Journal of Computer Science

Research Paper

Hematologic Cancer Cell Detection and Classification Using Optimized VGG-19 with Stratified K-fold Cross-Validation

Hema Patel[1], Himal Shah[2], Gayatri Patel[3], Atul Patel[1]

[1]*Smt. Chandaben Mohanbhai Patel Institute of Computer Applications, Charotar University of Science and Technology, CHARUSAT - Campus, Anand, India*
[2]*QURE Haematology Centre, Ahmedabad, India*
[3]*Ramanbhai Patel College of Pharmacy, Charotar University of Science and Technology, CHARUSAT - Campus, Anand, India*

Article history
Received: 14 May 2025
Revised: 5 June 2025
Accepted: 16 June 2025

*Corresponding Author:
Hema Patel,
Smt. Chandaben Mohanbhai Patel Institute of Computer Applications, Charotar University of Science and Technology,
CHARUSAT - Campus,
Anand, India
Email: hemapatel.mca@charusat.ac.in

Abstract: The most prevalent pediatric blood malignancy is acute lymphoblastic leukemia (ALL). ALL is a lethal disease in which patients have a lower survival rate. Its prompt detection and precise categorization are essential for successful treatment. Manual microscopic diagnosis is laborious, prone to mistakes, and heavily reliant on specialized knowledge. With stratified 7-fold cross-validation, which assures an equal ratio of normal and malignant cells in each fold, this research provides an enhanced VGG-19-based deep learning model for reliable binary categorization of leukemic vs normal cells to overcome the drawbacks of the manual detection procedure. For this study, the C-NMC leukemia dataset used comprises single-cell images of normal (HEM) and cancerous (ALL) types. For the binary classification test, transfer learning was utilized by keeping the initial convolutional layers of the pre-trained VGG-19 model and swapping out its last few layers. This resulted in an accuracy of 98.87%, a sensitivity of 98.97%, 98.82% specificity, 99.68% precision, and an F1-Score of 99.23%. The outcomes demonstrate how well the model handles morphological differences and class imbalance in leukemic cell images. In addition, the proposed model also outperformed the other pre-trained neural networks, viz., ResNet-18, ShuffleNet, and GoogleNet, in terms of accuracy.

Keywords: Leukemia, VGG-19, Stratified K-fold Cross-Validation, Acute Lymphoblastic Leukemia, Deep Learning, Detection

Introduction

A hematological illness, leukemia is defined by a rapid expansion of aberrant white blood cells. According to the National Cancer Institute, an estimated 66,890 new instances of cancer disease are expected to be diagnosed in 2025, making up roughly 3.3% of all new cases of cancer. Furthermore, it is anticipated that this form of cancer will be responsible for roughly 23,540 deaths, or 3.8% of all cancer-related deaths, as per the National Cancer Institute, USA. It can be recognized manually and using standard techniques. The manual method typically takes a long time, and the quality of the camera, the pathologists' experience, fatigue, and many other factors can sometimes change the results. In order to address these problems, digital microscopic analysis of blood smear images utilizing deep learning (DL) techniques has made it possible for data to be automatically identified and analyzed (Fatichah et al., 2012). Compared to manual alternatives, these methods are more dependable, quick, economical, and effective.

Leukemia is mostly caused by a lack of leukocyte production, which is also known as white blood cells. Two distinct types of WBCs, Lymphocytes and myeloid, are distinguished by the manifestation of their cytoplasm (Glenn et al., 2019) and are responsible for Lymphocytic Leukemia and Myeloid Leukemia.

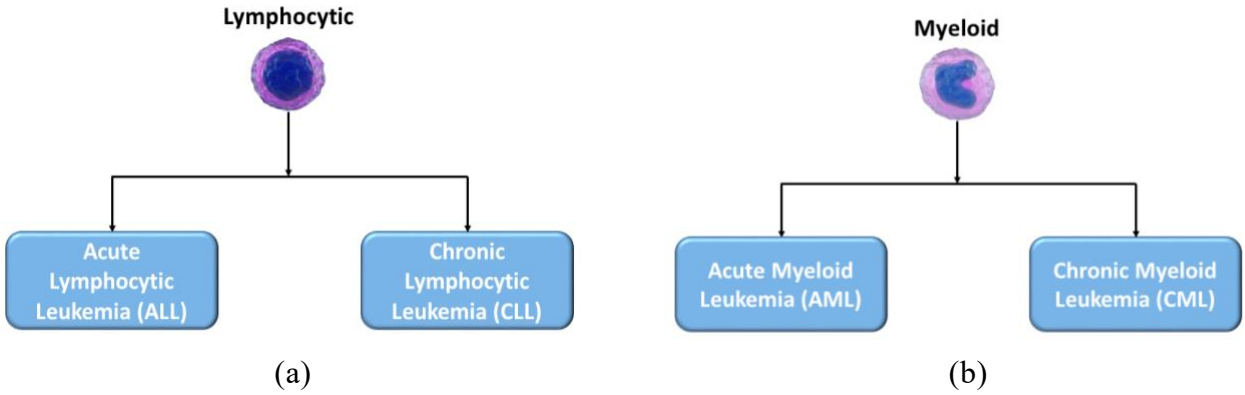


Fig. 1. Types of Leukemia

Table 1. The Characteristics of Various Leukemia Types

Characteristics	ALL	AML	CLL	CML
Cause	Due to neoplastic and abnormal development of B cells or T cells	The formation of adolescent monocyte cells in the blood	The rise in the progression of B-cell lymphocytes	Population of granulocytic exhibits because of the genetic mutation in the bone marrow's DNA
Variants	L1, L2 and L3 (FAB) (French American British) OR B-Cell and T-Cell WHO (World Health Organization)	M0, M1, M2, M3, M4, M4eo, M5a, M5b, M6, and M7	B-cell and T-cell CLL and prolymphocytic leukemia	PHASES: chronic and accelerated
Impacted People	Children under the age of 5 and adults above 50	People around 68 or above	Most often diagnosed in older adults (an average age of 70), and extremely rare among individuals under 40	The highest incidence is in adults, especially those aged 40 to 60
Statistical Details	Among children and people under the age of 20, this kind of leukemia is common, making up 75% of total leukemia cases	Accounts for about 1% of all cancers	It accounts for 25% of newly diagnosed leukemia instances and is mainly found in individuals over the age of 19, making up 38% of all leukemia diagnoses	Around 9560 (15% of) new instances detected yearly
Survival Rate (around 5-years)	About 65% individuals can survive with this for five years	People below 65 years have around 25% chances whereas above 65 years have only 2% chances	Patients with CLL has a 5- year relative survival rate of more than 92%	90% chances to survive for 5-years after diagnosis

Lymphocytic and myeloid leukemia can be further subcategorized as acute and chronic. In acute leukemia, the abnormal cells grow rapidly, whereas in chronic leukemia, the growth rate of malignant cells slows down (Mittal et al., 2022). So, on the basis of these two factors, there are basically four types of leukemia, namely, Acute Lymphocytic Leukemia (ALL), Acute Myeloid Leukemia (AML), Chronic Lymphocytic Leukemia (CLL) and Chronic Myeloid Leukemia (CML) as shown in Fig. (1). Fig. (1) (a) presents two types of lymphocytic leukemia, and Fig. (1) (b) shows two types of myeloid leukemia. Characteristics of each leukemia category are presented in Table (1).

