ENSEMBLE LEARNING FOR PRECISE DIAGONSIS OF ACUTE LYMPHOBLASTIC LEUKEMIA

Sai Samyuktha N[1], Harshitha Devina Anto[1], Ayesha Shaik[2*] and A Balasundaram[3]

[1,2]School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India

[3]School of Computer Science and Engineering, Center for Cyber Physical Systems, Vellore Institute of Technology, Chennai, India

Email: saisamyuktha.n2020@vitstudent.ac.in [1], harshithadevina.anto2020@vitstudent.ac.in [1] ayesha.sk@vit.ac.in [2*] (corresponding author), balasundaram.a@vit.ac.in [3]

ABSTRACT

Acute Lymphoblastic Leukemia (ALL) is characterized as a cancerous condition originating in the bone marrow, impacting the B or T lymphoblasts. ALL is characterized by an uncontrolled growth of aberrant, immature lymphocytes and their progenitors, which leads to the replacement of bone marrow segments and other lymphoid organs and results in the development of specific disease patterns. Pediatric leukemia, the most prevalent form of pediatric cancer, is mostly caused by ALL, accounting for 75–80% of cases. Recent reports indicate that male age-adjusted incidence rates in India may vary by region and could reach as high as 10.4 cases per million, whereas for females, the incidence rates can reach up to 62.3 cases per million, depending on the specific area. This research develops an ensemble model trained with ALL image dataset from Kaggle, and involves transfer learning deep learning architectures - ResNet50, MobileNet, VGG16, Xception, and NASNet Large. The ensemble is further used to classify ALL images into four distinct stages: Benign, Early, Pre, and Pro. The developed ensemble model consistently performs well across different evaluation metrics, like F1 score, test loss, and accuracy of 96.87%. Ultimately, our research advances automated Acute Lymphoblastic Leukemia diagnosis, contributing to improved cancer screening and treatment outcomes.

Keywords: Leukemia, Deep Learning, Ensemble, CNN, ALL, ResNet50, MobileNet, XceptionNet, VGG16, and NasNet Large.

1.0 INTRODUCTION

Children are more commonly affected by acute lymphoblastic leukemia (ALL), as it has a lower survival rate in adults. The two types of lymphoblastic leukemia are B and T. Chromosomal abnormalities and genetic alterations play a significant role in the development of ALL, affecting the differentiation and proliferation of lymphoid precursor cells. Various subtypes of ALL have distinct genetic characteristics, which can inform risk assessment and targeted treatments. MicroRNA expression profiles have also been identified as potential biomarkers for diagnosing, differentiating, prognosing, and treating hematologic cancers. While the overall survival rate for ALL has improved in developed countries, survivors still face significant risks and complications from treatment. Therefore, there is a need to develop new diagnostic and prognostic methods and less toxic therapies for ALL patients. [1]

Leukemia can be categorized according to the rate of disease progression and severity. Chronic leukemia typically begins with leukemia cells that resemble normal white blood cells, but over time, they undergo changes and evolve into a more severe form of chronic leukemia. On the other hand, Acute Leukemia is characterized by leukemia cells that are unable to function normally. A swift deterioration of the patient's condition arises as the leukemic cells rapidly increase in the body.[2][3]

The precise and prompt identification of acute lymphoblastic leukemia (ALL), a common and potentially fatal disease, is essential for efficient treatment. The present diagnostic procedures for ALL are invasive, expensive, and time-consuming, which makes it difficult to provide patients with immediate treatment. Every year in India, over 10,000 new cases of acute lymphoblastic leukemia (ALL) are identified below the age group of 25 years of age (50% of the population). The majority of patients succumb to the disease. This stands in stark contrast to the cure rates of over 70% seen in Western countries. [4]

Several methods have been implemented for the detection and classification of cells into Acute Lymphoblastic Leukemia (ALL) or normal using machine learning models by the implementation of the CART algorithm. The study suggests that the CART algorithm [5] performed ten-fold cross-validation to assess the accuracy of predicting pediatric Acute Lymphoblastic Leukemia (ALL) cases. The CART model's overall accuracy was 87.4%, and its sensitivity and specificity were 0.89 and 1, respectively. The clinical, phenotypic, and environmental aspects were examined, and platelet count abnormalities were found to be a significant characteristic for pediatric ALL prediction. The results imply that machine learning techniques can be successfully used to enhance prognosis and treatment outcomes. Further studies are recommended to explore long-term environmental exposure and phenotypic measurements for a more comprehensive analysis of ALL subtypes using machine learning techniques.

In the study of a hybrid classification model [6], combining Inception v3 as the feature extractor and XGBoost as the classifier, the suggested model achieves a high F1 score of 0.986, exceeding other approaches in the literature, for the identification of ALL from microscopic white blood cell images. The use of XGBoost as the classification head improves performance across various CNN backbones. Attention maps are utilized to interpret the features learned by the model, highlighting the significance of the nucleus and center of the cell in classification. Future research aims to enhance interpretability and collect larger datasets to further improve the model's performance and applicability in medical diagnostics. Other deep-learning-based approaches are discussed in related work.

In this study, a Kaggle dataset consisting of 3,256 PBS images from 89 patients suspected of having Acute Lymphoblastic Leukemia (ALL) is utilized to advance the field of ALL diagnosis through image classification. The dataset consists of three different kinds of malignant lymphoblasts, comprising both benign and malignant classes. The main objective is to create an automated image classification system that accurately identifies and distinguishes the many subtypes of ALL utilizing cutting-edge machine learning and image processing techniques. The goal of this research is to raise the diagnostic precision and effectiveness of ALL screening, which will benefit patients and medical professionals and improve medical technology.

Our specific focus in this research is on creating an ensemble model that combines different CNN architectures to enhance the accuracy and reliability of classifying images as either ALL-infected or normal cells. By capitalizing on the unique strengths of each architecture and utilizing a majority voting approach, the ensemble model significantly improves the overall accuracy of diagnosing Acute Lymphoblastic Leukemia (ALL). This increased accuracy is crucial in facilitating more precise and effective treatment interventions for patients.

The ensemble model incorporates a diverse set of CNN architectures, including ResNet50, XceptionNet, VGG16, MobileNet, and NasNetLarge. Each architecture contributes distinctive characteristics and knowledge about image analysis, resulting in a more thorough and reliable examination of the image features. This comprehensive technique helps minimize the likelihood of incorrect classification or false predictions, ensuring reliable and precise outcomes.

By utilizing the majority voting approach, the ensemble model combines the outputs of the individual architectures to reach a consensus on the final classification decision. This consensus-based approach further enhances the reliability and accuracy of the classification, providing healthcare professionals with more informed and precise insights for patient care. To avoid confusion in the majority voting process, we have included an odd number of CNN models in the ensemble. This prevents ties and ensures a clear outcome, enhancing the reliability and robustness of our ensemble model. With improved accuracy, medical practitioners can make better decisions regarding treatment strategies, resulting in improved patient outcomes and potentially reducing the risks associated with misdiagnosis or delayed treatment.

Our research significantly contributes to the advancement of medical image analysis by showcasing the effectiveness of ensemble techniques in improving diagnostic accuracy, particularly in complex medical conditions like Acute Lymphoblastic Leukemia (ALL). The ensemble model's ability to leverage the strengths of multiple CNN architectures highlights the importance of combining diverse approaches to achieve superior performance in image classification tasks. This research paves the way for future developments in medical image analysis and reinforces the value of ensemble techniques in enhancing diagnostic accuracy, ultimately benefiting both healthcare professionals and patients.

2.0 LITERATURE REVIEW

To improve the diagnostic process for Acute Lymphoblastic Leukemia (ALL), a study [7] focused on instance segmentation of white blood cell microscopy images was conducted. The methods included contrast enhancement using the exposure fusion framework (EFF), image augmentation through flipping, lymphoblast annotation using VIA, training Mask R-CNN with transfer learning, testing the trained model for instance segmentation, and ALL classification based on majority vote. The dataset consisted of 301 multi-cell ALL images, divided into training and testing sets. The EFF addressed lighting issues, and image augmentation techniques were applied due to the limited dataset size. Table 1 describes the methods and architectures discussed in this paper.

Table 1. Summary of architectures proposed and their performance.

Architecture	Data Augmentation	Accuracy	Dataset	Description
CNN with VGG-16, VGG-19, and Xception	Yes	92.48%	C-NMC dataset	It did not classify ALL subtypes
Attention-based CNN with an efficient channel attention module (ECA) and VGG-16 architecture	Used to enhance the dataset	91.10%		Classifies healthy and ALL cells but only works with single cell images.
A probability-based weight factor used in a deep CNN approach combining MobileNetV2 and ResNet18 architectures		99.39% and 97.18%	ALL-IDB1 and ALL-IDB2 respectively	Still only classifies healthy or ALL cells but could also handle multi-cell images.
Pretrained AlexNet fine-tuned for ALL detection & classification	Yes	99.50% and 96.06%	ALL-IDB2	Utilized augmentation techniques but still relied on single cell images and did not provide segmentation results.
CNN classifier model along with an initial segmentation step to separate lymphoblast cells in the image.		97.78%	Amreek Clinical Laboratory	This method used conventional segmentation separate from the CNN architecture.

