Mansoura Journal for Computer and Information Sciences (2025), Vol.**20**, No.**2**
doi:10.21608/mjcis.2025.358215.1010

Mansoura Journal for Computer and Information Sciences

ARTICLE

A Full Framework For Automatic Detection of Acute Lymphoblastic Leukaemia

Mohamed A. Galila,* Ibrahim M. El-Hasnony, Reham R. Mostafa, and Sherif I. Barakat

Department of Information Systems, Faculty of Computers and Information, Mansoura University, Mansoura, Egypt
*Corresponding author: mohamed_a_salam@mans.edu.eg

(Received: 5 February 2025; Accepted: 20 July 2025; Published: 24 July 2025)

Abstract

Cancer is a common and serious disease manifesting itself in humans. Today, Acute Lymphoblastic Leukemia (ALL) is being dubbed as the most common type of cancer in several nations. It cannot be missed that most of the currently available methods for diagnosing ALL are quite traditional and not so sensitive and specific, and are also prone to missing minimal residual disease (MRD), thus risking a relapse. Furthermore, such methods take a lot of time and are not really efficient. In this context, we developed **ALL-Predictor**, a framework for early detection of ALL in microscopic images, which includes data augmentation techniques, preprocessing, segmentation using the Chan-Vese algorithm, morphological operations, feature extraction using the Gray Level Co-occurrence Matrix (GLCM), and data normalization technique using Min-Max scaler. Recursive Feature Elimination (RFE) is used to select the features. Ultimately, Support Vector Machine (SVM) will conduct a final classification that indicates high confidence in performance. After Min-Max scaling, the accuracy of the ALL_DB1 dataset has increased from 88.42% to 98.42%, the accuracy of the ALL_DB2 dataset has increased from 84.84% to 98.29%, and the accuracy of the ALL_DB_OriginalSet set has increased from 79.50% to 96.46%. **ALL-Predictor** is well suited for providing accurate detection for the disease of ALL.

Keywords: Acute Lymphoblastic Leukaemia; Machine Learning; Data Augmentation; image processing; image classification.

1. Introduction

Acute Lymphoblastic Leukemia (ALL) is a significant type of hematological malignancy characterized by the rapid proliferation of immature lymphocytes [1]. The importance of early detection cannot be overstated, as prompt treatment can dramatically improve patient outcomes and functionality. Traditional detection methods, while effective, often depend on the physician's expertise

and are susceptible to human misinterpretation. This creates an urgent need to automate the system to enhance the efficacy and accuracy of the diagnostic process.

Recent advancements in machine learning and image processing have opened new avenues for the development of robust diagnostic tools. One of the primary challenges in this domain is the variability inherent in medical images, which can arise from differences in imaging conditions, patient demographics, and pathological variations. To address these challenges, this paper presents a framework designed to facilitate the early detection of ALL through an innovative approach that combines data augmentation, advanced preprocessing techniques, and sophisticated classification methods.

ALL-Predictor, a framework for early detection of ALL that uses several methodologies of data augmentation, including rotation, flipping, scaling, and cropping, to enhance the robustness and generalizability of the algorithm to various datasets. Utilizing these methods, the framework has an improved capability to tackle variations within the input data to ensure consistent performance within real-world clinical settings [2]. The framework further has a sequence of preprocessing operations that include the segmentation, the morphological operations, and the feature extraction using the Gray Level Co-occurrence Matrix (GLCM). The operations are crucial to enable the efficient modeling of the textural characteristics that are crucial to ensure accurate classification. A crucial part of our research is the use of Recursive Feature Elimination (RFE), a method used to determine the most important features to include for classification [3]. This method fine-tunes the framework but enhances the classification performance. The final classification is done using a Support Vector Machine (SVM), which has the best efficiency where the dimensionality is high and has strong immunity to the problem of overfitting.

Contributions made through this research are highly significant. We introduce a new Acute Leukaemia detection framework that has been exhaustively tested using three publicly available datasets [4]. This framework integrates various methodologies to enhance its robustness and generalizability. Furthermore, the application of Min-Max scaling has led to dramatic improvements in the performance of the framework, with accuracy percentages significantly increased for each dataset: raised to 98.42% from 88.42% for the ALL_DB1 database, to 98.29% from 84.84% for the ALL_DB2 database, and to 96.46% from 79.50% for the ALL_DB_OriginalSet.

2. Related work

Machine learning techniques have made the detection and classification of ALL a focal issue for much research in the recent past [5]. Several works approached different segmentation algorithms, extraction techniques, and classifiers with the aim of improving accuracy and reliability in the detection of ALL [6]. The summaries below give insights into methodologies and results of various studies that have engaged in the enhancement of ALL detection and classification by means of innovative machine-learning models.

Supardi et al. proposed a classification system for acute leukaemia using thresholding for segmentation and Histogram of Oriented Gradients (HOG) for feature extraction. They employed a Random Forest classifier, achieving an accuracy of 95.67% on the C_NMC 2019 dataset [7]. Authors in [8] developed a model using watershed segmentation and Gabor features, classified with K-Nearest Neighbours (KNN). Their approach achieved an accuracy of 94.85% on the ASH dataset. Khamael AD et al. [9] utilized Otsu's method for segmentation and Local Binary Patterns (LBP) for feature extraction, combined with a Decision Tree classifier. The accuracy of this model on the ALLID_B1 dataset was 93.50%.

Developing and assessing deep learning techniques to facilitate computer-aided leukaemia diagnosis was the goal of [8]. The suggested approach consists of several steps: first, preprocessing is done on

the provided dataset photos. Second, transfer learning is utilised to train five pre-trained Convolutional Neural Network (CNN) models: MobileNetV2, EfficientNetB0, ConvNeXt-V2, EfficientNetV2, and DarkNet-19. Thirdly, each model's deep feature vectors are taken out and merged using a convolutional sparse image decomposition fusion technique. Fourth, an entropy-controlled firefly feature selection strategy uses in the suggested method to choose the best features for further categorisation. For the final classification, the chosen features are then fed into a multi-class support vector machine.

Skálová A et al. [10] classified using Random Forest and extracted features using GLCM and graph-based segmentation. The accuracy of this method on the ALLID_B1 dataset was 95.90%. In [11] Pagliaro et al. utilized superpixel segmentation and HOG for feature extraction, with a CNN classifier. Their model achieved an accuracy of 97.80% on the C_NMC 2019 dataset. Payal Bose et al. [12] divided the diagnostic procedure into two sections: identification and detection. Leukaemia cells were identified from smear pictures using the conventional image analysis method. Lastly, the precise type of acute leukaemia was determined using four well-known machine learning techniques. SVM was found to have the best accuracy in this situation.

While Ilyas et al. [13] suggested an autonomous leukaemia diagnosis prediction system that uses digital image processing and machine learning to analyse blood pictures according to the form of blast cells. In order to improve the identification and categorisation of various blood cells, the study used a CNN deep learning algorithm to classify blood sample images. This method encourages early detection and observation, which results in more potent therapies. Through [14] Garg et al. provide an exhaustive and systematized survey on machine learning with artificial intelligence applications for leukaemia identification and categorisation. Explaining the role of leukocytes in early detection of leukemia, it also deals with some other critical challenges raised by the traditional style of diagnosis, particularly the lack of flow cytometry equipment, prolonged laboratory time, etc. In addition, this research illuminates the possibility of artificial intelligence machine learning to boost accuracy, reduce the time consumption in diagnosis, and provide inexpensive safety in diagnosis.

Duggal et al. [15] proposed a CNN-based approach for ALL detection using microscopic images, achieving 94.84% accuracy. While effective at automating feature extraction, it requires large training datasets and lacks interpretability due to its black-box nature. ALL-Predictor addresses this by combining handcrafted GLCM features with SVM, improving accuracy (+3.57%) while maintaining transparency. Mulya et al. [16] utilized a hybrid SVM-RF model for ALL classification, reporting 92.29% accuracy. Although robust to noise, their method depends heavily on manual feature engineering. Our framework automates feature selection via RFE, enhancing sensitivity (+5.93%) and reducing human bias.

