# Automated Leukemia Detection and Classification Using Deep Learning Techniques

**J. Sri Sai Samhitha**[1]
Dept. of Computer Science and Engineering,
Amrita School of Computing, Bengaluru,
Amrita Vishwa Vidyapeetham, India.
bl.en.u4cse21072@bl.students.amrita.edu

**K. Adarsh Sagar**[2]
Dept. of Computer Science and Engineering,
Amrita School of Computing, Bengaluru,
Amrita Vishwa Vidyapeetham, India.
bl.en.u4cse21075@bl.students.amrita.edu

**Mohammed Jaffer Ali**[3]
Dept. of Computer Science and Engineering,
Amrita School of Computing, Bengaluru,
Amrita Vishwa Vidyapeetham, India.
bl.en.u4cse21124@bl.students.amrita.edu

**Tripty Singh**[4]
Dept. of Computer Science and Engineering,
Amrita School of Computing, Bengaluru,
Amrita Vishwa Vidyapeetham, India.
tripty_singh@blr.amrita.edu

***Abstract*—Healthcare is a wide field, which offers many facilities where people can learn about particular diseases, develop new technologies, perform surgeries, investigate the patients state and guarantee that the appropriate treatment is provided to the patient. With the increase in the number of health issues, several new problems have emerged in the medical sector which involves identification of diseases, right prescription of drugs, reduction in the amount of money that is used in acquiring equipment for diagnosis, enhancing the quality of the services being delivered to the patients, increasing accessibility of health care to many people with opportunities presented by advances in medical technology. With an aim to solve the above challenges, deep learning methods have been developed and integrated into the medical domain and are very essential when it comes to medical image analysis to provide the correct input that will enable the correct disease diagnosis. An example of such application is detection of leukemia cancer where several Deep learning models are integrated and trained on Microscopic images of patients blood cells to differentiate between HEM and acute Lymphoblastic leukemia which is most common among children. The captured microscopic images are altered by different image pre-processing methods, which will help the model not only to analyze, but also to improve the diagnosis and efficiency in the HEM and ALL differentiation process.This paper involves various deep learning models including Resnet50, InceptionV3 and EfficientNetB0 that are trained and tested on the images. Finally a new model has been proposed which outperformed the traditional deep learning models by providing an accuracy of 98.22% in classifying as "hem" and "all".**

***Index Terms*—Healthcare sector, Leukemia diagnosis, Hematological Malignancies (HEM), Acute lymphoblastic leukemia (ALL), Deep Learning models.**

## I. Introduction

In the current generation, one of the most important sectors that play major role in the promotion and maintenance of sustainable human health is the health care sector that adapts and grows with improved knowledge of various diseases which can be chronic and acute in nature. The ever increasing population as well as life time expectancy increases the intensity of pressure put on research, medical advancements and the medical resources availability. Other challenges which still persist apart from these main breakdowns include improving accessibility, increasing complexity of diseases, the ever increase in the overall costs and limited healthcare facilities available in some of the regions in the country. To enhance the patient health and minimize the resources expended in medical treatments and get rid of these challenges innovation and collaborative efforts are needed.

Under the healthcare sector, oncology which deals with study and treatment of cancer hold a very important place due to the fatal and complex impact of Cancer on human health. Leukemia is a blood cancer condition that affects the blood forming tissues and bone marrow by an abnormal increase in the count of white blood cells which disrupt the normal bloods ability to fight the diseases.Various factors including genetic disorders like down syndrome can increase the risk of leukemia. Environmental factors including exposure to radiations,chemicals and certain viruses also can contribute to leukemia. In some cases family history of leukemia or even other blood disorders can also lead to leukemia. Tiredness, frequent infections, easy bruising,bleeding or sudden weight loss are some of the symptoms of leukemia. A complete blood count is done, or a blood sample is taken to find out or rule out leukemia and diagnose the correct treatment which could be chemotherapy or treatment by drugs that can look onto the specific characteristics of cancerous cells. Leukemia is categorized primarily into four categories : the Acute Lymphoblastic Leukemia (ALL), Acute Myeloid Leukemia (AML), Chronic Lymphocytic Leukemia (CLL), and Chronic Myeloid Leukemia (CML). Among these categories ALL and

AML are associated to the acute type where they progress very quickly hence require immediate attention and treatment. ALL effects the lymphocytes ( a type of white blood cells) and AML is associated with the abnormal growth in the number of myeloid cells. Both of the above mentioned conditions are life threatening and require accurate,timely diagnosis and treatment to improve the patient health and outcome.

Adopting artificial intelligence technology and applying deep learning into the field of oncology helps research and work to be endowed with useful tools and methods of diagnosis and disease classification including leukemia. The deep learning methods are advantageous in defining mild variations of the medical images and can work through the huge capacity data which may be bulky and tremendously time consuming to be analyzed by the common methods to identify the certain cellular changes which may be easily negated.The goal of this paper is to use Deep Learning models like RESNet-50, EfficientNet B0, InceptionV3 and the improved and efficient model proposed in order to effectively differentiate leukemia as "all" type or "hem" type.

