Yugoslav Journal of Operations Research
35 (2025), Number 4, 947-964
DOI: https://doi.org/10.2298/YJOR240315006N

Research Article

INTEGRATING MULTICLASS CLASSIFIERS FOR ENHANCED ACUTE LYMPHOBLAST LEUKEMIA DETECTION: A COMPARATIVE STUDY

Roopashree NAYAK*
Sahyadri College of Engineering & Management, Mangaluru
Affiliated to Visvesvaraya Technological University, Belagavi
roopashree.ec@sahyadri.edu.in, ORCID: 0000-0002-1653-3392

Anush BEKAL
Sahyadri College of Engineering & Management, Mangaluru
Affiliated to Visvesvaraya Technological University, Belagavi
ORCID: 0009-0003-4066-6720

Jennifer C SALDANHA
St. Joseph Engineering College, Mangaluru
Affiliated to Visvesvaraya Technological University, Belagavi
ORCID: 0000-0002-6048-3230

Received: March 2024 / Accepted: November 2024

Abstract: Acute lymphoblastic leukemia (ALL) is a blood and bone marrow malignancy that is characterized by the growth of many immature lymphocytes known as lymphoblasts. It primarily affects children, particularly those aged two to five years, and is the primary cause of death in pediatric cancer cases. The method of treatment is determined to ALL, the individual's age at the time of diagnosis, and other pertinent considerations. Regardless, early detection and diagnosis are critical for a good prognosis. It is critical to precisely detect malignant cells to make a diagnosis and assess the extent of the disease. However, due to physical similarities, identifying lymphoblasts from normal white blood cells under a microscope is often difficult. Using computer-aided techniques can be extremely valuable in automating the identification of cancerous cells, allowing histopathologists and oncologists to make decisions about the early stage. This paper demonstrates the usefulness of extensive image pre-processing, feature extraction from ResNet50 and VGG19 CNN models, and robust feature selection in an automated diagnostic technique for Acute Lymphoblastic Leukemia. Notably, on the C-NMC 2019 Dataset, ResNet50 with Random

*Corresponding author

Forest feature selection appears as the best combination. The ResNet50 model achieves maximal precision, Weighted F1 Score, F1-score accuracy, and recall of 84.18%, 80.4%, 86.15%, 80.83%, and 88.7% respectively when combined with ANOVA and Random Forest. The combination of VGG19+Random Forest+SVM achieves a maximum accuracy of 86.2%. These findings highlight its exceptional performance in recognizing and categorizing target labels, demonstrating its ability to extract relevant properties for improved leukemia identification.

Keywords: Lymphoblast, ANOVA, transfer learning, ResNet50, multi-class classification.

MSC: 62H35, 68T10, 62P10, 62C55.

1. INTRODUCTION

Leukemia is a kind of cancer that affects the bone marrow and blood, producing uncontrolled and abnormal multiplication of multiple blood cell types such as platelets (thrombocytes), white blood cells (leukocytes), and red blood cells (erythrocytes) [1]. This overproduction of immature cells can cause healthy cells to be displaced in both the bone marrow and the circulation. Chronic lymphocytic leukemia, acute lymphoblastic leukemia, chronic myeloid leukemia, and acute myeloid leukemia are the four main forms of leukemia [2].

The main category of grouping is how quickly leukemia advances. The aberrant blood cells in acute leukemia are unable to carry out their regular functions and grow rapidly, leading the sickness to become worse at a faster rate. It requires ideal therapy. Leukemia includes more immature blood cells. These blood cells multiply or accumulate at a slower rate and can function normally for an extended period. A few types of ongoing leukemia at first produce no early side effects and can go unobserved or undiscovered for quite a long time.

The second sort of grouping is based on white blood cell affection. If leukemia influences the lymphoid cells, then it's called Lymphocytic leukemia. If leukemia influences the myeloid cells, then it's known as Myelogenous leukemia. Myeloid cells are responsible for the formation of red blood cells, white blood cells, and platelet-creating cells.

A person can develop any of the following four forms of leukemia at any point in their lifetime:

- White blood cell counts are elevated by Chronic Lymphocytic Leukemia (CLL), which affects both blood cells and the bone marrow. CLL, which is typified by a rise in leukemic cells that resemble B lymphocytes and clonal proliferation, is more common in adults than in children.
- White blood cell proliferation is a feature of ALL, or acute lymphoblastic leukemia, which mainly affects youngsters.
- Chronic Myeloid Leukemia (CML) affects the ability of normal cells to efficiently fight off infections. It is caused by genetic alterations in myeloid cells. Chronic myeloid leukemia (CML) is a common adult cancer that progresses more slowly and has acceleration, burst, and chronic phases.
- Acute Myeloid Leukemia (AML) is characterized by a large increase in white blood cells (WBCs), which may be brought on by insufficient bone marrow growth. Early detection is crucial for treatment to be effective. Possible symptoms include bleeding, trouble breathing, and other signs.

Acute Lymphoblastic Leukemia is commonly seen in children. It's seen in the bone marrow. Later, it will spread to lymphoid nodes, the liver & spleen of the body. Acute myeloid leukemia (AML) is one of the most common leukemias in adults. As the abnormal Myeloblasts cells multiply, they overcome the normal cells in the bone marrow and blood. Chronic Lymphocytic Leukemia (CLL) is commonly seen in adults & the affected person can live like a normal person without treatment. Chronic Myelogenous Leukemia (CML) affects mainly adults & the affected person may have a few symptoms as it enters the stage where the affected cells increase rapidly.

Acute Lymphoblastic Leukemia, which is more commonly linked with childhood & adolescence, accounts for a significant percentage of all children's cancer cases. It is thought to be responsible for about 25 % of all kid's cancer cases and up to 75 % of pediatric leukemia cases [3]. ALL is most typically diagnosed in children between 2 to 5 years old [4]. While ALL is uncommon in adults, becoming older is a risk factor. Adult ALL is usually diagnosed beyond the age of 50. It is critical to understand that the exact sub-type of ALL has a significant impact on survival rates. Figure 1 depicts the National Cancer Institute's statistics for acute lymphoblastic leukemia.