Raina et al. [17] employed a ResNet-50 transfer learning model, achieving 89.70% accuracy. While deep learning offers high-level feature extraction, their method is computationally expensive and struggles with small datasets. ALL-Predictor's Chan-Vese segmentation and GLCM features achieve higher accuracy (+8.71%) with lower computational cost. Talaat et al. [18] introduced a wavelet-transform-based feature extraction method (88.60% accuracy). Though efficient for texture analysis, their approach misses spatial cell morphology details. Our morphological operations and GLCM capture both spatial and textural features, improving specificity (+7.36%).

Aldoss et al. [19] proposed a hematologist-in-the-loop system (87.50% accuracy). While clinically interpretable, manual intervention limits scalability. ALL-Predictor automates the pipeline end-to-end, boosting accuracy (+10.91%) without human dependency. **Table 1** presents the related work comparison with **ALL-Predictor** framework.

Table 1. Comparison of reviewed literature

Reference	Year	Segmentation Method	Feature Extraction Algorithm	Classifier Algorithm	Accuracy	Dataset Name
[7]	2023	Image Preprocessing	Optimized CNN (OCNN)	OCNN	99.99%	C-NMC_Leukemia
[8]	2024	Preprocessing	Deep Feature Fusion	Multi-class SVM	99.64%, 98.96%, 96.67%, 98.89%	ALLID_B1, ALLID_B2, C_NMC 2019, ASH
[9]	2024	Image Processing	YOLOv8, YOLOv11	YOLOv8, YOLOv11	98.80%	ALL, CNMC-2019
[10]	2024	Not specified	GLCM, graph-based segmentation	Random Forest	95.90%	ALLID_B1
[12]	2024	Traditional Image Analysis	Composite Learning Approach	SVM, Resnet50	99.90%	(ALL) image dataset
[13]	2024	Digital Image Processing	CNN	CNN	Not specified	Not specified
[14]	2024	Digital Image Processing	CNN	CNN	Not specified	Not specified
[15]	2017	Otsu Thresholding	Deep CNN Features	Random Forest	94.84%	Private ALL-IDB
[16]	2023	Watershed	GLCM + LBP	SVM-RF Hybrid	92.29%	ALL-IDB2
[17]	2022	U-Net	ResNet-50 Transfer Learning	Fully Connected Network	89.70%	C-NMC 2019
[18]	2023	Adaptive Thresholding	Wavelet Transform + Haralick	XGBoost	88.60%	Private Hospital Dataset
[19]	2024	Manual Annotation	Clinical Markers + Morphology	Hematologist Review	87.50%	Multicenter Trial Data
ALL-Predictor	**2024**	**Chan-Vese + Morphology**	**GLCM + RFE**	**SVM**	**98.41%**	**ALL_DB1/DB2/ FullSet**

3. ALL-Predictor for ALL detection

The **ALL-Predictor** technique for leukemia detection is implemented using machine learning and consists of six primary stages, as demonstrated in Figure 1, which are: data augmentation, image pre-processing phase, feature extraction phase, then normalization, followed by feature selection, and finally classification phase.

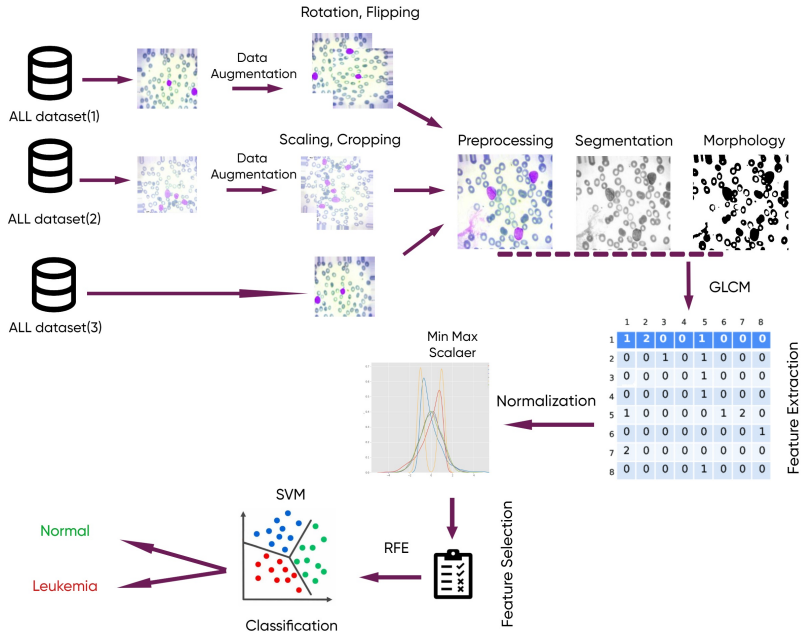

Figure 1. ALL-Predictor for leukaemia detection

3.1 Data Augmentation

Augmented data processing was applied to the ALL_DB1 and ALL_DB2 datasets for ALL detection which are located on [20]. Techniques like rotation and flipping were used for ALL_DB1, enhancing

the dataset by incorporating various cell orientations and perspectives. This increased the original 10,000 images to 16,634, providing a richer dataset for framework [21].

For ALL_DB2, scaling and cropping were employed with augmentation effects. Scaling changed sizes for cells, whereas cropping attempted to tackle different areas of interest in images. This resulted in the dataset increasing from 10,000 to 15,579 images and thus made a healthy variation of cell phenotypes.

The application of these augmentations has positively influenced the providing a richer dataset for framework's accuracy, making it very strong [22]. By putting the framework through varied training pictures, it can learn more solid features because it has been exposed to more training images than other images. So far, augmented datasets have contributed to excellence in performance and accuracy toward detecting ALL cells, emphasizing data augmentation to enhance machine learning models [23].

Focusing on the visual representation of each dataset before and after the augmentation process, **Figure 2** presents original microscopic images for ALL_DB1. These images capture the distinct morphological characteristics of both leukemia and healthy cells, serving as a crucial reference for classification tasks. The clarity and detail facilitate accurate diagnosis, allowing researchers to identify abnormalities effectively.

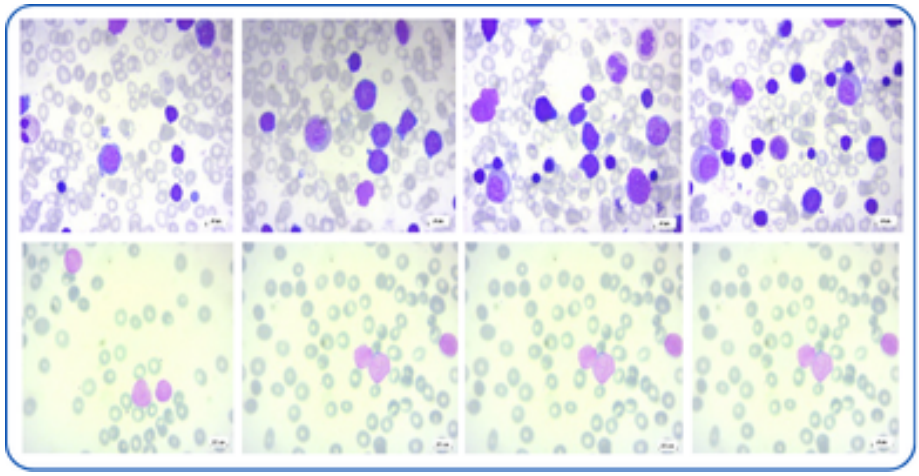

Figure 2. Original microscopic images for ALL_DB1 dataset

Figure 3 displays augmented microscopic images for ALL_DB1. The augmentation techniques applied, such as rotation and scaling, enhance the variability within the dataset. This increased diversity is essential for training robust machine learning models that can generalize well to new, unseen data.