Based on the fusion of CNN features, the study on hybrid methods for diagnosing ALL [8] utilized two datasets: C-NMC 2019 and ALL_IDB2, aiming to enhance the quality of microscopic blood images for improved diagnosis. Image enhancement techniques were applied, including color adjustment, noise removal, and edge enhancement of white blood cells. WBC cell segmentation was achieved using an active contour technique, and a hybrid strategy combining deep learning and machine learning was proposed for cell classification. Data augmentation, segmentation, and feature fusion techniques were employed, resulting in high accuracy and performance. The study emphasized the effectiveness of hybrid systems and the importance of these techniques.

The research [9] used a dataset of 130,002 cancer images categorized into eight classes and preprocessed for quality. Kaggle Notebook was utilized, offering tools, datasets, and GPU acceleration. The proposed method comprised a deep learning framework where DenseNet201 achieved the highest validation accuracy of 99.20%, followed by MobileNetV3 with 99.05% and VGG16 with 98.21%. DenseNet201 excels in accuracy, VGG16 handles objects of different scales efficiently, and MobileNetV3 performs well with limited training data. DenseNet201 achieved the highest accuracy among the three architectures. Further validation and collaboration are needed for reliable and safe clinical implementation. Computational efficiency, model size, and other architectures like ResNet and Inception can be explored for future work in multi cancer image classification.

The dataset collection, image pre-processing, feature extraction using pre-trained CNN models, feature selection using various methods, and optimization using nature-inspired algorithms make up the research methodology of blood cancer classification using deep CNN with optimized features [10]. The study focuses on applying machine learning approaches to categorize ALL. Different classifiers were employed, with the ResNet50 architecture and Logistic Regression achieving the highest accuracy of 99.53%. Accuracy was further increased by feature selection methods like Particle Swarm Optimization and Cat Swarm Optimization. Future work includes developing a real-time classification Android application and adapting the model for small IoT devices.

The proposed model [11] in CNN-Assisted Automatic White Blood Cancer Type Detection from Bone Marrow involves several steps, including image preprocessing, classification using the K-means algorithm, convolutional layer application, pooling layer, and a fully connected layer for classification. The model outperformed current algorithms in predicting cancer types with a high accuracy of 97.2%. However, the study highlights the need for further experiments with larger databases. The implementation was done using MATLAB 2013a. Future improvements include using the Fuzzy means algorithm, increasing the number of convolutional layers, and considering data points that may belong to multiple clusters.

In a web-based approach for detection of ALL [12], the ALL-IDB1, ALL-IDB2 microscopic blood samples are passed through the Median filter and Wiener filter for Image enhancement. Next, by utilizing K-means clustering and the L*a*b color space, image segmentation is then completed. The images are classified using a CNN model and deployed on a webpage designed using Flask API. Geometrical, Texture, Colour, and Statistical Features are used for classifying the cancer blood cell. The CNN network achieves high specificity (96.22%), recall (96.42%), precision (98.18%), and accuracy (96.36%) in classifying normal and ALL images, outperforming SVM, Naive Bayes, and VGG-16 models and demonstrating improved results.

A weighted ensemble of various CNN models, with weights set based on performance parameters, is used in the proposed framework for the automation of Acute Lymphoblastic Leukemia (ALL) detection [13]. Pre-processing methods and data augmentations are used to improve network generalization. The framework produces positive results in the preliminary test set, including a weighted F1-score of 88.6%, balanced accuracy of 86.2%, and AUC of 0.941. Addressing class imbalance and further enhancing the classification outcomes, notably for the Hem class, are future research topics. Future research is advised to use adversarial networks and examine the effects of data imbalance.

Initially, the White blood cells are detected from the ALL IDB 2 Dataset [14] which contains 260 cell images with a 50% distribution between normal and affected cells. Z-score, min-max, and grey-scaling are used to extract and normalize Shape, Color, and Texture features. The suggested approach distinguishes between healthy and damaged cells to address a binary classification problem. Different classifiers, including Naive Bayes, Support Vector Machines, Decision Trees, and k-Nearest Neighbor (k-NN), were examined with varying amounts of features. The classifiers that were selected produced acceptable outcomes when applied to previous pattern recognition problems. With the dataset randomly divided into ten folds, the models' performance was assessed using k-fold cross-validation, and they all produced good classification accuracy, especially K-NN.

An automated method for diagnosing Acute Lymphoblastic Leukemia (ALL) using microscopic images of blood cells was proposed in [15]. The method involves a three-stage classification scheme, utilizing a pre-trained neural network for feature extraction, PCA for dimensionality reduction, and a machine learning algorithm for binary classification. Through extensive experiments, the proposed approach outperforms previous works on the same dataset, achieving high accuracy. Future work includes exploring residual features from different convolutional layers, using other pre-trained networks, and applying image data augmentation techniques to address class imbalance.

Another study [16] presents a Hybrid Fuzzy C Means clustering-based Random Forest Classifier for classifying Acute Lymphocytic Leukemia as infected or normal. The proposed approach performs better than SVM, KNN, ANN, CNN, and Naive Bayes algorithms, with accuracy, sensitivity, and specificity values of 99.06%, 99.4%, and 97.8% respectively. Data augmentation techniques, including mirroring, shifting, and rotating, were used to generate diverse versions of the real dataset images, addressing limited training data and preventing overfitting. The proposed system utilizes the augmented Image Datastore function to resize, mirror, and rotate the training image data, effectively expanding the dataset and mitigating overfitting. The microscopic blood sample image is transformed from RGB to CMYK and Histogram equalization is performed afterward during image preprocessing. The proposed classification method for Acute Lymphocytic Leukemia was implemented using Matlab R2018 tools and compared to other related methodologies such as VCGNet, GLCM, CNN, MBAML, and HALM. To enhance classification accuracy and prevent overfitting, images from different sources, including ALL IDB and cancer imaging archives, were augmented in the initial stage.

This study [17] suggests a deep neural network-based model (ALNett) detects ALL by differentiating immature leukemic blast cells from healthy ones using depthwise convolution techniques by analyzing microscopic images of cells. In order to accurately forecast Acute Lymphoblastic Leukemia from these microscopic images, structural and contextual information is retrieved by convolution, max-pooling, and normalization of the relevant cluster layers that are used in extracting powerful local and global features. In order to learn the hierarchical properties required for exact categorization, the CNN architecture employs multiple stacked convolution layers. These layers are then modified for better classification by batch normalization added between these stacked clusters. The learning process and weights are improved by normalization, which results in a large reduction in the features present across the stacked hierarchical clusters and a faster classification. By achieving the highest classification accuracy of 91.13% and an F1 score of 0.96 with less computational complexity, the ALNett model performed better than other pre-trained models.

Acute lymphoblastic leukemia microscopic images from the C-NMC 2019 database, involving 101 cases, including 60 infected cases with ALL and 41 normal instances, are analyzed using an ensemble technique in this work [18]. In three stages, this information was utilized to categorize both normal WBCs and ALL cells. To handle the oversampling, image pre-processing techniques are applied to the images. The CNN architecture, which consists of the five networks ResNet-101, GoogleNet, SqueezeNet, MobileNetV2, and DenseNet-201, extracts deep spatial features from the image data. In order to extract long-distance dependent information features or temporal characteristics for better feature learning, these features are then transferred to a gated recurrent unit (GRU)-bidirectional long short-term memory (BiLSTM) architecture. Lastly, classification and optimization are performed using a multiclass SVM classifier and a softmax function. Using the multiclass SVM classifier on the test dataset, the DenseNet-201 model achieved an F1 score of 96.23% and an accuracy of 96.29%.

The suggested model [19] uses a lightweight, robust model based on EfficientNet-B3 that uses depth-wise separable convolutions to categorize immature ALL cells at an early stage. Tested on the two datasets C_NMC_19 and ALL datasets, the proposed model for classifying pictures outperforms the benchmark DL and other ensemble classifiers. The binary class dataset C_NMC_19 is utilized to distinguish between ALL cells and normal cells. The performance of the other pre-trained classifier models was compared to and evaluated using the ALL dataset, a multi-class dataset. According to comparative research, the EfficientNet-B3 greatly outperformed other methods for classifying leukemia into binary and multi-class categories, achieving detection rates of 95.62% and 97.27% for the binary and multi-classification datasets, respectively. EfficientNet-B3 improved the accuracy for both binary and multi-class classification based on the analysis of a multi-class dataset, attaining an accuracy of 96.81%, precision of 97.27%, recall of 97.87%, and F1 score of 97.57%.

This paper [20] introduced an ensemble deep-learning model combined with computer-assisted diagnosis methods to identify leukemia. Three models—VGG11, ResNet18, and ShufflenetV2—were trained and optimized using the ISBI 2019 C-NMC dataset. The weighted averaging technique is then used to form an ensemble, which is then evaluated using a binary classification criterion. An F1 score of 87.40% and an overall accuracy of 87.52% were generated by the proposed approach. As a result, it performs better than the majority of other methods for the same dataset.