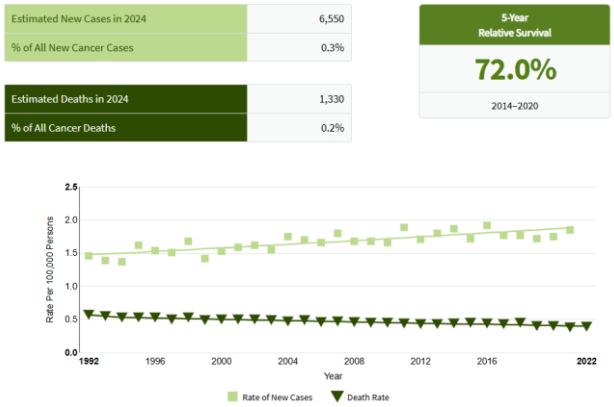

Figure 1: Statistics of ALL as per NCI [5]

2. RELATED WORK

Leukemia is a broad term for a group of blood malignancies that affect the bone marrow and blood cells. Due to the complexity and notable variety of this disorder, accurate diagnosis, classification, and therapy are required.

2.1. Image Processing Based Approach

Significantly more progress has been made with image processing in many other domains, including the healthcare industry. To obtain an improved or high-quality image, image processing involves applying several simple procedures to an image. This technique transforms an image into digital components. There are numerous commercial sectors in which this technology can be applied. Basic procedures are using an optical scanner to input a picture, managing and analyzing pictures (including image augmentation, compression methods, and pattern recognition in images like satellite photos), and analyzing and managing pictures. It is sometimes believed that picture

processing involves unnecessarily altering the image to achieve a certain level of beauty or to justify a widely accepted reality.

A substantial number of abnormal cells are produced when a person has leukemia in the bone marrow. CLL, ALL, and AML are the three most frequent forms of leukemia. A selection of commonly used segmentation of image algorithms is provided here. These three algorithms are HSV color-based categorization, x-means clustering, and marker-controlled watershed. To categorize leukemia into its categories and subtypes, an SVM classifier was employed in this instance. Classifying leukemic cells into CML, AML, CLL, and ALL is the claim made in this thesis. The limits are selected automatically when using the Otsus technique for automated picture segmentation. Three technologies for enhancing contrast in colored images utilizing RGB components are suggested in [6]: partial contrast, dark stretching, and bright stretching. Grayscale images are less dependable for image segmentation than colored images. Color models are essentially divided into two categories: HSI and RGB. Here is a straightforward conversion from RGB to HSV (Hue, Saturation, and Value). This section explains the two sturdy and highly effective texture descriptions, GLCM and LBP. The process flow will start with the image acquisition phase, in which a device with the appropriate pixel values and bounds is used to identify the image. Preprocessing, which involves doing some basic operations on the image and analyzing the data obtained during image acquisition, will be the following step. Subsequently, segmentation and feature extraction constitute the primary stages of image processing, during which necessary algorithmic adjustments are carried out. The preprocessing stage increases or improves the image quality; the images that are left over are obtained in the form of Cyan, Magenta, Yellow, and Key. Partitioning the image into manageable chunks makes it easier to interpret and more meaningful.

Within the K means clustering technique, the image data is arranged exclusively to each region, meaning that a given piece of data cannot be assigned to another. To use the K-means clustering approach, several steps must be completed. For example, RGB images from the database are used as input pictures; because color space is difficult to discern visually, RGB images are transformed to grayscale. In this case, you can convert a grayscale image to the product of the letter's "L" for luminosity, "a" for chromaticity, and "b" for the axis along which the color falls on the blue-yellow axis.

The study [7] focuses on diagnosing leukemia and provides a more thorough classification of leukemia into its four primary forms. Three different segmentation strategies were applied. To increase the precision and accuracy of the identification procedure, a vast amount of attributes were gathered. By discovering leukemia subtypes, such as AML M3, a variation of AML, this research may be furthered. To improve upon previous iterations, other categorization techniques might be researched.

2.2. Machine Learning based Approach

Leukaemia classification methods and several preprocessing techniques are covered in the paper [8]. In addition to results and conclusions, the study offers a succinct description of a contemporary classification technique. In summary, we may be able to classify leukemia sickness using a variety of modern machine-learning techniques. Deep learning models perform better for categorization when we have a huge collection of photographs. To distinguish between cells with and without leukemia, the SVM classifier was employed. Identify the hyper-plane dividing the two groups to classify anything.

SVMs use a non-linear classification method called the kernel technique. Testing and training data sets are created using all of the data.

Leukemia detection is the main focus of the research, which also offers a more thorough taxonomy of the disease through its four main forms. There were three segmentation techniques applied. The retrieval of several characteristics improved the precision and dependability of the detection process. Finding unique subtypes of leukemia, such as AML M3 [9], which is a subtype of AML, can support this research. To improve on the outcomes of the previous categorization techniques, it is possible to investigate other segmentation strategies. Staining and morphology are used in the study by [10] to identify particular blood cells. These are the four parts of classifier evaluation, SVM classification building, and feature extraction. Pre-processing, feature extraction, categorization model creation, and classifier evaluation are the general approaches for recognizing ALL cells. Certain subjects are covered in the following subsections under feature extraction. 2) Statistical feature, 1) color-based characteristic 3) A geometrical trait Images moments and the Haralick texture feature [10] are the other two. They generated features from image data by utilizing SVM and GBDT to build a platform-independent classifier. Apart from offering state-of-the-art results for identifying malignant cells, their suggested image classifier uses GBDT. the work necessary to improve specificity, increase the effectiveness of the classification, and assess these methods on a variety of datasets. The COVID-19 pandemic control-achieving methods using AI, and IOT are explained by Alrashdi et al. [11]. Khalil et al. [12] used different ML models for the prediction of kidney diseases.

2.3. Deep Learning Based Approach

Image classification, pattern recognition, and object tracking are a few examples of vision tasks where deep learning techniques have shown exceptional performance. DL technology is based on the principle that a general-purpose learning methodology is used to teach the feature extraction method from data, instead of having human engineers build it. Deep learning applications of convolutional neural networks (CNNs) have demonstrated impressive performance in image processing. Creating an end-to-end model for using a CNN-based workflow is rather easy. Additionally, while maintaining the necessary speed, CNN-based deep learning architecture enables the avoidance of intricate hand-crafted feature design. Consequently, CNN-based methods for identifying leukocytes have been created.

