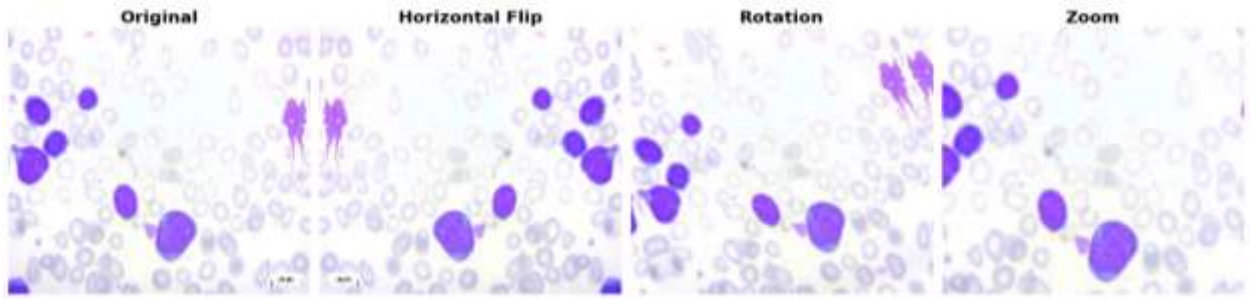


Figure 7: Data Augmentation

ResNet50 processes the augmented images to extract deep feature representations of blood cell structures from the images. The t-SNE visualization of the extracted feature vectors shows distinct clustering patterns that separate the benign, early, pre, and pro classes from each other, as shown in Figure 8. The model successfully identifies discriminative features for classification since the clusters exhibit clear separation from each other.

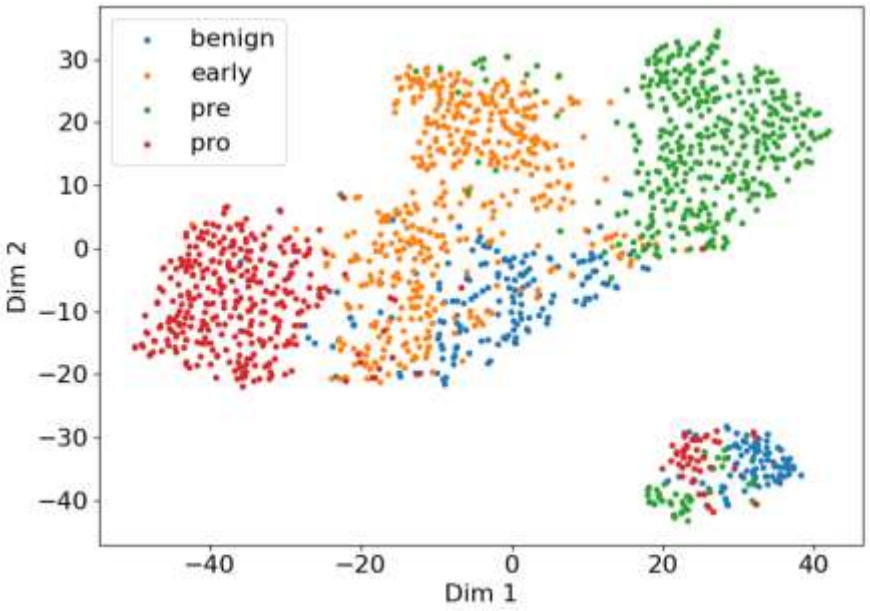


Figure 8: Feature Extraction using ResNet50

The extracted feature vectors through ResNet50 undergo analysis to determine their interdependencies and shared characteristics. The selected features display positive and negative correlations, which the correlation heatmap in Figure 9 demonstrates. The feature pairs show a strong positive correlation of 0.67 together with two feature pairs that exhibit negative correlations that reach −0.39 to show redundant information. The existing redundant information requires dimensionality reduction to achieve a more effective and compact feature representation.

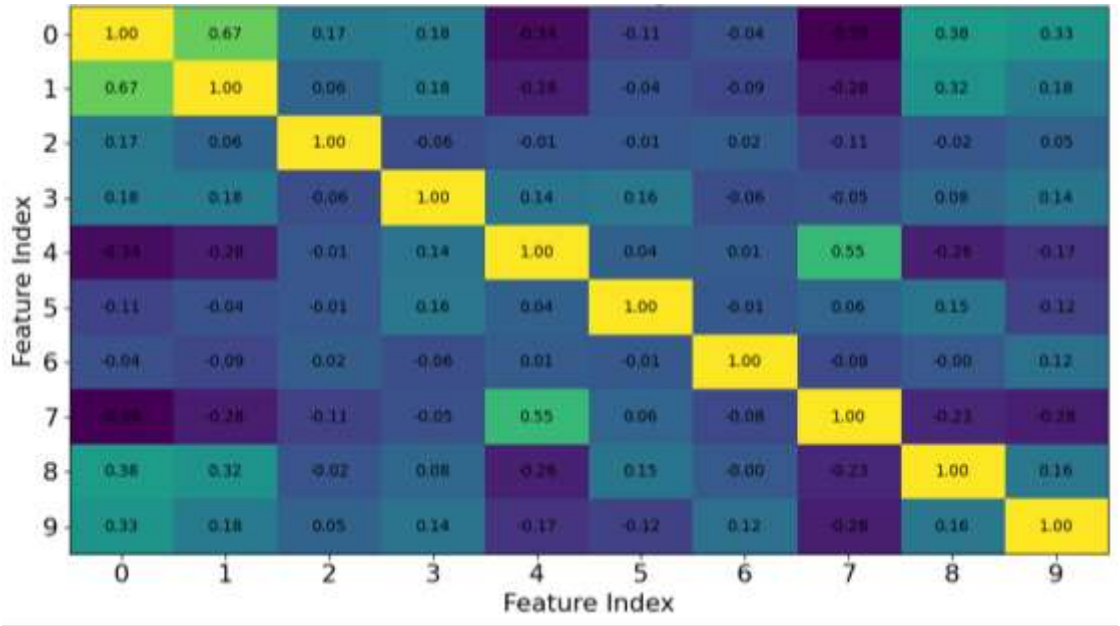


Figure 9: Feature Correlation Analysis

The extracted feature vectors require additional normalization processes, which standardize their value distribution. Figure 10 shows the distribution of features that existed before Z-score normalization compared to their distribution after the normalization process. The feature values after normalization display a distribution pattern that centers around zero and extends between −3 and +3. The standardization process decreases the impact of different feature scales while it enhances the consistency of the learning procedure.