## II. Literature Review

This section give an overall view of the insights obtained by analyzing various papers related to Leukemia Detection and Classification Using Deep Learning Techniques. [1] Payel Patra, Tripty Singh. In "Diabetic Retinopathy Detection using an Improved ResNet 50-InceptionV3 and hybrid DiabRetNet Structures", the authors give a deep learning framework to early detect and classify DR from eye fundus images with a much higher average accuracy of 93.79% that is, 7% higher than previous works. The approaches involve image preprocessing where fundus images are normalized and data augmented. Features of the image preprocessed were extracted using various versions of Convolutional Neural Network (CNN) structures inclusive of ResNet-50, Inception-V3, VGG- 19, DenseNet-121, and MobileNet V2. To checkout classification accuracy, the concept of DiabRetNet was initiated with use of multiple CNN models and activation functions. The proposed DiabRetNet architecture provided improved detection accuracy of 93.79% for the DR, which is helpful in early diagnosis of the disease which can prove helpful in preventing blindness.[2] Sharath Sunil et al., "An Effective Approach for Classifying Acute Lymphoblastic Leukemia Using Hybrid Hierarchical Classifiers," proposes an automated system that employs the hybrid hierarchical classifiers to diagnostic ALL cells into the several subtypes that includes L1, L2 and L3 with relative accuracy compared to those machine learning approaches. The proposed system eliminates the problem of variable accuracies that is a common thing with the previous systems. Out of all the algorithms tested, the two with the best result at classifying ALL cells were the KNN and Naïve Bayes Multinomial, with an overall accuracy of 60.8%. The system has segmentation module that uses image processing including converting image to grayscale, applying a 2D-order statistic filter, equalizing the histogram of the image, and applying thresholding using Otsu's method to arrive at a binary image. In feature extraction, the size, the boundary length, color values, form, and other morphological parameters of the image in the binary form are obtained. The classification module implies the use of hybrid hierarchical classifiers to determine initially the ratio of healthy and malignant cells and further – the division of malignant cells into L1, L2, and L3.[3] V. Lakshmi Thanmayi A et al., "Detection of Leukemia Using K-Means Clustering and Machine Learning," proposes an algorithm for developing a decision-support system in the early screening of leukemia with using image processing and machine learning classification. The proposed algorithm has achieved 95% accuracy and approximately 93% precision for leukemic and non-leukemic cell classification utilizing a dataset of 368 blood images. For the image segmentation part, United States uses the K- means clustering method; for the classification of the cells, it uses a linear Support Vector Machine (SVM) classifier. It comprises image pre-processing where noise is filtered and features emphasized, segmentation by K-means clustering to isolate the nucleus from the cytoplasm, further morphological processing of the nucleus image to optimize analysis, feature extraction from the segmented nucleus and classification by a linear SVM classifier.[4] K. S. Ananthu et al., "Acute Lymphoblastic Leukemia Detection Using Transfer Learning Techniques," proposes a comparison of various transfer learning models for identification of acute lymphoblastic leukemia from blood smear cells in which the MobileNet model offers the maximum accuracy of 97.88%. MobileNet was the best-scoring model out of all of the models that were used, delivering an accuracy level of 97.88%.Having undergone training, the accuracy on the validation set of MobileNet as a model amounted to 97.88%. MobileNet, Xception, InceptionV3, ResNet50 and DenseNet201 are the transfer learning models utilized for the experiment with data being split in a ratio of 80%, for training the model, and 20% for validation. Training was performed on a batch size of 32, 20 epochs, a learning rate of 0.0001 and the Adam Optimizer.[5] Arjun Abhishek et al., "Automated Classification of Acute Leukemia on a Heterogeneous Dataset Using Machine Learning and Deep Learning Techniques," proposes a new dataset of images of blood smears and applies machine learning and deep learning approaches in a binary and three-class problem to detect acute leukemia. The authors accomplished the leukemia classification task, as a binary classification problem of cancerous and normal, with an accuracy of 96% using a fine-tuned VGG16 model. Further, they achieved 95% accuracy in three classifying acute lymphoblastic leukemia (ALL) acute myeloid leukemia (AML) and normal samples using ResNet50. . The study also features binomial classification using workhorse identifiers like LBP, HOG, and SVM besides binomial categorization through transfer learning; SVM on the proposed and mixed datasets and fine-tuned CNN models.[6] Chaudhary Hassan Abbas Gondal et al., "Automated Leukemia Screening and Sub-types Classification Using Deep Learning," proposes an effective AC for the identification of leukemia and the classification between L1, L2, L3 using deep learning approaches with better accuracy than all existing approaches. The authors

utilized transfer learning with pre-trained convolutional neural networks (CNNs) to extract features from the input images and compared two transfer learning approaches: The fine tuning of the pre-learned CNNs and the stripping off of all the layers that feeds data to the final layer that feeds data to the shallow classifiers such as SVM and KNN. They do not carry out segmentation in their study, therefore, the methods described here extract features from the input images directly where the feature is extracted from the last full connected layer of pre-trained CNNs, and then passed to the SVM and KNN classifiers.[7] Pradeep Kumar Das et al., "An Efficient Deep Convolutional Neural Network Based Detection and Classification of Acute Lymphoblastic Leukemia," introduces an enhanced deep convolutional neural network (CNN) architecture that is designed to detect and classify Acute Lymphoblastic Leukemia (ALL) by adopting a probability-based weight factor to integrate MobileNetV2 and ResNet18 and produce exceptional performance with reduced computational complexity. The above deep learnt CNN framework gives the maximum accuracy on ALLIDB1 and ALLID test protocol of 99.39% and 97.18% respectively achieved on 70% training and 30% test data. The authors have presented a transfer learning model that integrates the MobileNetV2 and ResNet18 architectures, and use efficient bottleneck layer blocks of MobileNetV2 for better computational effectiveness. Furthermore, the authors proposed a new weight factor (Wf) as a way to combine the class probabilities in effort to achieve the best working balance. Kodavati, T. et al. [8] The investigation examines arrhythmias, a category of cardiovascular disorders, utilizing a hybrid model that integrates Gated Recurrent Units (GRU) with MobileNet V1, achieving a detection accuracy of 99.4% based on the MIT-BIH Arrhythmia dataset. This innovative approach proves to be a valuable asset in clinical diagnostics and patient management, addressing the critical need for prompt identification and classification of arrhythmias, which fundamentally disrupt cardiac rhythm and function amidst the ongoing global cardiovascular health crisis. Akhil et al. [9] study investigates the application of CT imaging for the identification of ischemic strokes, highlighting the automated delineation of infarct core areas through advanced deep transfer learning methodologies. Notably, the VGG-16 architecture demonstrated promising efficacy, achieving a Dice coefficient of 0.79 and an Intersection over Union (IoU) of 0.76, with the overarching aim of enhancing patient outcomes and clinical decision-making regarding the significant public health challenge posed by ischemic strokes. Ranjitha et al. [10] deals about blood cancer, specifically leukemia, significantly contributes to mortality worldwide, accounting for one in six deaths. Timely diagnosis is essential for successful treatment interventions. The pathogenesis of leukemia involves the aberrant proliferation of white blood cells in the bone marrow. Hematologists employ advanced microscopic imaging techniques for precise diagnosis. The methodologies encompass image segmentation, clustering, and classification, with K-means being a pivotal tool for determining cancer stages. The proposed approach demonstrates a commendable accuracy of 90% in the identification of cancerous cells. Gopigari et al. [11] deals about Leukemia, a rapidly proliferating hematological malignancy, significantly compromises immune functionality and predominantly affects pediatric and young adult populations. The present study evaluates various cellular segmentation methodologies utilizing performance metrics such as PSNR and MAE, while acknowledging that leukemia's etiology may involve both environmental factors and genetic susceptibilities, with standard therapeutic approaches encompassing chemotherapy and bone marrow transplantation. Tripathi et al. [12] discusses regarding lung disease ranks as the third primary cause of mortality worldwide. Early identification is essential to mitigate adverse health effects. Radiographic imaging is the standard method for initial lung disease assessment. Artificial intelligence significantly improves diagnostic precision in medical contexts. This research introduces two deep learning frameworks for lung disease categorization. The proposed models demonstrate superior performance compared to conventional methodologies in accuracy evaluations. Zolfaghari et al.[13] manuscript comprehensively reviews the automation of leukemia identification and white blood cell stratification. It compiled data from research articles written in indexed journals between 2012 and 2020. To achieve this, the research categorises previous work into six different classification categories. It emphasizes the importance of the feature extraction and the classification methods applied on feature spaces in the learning machine frameworks. Performance analysis of the metrics indicates that while traditional and deep learning methods work differently, they are not of the same efficiency. Thus, the study has evidence that future research utilising more complex CNNs may improve the diagnostic accuracy of this kind of algorithms. Anilkumar et al. [14] work examines the use of deep learning in differentiating leukemia through automated analysis. Traditional techniques of diagnostics are known to be a time-consuming and cost-Effective of cost. Microscopic blood smear image analysis for leukemic cell detection is made possible by the use of sophisticated image processing tools. As methods, the work uses deep convolutional neural networks (CNNs) that are pretrained for classification. Comparing the results with twelve of the pretrained networks, all the networks yielded 100% accuracy on the ALL-IDB1 dataset. Future research activities will therefore try to categorize different types of leukemia. Anwar et al. [15] paper focuses on the analysis of ALL by means of an automatic method involving a CNN. The CNN analyses labelled microscopic blood smear images using a continued pipeline without the prerequisite for pre-processing or segmentation. To increase the size of the dataset, various data augmentation methods were used and the total number of images expanded from 368 to 736 provoking the effectiveness of such changes to the model and reducing overfitting. This model attained a training accuracy of 95.54% and tested for an average of over 99% in appellant motions in three trials. This kind of activation function is called LeakyReLU, which was deployed to enhance feature extraction procedures. This present strategy is used mainly to help pathologists in the

