ALNet: A Cluster Layer Deep Convolutional Neural Network for Acute Lymphoblastic Leukemia classification

Malathy Jawahar[1], Sharen H[2], L. Jani Anbarasi[2], Amir H Gandomi[3]

[1]Leather Process Technology Division, CSIR-Central Leather Research Institute, Chennai, India.

[2]School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India.

[3]Faculty of Engineering & Information Technology, University of Technology Sydney, Australia.

ABSTRACT

Acute Lymphoblastic Leukemia (ALL) is cancer in which bone marrow overproduces undeveloped lymphocytes. Over 6500 cases of ALL are diagnosed every year in the United States in both adults and children, accounting for around 25% of pediatric cancers, and the trend continues to rise. With the advancements of AI and big data analytics, early diagnosis of ALL can be used to aid the clinical decisions of physicians and radiologists. This research proposes a deep neural network-based (ALNet) model that employs depth-wise convolution with different dilation rates to classify microscopic white blood cell images. Specifically, the cluster layers encompass convolution and max-pooling followed by a normalization process that provides enriched structural and contextual details to extract robust local and global features from the microscopic images for the accurate prediction of ALL. The performance of the model was compared with various pre-trained models, including VGG16, ResNet-50, GoogleNet, and AlexNet, based on precision, recall, accuracy, F1 score, loss accuracy, and receiver operating characteristic (ROC) curves. Experimental results showed that the proposed ALNet model yielded the highest classification accuracy of 91.13% and an F1 score of 0.96 with less computational complexity. ALNet demonstrated promising ALL categorization and outperformed the other pre-trained models.

Keywords: Convolutional Neural Network, Deep learning, Leukemia, Transfer Learning Models, Computer-aided Diagnostic.

1. Introduction

Leukemia is a type of blood cancer that causes the body's white blood cells to become cancerous. The immune system is placed at risk by these abnormal blood cells, which affect the bone marrow and blood and can suppress red blood cell and platelet formation in the bone marrow[1]. Leukemia is classified into two types depending on the affected white blood cell: myelogenous or acute myeloid leukemia (AML) if the affected cells are monocytes or granulocytes, or lymphoblastic or acute lymphoblastic leukemia (ALL) if the affected cells are lymphocytes. ALL is a dangerous hematological condition categorized by the overproduction

and continued multiplication of malignant and immature white blood cells. The disease can become fatal if left untreated, where the blasts rapidly spread into the bloodstream and other vital organs. Thus, it is pertinent to diagnose the disease as early as possible to improve treatment rates and recovery [2].

ALL is further divided into three subtypes: L1, L2, and L3. L1 cells are typically tiny and have a comparable shape, very little cytoplasm, and a well-structured and discoid nucleus. In comparison to L1, L2 cells exhibit more form variety and are larger with an irregular nucleus and different cytoplasm. L3 cells are round or oval, similar in form and size, have an abundance of cytoplasm, which includes vacuoles, and are usually bigger than L1 [1]. ALL is usually detected and diagnosed using a complete blood count test based on the number of white blood cells. A bone marrow aspiration, a microscopic examination of a blood smear, is used to confirm that a patient has been diagnosed with leukemia [1]. However, these manual detection approaches are both time-consuming and costly.

Considering the numerous limitations of manual detection methods, computer-aided detection and diagnostic methods using deep learning are a promising alternative. Computer-aided methods have proven to be effective, accurate, cost-effective, and fast compared to the traditional approaches. For many years, physicians have classified and counted cells using cytometers for diagnosis. [3]. As an alternative, automatic white blood cells (WBC) identification technology, which is recognized for its low cost and homogeneous accuracy, is gaining attention in hematological illness diagnostics. WBC includes segmentation, feature extraction, classification, and numeration as the four main processes, whereby segmentation directly impacts categorization and counting accuracy [4]. In recent years, many artificial intelligence experts have combined deep learning with Convolutional Neural Networks (CNNs), enabling the development of powerful computer systems for medical assistance. CNNs are also employed in signal processing, natural language processing, image, video processing, medical fusion [24], and tumor diagnosis [26,27]. However, substantial processing capacity and a large quantity of data are required to train and adjust neural networks to perform the required task [9].

Deep neural networks have been shown to provide better results than traditional neural networks. The depth of a neural network is crucial in the performance of a model, which effectively increases when the network layers are increased due to the high ability to extract more complex feature patterns. The complex features help to improve the recognition accuracy of the model. A deep neural network enhances the automation of feature generation resulting in better self-learning capabilities even in unstructured data.

To our knowledge, there are only a few publicly available supervised image datasets that can be used to test and validate methodologies for cell classification and segmentation of ALL. This work utilized the ALL Challenge dataset of ISBI 2019, a dataset created from segmenting microscopic images of white blood cells, for analysis. The strained noise and illumination errors have been rectified after the acquisition process through pre-processing [5].

The proposed custom ALNet model addresses the following research problems:

- The ability to diagnose ALL in datasets with distinctive attributes is a key difficulty for reliable detection and classification.
- ALL and HEM (normal WBCs) classification are based on blast shape and cytochemical staining, where an ideal classification should detect elite features for biological distinctions.
- Issues with optimal image categorization based on morphology, staining, molecular analysis, and flow cytometry must be fixed.
- Improved performance must be generated from the enhanced learning process through identifying the best-needed patterns.
- The selection of suitable features must be obtained for localization and area extraction of leukocytes, which separates distinct cell components, i.e., cytoplasm and nucleus.

Pre-trained models and conventional deep learning architectures are used in most existing research works to diagnose Acute Lymphoblastic Leukemia. Therefore, this work analyzed the need for an accurate diagnosis of ALL through distinguishing immature leukemic blasts from normal cells (cancer). Deep neural custom model ALNet have been proposed to test and identify whether or not a patient is positive for ALL through microscopic image cells. A multi-objective fitness function is created for all and normal patients by considering sensitivity, specificity, precision, and recall metrics. This research presents an ALL analysis/detection model based on a custom convolutional neural network: ALNet with enhanced clustered layers and hyper-tuned normalized parameters. The proposed model identified distinct key features for better classification even though blast shape and cytochemical staining is difficult to differentiate in images. The involved process like morphology, staining, molecular analysis, and flow cytometry identification is overcome by the deep neural network with an enhanced learning process resulting best patterns for ALL.

The significant contributions of the ALNet model are described as follows:

- The custom CNN architecture employs multiple stacked convolution layers that learn the hierarchical features for accurate classification. The batch normalization introduced between these stacked clusters enhances the obtained features for better classification.
- The weights and learning process are enhanced through normalization, resulting in significantly reduced features across the stacked hierarchical clusters and subsequently the accurate and fast classification of ALL microscopic images.
- In order to attain optimal accuracy, data were evaluated on multiple pre-trained transfer learning models, and also hyperparameter tuning was performed.
- The results for both pre-trained and custom models were obtained for training and testing based on ROC curve, accuracy, and loss.

The rest of the paper is organized as follows. The detection of Acute Lymphocytic Leukemia from microscopic images in previous research along with the existing limitations are discussed in section 2. The proposed custom ALNet model and its architecture are presented in section 3. Section 4 discusses the experimental, and section 5 presents the comparison of the proposed

and pre-trained transfer learning models. The study is concluded with a discussion of the future scope in section 6.

2.   Related Works

A review of current work carried out by various researchers on microscopic WBC segmentation and classification is presented in this section. Wu et al. [6] used the Hue (H) and Saturation (S) components of the HSI model to develop an iterative Otsu's threshold technique for leukocyte segmentation based on a circular histogram. Experiments reveal that the approach is successful in segmenting the nucleus of WBCs, but it loses cytoplasm information. On a private dataset of 330 images, Rehman et al. [7] proposed a technique to categorize ALL subtypes using a fine-tuned AlexNet. The latter researchers combined robust segmentation and deep learning approaches and a convolutional neural network to train the proposed model on bone marrow images to obtain reliable classification results. Compared to other techniques, such as K-Neural Network (KNN), Support Vector Machine (SVM), and Naive Bayes, their proposed strategy achieved a good accuracy of 97.78%. Shafique and Tehsin [8] employed ALL-IDB and 50 private images to classify ALL subtypes, using a pre-trained AlexNet that had been fine-tuned for their data. The pretrained network's final levels were replaced with new layers that categorize the input images into four categories, while a data augmentation strategy was utilized to reduce overtraining. Their datasets were compared to several color models to evaluate their performance across a variety of color images. Results show that their strategy achieved a sensitivity of 100%, specificity of 98.11%, and accuracy of 99.50% for acute lymphoblastic leukemia detection and a sensitivity of 96.74%, specificity of 99.03%, and accuracy of 96.06% for acute lymphoblastic leukemia subtype categorization. To classify ALL on ALL-IDB, Vogado et al. [9] applied a number of different pre-trained CNNs as fixed feature extractors and used pre-trained CNNs to extract characteristics from a blood smear image in order to provide a unique visual description. Principal Component Analysis (PCA) was used to extract features to identify the images as healthy or diseased, and an ensemble of classifiers was built by combining three distinct classification algorithms: Random Forest (RF), Multilayer Perceptron (MLP), and SVM. The proposed strategy achieved 100% accuracy. Deep learning models can be used to identify biomarkers from functional Magnetic Resonance imaging [25].