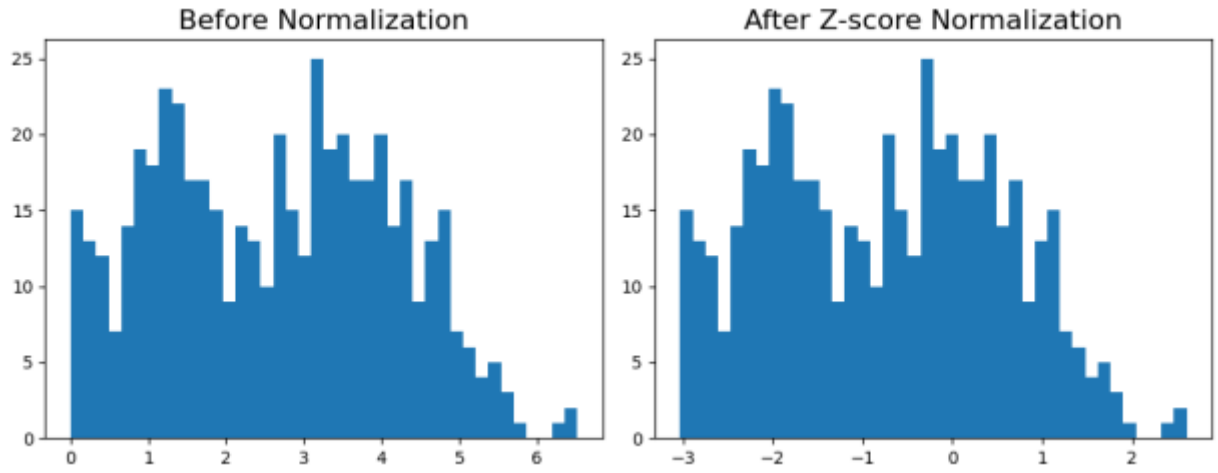


Figure 10: Feature Normalization (Z-score)

The normalized feature vectors undergo transformation through PCA, which reduces their dimensions while maintaining essential information. Figure 11 shows the feature space that has been reduced through PCA, which uses principal components PC1 and PC2 to display the data. The data points create compact distributions that show partial separation between different classes. The reduced spread and structured grouping show that PCA successfully identifies the major data variance, which results in better feature representation through its most important data points.

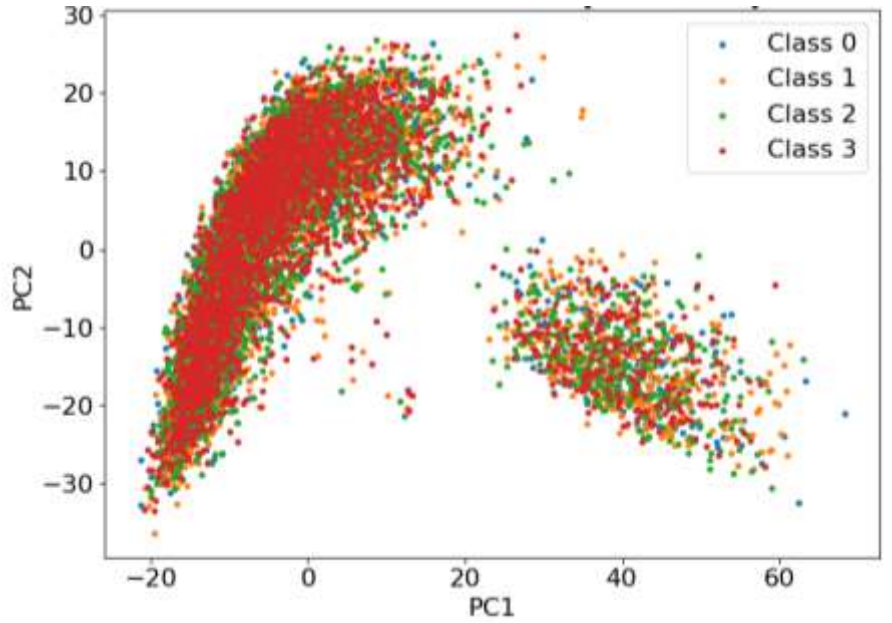


Figure 11: PCA-Based Feature Representation

The confusion matrices for the ViT, Swin Transformer, and proposed ensemble model show their classification results across four classes in Figure 12. The ViT model shows various misclassifications through its results in Figure 12a, which shows that the model misidentifies 13 benign samples as early and 15 samples as pro. Figure 12b shows better results because there are fewer mistakes, which include only 5 benign samples that the

system misclassified early. The ensemble model in Figure 12c demonstrates its best results through accurate classification of most samples, which includes only 2 benign samples that the model misidentified. The ensemble method shows its success through this continuous decrease in misclassification.