The research [13] regarded leukocyte identification as an object detection issue, and two remarkable object detection methods—the Single Shot Multibox Detector (SSD) and the Incremental Improvement Version of You Only Look Once (YOLO)—were used. A train set of 14,700 annotated photographs is used to develop detection models, and certain critical characteristics influencing different item identification approaches are studied to improve recognition performance. With an inference time of 53 milliseconds per image on the NVIDIA GTX1080Ti GPU, the best mean average precision of 93.10 percent and mean accuracy of 90.09 percent are attained [13].

Based on a straightforward threshold mechanism, the study [14] offers a basic segmentation methodology that produces good results. In order to classify ALL into subgroups and normal circumstances, this work offers a DL technique based on the CNN and Alexnet models. Data is used to build the model, with the final three layers being

fully integrated. classifier and softmax are created with new data and are well-tuned to the more recent classifier levels. The term "transfer learning" describes this.

Anil Kumar et al. [15] proposed the method for the categorization of b cell & t Cell ALL using the deep learning technique AlexNet and customized LeukNet model for the classification based on WHO standards. The accuracy of 94.12% is obtained with the smaller number of the dataset. Faruqui et al. [16] proposed the model named HAASNet for the diagnosis and classification of lung cancer. The designed model produced an accuracy of 96.07% but the limitation of this is unable to classify the stages of lung cancer. The IOMT concepts were used for the research so that it will be easy to make it portable on mobile devices.

Zhou et al. [17] used the CNN model for the classification of White Blood Cells and obtained an accuracy of 82.93%. in the proposed model, the classification of leukemia is performed based on the presence of the WBC. Akter et al. [18] used the CNN-based architecture to classify the four classes of brain tumours. U-net architecture is used for the segmentation of tumour images. The method gave an accuracy of 98.7%. The stages of the acute lymphocytic leukemia classification are achieved using the hybrid model and have achieved the best accuracy compared to the other studies, where the combination of MobileNet & SVM techniques were used for the categorization. [19].

3. MOTIVATION AND CONTRIBUTION

The use of multiclass classifiers in the detection of acute lymphoblastic leukemia (ALL) is inspired by a holistic approach to addressing the inherent complexities and limitations of leukemia diagnosis. Acute lymphoblast leukemia has a wide variety of genetic and clinical characteristics, needing a thorough treatment plan that can accommodate this variation. We want to exploit the complementary nature of their feature capture by merging multiple classifiers, resulting in a more nuanced and robust representation of the underlying data. This method enhances the detection system's reliability and makes it easier to deal with imbalanced data, a prevalent problem in medical datasets. The integrated system is designed to be more resilient to overfitting and capable of dynamic adaptation to evolving datasets by using the benefits of ensemble learning. Importantly, the comparative study contributes to ethical considerations by ensuring the system's reliability and minimizing the risk of false positives or false negatives that could have significant implications for patient care. The purpose of this effort is essentially to advance the clinical relevance and impact of leukemia detection systems, with the ultimate goal of improving patient outcomes and healthcare practices.

The approach of the present work is outlined as follows:

- First, the blood smear images were taken from the C-NMC2019 database.
- Then, the pre-processing is applied to find the region of intrest.
- The pre-trained deep learning models such as VGG19 & ResNet50 were used to get the features.
- Scale the extracted features using standardization or normalization techniques to ensure uniformity.
- The system efficiency is improved based on the feature selection methods ANOVA & Random Forest.
- To assess the performance of the model, divide the dataset into training and testing sets.

- The classification of the ALL is performed using machine learning (ML)algorithms such as Naive Bayes, SVM, and KNN.
- Finally, the proposed method excels over current state-of-the-art models by achieving impressive performance in accuracy.

4. PROPOSED METHODOLOGY

The outline of the proposed method is shown in Figure 2.

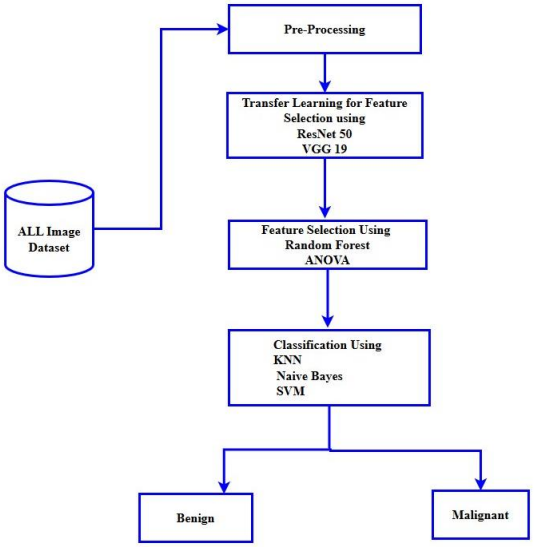

Figure 2: Transfer Learning-Feature Selection-Classification Framework

4.1. Dataset and its Prepossessing

The dataset utilized in the challenge, termed the C-NMC 2019 Dataset [20], was created by the AIIMS Laboratory of Oncology in New Delhi, India. It comprises 15114 images segmented from microscopic images of white blood cells with labels (normal versus malignant). The expert oncologist marked the ground reality. The individual microscopic images are 450 × 450 three-channel images. Stain color normalization has been utilized to mitigate staining noise and illumination flaws, thereby accurately representing real-world visuals. [21], [22], [23]. Images were taken from a dataset that included 49 people in good health and 69 people with cancer and was already split into training sets, validation sets, and test sets, as shown in Table 1.

In our case, however, due to the limitations of Google Colaboratory on the use of the GPU, it was not possible to exploit the entire dataset because the use of the CPU alone would have led to too long training times. To solve this problem, we decided to use only the training set as the entire dataset and split it in turn into a validation set, training set, and test set. A distribution of 60% went into the training set, 20% went into the validation set, and 20% went into the test set. The images to be included in each of the sets were chosen randomly to maintain the data distribution of the original dataset. The results are shown in Table 2.

