An Efficient Classification Technique of Acute Lymphoblastic Leukemia Cells

Hadeel Farag, Noura A.Semary, Khalid M.Amin, and Sondos Fadl

***Abstract*—An accurate and rapid leukemia diagnosis is essential due to the aggressive nature of the disease. Traditional methods relying on blood and bone marrow examination are time-consuming, error-prone, and heavily reliant on specialist experience. This study addresses these limitations by proposing a fast, high-resolution computer-aided method for classifying leukemia cells. The method utilizes pre-processing, data augmentation, and K-means clustering for image segmentation. Extracted features from a DenseNet-201 model are then fed into the random forest, extreme gradient boosting (XGBoost), and support vector machine (SVM) classifiers to categorize cells as Normal or acute lymphoblastic leukemia (ALL). We evaluated our method on two publicly available datasets: ALL-IDB2 (containing 260 images with an even distribution of Normal and ALL cells) and C-NMC-2019 (comprising 10,661 microscopic blood images: 7,272 ALL images from 47 patients and 3,389 Normal images from 26 healthy individuals). On C-NMC-2019, the DenseNet-201 and SVM combination achieved exceptional results, with 99.13% accuracy, 99.24% specificity, 99.01% sensitivity, 99.00% precision, 99.10% F1-score, and 99.96% AUC (area under the receiver operating characteristic (ROC) curve). Even more remarkably, on ALL-IDB2, the model achieved 100% across all metrics. Although this might indicate overfitting, the true test of the model lies in its generalizability. The outstanding performance across two datasets highlights the effectiveness and generalizability of the proposed method, surpassing the performance of existing well-established and state-of-the-art methods and suggesting its potential application in various medical diagnosis domains.**

***Index Terms*—Acute lymphoblastic leukemia, Computer-aided medical diagnosis, Deep learning, DenseNet-201, K-means clustering, Support vector machine (SVM)**

I. Introduction

CANCER, a disease characterized by uncontrolled cell growth, remains a significant global health burden. According to the World Health Organization (WHO), it was the second leading cause of death worldwide in 2018, with approximately 9.6 million deaths and 18.1 million new cases [1]. Statistics from the American Cancer Society (ACS) further emphasize the importance of early detection. In 2023 alone, they reported a substantial disparity between new cases and deaths among males (35,670 vs. 13,900) and females (23,940 vs. 9,810) [2]. These figures highlight the critical role of early diagnosis and treatment in combating cancer.

Among various types of cancer, leukemia—a blood cancer that affects the bone marrow and blood—stands out for its aggressiveness and widespread impact. It disrupts the production of healthy blood cells and weakens the immune system, making individuals more susceptible to infections. Leukemia arises from the uncontrolled proliferation of malignant white blood cells (WBCs) in the bone marrow, which hinders the production of red blood cells and platelets. These malignant cells can infiltrate organs such as the kidneys, liver, spleen, and brain, potentially leading to secondary cancers [3]. The World Health Organization (WHO) classifies leukemia into four main types: chronic myelogenous leukemia (CML), chronic lymphocytic leukemia (CLL), acute myeloid leukemia (AML), and acute lymphoblastic leukemia (ALL) [1].

Acute lymphoblastic leukemia (ALL), particularly concerning due to its rapid progression, necessitates prompt diagnosis and treatment for both children and adults. Traditional diagnosis relies on manual cell count and categorization by skilled personnel. However, this approach suffers from limitations such as time-consuming analysis and potential inconsistencies in accuracy due to subjective interpretation by the pathologist [4].

The demand for a fast, reliable, and cost-effective diagnostic method has driven the development of automated microscopy for blood sample analysis. This technology offers significant advantages in terms of speed and accuracy, complementing pathologists' assessments [5]. In recent decades, there has been a surge in computer-aided diagnostic

Manuscript received October 25, 2024; revised May 4,2025

H.Farag is a Teaching Assistant of the Department of Information Technology, Faculty of Computers and Information, Kafrelsheikh University, Kafrelsheikh, 33516, Egypt (corresponding author to provide phone: 00201555855836; e-mail: hadeel_farag@fci.kfs.edu.eg).

N. Semary is a Professor of the Department of Information Technology, Faculty of Computers and Information, Menoufia University, Shebin El-koom, 32511, Egypt (e-mail: noura.semary@ci.menofia.edu.eg).

K. Amin is a Professor of the Department of Information Technology, Faculty of Computers and Information, Menoufia University, Shebin El-koom, 32511, Egypt (e-mail: k.amin@ci.menofia.edu.eg).

S. Fadl is an Assistant Professor of the Department of Information Technology, Faculty of Computers and Information, Menoufia University, Shebin El-koom, 32511, Egypt (e-mail: sondos.magdy@ci.menofia.edu.eg).

(CAD) systems designed to differentiate normal and ALL cells, effectively addressing the limitations of manual diagnosis. Notably, the accessibility, speed, and affordability of computer-aided microscopy have revolutionized the field, reducing the need for specialized laboratory equipment in research settings [6].

Biomedical image processing has emerged as a powerful tool, leveraging computer algorithms for disease diagnosis [6]. For instance, in [7], skin cancer was classified into seven categories using machine learning (ML) techniques, while in [8], malaria was classified using deep learning (DL). Studies exploring ML and DL techniques have demonstrated promising results in classification, offering advantages in speed, simplicity, and accuracy [9], [10]. However, some existing methods face challenges such as computational inefficiency, suboptimal classification accuracy, and overfitting [9], [10]. In this work, we aim to address these limitations by proposing a novel hybrid technique for ALL diagnosis. Our key contributions are as follows:

1) *Efficient ALL diagnosis with high accuracy:* We propose a method that utilizes deep features to achieve highly accurate ALL detection on two publicly available blood cell classification datasets (ALL-IDB2 and C-NMC-2019).
2) *Reduced computational complexity:* By employing K-means clustering to isolate only the regions of interest (ROIs) in blood cell images, we significantly reduce processing time and computational burden compared to analyzing the entire image.
3) *Enhanced performance:* We leverage the Densenet-201 model for feature extraction and a support vector machine (SVM) for classification, surpassing the performance of existing well-established and state-of-the-art methods.

The remainder of the paper is structured as follows. Section II reviews relevant research on leukemia diagnosis and existing CAD methods. Section III details the proposed technique, encompassing image acquisition, preprocessing, segmentation, augmentation, feature extraction, and classification. Section IV presents and analyzes the experimental findings. Finally, Section V summarizes the key takeaways and potential future directions.

II. Related work

The increasing availability of medical image datasets has driven the application of machine learning (ML) and deep learning (DL) techniques in disease diagnosis, including leukemia detection [11], [12]. However, a major challenge remains the limited size and public accessibility of leukemia image datasets.

Several researchers have focused on developing automated methods for leukemia detection. Early approaches primarily relied on traditional machine learning techniques, such as k-means clustering. For example, Bouzid et al. [13] proposed a method that uses k-means clustering for the segmentation and classification of blood cells to detect leukemia. Their approach was tested on a relevant dataset and achieved an accuracy of 98%. Moshavash et al. [14] employed white blood cell (WBC) segmentation and SVM classification on the ALL-IDB1 and ALL-IDB2 datasets, reaching an accuracy of 89.81%. Umamaheswari and Geetha [15] presented a method that used thresholding and mathematical operations for WBC nucleus segmentation, followed by KNN classification on the ALL-IDB2 dataset, achieving an accuracy of 96.25%. Recent advancements have led to the successful application of deep learning (DL) techniques, particularly convolutional neural networks (CNNs), in leukemia classification. Elrefaie et al. [16] utilized K-means clustering and feature extraction with empirical mode decomposition (EMD) analysis, followed by a neural network classifier on the ALL-IDB2 dataset, achieving an accuracy of 98.7%. Balasubramanian et al. [17] further demonstrated the effectiveness of deep learning by using a modified U-Net for segmentation and a radial basis function (RBF) kernel SVM (RBF-SVM) for classification, achieving 99.42% accuracy on the ALL-IDB2 dataset. Saxena et al. [18] investigated the impact of different segmentation methods on lung tumor classification using a deep learning network, with K-means clustering yielding the best results.

Several studies have further explored deep learning for leukocyte classification using the C-NMC-2019 dataset. De Oliveira et al. [19] achieved an F1-score of 92.60% by utilizing modified VGG16, VGG19, and Xception architectures with data augmentation techniques such as mirroring, rotation, blurring, shearing, and salt-and-pepper noise to balance the training and validation sets. Chayan Mondal et al. [20] applied data augmentation and preprocessing on C-NMC-2019, training an ensemble classifier using five pre-trained networks (Xception, VGG-16, DenseNet-121, MobileNet, and InceptionResNet-V2), achieving an F1-score of 89.7%, an accuracy of 88.3%, and an area under the receiver operating characteristic (AUC) of 95% on the initial test set. These studies highlight the effectiveness of deep learning techniques in blood cell classification.

Several other studies have addressed class imbalance and

feature extraction in blood cell classification using public datasets. Mohammed et al. [21] tackled the imbalanced C-NMC-2019 dataset by integrating convolutional neural networks (CNNs) with a gated recurrent unit (GRU)-bidirectional long short-term memory (BiLSTM) architecture to capture long-range dependencies and enhance feature learning. Their approach, incorporating softmax and multi-class SVM classifiers, achieved an accuracy of 96.29% and an F1-score of 96.23%. Similarly, Ahmed et al. [22] proposed an image enhancement method for microscopic blood images (ALL-IDB2 and C-NMC-2019) using filters and active contours. They extracted white blood cell regions and processed them with CNN models, reducing feature redundancy via principal component analysis (PCA). Deep feature maps from CNNs (DenseNet121, ResNet50, and MobileNet) were then combined to create hybrid deep feature maps for classification using random forest (RF) and extreme gradient boosting (XGBoost) classifiers. Their approach achieved remarkable results, attaining 98.8% accuracy on C-NMC-2019 and 100% accuracy on ALL-IDB2.