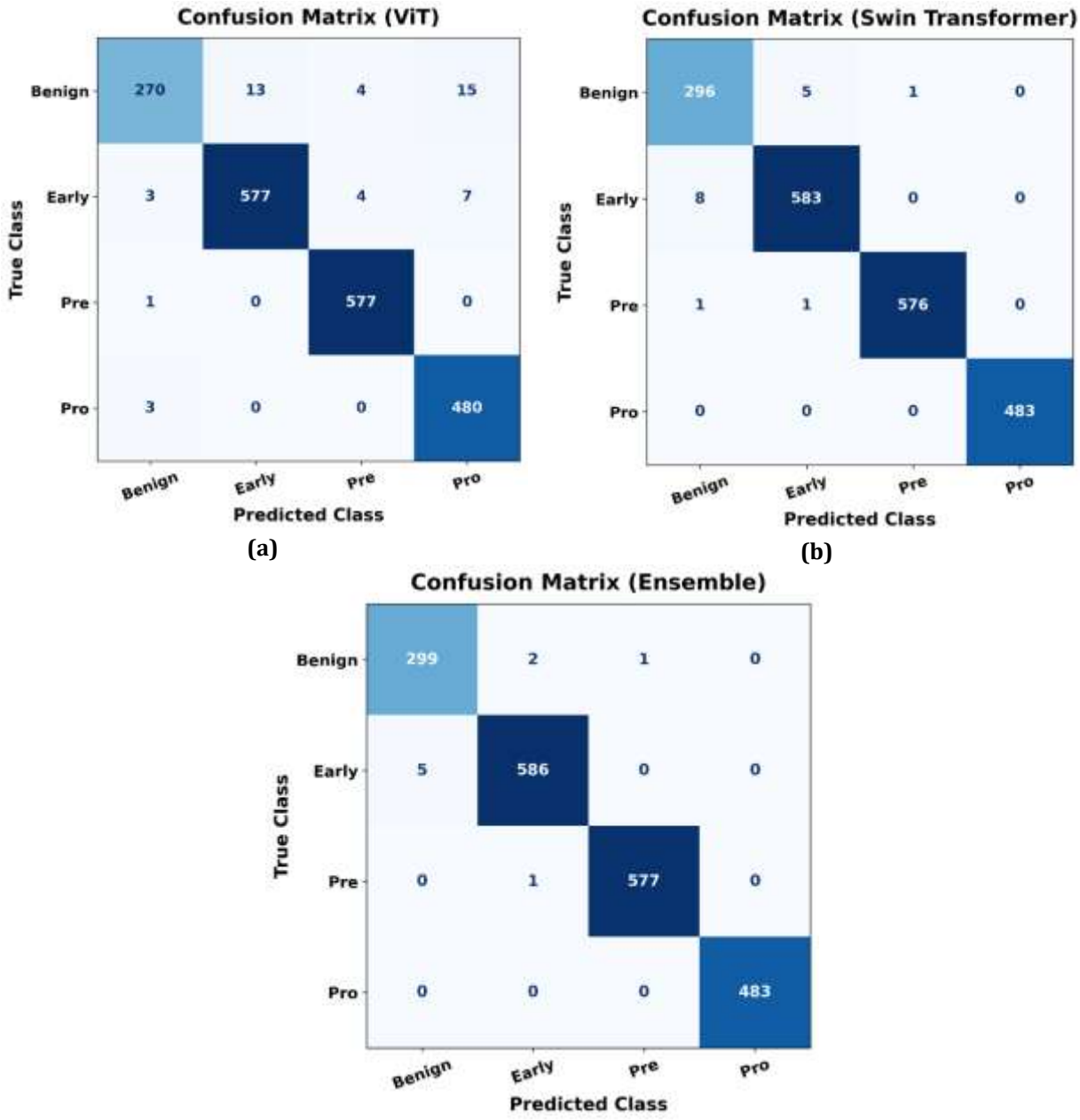


Figure 12: Confusion Matrices

The classification performance results of the Swin Transformer and ViT, and the proposed ensemble model appear in Figure 13 through precision metrics, recall metrics, F1-score metrics, and accuracy metrics across all class categories. The Swin Transformer shows high performance according to Figure 13a because the system achieved 98.6% precision, and 99.0% recall, and 98.8% F1-score, and 99.54% accuracy for the benign class, while other classes achieved values that approached 100%. The ViT model shows comparable performance to the Swin Transformer according to Figure 13b, although some metric values show slightly lower results than the Swin Transformer. The ensemble model shows its best performance according to Figure 13c because it achieved 98.6% precision, and 99.0% recall, and 98.8% F1-score, and 99.54% accuracy for the benign class,

while the other classes received almost perfect results. The ensemble approach demonstrates its effectiveness by producing better results in classification tasks.

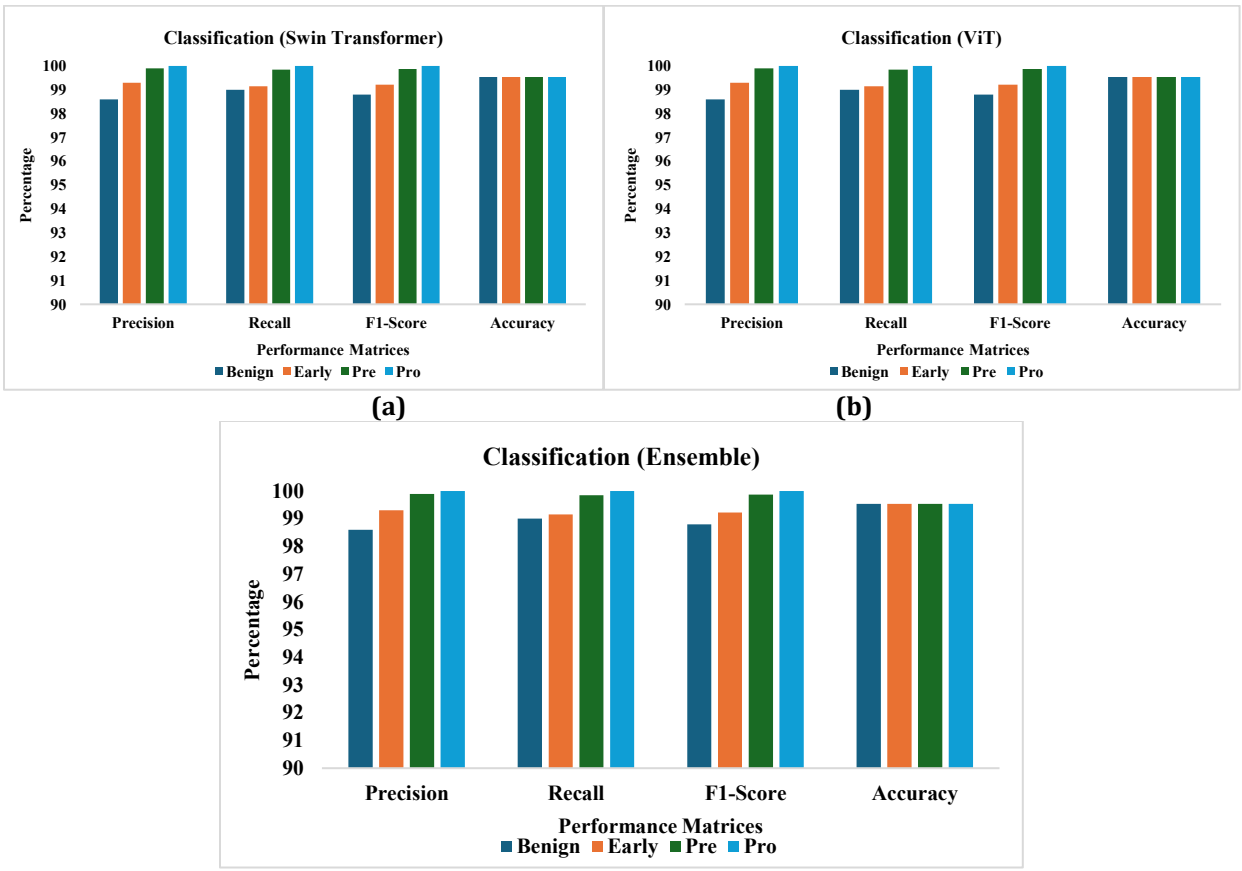
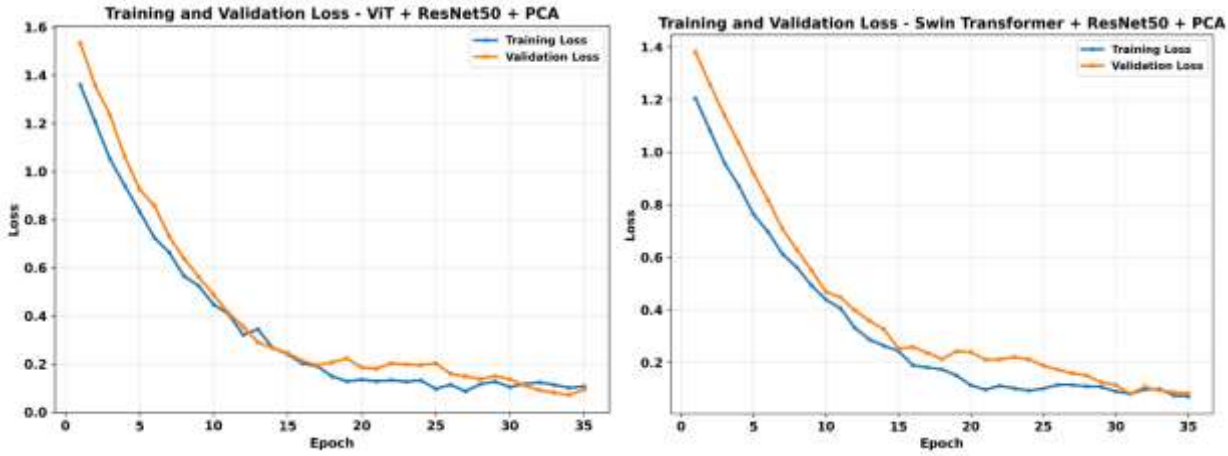
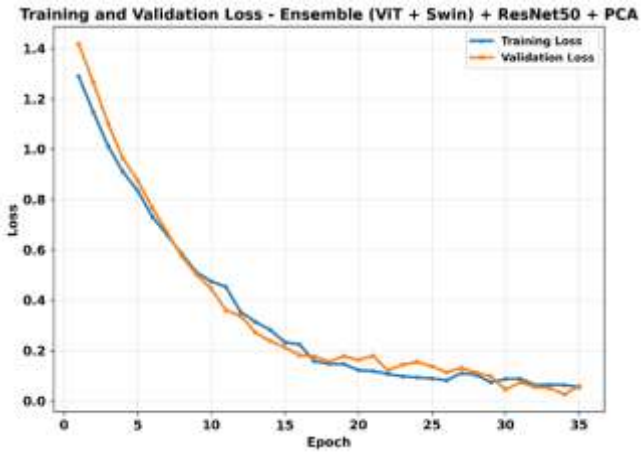


