Scientific Research Publishing

Journal of Computer and Communications, 2026, 14(3), 278-296
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

A Hybrid Deep and Machine Learning Framework with Feature Selection for Automated Classification of Acute Lymphoblastic Leukemia

Husne Farah[1,2*], Fahmida Islam[1,3], Shuvo Biswas[4,5], Mohammad Shorif Uddin[2]

[1]Department of Computer Science and Engineering, The People's University of Bangladesh, Dhaka, Bangladesh
[2]Department of Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh
[3]Department of Information and Communication Engineering, Islamic University, Kushtia, Bangladesh
[4]Department of Computer Science and Engineering, Bangladesh University of Business and Technology, Dhaka, Bangladesh
[5]Department of Information and Communication Technology, Mawlana Bhashani Science and Technology University, Tangail, Bangladesh
Email: *husnefarahcse.pub@gmail.com, fahmidaislam.ice@gmail.com, shuvo@bubt.edu.bd, shorifuddin@juniv.edu

How to cite this paper: Farah, H., Islam, F., Biswas, S. and Uddin, M.S. (2026) A Hybrid Deep and Machine Learning Framework with Feature Selection for Automated Classification of Acute Lymphoblastic Leukemia. *Journal of Computer and Communications*, **14**, 278-296.
https://doi.org/10.4236/jcc.2026.143014

Received: March 1, 2026
Accepted: March 28, 2026
Published: March 31, 2026

Abstract

Acute Lymphoblastic Leukemia (ALL), a highly aggressive subtype of blood cancer, demands early and accurate diagnosis to improve patient outcomes. Traditional diagnostic methods relying on manual microscopic analysis are frequently tedious and susceptible to human error. This study proposes a hybrid intelligent framework that combines deep learning (DL), machine learning (ML), and statistical feature selection to automate the classification of ALL from microscopic blood smear images. A publicly available dataset containing 3262 images is utilized. Deep learning models—VGG16, VGG19, ResNet50, and MobileNet—are used to extract high-level features, which are then fed into four ML classifiers: K-Nearest Neighbor, Naive Bayes, Random Forest, and Support Vector Machine (SVM). Among all configurations, the ResNet50 + SVM combination gave the top result of 99.54%. Further enhancement using Analysis of Variance (ANOVA) for feature selection increased the accuracy to 99.69%. The proposed hybrid approach demonstrates strong potential for clinical deployment as a reliable and efficient tool for automated leukemia diagnosis.

Keywords

Blood Cancer, Acute Lymphoblastic Leukemia, Machine Learning, Deep Learning, Feature Selection

1. Introduction

Cancer is a condition marked by the unregulated growth of unusual cells, which can quickly invade and spread to various organs in the body [1]. In general, the most familiar cancers are skin cancer, lung cancer, blood-related cancers such as leukemia and lymphoma, and breast cancer [2]. The World Health Organization (WHO) [3] proclaims that lung cancer accounts for about 9.2 million deaths, skin cancer accounts for about 1.7 million deaths, while breast cancer has led to approximately 627,000 deaths [4] [5]. Leukemia, in particular, has a very high fatality rate. It is an aggressive tumor that originates in bone tissue because of uncontrolled cloning of underdeveloped white blood cells (WBC). Leukemia [6] [7] ranks as one of the most commonly identified cancers in the United States, along with cancers of the prostate, breast, colon, and lung. Estimates from the End Results (SEER) Program, Epidemiology, and U.S. Surveillance indicate that around 60,650 new leukemia cases were identified in the US in 2022, resulting in approximately 24,000 deaths. According to WHO cancer databases [8], the likelihood of developing leukemia differs greatly depending on the region and the specific subtype of the disease. In India, childhood cancers account for approximately 4% of all cancer cases, with leukemia being the most prevalent type, representing nearly 24% - 29% of pediatric cancers [9]. In 2019, there were around 61,780 diagnosed cases of leukemia in the Americas, with an additional 9900 cases in the UK. According to the National Institutes of Health (NIH), the global number of currently detected leukemia cases worldwide rose between 1990 and 2017/2018. The estimated cases increased from approximately 345,000 - 354,000 to more than 518,000, as reported in Global Burden of Disease studies. Despite this increase in the total number of cases, the Age-Standardized Incidence Rate (ASIR) of leukemia showed a gradual decline during the same period, decreasing by about 0.43% annually [10] [11].

Leukemia is separated into two primary types: chronic leukemia (CL) and acute leukemia (AL). While CL develops slowly over time, AL, if untreated, typically results in a life expectancy of just three months [6]. AL is classified into two forms under the French-American-British (FAB) classification system: Acute Lymphoblastic Leukemia (ALL) and Acute Myeloid Leukemia (AML) [12]. Similarly, CL is classified as Chronic Lymphocytic Leukemia (CLL) and Chronic Myeloid Leukemia (CML) [12]-[15]. ALL is a very aggressive tumor that affects both children and adults, resulting in about 25% [16] of all pediatric cancer cases. In this case, "acute" reflects its rapid progression, which can be fatal within months if untreated. "Lymphocytic" denotes its origin from lymphocyte progenitors, one kind of white blood cell. ALL spreads primarily in bone marrow stem cells and contaminates quickly throughout the body, damaging organs including the nervous system, brain, liver, lymph nodes, and spleen [17]-[20]. Symptoms of ALL include nose bleeding, gum bleeding and fatigue, frequent infections, joint pain, enlarged lymph nodes, and fever [21] [22]. This type of leukemia primarily affects the blood and bones [8] [23] [24]. The disease has been referred to as "acute young adult-

hood leukemia" since it is more frequent in children than adults. When detected early, ALL is treatable; however, if left untreated, it can be fatal within months [25]-[27].

The FAB classification system divides ALL into three types of categories: L1, L2, and L3. L1 cells are small, with uniform nuclei, few nucleoli, and little cytoplasm. L3 cells are usually of standard to large size, with prominent cytoplasmic vacuoles, whereas L2 cells are larger with irregular nuclear structures. Early and accurate detection of ALL significantly improves survival rates [21] [28].

Hematologists typically diagnose ALL using blood test samples and bone marrow biopsy examinations under a microscope. However, the reliability of these experiments depends on medical expertise, and prolonged microscope use can compromise accuracy [21] [29]-[31]. Moreover, human-dependent diagnoses often result in errors and delays. To overcome these limitations, automated diagnostic approaches are essential for improving speed, accuracy, and efficiency. Recently, artificial intelligence (AI), deep learning (DL), and machine learning (ML) have emerged as novel technologies for assisting in medical decision-making. Several automated diagnostic algorithms have been presented to identify ALL from blood smear images without human intervention [8] [32]. However, this study focuses on a smart framework for classifying ALL and identifying affected tissues by leveraging ML, DL, and feature selection methods. The important contributions of this manuscript are as follows:

- We employ multiple deep learning models (VGG16, VGG19, MobileNet, and ResNet50) to automatically retrieve rich feature representations from blood smear images for ALL detection.
- We integrate a statistical feature selection method (ANOVA) to identify and retain the most relevant features, thereby enhancing classification accuracy and reducing computational overhead.
- We apply four machine learning classifiers to the selected features and demonstrate that the ResNet50-SVM combination achieves superior performance, offering a robust and scalable solution for automated ALL classification.

The other parts of this work are organized as follows. Section 2 outlines the literature review. Section 3 provides the projected methodology. Section 4 presents the experimental results and analysis. Finally, Section 5 discusses the conclusion.