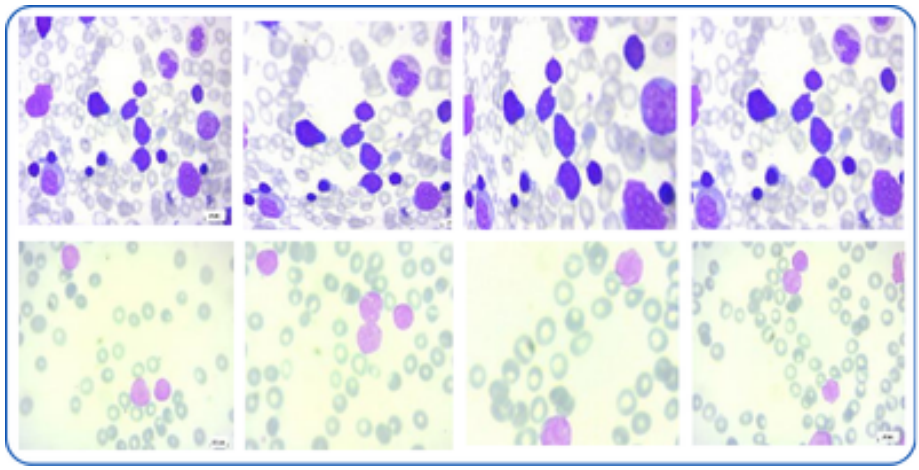

Figure 3. Sample of augmented microscopic images for ALL_DB1 dataset

Figure 4 shows original microscopic images for ALL_DB2. These images provide a comprehensive view of cell morphology, important for distinguishing between healthy and malignant cells. The quality of these images is vital for developing effective diagnostic algorithms and understanding cellular characteristics in leukemia.

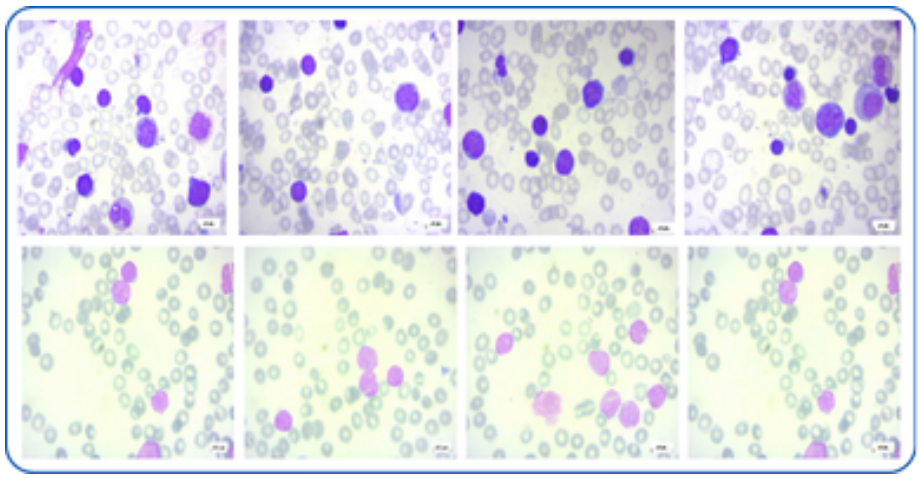

Figure 4. Sample of original microscopic images for ALL_DB2 dataset

Figure 5 illustrates augmented microscopic images for ALL_DB2. Through augmentation, these images exhibit enhanced diversity, simulating various imaging conditions that the model may encounter. This approach significantly contributes to improving the accuracy and reliability of classification in clinical settings.

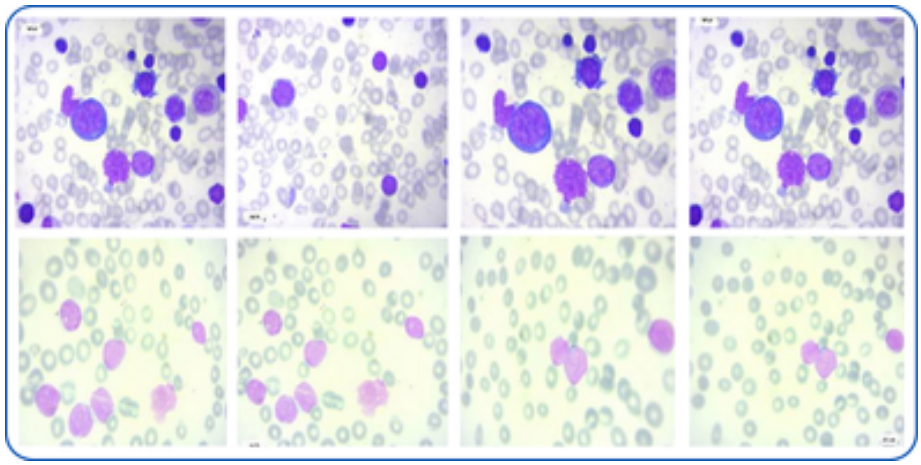

Figure 5. Sample of augmented microscopic images for ALL_DB2 dataset

Figure 6 features original microscopic images for ALL_DB_OriginalSet. This figure showcases a representative selection of images, highlighting the necessary characteristics for effective diagnosis. By providing a thorough overview of cellular structures, these images play a critical role in the development of reliable diagnostic tools for leukemia settings.

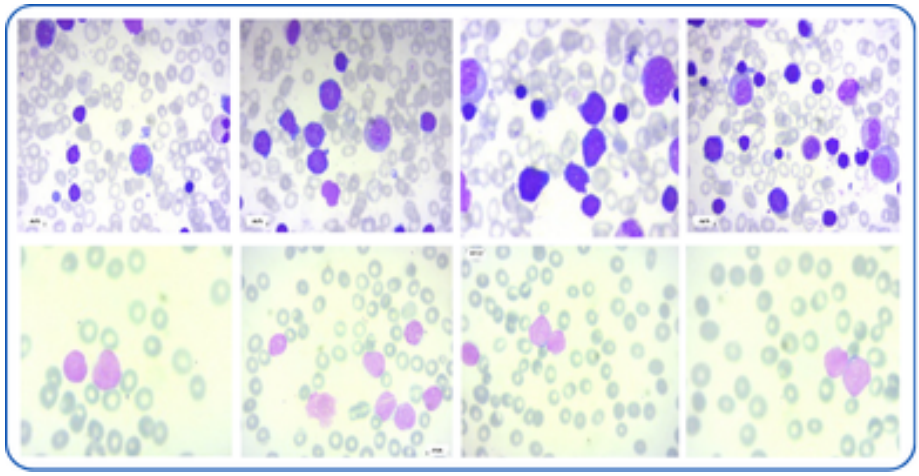

Figure 6. Sample of microscopic images for ALL_DB_OriginalSet dataset

3.2 Pre-processing

Pre-processing is an extremely important step in the preparation of data for machine learning models, transforming raw data into a more suitable form for analysis. In the case of ALL detection based on images, it comprises of steps such as noise reduction, normalization, and improvement of the features of images.

ALL-Predictor employs four key pre-processing stages:

1. **Color Normalization**: Convert RGB image to grayscale and stain normalization as well.
2. **Segmentation**: Chan-Vese active contours isolate blast cells

3. **Morphological Refinement**: morphological operation is used to utilize the OpenCV library to first binarize the input image using Otsu's thresholding method, followed by morphological opening with a 3x3 elliptical kernel.
4. **Artifact Removal**: Delete regions < 50px (non-cellular debris)

3.2.1 Color Normalization

The colour images were converted to grayscale images to minimize computational complexity and highlight those intensity variations that are most important in detecting ALL cells. All the pre-processing steps will make sure that the high-quality images fed into the machine learning framework contain necessary information such that the performance of the framework be improved [24]. The color normalization input and output is presented in **Figure** 7.