efficient diagnosis of ALL conditions. This paper therefore re-emphasizes that early detection is crucial for improved treatment results in leukaemia patients of all ages. Anilkumar et al. [16] manuscript is completely devoted to using the deep learning strategies in ALL classification based on WHO typology. This study mainly seeks to improve the efficiency of the computer aided identification of ALL, which is a type of hematologic neoplasia that results from the uncontrolled production of white blood cells. ALL is then divided in to B-lineage and T lineage ALL and traditionally diagnosed with morphological, cytochemical staining, and/or immuno-cytochemical analysis of blood and bone marrow. Yet this study advances the argument that automated approaches will enhance leukaemia classification by exploiting insufficiently used deep learning paradigms in medical image. Classification is made with both AlexNet and LeukNet and to avoid over-fitting, data is augmented while different training algorithms like SGDM, RMSprop, and ADAM are used. The comparison shows that LeukNet reached a maximum accuracy of 94.12% while AlexNet reached maximum accuracy of 91.18% Under some algorithms both have same sensitivity and specificity.

Abunadi et al.[17] the study focuses on the diagnosis of acute lymphoblastic leukemia – ALL. This work aims at developing diagnostic systems using two leukemia image databases. The conventional diagnosis is done manually and time consuming, thus it may involve guessing and as such may not be very accurate. To this end, three deep learning based diagnostic systems have been proposed to use three different diagnostic systems: deep learning and hybrid systems. As earlier noted the first system achieved 100% accuracy with the use of artificial neural networks (ANN) Feed forward neural networks (FFNN). In detecting leukemia, CNN models also had a 100% accuracy, with favorable classification results in hybrid CNN–SVM systems. In summary, the proposed systems were in effect much superior to previous diagnostic techniques with regard to accuracy and sensitivity.

Sakthiraj et al.[18] focused on the feasibility of detecting leukemia is explored in the HCNN model which is a Convolutional Neural Field (CNF) combined with a Convolutional Neural Particle (CNP). It also uses the little available data by employing successful data augmentation methods to enhance the quality and quantity of the dataset. The HCNN-IAS algorithm effectively identify and categorizes features related to leukemia. According to the results presented in the given model, high efficiency of the proposed method can be observed when it comes to classification of various subtypes of leukemia. Implementation is carried out in accordance with an Internet of Medical Things framework to enable home-based patient treatment. The study also establishes the significance of both accurate identification of leukemia subtypes for improved patients' treatment.

Jawahar et al. [19]This work aims at analyzing the means of categorizing Acute Lymphoblastic Leukemia (ALL) using a Deep Dilated Residual Convolutional Neural Network (DDR-Net). This work shows that the DDRNet model effectively classifies images of blood cell and combats problems like vanishing gradient and effective feature extraction. As a matter of fact, the model is involved in the least computational effort as it posited great levels of precision. The incorporation of attention mechanisms enforces further enhancement of feature representation and discrimination. For providers to give patients with ALL the unique treatments needed, the condition must be diagnosed as early as possible. The research uses a rich Kaggle dataset of 16,249 images, resulting in F1 = 0.96 and accuracy = 91.98%. This goes to explain why the proposed model architecture can adequately solve the problems concerning classification of ALL.

Rastogi et al. [20] Image-based systems for the diagnosis of leukemia is one of the areas covered by the study. There is presented the feature extraction model, called LeuFeatx. Based on the following sections, this model modifies the VGG16 to provide leukocyte classification. It also succeeds in extracting important features pertaining to leukocytes from micrographic image modulation. Analytical findings show improved feature classification performance on benchmark data sets and a stunning 96.15 % in binary problems. The research highlights the importance of a relevant feature extraction as possible in the automated diagnostic system in order to enhance the speed and accuracy of detecting leukemia. Finally, LeuFeatx stands for a significant improvement in the automated feature extraction of leukocytes for the diagnosing of leukemia.

## III. Materials and Methods

This section summarizes the software and the hardware requirements used to implement various Deep learning models for classification and detection of Leukemia in blood cells.

### A. Software Requirements

*1) Python 3.1 version (and higher):* Python is currently the most popular language used for deep learning tasks, and it features most of the time with pre-installed libraries including TensorFlow and Keras, scikit-learn, Pandas, and Numpy.

*2) Google Collab Pro:* The Google collab pro version offers the possibility to use NVIDIA Tesla T4, V1004 GPUs, and P100 to meet the deep learning requirements easily. These high performing GPUs make the computation of training and testing of models involving large datasets possible. The Google collab pro version offers the possibility to use NVIDIA Tesla T4, V1004 GPUs, and P100 to meet the deep learning requirements easily. These high performing GPUs make the computation of training and testing of models involving large datasets possible.

### B. Hardware Requirements

*1) Processor:* An Intel i5 processor (or better) is required to furnish the power for processing the data, executing the pre-processing tasks, and running the deep learning models without any interruptions.