2. Literature Review

Recently, ML and DL approaches have been used to detect and categorize ALL. A summary of these approaches is provided below.

Researchers in the paper [8] developed a unique Bayesian-based optimal CNN approach for recognizing ALL in microscopically smeared pictures. The improved CNN model obtained flawless performance on the evaluation set thanks to Bayesian optimization. In the study [19], a hybrid InceptionV3-XGBoost structure was designed to identify ALL using microscopic pictures of white blood cells. The sim-

ulation used InceptionV3 for feature extraction and XGBoost for classification. The mixed approach obtained an F1 score of 0.986. In author [22], an intelligent model was created by integrating Support Vector Machine (SVM) classification. They conducted the study using 4000 lymphocyte samples from the Hayatabad Medical Complex. In their publication [33], the researchers suggested a ViT-CNN ensemble model that combines Vision Transformer and CNN for ALL diagnosis. Using the ISBI 2019 dataset containing 10,661 cell pictures, the model provided a higher degree of accuracy (0.991). In this paper [34], a 2 × 2 max-pooling and ten convolutional layers with 6 ML approaches were presented for ALL categorization. The DL systems achieved three distinct accuracy levels: 81.63% (ResNet50), 84.62% (VGG16), and 82.10% (proposed model). The ML models had the following precision: 81.72% for RF, 79.88% for LR, 79.28% for SVM, 77.89% for KNN, 68.91% for SGD, and 27.33% for MLP. In research [35], the researchers presented an attention-based CNN model that included the Efficient Channel Attention block and VGG16 classifier. The predicted result scored 91.1% reliability. Contrary to this, the paper [36] used Mask R-CNN to segment white blood cells and contrast augmentation methods to increase picture quality. In their publication [37], the authors devised an intelligent CNN-based technique for automated lymphoblast recognition in single-cell pictures. The proprietary ALL-NET model, trained on the C-NMC 2019 dataset, has a peak precision of 95.54%. In publication [38], the researchers developed a DL structure for leukemia identification, combining the adam optimizer and tversky loss function. The algorithm was developed on an array of data gathered by the Shahid Ghazi Tabatabai Cancer Institute and correctly recognized 99% of ALL and Acute Myeloid Leukemia (AML) patients. Author [39] created an automatic ALL recognition classifier by leveraging EfficientNet-B3 architecture. They built their model utilizing the C-NMC Leukemia data set, which has 27,558 RBC clinical data. The suggested model was 98.31% accurate with a Dice similarity coefficient of 0.981. Study [40] describes the development of machine learning-based approaches for predicting colorectal cancer survival using SEER data, achieving an AUC of 0.804 for 5-year survival prediction, outperforming conventional staging systems. In research [3], an AlexNet-GRU model was proposed for breast cancer detection in lymph nodes by combining CNN-GRU and CNN-LSTM systems. **Table 1** summarizes some of the existing publications for ALL categorization.

Based on the research previously mentioned, we can infer that certain studies [36]-[38] used only DL approaches, whereas other research [22] [34] used both ML and DL to identify ALL. In some studies [36] [38] [40], optimization strategies were used to improve the model's efficiency by lowering loss. The majority of research concentrated on binary categorization of contaminated blood cells, with little attention to multiclass categorization. To address the aforementioned restrictions, this study proposed a framework for multiclass classification by leveraging ML and feature selection algorithms. The offered framework not only enhances the classification performance but also minimizes the operational cost.

Table 1. Summary of some published papers on blood cancer classification.

Ref.	Method	Strength	Drawbacks
Atteia *et al.* (2022) [8]	Bayesian-based DL architecture	Achieved 100% accuracy on the test set through Bayesian optimization.	A limited dataset size may impact generalizability.
Ramaneswaran *et al.* (2021) [19]	Hybrid InceptionV3-XGBoost	High F1-score of 0.986, leveraging transfer learning for effective classification.	Missing dataset-specific nuances.
Arbab *et al.* (2022) [22]	AlexNet with SVM	Achieved 98% accuracy by combining a CNN feature extractor with an SVM classifier.	Small amount of data set.
Jiang *et al.* (2021) [33]	ViT-CNN	Achieved 99.03% accuracy with robust ensemble capabilities.	High computational complexity.
Rezayi *et al.* (2021) [34]	10-layer CNN, ResNet50, and VGG16	Versatility in combining deep learning and traditional ML techniques.	Lower accuracy compared to advanced models.
Zakir *et al.* (2021) [35]	Attention-based CNN with ECA module + VGG16	Extracting high-quality deep features.	Accuracy is lower than that of other reviewed models.
Revanda *et al.* (2022) [36]	Mask R-CNN	Enhanced segmentation quality and improved diagnosis in low-light conditions.	Lack of ensemble techniques
Sampathila *et al.* (2022) [37]	ALL-NET with a custom CNN	Achieved the highest accuracy (95.54%) for binary classification.	Applicable only to binary classification.
Ansari *et al.* (2023) [38]	Customized CNN model	Detecting both ALL and AML cases effectively.	Dataset size is not specified.
Abd *et al.* (2023) [39]	EfficientNet-B3	Achieved 98.31% accuracy and 98.05% Dice Similarity Coefficient, demonstrating superior performance.	Focused only on binary classification.

3. Methods and Methodology

The following part describes the building process for the suggested structure. Figure 1 exhibits the general design of the proposed method. This structure has four stages: 1) data preparation, 2) extracting features utilizing DL techniques, 3) selecting features utilizing feature pickers, and 4) final classification utilizing an ML classifier. In stage 1, the initial information is processed utilizing methods for image processing such as cropping and scaling. In stage 2, four deep learning algorithms (VGG16, VGG19, ResNet50, and MobileNet) are used to obtain high-quality characteristics from the previously processed information. In stage three, we use a well-known optimization approach called analysis of variance (ANOVA) to select rich characteristics from the extracted characteristics. Finally, in stage 4, four ML classifiers (NB, SVM, RF, and KNN) are employed to categorize ALL using the characteristics that were selected. The following subsections detail each setup in consecutive order.

3.1. Dataset Description

This study used a set of data from the Kaggle repository [41]. The photos in this

collection were created in the bone marrow laboratory of Taleqani Hospital (Tehran, Iran) and include 3262 genuine peripheral blood sample images. These photos were obtained from 89 patients, 25 of whom were recognized as fit, whereas the other 64 were confirmed with ALL. The set of data is classified into two main categories: malignant and benign. The malignant group is further split into three subgroups: early, pro-B, and pre-B. The photographs were taken with a Zeiss camera fitted to a microscope at 100 × magnification and saved in the form of a JPG. A professional used the technique of flow cytometry to precisely classify these photos. Figure 2 depicts one of the sample photos utilized in this investigation. The dataset used in this work was divided into two parts: train (80%) and test (20%) based on patient-wise split. This strategy ensures that images belonging to the same patient do not appear in both the training and testing sets. This approach prevents information leakage and ensures a more realistic evaluation of model performance. Table 2 provides an overview of the entire dataset.

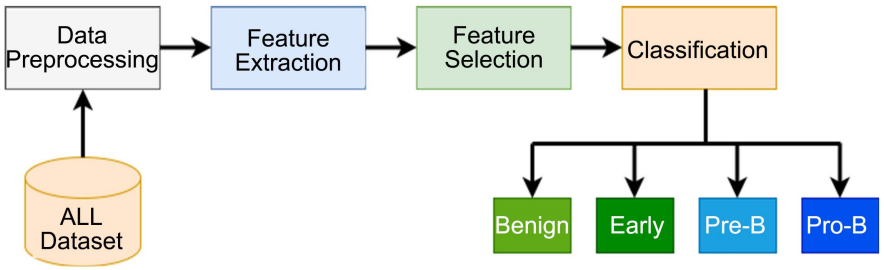


Figure 1. Overall architecture of the proposed system.