Prellberg and Kramer [11] presented a classification model based on a ResNeXt convolutional neural network with Squeeze-and-Excitation modules and attained an F1-score of 88.91% on the C-NMC dataset. Madhloom et al. [10] utilized and classified a private dataset and classified using a combination of KNN, Naive Bayes, SVM, and MLP, where KNN outperformed with 99% accuracy compared to other conventional classifiers. Shi et al. [12] employed deep convolutional neural networks to solve the cell categorization challenge (CNNs). An ensemble of state-of-the-art CNNs was implemented to enhance the CNN classifier's generalization, and various augmentation techniques were explored. In addition, the Grad-CAM method was used to identify the discriminative region forecasting, thus predicting image category. Their model obtained an accuracy of 86.9% in early testing and 87.9% in final testing. Krishna Kanth [13] suggests the fuzzy hypersphere neural network (FHSNN) classifier for ALL identification

where the number of genes in microarray datasets is greater than the number of samples available, achieving 100% accuracy with fewer genes than previously reported approaches.

On the ISBI 2019 dataset, Kulhalli et al. [14] employed the ResNeXt101 Convolutional Neural Network architecture trained for varied periods by modifying the decision threshold. They attained an F1 score of 0.825 on the testing data. Pan et al. [15] introduced a neighborhood-correction algorithm (NCA), fine-tuned a residual network to generate feature maps, and constructed a fisher vector for classification using similar weighted majority neighbors. Microscopic Normal and cancer cells (C-NMC) images were tested using NCA, which achieved an F1 score of 92.50 and an accuracy of 91.73% during testing. For lymphocyte/lymphoblast classification, Neoh et al. [16] suggested stimulating discriminant measures (SDM) based clustering which attained superior results than SVM, multi-layer perceptron, and Dempster-Shafer ensemble. SDM-based clustering method outperformed Fuzzy C-means, which focuses solely on within-cluster scatter variance when tested with the ALL-IDB2 database. It also outperformed Latent Dirichlet Allocation (LDA) during nucleus-cytoplasm separation, attaining a recognition rate of 96.72%.

Saber et al. [20] proposed a deep learning model based on the transfer learning technique to aid in the detection of Breast Cancer detection. The features are retrieved from the MIAS dataset using a pre-trained convolutional neural network architecture such as Inception V3, ResNet50, Visual Geometry Group networks (VGG)-19, VGG-16, and Inception-V2 ResNet in the proposed model. The transfer learning of the VGG16 model is effective for breast cancer diagnosis, with overall accuracy, sensitivity, specificity, precision, F-score, and AUC of 98.96%, 97.83%, 99.13%, 97.35%, 97.66%, and 0.995% for the 80-20 method and 98.87%, 97.27%, 98.2%, 98.84%, 98.04%, and 0.993% for the 10-fold.

Kumar et al[28] proposed salp swarm and cat swarm method to optimize convolution neural network (SSPSO-CNN) and classify blood cells. Training was carried out using VGG19 model with 10674 images. Particle swarm optimization (PSO) was used along with salp swarm optimization (SSPSO) to improve the classification accuracy. The proposed model attained classification accuracy of 99% using the augmented dataset.

Deep learning has been used in the diagnosis of various diseases including plant diseases and insect pests [29]. The performance is said to be affected by the structure of the network and the hyperparameters which are manually selected. Yu et al. proposed an improved fruit fly optimization algorithm that searches for a better learning rate. Among all the models IResNet50 has better diagnostic accuracy for tomato pests with 94.4%. Generative Adversarial Networks have been used in image processing due to image generation capabilities and have been used in data augmentation. Guan et al.[30] proposed Texture-constrained multichannel progressive generative adversarial network (TMP-GAN), a medical image augmentation method. Joint training of Multiple channels is used to avoid shortcomings of generation methods. To improve the fidelity of synthesized images adversarial learning-based texture discrimination loss was used. To improve the accuracy TMP-GAN uses a progressive generation mechanism of the medical image synthesizer. This method experimented on the CBIS-DDMS and their pancreatic dataset which was observed to show an improved F1 score of 2.77% and 2.36%.

An adaptive learning network called the neuro-fuzzy-based group method of data handling (NF-GMDH) was utilized to forecast the scour process at pile groups owing to waves [31]. The particle swarm optimization (PSO) and gravitational search algorithms were used to create the NF-GMDH network (GSA). Sediment size, geometric property, pile spacing, pile group organization, and wave characteristics upstream of group piles are all effective criteria for scour depth. NF_GMDH is a hierarchical structured adaptive learning network where each neuron has two input variables and one output variable. The final output is the average output of the last layer. The Comparative analysis of techniques used for classification in the existing literature is detailed in Table 1.

Table 1: Comparative analysis of techniques used for classification in the existing literature

Ref	Data	Method	Advantage	Disadvantage	Accuracy (%)
[35]	ALL-IDB1	ShuffleNet	The performance surpasses most of the pre-trained models like Xception and ResNet50.	Results may have misclassification since the rules of shuffling are manually set rather than adaptive rule setting.	96.67
[34]	ALL_IDB2	LeuFeatx – VGG16- adapted fine-tuned feature-extracted model.	High performance of multiple classifiers employing LeuFeatx deep features.	Data Imbalance can cause reduced performance	96.15
[33]	AI-cell platform	CNN	Proposed model classifies 19 types of WBCs and achieves better results than pre-trained models.	Lack of dataset	82.93
[32]	WBC	Capsule Network	Effectively learned training data and achieved high accuracy	WBC's were only classified under five classes	96.86
[28]	WBC	CNN	VGG19 CNN architecture was	Multiple optimization	99.00

			optimized using SSPSO algorithm to achieve high classification accuracy	algorithm is a lengthy process to fine tune CNN	

2.1   Convolutional Neural Network

Advancements in a diversity of pattern recognition fields have been achieved over the last decade, from image processing to speech recognition, due to Convolutional Neural Networks. The most significant benefit of CNNs is that they reduce the number of parameters in Artificial Neural Networks (ANNs). This achievement has prompted researchers and developers to examine larger models in order to address difficult problems that were previously impossible to solve with regular ANNs [17]. Biological activities influence convolutional networks, where the pattern of communication between neurons mirrors a neuron's reaction to a specific stimulus in the visual cortex. Individual neurons react to stimuli based on their receptive zone. The receptive fields of the various neurons somewhat overlap when spanning the complete field of vision. The term "multilayer perceptron" refers to an entirely interconnected network, where neurons in each layer are coupled to all neurons in the layer below it. [18].

CNN is comprised of two neural networks: one that extracts features from the input to be analysed and another that classifies the extracted features. The feature extraction network receives the input image, then the extracted feature signals are fed into the neural network and used to classify the image. The classification neural network then generates an output based on the image's properties [19]. The convolution operation is given in equation (1):

$$F_{\text{all}}\,(A, B) \; = \; (Z * F)(A, B) \; = \; \Sigma_i \Sigma_j \; Z(A + l, B + m)F(l, m) \qquad (1)$$

where Z is the input image matrix; F denotes the two-dimensional filter of sizes A and B; and F_{all} denotes the output feature map. The convolution operation is represented by $(Z * F)$.

CNN offers a variety of alternatives, such as stride, for reducing adverse effects and settings. Controlling the stride allows us to modify the overlap with the following layer's node [17]. Given an image of I×I dimension and L×L filter size, the output size O is determined using equation (2), as illustrated in Figure 1.
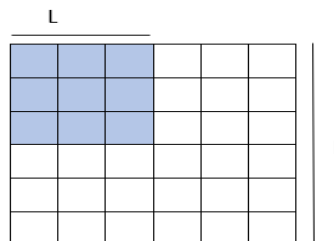

Figure 1. Movement of filter over the image

$$O = 1 + \frac{I-L}{X} \tag{2}$$

where I is the input size; L is the filter size; and X is the stride size. One disadvantage of the convolution method is the loss of information that may be present on the image's border. Because edge features are only captured when the filter slips, they never have a chance to be recognized. The use of zero padding is a simple but effective approach to solve the issue, which is determined by modifying equation (2) into equation (3) with zero-padding:

$$O = 1 + \frac{I+2Z-L}{X} \tag{3}$$

This padding concept allows us to avoid decreasing the network's output size with depth, where Z is the number of zero-padding layers. As a result, any number of deep convolutional networks may be used. Following convolution, non-linearity is the next layer that can be used to regulate or turn off the output and to saturate or limit the output that is generated.

Deep learning networks employ the rectified linear activation unit (ReLU) function to train deep convolutional networks. ReLU returns the value provided as input (z) linearly for values greater than 0 or the value 0 if the input value is 0 or less than 0. The mathematical formulation of ReLU is given by equations (4) and (5):

$$ReLU(z) = \max(0, z) \tag{4}$$

$$\frac{d}{dz} ReLU(z) = \begin{cases} 1 \; if \; z > 0 \\ 0 \; otherwise \end{cases} \tag{5}$$

The pooling layer reduces the size of the image by combining neighboring pixels in a specified area into a single representative value. The three forms of pooling include max, average, and sum. The preceding layer's dimensionality is reduced by half when it is pooled. In a normal neural network, the fully-connected layer is compared to how neurons are constructed and are related to every other node in the subsequent layers.

The Softmax function (F^{all}) computes the relative probabilities of each 'm' output classes using equation (6).

$$F^{all} = \frac{e^{Z_l}}{\sum_{i=1}^{m} e^{Z_i}} \tag{6}$$

where m is the number of classes, e^{Z_l} is the standard exponential function for input vector and e^{Z_i} is the standard exponential function for the output vector. In most cases, the total of the outputs equals 1.