These studies demonstrate significant advancements in addressing data imbalance and feature extraction for blood cell classification, motivating our exploration of a high-resolution method for this task.

Other studies have explored pre-trained models and feature extraction techniques for leukemia cell classification. Renuka et al. [23] achieved an accuracy of 96.15% by utilizing features from the AlexNet model with an SVM classifier on an unspecified dataset. Rehman et al. [24] modified the AlexNet architecture and combined it with various classifiers (Naïve Bayes, KNN, SVM) using local binary pattern (LBP) features, achieving an accuracy of 97.78%. On the ALL-IDB1 dataset, TTP et al. [25] developed a CNN for both feature extraction and classification, reaching 96.43% accuracy. Prellberg and Kramer [26] employed a pre-trained ResNeXt50 model on the C-NMC-2019 dataset, achieving an F1-score of 88.91%. Ananthu et al. [27] compared various pre-trained models (Xception, InceptionV3, DenseNet201, ResNet50, MobileNet) on the ALL-IDB2 dataset, with the highest accuracy reaching 97.88%. These studies demonstrate the effectiveness of pre-trained models and different feature extraction methods for blood cell classification, motivating our investigation of a DenseNet-201 model combined with various classifiers for this task.

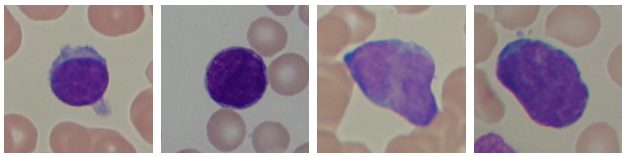
Recent studies highlight a growing preference for deep learning (DL) techniques in automatic feature extraction for leukemia cell classification (e.g., [25], [26], [27]). However, some methods continue to rely on manually crafted features (e.g., [14], [15]). Additionally, hybrid approaches that integrate manual and deep features have been explored (e.g., [23], [24]). While these techniques have shown effectiveness, they can be constrained by robustness issues, computational complexity, and susceptibility to overfitting, potentially limiting their real-time applicability.

Our proposed method addresses these limitations by introducing a novel hybrid approach that leverages the strengths of both machine learning (ML) and DL. This approach aims to achieve efficient and accurate ALL diagnosis on two publicly available blood cell classification datasets (ALL-IDB2 and C-NMC-2019). By demonstrating superior performance compared to existing methods, our model holds significant promise for real-world applications.

III. Materials and Method

A. Dataset description

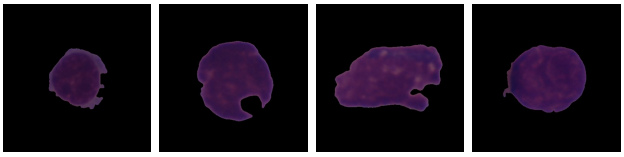
1) ALL-IDB2 dataset: This study utilizes the publicly available ALL-IDB2 dataset [28], which consists of 260 images evenly distributed between Normal and ALL cells (130 images each). The images were captured using a Canon PowerShot G5 camera coupled with an optical laboratory microscope, resulting in a resolution of 257×257×3 pixels. The images are stored in ".tif" format with a 24-bit color depth. Figure 1 presents sample images of both Normal and ALL cells from the dataset.



(a) (b) (c) (d)

Fig. 1: Sample images from the ALL-IDB2 dataset: (a) Normal cell, and (b), (c), and (d) ALL cell

2) C-NMC-2019 dataset: The C-NMC-2019 dataset, publicly available from The Cancer Imaging Archive (TCIA) [29], was used for this study. Originally released for the International Symposium on Biomedical Imaging (ISBI) competition on leukemia detection, it contains 10,661 microscopic blood images (450x450 pixels, 24-bit RGB resolution). Notably, these images depict single, pre-segmented cells from both ALL patients (7,272 images from 47 individuals) and normal individuals (3,389 images from 26 healthy individuals). This pre-segmentation ensures the analyzed cells are malignant or benign lymphocytes, confirmed by experienced oncologists (as shown in Fig. 2).



(a) (b) (c) (d)

Fig. 2: Sample images from the C-NMC-2019 dataset: (a) Normal cell and (b), (c), and (d) ALL cell

B. Framework of proposed method

Proposed Framework Our proposed method for ALL diagnosis follows a five-step process: preprocessing, segmentation, data augmentation, feature extraction, and classification (as illustrated in Fig. 3).

C. Preprocessing

Preprocessing is a crucial initial step in image processing tasks. It aims to enhance the quality of visual information within each image and remove artefacts to reveal hidden details. This, in turn, improves the model's final results and accuracy. In this study, the preprocessing stage comprises three key steps:

1) RGB to LAB conversion: Blood smear images from the ALL-IDB2 dataset undergo an initial conversion from the RGB color space to the LAB color space to facilitate segmentation. Segmentation in the RGB domain can be challenging due to variations in image quality and brightness. For instance, aging stains can significantly alter the colors and intensities of both blood cells and the background within the RGB spectrum.

The LAB color space offers distinct advantages for segmentation, as it separates luminance information (represented by the 'L' channel) from chrominance information (represented by the 'A' and 'B' channels). This separation simplifies segmentation by allowing focus on the 'A' and 'B' channels, which retain the essential color information needed to accurately isolate cells (Fig. 4). Notably, this conversion step is not applied to the C-NMC-2019 dataset, as it already contains pre-segmented cells.

2) Decoding and resizing: Each image in both datasets undergoes decoding and resizing. In convolutional neural networks (CNNs), each model architecture has a specific input image size optimized for network efficiency and reduced computational burden. Therefore, input images must be adjusted to match the target network's requirements.

Our proposed method employs the DenseNet-201 architecture for feature extraction. Compared to other architectures, DenseNet-201 has a lower learning capacity; however, this results in reduced computational costs, enabling the use of high-resolution input images with dimensions of 224×224×3. Consequently, all images in both datasets are resized to 224×224 pixels to ensure compatibility with the DenseNet-201 network.

3) Data normalization: Normalization aims to standardize the data by scaling all features to a similar range. This step enhances model performance and training stability. In essence, data normalization ensures all data points fall within a predefined range. Here, a simple normalization technique is employed: all pixel values are divided by 255, effectively rescaling them to the range of $[0,1]$.

D. Segmentation

Image segmentation plays a critical role in image analysis tasks like cell identification. Its primary objective is to partition an image and isolate specific ROIs containing the objects of interest (in our case, blood cells). These ROIs are then used for feature extraction, a crucial step in improving classification accuracy. During this stage, we evaluated various segmentation techniques commonly employed in deep learning and cell image analysis. K-means clustering emerged as the most efficient method for our proposed model compared to the alternatives. It is important to note that the C-NMC-2019 dataset is already pre-segmented, containing single-cell images. Therefore, the segmentation process was solely applied to the ALL-IDB2 dataset to isolate individual cells from each image, as shown in Fig. 5. This step ensures consistency in image format across both datasets for our analysis.

K-means clustering technique: K-means clustering is an unsupervised learning algorithm that groups data objects based on their inherent similarities. It excels at identifying clusters within unlabeled data. The core principle involves defining a pre-determined number of clusters (denoted by 'K') and assigning each data point to the closest cluster centroid (central point). We applied K-means clustering to the LAB color space representation of the images (converted from RGB) for segmentation. Specifically, we focused on the 'A' channel, which carries essential color information for cell isolation. The algorithm aims to minimize an objective function, represented by

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} \|x_n - t_k\|^2, \tag{1}$$

where J represents the distance between a pixel (x) and its assigned cluster center (t_k), N denotes the number of data points, and K is the number of clusters. In our case, we found that $K = 3$ yielded the optimal segmentation results compared to $K = 7$.