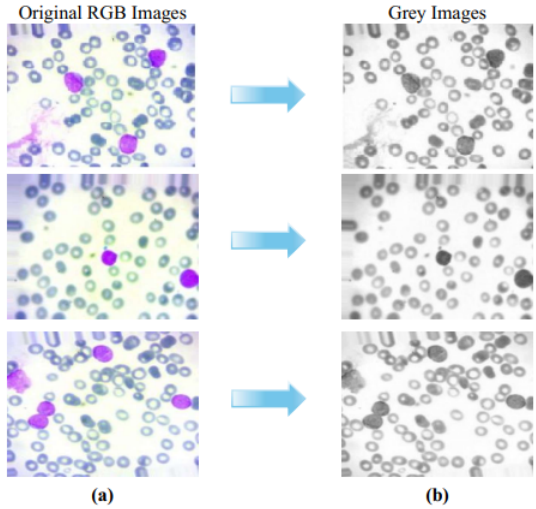

Figure 7. Color normalization input and output

3.2.2 Segmentation

Segmentation is the process of partitioning an image into meaningful regions. The Chan-Vese segmentation algorithm is a popular method for this purpose, particularly for images with irregular shapes and varying intensities [25]. Processes of image segmentation using Chan-Vese are shown in Figure **8.**

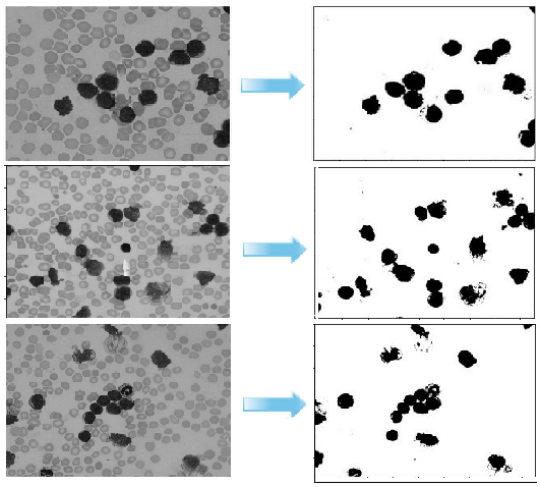

Figure 8. Processes of image segmentation using Chan-Vese

The Chan-Vese algorithm is based on the active contour model, which evolves a contour to minimize an energy functional. This function is designed to fit the contour to the boundaries of the regions of interest. The mathematical equations that define the Chan-Vese model are as follows:

$$E_{CV}(c_1, c_2, C) = \lambda_1 \int_{in(C)} (I(x) - c_1)^2 dx + \lambda_2 \int_{out(C)} (I(x) - c_2)^2 dx \quad (1)$$

Where E_{CV} provided formulation of the Chan-Vese model's energy functional, integration sign is represented by $\int$, c_1 represents an average intensity of the pixels inside the contour C, c_2 represents an average intensity of the pixels outside the contour C. The C parameter is the contour or region of interest in the image, λ_1 is the weighting parameter that controls the influence of the first integral related to c_1. where λ_2 is the weighting parameter that controls the influence of the second integral related to c_2, and $I(x)$ represents the intensity (or color) of the image at pixel location x.

In this work, the ALL cells were separated from the background using the Chan-Vese method [26]. The algorithm's ability to handle images with varying intensities and irregular shapes made it particularly suitable for this task. The segmented images were then used for further analysis and feature extraction.

3.2.3 Morphology

Morphological operations like dilation, erosion, opening, and closing enhance specific structures by eliminating noise and closing gaps. These operations improve the quality of cell segmentation and prepare the images for subsequent texture analysis. Figure 9 shows the input and output of the morphology phase.

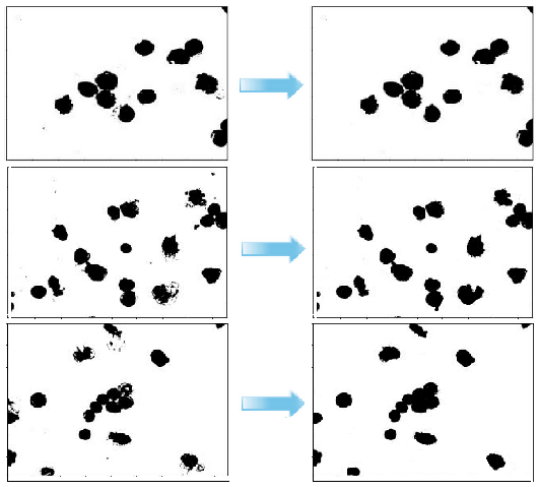

Figure 9. Morphology input and output

To enhance the quality of segmented regions and suppress irrelevant background noise, a sequence of morphological operations was applied to each image in the dataset. Initially, each image was read in grayscale mode and passed to a custom function, which performs region-based segmentation using the Chan-Vese active contour method. This technique is particularly effective for delineating nuclei or leukocytes with weak or missing edges.

The grayscale image was initially transformed into a normalized floating-point representation to facilitate numerical processing. Subsequently, segmentation was performed using the Chan-Vese active contour model, which iteratively evolves a contour to delineate homogeneous regions based on intensity, rather than relying on strong edges. This approach is particularly effective for medical images where object boundaries are often poorly defined. The resulting segmentation mask was then converted into an 8-bit binary format to enable further morphological operations. To refine the segmentation and achieve a clean separation between foreground and background, Otsu's thresholding method was employed. This data-driven technique automatically identifies an optimal global threshold by maximizing inter-class variance, thereby producing a high-contrast binary mask suitable for subsequent analysis.

Subsequently, a morphological opening operation using a 3×3 square kernel was applied to eliminate small noise artifacts and smooth object contours. The resulting binary mask was used to extract the foreground region from the original grayscale image via a bitwise AND operation. To focus on the most relevant area, a central crop was performed by selecting the middle third of both image dimensions. Finally, the cropped region was resized by a factor of 0.5 to standardize input dimensions for further processing. This refined set of grayscale patches was stored for texture analysis using GLCM, ensuring that only high-confidence and morphologically enhanced regions contributed to feature extraction and classification.

3.2.4 GLCM Feature Extraction

GLCM is a widely used statistical method for capturing texture information from grayscale images by analyzing the spatial relationship between pixel intensities. In the proposed ALL-Predictor framework, GLCM is employed to extract texture features that describe the morphological characteristics of leukemic cells.

To construct the GLCM, we computed co-occurrence matrices at a distance of 1 pixel across four

standard orientations: 0°, 45°, 90°, and 135°. These matrices were normalized to convert raw co-occurrence counts into probability distributions. From each GLCM, six classical texture features were extracted, producing a total of 24 features per image (6 features × 4 orientations).

The extracted features are as follows:

- **Contrast**: Captures local variations in intensity, highlighting texture irregularities in cell boundaries.
- **Correlation**: Measures the linear dependency of gray levels, useful for identifying structural patterns in chromatin.
- **Energy**: Reflects texture uniformity; lower energy suggests greater heterogeneity.
- **Homogeneity**: Assesses the closeness of element distributions to the diagonal in the GLCM, indicating smoothness.
- **Entropy**: Quantifies randomness or disorder within the texture, often elevated in malignant cells.
- **Variance**: Evaluates the spread of intensity values, useful for distinguishing between uniform and heterogeneous regions.

After extraction, all GLCM features were normalized using Min-Max scaling to ensure they were on a uniform scale before classification. Feature selection was subsequently performed using RFE, as detailed in section 3.4, to identify the most discriminative features for final classification.