The cross-entropy (Z(PQ)) between two probability distributions (P, Q) defined in equation (7) is employed as the loss function:

$$Z(PQ) = -\Sigma_i P_i \log(Q_i) \tag{7}$$

where P_i represents the probability of the event 'i' in P (target probability distribution), Q_i represents the probability of the event 'i' in Q (approximation of the underlying / target distribution).

2.2   Transfer Learning Models

In its most basic form, transfer learning uses information gained from one activity to solve other related issues. In this study, the knowledge collected from pre-trained networks was used to detect ALL in patients. ALL images were classified from normal images using four pre-trained models: ResNet-50, VGG-16, AlexNet, and GoogleNet. These deep transfer learning pre-trained models were used as a feature extractor, and the classifier was trained on top of this extractor. Residual Networks, or ResNet, is a well-known neural network that is used to perform various computer vision tasks. ResNet-50 is a deep convolutional neural network with 50 layers. AlexNet, the model credited with igniting interest in deep learning, has only eight convolutional layers, including five convolutional layers and three fully connected layers. Comparatively, the VGG network has 19 layers, Inception or GoogleNet has 22, and ResNet 152 has 152. AlexNet is implemented with multiple GPUs, Overlapping Pooling, and ReLU Nonlinearity. Visual Geometry Group (VGG) is a deep convolutional network with 16 or 19 layers in VGG-16 or VGG-19, respectively. However, Simonyan and Zisserman's VGG-16 and VGG-19 architectures include five convolution blocks and three fully connected layers. Inception-V1 is a 27-layer deep CNN, and Inception-V3 is the third version of Google's Inception CNN, which is made up of many inception blocks, each of which contains various sized convolution and pooling layers. Each inception block performs multi-level feature extraction. The InceptionResNet-50V2, which was derived from Francois Chollet's Xception model, has 164 layers, substantially more layers than Inception-V3. Xception, unlike Inception, uses depth-wise separable convolutions instead of inception modules.

2.3 Research Gaps

The predominant gaps in the research on Acute Lymphocytic Leukemia are addressed below.

1. The majority of studies were done using small and private datasets [7,8], which limits the performance of deep learning models. To increase the performance of trained models, validation and testing must be done on large datasets to improve the model generalization.

2. Using typical image processing techniques, it is difficult to describe and extract certain features from microscopic images in order to discern the properties of ALL.

3. The challenge of distinguishing immature leukemic blasts from normal cells under the microscope is difficult since the images of the two cells are morphologically identical; hence, automatic feature extraction must be conducted. Although most deep learning models offer autonomous feature learning, they require many computations to obtain higher classification accuracy. As a result, this trade-off between accuracy and complexity must be handled through customized deep architectures.

3. Methodology

Figure 2 depicts the proposed system's overall workflow, illustrating data acquisition followed by training and testing the dataset using custom and transfers learning models. The dataset was split into training, validation, and testing, where ResNet 50, VGG-16, AlexNet, GoogleNet, and the proposed custom CNN named ALNet were employed for the classification of Acute

Lymphoblastic Leukemia microscopic images. The final output includes the enhanced classified images, which demonstrates good accuracy.
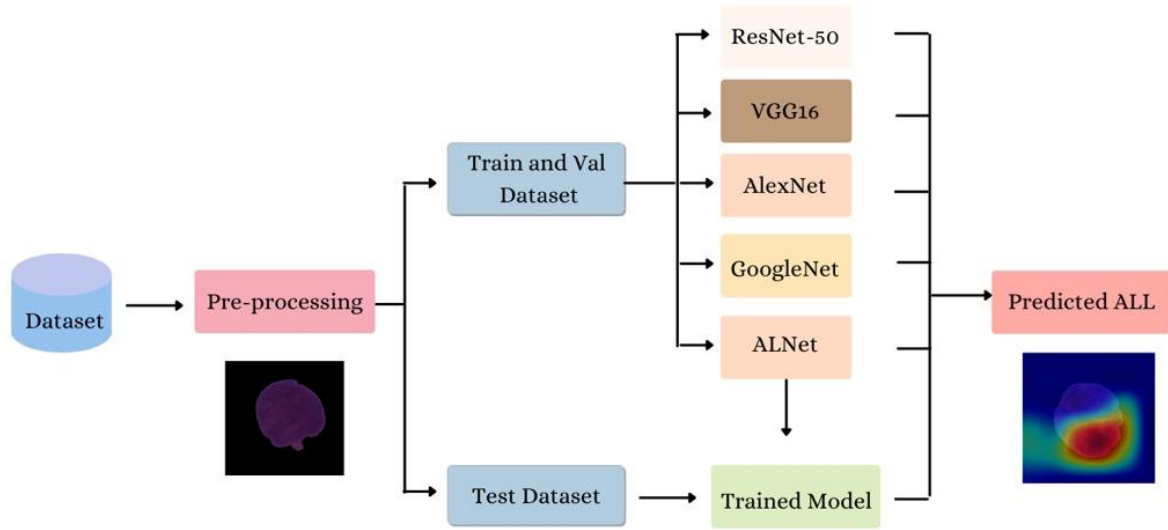

Figure 2. Overall workflow of the proposed system

In recent years, several state-of-the-art convolution models have been used to classify Acute Lymphocytic Leukemia, which primarily include transfer learning and convolution models. This research presents a custom ALNet model and seven pre-trained models to classify Acute Lymphocytic Leukaemia from microscopic images. The proposed ALNet Model includes clustered convolution and pooling layers normalized by a batch normalization layer to reduce features and computational complexity. Batch normalization transforms the mean output close to 0 while keeping the standard deviation close to 1. This layer's output is normalized using the mean and standard deviation of the current batch of inputs.

$$B_N = \gamma * \frac{b - mean(b)}{\sqrt{var(b) + \varepsilon}} + \beta \tag{8}$$

where γ is the learned scaling factor; b is batch; β is a learned offset factor resulting in lesser trainable parameters; and ε is a small configurable constant that is part of the constructor arguments.

3.1   ALNet Model

The proposed ALNet model includes five 2-D convolution layers, max-pooling layers, and an activation layer using ReLU as the activation function. These layers are clustered for generating improved features followed by batch normalization, resulting in elite features for better classification. Figure 3 depicts the suggested flow diagram and features of the ALNet model. Equation (9) is used to measure and evaluate the similarity between test and training images in order to improve their robust detection of ALL and HEM:

$$O = X(n)\varphi(i(n)) \tag{9}$$

where O is the neural network's output; I is the input vector; n is the number of input training samples; and φ is the radial basis function.

The evaluation of classification error e(n) is shown in equation (10), where L(n) is the output response, and n is the number of iterations of the neural network. Equation (11) gives the objective function f(n), and equation (12) is used to determine the weights (z) that are updated during the computation process:

$$e(n) = L(n) - O(n) \tag{10}$$

$$f(n) = \frac{1}{2} e^2(n) = \frac{1}{2} [L(n) - O(n)]^2 \tag{11}$$

$$\Delta z(n + 1) = (n + 1) - \Delta z(n) \tag{12}$$

Equation (13) is calculated to obtain the weights based on the gradient descent:

$$\Delta z(n + 1) = \delta_n \Delta z(n) - \gamma_c P_{z(n)}^\beta f(n) \tag{13}$$

where $0 < \beta < 1$, $0 < \delta < \gamma$; γ is the learning rate; δ is the momentum factor; and δ_n is the momentum coefficient and $P\ refers\ the\ approximation$.
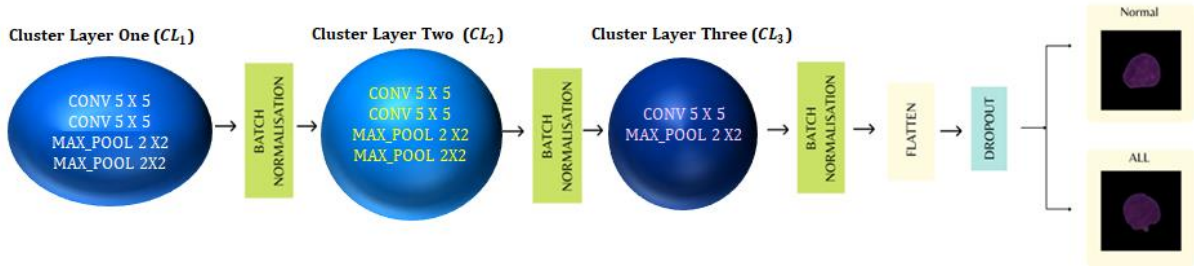
Figure 3. Architectural overview of the proposed network ALNet

3.1.1 Cluster Layer One (CL_1)

The hierarchical structure is maintained for the included layers in each cluster group. The Acute Lymphoblastic Leukemia images are used as the input of the first convolution layer, which is followed by another stacked hierarchical layer, and the generated features are fed to two stacked pooling layers. The convolutional layer processes the input of size 224×224×3 with 5×5 kernel size, yielding a 32-feature map, followed by another convolutional layer that processes the output of the preceding convolution layer with 5×5 pixel filters, yielding a 32-feature map. A feature dimensionality reduction is achieved by employing two max-pooling layers with a 2×2 kernel size. A batch normalization layer is used to normalize the outputs of the previous levels.

