Mathematical and Computational Applications

MDPI

Article

A Comprehensive Assessment and Classification of Acute Lymphocytic Leukemia

Payal Bose [1] and Samir Bandyopadhyay [2,*]

[1] Department of Computer Science and Engineering, Swami Vivekananda University, Kolkata 700120, India; bosepayal91@gmail.com
[2] Department of Computer Science and Engineering, The Bhawanipur Education Society College, Kolkata 700020, India
* Correspondence: 1954samir@gmail.com

Abstract: Leukemia is a form of blood cancer that results in an increase in the number of white blood cells in the body. The correct identification of leukemia at any stage is essential. The current traditional approaches rely mainly on field experts' knowledge, which is time consuming. A lengthy testing interval combined with inadequate comprehension could harm a person's health. In this situation, an automated leukemia identification delivers more reliable and accurate diagnostic information. To effectively diagnose acute lymphoblastic leukemia from blood smear pictures, a new strategy based on traditional image analysis techniques with machine learning techniques and a composite learning approach were constructed in this experiment. The diagnostic process is separated into two parts: detection and identification. The traditional image analysis approach was utilized to identify leukemia cells from smear images. Finally, four widely recognized machine learning algorithms were used to identify the specific type of acute leukemia. It was discovered that Support Vector Machine (SVM) provides the highest accuracy in this scenario. To boost the performance, a deep learning model Resnet50 was hybridized with this model. Finally, it was revealed that this composite approach achieved 99.9% accuracy.

Keywords: blood cancer; composite learning; deep learning (DL); hybrid model; acute lymphoblastic leukemia (ALL); machine learning (ML); Resnet50; support vector machine (SVM)

check for updates

Citation: Bose, P.; Bandyopadhyay, S. A Comprehensive Assessment and Classification of Acute Lymphocytic Leukemia. *Math. Comput. Appl.* **2024**, *29*, 45. https://doi.org/10.3390/mca29030045

Academic Editor: Leonardo Trujillo

Received: 26 April 2024
Revised: 5 June 2024
Accepted: 7 June 2024
Published: 9 June 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Leukemia is a form of blood cancer [1] that results in an increase in the number of white blood cells in the body. Whenever these white cells force away the red blood cells and platelets that the human body need to function correctly, the chances of having this disease increase. It is a diverse category of hematopoietic cancers caused by abnormal multiplication of growing leukocytes. These are categorized as acute or chronic and myeloid or lymphoid depending on the originating cell. The most common variants are acute myeloid leukemia (AML) and chronic myeloid leukemia (CML), both of which include the myeloid branch. The excessive production of red blood cells, white blood cells, or platelets by the bone marrow characterizes disorders of these myeloid origins. As more excess cells accumulate in the bone marrow or blood, it frequently becomes worst with time. A viral infection, exhaustion, anemia, hemorrhage issues, and other complications could result from this [2,3].

However, acute lymphoblastic leukemia (ALL) and chronic lymphocytic leukemia (CLL), both of which involve the lymphoid sequence. Cells that develop into lymphocytes or white blood cells are the source of the lymphocytic leukemia stream. In simple terms, its cancerous cells are found in blood and bone marrow, respectively. The most prevalent type of leukemia malignancy that affects adults is called CLL. Despite the fact that it can strike grownups, ALL mainly affects adolescents [4,5].

It was estimated that in 2020 that there roughly 500,000 new cases of leukemia were found globally. Globally, there was an approximately five-fold fluctuation in the age-standardized prevalence of occurrence, which was reported to be 5.4 per 100,000. The number of linked deaths reported in 2020 was nearly 350,000 in the context of death rates. Leukemia-related mortality varied less by region globally. Most parts of Asia, Europe, America, Australia, and New Zealand recorded death rates between 2.5 and 4.0 per 100,000 people [6–8]. The worldwide cancer research organization revealed that by 2023, there would be 2.4 additional instances per 100,000 people, and the fatality rate will be increased to between 1.4 and 1.8 per 100,000 people. According to reports, the total percentage of survivors increased by more than 70% during the preceding five years due to the use of contemporary medical advances [9,10].

In this investigation, acute lymphoblastic leukemia (ALL) was used as the research subject.

Microscopic image analysis is very important in early leukemia testing and accurate diagnoses for this illness [11]. This analysis describes how testing blood sample appears under a microscope, including the size of blood cells, shape, and the amounts of different types of blood cells, including red blood cells, white blood cells, and platelets.

The correct identification of leukemia at any stage is essential. Since current traditional approaches rely mainly on microscopic inspection, which itself is time-intensive and highly dependent on field experts' knowledge, it is possible that a poor understanding and a long period of examination might damage the human body. In this situation, an automated leukemia identification provides a new avenue for reducing human participation while delivering more reliable diagnostic information. ALL prediction is a challenging process. Normal physical examinations and information gathered from groups of specimens are time-dependent and money-consuming methods for identifying and predicting leukemia. The condition has occasionally been seen to advance from the premature time to noticeably greater levels due to an inadequate evaluation. In contrast to general medical inspection, digital image analysis is now more successful at identifying this condition. The medical community greatly benefits from ML. It is a technology utilized by the medical industry to assist medical practitioners in managing critical data and delivering clinical outcomes. Finding patterns and insights from a picture that would be hard to detect intuitively can be assisted by employing ML techniques used in healthcare. In contrast to conventional methods, ML models offer a forecast that is reliable and efficient in terms of effectiveness, cost, and time.

In this work, a hybrid learning methodology is employed to build an automated leukemia monitoring system. This technique will analyze blood smear images for the existence of leukemia. During the initial phase, three well-known single learning models are employed to predict ALL categories. The best single learning model discovered in the first phase is fused with a deep neural model network in the following phase to improve model forecasting accuracy. The relevance of this work is highlighted below:

1. To assess and forecast the type of leukemia, an automated technique is developed;
2. A DL model called Resnet50 is used to acquire characteristics, and an ML model called SVM is used to classify data in order to establish this technique;
3. This system uses digital blood smear images for detection and prediction;
4. The constructed smart strategy uses known sets of information to forecast and monitor the form of Leukemia;
5. Whenever it detects a potentially malicious blood cell, this automatic system will send an alarm;
6. This approach is significantly more precise and faster than traditional techniques.

2. Literature Survey

In this section, the scientific research related to the inquiry is comprehensively assessed. The suggested process investigated how to detect and classify acute lymphoblastic leukemia. As a result, a comprehensive review focusing on the mentioned principles, as well as an overview of the associated literary work, is offered below.

Blood cancer [12] has become an increasing problem in recent decades, necessitating earlier detection in order to commence appropriate treatment. The therapeutic diagnosis process is expensive and time-consuming, requiring the participation of healthcare professionals and a series of examinations. As a result, an automatic detection method for precise prognosis is far more important than the conventional approach. With the advent of technology, finding abnormal cells from blood smear images has become considerably easier, even more reliable, and significantly less time consuming than the traditional approach. Many scientists and researchers throughout the worldwide have focused on developing progressively inventive and accurate ways for such systems and related solutions for these scenarios.

Various investigators employ a variety of computer vision approaches for identification and machine learning models for prediction. To identify the kind of leukemia from blood smear pictures, they adopted a support vector system based on radial kernels [13]. To identifying the characteristics of these cancerous cells, other studies also employed a few different ML models [14–16]. Several ML-based models for leukemia detection and classification are presented in depth by the authors in their review paper. They provide as a concise summary of various performance metrics, benefits, and drawbacks of several related studies that will be informative for other authors [17,18].