Table 2. Working dataset distribution.

Dataset	Class	Train (80%)	Test (20%)
ALL	Benign	410	102
	Early	783	196
	Pre-B	764	191
	Pro-B	637	159
	Total	2594	648

3.2. Data Preprocessing

The data pretreatment pipeline consists of various processes that guarantee the input photos are appropriately structured for further examination. First, the photos are loaded into OpenCV, and any illegible ones are discarded. Each picture is then transformed to grayscale, which reduces complexity yet retains fundamental integrity. Otsu's thresholding approach is used to separate both foreground and background objects. This approach provides a binary picture in which the item appears in white and the background is black. The foreground pixels are subsequently recognized so that only the area of interest remains. If no foreground pixels are identified, the picture is discarded. The captured item is cropped to elimi-

nate any superfluous background areas. After cropping, the picture is enlarged to 224 × 224 pixels to ensure uniformity in the input dimension. This preliminary processing method improves the overall quality of the data being input by removing noise, standardizing image parameters, and focusing on specific areas. We handle the grayscale images for training the CNN architecture through image resizing and normalization steps. Firstly, all images are resized to 224 × 224 pixels to match the input requirements of the CNN architectures. Then, the pixel values of each image are normalized to the range [0, 1] prior to feature extraction.

3.3. Feature Extraction Using the DL Model

In this research, four DL-based feature extractors (ResNet50, VGG19, MobileNet, and VGG16) are applied to retrieve the DL features from the blood microscopic images. In the feature-extraction setup, features are extracted from the dataset by following different steps so the feature vectors are reproducible. Firstly, all CNN backbones were initialized with ImageNet pre-trained weights. Then, the convolutional base of each network was used as a feature extractor. After that, the final classification layers were removed, and features were extracted from the Global Average Pooling (GAP) layer. During feature extraction, the convolutional layers were kept frozen to preserve the learned representations and reproduce the feature vector. Each DL model extracts a different number of features, such as the VGG model extracts 512 features, ResNet50 extracts 2048 features, and MobileNet extracts 1024 features. Algorithm 1 shows the feature extraction procedure of each model. The description of each DL model is given in the next subsection.

Algorithm 1. Feature extraction procedure using DL models

Input: 2D microscopic data

Output: DL Feature Map

Initialization:

1. $p = 2P - 1$, *for* $P = 1, 2, 3 \ldots \ldots \ldots \ldots p$

2. $D \leftarrow$ *Input data*

3. $X_p \leftarrow$ *Use the median filter on the input data D with kernel size* $p \times p$

4. $M_f \leftarrow$ *Extracted Feature Map*

Start:

1. For P = 1 to n:

2. Calculate X_p

3. Apply (D, X_p) *to find* $M_f \mid M_f \{R_0, R_1, \ldots\ldots R_{14}\}$

4. $M_f \leftarrow M_f$

5. end for

6. display M_f

End

1) VGGNet

VGGNet [42] is a CNN classifier developed by the Visual Geometry Group at the University of Oxford in 2014. This architecture employs a series of 3 × 3 convolutional layers stacked deeper compared to earlier architectures. This architec-

ture uses max-pooling layers to sequentially minimize spatial dimensions while enhancing characteristic richness. It utilizes fully connected layers at the end for classification. In this work, we utilized two VGGNet architectures named VGG16 and VGG19 to extract features, which have 16 and 19 weight layers, respectively. These two networks extract 512 features separately from the input data.

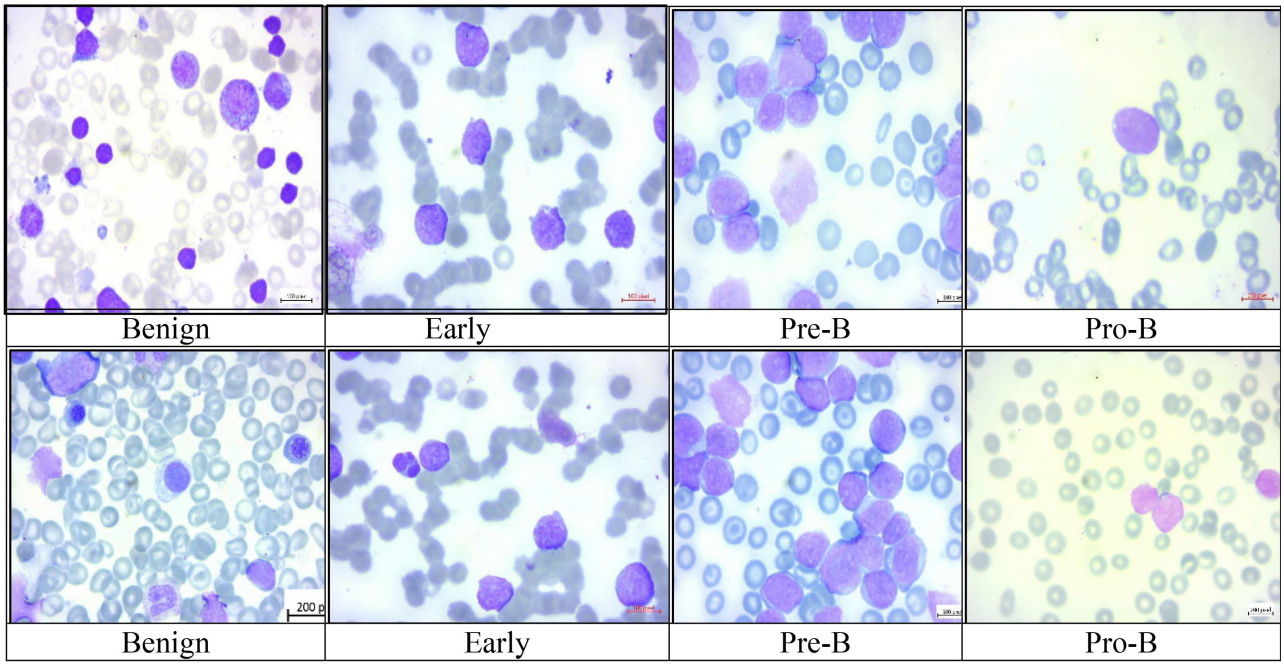


Figure 2. Example of some microscopic images.

2) ResNet50

The ResNet50 [43] model comprises 50 distinct layers with 2M variables. It consists of several components: 64 kernels, convolution level, dense layer, and max-pooling level. The residual block of a ResNet50 model allows for the deterioration issue and removes the vanishing issue. Furthermore, the skip connection block works as a super pathway. In this work, ResNet50 takes ALL images as input and extracts 2048 DL features using the last layer.

3) MobileNet

MobileNet [44] is a lightweight network that is commonly applied in embedded strategies for diagnostic-based systems. The DC (depth-wise convolutions) allows this network to minimize the training time. The operation of the MobileNet network is first the DC, followed by the PC (point-wise convolution). The convolution process of the MobileNet is expressed by the following Equation (1).

$$TK = \sum_{j=1}^{c_j} T_j K \tag{1}$$

here, T indicates the input tensor, K indicates the kernel, T_j denotes the tensor's j-th component, and * indicates the CO (convolution operation). However, after performing the component-wise product and moving K over T in the convolu-

tional layer, the final result of the CO is calculated by combining T and K. However, this experiment applies the MobileNet DCNN model as the first feature extractor that retrieves 1024 high-impact features.