In this study [21], an ensemble hybrid model system with deep learning techniques is used to implement a leukemia detection system that is intended for use in a real clinical setting. This study creates two Hybrid models (trained at 1 epoch and 30 epochs), four Ensemble models (two Ensemble models of VGG19 and ResNet50 and two Ensemble models with an additional Hybrid model), four pre-trained models using ImageNet (VGG19 1 epoch, VGG19 30 epochs, ResNet50 1 epoch, and ResNet50 30 epochs), and four Hybrid models. Transfer learning allows the models to be improved on the leukemia domain much faster, hybrid models help identify different feature patterns by processing images through two different architectures. ResNet50 30 epochs, with an accuracy of 84%, is the independent model with the highest accuracy. Hybrid 30 epochs is the most accurate Hybrid model, with an accuracy of 84%. Ensemble 4 (Hybrid 30 epochs, VGG19 1 epoch, and ResNet50 1 epoch) was 2% more accurate than Ensemble 2, the second-best model, with an accuracy of 86%.

The automated deep learning weighted CNN ensemble for improved ALL classification is the main topic of this paper [22]. The ensemble candidates' respective metrics, such as the F1-score, area under the curve (AUC), and kappa values, are used to estimate the weights. The pre-trained models Xception, VGG-16, DenseNet-121, MobileNet, and InceptionResNet-V2 make up the ensemble. Several data enhancements and pre-processing techniques are utilized to increase the generalization of the network. The C-NMC-2019 ALL dataset was used to train and test the model to differentiate between images of ALL cells and images of normal cells. The proposed weighted ensemble model had a weighted F1-score of 89.7%, an AUC of 0.948, and a balanced accuracy of 88.3%.

This research [23] has developed a classification model trained by microscopic blood smear images using transfer learning techniques. Classification of Normal over B-ALL White Blood Cancer Microscopic Images of Malignant Cells from SBILab's publically accessible website served as the dataset for the model's training. To distinguish leukemia cells from healthy cells, the model combines Mobilenet V2 and Convolutional Neural Networks (CNNs). According to the study, the model has a 96.58% overall accuracy rate, a 95.17% sensitivity rate, and a 98.58% specificity rate.

A novel deep learning framework (DLF) based on convolution neural networks is proposed in this study to detect ALL [24]. Real-time applications can be used with this framework. This model has 41,626 tuning parameters, which is comparatively fewer learnable parameters than the other pre-trained models, and has an average accuracy of 98.17%.

In order to classify ALL smear images by extracting B and T cell lymphocytes, this paper [25] uses a pre-trained CNN AlexNet as well as a comparison with LeukNet, a custom-made deep learning network. For comparing each of their efficacy of classification, three training algorithms SGDM, RMSprop and ADAM were used. SGDM algorithm had better performance with LeukNet with an accuracy of 94.12% in classifying B and T cell lymphoblasts.

This paper [26] uses a computational deep learning technique for pattern identification that monitors the white blood cell count, compositional changes, and protein alterations that may be associated with ALL. It does this by employing mathematical algorithms to identify distinct patterns on images from large datasets. The management of chemotherapy by the ALLDM model was 81.53% (DDS) and 87.92% (SDS), the management of stem cell transplantation was 79.16% (DDS) and 94.31% (SDS), the management of radiation treatment was 63.77% (DDS) and 87.37% (SDS), and the management of targeted therapeutic medications was 88.92% (DDS) and 85.86% (SDS).

3.0 PROPOSED SYSTEM

The proposed work for our research study is presented in this section. A flow chart of the entire process is illustrated in Figure 1., which aims to perform the classification of microscopic blood cell images and analyze if diagnosed with Acute Lymphoblastic Leukemia. Firstly, the dataset is collected and split into train, test, and validation sets. Image data generators are created for train, test and validation. Next, the models (ResNet50, VGG16, MobileNet, Xception, and NasNet Large) are programmed and parameters like the number of training epochs and image input dimensions are set. Finally, compile all models and train them with train and validation datasets. Export and save all the trained models to combine them into an ensemble. Perform the major voting technique and retrieve the output of the classification by testing the ensemble model.

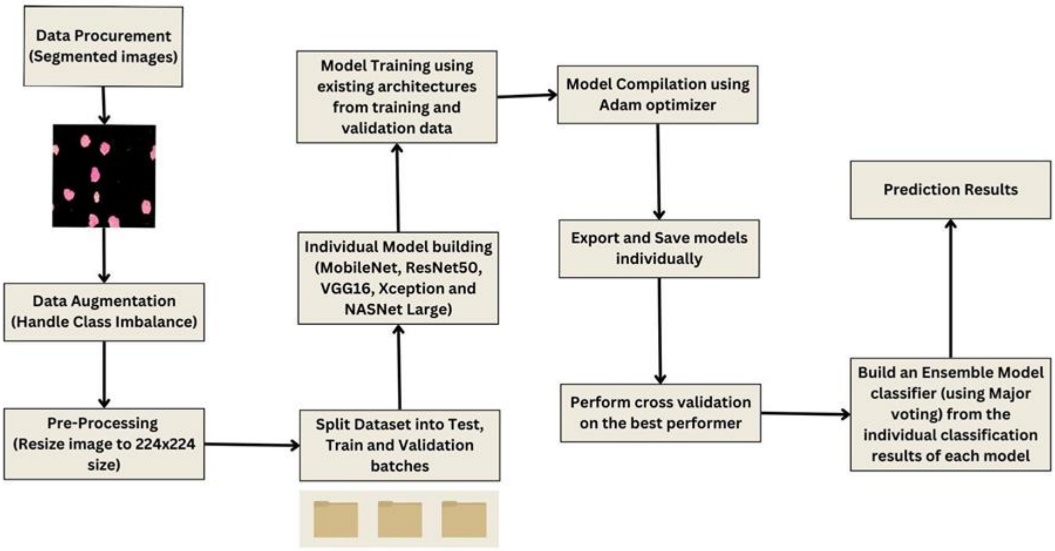

Data Procurement (Segmented images)
Data Augmentation (Handle Class Imbalance)
Pre-Processing (Resize image to 224x224 size)
Split Dataset into Test, Train and Validation batches
Individual Model building (MobileNet, ResNet50, VGG16, Xception and NASNet Large)
Model Training using existing architectures from training and validation data
Model Compilation using Adam optimizer
Export and Save models individually
Perform cross validation on the best performer
Build an Ensemble Model classifier (using Major voting) from the individual classification results of each model
Prediction Results

Figure 1. Flowchart showing the Proposed Methodology.

3.1 Dataset Procurement

The dataset consists of two folders - ‘Original’ and ‘Segmented’. The pre-processed images from the ‘Segmented’ folder of the dataset are split into train, test and validation batches. The description of split and the batches is tabulated in Table 2. The split batch is exported as such in gdrive and saved as Folders named - Test, Train and Validation. This split is used for training & improvising the model individually.

Table 2. Dataset Split Description.

Batch	Total
Test	330
Train	2263
Validation	649

3.2 Handling Class Imbalance

From Table 2, it is clearly evident that the number of image samples of ‘Benign’ class images were significantly lesser than the number of image samples of other classes leading to class imbalance. Hence, random samples of images were selected from the ‘Benign’ class as well as ‘Pro’ class to match the number of samples to a closer range of values. Using data augmentation techniques, selected images were processed with random rotation, shear, zoom and horizontal flip transformations to produce a larger number of samples. Table 3 depicts the improvised number of sample data from Table 2 after data augmentation.

Table 3. Split Description after Data Augmentation.

Batch	Class Label	Number of Images	Total
Test	Benign	98	385
	Early	99	
	Pre	97	
	Pro	91	
Train	Benign	682	2677
	Early	689	
	Pre	674	
	Pro	632	
Validation	Benign	195	764
	Early	197	
	Pre	192	
	Pro	180	

3.3 Development of Pre-Trained Models

A combined model incorporating different CNN architectures, namely ResNet50, VGG16, MobileNet, Xception, and NASNet Large, is constructed. The final classification is determined through majority voting, where the collective decision of the different models is considered as illustrated in Figure 2.