3.1.2 Cluster Layer Two (CL_2)

Similarly, two convolution layers are grouped into a cluster using 64 filters with a 5×5- kernel size trailed by two stacked 3×3 max-pooling layers resulting in 11×11×64 fine-scale features of the image. Following the clustered layers, a batch normalization layer is used to normalize the resulting output from the convolution layers.

3.1.3 Cluster Layer Three (CL_3)

The normalized features obtained from the stacked clustered convolution features are fed to a convolution layer followed by a pooling layer. In CL_3 layer the features are processed using 128 filters with 5×5 kernel size, resulting in a 128-feature map. In the max-pooling layer, the 3×3 kernel size is employed, resulting in a dimensionality reduction of 3×3×128. Using batch normalization, the time taken for the training process is reduced, achieving a better and faster convergence with optimal learning rate scheduling. The flattening layer serves as a utility layer for converting the output to a vector. To avoid the network from resulting in overfitting, the number of contributing neurons has been reduced by 30%, which helps to decrease the generalization error. The output obtained from the flattened layer is fed into a two-layer feed-forward network with 128 neurons and a ReLU activation unit on one layer and two neurons and a sigmoid activation unit on the other layer. ReLU preserves sparsity while lowering the probability of a vanishing gradient, which occurs more frequently in dense models. The final layer comprises a sigmoid activation function to overcome the dead ReLU problem, which prevents further learning. Fully connected layers act as classification layers, contributing to the learning of a non-linear function by mixing non-linear combinations of high-level features provided by the output of the convolutional layers. Binary cross-entropy is employed as a loss function during the training phase, while RMSprop is used as an optimizer. The average of gradient descent is determined via equation (14):

$$S_r = \propto . S_r + (1-\propto). \infty Q^2 \qquad\qquad (14)$$

where Q and ∞ are the smoothing parameters used to update the gradient descent (S$_r$) and $\propto$ is the learning rate. This optimizer achieves speedy convergence by using an exponentially decreasing average of squared gradients and discarding information from the previous state. This model achieved an accuracy of 94%, specificity of 0.98, and sensitivity of 0.86. Because it is a shallow network, it works well as a baseline model.

ALNet model consists of training set of images $\{l_i\}$ belonging to two classes $\{l_i\}$, the ultimate objective of the proposed ALNet model is to learn the effective feature mapping $F(i) = l$ into the three clustered levels of composition $F(i) = f(CL_1) \dots, f(CL_2) \dots f(CL_3)$. Each function defines the clustered layer of the processed input images. The features generated in the initial clustered layer processed by the kernel with a bias resulted in new discriminative high and low-level features. These discriminative features are further normalised resulting in fine features to be processed by the next cluster. Finally, the features are fine-tuned to generate the elite feature set to discriminate the ALL and HEM with high accuracy.

Table 2: Layer structure and the obtained features

	Layer Type	Shape	Param
Cluster Layer 1	Conv2D	(None, 220, 220, 32)	2432
	Conv2D_1	(None, 216, 216, 32)	25632
	MaxPooling2D	(None, 108, 108, 32)	0
	MaxPooling2D_1	(None, 54, 54, 32)	0
	BatchNormalization	(None, 54, 54, 32)	128
Cluster Layer 2	Conv2D_2	(None, 50, 50, 64)	51264
	Conv2D_3	(None, 46, 46, 64)	102464
	MaxPooling2D_2	(None, 23, 23, 64)	0
	MaxPooling2D_3	(None, 11, 11, 64)	0
	BatchNormalization	(None, 11, 11, 64)	256
Cluster Layer 3	Conv2D_4	(None, 7, 7, 128)	204928
	MaxPooling2D_4	(None, 3, 3, 128)	0
	BatchNormalization	(None, 3, 3, 128)	512
Classification Layer	Flatten	(None, 1152)	0
	Dropout	(None, 1152)	0
	Dense	(None, 128)	147584
	Dense_1	(None, 2)	258

4. Results and Discussions

The dataset, experiments, model training, and validation are covered in depth in this section. This section also includes a performance comparison of the suggested technique with previous works.

4.1 Experimental Setup

Google Collab notebooks, a cloud computing environment, were used to implement the proposed work. Google Collab offers a free tensor processing unit (TPU) and graphics processing unit (GPU) to develop a deep neural learning model. To run the dataset on the transfer learning model, MATLAB was employed. The custom ALNet was implemented using Phyton programming. The data augmentation was done with OpenCV libraries.

4.2 Data Acquisition and Pre-Processing

The data were downloaded from Kaggle [23], the ISBI 2019 dataset. The training dataset consisted of 73 subjects, including 47 ALL (cancer) patients and 26 healthy individuals. 10,661 cell images, 7272 ALL (cancer) images, and 3389 images were generated for the challenge. The test set consisted of 28 total subjects including 1219 ALL images for 13 ALL patients and 648 normal images for healthy patients, making a total of 1867 cell images. The final test set composition of the data included 2586 images for 17 subjects, including 9 ALL and 8 normal patients. The training data was pre-processed and split to train and test three folds. Table 3 shows the distribution of the training and test data. The size of images used for training and

validation was $450 \times 450 \times 3$ pixels. The images were pre-processed by cropping and resizing, resulting in a size of $224 \times 224 \times 3$ pixels. Cropping was performed around the original image's center to avoid the loss of the data.

Table 3: Distribution of images over train and test over folds

Fold	Fold 0	Fold 1	Fold 2	Total
Training	2821	2864	2843	8528
Test	706	717	711	2134
Total (Fold)	3527	3581	3554	10662

4.3   Data Augmentation

Images from the datasets were in BMP format and resized to $224 \times 224 \times 3$ pixels for better processing. Augmentation was done after splitting the images into the training, validation, and test sets. Some methods, such as 15-degree rotation, horizontal flipping, zooming, and random contrast enhancement, were performed to generate more images. Horizontal axis flipping is substantially more common than vertical axis flipping since it has been used on benchmark datasets like ImageNet CIFAR-10 and Dermnet, making it far more popular. Fold 0 consisted of 2821 training and 706 test images, which resulted in 5642 training and 1412 test images after augmentation. Fold 1 consisted of 2864 training and 717 test images and, after augmentation, 5728 training and 1434 test images. Fold 2 consisted of 2843 training and 711 test images and subsequently 5686 training and 1422 test images after augmentation. Figure 4 displays the ALL, normal, and augmented images.
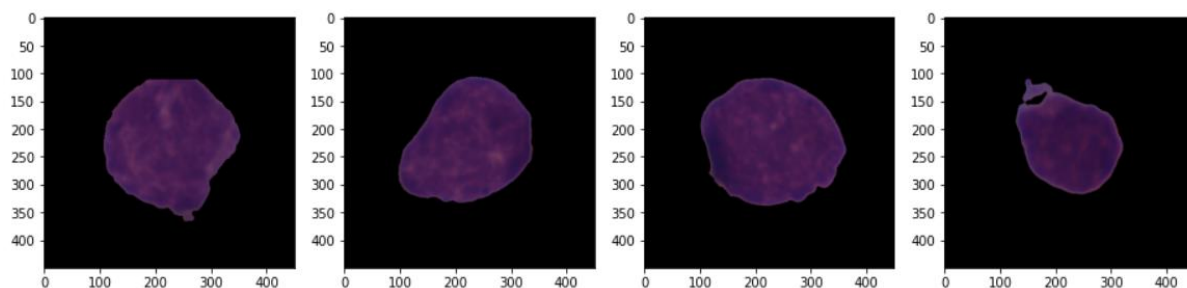
(a)
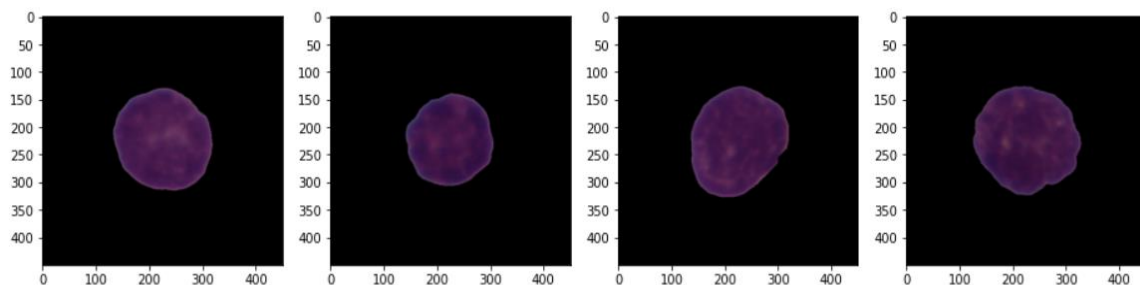
(b)
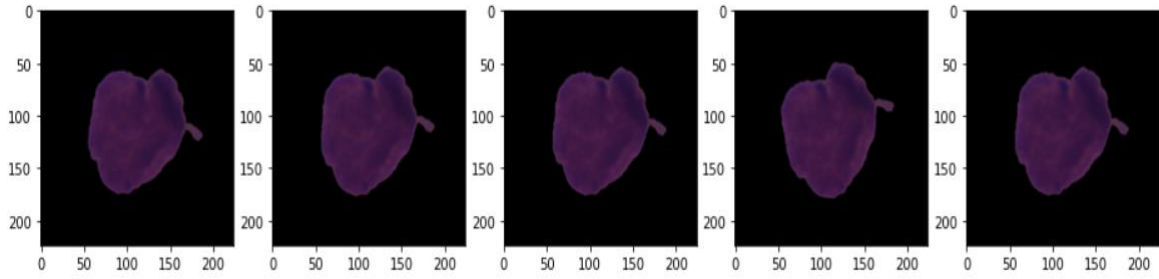
(c)

Figure 4. Examples of ALL, Normal, and Augmented images

4.3 Hyperparameter Tuning

Grid search was used to tune hyperparameters on ALNet models, which included the learning rate, epochs, dropout rate, batch size, and optimization units used in the gradient. The analysis included a dropout factor of 20%, 140 epochs, and gradient optimizers, such as Adams or RMSprop. In most versions of the ALNet models, a dropout likelihood of 20% was found to resonate well to minimize overfitting and improve the impact of generalization, and RMSProp was the best gradient optimizer.