3.4. Feature Selection

The operating technique of the offered feature selection method is described in this section. In this experiment, an updated feature selection algorithm named Analysis of Variance Feature Selection (ANOVA) is utilized to update the execution time and predicted results. The description of this algorithm is given below.

Analysis of Variance Feature Selection (ANOVA)

ANOVA [45] is an updated statistical feature selector that ranks characteristics by computing the variance ratios across and within categories. The ratio displays how closely the δ^{th} characteristic is related to the collective attributes. The ratio R for two working datasets is calculated by Equation (2).

$$R(y)=\frac{s_B^2(\delta)}{s_W^2(\delta)} \tag{2}$$

where $s_B^2(\delta)$ and $s_W^2(\delta)$ are the sample variances between classes and within classes. The formulas for these two sample variances are given in Equation (3) and Equation (4).

$$s_B^2(\delta)=\sum_{i=1}^{K} n_i \frac{\left(\left(\sum_{j=1}^{n_i} f_{ij}(\delta)/n_i\right)-\left(\sum_{i=1}^{K}\sum_{j=1}^{n_i} f_{ij}(\delta)/\sum_{i=1}^{K} n_i\right)\right)^2}{df_B} \tag{3}$$

$$s_W^2(\delta)=\sum_{i=1}^{K}\sum_{j=1}^{n_i} \frac{\left(f_{ij}(\delta)-\left(\sum_{i=1}^{K}\sum_{j=1}^{n_i} f_{ij}(\delta)/\sum_{i=1}^{K} n_i\right)\right)^2}{df_W} \tag{4}$$

the degrees of these two sample variances are defined as $df_B=k-1$ and $df_W=N-K$, where K indicates the number of classes and N is the total number of samples. The frequency of the δ^{th} characteristic in the j^{th} instance in the i^{th} class is indicated by $f_{ij}(\delta)$. However, the sum of all examples in the j^{th} class is indicated by n_i. The working principle of the ANOVA feature selector is given in Algorithm 2.

Algorithm 2. Feature selection mechanism using ANOVA

***Input**: Extracted feature map*

***Output**: Optimal feature map*

initialization:

1. Y = F-1, for F = No. of retrieved features of each DL algorithm.

2. L ← Data labels

3. Y_n ← Training samples

4. L_n ← Training labels

5. M_f ← Respective optimal feature map

6. G_f ← Number of generated optimal feature maps

start:

1. feature = ANOVA Feature Selector(Y_n, L_n)

2. if (Extracted features > 0.5) :
3. G_f *= features*
4. M_f *= sum (total no. of* G_f*)*
5. end if
6. display F_v
end

3.5. Classification

This part represents the final classification tasks utilizing several ML classifiers. Four ML classifiers (SVM, RF, KNN, and NB) are used to classify blood cancer types. Among them, the SVM classifier provided the best results. The working mechanism of this classifier is given below.

Support Vector Machine (SVM)

SVM [46] is an ML paradigm capable of classifying diseases by analyzing input data. It works by constructing an optimal hyperplane that separates data points into different classes within an n-order (D) space, which makes it easier to classify new instances. In this work, the classification process leverages the kernel trick (KT), a method that enables SVM to handle non-linear data. Specifically, for a two-dimensional dataset that is not linearly separable, the kernel trick maps the instances into a higher-order space where a linear partition becomes feasible. The corresponding mathematical expression is provided in Equation (5).

Kernel trick (KT): $$k\left(x_i, x_j\right) = x_i \cdot x_j \tag{5}$$

4. Results and Discussion

The suggested methodology for ALL classification was simulated on the Google Colab online platform, 64-bit Windows 11 operating system, Intel Core-i7 CPU, and 64 GB RAM. The Keras library was used to connect the DL model with the Python language.

However, several evaluation matrices like Accuracy (A), Recall (R), Precision (P), Area Under the Curve (AUC), and F1-score (F) are needed to test the performance of the proposed framework. Four parameters such as true positive (TP), false negative (FN), false positive (FP), and true negative (TN) are needed to calculate the performance matrix of the proposed work. The formulas of the evaluation matrices are given in Equations (6)-(9).

$$\text{Accuracy}\left(\text{A}\right) = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{6}$$

$$\text{Precision}\left(\text{P}\right) = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

$$\text{Recall}\left(\text{R}\right) = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

$$\text{Fl score}\left(\text{F}\right) = 2\frac{\text{PRE} \cdot \text{REC}}{\text{PRE} + \text{REC}} \tag{9}$$

4.1. Experimental Results without a Feature Selection Method

This part reflects the simulated results of the suggested work without feature selectors. We summarized the classification results of each DL model with four ML classifiers from Table 3-6. Table 3 demonstrates simulated results of the VGG16 feature extractor with different ML classifiers. From Table 3, the SVM classifier provides the best results with an accuracy of 98.00%, and the NB classifier provides the lowest performance with an accuracy of 80.58%.

Table 3. The overall performance of different classifiers with VGG16.

Feature Extractor	Classifier	Extracted feature	A (%)	P (%)	R (%)	F (%)
VGG16	KNN	512	95.07	94.93	94.66	94.75
	SVM		98.00	98.00	97.84	97.91
	RF		93.374	94.20	92.51	93.09
	NB		80.586	80.00	79.78	79.72

Table 4 demonstrates the classification outcomes of the VGG19 feature extractor with different ML classifiers. From Table 4, the SVM classifier provides the best results with an accuracy of 97.69%, and the NB classifier provides the lowest performance with an accuracy of 87.06%.

Table 4. Simulated results of four classifiers with VGG19.

Feature Extractor	Classifier	Extracted feature	A (%)	P (%)	R (%)	F (%)
VGG19	KNN	512	94.76	94.94	94.17	94.42
	SVM		97.69	97.67	97.70	97.68
	RF		92.604	93.46	91.75	92.34
	NB		87.057	87.16	86.70	86.85

Table 5 demonstrates the classification results of the MobileNet feature extractor with four ML classifiers. From Table 5, the SVM classifier provides the best results with an accuracy of 94.61%, and the NB classifier provides the lowest performance with an accuracy of 62.25%.

Table 5. Simulated results of four classifiers with MobileNet.

Feature Extractor	Classifier	Extracted feature	A (%)	P (%)	R (%)	F (%)
MobileNet	KNN	1024	92.30	92.48	92.45	92.32
	SVM		94.61	94.61	94.84	94.70
	RF		89.522	90.34	89.12	89.56
	NB		62.250	68.76	61.05	60.40

Table 6 shows the classification results of the ResNet50 model with different

ML classifiers. From Table 6, the SVM classifier provides the best results with an accuracy of 99.54%, and the NB classifier provides the lowest performance with an accuracy of 83.82%.

Table 6. The overall performance of different classifiers with ResNet50.

Feature Extractor	Classifier	Extracted feature	A (%)	P (%)	R (%)	F (%)
ResNet50	KNN	2048	98.00	97.81	98.14	97.96
	SVM		99.54	99.49	99.56	99.52
	RF		96.456	97.11	96.04	96.47
	NB		83.821	83.77	83.65	83.64

From the above comparison table, we see that the ResNet50 feature extractor with the SVM classifier obtained the best results among the feature extractors. ResNet50 with SVM classifier (ResNet50 + SVM) provided an accuracy of 99.54%, an F1-score of 99.52%, a recall of 99.56%, and a precision of 99.49%. Table 7 provides a summary of the ResNet50 feature extractor after extracting the features from the input data. The performance metrics of the ResNet50 + SVM network are tabulated in Figure 3. Figure 4 shows the evaluation chart of accuracy from different feature extractors with the SVM classifier.