Table 1: Dataset split for cancer and healthy subjects

	Cancer Cells	Normal Cells	Total
Training Set	7272	3389	**10661**
Validation Set	1219	648	**1867**
Test Set	Unknown	Unknown	**2586**
Total	Unknown	Unknown	**15114**

Table 2: Distribution of the original dataset

	Cancer Cells	Normal Cells	Total
Training Set	4432	1964	**6396**
Validation Set	1403	729	**2132**
Test Set	1437	696	**2133**
Total	7272	3389	**10661**

4.2. Pre-processing

This is an essential technique employed to improve image quality and enhance visualization, particularly in the context of medical imaging. In the field of medical imaging, image processing plays a vital role in enhancing image quality. It serves as a fundamental step that can significantly impact the success and accuracy of subsequent phases in the proposed methodology. Medical images often exhibit various issues that can adversely affect their clarity and visibility. If these images are of subpar quality, it can result in unsatisfactory outcomes and reduced accuracy in subsequent stages of analysis. The flow of the ROI extraction is shown in Figure 3.

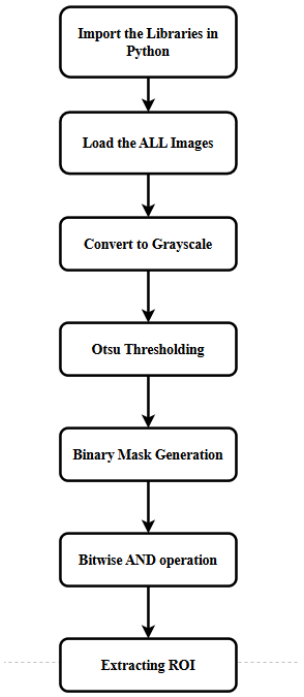

Figure 3: Flow of pre-processing for the extraction of ROI

4.2.1. Sample Image

This manuscript makes use of the C-NMC 2019 Dataset, sample images are shown in Figure 4.

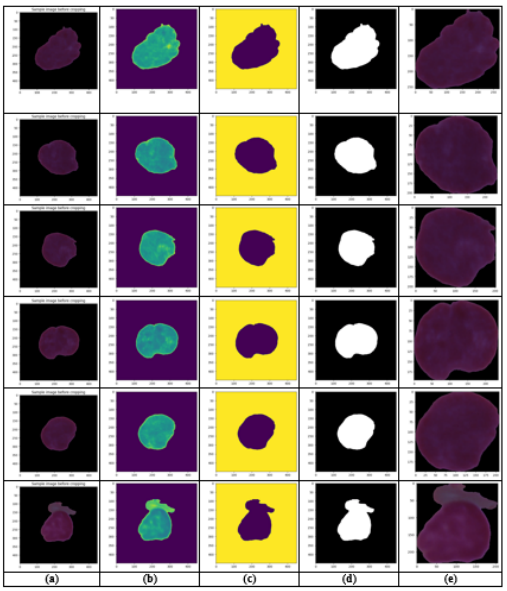

Figure 4: Pre-processing: (a) Sample image from the database (b) BGR to greyscale. (c) Applying Otsu's thresholding for binary mask generation (d) Masked image with bitwise AND operation. (e) Extracting and displaying region of interest

4.2.2. Converting the Loaded Image to Grayscale

In this step, the initially loaded image, which is in the BGR color format, is transformed into a grayscale image as shown in Figure 4. Grayscale conversion is a common image processing operation that represents each pixel's intensity as a single value, rather than as a combination of color channels. This simplifies further image analysis and is often used to extract important features or information from the image while discarding color information. The result is a grayscale image where each pixel's value represents its brightness or intensity, making it easier to work with in various image processing tasks.

4.2.3. Applying Otsu's Thresholding for Binary Mask Generation

In this step, Otsu's thresholding method is utilized to perform thresholding on the grayscale image, resulting in the creation of a binary mask as shown in Figure 4

4.2.4. Creating a Masked Image with Bitwise AND Operation

In this step, a masked image is generated. The image serves as the source, and the binary mask is applied through a bitwise AND operation. The resulting image retains regions where both the image and binary mask generated from the previous step have non-zero values, effectively isolating the object or region of interest while masking the background as shown in Figure 4.

4.2.5. Extracting and Displaying Region of Interest

In this section it calculates the minimum and maximum x and y coordinates from these non-zero pixels, defining a bounding box around the region of interest. Subsequently, a cropped image is extracted from the original image based on these calculated boundaries. Finally, it resizes the cropped image to a fixed size of 224 × 224 pixels as shown in Figure 4.

4.3. Feature Extraction using Transfer Learning

4.3.1. ResNet50

ResNet50 is a DCNN architecture from the Residual Networks family that was introduced by [24]. The ResNet50 model has fifty layers, the bulk of which are forty-eight convolutional layers. A max pool layer and an average pool layer make up the remaining two layers. This architecture is based on stacked residual blocks, a significant advancement in deep learning. ResNet50, being a pre-trained deep learning model, is extremely adaptable and may be used in a variety of computer vision projects. ResNet50 excels at feature extraction, in particular, using pre-trained weights to extract high-level features from input images.

4.3.2. VGG19

As described in the work by Simonyan et al. [25], VGG19 is a member of the Visual Geometry Group (VGG) family of Deep Convolutional Neural Networks (DCNN) introduced by the University of Oxford's Visual Geometry Group. Expanding upon the original VGG16 design, this architecture consists of 19 layers, including 16 convolutional layers and 3 fully connected layers.

4.4. Feature Scaling of Data

Feature scaling is a crucial pre-processing step in many ML algorithms, ensuring that all features have the same scale and fall within a specified range (typically between 0 and 1) to prevent any one feature from dominating the others. Each feature's minimum and maximum values in x will be used for scaling. The scaling of the dataset is computed based on the minimum and maximum values. This scaled dataset is frequently used in machine learning models to improve model performance by ensuring that all attributes are scaled uniformly.