Acute Lymphocytic Leukemia (ALL), which is also identified as lymphoblastic leukemia or lymphoid leukemia. With over 25% of pediatric cancer cases being Acute Lymphoblastic Leukemia (ALL). Children under the age of five have the highest risk of developing ALL (Patel et al., 2024). The risk gradually drops until the mid-20s and then gradually rises once more near the age of 50. Approximately four out of ten ALL occurrences occur in adults. The severity can be judged by the statistics which state that each year, around 6,100 new instances of ALL are diagnosed, with approximately 3,450 cases occurring in the male category and 2,650 cases in the female category. About 1,400 individuals die with ALL each year, with 720 of those fatalities occurring in men and 680 in women, in spite of advancements in therapy according to the American Cancer Society. Timely therapy greatly increases the prospects of recovery and long-term survival; therefore, early and accurate diagnosis is essential for ALL. Timely therapy greatly increases the prospects of recovery and long-term survival; therefore, early and accurate diagnosis is essential for ALL.

Fever, exhaustion, joint pain, bruises, infections, rapid weight loss, swelling of the lymph nodes, patches on the skin, breathing difficulties, bone soreness, and paleness are a few common symptoms of leukemia. However, these symptoms can also be signs of other physical disorders. The subtle nature of these indicators makes it challenging to diagnose leukemia in its initial phase, which becomes a challenge for hematologists and medical professionals.

Analysis of Peripheral Blood Smear (PBS) images with the aid of a microscope is one of the most popular and less expensive procedures being used by hematopathologists to identify whether the cells are benign or leukemic. Experts may employ other tests to identify leukemia, such as Complete Blood Count (CBC), Flow Cytometry, biopsy of bone marrow sample, lumbar puncture, and Fluorescence in Situ Hybridization (FISH), etc.

Manual diagnostic procedures for leukemia present several significant concerns. First, the subjective nature of these methods can lead to inconsistent results, the same blood sample may yield different findings when assessed by different physicians, particularly if they have been trained under varying assessment methodologies. Moreover, diagnostic accuracy can be compromised by human factors such as fatigue, illness, or other cognitive stressors, which may hinder a physician's ability to precisely identify leukemia subtypes.

In addition to reliability issues, manual techniques are often inefficient and costly. The evaluation process tends to be time-consuming, delaying diagnosis and treatment planning. Furthermore, many of the required laboratory tests are prohibitively expensive, which can limit access to timely and accurate diagnostics, particularly in resource-constrained settings.

Nowadays, researchers are involved in the automation of computer-aided leukemia detection utilizing DL algorithms. Because deep neural networks are capable of extracting intricate structures from data, they are now well-suited for image classification and segmentation applications. Automated machine learning (ML) and deep learning-based models can quickly and accurately identify the abnormality in the specific type of cells with an accurate number of blasts, which can be helpful to hematopathologists in the detection and classification of leukemic cells, which is necessary for the proper treatment and, in this way, contribute to saving human lives.

This study utilized single-cell blood smear images. The proposed model was developed using an optimized VGG-19 architecture, trained and evaluated with stratified 7-fold cross-validation. The methodology included essential preprocessing steps including image resizing, normalization, augmentation, and cleaning. The performance of the proposed model was benchmarked against three other pre-trained convolutional neural networks: ResNet-18, ShuffleNet, and GoogleNet.

Literature Review

Several research works have been thoroughly reviewed in this section to identify leukemia and its subtypes. These studies have examined blood smear images using various ML and DL techniques.

Kadhim et al. (2023) suggested a convolutional neural network (CNN)-based model for the detection of leukemia. They have used ALL-IDB1 dataset. The model achieved an accuracy of 98% to differentiate malignant cells from normal cells. The proposed model categorized cancerous cells among “multiple myeloma (MM), acute myeloid leukemia (AML), and acute lymphoblastic leukemia (ALL) cells”. Baig et al. (2022) have used two CNN networks, CNN-1 and CNN-2. Researchers have used the dataset of 4150 images. The bagging ensemble classification technique was employed and achieved 97.4% accuracy. Similarly, Kalaiselvi et al. (2020) employed a model. The model was based on Modified Convolution Neural Networks (CNNs). The modified CNN was used to optimize the categorization process of leukemia into four categories AML, ALL, CML and CLL. The model utilized the dataset of around 10,000 blood smear images and obtained 98% accuracy. Setiawan et al. (2018) employed a support vector machine classifier to accurately identify M4, M5, and M7. The model used the dataset of 1710 images. Arif et al. (2022) have also implemented CNN for the identification of leukemic cells. The researchers used two datasets- ALL-IDB1 and ALL-IDB2, and obtained 98.05% accuracy. Likewise, Sahlol et al. (2020), implemented the VGGNet and acquired 87.90% accuracy for the classification of leukemia.

Wang et al. (2019) implemented two leukocyte recognition methods, “Single Shot Multibox Detector and an Incremental Improvement Version of You Only Look Once”, and obtained “the best mean average precision of 93.10% and mean accuracy of 90.09%”. Park et al. (2024) suggested a DL-based model. This model could categorize blood smear images of 12 distinct cell types. 42,386 single-cell blood smear images were used by the model. This model obtained an accuracy of 87.79% with EfficientNet-V2 (B2) and “Chrono-SCA-ACNN (Chrono-Sine Cosine Algorithm-based actor-critic neural network)”. Subsequently, the leukemia was detected by applying the chosen features to the suggested classifier on the ALL-IDB2 database (Jha & Dutta, 2020). This approach obtained an accuracy of 99%. A DL framework for differentiating between healthy and malignant cells was presented by Mourya et al. (2018). The name of the model was Leukonet. Leukonet was a hybrid architecture that combines CNN with “discrete cosine transform (DCT)” characteristics. In order to assess Leukonet, 312 cancer cell images and 324 healthy cell images were included in the test results. This obtained an accuracy of 89.70% for the cancer cell class. Arivuselvam et al. (2022) proposed a model, which is based on CNN-based Deep Learning (DCNN), for identifying leukemia categories. In this work, DCNN worked as a feature extractor. The researcher employed "adaptive gamma correction based on histogram" equalisation, along with assisted filtering, to analyse the enhancement in intensity while maintaining the important details of the images.