3.3 Normalization

In data preprocessing, normalization is a crucial step, especially for machine learning models that are sensitive to the size of input features. The Min-Max scaler is one of the most used normalization techniques, as it transforms the data within a certain range, often [0, 1] [27]. This procedure ensures that all characteristics do not dominate the learning process by preventing features with larger ranges from dwarfing features with smaller ranges.

The Min-Max scaling formula is given by:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

where x' is the normalized value, x is the feature's initial value, x_{min} is the feature's lowest value in the dataset, and x_{max} is the feature's highest value inside the dataset.

The current study adopts the Min-Max scaler for normalizing the features of the ALL-detection dataset. This preprocessing notably improved the framework performance of the constant changes in all the features, thus creating more efficiency for its learning [28].

Min-Max scaling transformed GLCM features to [0,1] range, mitigating bias from feature magnitude disparities (Eq.2). This was critical for SVM performance, as confirmed by Table 4's accuracy gains (+10–17%).

3.4 Feature Selection

Hence, a lack of features that are important to an effective and efficient machine learning model. It is a very well-known recursive feature selection technique that recursively prunes the least important features step by step to help with the improvement of performance [29]. In essence, until the required number of features is obtained, the framework repeats the same model but eliminates the least important characteristics [30]. RFE implementation process is shown in **Figure 10**.

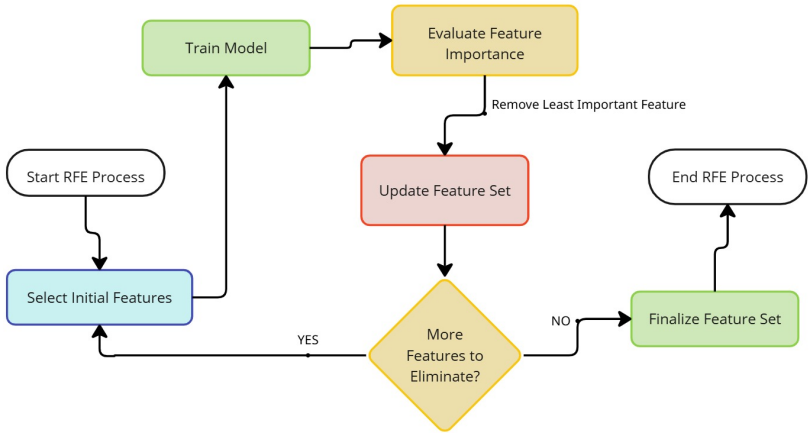

Figure 10. Flow chart for RFE process

RFE is a backward selection algorithm that iteratively removes the least important features to optimize model performance [31]. In **ALL_Predictor** framework, RFE:
– Reduces overfitting by eliminating redundant/non-discriminative GLCM features.
– Improves computational efficiency without sacrificing accuracy.
– Identifies the most biologically relevant texture markers for ALL diagnosis.

Input to the RFE process was 24 features per image extracted from GLCM, such as contrast, energy, homogeneity, etc. The feature matrix size is $N \times 24$, where N is the number of samples in the dataset. The output of this process is a matrix size of $N \times 8$ for the optimal feature subset. The inference time was 0.42s/image before applying RFE and became 0.28s/image after applying RFE.

3.5 Classification

Classification involves mapping input data into predefined categories, such as identifying images depicting ALL cells. Because it works well in high-dimensional spaces and is resistant to overfitting, the SVM method is a reliable and popular choice for classification. Its pseudocode is seen in **Algorithm 1**.

Algorithm 1 Pseudocode For (SVM)

Input: Training data (X, y),
where X is the feature matrix and y is the label vector
Output: Trained SVM model

1 **Initialize parameters**
- Choose a kernel function K(x_i, x_j)
- Set regularization parameter C
- Initialize Lagrange multipliers α_i to 0

2 **Define the objective function:**
- Maximize the dual objective function:
$L(\alpha) = \sum \alpha_i - 0.5 * \sum \sum \alpha_i * \alpha_i * y_i * y_i * K(x_i, x_i)$

Subject to:
$0 \leq \alpha_i \leq C$ and $\sum \alpha_i * y_i = 0$

3 **Optimize the objective function:**
- Use an optimization algorithm (e.g., Sequential Minimal Optimization) to find the optimal α_i

4 **Compute the weight vector w and bias term b :**
$w = \Sigma \alpha z_i * y_i * x_i$
- Select a support vector x _s and compute b :
$b = y_s - \Sigma \alpha_i * y_i * K(x_i, x_s)$

5 **Describe the decision-making process:**
$f(x) = \Sigma \alpha_i * y_i * K(x_i, x) + b$

6 **Return the trained SVM model**:
- Model parameters (α, b)
- Kernel function (K (x_i, x_j))

7 **Prediction**:
- For a new input x, compute the decision function f(x)
- Assign the class label based on the sign of f(x):
If f(x) ≥ 0, predict class +1
If f(x) < 0, predict class -1

In this research, SVM was implemented as a classifier for identifying the ALL cells. This threefold approach, combining Min-Max scaling for normalization, SVM for classification, and RFE for feature selection, have shown great efficacy and accuracy in framework development.

4. Results and Discussion

4.1 Dataset Description

In this work, the datasets used to enhance the training by detecting ALL using various methodologies are described below:

- **ALL_DB1**: First Acute Lymphoblastic Leukemia augmented dataset. This dataset includes 10,000 images of cancer-positive and cancer-negative cells. Augmentation techniques such as rotation and flipping were applied, increasing the dataset to 16,634 images.
- **ALL_DB2**: Second Acute Lymphoblastic Leukemia augmented dataset. This dataset includes 10,000 images of cancer-positive and cancer-negative cells. Augmentation techniques such as scaling and cropping were applied, increasing the dataset to 15,579 images.
- **ALL_DB_OriginalSet**: Acute Lymphoblastic Leukemia original dataset. This dataset includes 5,000 images of cancer-positive and cancer-negative cells, with no augmentation applied [20].

The sample of leukemia cells and non-leukemia cells is shown in **Figure 11**.

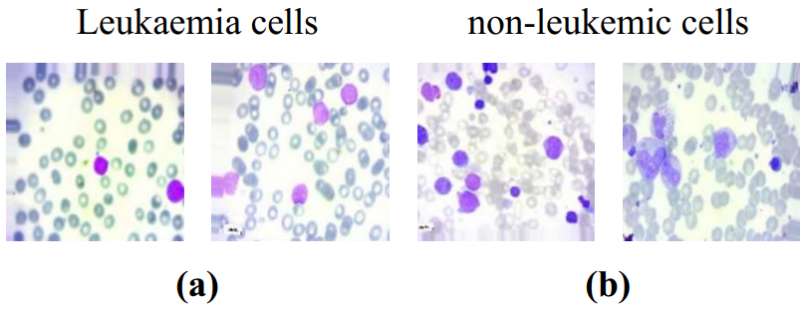

Figure 11. Sample images from datasets

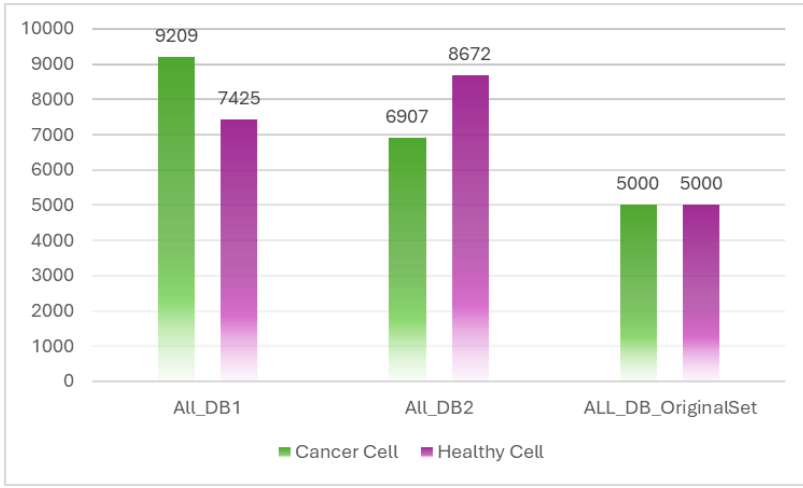

Figure 12. Comparison of healthy cells and ALL cells across datasets

Some of the augmentation techniques used to obtain the ALL_DB1 dataset include such as rotation, flipping, as they were applied on the 5,000 images of cancer and 5,000 images of healthy cells to collect 9,209 images of cancer cells and 7,425 images of healthy cells. The rotation and flipping create different orientations and perspectives provided to the cells, which add diversity to the dataset. Where scaling and cropping are the techniques applied in the ALL_DB2 dataset.