Figure 13: Classification Performance Metrics

The training and validation loss for (a) ViT + ResNet50 + PCA, (b) Swin Transformer + ResNet50 + PCA, and (c) the proposed ensemble model are illustrated in Figure 14. The training and validation loss curves of Figure 14(a) show a continuous decline from approximately 1.5 to below 0.1, which demonstrates that effective learning occurred with small variations during the later training periods. The loss in Figure 14(b) shows a similar pattern to previous results because it decreases from approximately 1.4 to almost 0.08, which demonstrates better convergence and stability than ViT. The performance of Figure 14(c) shows its best stable performance because both training and validation loss decrease from approximately 1.4 to almost 0.05, with the two curves maintaining a close distance during the process. The ensemble model achieves better convergence and generalization performance because its training and validation losses show consistent reduction while maintaining close alignment.



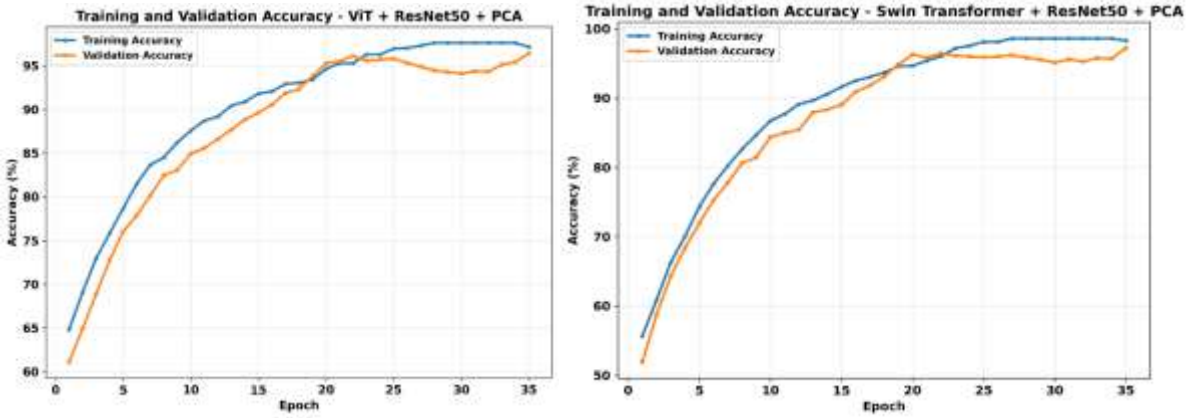
(a) **(b)**



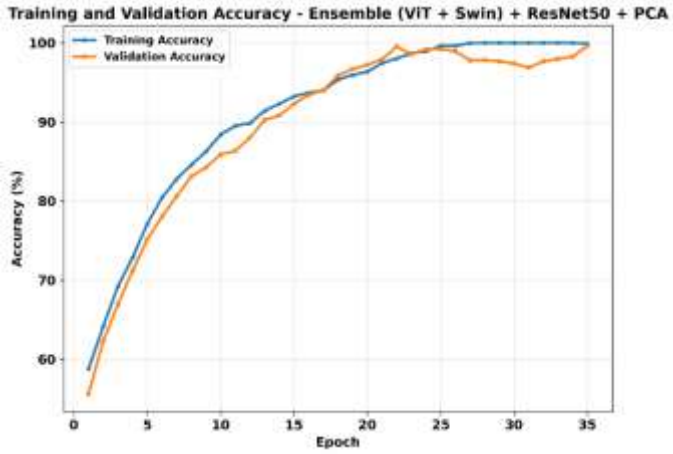
(c)

Figure 14: Training and Validation Loss

The training and validation accuracy show the performance of (a) ViT, (b) Swin Transformer, and (c) the proposed ensemble model throughout their training process. The ViT model shows a continuous accuracy improvement from 65% to 97% according to Figure 15(a) while maintaining a small accuracy gap between its training and validation results. The Swin Transformer shows better performance in Figure 15(b) because it achieves around 98% accuracy while showing dependable performance throughout its training process. The ensemble model shows its highest accuracy performance in Figure 15(c) through its exceptional 99% accuracy achievement, which presents almost no difference between training and validation results. The method shows effective learning ability because its performance improves continuously while maintaining strong generalization skills.