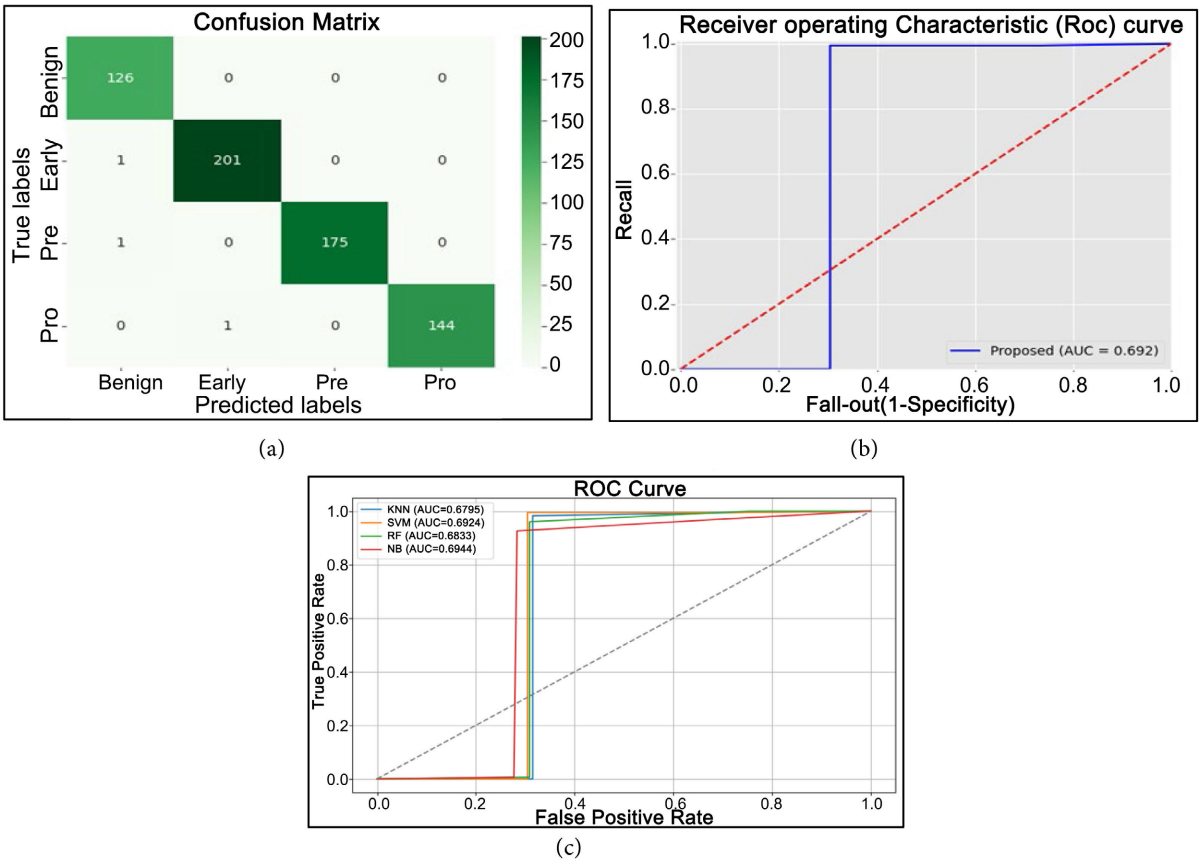


Figure 3. Performance matrix of the ResNet50 + SVM model: (a) confusion matrix, (b) AUC curve, (c) ROC-AUC curve.

The following parameters are used in each ML classifier to classify different subtypes of blood cancer. In the KNN ML classifier, the best nearest neighbor is K = 14. In SVM, we apply grid search as a cross-validation (CV) technique to find the optimal solution. The best parameters of the SVM classifier are C = 1, gamma = 0.1, and kernel = polynomial. In the RF classifier, the best parameters are max depth = 7, and number of estimators = 200. In this research, we use the Gaussian Naïve Bayes variant to find the optimal solution.

Table 7. Summary of the input layer with output shape and parameters of the ResNet50 model.

Layer	Output	Param #
Input	(None, 224, 224, 3)	0
Resnet50	(None, 7, 7, 2048)	23, 587, 712
Global Average Pooling	(None, 2048)	0

Total params: 23,587,712 (89.98 MB); **Trainable params:** 23,534,592 (89.78 MB); **Non-trainable params:** 53,120 (207.50 KB).

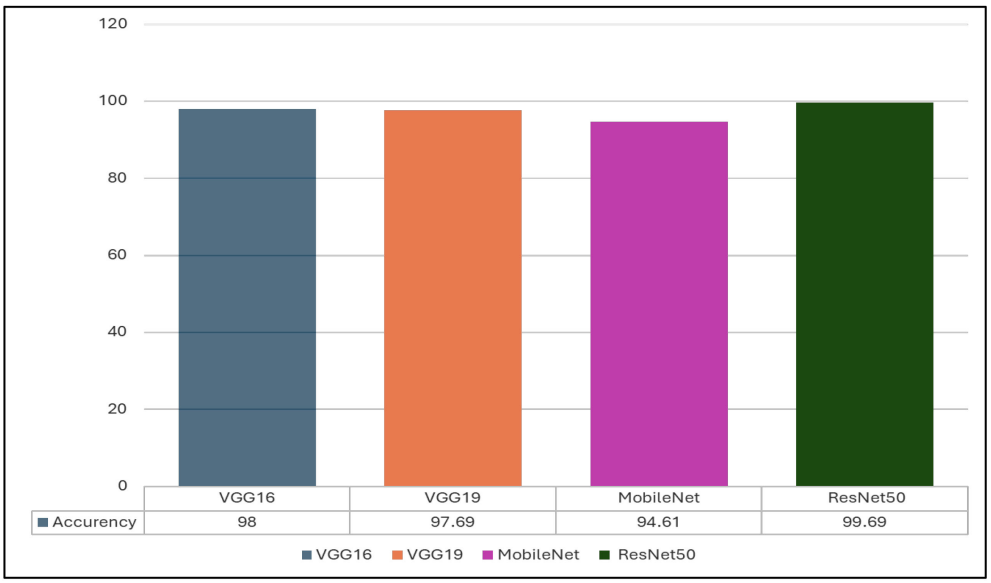


Figure 4. Comparison of the accuracy of different feature extractors with the SVM classifier.