Modern society is very interested in deep learning models. These algorithms are capable of processing complicated and huge datasets that would be challenging for conventional ML methods to comprehend. A technique for leukemia diagnosis using labeled bone marrow pictures was put forth by certain researchers. In order to deliver trustworthy prediction performance, they employed a strong classification methodology using the deep convolutional model approach [19–22]. A deep convolution model with a distinct ALLNET structure was suggested by some other authors for forecasting. The suggested framework has the maximum level of precision [23]. Several authors have suggested a multi-step DL strategy. They used this strategy to effectively separate the cells from the pictures and make reliable predictions [24].

A brief literature analysis is illustrated in Table 1.

Table 1. Brief analysis of the literature survey.

Authors (Year)	Method Used	Dataset Type	Performance Achieve
Erum Yousef Abbasi et. al. (2024) [25]	ML Models (RF, NB, DT, LR, GB) and DL Models (RNN, FNN)	Large	ML achieve = 97% DL achieve = 98%
Althaf Ali A et al. (2023) [26]	ANN	Large	92.1%
A. Khuzaim Alzahrani et al. (2023) [22]	UNET	Large	97.82%
W. Rahman et al. (2023) [27]	CNN (RESNET 50)	Medium	99.84%
Almadhor et al. (2022) [28]	Single learning model, ensemble model and pre-trained CNN model	Large	90%
Tawfeeq Shawly, Ahmed A. Alsheikhy (2022) [29]	CNN with AlexNet	Large	98%
Zhou et al. (2021) [30]	CNN (Clinical data)	Large	82.93%
Ansari, Sanam, et al. (2023) [31]	Fuzzy deep neural network	Large	98.8%
Sampathila et al. (2022) [32]	CNN with ALLNET	Large	95.54%

3. Methods Details

Leukemia is a kind of cancerous disease that affects blood production elements, particularly bone marrow. In therapeutic terminology, there are several distinct examinations are available that may be employed to detect leukemia. The amounts of white blood cells (WBCs), red blood cells (RBCs), and platelets in the bloodstream are determined by a complete blood count. Cell examinations can be carried out from the bone marrow or lymphatic vessels to search for signs of leukemia and the rate at which it is growing. However, it takes time and requires skilled medical specialists. Numerous computer vision algorithms are employed in Digital Image Processing Techniques for the identification of leukemia. To locate disease-affected tissues, different color intensification and color segmentation approaches were used in this investigation.

3.1. Brightness, Contrast, Sharpness, and Color Intensity Enhancement

In machine vision, brightness [33] is defined as the measurable amplitude of all the image pixels that compose an assembly that made up the digital picture once it has been taken, processed, and presented. To modify the intensity of the brightness of an image, the image pixel intensities should be adjusted by a fixed value. Simply adding a positive fixed value to all the image pixels increases the brightness level of the image. Deducting a positive number from all the picture pixels, on the other hand, darkens the image.

$$Adjustment_{Brightness} = \begin{cases} Brighter\ when\ Pixel_{value} + K \\ Darker\ when\ Pixel_{value} - K \end{cases} \quad (1)$$

where K = a constant value for brightness adjustment.

Improving the contrast [33] level of an image improves the range between black and white pixels, making white parts lighter and black ones darker. It simply redesigned the pictures pixel intensity values. A well-contrasted photograph features prominently black and white distinctions.

$$Adjustment_{Contrast} = \begin{cases} Dark\ when\ Pixel_{value} + M \\ White\ when\ Pixel_{value} - M \end{cases} \quad (2)$$

where M = a constant value for contrast enhancement.

The degree of clarity that an imaging modality can recreate is determined by the sharpness [34] of a picture. It is characterized by the margins between distinct hues or colors in each region. Image sharpening is a technique used to make digitized photos look sharper or clearer. It is a crucial tool in the image processing system. Appropriate sharpening of an image makes it appear more noticeable and livelier. Figure 1 shows the definition of sharpness level graphically.

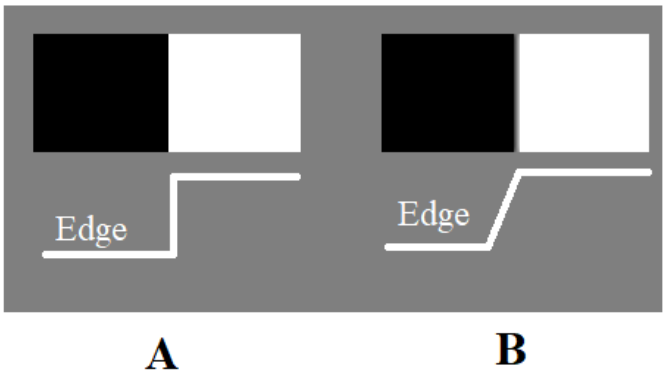


Figure 1. Definition of sharpness level: (**A**) high sharpness; (**B**) low sharpness.

During analysis, an image enhancement approach is employed to increase the image's appearance and save the informative characteristics of the original data. Color augmenta-

tion is an important aspect of it. This approach is a set of processes that strive to improve the visual look of a picture or transform the picture to a state that is more suitable for analysis by a person or computer. This procedure also included brightness and contrast adjustments, as well as histogram equalization adjustments.

3.2. Image Segmentation

Image segmentation is a technique [35,36] used in digital image processing. It analyzes a picture and divides it into distinct segments or sections based on the pixels in the object's attributes. It is widely used to discern and properly identify foreground and background areas.

An image can be segmented using several different techniques. One of these, the thresholding-based segmentation method [37], is both fast and significant. The intensity histogram of each pixel in the image is considered throughout this approach. After that, a specific threshold value is set to segment the image. Global thresholding is one of the most well-liked techniques for segmenting images based on thresholds. The idea behind global thresholding is that the subject can be separated from the background using a straightforward process that compares image contents with a predetermined threshold value when the image has a bimodal histogram. Figure 2 depicts the histogram distribution of the global thresholding model.

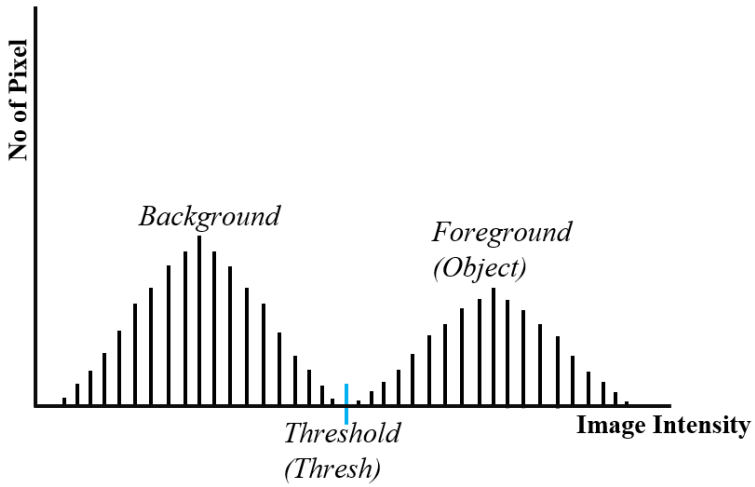


Figure 2. Histogram distribution of the global thresholding model.

Let us consider $(m,\ n)$ to be the coordinate of an image pixel, and the threshold value of an image is defined as $Thresh$. Then, the threshold image $T(m,n)$ is defined as follows:

$$T(m,n) = \begin{cases} 0\ if\ (m,n) \leq Thresh \\ 1\ if (m,n) > Thresh \end{cases} \quad (3)$$

The result of the thresholding approach is a binary image. The pixels with an intensity value of 1 are specified as foreground objects, and pixels with an intensity value of 0 are specified as background objects.