An effective technique for the detection of CLL was created by (Sivalingam et al. 2022). In this model, fuzzy and cuckoo search-based filters were used for preprocessing the images to remove noise and artefacts. To create the segments from the pre-processed blood smear images, deep fuzzy clustering was used for cell segmentation. Using Random Multimodal Deep Learning (RMDL), CLL was discovered. Using the suggested Jaya-Competitive Swarm Optimisation (Jaya-CSO) algorithm, RMDL training was carried out. The study proposed by Surya et al. (2021) used the ALL-IDB2 dataset for the detection of ALL. Following image preprocessing, AlexNet was used for both feature extraction and classification. Transfer learning was utilised by Vogado et al. (2018) to extract features from images for categorisation. The characteristics chosen for the Support Vector Machine classifier were based on their gain ratios, and the authors investigated three CNN frameworks. The suggested model did not require a segmentation process and categorised images with features generated from different image datasets.

Narayanan et al. (2025) implemented a hybrid classifier on the dataset of 8637 to identify and classify ALL. Ouyang et al. (2021) suggested a CNN for object detection and instance segmentation. It mixes a pre-trained model with a model that was created from scratch. The selection of instance segmentation was based on its ability to tolerate a certain amount of cell overlap and its average precision of 62.5% and recall of 84.1%. In order to concurrently collect global and local attributes, Mathur et al. (2020) introduced a unique architecture called the Mix-up Multi-Attention Multi-Task Learning Model (MMA-MTL). Similarly, the two-stage artificial neural networks along with the Particle Swarm Optimisation approach were proposed by (Agustin et al., 2021). The purpose of the model was to categorise the immature WBCs in patients with ALL. Binary classification of lymphoid cells in the first stage and binary classification of lymphoblast cells in the second stage were both included in the suggested method and achieved accuracy of 86.92%. Acharya et al. (2023) proposed a model to classify subtypes of AML. Authors used Random Forest (RF)

classifier to categorize AML into subtypes. The morphologic characteristics of leukemic blasts, including Auer rods, and cytoplasm, were retrieved. Two datasets - ALL-IDB and Cell Vision were combined and used by (Sipes & Li, 2018). Researchers developed a three-layer CNN by adding a convolutional layer, ReLU, and pooling layers to the two-layer NN-1 architecture for the classification.

This information provides a strong basis for recommending advanced CNN-based models for the detection and classification of leukemia, which have better generalization, greater accuracy, and practical application.

Materials and Methods

Proposed Method

The suggested model has been developed using optimized VGG-19 framework. The purpose of the proposed method is to classify single-cell blood smear images of the CNMC Leukemia dataset as either healthy (HEM) or cancerous (ALL) cells. The suggested ensemble model automatically learns the features needed for image classification during training by analyzing the images and then classifies the images as depicted in Fig. (2).

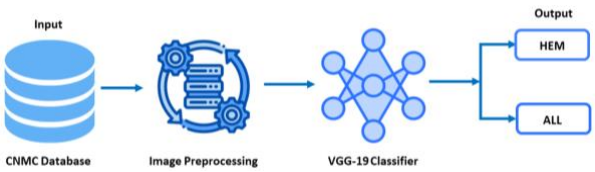


Fig. 2: Flow of Proposed Model

Data Repository

Image acquisition is the initial phase of image processing. This entails acquiring the unprocessed data in the image format, which serves as the basis for all further processing stages. For this research, we have used the C-NMC Leukemia dataset of blood smear images that comprises a total of 15114 images of basically two types: normal (HEM) and cancerous (ALL), available on Kaggle's website. These single cell blood smear images were obtained from total 118 individuals and categorised into three distinct folders called "C-NMC training data" (10661 cells, 7272 cancerous cell images from 47 patients, and 3389 benign single cell images from exact 26 individuals); "CNMC test preliminary phase data" (1867 images, 1219 malignant cell images from 13 patients, and 648 healthy cell images from 15 individuals); and "C-NMC test final phase data" (2586 images from 17 individuals) as shown in Fig. (3).

The proposed model is trained on an unlabeled dataset. The training, validation, and testing datasets of the CNMC Leukemia dataset are imbalanced. In this research, a training dataset comprising 10661 cell images was used. Out of which, 3389 are normal cell images, whereas 7272 are malignant images. The division of both types of datasets was as follows: 70% for training, 15% for validation, and 15% for testing. Thus, for the normal (Hem) category, the model used 2373 cell images for training, 508 images for validation and testing purposes, whereas 5090 cancerous cell images were used to train the proposed model, and 1091 cell leukemic images were used to test and validate the suggested model, as shown in Fig. 4.

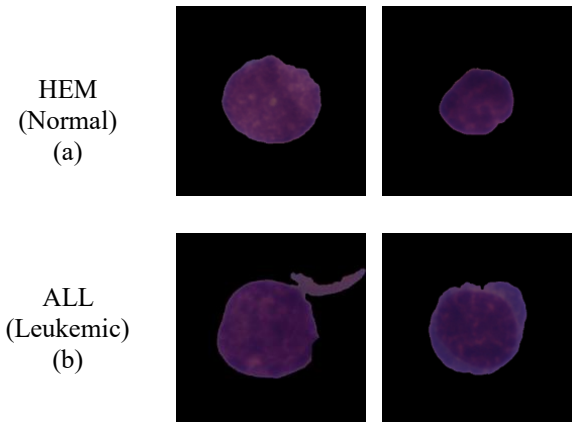


Fig. 3: CNMC Leukemia Dataset Images

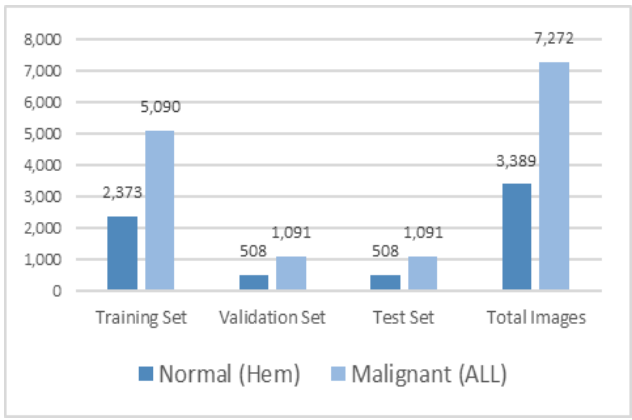


Fig. 4: Division of Dataset

Image Preprocessing Phase

Preparing raw images for further evaluation is known as preprocessing. In order to make it simpler to extract pertinent features and carry out precise detection, this

phase attempts to improve image quality, eliminate noise, and normalize various image properties.

Resizing Images

To ensure uniformity in later processing stages, resize images to a standard size. Also, Images are commonly represented in deep learning as matrices of pixel data. Larger images might cause a bottleneck when processing with limited computational capabilities because they take more memory to store and process. An image's memory footprint can be decreased by resizing it to a smaller size, which facilitates manipulation. The image size of the training and validation dataset folders is 450 x 450 pixels, and the size of images for the testing data directory is 600 x 600 pixels. The size of the training dataset images from the training folder is resized to 224 x 224 x 3 pixels.