The customized CNN model was trained for 140 epochs using the RMSprop optimizer with a learning rate of 0.0001. The CNN training process using batch normalization acquired better and faster convergence with optimal learning rate scheduling. Batch normalization provides the required regularisation, but it is not assured. Depending on the complexity of the problem dropout is utilized. In the proposed ALNet, addition of dropout before the pooling layer performed better than the batch normalisation with dropout after the pooling layer.

4.4 Visual Interpretation of Trained Model Features

Figure 5 depicts the characteristics retrieved by ALNet's inner layers throughout the training phase. The visual interpretation of the intermediate feature layers shows how the extracted features trains the model.

Visual interpretation of feature map of the first layer	Visual interpretation of feature map of the 6th layer

| Gradcam | Activation map |
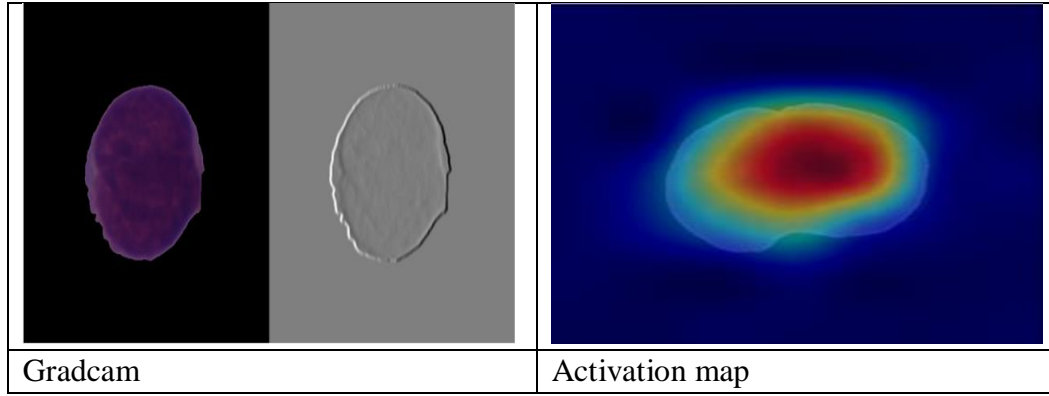
Figure 5. Visualization of the most salient features of the convolution and pooling layers

4.5 Evaluation Metrics

Evaluation metrics are critical for determining the performance of a trained model. The ALNet model's performance was assessed in terms of precision, accuracy, F1-score, recall, and confusion matrix. The testing accuracy was determined by estimating the outcome of the trained model on the test set. The confusion matrix was used to examine the model's performance in each class. The testing accuracy was determined by estimating the outcome of the training phase on the test set that was allocated after the dataset was partitioned. Equations (15) to (18) include the equations for the evaluation metrics:

$$Accuracy = \frac{X}{X+Y+Z+W} \tag{15}$$

$$Precision = \frac{X}{X+Y} \tag{16}$$

$$Recall = \frac{X}{X+W} \tag{17}$$

$$F1\ Score = \frac{2*Precision*Recall}{Precision+Recall} \tag{18}$$

Where X is True Positive, Y is True Negative, Z is False Positive, and W is False Negative. True positives(X) are outcomes when the model correctly predicts the positive class. True Negative (Y) are outcomes when the model correctly predicts the negative class. False positives (W) are outcomes when the model predicts the positive class as the negative class and false negative (Z) is outcomes when the model predicts the negative class as the positive class.

4.6 Ablation Studies

The efficiency of the various components in the proposed design was tested using ablation experiments. The impact of clustering the convolution and pooling layers, followed by batch normalization, is critical. To perform the latter processes, custom ALNet and transfer learning models were trained, and their performance was evaluated using training and testing data grouped into three-folds. The ALNet model is tested for all three-fold test data and transfer learning model is tested with 645 images.**4.7 Analysis of ALNet Architecture**

To test the ALNet model's efficacy for classification, it was trained and tested in three folds. As mentioned, the model's architecture contains three clustered convolutions and max-pooling layers, followed by batch normalization. Each fold contains 535,010 trainable parameters. The model was trained over 140 epochs on each fold using the RMSprop optimizer. Table 4 shows each fold's accuracy and loss curves for both the training and validation sets. An inversely proportional relationship is noticed between the training loss and training accuracy with the increasing number of epochs. The training loss curve keeps declining until 60 epochs for all the three fold data. Eventually, after 60 epochs, the training loss curve saturates and no further significant change is noticed. Similar behaviour in the opposite trend is noticed for training accuracy. A sharp positive slope is observed from 0 epoch to 60 epochs. The rate of change of the positive slope decreases and eventually becomes flat after 60 epochs. The ALNet training process involved batch normalization which acquired better and faster convergence with 0.0001 learning rate. While training accuracy improved with time, where the model converged faster and attained saturation. Although the validation accuracy improved over time, it varied during the training. The model had a mean validation accuracy of 91.13% on all folds. The architecture of ALNet is the simplest among all the other mentioned transfer learning models. AlNet is robust to learning wide range of features relevant to microscopic cells. The architecture of ALNet is relatively smaller and simpler compared to other pre-trained models used in this study. The use of transfer learning model has a disadvantage of being time-consuming and computationally costly. However, the architecture of ALNet is simple and requires fewer parameters and less time.

Table 4: The ALNet model's learning process in terms of the epochs, models loss, and model accuracy curve

	Loss	Accuracy
Fold 0		

Fold 1		
Fold 2		
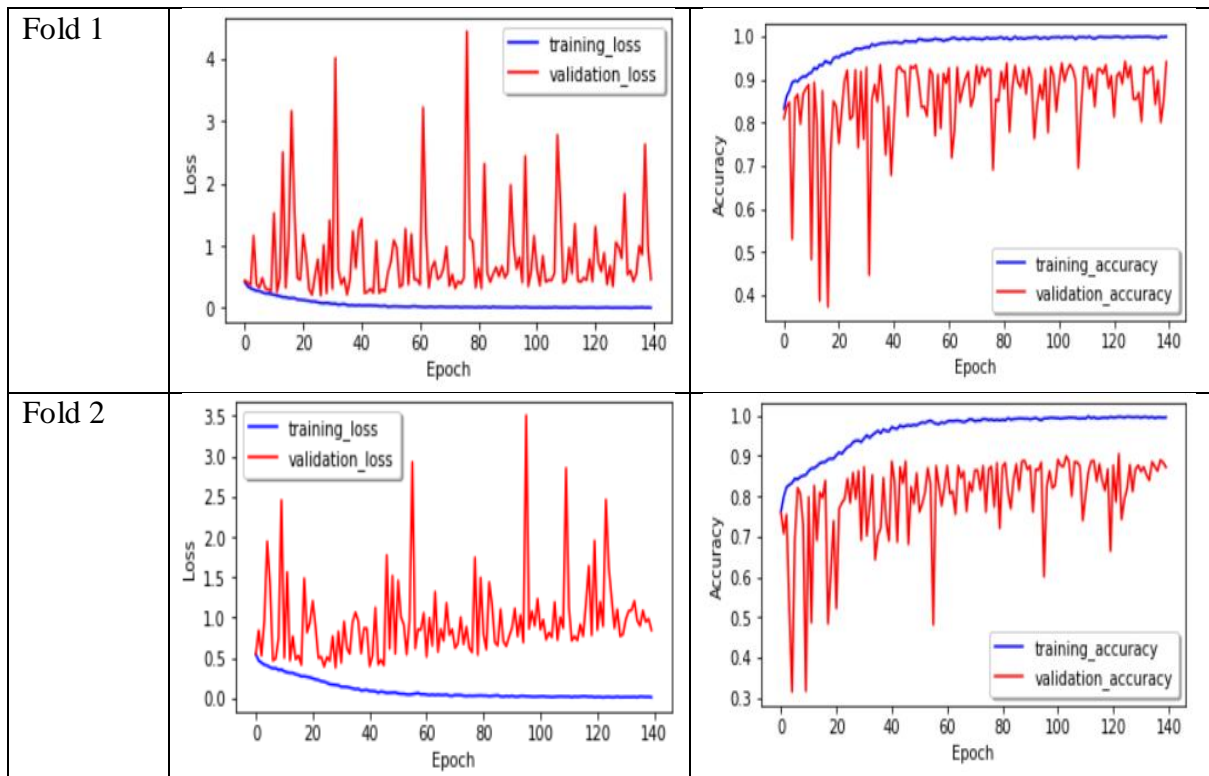
Table 5: Learning process recorded by the AlexNet, VGG16, ResNet50, GoogleNet model with respect to epochs, loss, accuracy, respectively