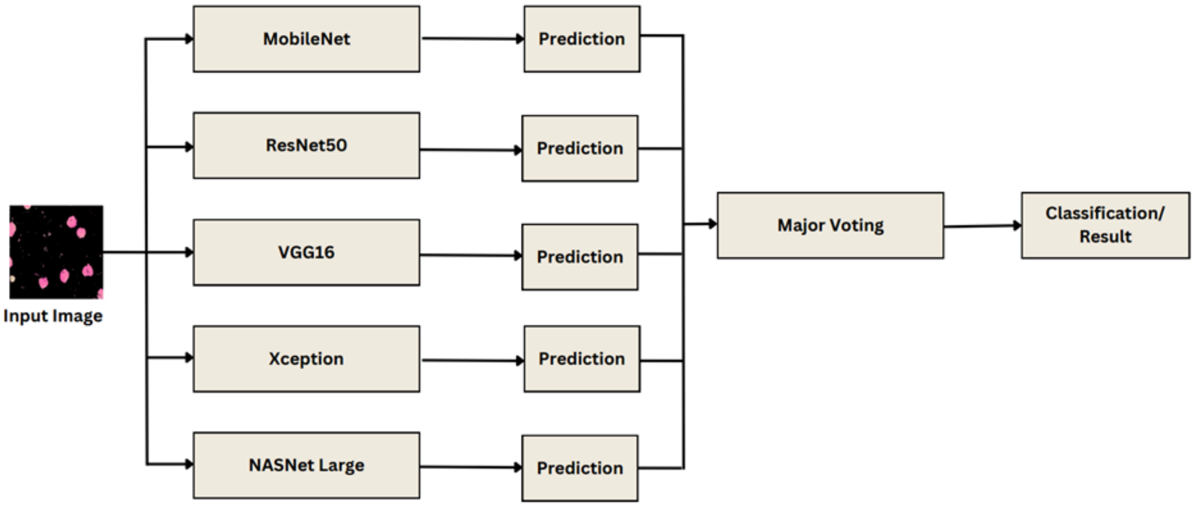


Figure 2. Ensemble Model.

MobileNetV2 [27] is a convolutional neural network architecture designed for mobile devices to operate at their best. The residual connections take the form of an inverted residual structure, occurring between the bottleneck layers. In order to filter features and add non-linearity, lightweight depthwise convolutions are utilized in the intermediate expansion layer. The MobileNetV2 architecture is made up of 19 residual bottleneck layers after an initial fully convolutional layer with 32 filters.

ResNet50, also known as Residual Networks [28], takes a unique approach to learning by focusing on residual functions with respect to the layer inputs. Rather than expecting each set of stacked layers to directly learn the desired mapping, ResNets allow these layers to learn a residual mapping. This is achieved by stacking residual blocks, resulting in networks such as ResNet-50 with fifty layers composed of these blocks.

In a formal sense, h(x) denotes the required underlying mapping. The objective of the stacked nonlinear layers is to learn a different mapping, f(x), which is defined as f(x) = h(x) - x. ResNets allow the network to learn the residual information by recasting the original mapping into the form of f(x) + x.

According to empirical data, ResNets are simpler to optimize and can obtain higher accuracy by significantly increasing the network's depth.

VGG16 is a CNN model [29] developed with a 16-weighted layer deep neural network and 138 million parameters. It consists of 5 max-pooling layers, 3 dense layers, and 13 convolutional layers. It has a 224 x 224 input tensor with three RGB channels. In the design, the max pool and convolutional layers are arranged in a pattern as a 3 x 3 filter with stride 1 and a 2 x 2 filter with stride 2 respectively, with the softmax layer as its final layer.

Xception is a modified depth-wise separable set of convolutions, its an extreme version of the Inception model [30]. The filters are initially applied to each depth map by the model before compressing the input space using 1 x 1 convolution and applying it across the depth maps.

Neural Search Architecture (NAS) Network is composed of blocks of normal and reduction cells[31], which is searched using reinforcement learning search techniques. An encoder and a decoder are present in the network, followed by a classification layer. Reduction cells return feature maps that have had their height and breadth decreased by a factor of two as opposed to normal cells, which return feature maps with the same dimensions. Instead of building the block using hand-crafted conclusions, NASNet often uses the controller RNN to find the optimal combinations from a collection of operations to construct a cell with the best performance.

4.0 EXPERIMENTAL RESULTS AND DISCUSSION

The proposed framework was created using Python and a variety of its tools, including tensorflow and numpy. The ensemble model's development environment was Google Colab with a GPU processor. The performance of the models has been assessed using performance indicators like accuracy, F1-score, precision and recall.

4.1 Dataset Description

Peripheral blood smear (PBS) images from 89 patients suspected of having Acute Lymphoblastic Leukemia (ALL) make up the dataset [32] used in this study. The bone marrow laboratory staff at Taleqani Hospital in Tehran, Iran, skillfully prepared and stained these images. The dataset had a total of 3256 PBS images that were split in the following manner as shown in Figure 3. The segmented images were split into two major categories: benign and malignant. Hematogones make up the benign class, while the ALL group—which is further divided into three subgroups of malignant lymphoblasts: Early, Pre-B, and Pro-B ALL—represents the malignant class.

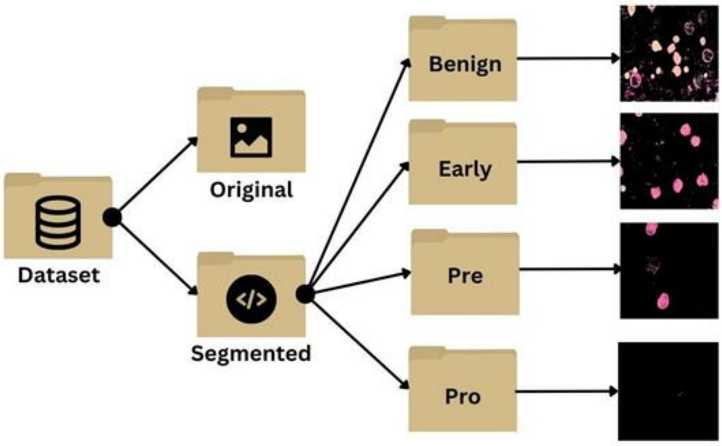


Figure 3. Dataset Structure.

Each image was taken using a Zeiss camera mounted on a microscope at a 100x magnification and saved as JPG files. To provide more information and facilitate analysis, the dataset also includes segmented images. Segmentation was performed using a color thresholding technique in the HSV color space, which allowed for the isolation and highlighting of specific regions of interest within the images. The definitive determination of cell types and subtypes within the dataset was made by a specialist using the flow cytometry tool. This dataset serves as a valuable resource for studying and developing algorithms for the automated diagnosis and classification of ALL based on peripheral blood smear images. The dataset consists of two directories:

- Original: consists of raw images
- Segmented: consist of pre-processed images

Segmented images are used for model training and classification. Acute Lymphoblastic Leukemia can be categorized into four classes - Benign, Early, Pre, Pro. The dataset structure, labels and the corresponding description is given in Table 4.

Table 4. Dataset Labels and Description.

S. No.	**Class Label**	**Number of Images**		**Class Label Description**
		Original	**Data Augmented**	
1	Benign	504	975	Benign ALL subtype includes hematogones, non-threatening lymphoblasts found in healthy individuals, particularly children, during normal immune system development.
2	Early	985	985	Early pre-B ALL is an intermediate stage of lymphoblast development, showing partial maturation towards pre-B cells. They exhibit specific markers and genetic abnormalities distinguishing them from other subtypes.
3	Pre	963	963	Pre-B ALL cells are more mature than early pre-B cells but still exhibit incomplete maturation. They display specific surface markers and genetic alterations. Pre-B ALL is the most common subtype of childhood ALL.

4	Pro	804	903	Pro-B ALL is an early stage of lymphoblast development, preceding the differentiation into pre-B cells. Pro-B ALL cells lack certain characteristic markers seen in more mature subtypes and may possess distinct genetic abnormalities.
Total		3256	3826	

4.2 Pre-processing of Segmented Images

The dataset source had segmented images of the original PSB images, this project uses these preprocessed images to extract distinct features efficiently. These images are obtained by enhancing the contrast, eliminating noise and modifying its size and variability of the original images; this is done so that the model can quickly learn the features. Added to this, some image processing techniques are also applied to systematically achieve the same and improve the model's accuracy. Using the Python Imaging Library (PIL), the images of all 4 stages of ALL: Benign, Early, Pre, and Pro, are loaded. By introducing contrast adjustment, the leukemia stains inside the cells are made darker. Images are then resized and pixel values are normalized. Next data augmentation is done on the images by an Image Data Generator that randomly applies rotation, shifting, flipping, zooming, on the images during training, and the resultant is converted into a grayscale format using the OpenCV libraries. Otsu's thresholding aims to take the grayscale histogram obtained as input and identify an optimal threshold value that segregates the image into black and white with maximum inter-class variance. For image pixels whose saturation is greater than the optimal threshold, are set to white whereas the others are set to black. The resulting segmented images can also be modified by changing the lower and upper bounds of different color channels in the image, based on the user's needs. The images and their respective segmented illustrations are represented below in Figures 4,5,6 and Figure 7.

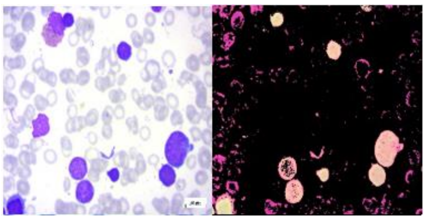


Figure 4. Original (left) and Segmented (right) image of Benign class.

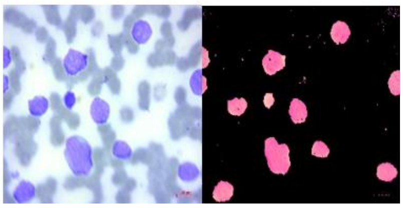


Figure 5. Original (left) and Segmented (right) image of Early class.