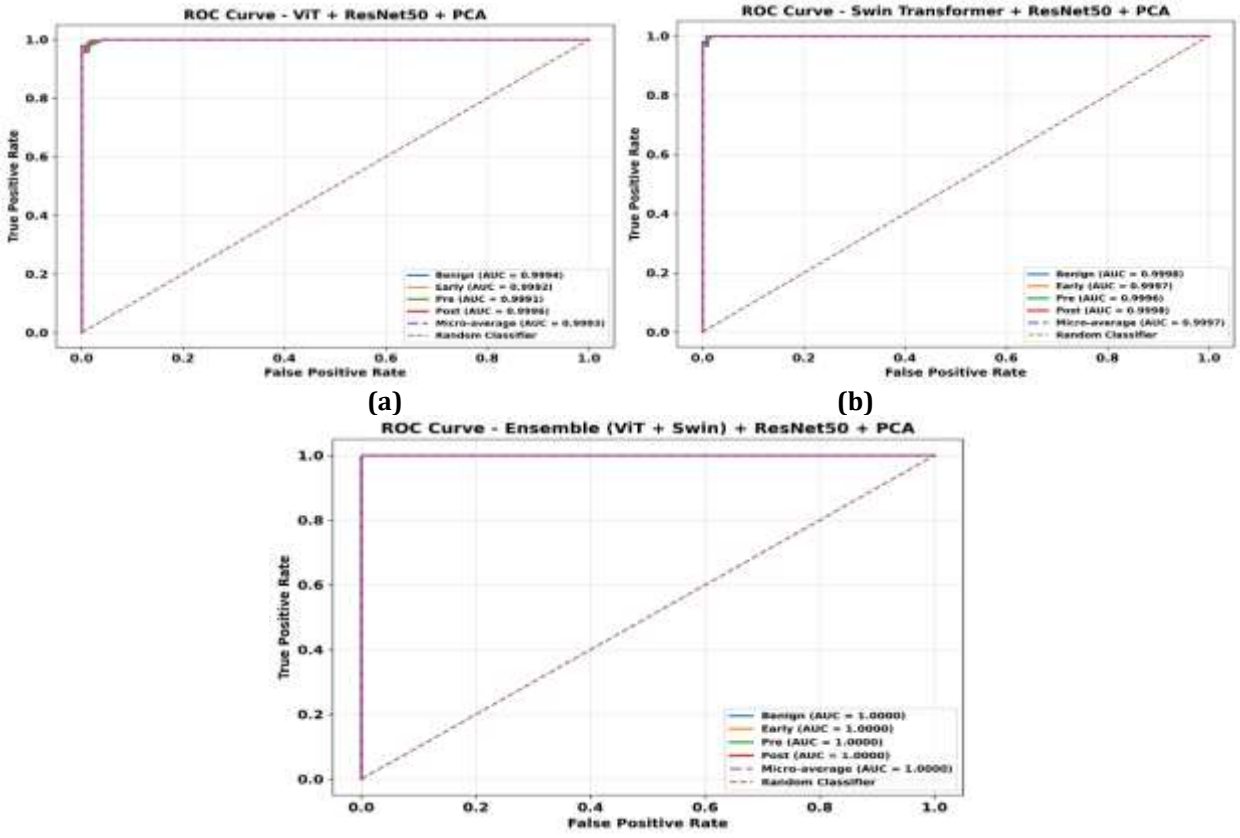
(a) **(b)**



(c)

Figure 15: Training and Validation Accuracy

Figure 16 displays the ROC curves for three models, which include (a) ViT, (b) Swin Transformer, and (c) the proposed ensemble model, to demonstrate their classification results through their TP rate and FP rate metrics. The ViT model in Figure 16(a) achieves excellent AUC results because it scored AUC values of 0.9994 for benign, 0.9992 for early, 0.9991 for pre, and 0.9996 for pro classes together with a micro-average AUC of 0.9993. The Swin Transformer shows improved results in Figure 16(b) because it achieves AUC scores of 0.9998 for benign and 0.9997 for early and 0.9996 for pre and 0.9998 for pro categories while achieving a micro-average AUC of 0.9997. The ensemble model shows its highest performance in Figure 16(c) because it achieves a perfect AUC of 1.0000 for all classes and the micro-average. The results demonstrate that the proposed ensemble model possesses a better ability to differentiate between different classes than other models.



(a) **(b)**

(c)

Figure 16: ROC Curve Analysis

The ablation study in Table 2 demonstrates how various model configurations affect classification results, which show that ResNet50 and PCA both contribute to classification accuracy. The baseline ViT and Swin Transformer models achieve accuracies of 94.85% and 96.38%, respectively. The ResNet50 integration leads to improved results, which bring ViT accuracy to 96.04% and Swin Transformer accuracy to 97.69%. The Swin Transformer achieves an accuracy of 99.18% through PCA implementation, which creates better results. The proposed ensemble model, which operates with ResNet50 and PCA, delivers optimal results by achieving an accuracy of 99.54%, precision of 99.45%, recall of 99.50%, and F1-score of 99.47%. The system performance enhancement process shows how feature extraction methods work together with dimensionality reduction techniques to enhance system effectiveness.

Table 2: Ablation Study

Model Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ViT	94.85	94.62	94.10	94.36
Swin Transformer	96.38	96.15	95.82	95.98
Ensemble (Swin + ViT)	97.12	96.98	96.75	96.86
ViT + ResNet50	96.04	95.87	95.35	95.60
Swin Transformer + ResNet50	97.69	97.51	97.28	97.39
Ensemble (Swin + ViT) + ResNet50	98.42	98.27	98.10	98.18
ViT + ResNet50 + PCA	97.44	97.38	96.56	96.91
Swin Transformer + ResNet50 + PCA	99.18	98.97	99.08	99.02
Ensemble (Swin + ViT) + ResNet50 + PCA	99.54	99.45	99.50	99.47

The proposed model comparison study in Table 3 assesses how existing methods perform against the proposed method. Hita et al. reached an accuracy level of 98.65% in 2026 through their use of VGG16. Jain et al. achieved a 92.50% accuracy rate in 2025 through their implementation of a Swin Transformer system, which combined MoCo and bi-directional encoder transformer technology. Anand et al. used an optimized CNN system with Adam and Adamax hyperparameter tuning methods to achieve a 96.00% accuracy rate in 2024. Atteia et al. developed a combined PCA PSO CNN Bayesian SVM system, which achieved a 97.40% accuracy result in 2023. Sampathila et al. used DL methods in 2022 to achieve an accuracy result of 95.54%. The suggested technique achieves an accuracy of 99.54%, which surpasses all existing methods through its successful combination of Swin Transformer, ViT, ResNet50, and PCA.