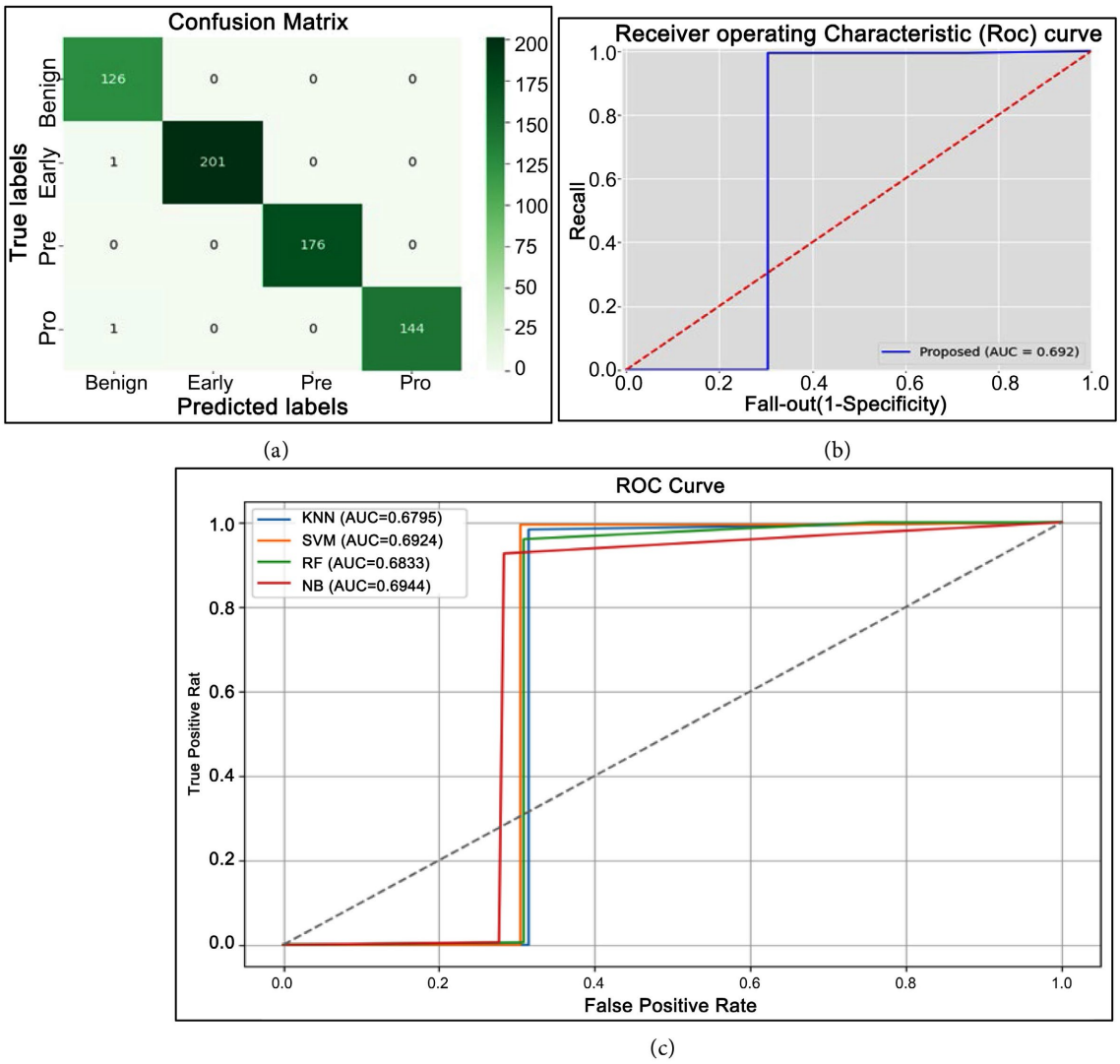
4.2. Experimental Results with the Feature Selection Method

This part presents the experimental outcomes using a feature selection approach named Analysis of Variance Feature Selection (ANOVA). This feature selection algorithm is applied to the best model (ResNet50 + SVM) to improve the classification results of this model. The ANOVA feature selector selects 1678 best features from the extracted feature set. Table 8 shows the corresponding results of the ResNet50 + SVM model after applying the ANOVA feature selector. Table 8 reflects that the feature selector method improved the previous accuracy from 99.54% to 99.69%. The evaluation matrices of the ResNet50 + ANOVA + SVM

network are tabulated in Figure 5. Table 8 reflects the corresponding comparison between the two classification methods. The ANOVA algorithm selects 1678 best features from the extracted 2048 features. In the feature selection technique, the proposed framework makes classification results based only on the best features rather than all features. That is why the ResNet50 + ANOVA + SVM model provides better results than the ResNet50 + SVM model.

Table 8. Performance comparison of pneumonia prediction using the ANOVA feature selector for the ResNet50 model.

Feature Extractor	Feature Selector	Selected feature	Classifier	A (%)	P (%)	R (%)	F (%)
ResNet50	ANOVA	1678	SVM	99.69	99.61	99.70	99.65
	-	2048		99.54	99.49	99.56	99.52



(a) (b)

(c)

Figure 5. The performance measurement ANOVA feature selection with ResNet50+SVM (a) Confusion matrix (b) AUC-ROC curve (c) ROC curve for all classifiers.

Validation Protocol

To ensure the robustness and reproducibility of the results, we employed a 10-fold stratified cross-validation protocol on the training set. Unlike standard shuffling, stratification ensures that the class distribution of ALL subtypes in each fold mirrors the original dataset, preventing bias in smaller sub-classes. For each of the K = 10 iterations, nine folds were used for feature extraction and classifier training, while the remaining fold served as the validation set. Table 9 shows the reporting stability of this work with Mean ± Variability score. Table 10 shows the comparative analysis for the SOTA approach on the ALL dataset.

Table 9. The stability reporting of this work with Mean ± Variability score.

Model Pipeline	Mean Acc (%)	Mean Pre (%)	Mean Rec (%)	Mean F1 (%)
VGG16 + SVM	98 ± 0.62	98 ± 0.55	97.84 ± 0.72	97.91 ± 0.0.47
VGG19 + SVM	97.69 ± 0.73	97.67 ± 0.81	97.70 ± 0.80	97.67 ± 0.73
MobileNet + SVM	94.61 ± 1.22	94.61 ± 1.15	94.84 ± 1.08	94.70 ± 1.10
ResNet50 + SVM	99.54 ± 0.28	99.49 ± 0.32	99.56 ± 0.24	99.52 ± 0.30
ResNet50 + ANOVA + SVM	99.69 ± 0.11	99.61 ± 0.14	99.70 ± 0.11	99.65 ± 0.12

Table 10. Comparative results for the SOTA approach on the ALL dataset. The highest outcomes are shown in bold. Here, A stands for accuracy.

Authors/Ref.	Algorithms	Methods	Optimizer	A (%)
Arbab *et al.* (2024) [22]	SVM, CNN, AlexNet	ML + DL	–	98
Rezayi *et al.* (2021) [34]	CNN, ResNet50, VGG16, RF, MLP, LR, KNN, SVM	DL + ML	Adam	84.62
Revanda *et al.* (2022) [36]	Mask RCNN	DL	SGD	83.72
Sampathila *et al.* (2022) [37]	cross-entropy loss function, CNN	DL	Adam	95.54
Ansari *et al.* (2023) [38]	Tversky loss function, CNN	DL	Binary cross-entropy, Adam	99
Safuan *et al.* (2020) [47]	AlexNet, CNN, GoogleNet, VGG	DL	–	99.13
Pallegama *et al.* (2020) [48]	CNN	DL	–	98.53
Abunadi *et al.* (2022) [49]	CNN, FFNN, SVM, ANN, GoogleNet, AlexNet, ResNet18	DL + ML	Adam	100
Rahman *et al.* (2023) [50]	VGG19, ResNet50, InceptionV3, SVM, RF, DT, NB, XGB, KNN, LR	ML + DL	PSO, CSO	99.84
Proposed	VGG16, VGG19, ResNet50, MobileNet, SVM, RF, NB, KNN	ML + DL	ANOVA	**99.69**

5. Conclusion and Future Scope

This study offers a hybrid framework that integrates deep learning for feature ex-

traction, an ANOVA feature selector, and machine learning for prediction to automate the diagnosis of Acute Lymphoblastic Leukemia. The framework effectively classifies four ALL subtypes using microscopic blood smear images. The highest accuracy is obtained by the ResNet50+SVM model, which exhibits an accuracy of 99.69% after applying ANOVA feature selection. The results demonstrate the model's high diagnostic potential and clinical relevance. Future work will focus on implementing this framework in real-time environments through mobile or IoT-based systems for broader clinical applicability.