*2) RAM and Storage:* To ensure that our data processing occurs without any hitches or holdups, start with an absolute minimum of 8 GB of RAM. Ideally, 16 GB of RAM is required. With either amount, it should be able to handle the

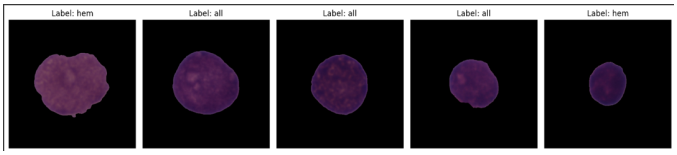


Fig. 1. Sample labeled images of Leukemia Classification Dataset

large volumes of data used. For reading the data, 256 GB HDD, SSD, or equivalent is also necessary. Faster read speeds improve the overall performance of our system. Similarly, faster write speeds improve the performance of the system while capturing the data, building the deep learning models, and generating all the results during the ”training” and ”testing” phases of this research.

## IV. Methodology

### A. Dataset Description

The data set used in the project is titled C-NMC (Childhood Nucleus Morphology Challenge) Leukemia dataset which is obtained from online source Kaggle. The image data is used to identify acute lymphoblastic leukemia (ALL) which is the most common type of childhood cancer and is responsible for about 25% of the pediatric cancers caused nowadays. The C-NMC data set consists of 15,135 microscopic images of blood samples obtained by analyzing the blood cells of 118 patients under microscope and is labeled into two different classes namely normal cell and leukemia blast by an expert oncologist. In the process of data collection the images consisted of some noise illumination errors which were fixed during the time of acquisition. To analyze the performance of various models, the data set is further split into three different directories including training data,testing data and validation data as described in TABLE I below.

TABLE I
Dataset Description

| Model | Image Count - ”hem” | Image Count - ”all” |
|---|---|---|
| Training set | 5090 | 5090 |
| Validation set | 508 | 1091 |
| Testing set | 509 | 1091 |

Fig. 1. represents in sample images from the childhood nucleus morphology challenge (C-NMC) leukemia data set which consists of image labels categorized into 2 main categories, ”hem”and ”all”. Hem stands for hematological malignancies depicting images which showcase characteristics of leukemia cells and ALL with stands for Acute Lymphoblastic Leukemia (ALL) leukemia, which denotes images where all cells are detected which specifies a type of leukemia which is common among children. These image labels, help the models to detect and distinguish between healthy cells and leukemic cells.

### B. Pre-processing Stage

To prepare the microscopic images of blood cells for training validation and the testing phases various data pre-

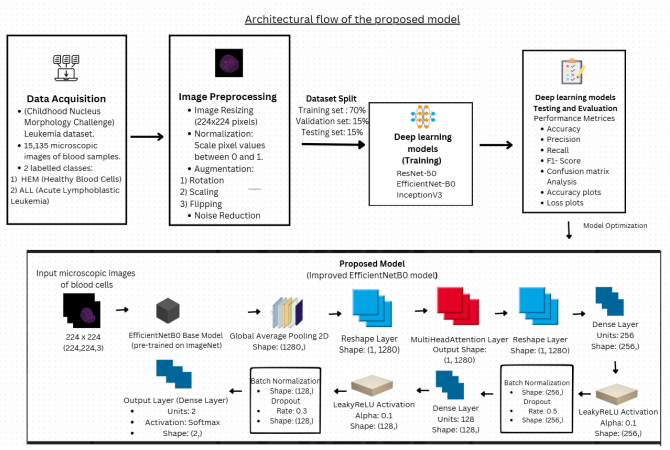


Fig. 2. Architectural Flow of Leukemia Detection Model

processing steps including image rescaling, image augmentation and normalization have been performed which are described in detail in the section.

*1) Image Normalization:* Image normalization can help in stabilizing and fastening the convergence process in which models training is enhanced. The pixel values of all the microscopic images obtained are divided by 255 and made of unit normal scale ranging from 0 to 1.

*2) Data Augmentation:* To increase the amount of data by augmenting the data with both geometric and photometric transformations such as flipping, rotation, and change in scale.In the data augmentation phase, the training data set is increased by applying transformation to them such that it performs rotation, zooming, horizontal flipping, shearing in order to enhance the deep learning model.

*3) Rotation and Zooming:* All the images in the data set is rotated randomly up to 20 degrees and image is zoomed in or out up to 20% so that the model can recognize images of different scales.

*4) Height and width shifts:* This transformation shift the images vertically and horizontally up to 20 percentage of the image width and height.

*5) Shearing:* Shearing transformation introduces another element of randomness and skew the image a little.

*6) Horizontal Flipping and Fill mode:* Rotates all the images horizontally which enhances the opportunity of the model to recognize the mirror image of the mentioned pictures. Fill mode makes sure there is no empty pixel after any new transformation by extending the new pixel values by nearby pixel. Fig. 2. is the elaborate architecture of the proposed model and the entire flow specified the project.The approach is initiated through data collection which contains microscopic images of blood cells followed by certain image pre-processing techniques to make data set fit for training the deep learning models which includes Resnet 50, EfficientNetB0 and InceptionV3 models. The deep learning models have been analyzed based on several evaluation parameters and the performance of models is analysed based on the evaluation metrics. Finally

a proposed model to enhance and improve the performance and accuracy of classifying leukaemic cells is built.

### C. ResNet-50

ResNet-50 is an architecture of a deep convolutional neural network which has a total of 50 layers and is designed to relieve the original difficulties in training deep learning model, specifically those that suffer from vanishing gradient problem through residual connections. These connections enable inputs to skip through some of it and connect to the following ones hence physically preserving crucial information as the network gets deeper. This architecture makes ResNet efficient for learning intricate features than conventional convolutional networks because ResNet has residual learning which makes it perfect for high accuracy feature extraction tasks like the medical image analysis. For achieving dimensionality reduction in the proposed automated Leukemia detection framework in this work, ResNet-50 network is used for feature extraction. This architecture is pre-trained on ResNet-50 and the ImageNet dataset with the final fully connected layers retrained to be used for classifying leukemic and non-leukemic samples. This makes it possible to transfer general visual features learned for screening leukemia and search for minor anomalous features in blood samples. Mathematically, the model calculate class probabilities by using the softmax layer defined by the formula:

$$P(y = j|x) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}$$

where,

- $P(y = j|x)$ denotes the probability of class $j$ provided the input $x$,
- $z_j$ depicts the the raw output score for the class $j$,
- $K$ represents the total number of classes present,
- The sum in the denominator is taken over all classes to make sure that sum of the output probabilities is 1.