Table 3: Comparative Analysis

Author	Year	Technique	Accuracy (%)
Hita et al. [30]	2026	VGG16	98.65
Jain et al. [32]	2025	Swin Transformer + MoCo + Bi-directional Encoder Transformer	92.50
Anand et al. [37]	2024	Optimized CNN with Adam/Adamax hyperparameter tuning	96.00
Atteia et al. [40]	2023	Hybrid PCA + PSO + CNN + Bayesian SVM	97.40
Sampathila et al. [42]	2022	DL	95.54
Proposed Method	2026	Ensemble (Swin + ViT)	99.54

Conclusion

Blood cancer serves as a critical global health issue because early and precise detection methods establish essential pathways for enhancing patient outcomes. The existing studies face multiple challenges, which include data imbalance problems, insufficient feature representation, and their inability to handle multi-stage classification tasks. The research introduces a hybrid DL framework that enables early blood cancer detection and subtype classification through analysis of microscopic blood smear images. The approach combines data pre-processing and augmentation with deep feature extraction that uses ResNet50 for subsequent dimensionality reduction through PCA. The system uses both Swin Transformer and ViT classification techniques together with their combined model to achieve superior results through advanced feature extraction. The experimental outcomes show that the proposed method demonstrates outstanding classification accuracy, with its ensemble model achieving 99.54% accuracy, together with exceptional precision, recall, and F1-score performance. The ablation study shows that ResNet50 and PCA integration

enhance both feature representation and classification performance. The results show that the proposed system provides accurate and trustworthy automatic blood cancer diagnosis support. The model would expand its capabilities through future research, which would develop larger datasets with diverse samples and use explainable AI methods to enhance understanding and practical application.

References

[1]. Sharma, Raja, Satendra Kumar, Niranjan Kaushik, and Bahadur Singh. "Decoding the Mystery of Blood Cancer: Cause, Diagnosis, and Management." Current Cancer Therapy Reviews 21, no. 1 (2025): 40-53. https://doi.org/10.2174/0115733947263279231114111550

[2]. Shehta, Amr I., Mona Nasr, and Alaa El Din M. El Ghazali. "Blood cancer prediction model based on deep learning technique." *Scientific Reports* 15, no. 1 (2025): 1889. https://doi.org/10.1038/s41598-024-84475-0

[3]. Hemalatha, N., M. Sowjanya, and Y. Prapurna Chandra. "A review on leukemia: advances in diagnosis, treatment and prognosis." Journal of Innovations in Applied Pharmaceutical Science (JIAPS) (2025): 1-6. https://doi.org/10.37022/jiaps.v10i1.650

[4]. Liu, Xueru, Juan Zhang, Ting Yi, Hui Li, Xing Tang, Dan Liu, Daichao Wu, and Yukun Li. "Decoding tumor angiogenesis: pathways, mechanisms, and future directions in anti-cancer strategies." Biomarker Research 13, no. 1 (2025): 62. https://doi.org/10.1186/s40364-025-00779-x

[5]. Saha, Sayan, Shreya Ghosh, Suman Ghosh, and Sumit Nandi. "A Comprehensive Review on Unravelling the Complexity of Blood Cancer." blood 6 (2024): 7. https://doi.org/10.47583/ijpsrr. 2024.v84i03.020

[6]. Leuraud, Klervi, Dominique Laurier, Michael Gillies, Richard Haylock, Kaitlin Kelly-Reif, Stephen Bertke, Robert D. Daniels et al. "Leukaemia, lymphoma, and multiple myeloma mortality after low-level exposure to ionising radiation in nuclear workers (INWORKS): updated findings from an international cohort study." The Lancet Haematology 11, no. 10 (2024): e761-e769. https://doi.org/10.1016/S2352-3026(24)00240-0

[7]. Farshi, Esmaeil. "Comprehensive overview of 31 types of cancer: Incidence, categories, treatment options, and survival rates." International Journal of Gastroenterology and Hepatology 5, no. 3 (2024): 1-11. Comprehensive Overview Of 31 Types Of Cancer: Incidence, Categories, Treatment Options, And Survival Rates

[8]. Clough, Sydney, Matthew Wheeler, James Stanley, Virginia Signal, Myra Ruka, Jonathan Koea, and Jason Gurney. "Blood cancer incidence, mortality and survival for Māori in New Zealand." Cancer epidemiology 93 (2024): 102656. https://doi.org/10.1016/j.canep.2024.102656

[9]. Ghaderzadeh, Mustafa, Mehrad Aria, Azamossadat Hosseini, Farkhondeh Asadi, Davood Bashash, and Hassan Abolghasemi. "A fast and efficient CNN model for B-ALL diagnosis and its subtypes classification using peripheral blood smear images." International Journal of Intelligent Systems 37, no. 8 (2022): 5113-5133. https://doi.org/10.1002/int.22753

[10]. Ananth, Christo, K. Mystica, M. Sridharan, S. Aaron James, and T. Ananth Kumar. "An Advanced Low-cost Blood Cancer Detection System." International Journal of Early Childhood Special Education 14, no. 5 (2022). https://doi.org/10.9756/INTJECSE/V14I5.475

[11]. Ahad, Md Taimur, Sajib Bin Mamun, Sumaya Mustofa, Bo Song, and Yan Li. "A comprehensive study on Blood Cancer detection and classification using Convolutional Neural Network." arXiv preprint arXiv:2409.06689 (2024). https://doi.org/10.48550/arXiv.2409.06689

[12]. Yan, Geng, Gao Mingyang, Shi Wei, Liang Hongping, Qin Liyuan, Liu Ailan, Kong Xiaomei, Zhao Huilan, Zhao Juanjuan, and Qiang Yan. "Diagnosis and typing of leukemia using a single peripheral blood cell through deep learning." Cancer Science 116, no. 2 (2025): 533-543. https://doi.org/10.1111/cas.16374