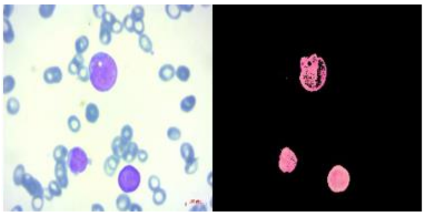


Figure 6. Original (left) and Segmented (right) image of Pre class.

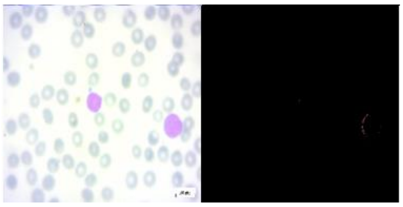


Figure 7. Original (left) and Segmented (right) image of Pro class.

4.3 Performance Analysis

The batch size for the ensemble model was set to 32 and the image input size is 224 x 224. The number of epochs set for ResNet50 is 15, Xception is 10, VGG16 is 5, MobileNet is 10 and for NASNetLarge its 15 epochs.

A comparison between the performance of different models used for the final majority-voting Ensemble model:

Table 5. Comparison of 5 CNN Architectures.

S. No.	Model Name	Accuracy	Precision	Recall	F1 Score
1	MobileNet	0.96	0.97	0.95	0.96
2	ResNet50	0.84	0.85	0.85	0.84
3	Xception	0.93	0.95	0.93	0.94
4	VGG16	0.93	0.93	0.93	0.93
5	NASNetLarge	0.93	0.96	0.93	0.94
6	Ensemble	0.96	0.97	0.95	0.96
7	MobileNet after cross validation (model 1)	0.84	0.80	0.85	0.81
8	MobileNet after cross validation (model 2)	0.96	0.97	0.95	0.95

From Table 5 we infer the evaluation metrics for different models. ResNet50 has relatively low scores across all metrics, suggesting lower overall performance. Xception, VGG16, NasNetLarge perform almost at the same level with moderate accuracies, their precision, recall, and F1 scores are about the same. Since MobileNet was the best performer of the entire ensemble, by comparing with its other metrics as well, the ensemble model thereby achieved the highest accuracy of 96% with a high precision, recall, and F1 scores. Precision is the percentage of the instances predicted as positive by the model are actually true positives, suggesting a high rate of false positives. The ensemble model uses a major voting technique to determine its final performance. The majority-voting Ensemble model achieves 96.87% accuracy, 97.91% precision, 95.83% recall, and 96.64% f-score. The individual predictions of all models are as shown in the figures. Since MobileNet is the best performer among the 5 models used in the ensemble, K-fold cross validation techniques are applied to prevent overfitting. The result after applying cross validation resulted in 2 models in which the 2nd model gave the highest accuracy of 96%.

Model predictions of each model and the validation of the same is shown in Figures 8 to Figure 15. Images titles represent the prediction class label. Color of the image title represents the correctness of the prediction - red signifies wrong prediction and green signifies correct prediction. Figure 16 to Figure 23 represents the confusion matrix of the individual models.

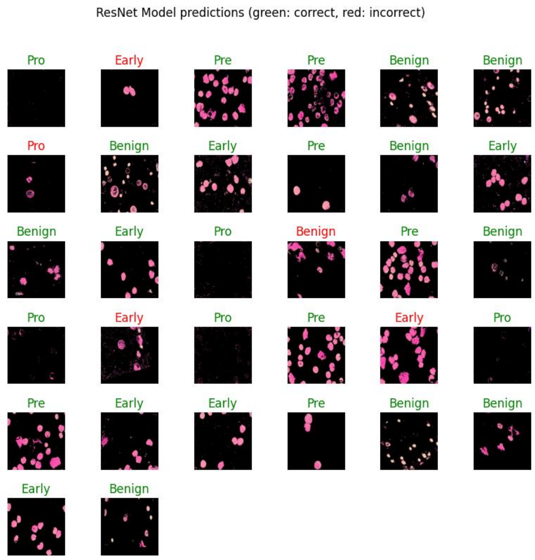


Figure 8. Predictions of ResNet50 Model.

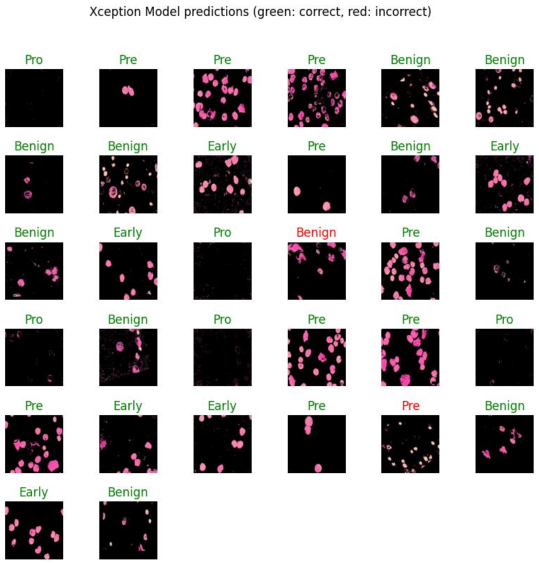


Figure 9. Predictions of Xception Model.

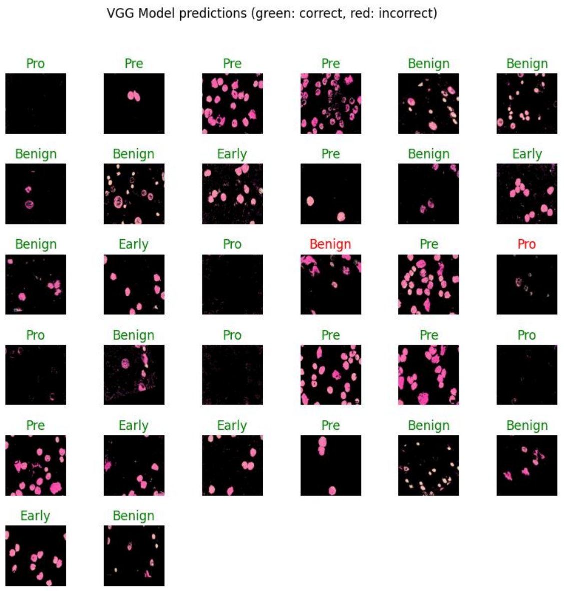


Figure 10. Predictions of VGG16 Model

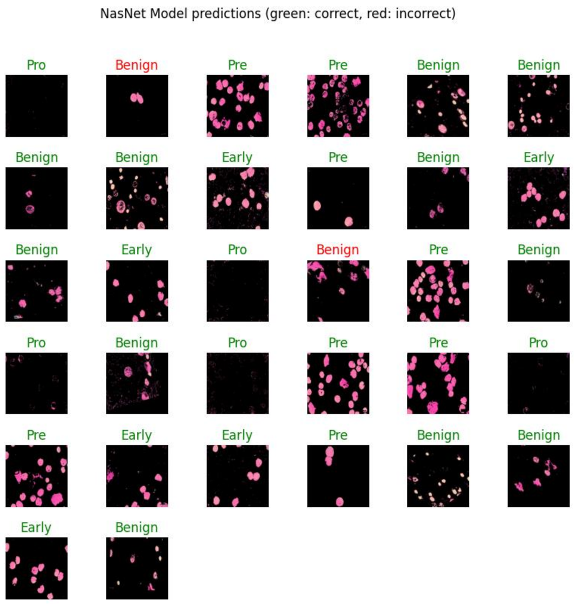


Figure 11. Predictions of NasNetLarge Model

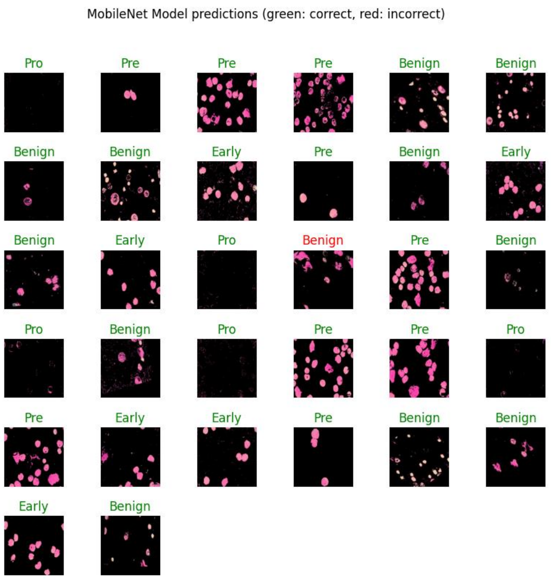


Figure 12. Predictions of MobileNet.