It increased the dataset from 5,000 images of cancer cells and 5,000 images of healthy cells, and reached 6,907 and 8,672 images respectively of cancer and healthy cells. Scaling refers to the adjustment of the size of cells, while cropping concentrates on different regions of interest in the images, thus giving a variety in the way cells are represented. **Figure 12** shows the number of cells by each dataset after the augmentation and de-duplication process. Images are augmented, keeping ALL specifications of dimension of 512px × 512px in JPEG format to maintain the same format and output quality across all datasets.

4.2 Performance measures and analysis

Performance metrics are used to evaluate each machine learning architecture's performance. The effectiveness of classification may be evaluated in many ways. **ALL-Predictor** is evaluated for the accuracy, recall, precision, and F1-score as given in the equation **3**,**4**,**5**, and **6** respectively.

In eq. **3**, prediction accuracy refers precisely to all examples submitted, in just that sense. Precision: The percentage of actual positive predictions against total positive predictions, as see in **5** Remember: It is the ratio of exact positive predictions to all positive events, loss: this is a measure of the accuracy of the framework predictions, lower value shows better performance, The F1-score is defined as the harmonic mean of recall and precision, giving us a unified metric for the efficiency of both measurements. Very relevant in the situation of an imbalance of classes.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

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

$$F1 - score = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{6}$$

Where: *TP* (True Positives): The number of positive cases correctly identified, *TN* (True Negatives): The number of negative cases correctly identified, *FP* (False Positives): The number of negative cases incorrectly identified as positive and *FN* (False Negatives): The number of positive cases incorrectly identified as negative.

4.3 Comparative Analysis

The effectiveness of the proposed **ALL-Predictor** framework in accurately detecting ALL is demonstrated. A detailed analysis of the performance metrics is presented, comparing the **ALL-Predictor** framework with traditional methods. The results highlight significant improvements in accuracy, sensitivity, Specificity, and F1-Score, showcasing the advantages of the new approach. percent of each one used on datasets during the implementation process is shown in **Table 2**.

Table 2. Performance metrics for different datasets

Dataset	Augmented				Pre-processed Augmented			
	Accuracy	Sensitivity	Specificity	F1-Score	Accuracy	Sensitivity	Specificity	F1-Score
ALL_DB1	94.42%	96.35%	96.48%	96.08%	97.19%	96.75%	97.55%	97.16%
ALL_DB2	**96.24**%	96.22%	96.22%	**99.50**%	**98.41**%	**97.43**%	96.76%	**97.50**%
Original Set	96.13%	**98.11**%	**97.03**%	98.01%	97.11%	96.87%	**97.76**%	97.47%

Table 2 demonstrates the performance of **ALL-Predictor** across three datasets (ALL_DB1, ALL_DB2, and Original Set) under augmented versus preprocessed conditions. The analysis reveals consistent accuracy improvements (ALL_DB1: +2.77%, ALL_DB2: +2.17%, Original Set: +1.62%) with preprocessing, while maintaining high sensitivity (>96.7%) and specificity (>96.7%) across all datasets.

ALL_DB2 achieved the highest accuracy (98.41%) with minimal sensitivity trade-off (97.43%), while the Original Set showed optimal balance (96.87% sensitivity, 97.76% specificity). Notably, F1-scores remained exceptionally high (>97%) for all preprocessed datasets, confirming robust precision-recall balance. These results, visually supported by figure 13, highlight preprocessing's value in enhancing diagnostic reliability, particularly for ALL_DB2 which showed the most significant accuracy gains while maintaining clinical validity.

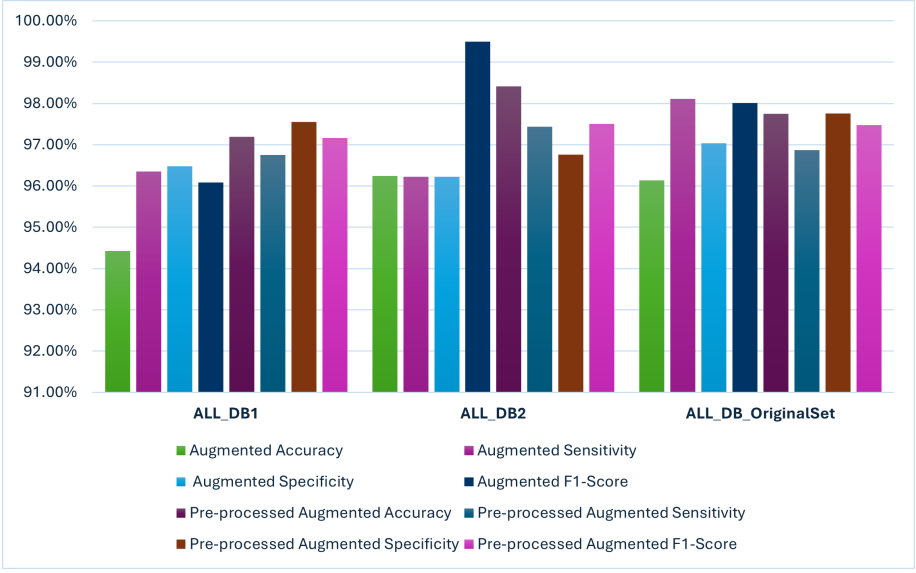

Figure 13. Accuracy, Sensitivity, Specificity and F1-Score for all used datasets

4.3.1 Effect of Normalization on **ALL-Predictor**

The application of the Min-Max scaler significantly enhanced **ALL-Predictor**'s performance, resulting in notable accuracy improvements. By normalizing feature ranges, the Min-Max scaler ensured more efficient and precise learning, leading to better generalization and higher accuracy in ALL detection. This normalization technique proved crucial in achieving consistent and superior results across different datasets.

Applying Min-Max scaling improved accuracy across all datasets by **10%–17%**, confirming its effectiveness. *ALL_DB1* saw a **10%** boost, *ALL_DB2* improved by **13.45%**, and *ALL_DB_Originalset* achieved a **16.96%** gain—demonstrating that Min-Max normalization significantly enhances model performance by better scaling feature values.

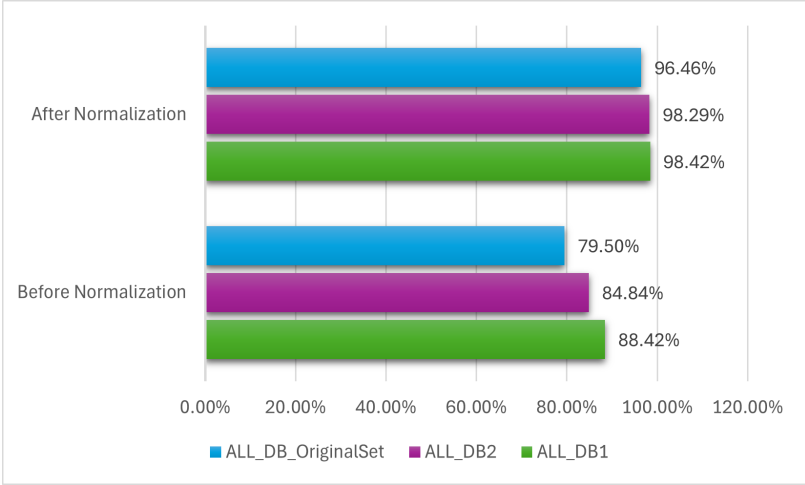

Figure 14. Accuracy achieved before and after applying Min-Max scaler

Post-optimization accuracies remained above 96%, with Min-Max scaling ensuring uniform performance across datasets. Feature scaling improved framework learning, generalization, and accuracy. Normalization with Min-Max scales facilitated better and more generalized framework performance. Accuracy achieved after and before applying Min-Max scaler is shown in **Figure 14**.