Normalization

Data has been normalized at the second step, in which image pixel values are frequently centered to have a zero mean or normalized between 0 and 1 to ensure quicker convergence. This enhances the model's performance and reliability because every significant feature makes an equal contribution to the network's learning process. We have employed Z-score normalization and Eq. 1 to represent this as follows:

$$X_n = \frac{x-\mu}{\sigma} \quad (1)$$

X_n represents the normalized form of the data, which can be calculated by computing the standard deviation σ and mean μ of the PBS images. Then, centers the data around zero by subtracting the average of all the image's pixel values and eventually computes the data's variance by dividing it by the standard deviation of each pixel value.

Image Augmentation

As deep learning is capable of processing ample data, some operations like rotation, scaling, and flipping are employed to produce additional images from the ones already present. This could contribute to the expansion and diversification of the dataset. This would eventually enhance the model's overall efficiency. This phase includes some methods, such as:

a. *Cropping and Scaling:* Creating numerous renditions of an image by cropping a portion of it or altering its size.

b. *Flipping:* To create new versions, flip an image in either a horizontal or vertical plane.

c. *Rotation:* Rotation is the method of turning an image at a particular angle to create new representations of the source image.

d. *Adjustment of Brightness:* It adjusts the brightness of the images. With different intensities of colours and brightness, a new version of images can be created.

e. *Shearing:* Shearing is the process of altering an image along a specified axis. By doing so, it creates new copies of the image.

Diminishing Noise

Removing noise and blemishes from the image is an essential process because these may affect the training of the model and eventually alter the outcome. To do so, numerous methods can be used to improve the quality of the image. A few instances of typical methods are as follows:

a. *Sobel Filtering:* It is one of the popular edge detection techniques. It is used to highlight areas where the image's intensity fluctuates or the margins of the image. Furthermore, it is also used to ascertain the gradient of the image intensity at every pixel.

b. *Thresholding:* Thresholding is another effective technique for removing noise from images. It involves setting a threshold value that divides the image into background and areas of interest to reduce noise.

Convolutional Neural Networks

A type of artificial neural network that is utilized for processing and image identification is referred to as a convolutional neural network (CNN). CNNs are designed with the hierarchical organization of the human visual system in mind. CNN is responsible to learn features form the images and for this, CNN uses the concept of dimension reduction. Identifying patterns in images is a significant application of CNNs. By filter optimization, CNNs automatically acquire new features. Regularized weights with fewer connections help to alleviate problems that were common in previous neural networks, such as vanishing and exploding gradients.

A collection of filters is combined with the input image in each layer of a CNN to create a set of activation maps that indicate which features exist in the image. After that, pooling layers minimize the activation maps to extract the most essential details while minimizing the volume of data. The visual representation of the flow of the deep CNN is illustrated in Fig. (5). The proposed model has been developed by the optimized VGG-19 neural network.

VGG-19 Classifier

Early in 2014, Simonyan & Zisserman (2014) from the University of Oxford in the United Kingdom proposed a pre-trained CNN model that is now known as the VGG network.

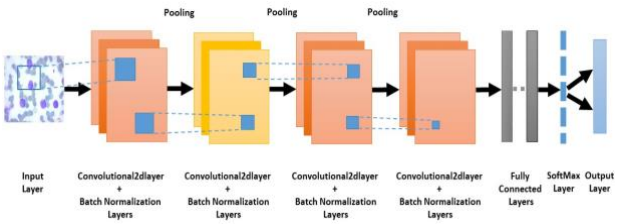


Fig. 5: Architecture of CNN

VGG (Visual Geometry Group) used the ImageNet ILSVRC dataset for training. Improved feature extraction and the utilization of maxpooling to downsample former classification using the SoftMax activation function are made possible by the model's highly linked convolutional and fully-connected layers. Fig. (6) depicts the VGG-19 architecture.

VGG-19 comprises a total of 19 layers, which include 16 convolutional layers, five max pooling layers, which are followed by three fully connected layers and a softmax layer for the classification at the end. All 16 convolutional layers are arranged in five different groups. Each groups are separated by a maxpool layer. When the input image is reduced by two, it is the responsibility of a maxpool layer to increase the number of filters by two in the convolution layer.

The input layer of the VGG-19 architecture receives an image with the dimensions of 224 × 224 × 3. Then, the input image is passed through a number of convolutional layers, each of which applies a number of filters (Sudha & Ganeshbabu, 2021). After this, it is followed by the application of a rectified linear unit (ReLU). ReLU is an activation function that is responsible for providing the non-linearity.

VGG-19 model employs max-pooling layers. These layers follow a sequence of convolutional layers. Max pooling layers are responsible for minimizing the spatial dimensions of the feature maps. Each time the feature map is applied, its size becomes half of the previous one, due to the normal pooling size of 2x2 with a stride of 2. This is known as downsampling. Downsampling makes the network more efficient while maintaining key features. This helps to reduce computational time and also helps to avoid the issue of overfitting. Then, to make the network resilient to spatial alterations to input images, pooling layers lower the variance of translations.

Eventually, the final maxpool layer sends the output to fully connected layers that perform the final classification procedure. Despite its size and computational demands, the model exhibits exceptional performance in image classification.

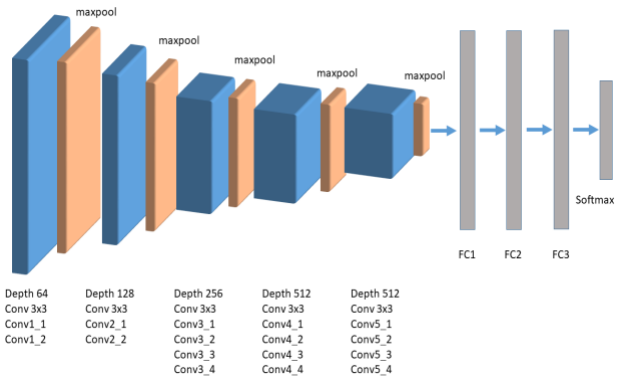


Fig. 6: Architecture of VGG-19

In the VGG-19 neural network, the convolutional layer follows the input layer, which is responsible for the generation of the output feature maps by organizing the input image or feature map with a collection of numerous filters (Ikechukwu et al., 2021). The input image passes through the first convolutional layer in VGG-19 using a 3 × 3 kernel with a stride of 1 and 1 padding with 3 input channels. This can be expressed mathematically by Eq. 2:

$$O(x,y,k) = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{c=1}^{3} I(x+i-2,y+j-2,c) \cdot W(i,j,c,k)+b_k \quad (2)$$

Here, I represent the size of the filter, c indicates the number of input channels, P shows the padding for VGG-19, and k is used to provide the information of the output channel index.