3.3. Feature Extraction and Machine Learning Models for Classification

3.3.1. Feature Selection

To develop a ML model, only a few variables in the dataset play a main role. As a result, the remaining features either become unnecessary or irrelevant. Finding and choosing the best characteristics from the dataset is very crucial for reducing this redundancy and improving predictive accuracy. A feature is a characteristic that affects or helps researchers to solve a problem, and selecting the key characteristics for the model is referred to as feature selection.

For feature selection [38], typically, two models—one supervised and the other unsupervised—are considered. Supervised feature selection approaches are employed when the dataset is categorized and aids researchers in finding the pertinent features to improve the model's efficacy. This is the main justification for using a supervised feature selection model in this research. One of the broadly used supervised feature selection techniques is the histogram-oriented gradient approach [39]. In the confined area of an image, this process counts instances of gradient direction. It concentrates on an object's structure and extracts the information from it. By employing the gradient's dimensions and orientation to create histograms, it collects the characteristics from those areas of the image. If g_x and g_y are the gradient of a pixel of an image I, then it is calculated as follows:

$$g_x = I(row, column + 1) - I(row, column - 1) \tag{4}$$

$$g_y = I(row - 1, column) - I(row + 1, column) \tag{5}$$

And the magnitude of that pixel is represented as follows:

$$Magnitude(m) = \sqrt{g_x^2 - g_y^2} \text{ and } Angle(\theta) = \left| tan^{-1} \frac{g_x}{g_x} \right| \tag{6}$$

3.3.2. Machine Learning Models for Classification

ML [40] is a type of artificial intelligence. It concentrates mostly on developing algorithms that allow a computer to autonomously learn from available knowledge and prior experiences. The identification of correlations from given knowledge is one of the objectives of a machine learning model. The patterns are then discovered from the training data using a learning algorithm, which creates a model that recognize the patterns and forecasts the results of new data.

Depending on the kind and characteristics of the task, three different types of machine learning models are available: (1) supervised, (2) unsupervised, and (3) reinforcement. The most straightforward one is the supervised model. It is mostly applied to training data with label information. The input–output combination principle explains how it functions. It is important to create an operation that can be trained using a learning set of data before being used on selections of unidentified data to execute forecasting. The effectiveness of supervised learning is evaluated using sets of labeled data. Three supervised ML models—Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Naïve Bayes (NB)—are employed to carry out this investigation. These three models are widely used and effective for multi-class classification problems.

The SVM [41] method aims to generate a decision boundary that can characterize the n-dimensional space. Depending on that, it can quickly classify the latest data point. The term "hyperplane" refers to this optimal decision threshold. This approach generates the extreme vectors or points that assist in the formation of this hyperplane. Support vectors are referred to as these extreme points. Whenever there is a distinct line distinguishing the classes, it functions much as intended. SVM becomes a potent tool to create forecasts for all the data that cannot be characterized by linear decision functions. In addition to using relatively low memory, it works well in high-dimensional spaces.

The KNN [42] model is a highly straightforward and efficient learning approach. By placing the new case in the category that closely resembles the currently accessible categories, it assumes that the new instances and existing instances are alike. Every new instance is categorized based on these similarities after the system stores all the existing information. In other words, when data appear, this algorithm can quickly classify the new information into a suitable category.

Naïve Bayes [43] is a simplistic learning algorithm. It makes predictions using Bayes principles. It is termed Naïve since it is dependent on the assumption of conditional probability. Since it operates independently with each characteristic, it may be employed for big

datasets to create forecasting models. It is particularly sensitive to other aspects, which indicates that it is not significantly affected by other components due to its Naïve characteristic.

3.4. Hybrid Approach

The hybrid approach [44] is a combination of two or even more computational approaches that outperform any single approach. It aids in the improvement of data analysis. The advantage of implementing this approach is that it improves performance by increasing model effectiveness. The DL and ML models are integrated to carry out this study. In order to categorize the input photos depending on these deep characteristics, the deep neural model first extracts these characteristics from the input images. Then, the ML model is employed to classify the images based on these features. The extraction of deep features makes this hybrid approach particularly effective. The deep features gathered all the essential data for classification and performed considerably better than any generative model. Figure 3 provides an illustration of the proposed hybrid approach preferred for this study.

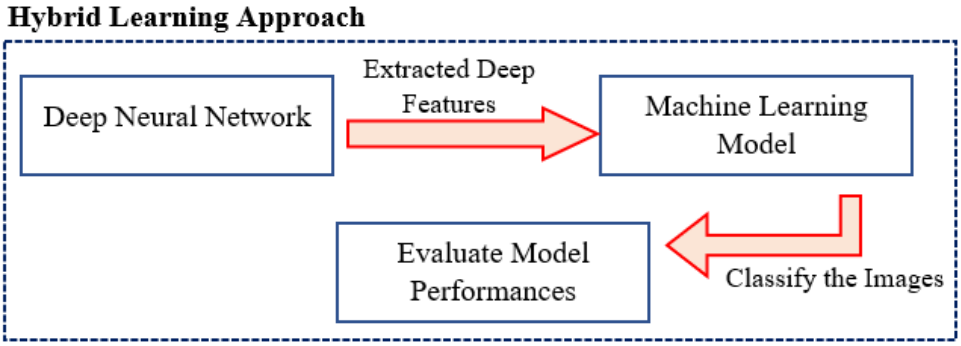


Figure 3. Outline of the proposed hybrid learning approach.

4. Proposed Methodology

Leukemia is a kind of blood cancer that is in its initial phases. At this point, red blood cells are diminishing, while the white blood cells are developing. In this situation, if a digital image of the blood cells can be made, it will be discovered that the amount of non-blood cells is significantly greater than the proportion of true blood cells. Such aberrant non-blood cell occurrences and counting the numbers can be used to determine whether an individual has leukemia. To identify if a person has leukemia, wide varieties of clinical evidence, hematological and bone marrow observations, and outcomes of more specialized definitive tests studies are required. These procedures are both time-consuming and expensive. Therefore, in response to these concerns, this study describes a strategy for forecasting leukemia using digital image evaluation.

To assess and anticipate these digital photos, the following procedures were undertaken: (1) detection and (2) identification. Figure 4 depicts the prediction process graphically.

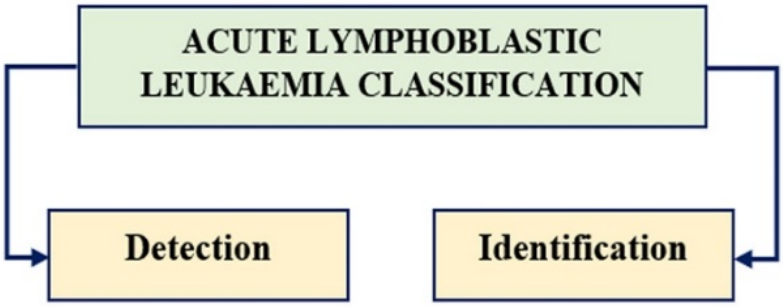


Figure 4. The prediction process for acute lymphoblastic leukemia prediction.

4.1. Detection Process

The fundamental and, perhaps, most successful approach for identifying leukemia is to count the amounts of white blood cells, platelets, and red blood cells in a person's blood. Medical testing research revealed that the non-blood cells detected in leukemia patients' blood are substantially darker than natural blood cells. This color contrast [45] principle may now be used to locate leukemia cells by examining a digital picture.