	Accuracy	Loss
AlexNet		
VGG16		

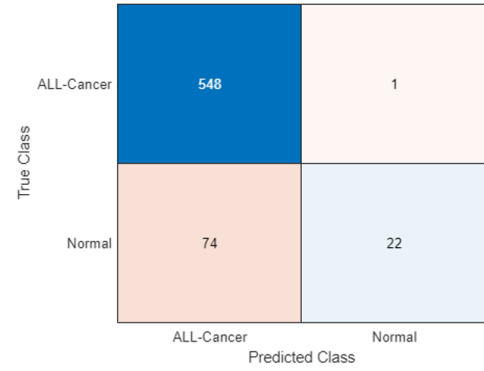|		
GoogleNet		
ResNet-50		
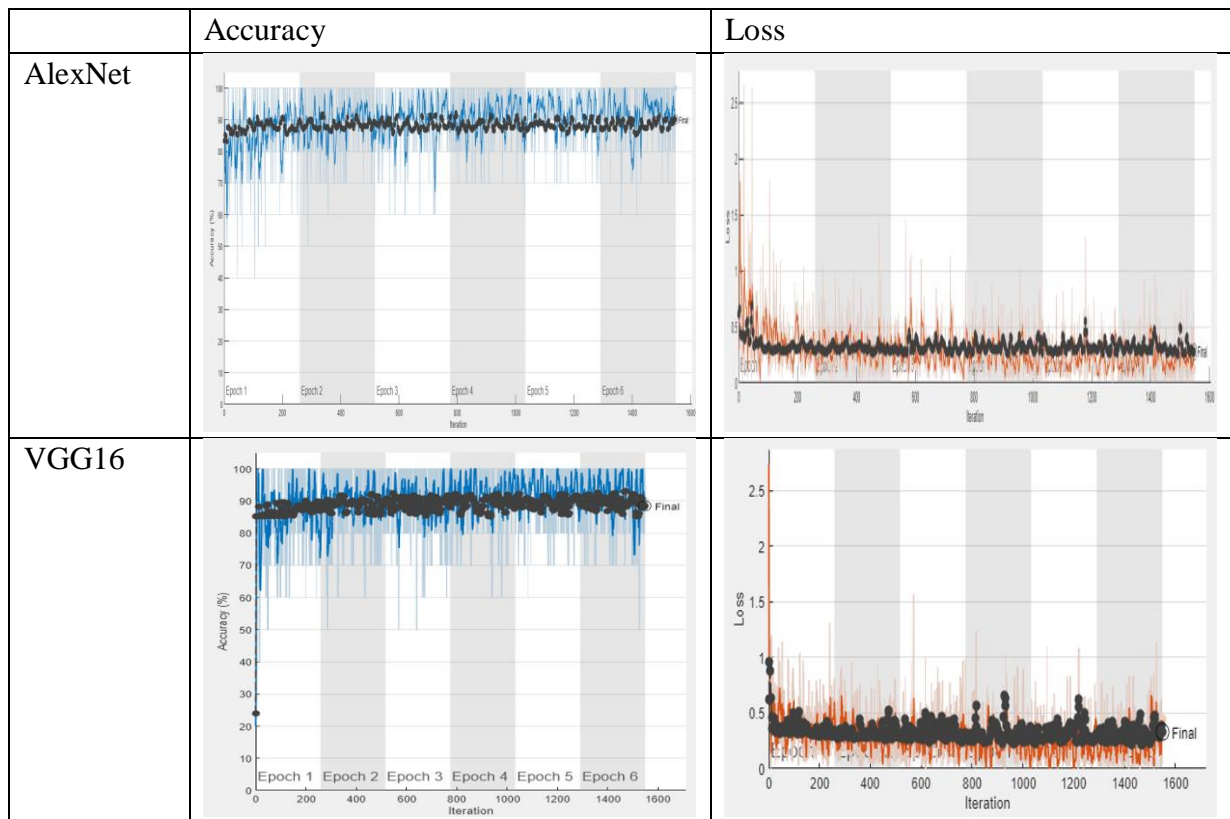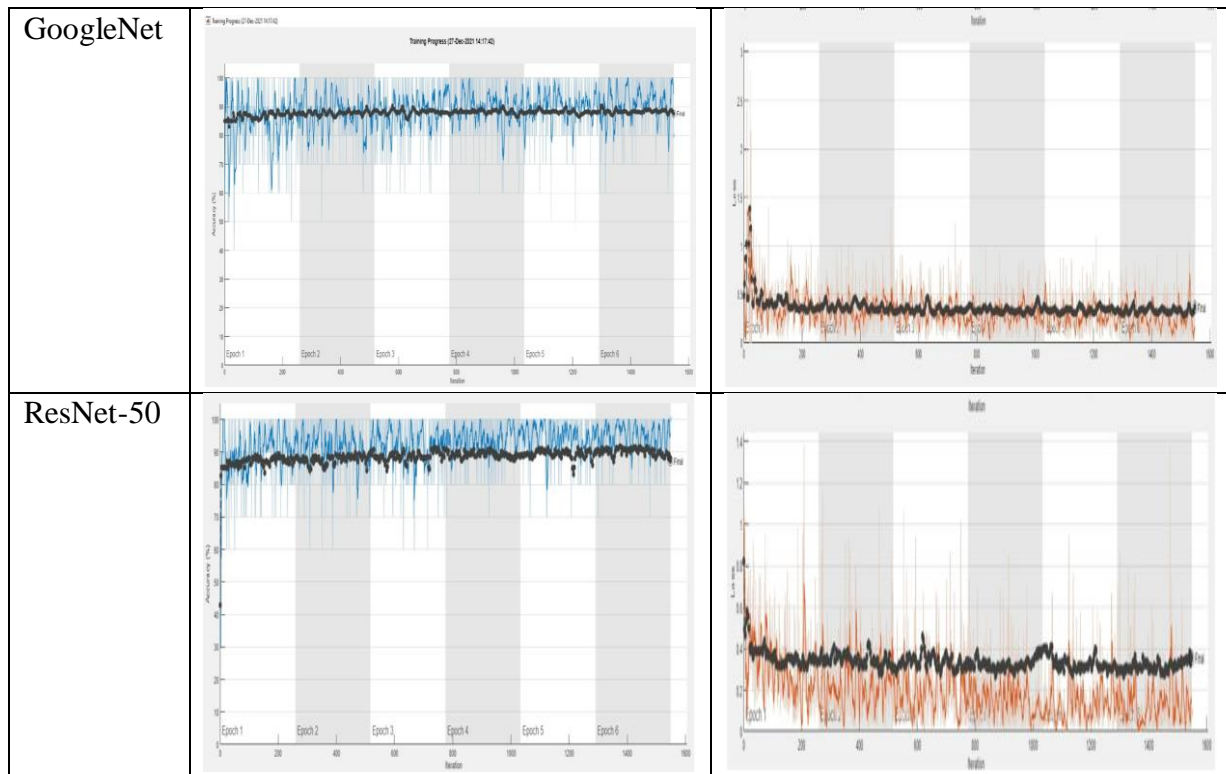
4.8 Analysis of Transfer Learning Architecture
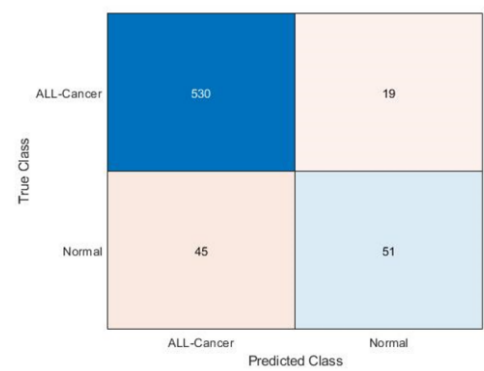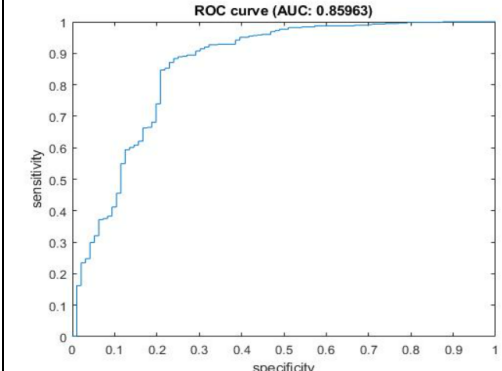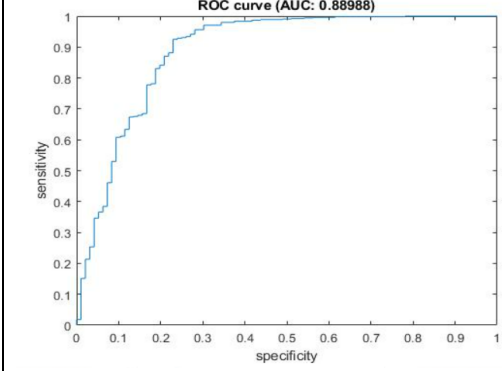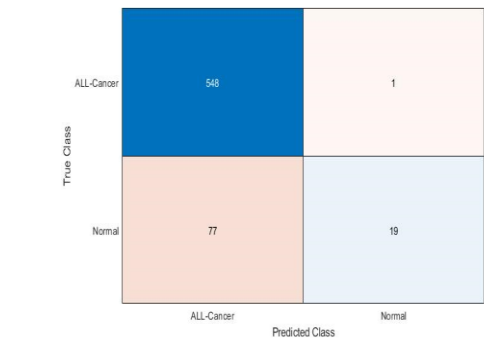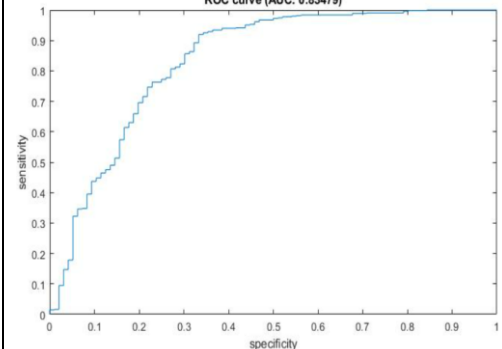
The performance of the pre-trained model with respect to the training and validation accuracy curves is shown in Table 5, and the interpretations are detailed in this section. As the loss curve descends, ResNet-50 yields strong fits until 400 epochs. Following that, the loss function increases, generating a disruption in the accuracy curve, indicating that the model overfits the data, resulting in an accuracy of 89 %. ResNet-50, AlexNet, GpoogleNet and VGG16 was fit over 1600 epochs for better results. The AlexNet model fits well, with an acceptable accuracy of around 90%. The accuracy of GoogleNet and VGG19 is 87 and 88 %, respectively. AlexNet outperformed all other pre-trained models, with a 90% accuracy rate. On the relevant validation sets, the Receiver operating characteristic (ROC) curve and the confusion matrix are illustrated after training. The ROC curve, along with the confusion matrix, is also derived for the transfer learning models and is displayed in Table 6.