An activation function known as a ReLU (Rectified Linear Unit) is applied after every convolution operation. It can be mathematically denoted as in Eq. 3:

$$f(z) = max\ (0,z) \quad (3)$$

Thus, it stays the same if z is positive. If z is negative, 0 is used in its place.

It is followed by an activation function as shown in Eq. 4:

$$A(x,y,k) = max\ (0,\ O(x,y,k)) \quad (4)$$

For down-sampling, max-pooling is employed. A fully connected layer is used to implement matrix-vector multiplication, as shown in Eq. 5:

$$O(i,j,c) = \max_{0\leq m<k} \max_{0\leq n<k} \{I(s.i+m, s.j+n,c)\} \quad (5)$$

Here, O is the output vector, whereas I is the input vector. k indicates the pooling window size, and s shows the no. of strides.

Softmax layer is employed to execute the final classification, as given in Eq. 6:

$$S(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_i}} \quad (6)$$

Here, k denotes the total number of classes, whereas the output from the last fully connected layer for a particular class i can be denoted by z_i.

Following the application of the softmax function in VGG-19, the output neurons' highest probability value is used to assign a class label as the last stage of classification, as represented in Eq. 7:

$$\hat{y} = arg \max_{i \in C} P_i \quad (7)$$

VGG-19's deep architecture demonstrates effectiveness in analysing medical images, making it a good choice for cancer cell detection and classification, especially when applied to a dataset with two classes: benign and malignant. VGG-19 is capable of discovering complex features from images. Effective feature extraction is achieved through a consistent architecture that includes repetitive blocks of convolutional layers, ReLU (Rectified Linear Unit) activations, and pooling layers. This structure uses small 3×3 convolution filters while ensuring computational efficiency. Under a microscope, distinguishing abnormal leukemic blasts from normal cells can be difficult since the two cells' appearances are identical. In such instances, the depth of VGG-19 facilitates the significant extraction of fine-grained texture, boundary, size, and shape information for cancer detection, where subtle morphological deviations distinguish healthy cells from malignant cells.

Model Training

In this study, the VGG-19 neural network, which is well-known for image classification tasks, has been used. Setting up CNN's architecture is an initial step in developing a model. CNN design is typically composed of a number of different kinds of layers and modules, which are frequently customised to the particular application and data properties. First of all, the input layer retrieved the images with dimensions of 450 x 450 from the database and scaled them to 224 x 244 x 3. This step is crucial because the VGG-19 pre-trained model works on the images with a fixed dimension of 224 x 244 x 3.

The recommended classifier, VGG-19's convenience, has been taken into consideration when determining the image size. It is followed by the normalisation of images using the z-score method to create uniform images, along with an augmentation mechanism to elevate the count of the training dataset in order to boost the model's performance. Afterwards, 7-fold cross-validation was implemented to avoid bias caused by the imbalanced dataset, and eventually, the implementation of optimised, fine-tuned VGG-19 was done to advance feature learning and classification.

In this way, this procedure increases the model's exactness and robustness. Thus, the model suggested for cancer cell detection and classification was optimised through various stages to guarantee robust performance of the proposed model.

Hyperparameters Tuning

To train the model, the proposed approach used a batch size of 32, which implied that 32 images were processed in a single iteration. The number of epochs selected was 50. An epoch is a complete run of the whole training dataset through the neural network. The selected value indicates that the model underwent 50 epochs of training and traced the entire training dataset this many times. Value 50 is an appropriate count for the proposed model to learn about the different patterns of healthy and leukemic cells and then differentiate the cells properly. To regulate the step size for updating weights, an initial learning rate of 0.0001 was chosen. In the proposed model, Stochastic Gradient Descent (SGD) optimizer was employed for optimization purposes. SGD is an iterative optimisation technique that minimises an objective function, usually a loss function. It is an efficient version of the classic Gradient Descent technique. It works effectively, particularly when working with big datasets. In each iteration, it uses a single, randomly chosen data point or mini-batch to approximate the gradient, which leads to frequent updates of model parameters such as weights and biases and lessens computation time.

In DL, the purpose of a loss function is essential to neural network training and optimisation. It functions as a measurable indicator of the difference between the outcome that a model forecasts and the output that is actually obtained. By considering this, the binary cross-entropy loss function was implemented, which compares the probability value 0 and 1 for each class, generated by the classification model. This loss penalizes the anticipated probability y's deviation from Eq. 8, which defines its target probability distribution, x.

$$L = \sum_{i=1}^{n} y_i \log(\hat{y}_i) \quad (8)$$

Where, n is the total number of classes, y_i is the actual label and $\hat{y}_i$ is the class 's predicted probability.

Transfer Learning and Fine-Tuning Strategy

In the context of transfer learning with CNNs, a common and effective training policy involves modifying the final layers of a pre-trained network to adapt it to a new classification task. In this research, the last three layers of the VGG-19 network have been replaced with new layers, namely, a fully connected layer, softmax, and classification layer, specially customized for the classification of two C-NMC Leukemia dataset groups, normal (HEM) and malignant (ALL).

The two parameters WeightLearnRateFactor and BiasLearnRateFactor were set to a higher value of 10 in order to speed up the learning of the newly implemented layers. These variables essentially let the new layers update their weights more quickly than the previously trained layers by acting as multipliers to the base learning rate. Faster convergence and more efficient learning of dataset-specific characteristics are made possible by this modification, which enabled the model to apply a higher effective learning rate to the new layers. Since freezing of earlier layers was not performed in this study, the VGG-19's earlier layers maintained their pre-trained weights, protecting generalised visual features acquired from huge datasets. In this way, the model gains from both past knowledge and task-specific adaptation through the amalgamation of transfer learning and fine-tuning, which enhances classification performance and also allows the model to achieve high accuracy with very little overfitting, even on a small, imbalanced dataset.

Stratified 7-fold cross-validation

A stratified cross-validation strategy has been implemented in the proposed model which utilizes imbalanced classes for normal and ALL cells, because it is helpful in assessing a model's performance, particularly when working with unbalanced datasets. For cross-validation, it entails dividing the dataset into several folds, also known as subsets, and sustaining the initial class distribution in each fold. This indicates that the percentage of samples of each class in each fold is roughly equal to that of the overall dataset. For instance, each stratified cross-validation fold will almost retain the 80-20 proportion in a binary classification task with 80% benign and 20% cancerous samples. Because it enables the model to learn and be fairly assessed across all classes, this method is essential in medical datasets or any other domain where class imbalance arises.

In k-fold cross-validation, the number of folds is chosen by balancing bias, variance, and computational expense. Because each training subset is small compared to the entire dataset, using too few folds (e.g., 3 or 5) can increase bias in performance predictions. On the other hand, an excessive number of folds (e.g., 10 or more) can decrease bias, but they also raise the deviation of the estimate, computational time, and cost because the model needs to be trained and verified several times.