4.5. Feature Selection

4.5.1. ANOVA

According to studies [26], [27], ANOVA feature selection is a machine-learning technique used to discover the most important qualities within a dataset. Based on Analysis of Variance (ANOVA), this method focuses on identifying variables that play a major influence in predicting the target variable. The F-value is generated for each feature in the ANOVA technique, and it serves as a measure of how well the variation of the target variable can be explained by the variance of that particular feature. Higher F-values are considered more important, showing a stronger association with the target

variable. The ANOVA F-value is calculated by examining how the variance of a particular characteristic varies across multiple values of the target variable.

As an early stage in this method, the F-statistic and related p-value for each feature are computed. These attributes are then ordered in order of importance based on their respective F-statistic values. Higher F-statistic values indicate more significant differences in mean values between groups. Setting a cutoff value for either the F-statistic or the p-value determines the selection of top characteristics. This comprises selecting features that exceed a preset threshold, ensuring that only those with the most significant statistical connections are included in the analysis.

The ResNet50 encoding network extracts 2048 features. ANOVA F-values are produced to determine the importance of these features concerning the target variable. The top 500 features with the greatest F-values, showing significant variation across distinct target variable groups, are found and chosen from among the 2048 features. Following that, these 500 high F-value features are fed into several classifiers for further classification tasks. The most significant features are chosen from a pool of 2048 features using two methods: ANOVA and the Random Forest selection method. The top 500 features are identified by ANOVA, whereas the Random Forest technique selects 632 features. The performance of these two sets of features is evaluated and SVM, Nave Bayes, and KNN classifiers are used for classification.

4.5.2. Random Forest

In this method, the feature selection is performed with feature embedding. It initializes a feature selector, which employs a Random Forest Classifier with 200 estimators and a specified random state to determine feature importance. Features are selected based on a threshold set to 1.25 times the median of feature importance. The function fits the selection of the input data and corresponding labels, identifies the selected features, and creates a new Data Frame containing only these chosen features. In essence, this code automates the process of feature selection by utilizing the importance scores provided by the Random Forest Classifier, helping to reduce dimensionality and improve model efficiency for machine learning tasks.

5. RESULTS AND DISCUSSION

This section delves into a thorough explanation and comparison of the outcomes generated from the strategy proposed in this study.

5.1. Feature Extraction Results Using ResNet50

In this, we discussed the proposed methodology where the feature was extracted from the dataset using ResNet50. After these extracted features were given to the feature selection using ANOVA and random Forest methods, these selected features were given to the KNN, Naive Bayes, and SVM Algorithms for classification. ResNet50, a deep convolutional neural network, is commonly employed in a transfer learning context for various image-related tasks such as leukemia identification. In this, 2048 features were extracted from the dataset. From this ANOVA model is used to select 500 features for the classification task. Table 3 provides statistics data for the ResNet50 feature extractor when used with different classifiers and feature selection methods. Figure 5 illustrates the contrasting classifier confusion matrices obtained using the ANOVA feature selector.

This instance involved running a dataset using ResNet50, which produced the extraction of 2048 features. Following that, about 632 of these features were chosen for classification using a Random Forest model. The Random Forest feature selection yielded contrasted classifier confusion matrices displayed in Figure 6. Table 4 summarizes the ResNet50 feature extractor's general statistics across all feature selection techniques.

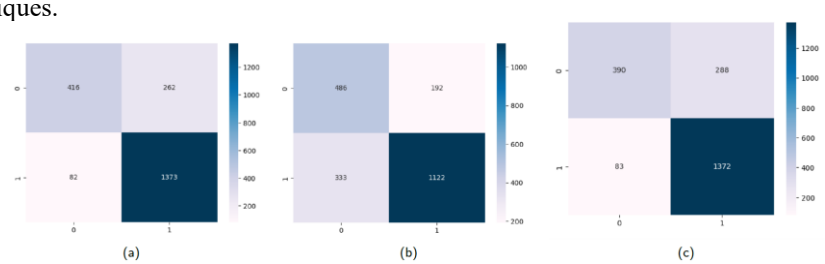

Figure 5: Comparison of classifier confusion matrices with ANOVA selector: (a) SVM classifier; (b) Naïve Bayes; (c) KNN classifier using ResNet50

Table 3: Comparative statistics of ResNet-50 feature extractor across various classifiers with diverse feature selection methods

Feature Selector	Classifier	Accuracy	Precision	Recall	F1-Score	Weighted F1-Score
ANOVA	SVM	0.839	0.840	0.944	0.899	0.831
	Naïve Bayes	0.754	0.854	0.771	0.810	0.759
	KNN	0.826	0.827	0.943	0.881	0.816
Random Forest	SVM	0.835	0.840	0.935	0.885	0.827
	Naïve Bayes	0.766	0.862	0.781	0.820	0.771
	KNN	0.830	0.828	0.948	0.884	0.820

Table 4: Overall statistical parameter of ResNet50 feature extractor of All feature selector

Statistical Parameter	Results%
Accuracy	80.83
Precision	84.18
Recall	88.7
F1-Score	86.15
Weighted F1-Score	80.4

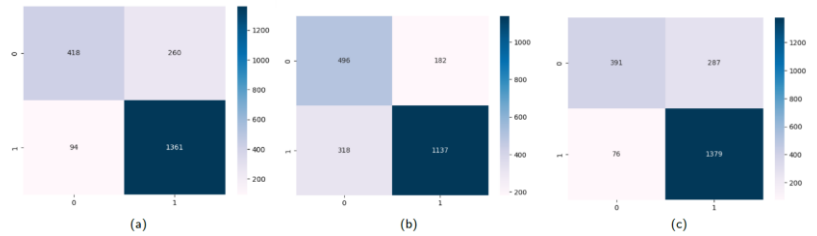

Figure 6: Comparison of Classifier Confusion Matrices with Random Forest Feature Selector: (a) SVM classifier; (b) Naïve Bayes; (c) KNN classifier using ResNet50

5.2. Feature Extraction Results Using VGG19

In this section, we will look at the results of feature extraction using the VGG19. VGG19 was used to extract 512 features from the dataset in this context. Following the extraction of these features, ANOVA was used to choose roughly 500 features for the classification process. Table 5 displays statistical data for the VGG19 feature extractor when applied to several classifiers using different feature selection strategies. The contrasted classifier confusion matrices created using the ANOVA feature selector are shown in Figure 7. The contrasted classifier confusion matrices obtained using the Random Forest feature selection are shown in Figure 8 and Table 6 summarize the VGG19 feature extractor's general statistics across all feature selection methods.