4.7.1 Performance Analysis

The CNN architecture's model complexity and trainable parameters were greatly decreased by using the clustered block instead of single layers. The confusion matrix and ROC obtained for the proposed ALNet are presented in Table 7. The performance of the ALNet on the test data set for each fold is illustrated in Table 7. In fold 0, ALNet classifies 180 images as true positive, 470 as true negative, 49 images as false-positive and 6 images as false negative which is close to 92.20% classification accuracy. On fold 1, ALNet achieves 94% classification accuracy with 200 images as true positive, 470 as true negative, 33 images as false-positive and 10 images as a false negative. For the fold 2 dataset, the ALNet model classifies 170 images as true positive, 450 as true negative, 46 images as false-positive and 45 images as a false

negative, which is close to 87.20% classification accuracy. Overall, the average classification accuracy of three-fold data, ALNet achieved 91% accuracy with the highest F1-score of 96% on the test set, whereas the accuracy of AlexNet, VGG16, GoogleNet, and ResNet-50 is found to be 90%, 88%, 87%, and 87% respectively. The findings of the experiments conducted to compare the performance of ALNet with AlexNet, VGG16, ResNet50, and GoogleNet transfer learning model demonstrate the efficiency of ALNet. ALNet model outperformed the transfer learning models with high recall and precision values.

Table 6: Findings recorded by the AlexNet, VGG16, ResNet50, and GoogleNet transfer learning model predicting ROC curve and confusion matrix, respectively

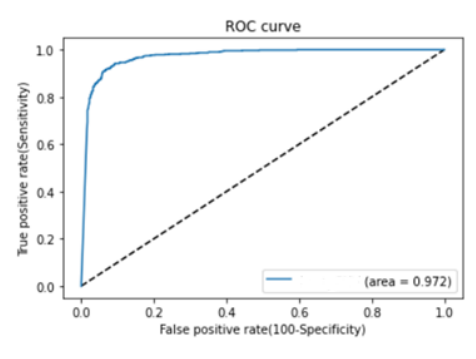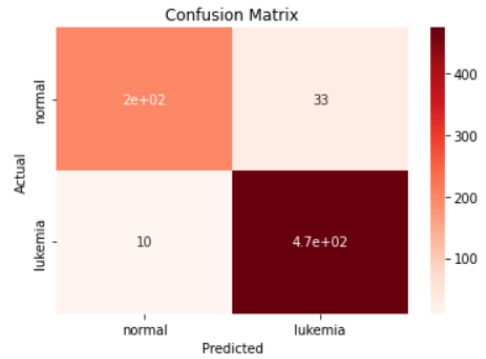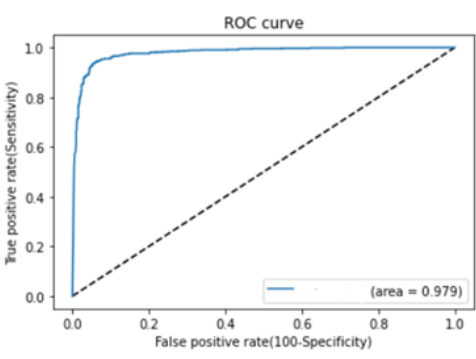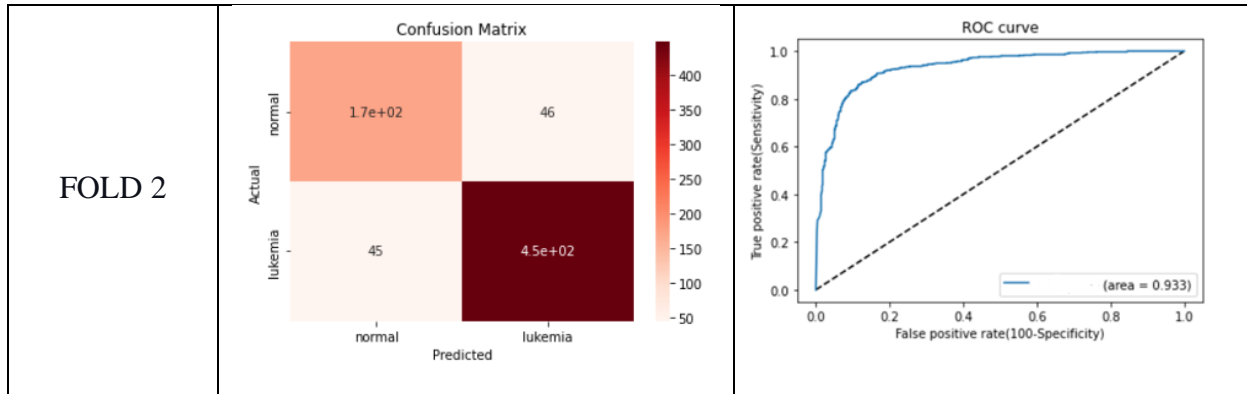|	Confusion Matrix	ROC Curve
AlexNet		
VGG16		
GoogleNet		

ResNet-50		

Table 7: Findings recorded by the ALNet model predicting ROC curve and confusion matrix, respectively

ALNet	Confusion Matrix	ROC Curve
FOLD 0		
FOLD 1		

| FOLD 2 | |

4.7.2 Comparison with State-of-the-art Architectures

The proposed model's performance was compared to other existing methods for classifying ALL microscopic images. The test dataset was used to assess the proposed ALNet and transfer learning models. Performance was compared based on recall, precision, F1-score, and testing accuracy. Table 8 shows the comparison of AlexNet, VGG16, GoogleNet, and ResNet-50, which demonstrate that ALNet had the greatest accuracy and F1-score of all the models and used fewer parameters. Table 9 shows the fold-wise training and testing accuracy of the proposed ALNet model.

Table 8: Comparison of ALNet with other transfer learning models on the dataset

Model	Training parameters	Recall	Precision	Accuracy	F1-score
AlexNet	62,378,344	0.75	0.83	0.90	0.78
VGG16	138,423,208	0.61	0.92	0.88	0.74
GoogleNet	6,414,360	0.60	0.89	0.87	0.71
ResNet-50	25,636,712	0.81	0.78	0.89	0.79
Proposed ALNet	535,010	0.98	0.93	0.91	0.96

Table 9: Fold-wise training and testing classification accuracy

Fold	Accuracy in %	
	Training	Testing
Fold 0	99.75	92.20
Fold 1	99.89	94
Fold 2	99.54	87.20
Average	**99.73**	**91.13**

4.7.3. Comparison with Existing Models for ALL Classification

Table 10 compares the results of the proposed network with the other methods assessed on C-NMC 2019 dataset in recent years. Shi et al [12] used a deep convolution neural network along with various data augmentation and preprocessing techniques and obtained 87.9% accuracy. An ensemble of recurrent and convolution neural networks was proposed by Shah et al [21] to exploit the spectral features using cosine transform. The combination of convolution and RNN model achieved 86.6% accuracy. A weighted ensemble model utilizing the weights from the ensemble candidates' kappa values was investigated by Mondal et al. [22]. As illustrated in Table 8, the proposed ALNet model is simpler in terms of the number of parameters (535,010) compared to other models. Overall, the proposed ALNet model achieved the state-of-the-art highest training (99.73%) and testing accuracy (91.3%) with fewer model parameters performed better than the other methods.