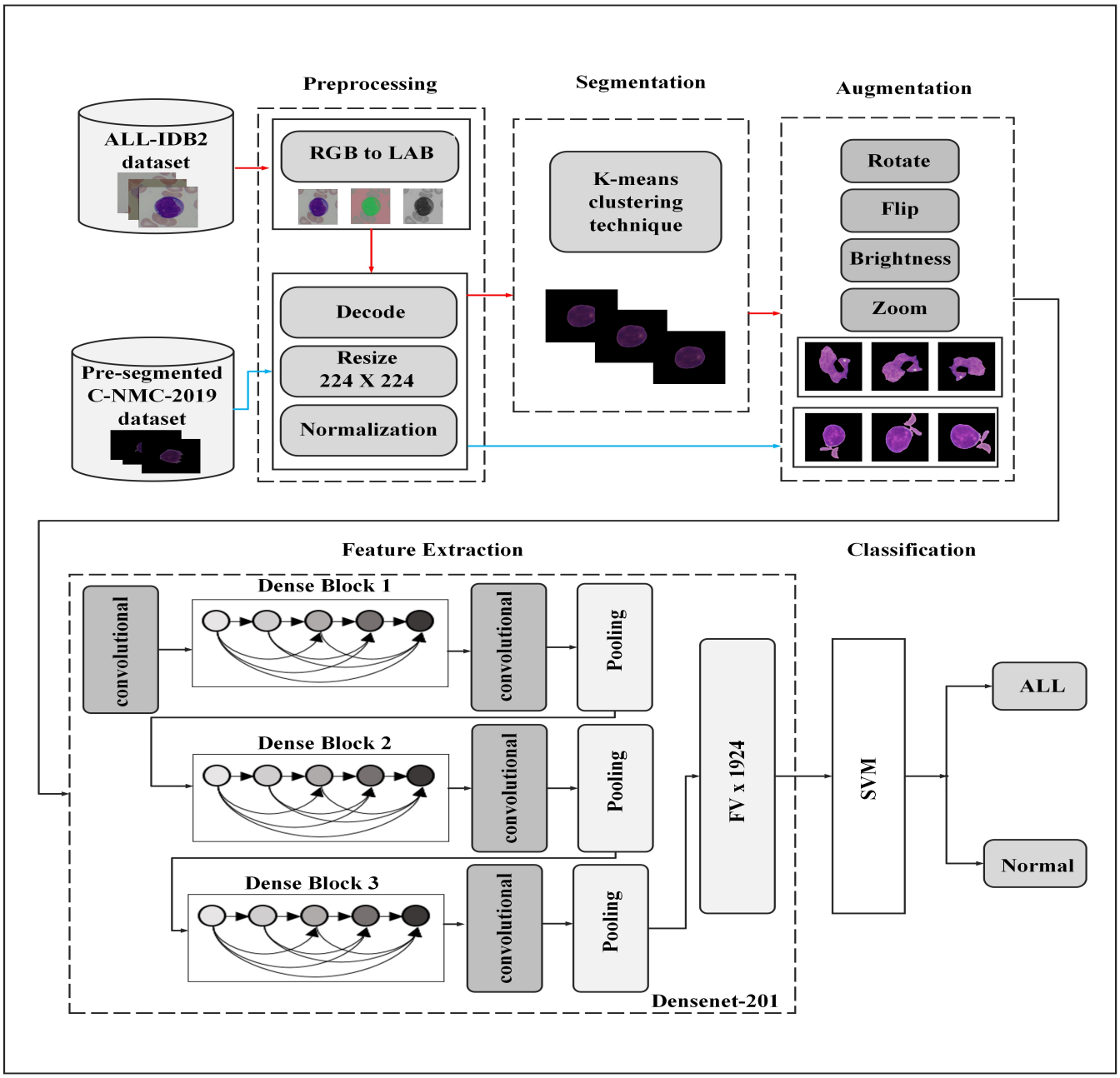
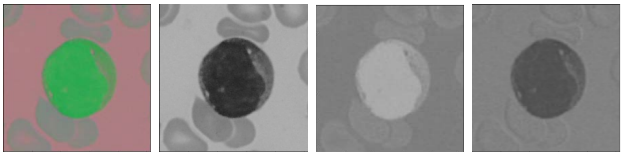
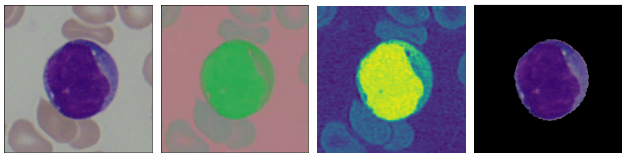


Fig. 3: The framework of the proposed method



(a) (b) (c) (d)

Fig. 4: Samples of (a) an LAB image, (b) its 'L' channel, (c) its 'A' channel, and (d) its 'B' channel



(a) (b) (c) (d)

Fig. 5: Segmentation process: (a) Original RGB cell, (b) Converted LAB image, (c) 'A' channel used for segmentation, (d) Segmented cell

Following K-means clustering, a binary threshold with values of 141 and 255 is applied to refine the segmentation and create well-defined cell regions (as shown in Fig. 5). Finally, the segmented cells are converted back to RGB format for the subsequent processing stage.

E. Augmentation

Data augmentation is a commonly used technique to artificially expand the size and diversity of a dataset. This is especially beneficial for DL models, which require vast amounts of data for optimal training and generalization.

Given the limitations of our relatively small datasets, data augmentation plays a crucial role in enhancing model performance. We employed various augmentation techniques on both datasets to introduce controlled variations that mimic real-world scenarios. These variations help the model learn robust features and improve its ability to identify cells across a broader range of appearances. The specific techniques used include:

- *Rotation augmentation (0° and 45°)*: Exposes the model to cells in different orientations, enhancing generalizability.
- *Flipping augmentation (horizontal and vertical)*: Introduces additional complexity by simulating real-world scenarios where cells may appear flipped.
- *Resizing augmentation (scaling factor 0.2)*: Ensures the model can effectively handle objects of varying sizes.
- *Brightness augmentation (factors of 0.2 and 1.0)*: Improves the model's robustness and adaptability to different lighting conditions.

Data augmentation addressed two key challenges in both datasets (as illustrated in Table I):

1) *Limited dataset size in ALL-IDB2*: This dataset initially contained only 260 images (130 ALL and 130 Normal). Augmentation expanded the dataset to 1,500 images (750 ALL and 750 Normal), as shown in Fig. 6.
2) *Class imbalance in C-NMC-2019*: This dataset exhibited a class imbalance, with significantly more ALL (7,272) than Normal (3,389) cells. Augmentation was applied to create a balanced dataset of 8,000 images per class (ALL and Normal), as shown in Fig. 7.

TABLE I: Number of samples before and after augmentation for both datasets

	Before Aug.		After Aug.	
Dataset	ALL	Normal	ALL	Normal
ALL-IDB2	130	130	750	750
C-NMC-2019	7,272	3,389	8,000	8,000

These augmentations effectively addressed the limitations of our datasets, improving the generalizability and robustness of the proposed model.

F. Feature extraction

Transfer learning is a widely adopted technique in ML and DL techniques, particularly for computer vision and natural language processing tasks. It leverages the knowledge gained by a pre-trained model on a large dataset and applies it to a new, related task. In our approach, we utilize transfer learning for feature extraction. Instead of training a DL model from scratch, we extract relevant features from the intermediate layers of a pre-trained model. These intermediate layers capture high-level abstractions and representations that serve as valuable input for the classification stage.

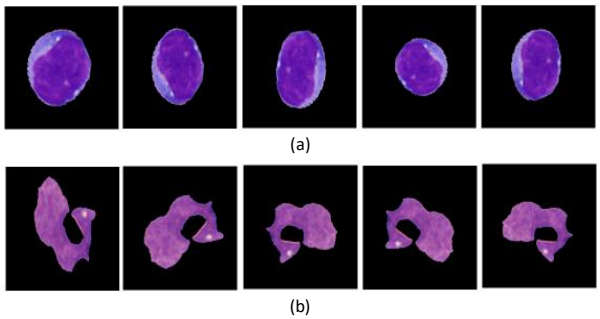


Fig. 6: Examples of augmented images from the ALL-IDB2 dataset: (a) ALL cells, (b) Normal cells

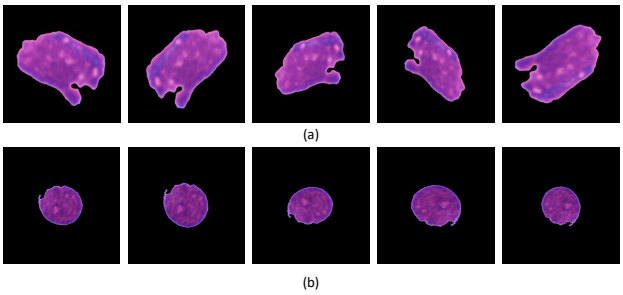


Fig. 7: Examples of augmented images from the C-NMC-2019 dataset: (a) ALL cells, (b) Normal cells

DenseNets: DenseNet is a deep CNN architecture known for its dense connectivity between layers. This dense connectivity pattern promotes feature reuse and improves gradient flow within the network. The architecture consists of several dense blocks, which are interconnected layers that facilitate feature propagation throughout the network (Fig. 8). DenseNets also incorporate transition layers between these dense blocks to adjust feature map sizes. Several DenseNet variants exist, including DenseNet-169, DenseNet-121, DenseNet-201, and DenseNet-264. These variants differ in their number of layers and are often pre-trained on the ImageNet image dataset. They all share a common growth rate of 32, where each convolutional layer within a dense block involves a sequence including a ReLU activation function and a convolutional layer [30].

In our proposed method, we evaluated various DenseNet architectures for feature extraction and found DenseNet-201 to yield the most promising results. This selection is partly attributed to DenseNet-201's lower number of

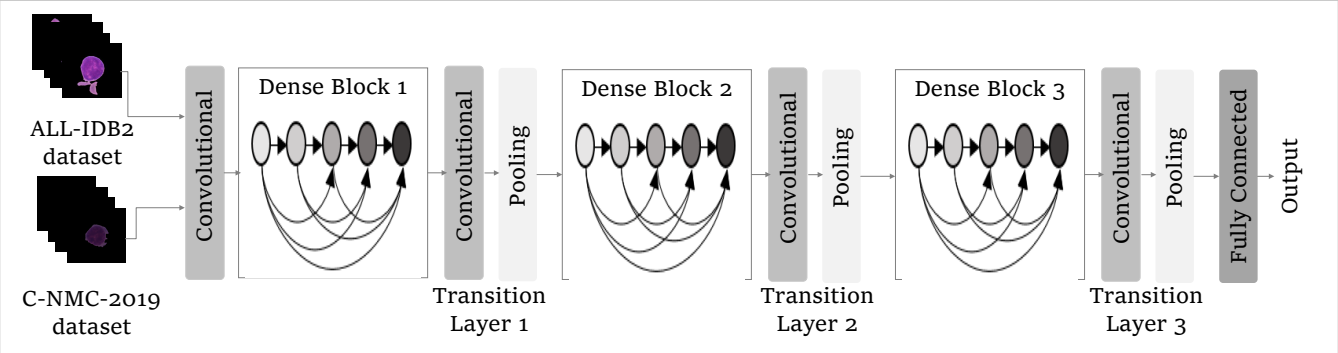


Fig. 8: A DenseNet with three dense blocks

trainable parameters compared to other architectures [30]. With approximately 20 million parameters, it is less computationally expensive than models like VGG-19 (which has 144 million parameters). Furthermore, DenseNet-201 demonstrated superior performance on the ALL-IDB2 dataset compared to other tested networks.