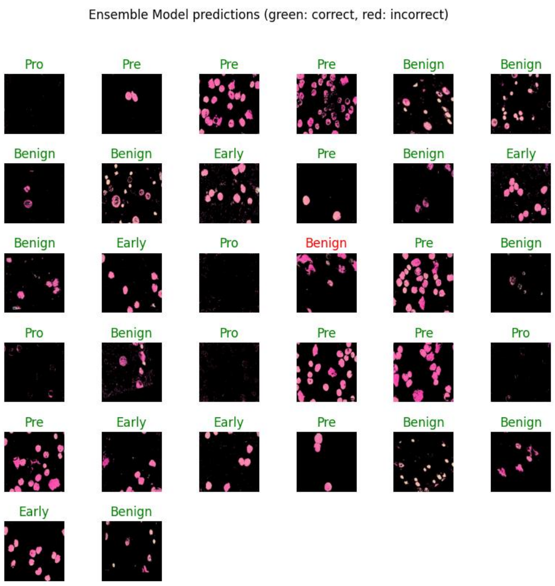


Figure 13. Predictions of Final Ensemble Model

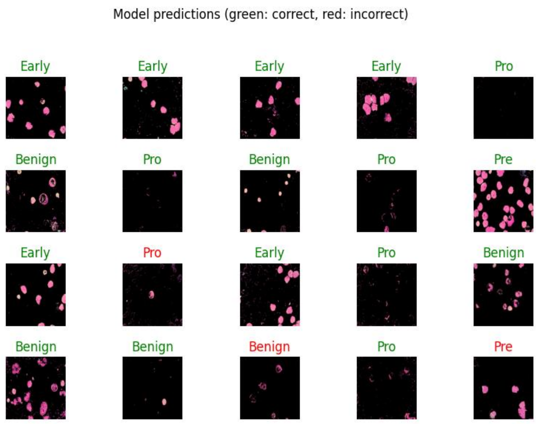


Figure 14. Predictions of MobileNet after cross validation (model 1).

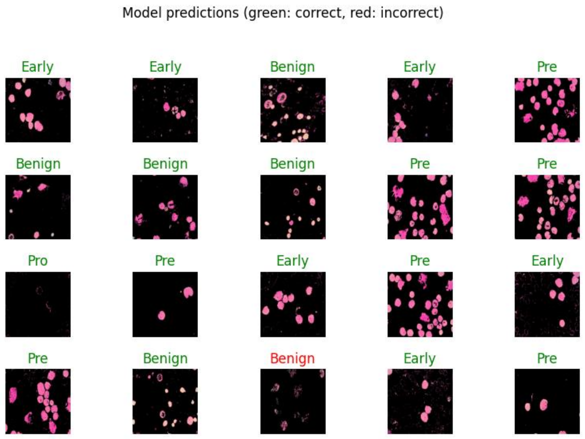


Figure 15. Predictions of MobileNet after cross validation (model 2).

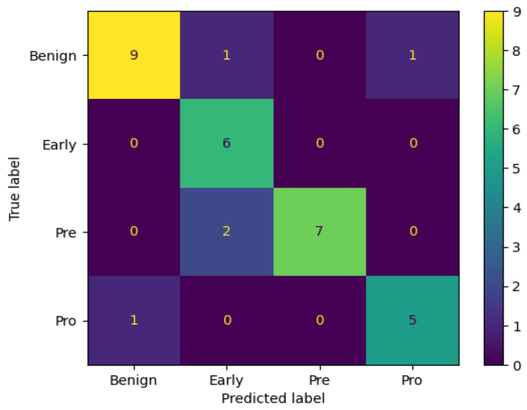


Figure 16. Confusion matrix of ResNet50 model

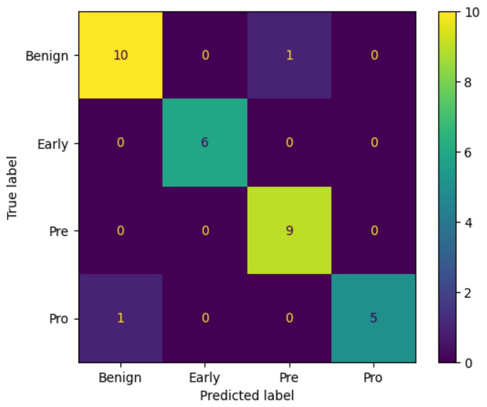


Figure 17. Confusion matrix of Xception model

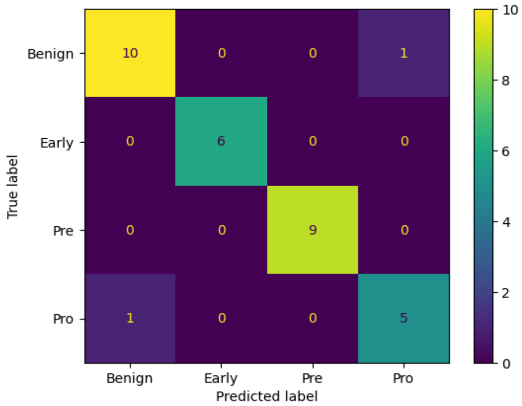


Figure 18. Confusion matrix of VGG16 model.

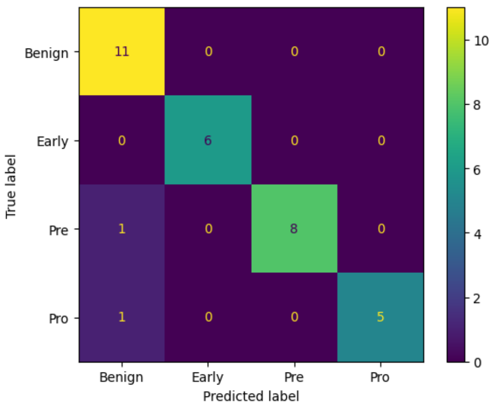


Figure 19. Confusion matrix of NasNetLarge model.

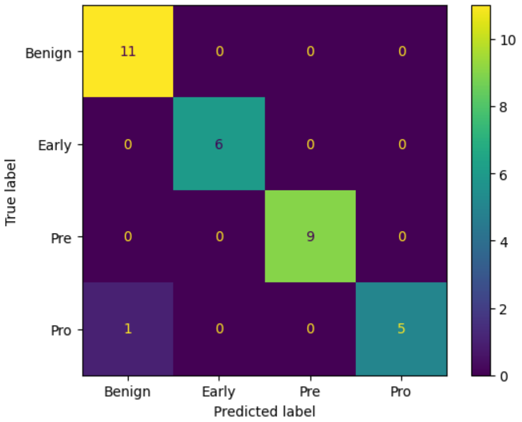


Figure 20. Confusion matrix of MobileNet model

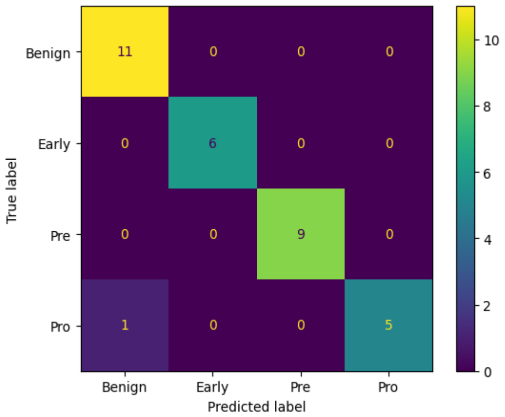


Figure 21. Confusion matrix of Final Ensemble Model.

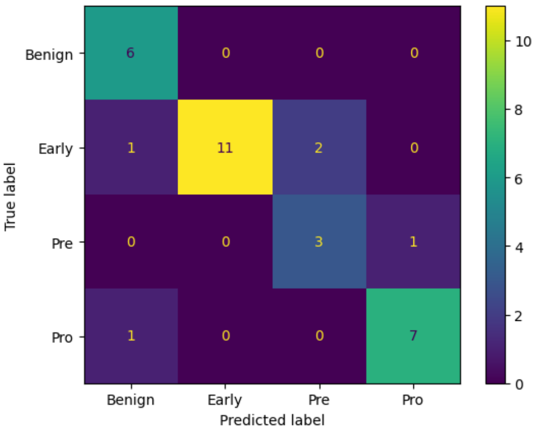


Figure 22. Confusion matrix of MobileNet after cross validation (model 1)

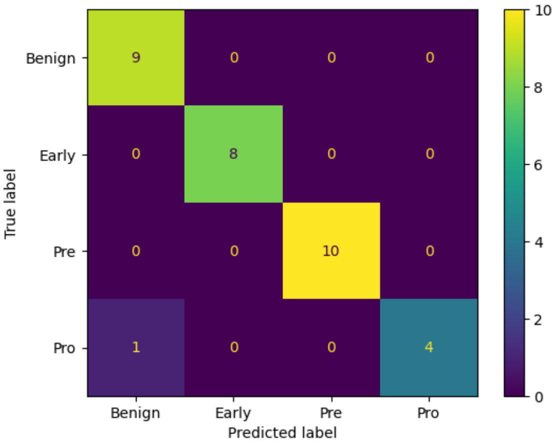


Figure 23. Confusion matrix of MobileNet after cross validation (model 2)

The confusion matrix for the ensemble model for predicting Acute Lymphoblastic Leukemia stage is shown in Figure 21. For the confusion matrices depicted above in Figures 16 to Figure 23 the labels are Benign, Early, Pre, and Pro classes. Figure 22 and Figure 23 denote the confusion matrices of results of MobileNet model after cross validation. The x-axis indicates the predicted labels, while the y-axis represents the true labels. The correctly predicted samples are shown by the diagonal elements in the plot. Inferring from the matrix we find that the classification boundary between almost all the classes is learned well by the classifier since most of the values are 0. Out of a total of 32 samples, 31 of them were properly predicted. The overall accuracy is therefore 96.87%.