This study employed a machine vision technique to identify leukemia from digital pictures. It has been noted that appropriate non-blood cells are often difficult to detect due to low image clarity and details. To correctly identify leukemia from blood smear pictures, the affected cells had to be precisely distinguished. This was one of the most crucial procedures. To address this issue, a considerable pre-processing procedure was implemented in this study. This pre-processing procedure consisted of four stages: (1) equalizing the color intensity level, increasing the contrasting level, and sharpening the objects' boundaries to make the impacted cells more visible; (2) using the color clustering model to detect the cell boundaries; (3) applying the morphological operators to segment the appropriate regions; and (4) enumerating the damaged cells.

4.2. Identification Process

The procedure of predicting acute leukemia is difficult. The identification and prediction of leukemia through standard physical evaluations, as well as the way of gathering information from collective samples, are both time-consuming and expensive. In the case of a lack of an adequate diagnosis, the disease has been known to progress from the early stage to considerably higher levels at times. As a result, the patient's life might be jeopardized.

However, diagnosing this condition through digital image evaluation is more effective than conventional healthcare examination. This procedure is considerably more significant, requires less time, is much less expensive, and is far more precise than the conventional one. This investigation employed machine learning and fusion learning methods to detect leukemia from digital images. One of the benefits of employing these learning techniques is their ability to rapidly and precisely detect the type of condition.

To carry out this investigation, first, three well-known single learning models were utilized. The best model was selected among them based on its precision. In order to boost the effectiveness of the forecasting model, the selected model was paired with a deep convolution system Resnet50 [46,47].

A complete workflow structure and pseudo-algorithm of the proposed model is illustrated in Figure 5, and the pseudo-codes for the proposed model are illustrated in Algorithms 1–3.

Algorithm 1: Pseudo-code for image pre-processing.

```
Input: RGB image (img)
For each pixel in img i = 0 to n
      img_b ← img[pixel(i)] + A          //Where A = the factor of image brightness level
      img _c ← img_b-B                   //Where B = the factor of image contrast level
end
img ← image_sharpen(img_c, S)          //Where S = the factor of image sharpen level

red_c ← img(:,:,1) + L1                //Where L1, L2, L3 are the R, G, B color level
green_c ← img(:,:,2) + L2                    adjustment factor
blue_c ← img(:,:,3) + L3

img ← concate(red_c, green_c, blue_c)
```

Algorithm 2: Pseudo-code for leukemia cell detection and counting the no. of cells.

Input: RGB pre-processed image (img)
Initialize k = random values, iteration = n //K-means clustering method
For each pixel of img:
- Start iteration from 0 to n
  - Find the mean closest to the pixel using the Euclidean distance measurement
  - Assign an item to mean
  - Update the mean in that cluster depending on the k value
  - End iteration

End

img_cluster ← clustered_classified_image(img)
img_b ← convert_Binary(img_cluster) //Binary Conversion
img_e ← Dilate_image(img) //Morphological operation for segmentation to
img_seg ← erode_image(img_e) detect leukemia cells

img_masked ← mask(img, img_seg)
region_count = 0
For all segmented regions in img_masked j = 1 to m:
- img ← give _boundary[img_masked(m)]
- region_count = region_count + 1

End

Algorithm 3: Pseudo-code for leukemia prediction.

Features ← HOG_Features_Extraction(img) //feature extraction module
For all images in the dataset:
Training_data← Features 70%
Test_data ← Features 30%

Build machine learning models SVM, KNN, Naïve Bayes
For SVM, initialize no _of_iteration = random value
- train_svm(Training_data, class_lable, no_of_iteration)

For KNN, initialize k = random value
- Calculate_nearest_neighbor(Training_data, k)

For Naïve Bayes
- Feature_probability ← probability_of_occurance(Training_data)
- Feature_likelihood ← greatest_likelihood(Feature_probability)

Measure Accuracy of all classifiers
End
Best_predictive_model ← Best_Acuuracy(KNN, SVM, Naïve Bayes)

Deep_Feutures ← feature_extraction_Resnet50(features, labels)
For all images in the dataset:
- Training_data ← Deep_Features 70%
- Test_data ← Deep_Features 30%
- Train_Model ← Training (Training_Data, Best_predictive_model)
- Test_Model ← Testing (Train_Model, Best_predictive_model)
- Calculate Performance_Matrices (Accuracy, Precision, F1-Score)

End

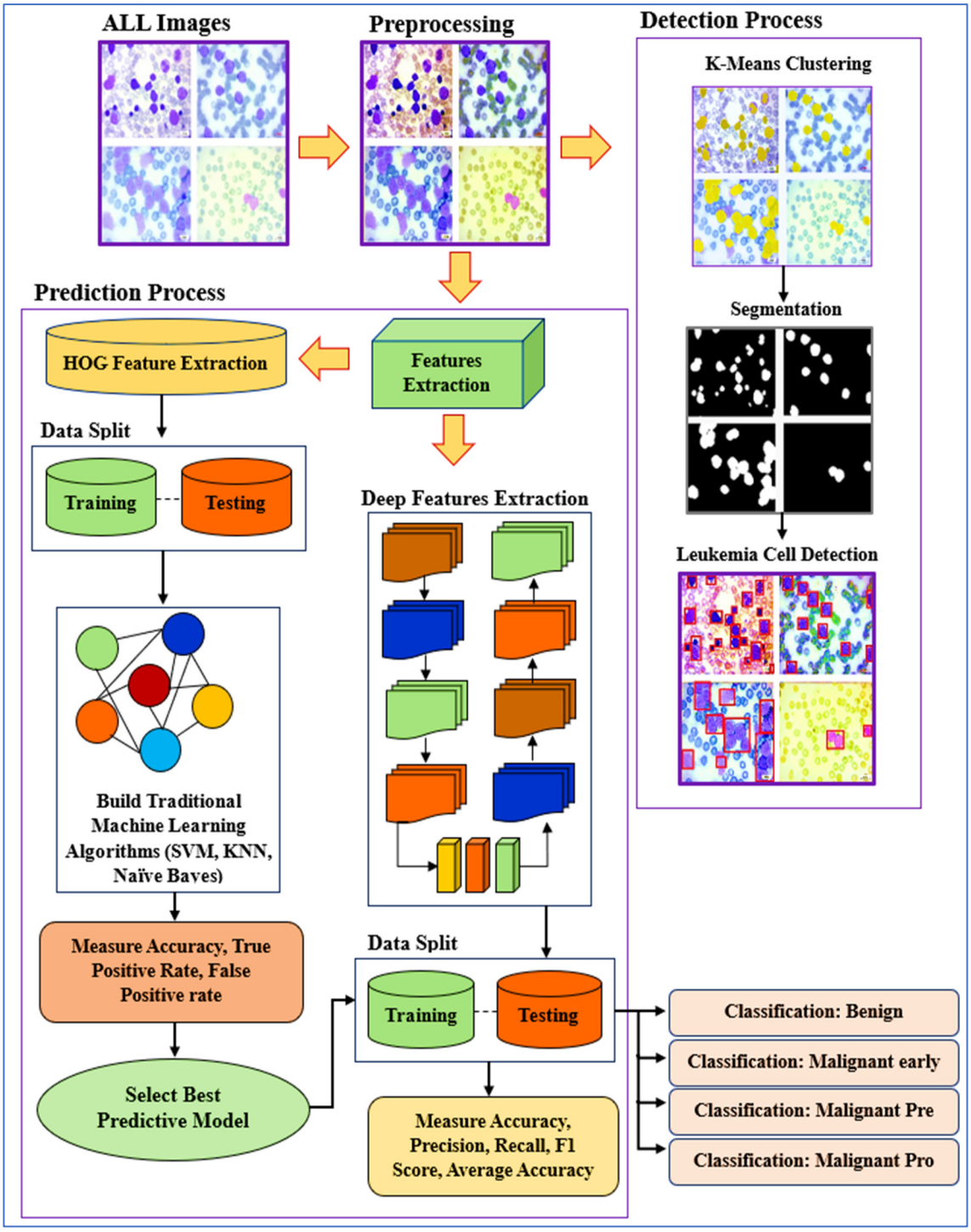


Figure 5. A complete workflow diagram for leukemia detection and prediction using the proposed methodology.