Prior to feeding the data into the DenseNet-201 architecture for feature extraction, we resize all images to a uniform size of 224x224 pixels. This ensures compatibility with the network's input requirements and facilitates the extraction of critical features for subsequent classification.

G. Classification

Following feature extraction using DenseNet-201, the extracted features are fed into a classification block to categorize the cells as either Normal or ALL. We employ an SVM classifier for this task.

GridSearch, a hyperparameter optimization technique, is used to identify the optimal SVM hyperparameter configuration that maximizes the classification accuracy compared to other approaches. We also explored the application of alternative ML classifiers, including RF and XGBoost. However, our evaluations revealed that the SVM achieved the most effective classification performance on both the ALL-IDB2 and C-NMC-2019 datasets.

IV. Results and Discussion

This section evaluates the performance of the proposed method, presents the achieved results, and conducts a comparative analysis of feature extraction using various neural networks (NNs). Additionally, we assessed multiple classifiers to determine the most effective one for this task.

A. Experimental setup

We implemented the proposed method using Python and the Keras library, a high-level API for the TensorFlow machine learning framework. Each dataset was split into training (80%) and testing (20%) sets for model training and evaluation. The experiments were conducted on a PC with an Intel(R) Core(TM) i7-7500U CPU, 16GB of RAM, and a 64-bit Windows operating system.

B. Evaluation metrics

To comprehensively assess the proposed method's performance in cell classification, we employ a set of performance metrics beyond just accuracy. These metrics include accuracy, precision, sensitivity (recall), specificity, F1-score, AUC, and receiver operating characteristic (ROC) curve, Relying solely on accuracy can be misleading, so using a combination of metrics provides a more robust evaluation of the model's effectiveness [31], [32]. Here, we present the chosen metrics and their mathematical formulas for clarity as shown in the following equations: [33]:

- *Accuracy:* The proportion of correctly predicted cases. It is calculated by dividing the number of true positives (TP) and true negatives (TN) by the total number of cases.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

- *Specificity:* The percentage of negative samples correctly identified. A high specificity indicates the model's ability to correctly classify negative examples, while a low specificity suggests issues with correctly identifying negative cases.

$$Specificity = \frac{TN}{TN+FP} \tag{3}$$

- *Sensitivity (Recall):* The proportion of positive samples correctly identified. It reflects the model's ability to identify true positives and avoid false negatives.

$$Sensitivity = \frac{TP}{TP+FN} \tag{4}$$

- *Precision:* The ratio of correctly predicted positive cases to the total predicted positive cases.

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

- *F1-Score:* A harmonic mean that combines precision and recall into a single metric, providing a balanced view of model performance.

$$F1-score = 2 * \frac{precision * recall}{precision + recall} \quad (6)$$

- *ROC Curve:* A visual representation of a classification model's performance across various classification thresholds. It utilizes the True Positive Rate (TPR) and False Positive Rate (FPR) to assess the model's effectiveness. A model with a higher TPR and lower FPR at each threshold demonstrates superior performance.
- *AUC:* The area under the ROC curve quantifies the model's ability to distinguish between positive and negative classes. A higher AUC value indicates better model performance.

These metrics rely on the concept of a confusion matrix, which is a table summarizing the model's prediction outcomes (Fig. 9). The confusion matrix includes:

- *True Positive (TP):* A case where the model correctly predicts the positive class.
- *True Negative (TN):* A case where the model correctly predicts the negative class.
- *False Positive (FP):* A case where the model incorrectly predicts the positive class (also known as a Type I error).
- *False Negative (FN):* A case where the model incorrectly predicts the negative class (also known as a Type II error).

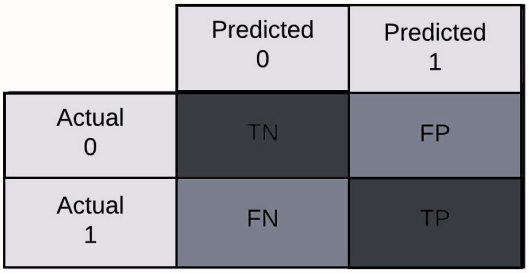


Fig. 9: Illustration of a confusion matrix

By evaluating these metrics, we gain a comprehensive understanding of the proposed method's strengths and weaknesses in classifying blood cells across the ALL-IDB2 and C-NMC-2019 datasets.

C. Hyperparameter tuning

Machine learning (ML) models rely on parameters that are learned from data during the training process. However, certain parameters, known as hyperparameters, cannot be directly learned from the data. These hyperparameters require manual selection and are typically determined through experience, trial-and-error, or dedicated tuning techniques. Selecting optimal hyperparameters plays a crucial role in improving the model's efficiency and overall performance.

In this study, GridSearchCV, a hyperparameter tuning tool, was employed to optimize the performance of the support vector machine (SVM) model. GridSearchCV systematically evaluates a predefined grid of hyperparameter values and identifies the combination that yields the best results based on the selected scoring metrics. This approach automates the hyperparameter selection process, ensuring the most effective configuration for the SVM model.

Specifically, we used GridSearchCV to tune the following key hyperparameters of the SVM model:

- *C:* This parameter controls the trade-off between training error and model complexity. A higher *C* value penalizes training errors more heavily, potentially leading to overfitting.
- *Gamma:* This parameter determines the influence of training points on the decision boundary. A lower gamma value implies that data points have a broader influence, while a higher gamma value restricts their influence to a narrower region.
- *Kernel:* This parameter defines the kernel function used by the SVM model. We explored both the RBF and polynomial kernels.

The results of the GridSearchCV optimization are presented in Table II. As shown, the best-performing hyperparameter configuration consists of C=10, Gamma=0.01, and the RBF kernel.

D. Numerical results

Following the preprocessing and data augmentation stages, K-means clustering was applied to isolate nuclei from the background, specifically in ALL-IDB2 images. The segmented images were then processed using the pre-trained DenseNet-201 model for feature extraction. Subsequently, multiple classifiers—SVM, RF, and XGBoost—were evaluated to assess classification performance.

As presented in Table III, the SVM classifier achieved the highest accuracy on both datasets during the testing phase. The proposed method was further applied to the C-NMC-2019 dataset, where SVM again demonstrated superior performance.

Table III presents the classification accuracy of the XGBoost, RF, and SVM models combined with the pre-trained DenseNet-201 model for both datasets. Notably, the model achieved 100% accuracy on the training sets,

TABLE II: Best values of SVM's hyperparameters based on GridSearchCV

Hyperparameter	Value range	GridSearchCV result
C	[0.1, 1, 10, 100, 1000]	10
Gamma	[1, 0.1, 0.01, 0.001, 0.0001]	0.01
Kernal	[rbf, poly]	rbf

TABLE III: Classification accuracy of XGBoost, RF, and SVM models with the pre-trained DenseNet-201 model for both datasets

	XGBoost		RF		SVM	
Dataset	Training	Testing	Training	Testing	Training	Testing
ALL-IDB2	100	97.6	100	95.97	100	**100**
C-NMC-2019	100	98.0	100	97.19	100	**99.13**

indicating that all training samples were classified correctly. However, this may suggest overfitting, where the model memorizes training data but struggles with generalization to unseen data.

The true measure of model performance lies in its generalizability. In this regard, the proposed model demonstrated excellent performance on the testing sets, achieving 100% accuracy on ALL-IDB2 and 99.13% accuracy on C-NMC-2019. These results highlight its effectiveness in classifying leukemia cells.

E. Quantitative evaluation

We conducted a quantitative evaluation to assess the performance of the proposed method. This evaluation involved comparing various neural networks (NNs) as feature extractors and different machine learning (ML) algorithms as classifiers to identify the most effective combination for the ALL-IDB2 dataset.

Table IV presents a summary of the performance comparison. As observed, the combination of DenseNet-201 for feature extraction and SVM as the classifier achieved the highest accuracy on both datasets. Figures 10–12 visually depict the confusion matrices corresponding to each model combination reported in Table IV for the ALL-IDB2 dataset. In addition, Figure 13 presents the confusion matrices for Random Forest, XGBoost, and SVM, with DenseNet-201 as the feature extractor for the C-NMC-2019 dataset. These confusion matrices provide a detailed breakdown of the classification results for each class.

We observed that DenseNet-based architectures consistently outperformed VGG-19 and ResNet-50. This may be attributed to DenseNet's inherent ability to promote feature reuse and enhance gradient flow within the network, leading to more effective feature extraction.