[13]. Karar, Mohamed Esmail, Bandar Alotaibi, and Munif Alotaibi. "Intelligent medical IoT-enabled automated microscopic image diagnosis of acute blood cancers." Sensors 22, no. 6 (2022): 2348. https://doi.org/10.3390/s22062348

[14]. Tripti R, Kulkarni, Gururaj Bharathi, and Jaiswal Aditi. "Comparison of Implementation in Blood Cancer Causes and Diseases." International Journal of Trend in Scientific Research and Development 9, no. 1 (2025): 503-512. http://eprints.umsida.ac.id/id/eprint/15300

[15]. Cheng, Wangxinjun, Jingshuang Liu, Chaofeng Wang, Ruiyin Jiang, Mei Jiang, and Fancong Kong. "Application of image recognition technology in pathological diagnosis of blood smears." Clinical and Experimental Medicine 24, no. 1 (2024): 181. https://doi.org/10.1007/s10238-024-01379-z

[16]. Waggiallah, Hisham Ali, Abdulkareem Al-Garni, Aisha Ali M. Ghazwani, Abdulkarim S. Bin Shaya, Humood Al Shmrany, and Yousif Mohammed Almosaad. "The future insights of AI Applications in Hematology diseases diagnosis and prognosis." (2025). doi: https://doi.org/10.56294/saludcyt20251430
[17]. Alshdaifat, Nawaf, Hamza Abu Owida, Zaid Mustafa, Ahmad Aburomman, Suhaila Abuowaida, Abdullah Ibrahim, and Wafa Alsharafat. "Automated blood cancer detection models based on efficientnet-b3 architecture and transfer learning." Indonesian Journal of Electrical Engineering and Computer Science 36, no. 3 (2024): 1731-1738. https://doi.org/10.11591/ijeecs.v36.i3.pp1731-1738
[18]. Oybek Kizi, Rakhmonalieva Farangis, Tagne Poupi Theodore Armand, and Hee-Cheol Kim. "A review of deep learning techniques for leukemia cancer classification based on blood smear images." Applied Biosciences 4, no. 1 (2025): 9. https://doi.org/10.3390/applbiosci4010009
[19]. Shen, Dinggang, Guorong Wu, and Heung-Il Suk. "Deep learning in medical image analysis." Annual review of biomedical engineering 19, no. 1 (2017): 221-248. https://doi.org/10.1146/annurev-bioeng-071516-044442
[20]. Li, Mengfang, Yuanyuan Jiang, Yanzhou Zhang, and Haisheng Zhu. "Medical image analysis using deep learning algorithms." Frontiers in public health 11 (2023): 1273253. https://doi.org/10.3389/fpubh.2023.1273253
[21]. Tsuneki, Masayuki. "Deep learning models in medical image analysis." Journal of Oral Biosciences 64, no. 3 (2022): 312-320. https://doi.org/10.1016/j.job.2022.03.003
[22]. Li, Xiang, Minglei Li, Pengfei Yan, Guanyi Li, Yuchen Jiang, Hao Luo, and Shen Yin. "Deep learning attention mechanism in medical image analysis: Basics and beyonds." International Journal of Network Dynamics and Intelligence (2023): 93-116. https://doi.org/ 10.53941/ijndi0201006
[23]. Baig, Raheel, Abdur Rehman, Abdullah Almuhaimeed, Abdulkareem Alzahrani, and Hafiz Tayyab Rauf. "Detecting malignant leukemia cells using microscopic blood smear images: a deep learning approach." Applied Sciences 12, no. 13 (2022): 6317. https://doi.org/10.3390/app12136317
[24]. Abed, Amira Hassan, and Moshira A. Ebrahim. "Blood Cancer Detection from Blood Smear Images using Machine Learning Algorithms." SciNexuses 1 (2024): 160-173. https://doi.org/10.61356/j.scin.2024.1509
[25]. Xu, Yan, Rixiang Quan, Weiting Xu, Yi Huang, Xiaolong Chen, and Fengyuan Liu. "Advances in medical image segmentation: A comprehensive review of traditional, deep learning and hybrid approaches." Bioengineering 11, no. 10 (2024): 1034. https://doi.org/10.3390/bioengineering11101034
[26]. Abed, Amira Hassan. "Classifying Blood Cancer from Blood Smear Images using Artificial Intelligence Algorithms." Int. J. Advanced Networking and Applications 16, no. 05 (2025): 6602-6614. V16I5_10-libre.pdf
[27]. Thakur, Tanima, Isha Batra, Arun Malik, Deepak Ghimire, Seong-Heum Kim, and ASM Sanwar Hosen. "RNN-CNN based cancer prediction model for gene expression." IEEE access 11 (2023): 131024-131044. https://doi.org/10.1109/ACCESS.2023.3332479
[28]. Khan, S., Sajjad, M., Hussain, T., Ullah, A. and Imran, A.S., 2020. A review on traditional machine learning and deep learning models for WBCs classification in blood smear images. Ieee Access, 9, pp.10657-10673. https://doi.org/10.1109/ACCESS.2020.3048172
[29]. Deo, Arpit, Ish Pandey, Safdar Sardar Khan, Aditi Mandlik, Nitika Vats Doohan, and Bhupendra Panchal. "Deep learning-based red blood cell classification for sickle cell anemia diagnosis using hybrid CNN-LSTM model." Traitement du signal 41, no. 3 (2024): 1293. https://doi.org/10.18280/ts.410318
[30]. Hita, Xhesina, Farrukh Javed, and Stefano Lodi. "Reliable leukemia detection via transfer-enhanced Bayesian CNNs." Computers in Biology and Medicine 202 (2026): 111419. https://doi.org/10.1016/j.compbiomed.2025.111419
[31]. Mollick, Md Abu Ahnaf, Md Mahfujur Rahman, D. M. Asadujjaman, Abdullah Tamim, Nosin Anjum Dristi, and Md Takbir Hossen. "Detection and Classification of Acute Lymphoblastic Leukemia Utilizing Deep Transfer Learning." arXiv preprint arXiv:2501.14228 (2025). https://doi.org/10.48550/arXiv.2501.14228
[32]. Jain, Eshika, Pratham Kaushik, Vinay Kukreja, Modafar Ati, Shanmugasundaram Hariharan, Vandana Ahuja, Abhishek Bhattacherjee, and Rajesh Kumar Kaushal. "Swin Transformer and Momentum Contrast (MoCo) in Leukemia Diagnostics: A New Paradigm in Al-Driven Blood Cell Cancer Classification." IEEE Access (2025). https://doi.org/10.1109/ACCESS.2025.3563256