5. Experimental Results

5.1. Dataset Details

Image Samples for this investigation were gathered from the Kaggle library [48]. This collection included 3256 blood smear pictures from 89 people who were thought to have acute lymphocytic leukemia. The bone marrow research lab at Taleqani Hospital generated the samples for this collection. The collection is categorized into two primary

classes—'Benign' and 'Malignant—and three additional malignant classes—'Early', 'Pre', and 'Pro'. Figure 6 illustrates the distribution of images across all categories.

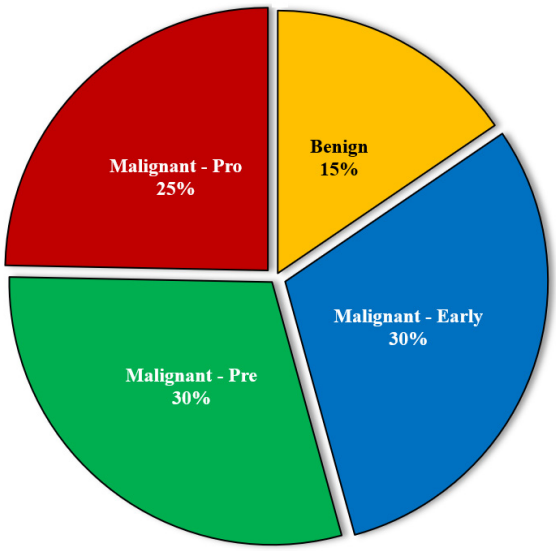


Figure 6. Visualization of Dataset.

5.2. The Outcome of Leukemia Detection

The results of leukemia identification from blood smudge images are shown in Figure 7. In the identification process, a number of algorithms based on computer vision were employed to recognize the non-blood components. The detailed descriptions of each stage in this procedure are provided in the illustrations beneath.

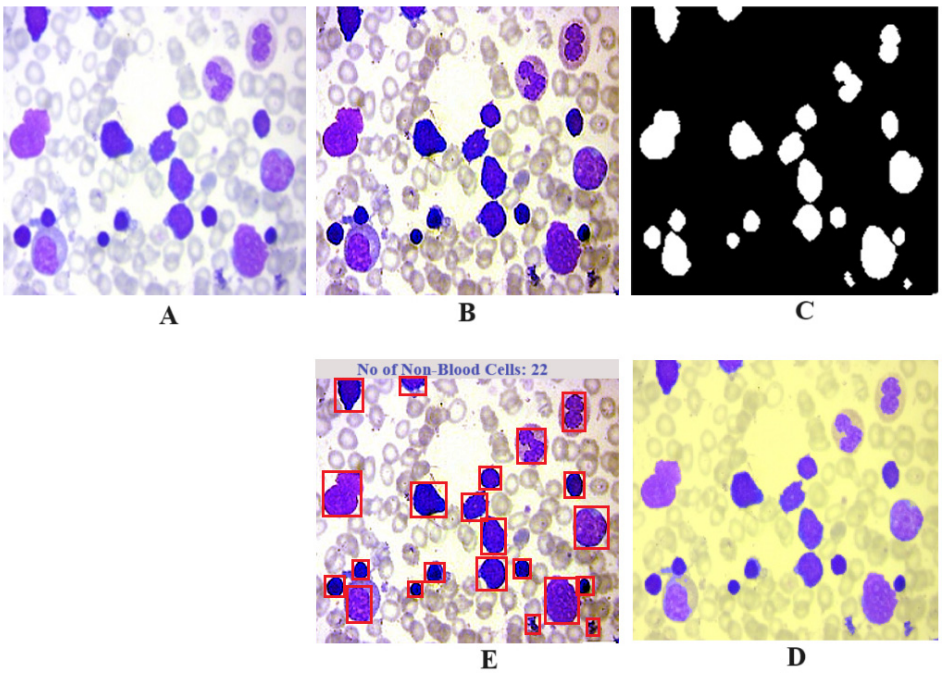


Figure 7. Leukemia detection process. (**A**) Original image: category benign; (**B**) brightness, contrast, and sharpness adjustment image; (**C**) binary segmented image; (**D**) highlighted non-blood cells after applying the K-means clustering method and morphological operators; and (**E**) no. of non-blood cells calculation.

5.3. The Outcome of Leukemia Identification

There are two stages in the leukemia prediction method. Three well-known machine learning processes are employed during the first stage. The specifications for each model prior to the training procedure are shown in Table 2. The effectiveness levels of the three single learning models that have been proposed are represented in Table 3. The effectiveness levels of the models were assessed using the following metrics: model accuracy, area under the curve (AUC), the rate of true positives (TPR), the rate of false negatives (FNR), the positive predicted value (PPV), and the false detection rate (FDR). The highest-scoring model is chosen from stage one based on these performance benchmarks. The receiver waveform evaluation (ROC) for the top model is illustrated in Figure 8. In the following step, a deep residual network is constructed to retrieve deep features, and the best predictive model from stage one is used for classification. The performance of this hybrid strategy is displayed in Table 4, and the receiver curve evaluation of each classification category is illustrated in Figure 9.

Table 2. Experimental specifications of applied single learning models.

Model Name	Model Details for Classification
Support Vector Machine (SVM)	Kernel: Linear Kernel scale: Automatic Box Constrain Level: 1 Standardize data: True PCA Disabled Multi-Class Method: One vs. One
K-Nearest Neighbor (KNN)	Preset: Weighted Number of Neighbors: 10 Distance Metric: Euclidean Distance Weighted: Squared Inverse Standardize data: True PCA Disabled
Naïve Bayes (NB)	Preset: Gaussian Distribution name for numeric predictor: Gaussian Distribution name for categorical predictor: MVMN (Multi-Variate Normal Distribution)

Table 3. Performance analysis of single learning models.

Model Name	Cross-Validation Value	Model Accuracy (%)	Model AUC (%)	Performance Matrices (%)			
				TPR	FNR	PPV	FDR
SVM	5	68.22	93.12	42.3	57.7	85.22	14.8
KNN	5	30.23	58.26	70.6	29.4	20.6	79.4
NB	5	60.11	86.20	89.3	10.7	52.3	47.7

Table 4. Performance analysis of the hybrid learning model.

Method Used	Class Names	Performance Metrices Analysis (%)				Classifier Average Accuracy (%)	Classifier Overall Accuracy (%)
		Accuracy	Precision	Recall	F1_Score		
Resnet50 with Support Vector Machine	Benign	98.55	92.62	91.44	91.99	99.42	99.98
	Malignant Early	99.35	95.00	95.91	95.44		
	Malignant Pre	99.88	98.63	98.62	98.62		
	Malignant Pro	99.91	99.61	99.24	99.39		

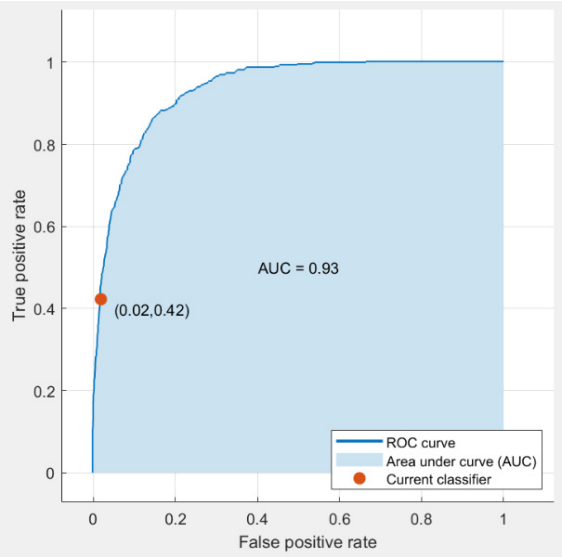


Figure 8. The ROC analysis for the top model SVM.