To further analyze the performance of DenseNet-201 as a feature extractor, we evaluated its effectiveness

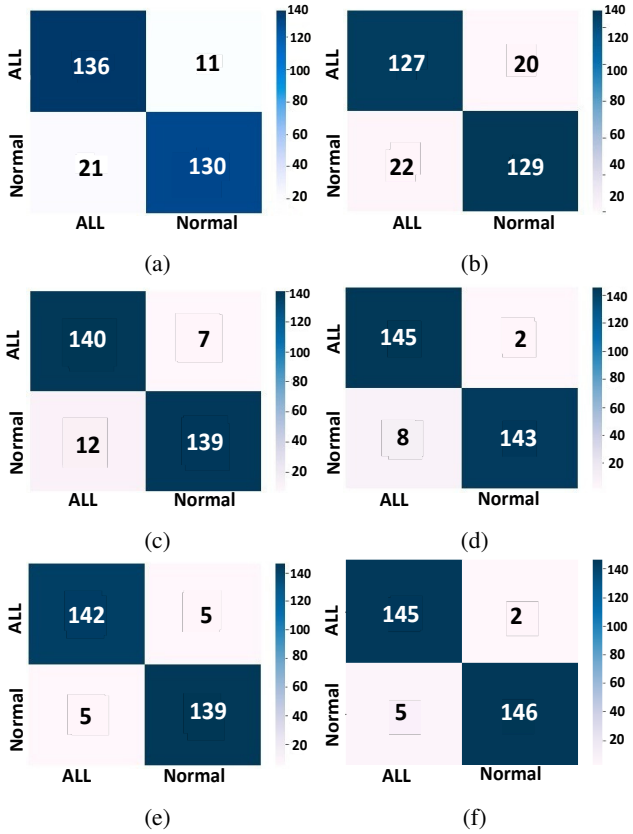


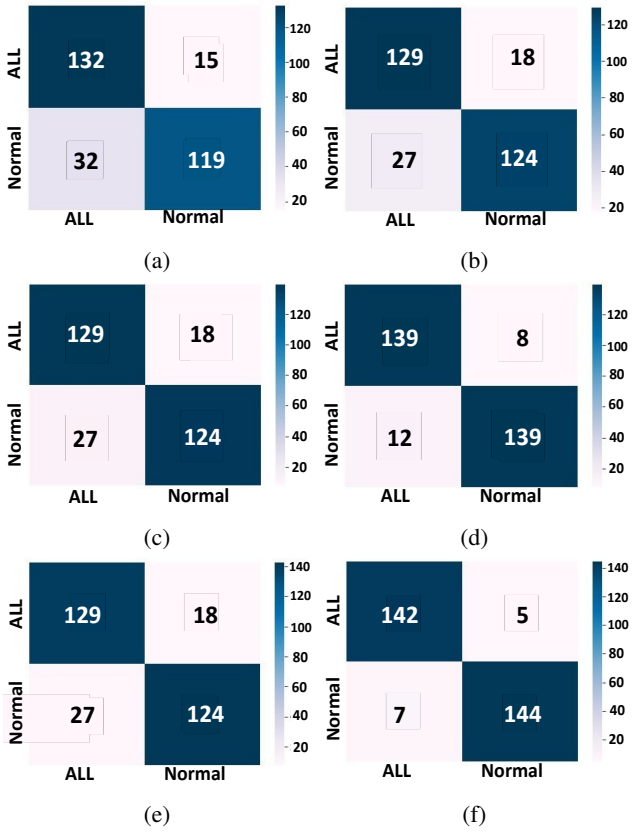
Fig. 10: The ALL-IDB2 confusion matrices for XGBoost with different feature extractors: (a) VGG-19, (b) ResNet-50, (c) MobileNetV2, (d) DenseNet-169, (e) DenseNet-121, and (f) DenseNet-201.

using different classifiers (RF, XGBoost, and SVM) on two datasets: ALL-IDB2 and C-NMC-2019. Tables V and VI provide a comprehensive evaluation of these results, including accuracy, specificity, sensitivity, precision, F1-score, and area under the curve (AUC). Notably, SVM achieved the highest performance across all metrics on both datasets.

Figures 14 illustrates the ROC curves for each classifier applied to DenseNet-201 features for both datasets. It presents the ROC-AUC curves, demonstrating the

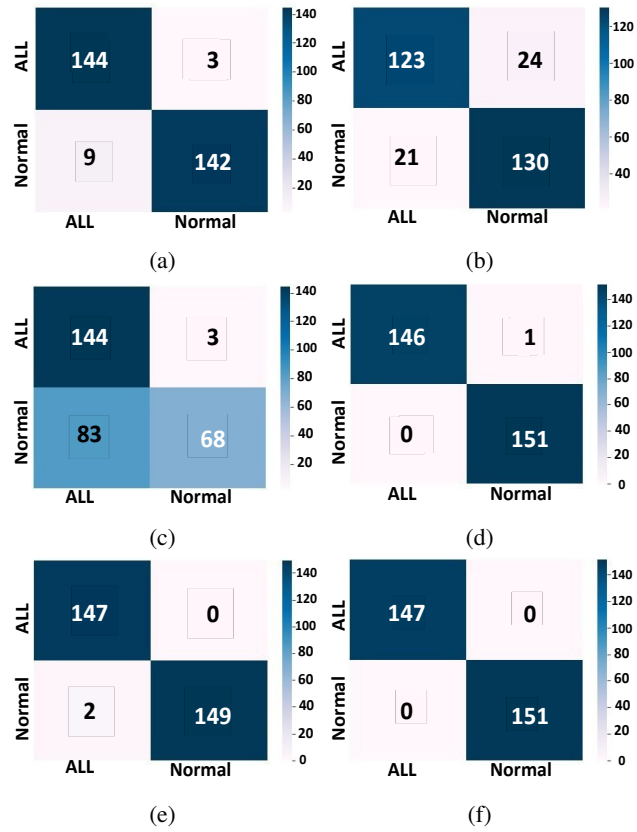
TABLE IV: Performance assessment of different classifiers and feature extraction networks of both datasets

Dataset	ALL-IDB2			C-NMC-2019		
	XGboost	RF	SVM	XGboost	RF	SVM
VGG-19	89.26	84.22	96	96.0	95.7	96.2
Resnet-50	85.90	84.89	84.4	95.66	95.88	95.5
MobileNetv2	93.62	92.95	71.90	97.25	96.84	98.6
DenseNet-121	96.6	94.3	99.3	97.8	97.0	98.7
DenseNet-169	96.64	93.3	99.7	97.87	97.1	99.00
DenseNet-201	97.68	95.97	**100**	97.53	97.19	**99.13**



(a) (b) (c) (d) (e) (f)

Fig. 11: The ALL-IDB2 confusion matrices for RF with different feature extractors: (a) VGG-19, (b) ResNet-50, (c) MobileNetV2, (d) DenseNet-169, (e) DenseNet-121, and (f) DenseNet-201.



(a) (b) (c) (d) (e) (f)

Fig. 12: The ALL-IDB2 confusion matrices for SVM with different feature extractors: (a) VGG-19, (b) ResNet-50, (c) MobileNetV2, (d) DenseNet-169, (e) DenseNet-121, and (f) DenseNet-201.

performance of different models in the classification process. 14a shows that SVM achieves the best performance (AUC = 0.99) compared to Random Forest (AUC = 0.95) and XGBoost (AUC = 0.90), highlighting the robustness of the proposed model that combines DenseNet and SVM. Figures 14b and 14c display the model's performance on the ALL-IDB2 and C-NMC-2019 datasets separately, further emphasizing the superiority of SVM for classification in both cases. On the ALL-IDB2 dataset, SVM achieved the highest performance (AUC = 0.99), followed by Random Forest (AUC = 0.95) and XGBoost (AUC = 0.90). Similarly, for the C-NMC-2019 dataset, SVM attained the highest AUC (0.99), outperforming Random Forest (AUC = 0.94) and XGBoost (AUC = 0.89). Here, we can see that SVM outperforms both RF and XGBoost in all evaluation metrics when using DenseNet-201 features. This suggests that SVM might be particularly well-suited for this specific classification task due to its ability to learn complex decision boundaries between different cell types.

F. Ablation study

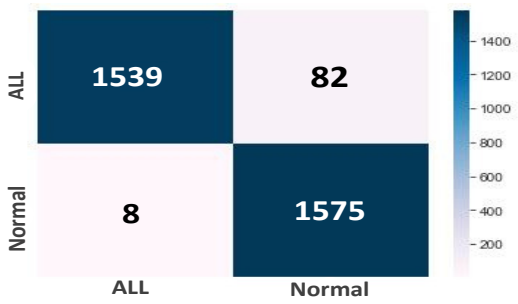
To gain a deeper understanding of the impact of different processing steps on the overall performance, we conducted an ablation study. This involved evaluating the model's performance with and without specific procedures. More specifically, this section evaluates the impact of two key techniques employed in our proposed method: image segmentation and data augmentation.