7-fold cross-validation was chosen for this research as an appropriate compromise between these issues. 7-fold offers enough test folds to give a consistent and trustworthy estimate of performance, as well as large enough training subsets in each iteration to enable efficient model learning and minimise bias. 7-fold, as opposed to 3 or 5, lessen the possibility of overfitting to any specific subset and better capture data variability. Additionally, 7-folds cross-validation is less computationally demanding than 10-fold cross-validation, which makes the training process more feasible.

The proportions of normal and cancer cell images remain the same in each fold due to the stratified nature of the dataset. Around 72 normal cell images and 156 malignant images were included in each test set. The average of the outcomes from each of the seven folds yields the final result.

In conclusion, 7-fold offers a solid balance between optimising training data, helps minimize bias brought on by particular data splits, and guarantees equitable, representative evaluation on a variety of subsets in the employed C-NMC leukemia dataset, where data availability is constrained and each sample is important.

Hardware Specification

Training and testing of the suggested approach was carried out on an AI supercomputer, the "NVIDIA DGX Station with four NVIDIA A100 Tensor Core GPUs and 40GB of RAM each" using the software MATLAB 2024a.

Evaluation Metrics

Following training, an analysis of the model's overall performance is conducted on the testing dataset to assess the efficacy of the suggested model. The most common performance parameters, such as accuracy, sensitivity, specificity, precision, and F1-Score, have been used in the evaluation. The statistics required to calculate any suggested system's success rate are provided by the equations below.

$$accuracy = \frac{T_P + T_N}{T_P + F_P + T_N + F_N} \quad (9)$$

$$sensitivity/recall = \frac{T_P}{T_P + F_N} \quad (10)$$

$$specificity = \frac{T_N}{T_N + F_P} \quad (11)$$

$$precision = \frac{T_P}{T_P + F_P} \quad (12)$$

$$f_1 score = \frac{2 \times T_P}{(2 \times T_P) + F_P + F_N} \quad (13)$$

Results and Discussion

Using blood smear images, the proposed model has been developed by combining the concept of stratified 7-fold cross-validation, which shows how robust and reliable it is at differentiating between normal and cancerous cells.

In four out of five assessment metrics, the suggested model outperforms popular pre-trained networks, including ResNet-18, ShuffleNet, and GoogleNet as shown in Table (2).

With a noteworthy accuracy of 98.87%, the suggested model outperforms ResNet-18 (98.75%), ShuffleNet (97.24%), and GoogleNet (98.56%). This shows that because the stratified cross-validation strategy guarantees a balanced class distribution and minimises model bias, the optimised VGG-19 generalises well over a variety of data folds.

With a sensitivity of 98.97%, the recommended approach outperforms ResNet-18 (98.80%), ShuffleNet (97.34%), and GoogleNet (98.25%). This illustrates how well the model can identify samples that are cancerous, whereas the model's competitive specificity of 98.82%, which surpasses that of ResNet-18 (98.62) and ShuffleNet (97.05) but falls short of GoogleNet (99.21), demonstrates its ability to accurately identify normal samples and minimise false positives. The F1-score, which strikes a compromise between sensitivity and precision, is 99.23%. It displays a strong harmonic mean and shows that the suggested model consistently handles imbalanced classes. Additionally, the suggested model's precision of 99.68% is the greatest of all the architectures that were compared. Higher precision lowers the likelihood of misclassifying normal cells as malignant, which lowers the risk of unneeded medical procedures. This makes it an essential parameter in the diagnosis of leukemic cells.

Table (3) presents the performance comparison of the proposed model with other research work carried out for the detection of ALL cells. Accuracies of various classifiers: Convolutional Neural Network (with Gray Level Co-occurrence Matrix (GLCM)) (Agrawal et al., 2019), neighborhood-correction algorithm (Pan et al., 2019), Chronological SCA-based Deep CNN (Oliveira & Dantas, 2021), VGG-16 (Oliveira & Dantas, 2021), CNN (Genovese et al., 2021), AlexNet (Arif et al., 2022), CNN (Claro et al., 2022), VGG16 (Abhishek et al., 2023), and hypercomplex-valued convolutional neural networks (HvCNNs) (Vieira & Valle, 2022) have been compared with the suggested model that shows that this research achieved the highest accuracy.

Table 2. Performance Evaluation and comparison of the Proposed model

Method	Accuracy	Sensitivity	Specificity	Precision	F1-Score
ResNet-18	98.75%	98.80%	98.62%	99.35%	99.08%
ShuffleNet	97.24%	97.34%	97.05%	98.60%	97.97%
GoogleNet	98.56%	98.25%	99.21%	99.62%	98.93%
Proposed	98.87%	98.97%	98.82%	99.68%	99.23%

Table 3. Performance Comparison of Proposed Model with other Research

Author and Year	Dataset	Accuracy
Agrawal et al., 2019	Medical Image & Signal Processing Research Centre	97.30%
Pan et al., 2019	CNMC	91.73%
Jha et al., 2019	ALL-IDB2	98.70%
Oliveira & Dantas, 2021	CNMC	92.48%
Genovese et al., 2021	ALL-IDB2	97.92%
Arif et al., 2022	ALL-IDB	98.05%
Claro et al., 2022	ALL IDB-1 and ALL IDB-2	94.73%
Abhishek et al., 2023	ALL-IDB and ASH	84%
Vieira & Valle, 2022	ALL-IDB2	96.6%
Proposed	CNMC	98.87%

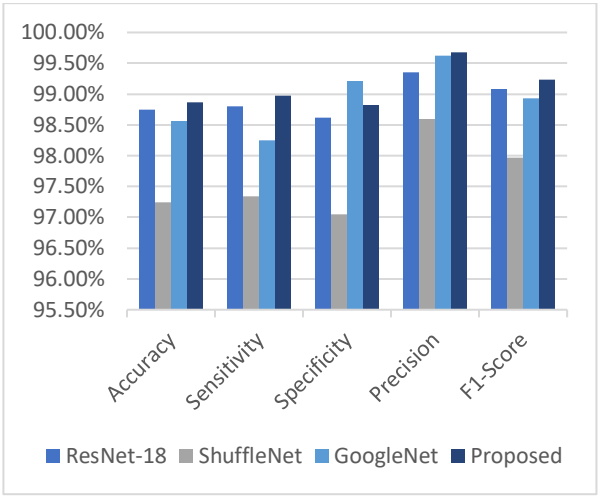
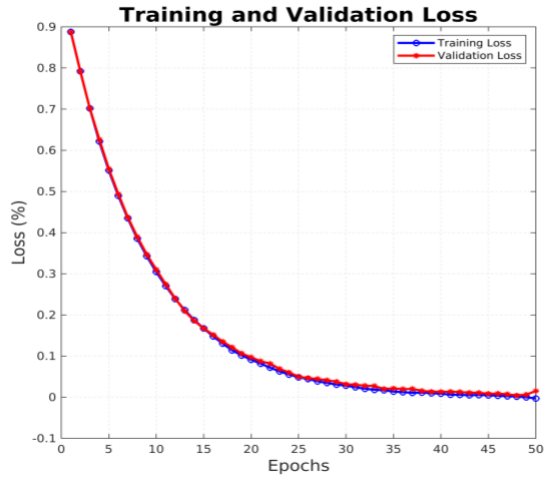
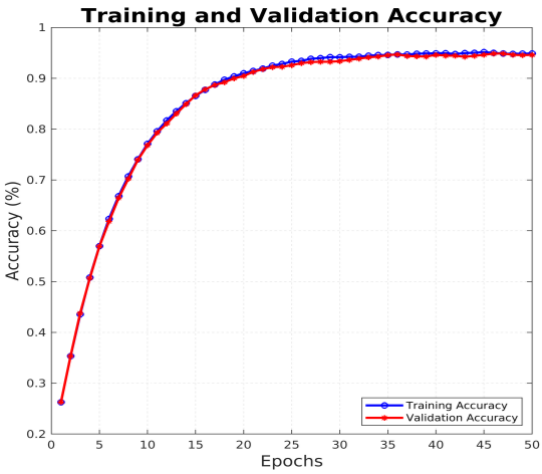