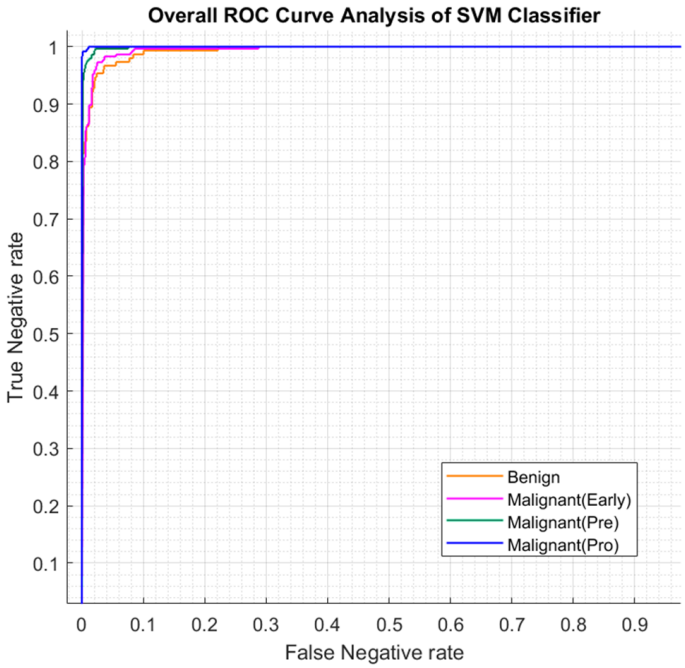


Figure 9. The ROC analysis for the hybrid learning approach [Resnet50 with SVM].

6. Discussions and Limitations

One of the main issues involved in assisting in and extending life is leukemia recognition at a preliminary phase. Therefore, creating a reliable detection method is one of the most significant priorities. Leukemia assessment and forecasting have proven to be challenging and time-consuming tasks based on traditional therapeutic approaches. Prediction using ML algorithms using just a digital blood smear image has become incredibly successful and efficient as a result of the high popularity of AI and ML in the medical industry. The major goal of this work is to create a framework that precisely forecasts acute lymphoblastic leukemia (ALL). Predicting ALL at its earliest stages is necessary to prevent it from spreading too widely. The integration of AI, ML, and DL into all of these domains has become crucial since they accelerate and automate the forecasting procedure.

Several researchers have employed ML and DL in various ways to predict leukemia. In the majority of scenarios, individual or combined neural networks for DL models and

individual or ensemble learners for ML models are employed for forecasting. In those instances, it has been discovered that the efficacy of a truly positive forecasting rate is compromised as a result of insignificant characteristics, an inadequate pre-processing procedure, and the improperly tuning of the categorization system.

An integrated model that combines ML and DL is created in this study to address these problems. Here, a DL model Resnet50 is employed to gather important deep characteristics that aid with creating a suitable features model for categorization. Resnet50, or residual learning, used to collect the residual characteristics compared to the specific characteristics. Furthermore, since it has 50 layers, it was able to extract greater amounts of residual information from the source images. To categorize the deep characteristics, SVM, a ML method, is used in this study. The main benefits of employing SVM are that it clearly separates the distinct classes with clear margins, is reasonably memory-efficient, and works efficiently in high-dimensional spaces. In light of these factors, the SVM model predicts leukemia with deep characteristics more effectively. Table 5 presents a comparative approach that highlights the advantages of utilizing the suggested approach in terms of effectiveness.

Table 5. Comparative analysis of several current research with the proposed model for leukemia prediction.

Author Details	Applied Algorithm	Dataset Size	Obtained Accuracy
Erum Yousef Abbasi et. al. (2024) [14]	ML Models (RF, NB, DT, LR, GB) and DL Models (RNN, FNN)	Large	ML achieve = 97% DL achieve = 98%
Althaf Ali A et al. (2023) [15]	ANN	Large	92.1%
A. Khuzaim Alzahrani et al. (2023) [16]	UNET	Large	97.82%
W. Rahman et. al. (2023) [17]	CNN (RESNET 50)	Medium	99.84%
Almadhor et. al. (2022) [18]	Single learning model, ensemble model and pre-trained CNN model	Large	90%
Tawfeeq Shawly, Ahmed A. Alsheikhy (2022) [19]	CNN with AlexNet	Large	98%
Zhou et. al. (2021) [20]	CNN (Clinical data)	Large	82.93%
A. Sanam, et. al. (2023) [21]	Fuzzy deep neural network	Large	98.8%
Sampathila et al. (2022) [22]	CNN with ALLNET	Large	95.54%
Proposed Model (Hybrid Approach)	**K-means clustering for Detection and Resnet50 for feature extraction with Multi-Class Classification using SVM**	**Large**	**99.98%**

The following real-world cases can benefit from using this automated decision-making method, depending on the predicted outcome:

1. This technique can be useful in the healthcare field since it allows doctors to effortlessly and precisely diagnose leukemia in its early stages. As a result, it offers a trustworthy option for beginning early treatment and reducing the severity of mortality situations.
2. To identify threatening non-blood cells, the color clustering strategy for monitoring is a very trustworthy and effective technique. This automated technique aids with analyzing the development of these non-blood cells so that the seriousness of this illness can be alerted upon reaching a benchmark.
3. The hybrid approach for determining the kind of acute leukemia also presents a successful and efficient method. Unlike the typical ML paradigm, this method integrated both ML and DL strategies. The DL method extracts the deepest characteristics from the images. This hybrid strategy has this as one of its main advantages. These in-depth features allowed the ML classification model to outperform any single learning methods in terms of effectiveness.

4. It has been found that the Resnet50 strategy for deep feature extraction and the SVM strategy for machine learning surpass all other hybrid models in terms of performance. This combined strategy offers accuracy levels exceeding 99%.

Despite achieving above 99% accuracy, this automated identification and prediction approach still has several boundaries:

- RGB picture format should be used for the input variable. Failure to do so will result in improper cluster operation.
- The quality of the picture must be high; otherwise, deep feature collection and detection will be impacted.
- The dataset should have a sufficient number of photos to enable improved predictions.
- To attain dependable effectiveness in classification, machine learning models necessitate substantial quantities of better-quality data. Thus, it is crucial to collect data for ALL. Biased or inadequate inputs might cause problems with adaptation and lower the desired efficiency of the models. In this article, a combined strategy is applied. The deep learning model applied is to the extraction of deep characteristics. So, the feature values also impact the efficiency of classification if the quality of the image degrades.