TABLE V: Overall performance assessment of classifiers with features extracted by DenseNet-201 of ALL-IDB2 dataset

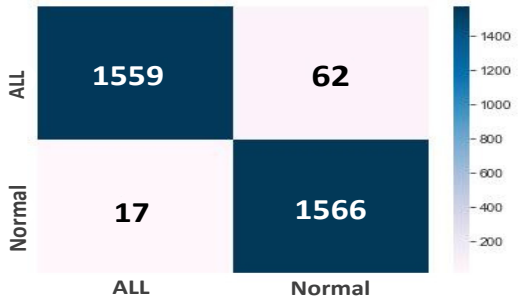
Classifier	Accuracy	Specificity	Sensitivity	Precision	F1-score	AUC
XGBoost	97.65	98.64	96.68	98.63	98.63	99.70
RF	95.97	96.60	95.40	95.36	96.00	99.38
SVM	**100**	**100**	**100**	**100**	**100**	**100**

TABLE VI: Overall performance assessment of classifiers with features extracted by DenseNet-201 of C-NMC-2019 dataset

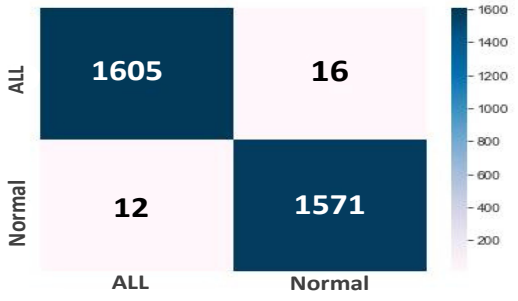
Classifier	Accuracy	Specificity	Sensitivity	Precision	F1-score	AUC
XGBoost	97.53	96.20	98.92	96.19	98.00	99.81
RF	97.19	95.05	99.48	95.10	97.20	99.40
SVM	**99.13**	**99.24**	**99.01**	**99.00**	**99.10**	**99.96**



(a)



(b)



(c)

Fig. 13: The C-NMC-2019 confusion matrices for DenseNet-201 as a feature extractor with different classifiers: (a) XGBoost, (b) RF, (c) SVM.

1) Impact of segmentation: Image segmentation plays a crucial role in isolating cells of interest (nuclei) from background noise in images. Table VII presents the performance of different feature extractors and classifiers (XGBoost, RF, and SVM) when applied with and without image segmentation (denoted as "Seg" and "NoSeg," respectively). Notably, segmentation was applied only to the ALL-IDB2 dataset, as the C-NMC-2019 dataset already consists of pre-segmented cells.

As observed in the table, applying image segmentation consistently improves the classification accuracy for all models. This suggests that isolating the relevant regions (nuclei) from the background noise enhances the feature extraction process and ultimately leads to better classification results. As shown, with DenseNet-201 as the feature extractor, SVM achieved an accuracy of 94.23% with segmentation, which increased to 100% with augmentation (as demonstrated in the later section).

2) Impact of augmentation: Data augmentation is another technique commonly used to improve the performance of deep learning (DL) models, particularly convolutional neural networks (CNNs). CNNs rely heavily on large amounts of training data to learn effective features. Data augmentation artificially expands the training dataset by generating variations of existing images through techniques such as flipping, rotation, and scaling. This helps the model become more robust to variations in real-world data and reduces the risk of overfitting.

Table VIII presents the impact of data augmentation on classification accuracy using DenseNet-201 and three different classifiers (XGBoost, RF, and SVM) with the segmented ALL-IDB2 dataset. "Aug" and "NoAug" indicate the results with and without data augmentation, respectively.

As shown in the table, data augmentation significantly enhances performance across all classifiers. Notably, when using SVM with DenseNet-201 features, accuracy increases from 92.30% without data augmentation to 100% with data augmentation.

G. Comparison with state-of-the-art methods

We compared the performance of our proposed method with several recent state-of-the-art approaches for blood cell classification reported in the literature. Table IX presents a

TABLE VII: The impact of data segmentation on classification accuracy with respect to different classifiers and different features extractors (the ALL-IDB2 dataset)

	XGBoost		RF		SVM	
Network	Seg	NoSeg	Seg	NoSeg	Seg	NoSeg
VGG-19	84.61	88.46	88.46	80.76	94.23	92.3
Resnet-50	82.69	82.69	88.46	76.92	80.8	80.76
MobileNetv2	90.38	90.38	90.38	90.38	94.23	90.38
DenseNet-169	90.38	92.30	98.07	98.03	96.15	92.30
DenseNet-121	86.53	82.69	96.15	88.46	92.30	90.38
DenseNet-201	90.38	88.46	94.23	92.30	94.23	92.30

TABLE VIII: The impact of data augmentation on classification accuracy was evaluated using DenseNet-201 as a feature extractor and multiple classifiers on the segmented versions of both datasets.

	XGBoost		RF		SVM	
Dataset	Aug	NoAug	Aug	NoAug	Aug	NoAug
ALL-IDB2	97.65	90.38	95.97	94.23	**100**	92.30
C-NMC-2019	97.53	89.17	97.20	87.40	**99.13**	90.20

comparative analysis for the ALL-IDB2 dataset, detailing the segmentation technique, feature extractor, classifier, and achieved accuracy for each method.

As shown in the table, our proposed method—utilizing K-means clustering for segmentation, DenseNet-201 for feature extraction, and SVM for classification—achieves the highest accuracy (100%) on the ALL-IDB2 dataset. Some previous methods achieved comparable accuracy (e.g., 99.42% by [17] and 98.7% by [16]).

For the C-NMC-2019 dataset, where the class imbalance is a concern, we further evaluated our model by using F1-score, a common metric for imbalanced datasets. Table X compares our method with other recent approaches. As can be seen, our proposed method using DenseNet-201 with SVM achieves the highest F1-score (99.10%) and accuracy (99.13%) on the C-NMC-2019 dataset.

our approach offers several advantages:

1) *Simpler segmentation:* Our method employs K-means clustering, a well-established and computationally efficient technique for image segmentation. In contrast, some other methods rely on more complex architectures, such as modified U-Net ([17]).
2) *Deeper feature extraction:* We leverage the capabilities of DenseNet-201, a deep CNN known for its robust feature extraction abilities. This contrasts with methods using shallower networks (e.g., Xception and DenseNet in [34]) or entirely different approaches, such as centering with SSOA ([35]).
3) *Effective classification:* Our choice of SVM as the classifier has proven to be highly effective for this specific classification task. While other methods explore various classifiers (e.g., logistic regression and RF [36]), SVM's ability to learn complex decision boundaries makes it particularly well-suited for blood cell classification.

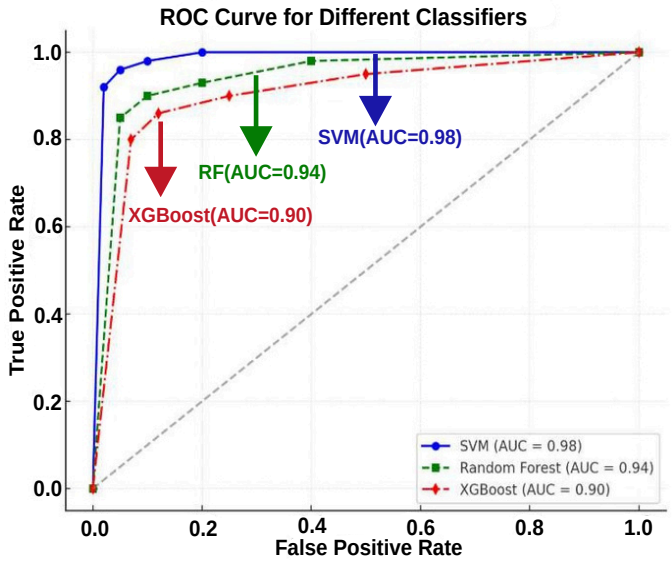
Figure 15a compares the performance of the proposed model with previous methods. The results demonstrate that the proposed model outperforms other approaches, such as Elrefaie et al. [16] (98.7%) and Balasubramanian et al. [17] (99.42%), achieving 100% accuracy on the ALL-IDB2 dataset. This improvement is attributed to a robust framework integrating pre-processing, segmentation, and data augmentation techniques, significantly enhancing the model's performance. Similarly, Figure 15b illustrates the superiority of the proposed model on the C-NMC-2019 dataset, achieving an accuracy of 99.13%.

V. Conclusion

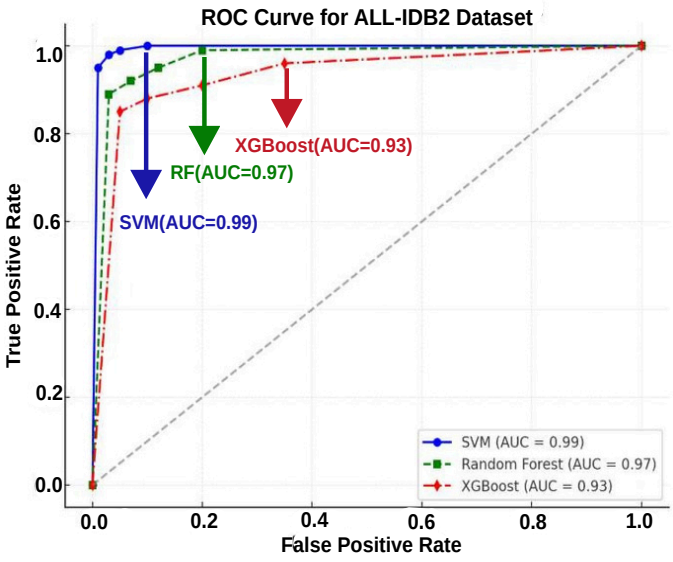
In this study, we proposed a novel method for blood cell classification by integrating image preprocessing, augmentation, segmentation, feature extraction, and classification techniques. Our method achieved an outstanding accuracy of 100% on the ALL-IDB2 dataset and 99.13% on the C-NMC-2019 dataset, outperforming several recent state-of-the-art approaches.

The ablation study highlighted the critical role of both image segmentation and data augmentation in enhancing overall performance. Isolating the cells of interest (nuclei) from background noise improved the feature extraction process, while data augmentation, a widely used deep learning technique, enhanced model robustness, reduced the risk of overfitting, and addressed data imbalance.