Fig.7. Performance Comparison of Pre-trained and Proposed Networks



(a)



(b)

Fig. 8: Loss and Accuracy Plot of Proposed Model

Fig. (7) presents the results in a visually appealing manner and depicts that the optimized VGG-19 classifier with the aid of stratified 7-fold cross-validation achieved the best results in comparison with other pre-trained frameworks.

Fig. (8) illustrates the convergence of the loss and accuracy of the suggested model during the training and validation procedures, with a validation loss of 0.0163 and a validation accuracy of 99.82%.

Conclusion

This research used stratified 7-fold cross-validation to offer an optimised VGG-19-based deep learning model for the efficient categorisation of normal and malignant leukemic cells. The model learnt discriminative features effectively and transferred learning with altered learning rates for the last three layers. Stratified 7-fold cross-validation was used to make sure the model was thoroughly tested on the complete dataset, resulting in a fair assessment that reduces bias brought on by class imbalance.

With a 98.87% accuracy, a sensitivity of 98.97%, a specificity of 98.82%, a precision of 99.68%, and an F1-score of 99.23%, the suggested model outperformed other pre-trained networks, including ResNet-18, ShuffleNet, and GoogleNet. In order to help medical professionals with early diagnosis and treatment planning, the results show that the optimised VGG-19 model is very effective and dependable in identifying and categorising leukemic cells. The practical implementation of the model can yet be improved, even with its great performance. The limitation of the proposed model is that an imbalanced dataset was used in this investigation. Despite the use of 7-fold stratified cross-validation to reduce sampling bias, the underlying class imbalance persisted. Furthermore, to enhance generalisation and lessen overfitting, heterogeneous data can be used to train the model in the future.

The suggested model could be used as a diagnostic tool to assist hematopathologists in identifying and classifying leukemia at an early stage using images from peripheral blood smears. It is suitable for integration into clinical decision support systems because of its high accuracy and generalisability, which were achieved through fine-tuning and transfer learning. By automating the detection of malignant cells, the model can reduce pathologists' errors and speed up diagnosis, which is required for timely treatment in order to save human lives. The methodological framework of the study also provides a scalable framework for further research in AI-powered medical diagnostics, promoting reproducibility and the development of similar tools for other hematological illnesses.

Acknowledgement

The management of Charotar University of Science and Technology deserves special recognition for their unwavering encouragement throughout the composition of this research work.

Funding Information

This research received no specific grant from any funding agencies.

Author's Contributions

Hema Patel: Responsible for conceptualization, material preparation, data collection, methodology, model development, writing the manuscript, and analysis of the results.

Himal Shah: As a Hematologist-Oncologist, he reviewed the manuscript and validated the clinical relevance of this research.

Gayatri Patel: As a co-supervisor, she reviewed and edited the manuscript. She critically evaluated the manuscript. Her valuable suggestions contributed to improving the overall quality of the manuscript.

Atul Patel: As a supervisor, he supervised the integrity and quality of the work by providing direction and guidance during this research. He reviewed the manuscript, validated the results and edited the manuscript.

Ethics

This research uses a publicly available dataset. There are no ethical issues with this research as this study does not include the participation of humans.

Conflict of Interest

The authors declare no conflict of interest.

References

Abhishek, A., Jha, R. K., Sinha, R., & Jha, K. (2023). Automated detection and classification of leukemia on a subject-independent test dataset using deep transfer learning supported by Grad-CAM visualization. *Biomedical Signal Processing and Control, 83*, 104722. https://doi.org/10.1016/j.bspc.2023.104722

Acharya, V., Ravi, V., Pham, T. D., & Chakraborty, C. (2021). Peripheral blood smear analysis using automated computer-aided diagnosis system to identify acute myeloid leukemia. *IEEE Transactions on Engineering Management, 70*(8), 2760-2773. https://doi.org/10.1109/TEM.2021.3103549

Agrawal, R., Satapathy, S., Bagla, G., & Rajakumar, K. (2019). Detection of white blood cell cancer using image processing. In *2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)* (pp. 1-6). IEEE. https://doi.org/10.1109/ViTECoN.2019.8899602

Agustin, R. I., Arif, A., & Sukorini, U. (2021). Classification of immature white blood cells in acute lymphoblastic leukemia L1 using neural networks particle swarm optimization. *Neural Computing and Applications, 33*(17), 10869-10880.

Arif, R., Akbar, S., Farooq, A. B., Hassan, S. A., & Gull, S. (2022). Automatic detection of leukemia through convolutional neural network. In *2022 International Conference on Frontiers of Information Technology (FIT)* (pp. 195-200). IEEE. https://doi.org/10.1109/FIT57066.2022.00044

Arivuselvam, B., & Sudha, S. (2022). Leukemia classification using the deep learning method of CNN. *Journal of X-Ray Science and Technology, 30*(3), 567-585.

Baig, R., Rehman, A., Almuhaimeed, A., Alzahrani, A., & Rauf, H. T. (2022). Detecting malignant leukemia cells using microscopic blood smear images: A deep learning approach. *Applied Sciences, 12*(13), 6317. https://doi.org/10.3390/app12136317

Claro, M. L., de Veras, R. M. S., Santana, A. M., Vogado, L. H. S., Junior, G. B., de Medeiros, F. N., & Tavares, J. M. R. (2022). Assessing the impact of data augmentation and a combination of CNNs on leukemia classification. *Information Sciences, 609*, 1010-1029. https://doi.org/10.1016/j.ins.2022.07.059

De Oliveira, J. E. M., & Dantas, D. O. (2021). Classification of normal versus leukemic cells with data augmentation and convolutional neural networks. In *Proceedings of VISIGRAPP (4: VISAPP)* (pp. 685-692).