7. Conclusions

A form of blood cancer called leukemia produces a huge amount of abnormal blood cells and typically starts inside the bone marrow. As of today, four different forms of leukemia have been identified. One of the regular types of leukemia is ALL. Due to the rapid growth and build-up of malignant cells in ALL, timely medical attention is necessary. The traditional procedure, which includes blood test examination, genealogy research, and frequent medication, requires an extensive amount of time, and the results may not always be good. When a sickness was not properly diagnosed, it could potentially spread so quickly that it reached a very dangerous level. Artificial intelligence and machine learning are particularly helpful for resolving these challenges. This automated procedure is incredibly efficient and effective at both identifying non-blood cells and predicting the nature of diseases. This work uses a color grouping method to identify non-blood cells from digital blood photos. This method separates the darkish non-blood areas from the bloodstream and aids with counting the non-blood cells. Furthermore, the three most widely used ML models are used to estimate the category of non-blood cells. The SVM is model shown to have the best level of accuracy. After that, a combined approach is developed to improve prediction performance. In this architecture, the SVM is paired with the Resnet50 deep neural network framework and acquires accuracy levels exceeding 99%.

In the foreseeable future, a customized treatment plan must be developed in order to comprehend ALL. The evaluation can consider a range of private details, such as genomic characteristics, clinical criteria, and response to treatment statistics, in order to produce a prognosis of treatment outcomes that is specific to each patient. The combination of artificial intelligence (AI) and machine learning (ML) systems aids the prediction of the most beneficial remedies for each particular ailment. As the AI and ML system is the most beneficial in this prediction aspect, real-time system tracking is one of the most effective ways to manage ALL. It is essential to regularly monitor a patient's status and how they are reacting to treatment. Subsequent investigations could focus on developing automated monitoring systems that provide immediate information to medical professionals on the statuses of patients. This would allow the medical professionals to treat it swiftly before the illness progresses, as it has been noted to worsen at a fast pace.

Author Contributions: Conceptualization, P.B. and S.B.; methodology, P.B.; software, P.B.; validation, P.B.; formal analysis, P.B.; investigation, P.B.; resources, P.B.; data curation, P.B. and S.B.; writing—original draft preparation, P.B.; writing—review and editing, S.B.; visualization, P.B. and S.B.; supervision, S.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are publicly available at https://www.kaggle.com/datasets/mehradaria/leukemia (accessed on 26 April 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Sacks, M.S.; Seeman, I. A statistical study of mortality from Leukemia. *Blood* **1947**, *2*, 1–14. [CrossRef]
2. Vakiti, A.; Reynolds, S.B.; Mewawalla, P.; Acute Myeloid Leukemia. StatPearls—NCBI Bookshelf. Available online: https://www.ncbi.nlm.nih.gov/books/NBK507875/ (accessed on 27 April 2024).
3. Eden, R.E.; Coviello, J.M.; Chronic Myelogenous Leukemia. StatPearls—NCBI Bookshelf. Available online: https://www.ncbi.nlm.nih.gov/books/NBK531459/ (accessed on 16 January 2023).
4. Puckett, Y.; Chan, O.; Acute Lymphocytic Leukemia. StatPearls—NCBI Bookshelf. Available online: https://www.ncbi.nlm.nih.gov/books/NBK459149/ (accessed on 26 August 2023).
5. Mukkamalla SK, R.; Taneja, A.; Malipeddi, D.; Master, S.R.; Chronic Lymphocytic Leukemia. StatPearls—NCBI Bookshelf. Available online: https://www.ncbi.nlm.nih.gov/books/NBK470433/ (accessed on 7 March 2023).
6. Huang, J.; Chan, S.C.; Ngai, C.H.; Lok, V.; Zhang, L.; Lucero-Prisno, D.E., III; Xu, W.; Zheng, Z.-J.; Elcarte, E.; Withers, M.; et al. Disease Burden, Risk Factors, and Trends of Leukaemia: A Global Analysis. *Front. Oncol.* **2022**, *12*, 904292. [CrossRef]
7. Balta, B.; Gebreyohannis, T.; Tachbele, E. Survival and predictors of mortality among acute leukemia patients on follow-up in Tikur Anbessa Specialized Hospital, Addis Ababa, Ethiopia: A 5-year retrospective cohort study. *Cancer Rep.* **2023**, *6*, e1890. [CrossRef]
8. Du, M.; Chen, W.; Liu, K.; Wang, L.; Hu, Y.; Mao, Y.; Sun, X.; Luo, Y.; Shi, J.; Shao, K.; et al. The global burden of leukemia and its attributable factors in 204 countries and territories: Findings from the global burden of disease 2019 study and projections to 2030. *J. Oncol.* **2022**, *2022*, 1612702. [CrossRef]
9. Acute Lymphocytic Leukemia—Cancer Stat Facts. SEER. Available online: https://seer.cancer.gov/statfacts/html/alyl.html (accessed on 4 June 2024).
10. Facts 2022–2023. Available online: https://www.lls.org/sites/default/files/2023-08/PS80_Facts_2022_2023.pdf (accessed on 4 June 2024).
11. Dong, Y.; Shi, O.; Zeng, Q.; Lu, X.; Wang, W.; Li, Y.; Wang, Q. Leukemia incidence trends at the global, regional, and national level between 1990 and 2017. *Exp. Hematol. Oncol.* **2020**, *9*, 1–11. [CrossRef]
12. Rupapara, V.; Rustam, F.; Aljedaani, W.; Shahzad, H.F.; Lee, E.; Ashraf, I. Blood cancer prediction using leukemia microarray gene data and hybrid logistic vector trees model. *Sci. Rep.* **2022**, *12*, 1000. [CrossRef]
13. Dese, K.; Raj, H.; Ayana, G.; Yemane, T.; Adissu, W.; Krishnamoorthy, J.; Kwa, T. Accurate Machine-Learning-Based classification of Leukemia from Blood Smear Images. *Clin. Lymphoma Myeloma Leuk.* **2021**, *21*, e903–e914. [CrossRef]
14. Patil Babaso, S.; Mishra, S.K.; Junnarkar, A. Leukemia Diagnosis Based on Machine Learning Algorithms. In Proceedings of the 2020 IEEE International Conference for Innovation in Technology, INOCON 2020, Bangaluru, India, 6–8 November 2020; pp. 1–5. [CrossRef]
15. Salah, H.T.; Muhsen, I.N.; Salama, M.E.; Owaidah, T.; Hashmi, S.K. Machine learning applications in the diagnosis of leukemia: Current trends and future directions. *Int. J. Lab. Hematol.* **2019**, *41*, 717–725. [CrossRef]
16. Dharani, T.; Hariprasath, S. Diagnosis of Leukemia and its types Using Digital Image Processing Techniques. In Proceedings of the 3rd International Conference on Communication and Electronics Systems, ICCES 2018, Coimbatore, India, 15–16 October 2018; pp. 275–279. [CrossRef]
17. Ratley, A.; Minj, J.; Patre, P. Leukemia disease detection and classification using machine learning approaches: A review. In Proceedings of the 2020 1st International Conference on Power, Control and Computing Technologies, ICPC2T 2020, Raipur, India, 3–5 January 2020; pp. 161–165. [CrossRef]
18. Ghaderzadeh, M.; Asadi, F.; Hosseini, A.; Bashash, D.; Abolghasemi, H.; Roshanpour, A. Machine Learning in Detection and Classification of Leukemia Using Smear Blood Images: A Systematic Review. *Sci. Program.* **2021**, *2021*, 9933481. [CrossRef]
19. Rehman, A.; Abbas, N.; Saba, T.; Rahman SI ur Mehmood, Z.; Kolivand, H. Classification of acute lymphoblastic leukemia using deep learning. *Microsc. Res. Tech.* **2021**, *81*, 1310–1317. [CrossRef]
20. Baig, R.; Rehman, A.; Almuhaimeed, A.; Alzahrani, A.; Rauf, H.T. Detecting Malignant Leukemia Cells Using Microscopic Blood Smear Images: A Deep Learning Approach. *Appl. Sci.* **2022**, *12*, 6317. [CrossRef]
21. Genovese, A.; Hosseini, M.S.; Piuri, V.; Plataniotis, K.N.; Scotti, F. Acute Lymphoblastic Leukemia Detection Based on Adaptive Unsharpening and Deep Learning. In Proceedings of the ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 1205–1209. [CrossRef]
22. Elhassan, T.A.; Rahim, M.S.M.; Swee, T.T.; Hashim, S.Z.M.; Aljurf, M. Feature extraction of white blood cells using CMYK-moment localization and deep learning in acute myeloid leukemia blood smear microscopic images. *IEEE Access* **2022**, *10*, 16577–16591. [CrossRef]
23. Bukhari, M.; Yasmin, S.; Sammad, S.; Abd El-Latif, A.A. A Deep Learning Framework for Leukemia Cancer Detection in Microscopic Blood Samples Using Squeeze and Excitation Learning. *Math. Probl. Eng.* **2022**, *2022*, 2801227. [CrossRef]