Authors' Contributions

All authors contributed to this research's design, analysis, writing, and revision. All authors approved the submitted version of the manuscript.

Data Availability Statement

The working dataset is available on the Kaggle online database. The link to this database is: https://www.kaggle.com/datasets/mehradaria/leukemia.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

[1] Ou, F., Michiels, S., Shyr, Y., Adjei, A.A. and Oberg, A.L. (2021) Biomarker Discovery and Validation: Statistical Considerations. *Journal of Thoracic Oncology*, **16**, 537-545. https://doi.org/10.1016/j.jtho.2021.01.1616

[2] Chtihrakkannan, R., Kavitha, D.P., Mangayarkarasi, T. and Karthikeyan, R. (2019) Breast Cancer Detection Using Machine Learning. *International Journal of Innovative Technology and Exploring Engineering*, **8**, 3123-3126. https://doi.org/10.35940/ijitee.k2498.0981119

[3] Ahmad, S., Ullah, T., Ahmad, I., AL-Sharabi, A., Ullah, K., Khan, R.A., *et al.* (2022) A Novel Hybrid Deep Learning Model for Metastatic Cancer Detection. *Computational Intelligence and Neuroscience*, **2022**, 1-14. https://doi.org/10.1155/2022/8141530

[4] Chaurasia, V., Pal, S. and Tiwari, B. (2018) Prediction of Benign and Malignant Breast Cancer Using Data Mining Techniques. *Journal of Algorithms & Computational Technology*, **12**, 119-126. https://doi.org/10.1177/1748301818756225

[5] Solanki, Y.S., Chakrabarti, P., Jasinski, M., Leonowicz, Z., Bolshev, V., Vinogradov, A., *et al.* (2021) A Hybrid Supervised Machine Learning Classifier System for Breast Cancer Prognosis Using Feature Selection and Data Imbalance Handling Approaches. *Electronics*, **10**, Article 699. https://doi.org/10.3390/electronics10060699

[6] Fujita, T.C., Sousa-Pereira, N., Amarante, M.K. and Watanabe, M.A.E. (2021) Acute Lymphoid Leukemia Etiopathogenesis. *Molecular Biology Reports*, **48**, 817-822. https://doi.org/10.1007/s11033-020-06073-3

[7] Leukemia—Cancer Stat Facts. https://seer.cancer.gov/statfacts/html/leuks.html

[8] Atteia, G., Alhussan, A. and Samee, N. (2022) Bo-Allcnn: Bayesian-Based Optimized CNN for Acute Lymphoblastic Leukemia Detection in Microscopic Blood Smear Im-

ages. *Sensors*, **22**, Article 5520. https://doi.org/10.3390/s22155520

[9] Sathishkumar, K., Chaturvedi, M., Das, P. Stephen, S. and Mathur, P. (2022) Cancer Incidence Estimates for 2022 & Projection for 2025: Result from National Cancer Registry Programme, India. *Indian Journal of Medical Research*, **156**, 598-607. https://doi.org/10.4103/ijmr.ijmr_1821_22

[10] Dong, Y., Shi, O., Zeng, Q., Lu, X., Wang, W., Li, Y., *et al.* (2020) Leukemia Incidence Trends at the Global, Regional, and National Level between 1990 and 2017. *Experimental Hematology & Oncology*, **9**, 1-11. https://doi.org/10.1186/s40164-020-00170-6

[11] Cranaf, R., Kavitha, G. and Alagu, S. (2024) A Decision Support System for the Identification of Acute Lymphoblastic Leukemia in Microscopic Blood Smear Images.

[12] Mishra, S., Majhi, B. and Sa, P.K. (2019) Texture Feature Based Classification on Microscopic Blood Smear for Acute Lymphoblastic Leukemia Detection. *Biomedical Signal Processing and Control*, **47**, 303-311. https://doi.org/10.1016/j.bspc.2018.08.012

[13] Das, P.K., Jadoun, P. and Meher, S. (2020) Detection and Classification of Acute Lymphocytic Leukemia. 2020 *IEEE-HYDCON*, Hyderabad, 11-12 September 2020, 1-5. https://doi.org/10.1109/hydcon48903.2020.9242745

[14] Patel, N. and Mishra, A. (2015) Automated Leukaemia Detection Using Microscopic Images. *Procedia Computer Science*, **58**, 635-642. https://doi.org/10.1016/j.procs.2015.08.082

[15] Bennett, J.M., Catovsky, D., Daniel, M., Flandrin, G., Galton, D.A.G., Gralnick, H.R., *et al.* (1976) Proposals for the Classification of the Acute Leukaemias French-American-British (FAB) Cooperative Group. *British Journal of Haematology*, **33**, 451-458.

[16] Das, P.K., A, D.V., Meher, S., Panda, R. and Abraham, A. (2022) A Systematic Review on Recent Advancements in Deep and Machine Learning Based Detection and Classification of Acute Lymphoblastic Leukemia. *IEEE Access*, **10**, 81741-81763. https://doi.org/10.1109/access.2022.3196037

[17] Abbas, N., Mohamad, D., Abdullah, A.H., *et al.* (2015) Nuclei Segmentation of Leukocytes in Blood Smear Digital Images. *Pakistan Journal of Pharmaceutical Sciences*, **28**, 1801-1806.

[18] Abbas, N., Saba, T., Mohamad, D., Rehman, A., Almazyad, A.S. and Al-Ghamdi, J.S. (2018) Machine Aided Malaria Parasitemia Detection in Giemsa-Stained Thin Blood Smears. *Neural Computing and Applications*, **29**, 803-818. https://doi.org/10.1007/s00521-016-2474-6

[19] Ramaneswaran, S., Srinivasan, K., Vincent, P.M.D.R. and Chang, C. (2021) Hybrid Inception v3 XGBoost Model for Acute Lymphoblastic Leukemia Classification. *Computational and Mathematical Methods in Medicine*, **2021**, 1-10. https://doi.org/10.1155/2021/2577375

[20] Brownlee, J. (2016) Deep Learning with Python: Develop Deep Learning Models on Theano and TensorFlow Using Keras. Machine Learning Mastery.

[21] Rehman, A., Abbas, N., Saba, T., Rahman, S.I.U., Mehmood, Z. and Kolivand, H. (2018) Classification of Acute Lymphoblastic Leukemia Using Deep Learning. *Microscopy Research and Technique*, **81**, 1310-1317. https://doi.org/10.1002/jemt.23139

[22] Arbab, Q., Khan, M.Q. and Ali, H.J.T.S. (2022) Automatic Detection and Classification of Acute Lymphoblastic Leukemia Using Convolution Neural Network.