Fatichah, C., Tangel, M. L., Widyanto, M. R., Dong, F., & Hirota, K. (2012). Interest-based ordering for fuzzy morphology on white blood cell image segmentation. *Journal of Advanced Computational Intelligence and Intelligent Informatics, 16*(1), 76-86.

Genovese, A., Hosseini, M. S., Piuri, V., Plataniotis, K. N., & Scotti, F. (2021). Histopathological transfer learning for acute lymphoblastic leukemia detection. In *2021 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)* (pp. 1-6). IEEE.

Glenn, A., & Armstrong, C. E. (2019). Physiology of red and white blood cells. *Anaesthesia & Intensive Care Medicine, 20*(3), 170-174. https://doi.org/10.1016/j.mpaic.2019.01.001

Ikechukwu, A. V., Murali, S., Deepu, R., & Shivamurthy, R. C. (2021). ResNet-50 vs VGG-19 vs training from scratch: A comparative analysis of the segmentation and classification of pneumonia from chest X-ray images. *Global Transitions Proceedings, 2*(2), 375-381.

Jha, K. K., & Dutta, H. S. (2019). Mutual information based hybrid model and deep learning for acute lymphocytic leukemia detection in single cell blood smear images. *Computer Methods and Programs in Biomedicine, 179*, 104987.

Jha, K. K., & Dutta, H. S. (2020). Nucleus and cytoplasm-based segmentation and actor-critic neural network for acute lymphocytic leukaemia detection in single cell blood smear images. *Medical & Biological Engineering & Computing, 58*(1), 171-186. https://doi.org/10.1007/s11517-019-02071-1

Kadhim, K. A., Najjar, F. H., Waad, A. A., Al-Kharsan, I. H., Khudhair, Z. N., & Salim, A. A. (2023). Leukemia classification using a convolutional neural network of AML images. *Malaysian Journal of Fundamental and Applied Sciences, 19*(3), 306-312.

Kalaiselvi, T. C., Santhosh Kumar, D., Subhashri, K. S., & Siddharth, S. (2020). Classification of leukemia using convolution neural network. *European Journal of Molecular & Clinical Medicine, 7*(4), 1286-1293.

Mathur, P., Piplani, M., Sawhney, R., Jindal, A., & Shah, R. R. (2020). Mixup multi-attention multi-tasking model for early-stage leukemia identification. In *ICASSP 2020 - IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 1045-1049). IEEE.

Mittal, A., Dhalla, S., Gupta, S., & Gupta, A. (2022). Automated analysis of blood smear images for leukemia detection: A comprehensive review. *ACM Computing Surveys, 54*(11s), 1-37.

Mourya, S., Kant, S., Kumar, P., Gupta, A., & Gupta, R. (2018). LeukoNet: DCT-based CNN architecture for the classification of normal versus leukemic blasts in B-ALL cancer. *arXiv preprint* arXiv:1810.07961.

Narayanan, K. L., Krishnan, R. S., Robinson, Y. H., Vimal, S., Rashid, T. A., Kaushal, C., & Hassan, M. M. (2025). Enhancing acute leukemia classification through hybrid fuzzy C-means and random forest methods. *Measurement: Sensors, 39*, 101876.

Ouyang, N., Wang, W., Ma, L., Wang, Y., Chen, Q., Yang, S., Xie, J., Su, S., Cheng, Y., Cheng, Q., Zheng, L., & Yuan, Y. (2021). Diagnosing acute promyelocytic leukemia by using convolutional neural network. *Clinica Chimica Acta, 512*, 1-6.

Pan, Y., Liu, M., Xia, Y., & Shen, D. (2019). Neighborhood-correction algorithm for classification of normal and malignant cells. In *ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging* (pp. 73-82). Springer.

Park, S., Park, Y. H., Huh, J., Baik, S. M., & Park, D. J. (2024). Deep learning model for differentiating acute myeloid and lymphoblastic leukemia in peripheral blood cell images via myeloblast and lymphoblast classification. *Digital Health, 10*, 20552076241258079. https://doi.org/10.1177/20552076241258079

Patel, H., Shah, H., Patel, G., & Patel, A. (2024). Hematologic cancer diagnosis and classification using machine and deep learning: State-of-the-art techniques and emerging research directives. *Artificial Intelligence in Medicine, 152*, 102883. https://doi.org/10.1016/j.artmed.2024.102883

Sahlol, A. T., Kollmannsberger, P., & Ewees, A. A. (2020). Efficient classification of white blood cell leukemia with improved swarm optimization of deep features. *Scientific Reports, 10*(1), 2536. https://doi.org/10.1038/s41598-020-59215-9

Setiawan, A., Harjoko, A., Ratnaningsih, T., Suryani, E., & Palgunadi, S. (2018). Classification of cell types in acute myeloid leukemia (AML) of M4, M5 and M7 subtypes with support vector machine classifier. In *2018 International Conference on Information and Communications Technology (ICOIACT)* (pp. 45-49). IEEE. https://doi.org/10.1109/ICOIACT.2018.8350822

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint* arXiv:1409.1556.

Sipes, R., & Li, D. (2018). Using convolutional neural networks for automated fine-grained image classification of acute lymphoblastic leukemia. In *2018 3rd International Conference on Computational Intelligence and Applications (ICCIA)* (pp. 157-161). IEEE Computer Society.

Sivalingam, N. P., Chinnasamy, S., & Suruli Muniyandi, T. (2022). An effective chronic lymphocytic leukemia detection method using hybrid optimization-aware random multimodal deep learning. *Concurrency and Computation: Practice and Experience, 34*(18), e7012.

Sudha, V., & Ganeshbabu, T. R. (2021). A convolutional neural network classifier VGG-19 architecture for lesion detection and grading in diabetic retinopathy based on deep learning. *Computers, Materials & Continua, 66*(1).

Surya Sashank, G. V., Jain, C., & Venkateswaran, N. (2021). Detection of acute lymphoblastic leukemia by utilizing deep learning methods. In *Machine Vision and Augmented Intelligence—Theory and Applications* (pp. 453-467). Springer.

Vieira, G., & Valle, M. E. (2022). Acute lymphoblastic leukemia detection using hypercomplex-valued convolutional neural networks. In *2022 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.

Vogado, L. H., Veras, R. M., Araujo, F. H., Silva, R. R., & Aires, K. R. (2018). Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification. *Engineering Applications of Artificial Intelligence, 72*, 415-422.

Wang, Q., Bi, S., Sun, M., Wang, Y., Wang, D., & Yang, S. (2019). Deep learning approach to peripheral leukocyte recognition. *PLOS ONE, 14*(6), e0218808. https://doi.org/10.1371/journal.pone.0218808