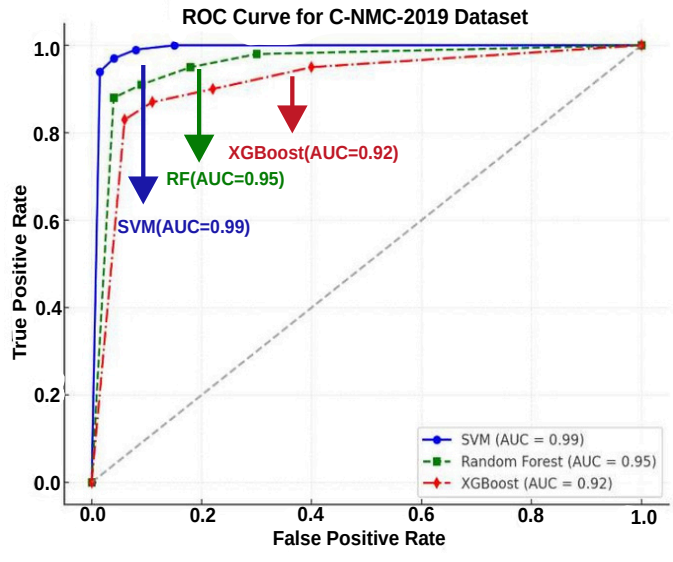
A comparative analysis demonstrated several advantages of our approach. Our method employs K-means clustering, a



(a)



(b)



(c)

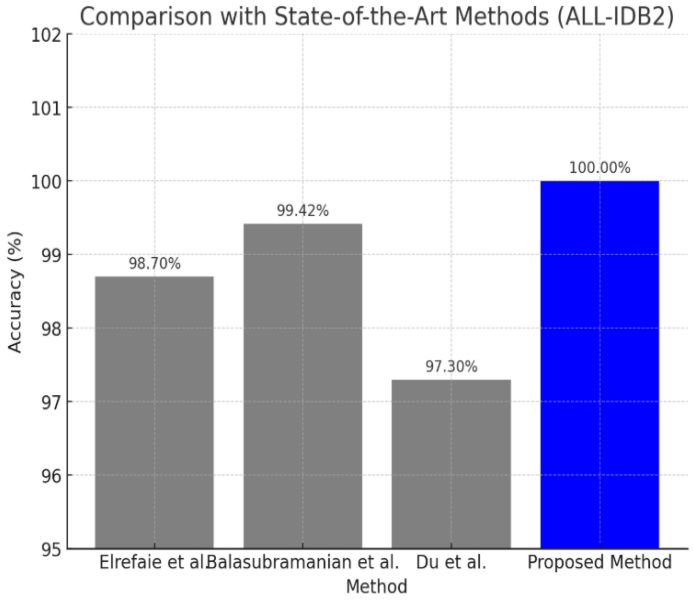
Fig. 14: ROC curves of the three classifiers: XGBoost, RF, SVM and DenseNet-201 as feature extraction.

TABLE IX: Comparison of the proposed model with state-of-the-art methods on the ALL-IDB2 dataset

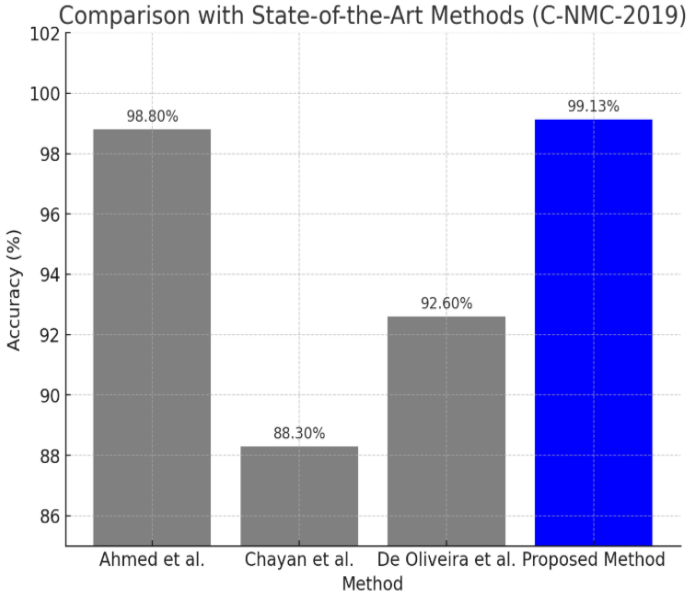
Reference	Segmentation	Feature extractor	Classifier	Accuracy
Elrefaie et al. [16]	K-means clustering	EMD	NNs	98.7%
Balasubramanian et al. [17]	Modified U-Net Architecture	RBF with SVM (RBF-SVM)	RBF-SVM	99.42%
Du et al. [34]	No segmentation	Xception and DenseNet	CNN	97.3%
Abdeldaim et al. [35]	Zack method	Centering with Social Spider Optimization algorithm (SSOA)	KNN	95.2%
Das et al. [36]	No segmentation	Three models based on modified ResNet50	Logistics regression, RF, and SVM (MKD-Net)	96.15%
Vishwakarma. [37]	No segmentation	ResNet-50	YOLOv8	97.44%
Baluabid et al. [38]	No segmentation	YOLOv8	YOLOv8	94.0%
Maqsood et al. [39]	No segmentation	MobileNetV2, EfficientNetB0, ConvNeXt-V2 EfficientNetV2, and DarkNet-19	SVM	98.96%
Proposed method	K-means clustering	DenseNet-201	SVM	**100%**

TABLE X: Comparison of the proposed model with state-of-the-art methods of C-NMC-2019 dataset

Reference	Model	F1-score	Accuracy
chen et al. [40]	Resnet101-9 ensemble model	88.94%	85.11%
Ahmed et al. [22]	CNN-RF and CNN-XGBoost	98.0	98.8%
Chayan et al. [20]	Ensemble model of five different pre-trained networks	89.7%	88.3%
De Oliveira et al. [19]	The VGG16 from scratch	92.60%	92.48%
Honnalgere et al. [41]	Modified VGG-16 network	91.7%	N/A
Dayata et al. [42]	integrate ConvNeXtTiny, MobileNetV2, EfficientNetV2B3, InceptionV3, and DenseNet121	95.80%	95.84%
Pan et al. [43]	Fine-tuned ResNet	92.50%	91.73%
Mohamed et al. [21]	hybrid CNN-GRU-BiLSTM and MSVM	96.23%	96.29%
Baluabid et al. [38]	YOLOv8	N/A	95.0%
Maqsood et al. [39]	MobileNetV2, EfficientNetB0, ConvNeXt-V2 EfficientNetV2, and DarkNet-19 and SVM as classifier	N/A	96.67%
Proposed method	DenseNet-201 with SVM	**99.10%**	**99.13%**



(a)



(b)

Fig. 15: comparison with State-of-the-Art Methods (a) ALL-IDB2, (b) C-NMC-2019 dataset.

well-established and computationally efficient segmentation technique, whereas some existing methods rely on more complex architectures. Additionally, we leverage the strong feature extraction capabilities of DenseNet-201, a deep CNN. Finally, our choice of SVM as the classifier proved particularly effective due to its ability to learn complex decision boundaries between different cell types.

While our model achieved 100% accuracy on the ALL-IDB2 dataset and 99.13% on the C-NMC-2019 dataset, such high test set performance may suggest potential overfitting. However, the significant performance improvement over existing methods underscores the effectiveness of our approach for blood cell classification.

Future work could focus on evaluating the generalizability of the model across additional blood cell classification datasets and exploring strategies to mitigate overfitting. Moreover, incorporating domain knowledge about blood cell morphology into the model could further enhance its accuracy and robustness.

REFERENCES

[1] World Health Organization, "Cancer report," 2022, [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer.

[2] American Cancer Society, "Leukemia statistics," 2023, [Online]. Available: https://cancerstatisticscenter.cancer.org/!/cancer-site/Leukemia.

[3] 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, no. 10, pp. 1–10, 2018.

[4] L. Putzu, G. Caocci, and C. Di Ruberto, "Leucocyte classification for leukaemia detection using image processing techniques," *Artificial Intelligence in Medicine*, vol. 62, no. 3, pp. 179–191, 2014.

[5] R. B. Hegde, K. Prasad, H. Hebbar, and B. M. K. Singh, "Feature extraction using traditional image processing and convolutional neural network methods to classify white blood cells: a study," *Australasian Physical & Engineering Sciences in Medicine*, vol. 42, no. 2, pp. 627–638, 2019.

[6] R. T. Raphael and K. Joy, "Segmentation and classification techniques of leukemia using image processing: An overview," in *Proc.International Conference on Intelligent Sustainable Systems (ICISS 2019)*. Palladam, India: IEEE, 12–14 December 2019, pp. 378–384.

[7] A. R. Shaik and P. R. Kumar, "Performance evaluation of machine learning algorithms on skin cancer data set using principal component analysis and Gabor filters," *IAENG International Journal of Computer Science*, vol. 51, no. 7, pp. 831–841, 2024.

[8] M. Turuk, R. Sreemathy, S. Kadiyala, S. Kotecha, and V. Kulkarni, "CNN Based Deep Learning Approach for Automatic Malaria Parasite Detection," *IAENG International Journal of Computer Science*, vol. 49, no. 3, p. 745–753, 2022.

[9] A. Murmu and P. Kumar, "DLRFNet: deep learning with random forest network for classification and detection of malaria parasite in blood smear," *Multimedia Tools and Applications*, vol. 83, no. 23, pp. 63 593–63 615, 2024.