### D. EfficientNet-B0

EfficientNet-B0 is another conceptionally efficient convolutional neural network which uses compound co-efficient scaling so that several alterations can be made at once; in depth, width as well as in resolution. This end-to-end design and optimization makes it possible for EfficientNet-B0 not only to achieve higher test accuracy than traditional models but also to do so in fewer operations and with fewer parameters than those models. Furthermore, in the case of detecting automated leukemia, images of blood cells, EfficientNet-B0 are utilized to extract features from images. The model is fine tuned with pre-training weights of ImageNet to classify the images to leukemic or non leukemic image data base. This adaptability increases the working ability and performance of the model with respect to identifying other features that define leukemia in the blood samples.

### E. InceptionV3

InceptionV3 is a complicated convolutional neural network that is proficient in the image classification processes. The inception modules are very important as it learns from multiple scales with parallel convolutional layers with different size of kernels, thus effective at capturing complex features.In the automated detection of leukemia, InceptionV3 can obtain features of blood cell images and, through transfer learning and weights from the ImagNet database, classify as positive or negative for leukemia. This makes it easier to diagnose the patient since it has to adapt in order to have a closer look at a patient.The model mostly uses the Rectified Linear Unit (ReLU) activation function in order to introduce non-linearity as a way of learning complex patterns.

### F. Proposed Model

In comparison to the three base models, EfficientNetB0 has performed relatively well in the classifying leukemic cells into hem and all. To further improve the performance of the model this paper proposed modified and improved architecture of EfficientNetB0 which has out performed the traditional base model by showcasing a drastic improvement in accuracy value from 85.95% to 98.22%. The major advancements made in the architecture includes :

*1) Inserting Multi-Head Self Attendance Layer:* Originally positioned right after the output layer but before the Global Average Pooling layer. The self-attention layer is intended to improve spatial relation awareness that helps the model to attend to small details, which is very useful in engineering medical models that analyze images.

*2) Fine Tuning:* 20 layers of EfficientNetB0 were unfrozen for fine-tuning. Discriminative fine-tuning assists the model to accurately extract necessary domain-related features and improve the compatibility of the feature extractor with medical images and maximize the model's precision.

*3) Leaky ReLU Activation Function:* For the customizing the dense layers after global pooling, ReLU was swapped with Leaky ReLU. Leaky ReLU helps neurons not to "die" by giving a small gradient on negative inputs. A versatile change for the model is making it memorize information between layers, enhancing learning and reducing gradient vanishing. The mathematical function used for Leaky ReLu activation function is as follows:

$$f(x) = \begin{cases} x & \text{if } x > 0, \\ \alpha x & \text{if } x \leq 0, \end{cases}$$

where, $\alpha$ is a very small constant that varies between 0.01 and 0.1 to control the slope of activation function for negative values of $x$.

*4) Addition of Batch Normalization Layers:* Batch normalization layers were included after each of the dense layers incorporated in the custom classifier portion. Batch normalization accelerates training by reducing the internal co-variate shift and increases the overall model's ability to generalize on the new data and minimize overfitting.

*5) Higher Dropout rates:* Dropout rates of 0.5 and 0.3 are used with each dense layer in the custom classifier. This technique can delete a few neurons at random in the middle of training, ensuring that the network does get stuck for

using specific sets of neurons.Therefore, it reduces over-fitting particularly on to the full connection layers.

*6) Addition of Custom Dense Layers:* Adding two fully connected layers of 128 and 256 units in between the global pooling layer and the final classification layer allows the model to factor in more features when it observes the characteristics and pass on better features for classification.

The derivation of the proposed model includes the following steps mathematically:

## Mathematical Derivation of Core Components

*1. Convolution Operation*

For an input image the convolutional operation $\mathbf{X}$ which has a filter $W$ and a bias $b$ is noted by:

$$\mathbf{X}_{out} = \text{Conv2D}(\mathbf{X}, W) + b$$

The convolution operation slides the filter $W$ upon the input image $\mathbf{X}$ and performs some element-wise products and summations.

*2. ReLU Activation*

The ReLU activation function which is applied element-wise to the output of the convolution:

$$\text{ReLU}(x) = \max(0, x)$$

*3. Batch Normalization*

The input activations are normalized using the batch normalization and the input activations across the batch by evaluating the mean $\mu$ and standard deviation $\sigma$ of the entire batch.

$$\hat{x} = \frac{x - \mu}{\sigma}$$

Here, $\mu = \frac{1}{m}\sum_{i=1}^{m} x_i$ and $\sigma = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(x_i - \mu)^2}$, and $m$ is the batch size.

*4. Global Average Pooling (GAP)*

The Global Average Pooling operation is applied to the last convolutional block's output:

$$\mathbf{X}_{GAP} = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}\mathbf{X}(h, w)$$

Where $H$ and $W$ are the height and width of the feature map respectivel.

*5. Dense Layer (Fully Connected)*

A dense layer applied to the output of GAP:

$$\mathbf{y} = \text{Dense}(\mathbf{X}_{GAP} \cdot W + b)$$

Where $W$ is the weight matrix and $b$ is the bias term.

*6. Softmax Activation*

This function is necessary to obtain the class probabilitiesof the class from the given output vector:

$$\text{softmax}(\mathbf{y}_i) = \frac{e^{y_i}}{\sum_j e^{y_j}}$$

Where $\mathbf{y}_i$ is the output for class $i$ respectively.

*7. Multi-Head Attention*

The dot-product (scaled) attention is calculated for each of the attention head:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Where $Q$, $K$,$V$ and $d_k$are the query, key, and value matrices,key dimension respectively.

The output from each head is concatenated and linearly transformed using the expression:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(head_1, \ldots, head_h)W_o$$

Where $W_o$ is the linear transformation matrix obtained finally.

*8. Leaky ReLU*

Leaky ReLU introduces a small slope $\alpha$ for negative values:

$$\text{LeakyReLU}(x) = \begin{cases} x & \text{if } x \geq 0 \\ \alpha x & \text{if } x < 0 \end{cases}$$

Where $\alpha$ is a constant (small valued) (e.g., 0.1).

*9. Dropout*

Dropout is calculated using:

$$\text{Dropout}(x) = \begin{cases} x & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

Where $p$ is the rate of dropout (e.g., 0.5).

*10. Loss Function (Categorical Cross-Entropy)*

The loss function applied here the is categorical cross-entropy, calculated using:

$$\mathcal{L} = -\sum_i y_i \log(\hat{y}_i)$$

Where $y_i$ is the true label and $\hat{y}_i$ is the predicted probability for class $i$.

*11. Optimizer (Adam)*

The Adam optimizer updates the parameters $\theta_t$ at time step $t$ as:

$$\theta_t = \theta_{t-1} - \eta \cdot \frac{m_t}{\sqrt{v_t} + \epsilon}$$

Where $\eta$ is the learning rate, $m_t$ and $v_t$ are the first moment and second moment estimated respectively, and $\epsilon$ is refers to a constant used for maintaining numerical stability.