[23] Mughal, B., Muhammad, N., Sharif, M., Rehman, A. and Saba, T. (2018) Removal of

Pectoral Muscle Based on Topographic Map and Shape-Shifting Silhouette. *BMC Cancer*, **18**, Article No. 778. https://doi.org/10.1186/s12885-018-4638-5

[24] Norouzi, A., Rahim, M.S.M., Altameem, A., Saba, T., Rad, A.E., Rehman, A., *et al.* (2014) Medical Image Segmentation Methods, Algorithms, and Applications. *IETE Technical Review*, **31**, 199-213. https://doi.org/10.1080/02564602.2014.906861

[25] Mughal, B., Muhammad, N., Sharif, M., *et al.* (2017) Extraction of Breast Border and Removal of Pectoral Muscle in Wavelet Domain. *Biomedical Research*, **28**, 5041-5043.

[26] Mughal, B., Sharif, M. and Muhammad, N. (2017) Bi-model Processing for Early Detection of Breast Tumor in CAD System. *The European Physical Journal Plus*, **132**, 1-14. https://doi.org/10.1140/epjp/i2017-11523-8

[27] Mughal, B., Sharif, M., Muhammad, N. and Saba, T. (2018) A Novel Classification Scheme to Decline the Mortality Rate among Women Due to Breast Tumor. *Microscopy Research and Technique*, **81**, 171-180. https://doi.org/10.1002/jemt.22961

[28] Mohapatra, S., Patra, D. and Satpathy, S. (2014) An Ensemble Classifier System for Early Diagnosis of Acute Lymphoblastic Leukemia in Blood Microscopic Images. *Neural Computing and Applications*, **24**, 1887-1904. https://doi.org/10.1007/s00521-013-1438-3

[29] Rehman, A., Abbas, N., Saba, T., Mahmood, T. and Kolivand, H. (2018) Rouleaux Red Blood Cells Splitting in Microscopic Thin Blood Smear Images via Local Maxima, Circles Drawing, and Mapping with Original RBCs. *Microscopy Research and Technique*, **81**, 737-744. https://doi.org/10.1002/jemt.23030

[30] Saba, T. (2017) Halal Food Identification with Neural Assisted Enhanced RFID Antenna. *Biomedical Research*, **28**, 7760-7762.

[31] Waheed, S.R., Alkawaz, M.H., Rehman, A., Almazyad, A.S. and Saba, T. (2016) Multifocus Watermarking Approach Based on Discrete Cosine Transform. *Microscopy Research and Technique*, **79**, 431-437. https://doi.org/10.1002/jemt.22646

[32] Ahmed, N., Yigit, A., Isik, Z. and Alpkocak, A. (2019) Identification of Leukemia Subtypes from Microscopic Images Using Convolutional Neural Network. *Diagnostics*, **9**, Article 104. https://doi.org/10.3390/diagnostics9030104

[33] Jiang, Z., Dong, Z., Wang, L. and Jiang, W. (2021) Method for Diagnosis of Acute Lymphoblastic Leukemia Based on ViT-CNN Ensemble Model. *Computational Intelligence and Neuroscience*, **2021**, Article 7529893. https://doi.org/10.1155/2021/7529893

[34] Rezayi, S., Mohammadzadeh, N., Bouraghi, H., Saeedi, S. and Mohammadpour, A. (2021) Timely Diagnosis of Acute Lymphoblastic Leukemia Using Artificial Intelligence-Oriented Deep Learning Methods. *Computational Intelligence and Neuroscience*, **2021**, Article 5478157. https://doi.org/10.1155/2021/5478157

[35] Zakir Ullah, M., Zheng, Y., Song, J., Aslam, S., Xu, C., Kiazolu, G.D., *et al.* (2021) An Attention-Based Convolutional Neural Network for Acute Lymphoblastic Leukemia Classification. *Applied Sciences*, **11**, Article 10662. https://doi.org/10.3390/app112210662

[36] Revanda, A.R., Fatichah, C., Suciati, N., *et al.* (2022) Classification of Acute Lymphoblastic Leukemia on White Blood Cell Microscopy Images Based on Instance Segmentation Using Mask R-CNN. *International Journal of Intelligent Engineering*, **15**, 625-637.

[37] Sampathila, N., Chadaga, K., Goswami, N., Chadaga, R.P., Pandya, M., Prabhu, S., *et al.* (2022) Customized Deep Learning Classifier for Detection of Acute Lymphoblastic Leukemia Using Blood Smear Images. *Healthcare*, **10**, Article 1812. https://doi.org/10.3390/healthcare10101812

[38] Ansari, S., Navin, A., Sangar, A., Gharamaleki, J. and Danishvar, S. (2023) A Customized Efficient Deep Learning Model for the Diagnosis of Acute Leukemia Cells Based on Lymphocyte and Monocyte Images. *Electronics*, **12**, Article 322. https://doi.org/10.3390/electronics12020322

[39] Abd El-Ghany, S., Elmogy, M. and El-Aziz, A.A.A. (2023) Computer-Aided Diagnosis System for Blood Diseases Using Efficientnet-B3 Based on a Dynamic Learning Algorithm. *Diagnostics*, **13**, Article 404. https://doi.org/10.3390/diagnostics13030404

[40] Osman M. H., Mohamed R.H., Sarhan H.M., Park E.J., Baik S.H., Lee K.Y. and Kang J. (2022) Machine Learning Model for Predicting Postoperative Survival of Patients with Colorectal Cancer. *Cancer Research and Treatment*, **54**, 517-524. https://doi.org/10.4143/crt.2021.206

[41] Aria, M. (2021) Acute Lymphoblastic Leukemia (ALL) Image Dataset. https://www.kaggle.com/datasets/mehradaria/leukemia

[42] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition.

[43] Wu, Z., Shen, C. and van den Hengel, A. (2019) Wider or Deeper: Revisiting the Resnet Model for Visual Recognition. *Pattern Recognition*, **90**, 119-133. https://doi.org/10.1016/j.patcog.2019.01.006

[44] Kumar Shukla, R. and Kumar Tiwari, A. (2023) Masked Face Recognition Using MobileNet V2 with Transfer Learning. *Computer Systems Science and Engineering*, **45**, 293-309. https://doi.org/10.32604/csse.2023.027986

[45] Nasiri, H. and Alavi, S.A. (2022) A Novel Framework Based on Deep Learning and ANOVA Feature Selection Method for Diagnosis of COVID-19 Cases from Chest X-Ray Images. *Computational Intelligence and Neuroscience*, **2022**, 1-11. https://doi.org/10.1155/2022/4694567

[46] Jakkula, V. (2006) Tutorial on Support Vector Machine (SVM). School of EECS, Washington State University.

[47] Safuan, S.N.M., Tomari, N.R.M., Zakaria, W.N.W., *et al.* (2020) Investigation of White Blood Cell Biomaker Model for Acute Lymphoblastic Leukemia Detection Based on Convolutional Neural Network. *Bulletin of Electrical Engineering and Informatics*, **9**, 611-618. https://doi.org/10.11591/eei.v9i2.1857

[48] Pallegama, R., Madhusanka, A., Priyankara, H.D.N.S., *et al.* (2020) Acute Lymphoblastic Leukemia Detection Using Convolutional Neural Network.

[49] Abunadi, I. and Senan, E.M. (2022) Multi-Method Diagnosis of Blood Microscopic Sample for Early Detection of Acute Lymphoblastic Leukemia Based on Deep Learning and Hybrid Techniques. *Sensors*, **22**, Article 1629. https://doi.org/10.3390/s22041629

[50] Rahman, W., Faruque, M.G.G., Roksana, K., Sadi, A.H.M.S., Rahman, M.M. and Azad, M.M. (2023) Multiclass Blood Cancer Classification Using Deep CNN with Optimized Features. *Array*, **18**, Article 100292. https://doi.org/10.1016/j.array.2023.100292