Table 10: Overall comparison of ALNet with existing models

S.no	Source	Dataset	Method	Accuracy (%)
1	Shi et al. [12]	C-NMC-2019	3 x PNASNet-5 + vote11	87.9
2	Shah et al [21]	C-NMC-2019	AlexNet + LSTMDENSE + DCT-LSTM	86.6
3	Mondal et al. [22]	C-NMC-2019	Ensemble of CNN	86.2
4	Proposed	C-NMC-2019	ALNet	99.73

5. Conclusion:

In this work, an efficient cluster layer guided by the normalization CNN model, namely ALNet, is proposed for the classification of Acute Lymphoblastic Leukemia from microscopic images. In general, distinguishing immature leukemic blasts from normal cells under the microscope is difficult since the two cells' appearances are morphologically identical. Importantly, the cluster layers of ALNet extract the structural and contextual details from images, resulting in a robust feature set that discriminates the multiple ALL classes. Experimental results indicated that the ALNet model has learned in-depth features capable of the precise classification of ALL from microscopic images. The proposed ALNet model achieved a high F1-score of 96% with overall training and test accuracy of 99.73% and 91.13% respectively. Overall, ALNet model provides an improved tool to aid pathologists and oncologists in their clinical decision for diagnosing leukemia.

Even though the current study provides high training accuracy, the test accuracy needs to be improved. In future, experiments will be carried out to improve the data augmentation using Generative Adversarial Networks (GANs). It is also proposed to construct a hybrid CNN to extract discriminate features by integrating different pre-trained CNNs and classification using Support Vector Machine (SVM), Random Forest and k-Nearest Neighbours(k-NN) classifiers.

REFERENCES

[1] Shafique, S., & Tehsin, S. (2018). Acute lymphoblastic leukemia detection and classification of its subtypes using pre-trained deep convolutional neural networks. *Technology in cancer research & treatment*, *17*, 1533033818802789.

[2] Li, Y., Zhu, R., Mi, L., Cao, Y., & Yao, D. (2016). Using the dual-threshold method, segmentation of white blood cells from acute lymphoblastic leukemia images. *Computational and mathematical methods in medicine*, *2016*.

[3] Labati, R. D., Piuri, V., & Scotti, L. (2011, September). All-IDB: The acute lymphoblastic leukemia image database for image processing. In *2011 18th IEEE International Conference on Image Processing* (pp. 2045-2048). IEEE.

[4] Li, Y., Zhu, R., Mi, L., Cao, Y., & Yao, D. (2016). Segmentation of white blood cells from acute lymphoblastic leukemia images using the dual-threshold method. *Computational and mathematical methods in medicine*, *2016*.

[5] Gupta, A., & Gupta, R. (2019). ALL Challenge dataset of ISBI 2019 [Data set]. The Cancer Imaging Archive. https://doi.org/10.7937/tcia.2019.dc64i46r

[6] Wu, J., Zeng, P., Zhou, Y., & Olivier, C. (2006, November). It is a novel color image segmentation method and its application to white blood cell image analysis. In *2006 8th international Conference on Signal Processing* (Vol. 2). IEEE.

[7] Rehman, A., Abbas, N., Saba, T., Rahman, S. I. U., Mehmood, Z., & Kolivand, H. (2018). Classification of acute lymphoblastic leukemia using deep learning. *Microscopy Research and Technique*, *81*(11), 1310-1317.

[8] Shafique, S., & Tehsin, S. (2018). Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks. *Technology in cancer research & treatment*, *17*, 1533033818802789.

[9] Vogado, L. H., Veras, R. D. M., Andrade, A. R., de Araujo, L. H., Silva, R. R., & Aires, K. R. (2017, October). Diagnosing leukemia in blood smear images using an ensemble of classifiers and pre-trained convolutional neural networks. In *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)* (pp. 367-373). IEEE.

[10] Mohapatra, S., Patra, D., & Satpathy, S. (2014). An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images. *Neural Computing and Applications*, *24*(7), 1887-1904.

[11] Prellberg, J., & Kramer, O. (2019). Acute lymphoblastic leukemia classification from microscopic images using convolutional neural networks. In *ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging* (pp. 53-61). Springer, Singapore.

[12] Shi, T., Wu, L., Zhong, C., Wang, R., & Zheng, W. (2019). Ensemble convolutional neural networks for cell classification in microscopic images. In *ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging* (pp. 43-51). Springer, Singapore.

[13] Kanth, B. K. A Fuzzy-Neural Approach for Leukemia Cancer Classification.

[14] Kulhalli, R., Savadikar, C., & Garware, B. (2019). Toward automated classification of b-acute lymphoblastic leukemia. In *ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging* (pp. 63-72). Springer, Singapore.

[15] 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, Singapore.

[16] Neoh, S. C., Srisukkham, W., Zhang, L., Todryk, S., Greystoke, B., Lim, C. P., ... & Aslam, N. (2015). An intelligent decision support system for leukemia diagnosis using microscopic blood images. *Scientific reports*, *5*(1), 1-14.

[17] Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017, August). Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)* (pp. 1-6). Ieee.

[18] Sanket, S., Vergin Raja Sarobin, M., Jani Anbarasi, L., Thakor, J., Singh, U., & Narayanan, S. (2021). Detection of novel coronavirus from chest X-rays using deep convolutional neural networks. *Multimedia Tools and Applications*, 1-26.

[19] Kim, P. (2017). Convolutional neural network. In *MATLAB deep learning* (pp. 121-147). Apress, Berkeley, CA.

[20] Saber,A., Sakr,M., Abo-Seida, O. M., Keshk,A and Chen,H., "A Novel Deep-Learning Model for Automatic Detection and Classification of Breast Cancer Using the Transfer-Learning Technique," in IEEE Access, vol. 9, pp. 71194-71209, 2021, doi: 10.1109/ACCESS.2021.3079204

[21] Shah, S., Nawaz, W., Jalil, B., & Khan, H. A. (2019). Classification of normal and leukemic blast cells in B-ALL cancer using a combination of convolutional and recurrent neural networks. In *ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging* (pp. 23-31). Springer, Singapore.

[22] Mondal, C., Hasan, M., Jawad, M., Dutta, A., Islam, M., Awal, M., & Ahmad, M. (2021). Acute Lymphoblastic Leukemia Detection from Microscopic Images Using Weighted Ensemble of Convolutional Neural Networks. arXiv preprint arXiv:2105.03995.

[23] Gupta, A., & Gupta, R. (2019). ALL Challenge dataset of ISBI 2019 [Data set]. The Cancer Imaging Archive. https://doi.org/10.7937/tcia.2019.dc64i46r

[24] Muhammad Adeel Azama, Khan Bahadar Khan, Sana Salahuddin, EidRehman, Sajid AliKhan, Muhammad AttiqueKhan, SeifedineKadry, Amir H.Gandomi. (2022) "A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics." Computers in Biology and Medicine 144 (2022): 105253.

[25] Amirhessam Tahmassebi, Amir H. Godomi, JanMCCann, Mieke H.J. Schulte, Anna E Goudriaan, Anke Meyer-Baese. (2018) "Deep learning in medical imaging: fMRI big data analysis via convolutional neural networks." Proceedings of the Practice and Experience on Advanced Research Computing. 2018. 1-4.

[26] Ambeshwar Kumar, ManikandanRamachandran, Amir H.Gandomi, RizwanPatan, SzymonLukasik, Ravichandran KatturSoundarapandian (2019) "A deep neural network based classifier for brain tumor diagnosis." Applied Soft Computing 82 (2019): 105528.

[27] Gopi Kasinathan, Selvakumar Jayakumar, Amir H.Gandomi, ManikandanRamachandran, Simon JamesFong, RizwanPatan (2019) "Automated 3-D lung tumor detection and classification by an active contour model and CNN classifier." Expert Systems with Applications 134 (2019): 112-119.

[28] Kumar, R, Joshi, S , Dwivedi, A, "CNN-SSPSO: A Hybrid and Optimized CNN Approach for Peripheral Blood Cell Image Recognition and Classification." International Journal of Pattern Recognition and Artificial Intelligence, 35(5), 2021

[29] Yu, H., Liu, J., Chen, C., Heidari, A. A., Zhang, Q., & Chen, H. (2022). Optimized deep residual network system for diagnosing tomato pests. *Computers and Electronics in Agriculture*, *195*, 106805.

[30] Guan, Q., Chen, Y., Wei, Z., Heidari, A. A., Hu, H., Yang, X. H., ... & Chen, F. (2022). Medical image augmentation for lesion detection using a texture-constrained multichannel progressive GAN. *Computers in Biology and Medicine*, *145*, 105444.

[31] Najafzadeh, M., & Azamathulla, H. M. (2015). Neuro-fuzzy GMDH to predict the scour pile groups due to waves. *Journal of Computing in Civil Engineering*, *29*(5), 04014068.

[32] Baydilli, Y. Y., & Atila, Ü. (2020). Classification of white blood cells using capsule networks. *Computerized Medical Imaging and Graphics*, *80*, 101699.

[33] Zhou, M., Wu, K., Yu, L., Xu, M., Yang, J., Shen, Q., ... & Zhao, L. (2021). Development and Evaluation of a Leukemia Diagnosis System Using Deep Learning in Real Clinical Scenarios. *Frontiers in Pediatrics*, *9*.

[34] Rastogi, P., Khanna, K., & Singh, V. (2022). LeuFeatx: Deep learning–based feature extractor for the diagnosis of acute leukemia from microscopic images of peripheral blood smear. *Computers in Biology and Medicine*, 105236.

[35] Das, P. K., & Meher, S. (2021, July). Transfer learning-based automatic detection of acute lymphocytic leukemia. In *2021 National Conference on Communications (NCC)* (pp. 1-6). IEEE.