[10] E. Y. Abbasi, Z. Deng, Q. Ali, A. Khan, A. Shaikh, M. S. Al Reshan, A. Sulaiman, and H. Alshahrani, "A machine learning and deep learning-based integrated multi-omics technique for leukemia prediction," *Heliyon*, vol. 10, no. 3, p. e25369, 2024.

[11] M. Rana and M. Bhushan, "Machine learning and deep learning approach for medical image analysis: diagnosis to detection," *Multimedia Tools and Applications*, vol. 82, no. 17, pp. 26 731–26 769, 2023.

[12] S. A. Ajagbe and M. O. Adigun, "Deep learning techniques for detection and prediction of pandemic diseases: a systematic literature review," *Multimedia Tools and Applications*, vol. 83, no. 2, pp. 5893–5927, 2024.

[13] A. Bouzid-Daho, N. Sofi, S. Soltane, and P. Siarry, "Automated detection in microscopic images using segmentation," *Brazilian Journal of Technology*, vol. 7, no. 2, p. e69317, 2024.

[14] Z. Moshavash, H. Danyali, and M. S. Helfroush, "An automatic and robust decision support system for accurate acute leukemia diagnosis from blood microscopic images," *Journal of Digital Imaging*, vol. 31, no. 5, pp. 702–717, 2018.

[15] D. Umamaheswari and S. Geetha, "A framework for efficient recognition and classification of acute lymphoblastic leukemia with a novel customized-knn classifier," *Journal of Computing and Information Technology*, vol. 26, no. 2, pp. 131–140, 2018.

[16] R. M. Elrefaie, E. A. Marzouk, M. A. Mohamed, and M. M. Ata, "Supervised acute lymphocytic leukemia detection and classification based-empirical mode decomposition," in *Proc. International Telecommunications Conference (ITC 2022)*. Aswan, Egypt: IEEE, 26–28 July 2022, pp. 1–7.

[17] K. Balasubramanian, K. Gayathri Devi, and K. Ramya, "Classification of white blood cells based on modified u-net and svm," *Concurrency and Computation: Practice and Experience*, vol. 35, no. 28, p. e7862, 2023.

[18] S. Saxena, S. Prasad, and D. M. TS, "Utilizing deep learning techniques to diagnose nodules in lung computed tomography (CT) scan images," *IAENG International Journal of Computer Science*, vol. 50, no. 2, pp. 537–552, 2023.

[19] J. E. M. de Oliveira and D. O. Dantas, "Classification of normal versus leukemic cells with data augmentation and convolutional neural networks," in *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2021)*, vol. 4. Setúbal, Portugal: SciTePress, INSTICC, 8–10 February 2021, pp. 685–692.

[20] C. Mondal, M. K. Hasan, M. Ahmad, M. A. Awal, M. T. Jawad, A. Dutta, M. R. Islam, and M. A. Moni, "Ensemble of convolutional neural networks to diagnose acute lymphoblastic leukemia from microscopic images," *Informatics in Medicine Unlocked*, vol. 27, no. 4, p. 100794, 2021.

[21] K. K. Mohammed, A. E. Hassanien, and H. M. Afify, "Refinement of ensemble strategy for acute lymphoblastic leukemia microscopic images using hybrid cnn-gru-bilstm and msvm classifier," *Neural Computing and Applications*, vol. 35, no. 23, pp. 17 415–17 427, 2023.

[22] I. A. Ahmed, E. M. Senan, H. S. A. Shatnawi, Z. M. Alkhraisha, and M. M. A. Al-Azzam, "Hybrid techniques for the diagnosis of acute lymphoblastic leukemia based on fusion of cnn features," *Diagnostics*, vol. 13, no. 6, pp. 1026–1049, 2023.

[23] T. V. Renuka and B. Surekha, "Acute-lymphoblastic leukemia detection through deep transfer learning approach of neural network," in *Proceeding of First Doctoral Symposium on Natural Computing Research (DSNCR 2020)*, vol. 169. Singapore: Springer, 14–15 December 2021, pp. 163–170.

[24] A. Rehman, N. Abbas, T. Saba, S. I. u. Rahman, Z. Mehmood, and H. Kolivand, "Classification of acute lymphoblastic leukemia using deep learning," *Microscopy Research and Technique*, vol. 81, no. 11, pp. 1310–1317, 2018.

[25] T. TTP, G. N. Pham, J.-H. Park, K.-S. Moon, S.-H. Lee, K.-R. Kwon *et al.*, "Acute leukemia classification using convolution neural network in clinical decision support system," in *Proc. SAI 2017 - 6th international conference on soft computing, artificial intelligence and applications (SAI-2017)*, vol. 7, no. 13. Sydney, Australia: Science and Information Organization, 28–29 December 2017, pp. 49–53.

[26] J. Prellberg and O. Kramer, "Acute lymphoblastic leukemia classification from microscopic images using convolutional neural networks," in *Proc. ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging,Part of IEEE 16th International Symposium on Biomedical Imaging*. Singapore: Springer, 8–11 April 2019, pp. 53–61.

[27] K. Ananthu, P. Krishna Prasad, S. Nagarajan, and E. Vimina, "Acute lymphoblastic leukemia detection using transfer learning techniques," in *Proc. 3rd International Conference on Intelligent Sustainable Systems (ICISS 2021)*, vol. 213. Singapore: Springer, 16–17 December 2022, pp. 679–692.

[28] F. Scotti, R. D. Labati, and P. V. Piuri, "ALL-IDB: Acute lymphoblastic leukemia image database," 2020. [Online]. Available: https://doi.org/10.21227/pm77-2n23

[29] S. Mourya, S. Kant, P. Kumar, A. Gupta, and R. Gupta, "All challenge dataset of isbi 2019 (c-nmc 2019)," 2019, version 1, dataset. [Online]. Available: https://doi.org/10.7937/TCIA.2019.DC64I46R

[30] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*. Honolulu, Hawaii, USA: IEEE, 21–26 July 2017, pp. 2261–2269.

[31] Ž. Vujović *et al.*, "Classification model evaluation metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 599–606, 2021.

[32] M. Ghaderzadeh, F. Asadi, R. Jafari, D. Bashash, H. Abolghasemi, and M. Aria, "Deep convolutional neural network–based computer-aided detection system for covid-19 using multiple lung scans: design and implementation study," *Journal of Medical Internet Research*, vol. 23, no. 4, p. e27468, 2021.

[33] A. Tharwat, "Classification assessment methods," *Applied computing and informatics*, vol. 17, no. 1, pp. 168–192, 2020.

[34] Z. Du, X. Xia, M. Fang, L. Yu, and J. Li, "A deep transfer fusion model for recognition of acute lymphoblastic leukemia with few samples," in *Proc. 19th International Conference on Intelligent Computing (ICIC 2023)*, vol. 14087. Cham, Switzerland: Springer, 10–13 August 2023, pp. 710–721.

[35] A. M. Abdeldaim, A. T. Sahlol, M. Elhoseny, and A. E. Hassanien, "Computer-aided acute lymphoblastic leukemia diagnosis system based on image analysis," in *Advances in Soft Computing and Machine Learning in Image Processing*. Cham, Switzerland: Springer, 2018, vol. 730, pp. 131–147.

[36] P. K. Das, A. Pradhan, and S. Meher, "Detection of acute lymphoblastic leukemia using machine learning techniques," in *Proc. Machine Learning, Deep Learning and Computational Intelligence for Wireless Communication (MDCWC 2020)*, vol. 749. Singapore: Springer, 3–4 October 2021, pp. 425–437.

[37] V. P. Vishwakarma and A. K. Yadav, "Mkd-net: A novel neuro evolutionary approach for blockchain-based secure medical image classification using multi-kernel dlm," *IEEE Access*, vol. 13, no. 1, pp. 29 900–29 913, 2025.

[38] R. Baluabid, H. Alnasri, R. Alowaybidi, R. Hafiz, A. Alsini, and M. Alharbi, "Detecting acute lymphocytic leukemia in individual blood cell smear images," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 19 167–19 173, 2025.

[39] S. Maqsood, R. Damaševičius, R. Maskeliūnas, N. D. Forkert, S. Haider, and S. Latif, "Csec-net: a novel deep features fusion and entropy-controlled firefly feature selection framework for leukemia classification," *Health Information Science and Systems*, vol. 13, no. 1, pp. 1–26, 2025.

[40] Y.-M. Chen, F.-I. Chou, W.-H. Ho, and J.-T. Tsai, "Classifying microscopic images as acute lymphoblastic leukemia by resnet ensemble model and taguchi method," *BMC bioinformatics*, vol. 22, no. 5, pp. 1–15, 2021.

[41] A. Honnalgere and G. Nayak, "Classification of normal versus malignant cells in b-all white blood cancer microscopic images," in *Proc. ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging*. Singapore: Springer, 8–11 April 2019, pp. 1–12.

[42] W. M. A. Dayata, S. Y. C. Yap, and C. D. Bandalan, "Automated blast cell detection for acute lymphoblastic leukemia using a stacking ensemble of convolutional neural networks," in *Proc. IEEE International Conference on Communication, Networks and Satellite (COMNETSAT 2023)*. IEEE, 5–7 July 2023, pp. 95–102.

[43] Y. Pan, M. Liu, Y. Xia, and D. Shen, "Neighborhood-correction algorithm for classification of normal and malignant cells," in *Proc. ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging*. Singapore: Springer, 8–11 April 2019, pp. 73–82.