24. Eckardt, J.N.; Middeke, J.M.; Riechert, S.; Schmittmann, T.; Sulaiman, A.S.; Kramer, M.; Sockel, K.; Kroschinsky, F.; Schuler, U.; Schetelig, J.; et al. Deep learning detects acute myeloid leukemia and predicts NPM1 mutation status from bone marrow smears. *Leukemia* **2022**, *36*, 111–118. [CrossRef]
25. Abbasi, E.Y.; Deng, Z.; Ali, Q.; Khan, A.; Shaikh, A.; Reshan MS, A.; Sulaiman, A.; Alshahrani, H. A machine learning and deep learning-based integrated multi-omics technique for leukemia prediction. *Heliyon* **2024**, *10*, e25369. [CrossRef]
26. A, A.A.; Hemalatha, K.; Priya, N.M.; Aswath, S.; Jaiswal, S. An Enhanced Analysis of Blood Cancer Prediction Using ANN Sensor-Based Model. *Eng. Proc.* **2023**, *59*, 65. [CrossRef]
27. Alzahrani, A.K.; Alsheikhy, A.A.; Shawly, T.; Azzahrani, A.; Said, Y. A novel deep learning segmentation and classification Framework for leukemia diagnosis. *Algorithms* **2023**, *16*, 556. [CrossRef]
28. Rahman, W.; Faruque MG, G.; Roksana, K.; Sadi AH, M.S.; Rahman, M.M.; Azad, M.M. Multiclass blood cancer classification using deep CNN with optimized features. *Array* **2023**, *18*, 100292. [CrossRef]
29. Almadhor, A.; Sattar, U.; Al Hejaili, A.; Ghulam Mohammad, U.; Tariq, U.; Ben Chikha, H. An efficient computer vision-based approach for acute lymphoblastic leukemia prediction. *Front. Comput. Neurosci.* **2022**, *16*, 1083649. [CrossRef]
30. Shawly, T.; Alsheikhy, A.A. Biomedical Diagnosis of Leukemia Using a Deep Learner Classifier. *Comput. Intell. Neurosci.* **2022**, *2022*, 1568375. [CrossRef]
31. Zhou, M.; Wu, K.; Yu, L.; Xu, M.; Yang, J.; Shen, Q.; Liu, B.; Shi, L.; Wu, S.; Dong, B.; et al. Development and Evaluation of a Leukemia Diagnosis System Using Deep Learning in Real Clinical Scenarios. *Front. Pediatr.* **2021**, *9*, 693676. [CrossRef]
32. Ansari, S.; Navin, A.H.; Sangar, A.B.; Gharamaleki, J.V.; Daneshvar, S. Acute Leukemia Diagnosis Based on Images of Lymphocytes and Monocytes Using Type-II Fuzzy Deep Network. *Lectronics* **2023**, *12*, 1116. [CrossRef]
33. Sampathila, N.; Chadaga, K.; Goswami, N.; Chadaga, R.P.; Pandya, M.; Prabhu, S.; Bairy, M.G.; Katta, S.S.; Bhat, D.; Upadya, S.P. Customized Deep Learning Classifier for Detection of Acute Lymphoblastic Leukemia Using Blood Smear Images. *Healthcare* **2022**, *10*, 1812. [CrossRef]
34. Maurya, L.; Lohchab, V.; Kumar Mahapatra, P.; Abonyi, J. Contrast and brightness balance in image enhancement using Cuckoo Search-optimized image fusion. *J. King Saud Univ.—Comput. Inf. Sci.* **2022**, *34*, 7247–7258. [CrossRef]
35. Hussien, R.M.; Al-Jubouri, K.Q.; Gburi MR a, A.; Qahtan AG, H.; Jaafar AH, D. Computer Vision and Image Processing the Challenges and Opportunities for new technologies approach: A paper review. *J. Phys.* **2021**, *1973*, 012002. [CrossRef]
36. Abdulateef, S.K.; Salman, M.D. A Comprehensive Review of Image Segmentation Techniques. *Al-Mağallat Al-'irāqiyyat Al-Handasat Al-Kahrabā'iyyat Wa-Al-Ilikttrūniyyat* **2021**, *17*, 166–175. [CrossRef]
37. Sundaram, A.; Sakthivel, C. Object detection and estimation: A hybrid image segmentation technique using convolutional neural network model. *Concurr. Comput. Pract. Exp.* **2022**, *34*, e7114. [CrossRef]
38. Niu, Z.; Li, H. Research and analysis of threshold segmentation algorithms in image processing. *J. Phys.* **2019**, *1237*, 022122. [CrossRef]
39. Suresh, S.; Newton, D.; Everett, T.H.; Lin, G.; Duerstock, B.S. Feature Selection Techniques for a Machine Learning Model to Detect Autonomic Dysreflexia. *Front. Neuroinform.* **2022**, *16*, 901428. [CrossRef]
40. Patwary MJ, A.; Parvin, S.; Akter, S. Significant HOG-Histogram of Oriented Gradient Feature Selection for Human Detection. *Int. J. Comput. Appl.* **2015**, *132*, 17. [CrossRef]
41. Ray, S.D. A Quick Review of Machine Learning Algorithms. In Proceedings of the International Conference Machine Learning, Big Data, Cloud and Parallel Computing 2019, Faridabad, India, 14–16 February 2019. [CrossRef]
42. Hearst, M.A.; Dumais, S.T.; Osman, E.; Platt, J.; Schölkopf, B. Support vector machines. *IEEE Intell. Syst. Their Appl.* **1998**, *13*, 18–28. [CrossRef]
43. Peterson, L.E. K-nearest neighbor. *Scholarpedia* **2009**, *4*, 1883. [CrossRef]
44. Rish, I. An empirical study of the naive Bayes classifier. In Proceedings of the IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, Seattle, WA, USA, 4–10 August 2001.
45. Miškovic, V. Machine Learning of Hybrid Classification Models for Decision Support. In Proceedings of the Sinteza 2014—Impact of the Internet on Business Activities in Serbia and Worldwide, Belgrade, Serbia, 25–26 April 2014. [CrossRef]
46. Cheng, H.; Jiang, X.; Ma, J.; Wang, J. Color image segmentation: Advances and prospects. *Pattern Recognit.* **2001**, *34*, 2259–2281. [CrossRef]
47. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**. [CrossRef]
48. Acute Lymphoblastic Leukemia (ALL) Image Dataset. Kaggle. Available online: https://www.kaggle.com/datasets/mehradaria/leukemia (accessed on 30 April 2021).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.