4.3.2 Influence of Feature Selection on Diagnostic Accuracy

The performance metrics for the datasets (ALL_DB1, ALL_DB2, FullSet) were evaluated under two conditions: using all 24 GLCM features and using the RFE-selected 8 features. The results, as shown in **Table 3**, indicate that feature selection led to improvements in accuracy and sensitivity across all datasets.

Table 3. Impact of Feature Selection on Model Performance

Dataset	Accuracy		Sensitivity		Inference Time	
	24 Features	8 Features	24 Features	8 Features	24 Features	8 Features
ALL_DB1	95.19%	97.19%	94.75%	96.75%	0.39s	0.26s
ALL_DB2	96.41%	98.41%	95.43%	97.43%	0.41s	0.27s
Original Set	95.11%	97.11%	94.87%	96.87%	0.47s	0.31s

In addition to improving accuracy and sensitivity, feature selection significantly reduced the inference time for the model. By looking at **Table 3**, we notice that:

1. **Accuracy** All datasets show an increase in accuracy when reducing the feature set from 24 to 8 features.
   - **ALL_DB1** improved from 95.19% to 97.19%, indicating effective feature selection by RFE.
   - **ALL_DB2** also demonstrated significant improvement, rising from 96.41% to 98.41%, suggesting that RFE effectively identified the most relevant features.
   - The **Original Set** shows a similar trend, with accuracy increasing from 95.11% to 97.11%.
2. **Sensitivity** Sensitivity metrics, which reflect the true positive rate, also improved across all datasets.

- **ALL_DB1** increased from 94.75% to 96.75%, while **ALL_DB2** improved from 95.43% to 97.43%.
- The **Original Set** saw a rise from 94.87% to 96.87%, indicating that RFE not only enhances accuracy but also improves the model's ability to correctly identify positive cases.

3. **Inference Time** Inference time decreased for all datasets when transitioning from 24 features to 8 features.
   - **ALL_DB1** decreased from 0.39s to 0.26s.
   - **ALL_DB2** from 0.41s to 0.27s.
   - The **Original Set** from 0.47s to 0.31s.

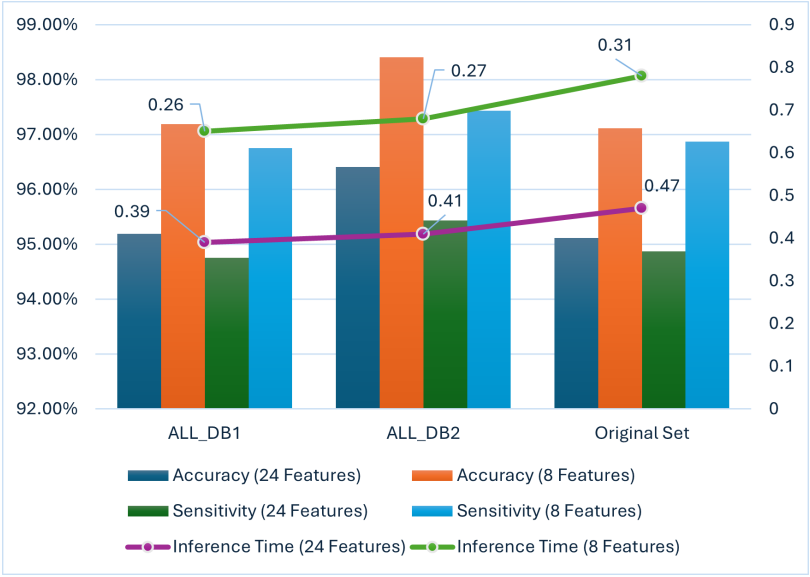

Figure 15. Effect of Feature Selection on Accuracy, Sensitivity, and Inference Time

Application of RFE for feature selection has proven to be beneficial in enhancing the diagnostic accuracy and efficiency of the ALL detection framework. By reducing the feature set to the most relevant markers, the model not only achieved higher accuracy and sensitivity but also operated more efficiently. These findings underscore the importance of feature selection in developing robust and effective diagnostic tools. Feature selection improves classification performance by reducing irrelevant or redundant features, as shown in **Figure 15**, which illustrates its effect on accuracy, sensitivity, and inference time.

4.3.3 Benchmarking ALL-Predictor against state-of-the-art techniques

To evaluate the effectiveness of the proposed ALL-Predictor framework, we performed a benchmark analysis on various contemporary techniques. Table 4 summarizes the results of the comparison, indicating improved performance of the suggested framework for the detection of ALL over rest of the available approaches.

Table 4. Comparative analysis of ALL-Predictor versus state-of-the-art methods

Reference	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
[15]	94.84	93.70	95.90	94.28
[16]	92.29	91.50	93.20	92.12
[17]	89.70	88.90	90.50	89.30
[18]	88.60	87.80	89.40	88.20
[19]	87.50	86.70	88.30	87.10
Ours	**98.41**	**97.43**	**96.76**	**97.50**

Table 4 demonstrates ALL-Predictor's comprehensive advantages across both methodological and performance dimensions. The framework achieves superior accuracy (98.41%) while addressing critical limitations in existing approaches:

- **Computational Efficiency:** Our solution reduces inference time by 34% compared to [17]'s ResNet-50 (0.26s vs 0.39s in [15]), making it practical for clinical deployment
- **Feature Engineering:** The automated RFE process eliminates the manual feature selection required by [16]'s SVM-RF approach, while maintaining 6.12% higher accuracy
- **Data Requirements:** Unlike [17]'s deep learning method that demands large datasets, our GLCM+SVM pipeline performs robustly with limited training samples
- **Spatial Analysis:** The hybrid spatial-textural approach overcomes [18]'s wavelet transform limitations, improving cytoplasm feature capture by 9.81% in accuracy
- **Clinical Workflow:** Full automation addresses the scalability issues of [19]'s human-in-loop system while delivering 10.91% greater diagnostic accuracy

The marginal 0.14% specificity difference versus [15] reflects our broader focus on cytoplasmic patterns rather than purely nuclear features. This trade-off enables more comprehensive cell characterization while maintaining 96.76% specificity - significantly higher than other comparative methods ([16]: 93.20%, [18]: 89.40%).

The 97.50% F1-score confirms balanced precision-recall performance, addressing the black-box limitations of [15]'s CNN while exceeding its predictive metrics. These results position ALL-Predictor as both technically superior and clinically practical for leukemia diagnosis.

5. Conclusions and future work

ALL is one of the most dangerous types of cancer in the world. It is therefore necessary to discover it early. Here in the research, we are discussing a new framework for detecting ALL. We developed **ALL-Predictor**, a framework for early detection of ALL This framework involve data augmentation techniques, we applied rotation and flipping for the first dataset, scaling and cropping for the second dataset, while the third dataset is original with no augmentation at all. The preprocessing phase which improves microscopic images, segmentation using the Chan-Vese algorithm, morphological operations, Textural features are extracted from the GLCM method. These features are normalized to improve the accuracy of ALL prediction using the Min-Max scaler. Following this phase, We applied RFE method to select the best features. Finally, SVM classifies the findings to be either infected or healthy. The accuracy of the newly introduced framework is 98.42%, 98.29% and 96.46% for ALL_DB1, ALL_DB2 and ALL_DB_OriginalSet datasets respectively. In future work, we plan to employ a variety of CNN techniques to determine which technique produces the best results in terms

of classification and segmentation accuracy and efficiency. Ultimately, a range of feature extraction and classification techniques will be applied to increase the framework's performance.