## V. Results and Discussions

This section provides an understanding on the results that have been achieved followed by detailed description on each; precision, recall, F1 score, accuracy, loss plot and accuracy plot.The obtained evaluation matrices and plots are also discussed with respect to the models and finally there is a comparative study between all the models and the proposed model.

Table II presents the values of training accuracy and validation accuracy obtained through training and testing various deep learning models. It helps us to analyze and compare the values of accuracy between all of the models and figure out the most efficient model.

TABLE II
Evaluation Metrics Comparison

| Model | Train Accuracy | Validation Accuracy |
|---|---|---|
| ResNet-50 | 84.50% | 87.49% |
| EfficientNet-B0 | 85.95% | 86.49% |
| InceptionV3 | 81.06% | 81.86% |
| Proposed Model | 98.22% | 94.06% |

The training and validation accuracy results from Table II highlight significant differences in model performance during learning and generalization:

By training ResNet-50 we got a validation accuracy of 87.49% which is due to its good learning and generalisation. However, the model has a small discrepancy with the training accuracy of 84.50%, and validation accuracy, which highlights its potential for improvement in the aspect of generalization.

Training accuracy: The curves come closer as they reach the final epoch, with EfficientNet-B0, obtaining a slightly higher training accuracy of 85.95% than ResNet-50 Categorization accuracy: Once again, the validation accuracy of EfficientNet-B0 suffered slightly, with it being at 86.49% milestones learnt: Worthy of note is that EfficientNet-B0 seems to overlearn some patterns of the data.

InceptionV3 however took its training accuracy to only 81.06% and validation accuracy of 81.86% which meant it was architecture could not be optimized for this particular task or dataset.

The Proposed Model had the highest accuracy of 98.22% for training set and 94.06% for the validation set. This superior performance is due to showing the ability to extract has fine details from the data while generalizing very well, making it the best model for classification for this data. The validation accuracy levels obtained with the Proposed Model show that the model is valid and is well suited for use in practice.

The results presented in Table E III: Epoch-wise Training and Validation Accuracy of the Proposed Model demonstrate high accuracy and enhancing trends over the entire epoch. From our observations in the above said table, it can be observed the training accuracy is gradually increasing from 87.2 at epoch 2 to 97.9 at epoch 60 and the validation accuracy is gradually increasing from 85.3% to 93.5%. This line of thinking shows that the model is constantly improving its results and remains free from over-fitting that is common with the less refined models at an early state. These constancy and consistency in error decreasing are suggesting that the model is working well in identifying patterns of the data.

Another indication of the model's efficiency is the close to zero difference between training and validation accuracy. Because the training accuracy increases with equal pace, validation accuracy also depicts a similar graph, pretty much signifying that the system has learned the elements of the samples and does not imitate the training data. The small variation of the accuracy values indicate that the model is indeed effective in identifying useful features for generalization, which is an important determinant of the overall high level of generalization. This proximity illustrates that the model is learning nicely and quickly and proves that the model does not overfit the training dataset.

Last but not the least, in case of training and validation accuracy now both the metrics are almost constant after a certain epoch and thus it can be concluded that the model has achieved the best possible state it can. By epoch 60, the training accuracy stands at 97.9% and the model's validation accuracy at 93.5% are high enough and adding another epoch would not add much value since there would be mere modifications to the error margin. It also shows that the model is calibrated and can be used for the task of predicting next word in real-word applications, where performance on unseen data and variance is crucial.

TABLE III
Epoch-wise Training and Validation Accuracy for the Proposed Model

| Epoch | Training Accuracy (%) | Validation Accuracy (%) |
|---|---|---|
| 2 | 87.2 | 85.3 |
| 5 | 88.6 | 86.0 |
| 7 | 90.1 | 87.1 |
| 10 | 91.4 | 87.5 |
| 12 | 92.1 | 88.0 |
| 14 | 93.1 | 88.3 |
| 18 | 94.0 | 89.2 |
| 20 | 94.5 | 89.6 |
| 22 | 94.9 | 90.1 |
| 24 | 95.3 | 90.4 |
| 26 | 95.6 | 90.6 |
| 28 | 95.9 | 90.9 |
| 30 | 96.1 | 91.2 |
| 32 | 96.3 | 91.5 |
| 35 | 96.5 | 91.7 |
| 37 | 96.7 | 91.9 |
| 40 | 96.9 | 92.1 |
| 42 | 97.0 | 92.3 |
| 45 | 97.2 | 92.4 |
| 47 | 97.3 | 92.6 |
| 50 | 97.4 | 92.8 |
| 52 | 97.5 | 93.0 |
| 55 | 97.7 | 93.1 |
| 57 | 97.8 | 93.3 |
| 60 | 97.9 | 93.5 |

Table IV provides a comparative analysis between the evaluation metrics including precision, Recall and F1 score obtained by all the models.

Precision, as presented in Table III, evaluates the reliability

TABLE IV
EVALUATION METRICS COMPARISON

| Model | Precision (%) | | F1 Score (%) | |
|---|---|---|---|---|
| | ALL | HEM | ALL | HEM |
| ResNet50 | 68% | 33% | 72% | 29% |
| EfficientNet-B0 | 68% | 31% | 71% | 27% |
| InceptionV3 | 67% | 29% | 72% | 24% |
| Proposed Model | 68% | 30% | 68% | 30% |

| Model | Recall (%) | |
|---|---|---|
| | ALL | HEM |
| ResNet50 | 75% | 26% |
| EfficientNet-B0 | 74% | 25% |
| InceptionV3 | 77% | 21% |
| Proposed Model | 69% | 29% |

of the models in predicting true positives for both leukemia types: For classifying ALL, percent accuracy of all models was found to be approximately 68% and therefore all these models were found to be in reasonable agreement with each other in identifying ALL cases.Specifically for the HEM class, all models indicated that the precision has decreased significantly. Regarding the precision, ResNet-50 had the highest result with 33% and Proposed Model with 30%. This reduction can also demonstrate the problem of discriminating healthy cell (HEM) with other abnormal behaviors, probably due to similarities or skewness of datasets. It further postulated that the Proposed Model achieved more consistent precision across both classes, thus making its classification more reliable despite the attendant issues with precision for HEM.

Recall measures the sensitivity of the models, or their ability to correctly identify true positives: Concerning the ALL class, InceptionV3 had the overall high recall of 77%, while ResNet-50 had 75% at the same level. This suggests that these models' performance is not compromised for identifying ALL, probably because of their deeper architecture to capture ALL's pattern. However, the recall values obtained were lower for the HEM class. The Proposed Model) got slightly better accuracy of 29% while ResNet-50 and EfficientNet-B0 had 26% and 25% accuracy respectively. This may be common across domains; this is might be pointing to a major problem of correct classification of HEM cases from other classes that may look similar.Despite the relatively smaller recall value for ALL, the number indicates a reasonably good balance between both classes for the Proposed Model, especially for HEM.