[33]. El Houby, Enas MF. "Acute lymphoblastic leukemia diagnosis using machine learning techniques based on selected features." Scientific Reports 15, no. 1 (2025): 28056. https://doi.org/10.1038/s41598-025-12361-4
[34]. Galila, Mohamed A., Ibrahim M. El-Hasnony, Reham Reda Mostafa, and Sherif Barakat. "A Full Framework For Automatic Detection of Acute Lymphoblastic Leukaemia." Mansoura Journal for Computer and Information Sciences 20, no. 2 (2025): 17-37. https://doi.org/10.21608/mjcis.2025.358215.1010
[35]. Kasim, Sazzli, Sorayya Malek, JunJie Tang, Xue Ning Kiew, Song Cheen, Bryan Liew, Norashikin Saidon, Raja Ezman, and Raja Shariff. "Multi-class leukemia cell classification using hybrid deep learning and machine learning with CNN-based feature extraction." Scientific Reports 15, no. 1 (2025): 23782. https://doi.org/10.1038/s41598-025-05585-x
[36]. Anand, Vatsala, Prabhnoor Bachhal, Deepika Koundal, and Arvind Dhaka. "Deep learning model for early acute lymphoblastic leukemia detection using microscopic images." Scientific Reports 15, no. 1 (2025): 29147. https://doi.org/10.1038/s41598-025-13080-6
[37]. Mustapha, Mubarak Taiwo, and Dilber Uzun Ozsahin. "Morphological Analysis and Subtype Detection of Acute Myeloid Leukemia in High-Resolution Blood Smears Using ConvNeXT." AI 6, no. 3 (2025): 45. https://doi.org/10.3390/ai6030045
[38]. Haque, Rezaul, Abdullah Al Sakib, Md Forhad Hossain, Fahadul Islam, Ferdaus Ibne Aziz, Md Redwan Ahmed, Somasundar Kannan, Ali Rohan, and Md Junayed Hasan. "Advancing early leukemia diagnostics: A comprehensive study incorporating image processing and transfer learning." BioMedInformatics 4, no. 2 (2024): 966-991. https://doi.org/10.3390/biomedinformatics4020054
[39]. Singh, Sanjay Kumar, Mamoon Rashid, Sultan S. Alshamrani, Mrim M. Alnfiai, Pranshu Saxena, and Aditya Khamparia. "Efficient Transfer Learning Approach for Acute Lymphoblastic Leukemia Diagnosis: https://doi.org/10.18280/ts.410409
[40]. Atteia, Ghada, Rana Alnashwan, and Malak Hassan. "Hybrid feature-learning-based PSO-PCA feature engineering approach for blood cancer classification." Diagnostics 13, no. 16 (2023): 2672. https://doi.org/10.3390/diagnostics13162672
[41]. Sheet, Joydeb, Chanchal Ghosh, and Biplab Kanti Das. "Deep learning-based transfer learning for the detection of leukemia." In 2023 International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC), pp. 1-6. IEEE, 2023. **DOI:** 10.1109/ISACC56298.2023.10084138
[42]. Sampathila, Niranjana, Krishnaraj Chadaga, Neelankit Goswami, Rajagopala P. Chadaga, Mayur Pandya, Srikanth Prabhu, Muralidhar G. Bairy, Swathi S. Katta, Devadas Bhat, and Sudhakara P. Upadya. "Customized deep learning classifier for detection of acute lymphoblastic leukemia using blood smear images." In Healthcare,vol. 10, no. 10, p. 1812. MDPI, 2022. https://doi.org/10.3390/healthcare10101812
[43]. https://www.kaggle.com/datasets/mehradaria/leukemia
[44]. Maharana, Kiran, Surajit Mondal, and Bhushankumar Nemade. "A review: Data pre-processing and data augmentation techniques." Global Transitions Proceedings 3, no. 1 (2022): 91-99. https://doi.org/10.1016/j.gltp.2022.04.020
[45]. Wang, Zaitian, Pengfei Wang, Kunpeng Liu, Pengyang Wang, Yanjie Fu, Chang-Tien Lu, Charu C. Aggarwal, Jian Pei, and Yuanchun Zhou. "A comprehensive survey on data augmentation." IEEE Transactions on Knowledge and Data Engineering (2025). https://doi.org/10.48550/arXiv.2405.09591
[46]. Huang, Sheng-Yao, Wen-Lin Hsu, Ren-Jun Hsu, and Dai-Wei Liu. "Fully convolutional network for the semantic segmentation of medical images: A survey." Diagnostics 12, no. 11 (2022): 2765. https://doi.org/10.3390/diagnostics12112765
[47]. Worsito, Singgih Bekti, and Suree Pumrin. "Image-Based White Blood Cell Subtype Detection Using ResNet-50 Convolutional Neural Network (CNN)[C]." In The 14th Regional Conference on Electrical and Electronics Engineering (RC-EEE 2021), p. 160. 2022. rc-eee2021-proceeding.indd
[48]. Kanimozhi, T., V. Rajeswari, R. Suguna, J. Nirmaladevi, P. Prema, B. Janani, and R. Gomathi. "RWHO: A hybrid of CNN architecture and optimization algorithm to predict basal cell carcinoma skin cancer in dermoscopic images." The Scientific Temper 15, no. 02 (2024): 2138-2142. https://doi.org/10.18174/680243
[49]. Hasan, Basna Mohammed Salih, and Adnan Mohsin Abdulazeez. "A review of principal component analysis algorithm for dimensionality reduction." Journal of Soft Computing and Data Mining 2, no. 1 (2021):20-30. https://doi.org/10.30880/jscdm.2021.02.01.003

[50]. Liu, Ze, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. "Swin transformer: Hierarchical vision transformer using shifted windows." In Proceedings of the IEEE/CVF international conference on computer vision, pp. 10012-10022. 2021. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows

[51]. Fan, Fan, Xiangfeng Zeng, Shunjun Wei, Hao Zhang, Dianhua Tang, Jun Shi, and Xiaoling Zhang. "Efficient instance segmentation paradigm for interpreting SAR and optical images." Remote Sensing 14, no. 3 (2022): 531. https://doi.org/10.3390/rs14030531

[52]. Amin, Sifatul, M. D. Jawed, M. D. Rashed Raj, M. D. Ahmed Saimoon, and M. D. Rayhan. "Vision transformer (ViT) approach in computer aided diagnosis of acute lymphoblastic leukemia." PhD., Brac University, 2023. http://hdl.handle.net/10361/20230
https://tintn.github.io/Implementing-Vision-Transformer-from-Scratch/