Table 5: Statistics of the VGG19 feature extractor across various classifiers and feature selection methods

Feature Selector	Classifier	Accuracy	Precision	Recall	F1-Score	Weighted F1-Score
ANOVA	SVM	0.843	0.859	0.921	0.889	0.839
	Naïve Bayes	0.649	0.859	0.581	0.693	0.661
	KNN	0.792	0.837	0.863	0.850	0.790
Random Forest	SVM	0.862	0.858	0.955	0.904	0.856
	Naïve Bayes	0.694	0.859	0.660	0.747	0.705
	KNN	0.798	0.831	0.885	0.857	0.794

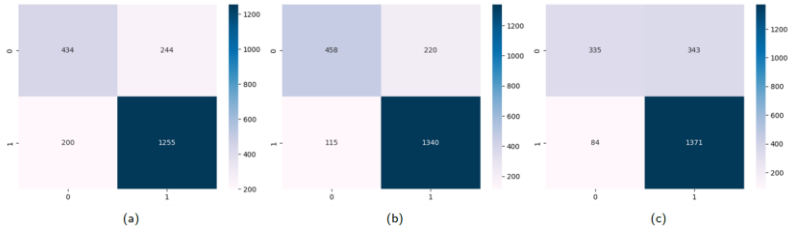

Figure 7: Comparison of classifier confusion matrices with ANOVA feature selector: (a) KNN classifier; (b) SVM; (c) Naïve Bayes classifier using VGG19

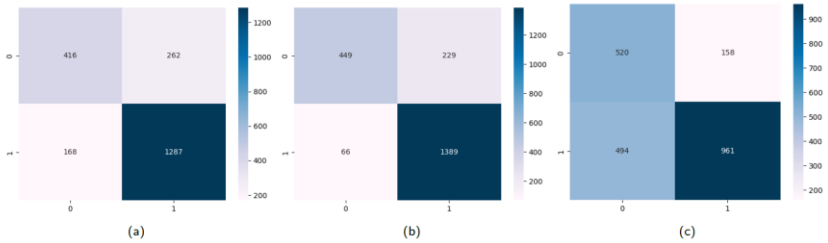

Figure 8: Comparison of classifier confusion matrices with ANOVA feature selector: (a) KNN classifier; (b) SVM; (c) Naïve Bayes classifier using VGG19

Table 6: Overall statistical parameter of VGG19 feature extractor of All feature selector

Statistical Parameter	Results%
Accuracy	77.3
Precision	85.05
Recall	81.08
F1-Score	82.33
Weighted F1-Score	77.42

5.3. Statistical Parameters

The Statistical Parameters estimated include precision, Weighted F1-Score, F1 score accuracy, and recall:

5.3.1. Accuracy

It is defined as the ratio of correctly predicted instances (true positives and true negatives) to the total number of instances in the dataset, as delineated in Eq. 1.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{1}$$

5.3.2. Precision

It is expressed as the ratio of true positive predictions to the overall number of positive predictions, demonstrating the model's accuracy in recognizing real positives, as stated in Eq. 2.

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

5.3.3. Recall

It is defined as the proportion of correct positive predictions to the total number of correct positive occurrences, and it serves as a measure of the model's ability to reduce false negatives, as explained in Eq. 3

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

5.3.4. F1 Score

The combination of recall and precision will form the performance parameter measured as the F1 score, which is given in Eq. 4

$$F1 - Score = \frac{2\times Precision\times Recall}{Precision+Recall} \tag{4}$$

5.3.5. Weighted F1-Score

It is a weighted average of the F1 scores for each class, with the weights based on the quantity of samples in each class, as described in Eq.

$$Weighted\ F1 - Score = \sum_i \left(\frac{F1-Score_i \times Number\ of\ Instances_i}{Total\ Number\ of\ Instances}\right) \tag{5}$$

6. DISCUSSION AND COMPARISON

In the following part, we'll look at the outcomes of the approach we have developed. The image processing methods discussed here are critical for preparing and retrieving useful information from the C-NMC 2019 Dataset. The dataset begins with a collection of sample photos. To facilitate subsequent analysis, the photos are first transformed into grayscale. Otsu's thresholding is then used to generate a binary mask, allowing objects to be separated from their surroundings. A masked image is generated through a bitwise AND operation, effectively highlighting the region of interest. To precisely extract the object, minimum and maximum non-zero-pixel coordinates are calculated, defining a bounding box. A cropped image is then obtained, which is resized to a standard 224 × 224 pixel size. In the subsequent phase, we employed two concurrent Convolutional Neural Network (CNN) models, specifically ResNet50 and VGG-19, for feature extraction from the dataset.

The ResNet50 model was programmed to extract 2048 unique features, while the VGG19 model focused on extracting 512 distinct features from the same dataset. These extracted features were later given for the feature selection process using ANOVA and Random Forest. The features extracted from ResNet50 undergo feature selection using Scikit-Learn's SelectKBest method with ANOVA (Analysis of Variance). This process involves selecting the 500 most relevant features from the input data and prioritizing their significance about the target labels. The result is a subset of the input features that consists of the 500 most informative ones, as determined through ANOVA analysis. In a similar vein, the features extracted from ResNet50 also undergo feature selection via ScikitLearn, but this time utilizing a Random Forest classifier. This method entails creating a feature selector that takes into account 200 decision trees within the Random Forest. The importance of features is determined by setting a threshold of 1.25 times the median importance score. A total of 632 features are chosen for their significance during this process. The retrieved features are then grouped into a new Data Frame.