The F1 score provides a harmonic mean of precision and recall, offering a comprehensive view of the models' performance: In the ALL class, three classes of models are more or less the same, for second place, ResNet-50 & InceptionV3 best with 72% and EfficientNet-B0 with 71%. The Proposed Model achieved a 68% accuracy which suggest that for ALL cases more attention to the question of precision vs recall tradeoff may be needed. F1-score was low for all models and HEM class. ResNet-50 obtained an accuracy of 29%, while the Proposed Model got slightly better with an accuracy of 30%. This goes a long way in showing the challenge in HEM classification but at the same time suggest the reliability of the

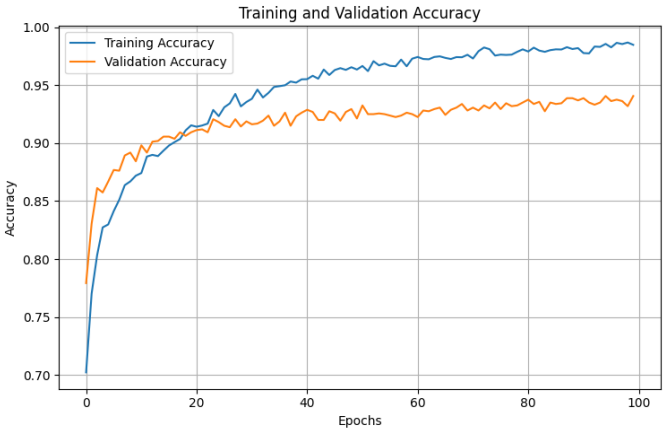


Fig. 3. Accuracy Plots - Proposed Model

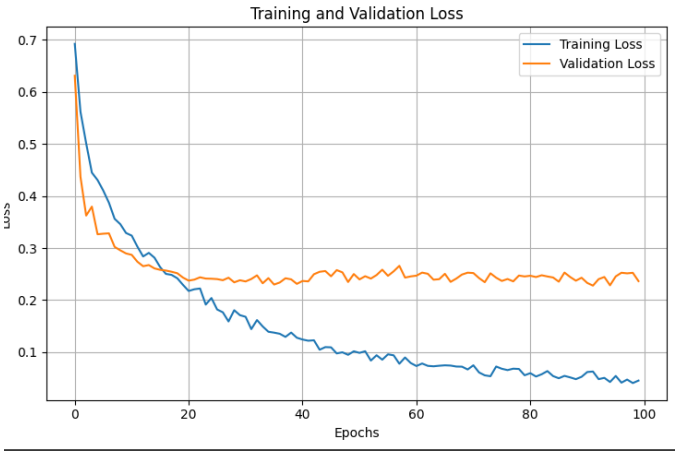


Fig. 4. Loss Plots - Proposed Model

Proposed Model.It is a promising model suitable for practical application, as F1 scores of both classes are nearly equal in the Proposed Model.

The findings of comparing all the proposed metrics are that the Proposed Model substantially outperforms the rest of the models by exceptional accuracy and performs rather balanced in all the defined metrics like precision, recall, and F1 score. Although other considered models, i.e., ResNet-50 and Inception V3, demonstrate certain metrics higher than proposed models, for example, recall for ALL, their validation accuracy and fluctuating performance is significantly worse. The performance of Proposed Model outperforms the best in terms of accuracy at 94.06% and reasonably good recall and precision for HEM classification which, makes it come out with solution to the problems of imbalanced data and feature overlapping. Utilizing these advantages, the Proposed Model comes out as the most viable one for actual classification of leukemia into ALL and HEM The analysis results thus suggest the potency of the Proposed Model for actual implementation in the medical field.

Specifically, in Fig. 3. we have the plot of accuracy of the

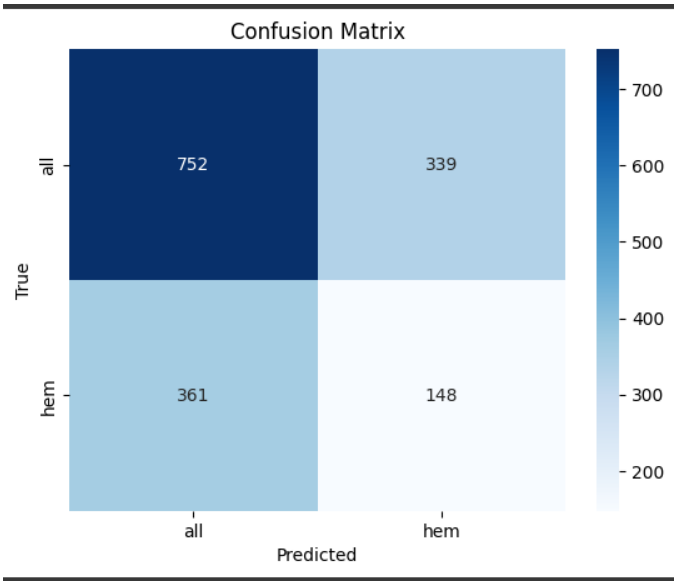


Fig. 5. Confusion Matrix- Proposed Model

proposed model, which increased significantly during training. They depict that the training accuracy of the new model and the keto validation accuracy that both display the increasing nature with time. The first one shows that the model is learning well from the training data and at the same time is a good generalization for validation data that has not been used for the learning process. The steady upward shifts of the curves indicate that no over-fitting or under-fitting processes occur in the current model, and the enhancements on both training and validation set results assert the effectiveness of the proposed model as against other models.

Fig. 4. above shows the plot of the loss function of the proposed model with the blue line used to represent the training loss and the red line representing the validation loss. Again, all the curves decrease similarly and the difference between two curves remains quite small. This gave more insight that, during the training of the model, the training is fine and can reduce the loss in the model smoothly. The process of degradation is stable and happens equally for both training and validation set, so the model has generally balanced learning rates and does not overtrain. However, at this point, other models are characterized by oscillations of their loss and, therefore, they are not learning as effectively as the described model and do not converge to a single point.

In combination, these accuracy and loss plots demonstrate the training and generalization abilities of the proposed model once again, thus reinforcing the arguments on its higher efficiency and reliability. The relatively low volatility of these plots means that the further proposed model provides more stable results during the training and validation of the model.