The corresponding ROC curves are also plotted. Figure 24 to Figure 28 represents the ROC curves of the individual models. The ROC for the ensemble model for predicting Acute Lymphoblastic Leukemia stage is shown in Figure 29. Figure 30 and Figure 31 represent the results of MobileNet cross validation ROC curves.

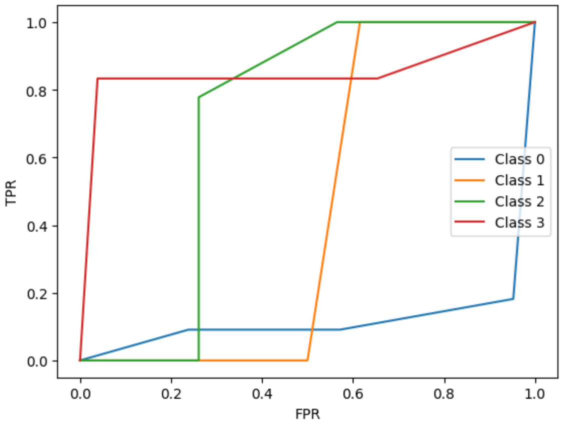


Figure 24. ROC Curve for ResNet50 model

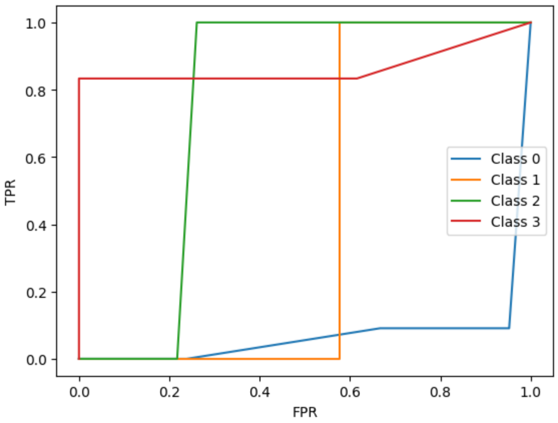


Figure 25. ROC Curve for Xception model

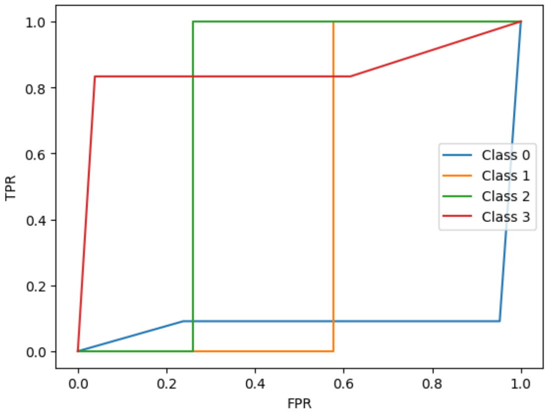


Figure 26. ROC Curve for VGG16 model

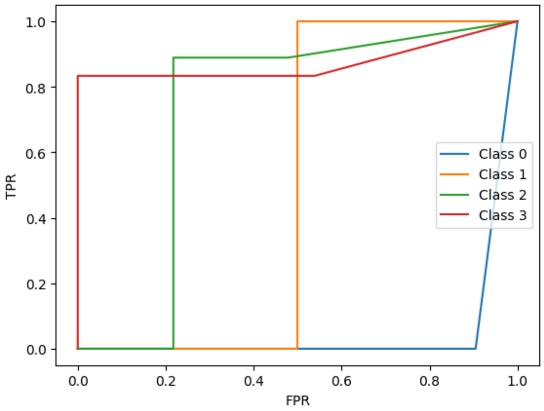


Figure 27. ROC Curve for NasNetLarge model

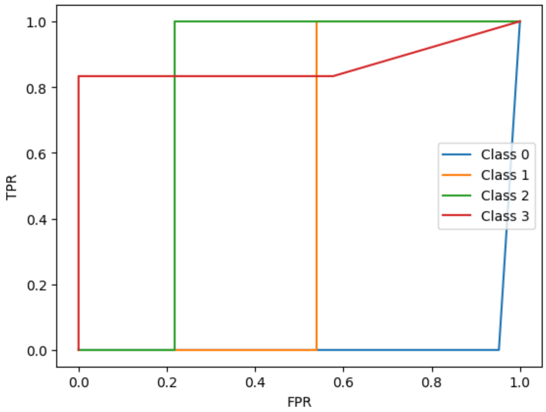


Figure 28. ROC Curve for MobileNet model

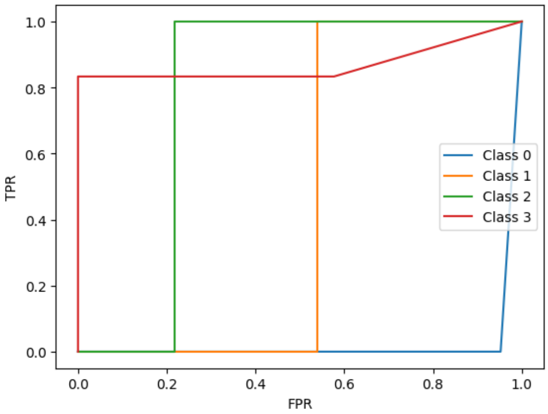
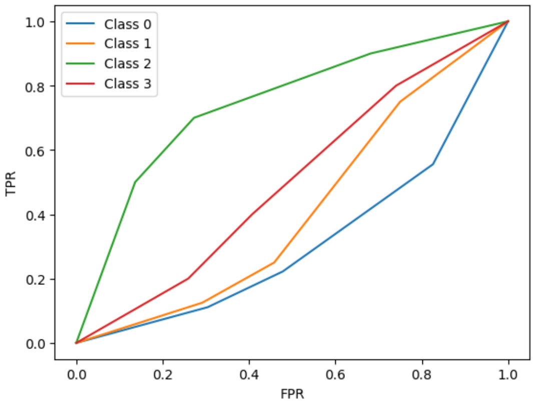


Figure 29. ROC Curve for Final Ensemble model



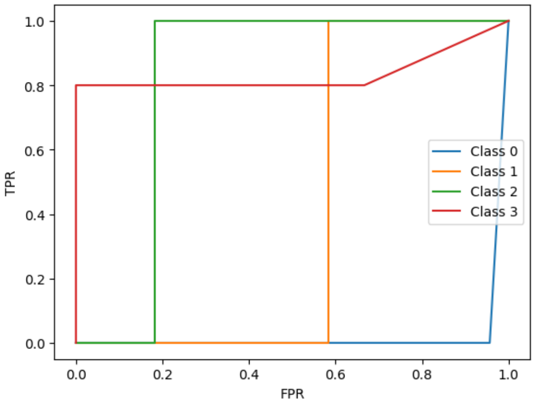


Figure 30. ROC Curve for MobileNet after cross validation (model 1)

Figure 31. ROC Curve for MobileNet after cross validation (model 2)

4.4 Comparative discussion of proposed work with Contemporary Work

The ResNet-101, GoogleNet, SqueezeNet, DenseNet-201, and MobileNetV2 networks are used as part of the CNN model specified in reference [18] to categorize the dataset. The model demonstrated excellent performance in terms of F1 score and accuracy by using an SVM classifier. Particularly, their method's use of the DenseNet-201 model, which had an F1 score of 96.23% and an accuracy of 96.29%, produced good results. In contrast, the ensemble model proposed in the current paper surpasses these achievements by achieving aN accuracy of 96.87%. This suggests that the ensemble model outperforms the DenseNet-201 system in terms of accuracy, showcasing its superiority in accurately classifying instances of Acute Lymphoblastic Leukemia. The ensemble model's remarkable accuracy highlights its potential as a robust and effective approach for the classification task at hand.

Xception, VGG-16, DenseNet-121, MobileNet, and InceptionResNet-V2 whose capabilities were combined to create a powerful ensemble model in reference [22]. To ensure an optimal combination, the weights assigned to each model in the ensemble were estimated based on their individual performance metrics.

An impressive set of results came from the weighted ensemble model's evaluation. A weighted F1-score of 89.7% obtained by the ensemble indicates that it performed well in terms of categorization, striking a good balance between precision and recall. Additionally, the ensemble model demonstrated a balanced accuracy of 88.3%, showcasing its ability to deliver consistent and accurate results across different classes. Additionally, the receiver operating characteristic curve's (AUC) area under the curve obtained a value of 0.948, illustrating the ensemble's high discriminatory power in distinguishing between positive and negative instances.

However, the proposed ensemble model described in the current study outperforms even this already impressive weighted combination. A remarkable 96.87% accuracy was attained using the ensemble model proposed in this paper. This signifies a significant improvement in performance compared to the previous weighted combination mentioned above.