The performance metrics produced by using KNN, SVM, and Nave Bayes classifiers on the features selected by ANOVA and Random Forest feature selection techniques were as follows: 80.83% accuracy,86.15% F1-Score, 84.18% precision, 80.4% of weighted F1 Score and 88.7% of recall rate. Two distinct strategies were used to pick a subset of features from the VGG19-extracted features. First, an ANOVA-based feature selection method was used, which resulted in the selection of 500 features from an initial 512. In addition, 154 features were chosen from the VGG19-extracted feature set using a Random Forest-based feature selection technique. The performance metrics achieved by using KNN, SVM, and Nave Bayes classifiers on features selected by ANOVA and Random Forest feature selection approaches were as follows: 77.3% accuracy, 85.05% precision, 81.08% recall rate, 82.33% F1-Score, and 77.42% weighted F1 Score.

6.1. Comparison with the other existing models

We have compared the classification results of the proposed method with various existing models. The outcomes of this analysis are systematically presented in Table 7. Among the models enlisted in Table 7, VGG19+RANDOM FOREST+SVM with ISBI 2019, the dataset achieves the best result with an accuracy of 86.2%. From Table 7 it is clear that the proposed method exhibits superior performance compared to another method. The graphical representation of the results is shown in Figure 9.

Table 7: Comparison of the proposed method with the other existing methods

Reference	Dataset	No. of Images	Method	Result(%)
Uddin et al. 2022 [28]	ISBI 2019	15114	VGG19	78
Khilji et al. 2020 [29]	C_NMC	3389	CNN Model	77.934
Ding et al. 2019 [30]	C_NMC	10661	Ensemble	85.6
Bold u et al. 2019 [31]	Clinic Hematology Service of the Hospital	5493	LDA	85
The proposed method for Deep Feature Selection	**ISBI 2019**	**10661**	**VGG19+Random Forest+SVM**	**86.2**

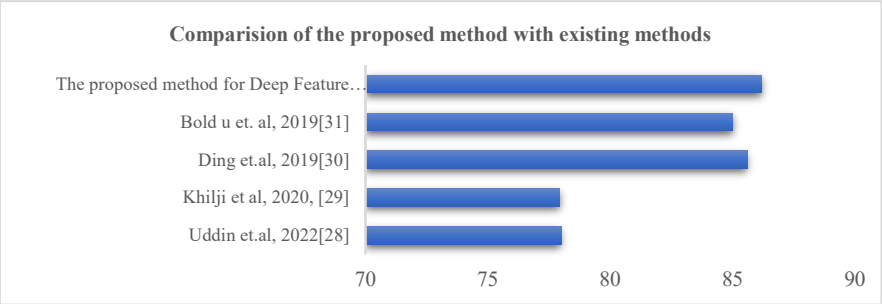

Figure 9: The graphical representation of the results with the existing methods

7. CONCLUSION AND FUTURE WORK

Adults and children can both be affected by leukemia, a type of blood cancer. The treatment for leukemia is determined by the exact subtype of the disease and the amount of its spread in the body. Initial stage detection of this ailment is critical to ensuring timely and appropriate treatment, ultimately leading to recovery. An automated diagnostic method for Acute Lymphoblastic Leukemia (ALL) was created in this study. Overall, this study shows that combining effective picture preprocessing, feature extraction from deep CNN models, and robust feature selection methods can greatly improve machine learning classifier performance. The results demonstrate the capability of ResNet50 and VGG19 in extracting informative features for the task at hand, as well as the relevance of adopting proper feature selection approaches for optimal classification performance. When the two feature extraction methods (ResNet50 and VGG19) and their related feature selection techniques (Random Forest and ANOVA) for the given C-NMC 2019 dataset are compared, it is clear that VGG19 with Random Forest feature selection outperforms other combinations in terms of classification performance. When combined with ANOVA and Random Forest feature selection, the ResNet50 model achieved 84.18 % of precision, 86.15% of F1-score, 80.83 % of accuracy, 88.7% of recall rate, and 80.4% of Weighted F1 Score. These results reveal that this combination performed exceptionally well in recognizing and classifying the target labels within the dataset, demonstrating its efficacy in extracting and selecting highly relevant characteristics.

Funding: This research received no external funding.

Data Availability: Datasets related to this article are taken from an open-source online data repository hosted at C_NMC 2019.

REFERENCES

[1] Roopashree, M. Suvarna and Dayakshini, "Categorization & classification of acute & chronic leukaemia using Visual Geometry Group-16 deep convolutional neural network architecture," in *Proceedings of the 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, Coimbatore, India, 2023, pp. 1–6. doi: 10.1109/ICECCT57124.2023.10179851.

[2] A. Shivathaya, A. M. S., K. V. Bhat, P. N., and R. Nayak, "Investigative study for identification & categorization of leukaemia cell," in *Proceedings of the 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, Coimbatore, India, 2023, pp. 1–7. doi: 10.1109/ICECCT56650.2023.10179792.

[3] M. Stanulla and A. Schrauder, "Bridging the gap between the north and south of the world: The case of treatment response in childhood acute lymphoblastic leukemia," *Haematologica*, vol. 94, no. 6, pp. 748–752, 2009. doi: 10.3324/haematol.2009.006783.

[4] P. M. Pillai and W. L. Carroll, "Acute lymphoblastic leukemia," in *Lanzkowsky's Manual of Pediatric Hematology-Oncology*, 8th ed., J. Lanzkowsky, J. M. Lipton, and J. F. Fish, Eds. Elsevier, Amsterdam, Netherlands, 2021, pp. 413–438. doi: 10.1016/B978-0-12-821671-2.00004-0.

[5] National Cancer Institute, "Cancer Stat Facts: Leukemia — Acute Lymphocytic Leukemia (ALL)," 2020. [Online]. Available: https://seer.cancer.gov/statfacts/html/alyl.html. Accessed: Feb. 13, 2025.

[6] W. Ladines-Castro, F.J. González-Castro, A.L. Fernández-Figueroa, E.O. Rodríguez-Pérez, L.M. Sánchez-Corona, and M.A. Chacón-Rodríguez, "Morphology of leukaemias," *Revista Médica del Hospital General de México*, vol. 79, no. 2, pp. 107–113, 2016. doi: 10.1016/j.hgmx.2015.06.007.

[7] R. B. Hegde, K. Prasad, H. Hebbar, and B. M. K. Singh, "Image processing approach for detection of leukocytes in peripheral blood smears," *Journal of Medical Systems*, vol. 43, no. 5, 2019. doi: 10.1007/s10916-019-1219-3.