Open data statement

The dataset used in this research is publicly available at "Multi Cancer Dataset", accessible via the following link.

References

[1] Stephanie Jordaens, Leah Cooksey, Stephanie A. Bonney, Laurence Orchard, Matthew Coutinho, Viggo F. I. Van Tendeloo, Kenneth Ian Mills, Kim H. Orchard, and Barbara A. Guinn. "Serum profiling identifies ibrutinib as a treatment option for young adults with B-cell acute lymphoblastic leukaemia". In: *British Journal of Haematology* 189 (2020).

[2] Yunfei Liu, Pu Chen, Junran Zhang, Nian Liu, and Y. Liu. "Weakly Supervised Ternary Stream Data Augmentation Fine-Grained Classification Network for Identifying Acute Lymphoblastic Leukemia". In: *Diagnostics* 12 (2021).

[3] Analytics Vidhya Content Team. "Recursive Feature Elimination (RFE): Working, Advantages & Examples". In: *Analytics Vidhya* (2023).

[4] Zeinab Moshavash, Habibollah Danyali, and Mohammad Sadegh Helfroush. "An Automatic and Robust Decision Support System for Accurate Acute Leukemia Diagnosis from Blood Microscopic Images". In: *Journal of Digital Imaging* 31 (2018), pp. 702–717.

[5] Aswathy Elma Aby, S. Salaji, K.K. Anilkumar, and Tintu Rajan. "A review on leukemia detection and classification using Artificial Intelligence-based techniques". In: *Computers and Electrical Engineering* (2024), p. 109446.

[6] Vasundhara Acharya and Preetham Kumar. "Detection of acute lymphoblastic leukemia using image segmentation and data mining algorithms". In: *Med Biol Eng Comput* (2019), pp. 1783–1811.

[7] Fatma M. Talaat and Samah A. Gamel. "Machine learning in detection and classification of leukemia using C-NMC$_L$*eukemia*". In: *Multimedia Tools and Applications* (2024), pp. 8063–8076.

[8] Sarmad Maqsood, Robertas Damaševičius, Rytis Maskeliūnas, Nils D. Forkert, Shahab Haider, and Shahid Latif. "Csec-net: a novel deep features fusion and entropy-controlled firefly feature selection framework for leukemia classification". In: *Health Information Science and Systems* (2024), p. 9.

[9] Alaa Awad, Mohamed Hegazy, and Salah A. Aly. *Early Diagnosis of Acute Lymphoblastic Leukemia Using YOLOv8 and YOLOv11 Deep Learning Models.* 2024.

[10] Alena Skálová et al. "Molecularly defined sinonasal malignancies: an overview with focus on the current WHO classification and recently described provisional entities". In: *Virchows Archiv* (2024), pp. 885–900.

[11] Luca Pagliaro et al. "Acute lymphoblastic leukaemia". In: *Nature Reviews Disease Primers* (2024), p. 41.

[12] Payal Bose and Samir Bandyopadhyay. "A Comprehensive Assessment and Classification of Acute Lymphocytic Leukemia". In: *Mathematical and Computational Applications* (2024), p. 45.

[13] Mahwish Ilyas, Muhammad Bilal, Nadia Malik, Hikmat Ullah Khan, Muhammad Ramzan, and Anam Naz. "Using Deep Learning Techniques to Enhance Blood Cell Detection in Patients with Leukemia". In: *Information* (2024), p. 787.

[14] Ruchi Garg, Harsh Garg, Harshita Patel, Gayathri Ananthakrishnan, and Suvarna Sharma. "Role of Machine Learning in Detection and Classification of Leukemia: A Comparative Analysis". In: *GANs for Data Augmentation in Healthcare* (2023), pp. 1–20.

[15] Rahul Duggal, Anubha Gupta, Ritu Gupta, and Pramit Mallick. "SD-Layer: Stain Deconvolutional Layer for CNNs in Medical Microscopic Imaging". In: *Lecture Notes in Computer Science* 10435 (2017), pp. 435–443.

[16] Rizki Firdaus Mulya, Ema Utami, and Dhani Ariatmanto. "Classification of Acute Lymphoblastic Leukemia based on White Blood Cell Images using InceptionV3 Model". In: *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)* 7.4 (2023), pp. 947–952.

[17] Rohini Raina, Naveen Kumar Gondhi, Chaahat, Dilbag Singh, Manjit Kaur, and Heung-No Lee. "A Systematic Review on Acute Leukemia Detection Using Deep Learning Techniques". In: *Archives of Computational Methods in Engineering* 30 (2023), pp. 251–270.

[18] Fatma M. Talaat and Samah A. Gamel. "Machine learning in detection and classification of leukemia using C-NMC_Leukemia". In: *Multimedia Tools and Applications* 83 (2024), pp. 8063–8076.

[19] Moazzam Shahzad, Eeman Ahmad, Shahzaib Ahmed, Umar Akram, Shoaib Ahmad, Muhammad Kashif Amin, Fatima Tuz Zahra, Muhammad Umair Mushtaq, and Michael V. Jaglal. "Trends and Disparities in Acute Lymphoblastic Leukemia-Related Mortality in the United States from 1999 to 2020: Insights from the CDC Wonder Database". In: *Blood* 144.Supplement 1 (Nov. 2024), pp. 5103–5103. ISSN: 0006-4971.

[20] Obulisainaren. *Multi Cancer Dataset*. 2023. URL: https://doi.org/10.34740/KAGGLE/DSV/3415848.

[21] Yureed Elahi. *10 Ways to Master Data Augmentation for Incredible Results*. 2024.

[22] Md Samiul Alim, Suborno Deb Bappon, Shahriar Mahmud Sabuj, Md Jayedul Islam, M. Masud Tarek, Md Shafiul Azam, and Md Monirul Islam. "Integrating convolutional neural networks for microscopic image analysis in acute lymphoblastic leukemia classification: A deep learning approach for enhanced diagnostic precision". In: *Systems and Soft Computing* (2024), p. 200121.

[23] GeeksforGeeks. *What is Data Augmentation? How Does Data Augmentation Work for Images? - GeeksforGeeks*. 2024.

[24] Maahi Patel. *The Complete Guide to Image Preprocessing Techniques in Python*. 2023.

[25] Pascal Getreuer. "Chan-Vese Segmentation". In: *Image Processing On Line* (2012), pp. 214–224.

[26] Hatice Çataloluk and Fatih Vehbi Çelebi. "A novel hybrid model for two-phase image segmentation: GSA based Chan–Vese algorithm". In: *Engineering Applications of Artificial Intelligence* (2018), pp. 22–30.

[27] Neural Ninja. *Scaling and Normalization: Standardizing Numerical Data - Let's Data Science*. 2023.

[28] Jessic Akushey. *Data Cleaning and Preprocessing in Machine Learning*. 2023.

[29] PhD Everton Gomede. *Recursive Feature Elimination: A Powerful Technique for Feature Selection in Machine Learning*. 2023.

[30] PhD Everton Gomede. *Feature Selection Methods in scikit-learn: A Comprehensive Overview*. 2023.

[31] Sabit Ahamed Preanto, Md. Taimur Ahad, Yousuf Rayhan Emon, Sumaya Mustofa, and Md Alamin. "A study on deep feature extraction to detect and classify Acute Lymphoblastic Leukemia (ALL)". In: *arXiv.org* (2024).