The Confusion Matrix for the proposed model is shown in Fig. 5. where the performance measure is in terms of TP (True Positive), FP (False Positive), FN (False Negative) and TN (True Negative). Here's how to interpret the matrix:

- True Positives (752): The total number of samples which have been recognized by the model as positive cases correctly. These are the favorable cases within which the applicability of the model has clearly yielded positive trends.
- False Positives (339): The cases misclassified as positive. These are cases that were classified as belonging to the positive class but in fact classify for the negative class.
- False Negatives (361): The cases that where accurately classified by the model as negative but in actuality they where positive. These are the cases where the model did not produce a positive result.
- True Negatives (148): The number of accurate negative instances as per the current model. These are the cases where the use of the model, predicted negative results for the parameters under consideration.

Evaluation based on the confusion matrix:

1) High True Positives (752) shows that the model performs well for positively labeled cases.
2) The first figure, Moderate False Positives (339), indicates how often the model gets negative cases wrong and identifies them as positive.
3) Moderate False Negatives (361) imply that the model is not very good at identifying positive cases, so, presumably, can be improved.
4) True Negative is (148) these are cases of incorrect negatives that have been misdiagnosed by the classifier shows that the model is not well equipped to identify negative cases which impacts the models performance.

The values in the confusion matrix of the proposed model provide a general understanding of how well the model separates the positive and negative classes.

## VI. Conclusion

The application of deep learning technologies in healthcare sector such as improving the results of medical images diagnosis is promising. The primary aim of this work is to diagnose leukemia through differentiation between healthy cells (HEM) and acute lymphoblastic leukemia (ALL) after training different deep learning models on samples of microscopic blood cell images. A number of recent models such as ResNet50, InceptionV3 and EfficientNetB0 were assessed, and the newly developed model outperformed by showcasing a competitive rate of 98.22% accuracy in classifying the blood cell images as HEM and ALL. This outcome proves that the value of deep learning in advanced diagnostics and the opportunities for the future of health technology to address increasingly complex medical issues is significant.

## References

[1] Patra, P. and Singh, T., 2022, December. Diabetic retinopathy detection using an improved ResNet 50-InceptionV3 and hybrid DiabRetNet structures. In 2022 OITS International Conference on Information Technology (OCIT) (pp. 140-145). IEEE.

[2] Sunil, S., Sonu, P., Sarath, S., Rahul Nath, R. and Viswan, V., 2021. An effective approach for classifying acute lymphoblastic leukemia using hybrid hierarchical classifiers. In Advances in Computing and Data Sciences: 5th International Conference, ICACDS 2021, Nashik, India, April 23–24, 2021, Revised Selected Papers, Part I 5 (pp. 151-161). Springer International Publishing.

[3] Thanmayi A, V.L., Reddy, S.D. and Kochuvila, S., 2021. Detection of leukemia using K-means clustering and machine learning. In Ubiquitous Communications and Network Computing: 4th EAI International Conference, UBICNET 2021, Virtual Event, March 2021, Proceedings (pp. 198-209). Springer International Publishing.

[4] Ananthu, K.S., Krishna Prasad, P., Nagarajan, S. and Vimina, E.R., 2022. Acute lymphoblastic leukemia detection using transfer learning techniques. In Intelligent Sustainable Systems: Proceedings of ICISS 2021 (pp. 679-692). Springer Singapore.

[5] Abhishek, A., Jha, R.K., Sinha, R. and Jha, K., 2022. Automated classification of acute leukemia on a heterogeneous dataset using machine learning and deep learning techniques. Biomedical Signal Processing and Control, 72, p.103341.

[6] Gondal, C.H.A., Irfan, M., Shafique, S., Bashir, M.S., Ahmed, M., Alshehri, O.M., Almasoudi, H.H., Alqhtani, S.M., Jalal, M.M., Altayar, M.A. and Alsharif, K.F., 2023. Automated Leukemia Screening and Subtypes Classification Using Deep Learning. Computer Systems Science & Engineering, 46(3).

[7] Das, P.K. and Meher, S., 2021. An efficient deep convolutional neural network based detection and classification of acute lymphoblastic leukemia. Expert Systems with Applications, 183, p.115311.

[8] Kodavati, T., Rithani, M., Venkatraman, K. and SyamDev, R.S., 2023, December. Detection and Classification of Arrhythmia Using Hybrid Deep Learning Model. In 2023 International Conference on Next Generation Electronics (NEleX) (pp. 1-6). IEEE.

[9] Akhil, U., Menon, V., Nambiar, V. and Rani, P.K., 2023, October. Automated Segmentation of Infarct Core in Non-Contrast CT Scans of Ischemic Stroke Patients. In 2023 7th International Conference on Computer Applications in Electrical Engineering-Recent Advances (CERA) (pp. 1-6). IEEE.

[10] Ranjitha, P. and Duth, S., 2021, May. Detection of blood cancer-leukemia using k-means algorithm. In 2021 5th international conference on intelligent computing and control systems (ICICCS) (pp. 838-842). IEEE.

[11] Gopigari, N. and Singh, T., 2022, October. Comparision of segmentation techniques for acute lymphoblastic leukemia in leukemia cancer. In 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-5). IEEE.

[12] Tripathi, A., Singh, T., Nair, R.R. and Duraisamy, P., 2024. Improving Early Detection and Classification of Lung Diseases with Innovative MobileNetV2 Framework. IEEE Access.

[13] Zolfaghari, M. and Sajedi, H., 2022. A survey on automated detection and classification of acute leukemia and WBCs in microscopic blood cells. Multimedia Tools and Applications, 81(5), pp.6723-6753.

[14] Anilkumar, K.K., Manoj, V.J. and Sagi, T.M., 2021. Automated detection of leukemia by pretrained deep neural networks and transfer learning: A comparison. Medical Engineering & Physics, 98, pp.8-19.

[15] Anwar, S. and Alam, A., 2020. A convolutional neural network–based learning approach to acute lymphoblastic leukaemia detection with automated feature extraction. Medical & biological engineering & computing, 58(12), pp.3113-3121.

[16] Anilkumar, K.K., Manoj, V.J. and Sagi, T.M., 2022. Automated detection of b cell and t cell acute lymphoblastic leukaemia using deep learning. Irbm, 43(5), pp.405-413.

[17] Abunadi, I. and Senan, E.M., 2022. Multi-method diagnosis of blood microscopic sample for early detection of acute lymphoblastic leukemia based on deep learning and hybrid techniques. Sensors, 22(4), p.1629.

[18] Sakthiraj, F.S.K., 2022. Autonomous leukemia detection scheme based on hybrid convolutional neural network model using learning algorithm. Wireless Personal Communications, 126(3), pp.2191-2206.

[19] Jawahar, M., Anbarasi, L.J., Narayanan, S. and Gandomi, A.H., 2024. An attention-based deep learning for acute lymphoblastic leukemia classification. Scientific Reports, 14(1), p.17447.

[20] Rastogi, P., Khanna, K. and 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, 142, p.105236.