[8] A. Ratley, J. Minj, and P. Patre, "Leukemia disease detection and classification using machine learning approaches: A review," in *Proceedings of the 2020 First International Conference on Power, Control and Computing Technologies (ICPC2T)*, 2020, pp. 161–165. doi: 10.1109/ICPC2T48082.2020.9071471.

[9] P. K. Das, A. Pradhan, and S. Meher, "Detection of acute lymphoblastic leukemia using machine learning techniques," in *Machine Learning, Deep Learning, and Computational Intelligence for Wireless Communication: Proceedings of MDCWC 2020*, 2021, pp. 425–437.

[10] S. Mandal, V. Daivajna, and V. Rajagopalan, "Machine learning-based system for automatic detection of leukemia cancer cell," in *Proceedings of the 2019 IEEE 16th India Council International Conference (INDICON)*, 2019, pp. 1–4.

[11] I. Alrashdi and A. Alqazzaz, "Synergizing AI, IoT, and blockchain for diagnosing pandemic diseases in smart cities: Challenges and opportunities," *Sustainable Machine Intelligence Journal*, vol. 7, pp. 1–28, 2024. doi: 10.61356/smij.2024.77106.

[12] N. Khalil, M. Elkholy, and M. Eassa, "A comparative analysis of machine learning models for prediction of chronic kidney disease," *Sustainable Machine Intelligence Journal*, vol. 5, pp. 1–3, 2023.

[13] Q. Wang, Y. Xie, M. Wang, H. Zhou, Y. Xu, L. Liu, and Z. Zhang, "Deep learning approach to peripheral leukocyte recognition," *PLoS One*, vol. 14, no. 6, pp. 1–18, 2018. doi: 10.1371/journal.pone.0218808.

[14] S. Shafique and S. Tehsin, "Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks," *Technology in Cancer Research & Treatment*, vol. 17, pp. 1–7, 2018. doi: 10.1177/1533033818802789.

[15] K. K. Anilkumar, V. J. Manoj, and T. M. Sagi, "Automated detection of B cell and T cell acute lymphoblastic leukaemia using deep learning," *IRBM (Institut de Recherche Biomédicale)*, vol. 43, no. 5, pp. 405–413, 2022. doi: 10.1016/j.irbm.2021.05.005.

[16] N. Faruqui, S. Kumar, M. Y. Khan, M. M. Ahmed, S. K. Gupta, and A. R. Al-Ahmad, "Healthcare as a service (HAAS): CNN-based cloud computing model for ubiquitous access to lung cancer diagnosis," *Heliyon*, vol. 9, no. 11, p. e21520, 2023. doi: 10.1016/j.heliyon.2023.e21520.

[17] M. Zhou et al., "Development and evaluation of a leukemia diagnosis system using deep learning in real clinical scenarios," *Frontiers in Pediatrics*, vol. 9, no. June, pp. 1–10, 2021. doi: 10.3389/fped.2021.693676.

[18] A. Akter et al., "Robust clinical applicable CNN and U-Net based algorithm for MRI classification and segmentation for brain tumor," *Expert Systems with Applications*, vol. 238, no. PF, p. 122347, 2024. doi: 10.1016/j.eswa.2023.122347.

[19] R. Nayak, A. Bekal, M. Suvarna, and D. Sathish, "Identifying subtypes of acute lymphoblastic leukemia using blood smear images: A hybrid learning approach," *Journal of The Institution of Engineers (India): Series B*, 2024. doi: 10.1007/s40031-024-01069-0.

[20] The Cancer Imaging Archive, "Cnmc2019 Dataset: All Challenge Dataset of ISBI 2019 (C-NMC 2019)," 2019. [Online]. Available: https://seer.cancer.gov/statfacts/html/alyl.html. Accessed: Feb. 13, 2025.

[21] A. Gupta et al., "GCTI-SN: Geometry-inspired chemical and tissue invariant stain normalization of microscopic medical images," *Medical Image Analysis*, vol. 65, p. 101788, 2020. doi: 10.1016/j.media.2020.101788.

[22] R. Gupta, P. Mallick, R. Duggal, A. Gupta, and O. Sharma, "Stain color normalization and segmentation of plasma cells in microscopic images as a prelude to development of computer-assisted automated disease diagnostic tool in multiple myeloma," *Clinical Lymphoma, Myeloma & Leukemia*, vol. 17, no. 1, p. e99, 2017.

[23] R. Duggal, A. Gupta, R. Gupta, M. Wadhwa, and C. Ahuja, "Overlapping cell nuclei segmentation in microscopic images using deep belief networks," in *Proceedings of the 10th Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP)*, 2016, pp. 1–8.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[25] K. Simonyan, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[26] R. K. Sachdeva et al., "A systematic method for breast cancer classification using RFE feature selection," in *Proceedings of the 2022 2nd International Conference on Advanced Computing and Innovative Technologies in Engineering (ICACITE)*, 2022, pp. 1673–1676. doi: 10.1109/ICACITE53722.2022.9823464.

[27] M. N. Triba et al., "PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters." *Molecular BioSystems*, vol. 11, no. 1 13-19, 2015.

[28] W. H. Abir, M. F. Uddin, F. R. Khanam, and M. M. Khan, "Explainable AI in diagnosing and anticipating leukemia using transfer learning method," *arXiv e-prints*, p. arXiv--2312, 2023.

[29] I. Q. Khilji, K. Saha, J. A. Shonon, and M. I. Hossain, "Application of homomorphic encryption on neural network in prediction of acute lymphoid leukemia," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 6, pp. 350–360, 2020. doi: 10.14569/IJACSA.2020.0110646.

[30] Y. Ding, Y. Yang, and Y. Cui, "Deep learning for classifying white blood cancer," in *ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging: Select Proceedings*, Springer, 2019, pp. 33–41.

[31] L. Boldú, A. Merino, S. Alférez, A. Molina, A. Acevedo, and J. Rodellar, "Automatic recognition of different types of acute leukaemia in peripheral blood by image analysis," *Journal of Clinical Pathology*, vol. 72, no. 11, pp. 755–761, 2019.