In order to effectively classify instances of the ISBI 2019 C-NMC dataset [20], the study trained and fine-tuned three models: VGG11, ResNet18, and ShufflenetV2. These models were carefully optimized to enhance their classification capabilities. To further improve the classification accuracy, the researchers proposed an ensemble model using the weighted averaging technique. The ensemble model was evaluated using the binary classification criterion.

According to the findings presented in reference [20], the proposed strategy achieved commendable results. An F1 score of 87.40% for the ensemble model indicates a good balance between recall and precision. Furthermore, the ensemble model's overall accuracy was 87.52%, demonstrating its ability to accurately classify instances of the C-NMC dataset.

However, the current paper introduces an even more accurate ensemble model. With a remarkable accuracy of 96.87%, the proposed ensemble model outperforms the previous model mentioned above by a significant margin.

5.0 CONCLUSION

In conclusion, this research project has indeed contributed to the field of medicine by instantly classifying images of a patient's blood cells, so that it can be treated at an early stage. The ensemble model ensures accurate prediction with precision with 96.87% accuracy. The future work involves extending our project to analyze Google images of different patients' blood cells to diagnose Acute Lymphoblastic Leukemia by training the model to bigger datasets. Further creating a website, where doctors who have access to their patient's records can upload images of their blood cells, which gives immediate test results of the chances the patient is diagnosed with Acute Lymphoblastic Leukemia. This model can also be further developed to identify other types of Leukemia.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments (All sources of funding of the study must be disclosed)

We would like to thank you for following the instructions above very closely in advance.

Conflict of interest

The authors declare there is no conflict of interest.

REFERENCES

[1] Saadi, M. I., Nikandish, M., Ghahramani, Z., Valandani, F. M., Ahmadyan, M., Hosseini, F., ... & Ramzi, M. (2023). miR-155 and miR-92 levels in ALL, post-transplant aGVHD, and CMV: possible new treatment options. Journal of the Egyptian National Cancer Institute, 35(1), 18.

[2] Mishra, S. J., & Deshmukh, A. P. (2014). Detection of Leukemia in Human Blood Sample based on Microscopic Images. International Journal of Advanced Research in Electronics and Communication Engineering, 1(3), 10-14.

[3] Menagadevi, M., Nirmala, M., Thiyagarajan, D., & Rajkumar, R. (2023). Web-based Approach for Detection of Acute Lymphoblastic Leukemia From Microscopic Blood Cell Images Using Convolutional Neural Network.

[4] Hunger, S. P., & Mullighan, C. G. (2015). Acute lymphoblastic leukemia in children. New England Journal of Medicine, 373(16), 1541-1552.

[5] Mahmood, N., Shahid, S., Bakhshi, T., Riaz, S., Ghufran, H., & Yaqoob, M. (2020). Identification of significant risks in pediatric acute lymphoblastic leukemia (ALL) through machine learning (ML) approach. Medical & Biological Engineering & Computing, 58, 2631-2640.

[6] Ramaneswaran, S., Srinivasan, K., Vincent, P. D. R., & Chang, C. Y. (2021). Hybrid inception v3 XGBoost model for acute lymphoblastic leukemia classification. Computational and Mathematical Methods in Medicine, 2021, 1-10.

[7] Revanda, A. R., Fatichah, C., & Suciati, N. (2022). Classification of acute lymphoblastic leukemia on white blood cell microscopy images based on instance segmentation using mask R-CNN. vol, 15, 625-637.

[8] Ahmed, I. A., Senan, E. M., Shatnawi, H. S. A., Alkhraisha, Z. M., & Al-Azzam, M. M. A. (2023). Hybrid techniques for the diagnosis of acute lymphoblastic leukemia based on fusion of CNN features. Diagnostics, 13(6), 1026.

[9] Desale, K., Asole, P., Bhosale, G., Bhos, S., & Ambhore, A. (2023). A Deep Learning Framework for Multi-Cancer Detection in Medical Imaging.

[10] Rahman, W., Faruque, M. G. G., Roksana, K., Sadi, A. S., Rahman, M. M., & Azad, M. M. (2023). Multiclass blood cancer classification using deep CNN with optimized features. Array, 18, 100292.

[11] Jani, R., & Tech, S. N. M. CNN-Assisted Automatic White Blood Cancer Type Detection from Bone Marrow.

[12] Menagadevi, M., Nirmala, M., Thiyagarajan, D., & Rajkumar, R. (2023). Web-based Approach for Detection of Acute Lymphoblastic Leukemia From Microscopic Blood Cell Images Using Convolutional Neural Network.

[13] Mondal, C., Hasan, M. K., Jawad, M. T., Dutta, A., Islam, M. R., Awal, M. A., & Ahmad, M. (2021). Acute lymphoblastic leukemia detection from microscopic images using weighted ensemble of convolutional neural networks. arXiv preprint arXiv:2105.03995.

[14] Abdeldaim, A. M., Sahlol, A. T., Elhoseny, M., & Hassanien, A. E. (2018). Computer-aided acute lymphoblastic leukemia diagnosis system based on image analysis. Advances in Soft Computing and Machine Learning in Image Processing, 131-147.

[15] Batool, A., & Byun, Y. C. (2023). Lightweight EfficientNetB3 Model based on Depthwise Separable Convolutions for Enhancing Classification of Leukemia White Blood Cell Images. IEEE Access.

[16] Narayanan, K., Krishnan, R., Robinson, Y., Vimal, S., Kaliappan, M., Rashid, T., ... & Ibrahim, N. J. (2023). A Hybrid Machine Learning Technique for Acute Lymphoblastic Leukemia Classification.

[17] Jawahar, M., Sharen, H., & Gandomi, A. H. (2022). ALNett: A cluster layer deep convolutional neural network for acute lymphoblastic leukemia classification. Computers in Biology and Medicine, 148, 105894.

[18] Mohammed, K. K., Hassanien, A. E., & Afify, H. M. (2023). Refinement of ensemble strategy for acute lymphoblastic leukemia microscopic images using hybrid CNN-GRU-BiLSTM and MSVM classifier. Neural Computing and Applications, 1-13.

[19] Batool, A., & Byun, Y. C. (2023). Lightweight EfficientNetB3 Model based on Depthwise Separable Convolutions for Enhancing Classification of Leukemia White Blood Cell Images. IEEE Access.

[20] Dhalla, S., Mittal, A., Gupta, S., & Singh, H. (2021). Multi-model ensemble to classify acute lymphoblastic leukemia in blood smear images. In Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10–15, 2021, Proceedings, Part I (pp. 243-253). Springer International Publishing.

[21] Kim, S. (2022, October). Developing a Leukemia Diagnostic System Based on Hybrid and Ensemble Deep Learning Architectures. In 2022 International Communication Engineering and Cloud Computing Conference (CECCC) (pp. 70-73). IEEE.

[22] Mondal, C., Hasan, M. K., Ahmad, M., Awal, M. A., Jawad, M. T., Dutta, A., ... & Moni, M. A. (2021). Ensemble of convolutional neural networks to diagnose acute lymphoblastic leukemia from microscopic images. Informatics in Medicine Unlocked, 27, 100794.

[23] Sheet, J., Ghosh, C., & Das, B. K. (2023, February). Deep Learning-Based Transfer Learning for the Detection of Leukemia. In 2023 International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC) (pp. 1-6). IEEE.

[24] Chand, S., & Vishwakarma, V. P. (2022). A novel deep learning framework (DLF) for classification of acute lymphoblastic leukemia. Multimedia Tools and Applications, 81(26), 37243-37262.

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

[26] Jiwani, N., Gupta, K., Pau, G., & Alibakhshikenari, M. (2023). Pattern Recognition of Acute Lymphoblastic Leukemia (ALL) Using Computational Deep Learning. IEEE Access, 11, 29541-29553.\

[27] Papers with Code - MobileNetV2 Explained. (n.d.). https://paperswithcode.com/method/mobilenetv2#:~:text=MobileNetV2%20is%20a%20convolutional%20neural,are%20between%20the%20bottleneck%20layers.

[28] Papers with Code - ResNet Explained. (n.d.). https://paperswithcode.com/method/resnet

[29] Great Learning. (2022, January 5). Everything you need to know about VGG16 - Great Learning - Medium. Medium. https://medium.com/@mygreatlearning/everything-you-need-to-know-about-vgg16-7315defb5918

[30] Akhtar, Z. (2021). Xception: Deep Learning with Depth-wise Separable Convolutions. OpenGenus IQ: Computing Expertise & Legacy. https://iq.opengenus.org/xception-model/

[31] Tsang, S. (2021, December 10). Review: NASNet — Neural Architecture Search Network (Image Classification). Medium. https://sh-tsang.medium.com/review-nasnet-neural-architecture-search-network-image-classification-23139ea0425d\

[32] Mehrad Aria, Mustafa Ghaderzadeh, Davood Bashash, Hassan Abolghasemi, Farkhondeh Asadi, and Azamossadat Hosseini, "Acute Lymphoblastic Leukemia (ALL) image dataset." Kaggle, (2021). DOI: 10.34740/KAGGLE/DSV/2175623.