ESKİŞEHİR TECHNICAL UNIVERSITY

FACULTY OF ENGINEERING

CLASSIFICATION MODELS FOR THE DIAGNOSIS OF ACUTE LYMPHOBLASTIC LEUKEMIA

Ali Han O.

Koray K.

Mechmet C. C.

A Bachelor of Science Project

Department of Computer Engineering

June 2023

ABSTRACT

Leukemia is a type of cancer that manifests itself with the proliferation of blood cells, especially white blood cells. This hematological disease is usually diagnosed using manual methods such as complete blood count, bone marrow aspiration, or microscopic examination of the blood sample. Manual methods of leukemia diagnosis are low-cost; however, they are less reliable, time-consuming, and labor-intensive methods. Abnormal white blood cells spread very quickly and also target other organs in the body. Therefore, early diagnosis of leukemia is critical. In addition, manual diagnosis methods results may differ according to the current psychological and physiological state of the medical personnel. Although the transition from manual methods to computer-aided diagnosis systems will increase the cost, it is of great importance in terms of providing doctors and laboratory technicians with a second opinion for definitive diagnosis. This study developed hybrid classification models that aim to achieve high performance by using transfer learning and ensemble learning algorithms on microscopic blood images in the ALL-IDB dataset, which is widely used in the literature for the diagnosis of Acute Lymphoblastic Leukemia (ALL). Furthermore, the original images for this disease classification, it is also used in different preprocessing techniques, which are frequently used in the literature and have proven to give satisfactory outcomes by increasing the overall performance of the system. Finally, after applying all the methods mentioned above, 100% accuracy was achieved.

Keywords: Acute Lymphoblastic Leukemia, Computer Aided Diagnosis, Data Augmentation, Ensemble Learning, Transfer Learning.

CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

ABBREVIATIONS

ALL	Acute Lymphoblastic Leukemia
CNN	Convolutional Neural Networks
LoG	Laplasian of Gaussian
CBC	Complete Blood Count

1. INTRODUCTION

Leukemia is caused by the rapid production of abnormal white blood cells by the body and affects the blood and bone marrow. The overproduction of abnormal and immature white blood, as a result of decreased production of red blood cells and increased damaged bone marrow, leads to immune system damage. The causes of leukemia are yet to be found. Scientists think that it is the result of hereditary and environmental factors. This hematological malignancy is usually diagnosed using manual methods such as complete blood count (CBC), bone marrow aspiration, or microscopic examination of the blood sample. Manual methods of leukemia diagnosis are low-cost; however, they are less reliable, time-consuming, and labor-intensive methods. Abnormal white blood cells spread very quickly and also target other organs in the body. Therefore, early diagnosis of leukemia is very important. In addition, manual diagnosis methods results may differ according to the current psychological and physiological state of the medical personnel. Although the transition from manual methods to computer aided diagnosis systems will increase the cost, it is of great importance in terms of providing doctors and laboratory technicians with a second opinion for definitive diagnosis. Computer-aided diagnostic system procedures for detecting lymphoblasts from colorized microscopic images consist of four steps. These are segmentation, leukocyte identification, lymphocyte identification, and identification of potential lymphoblasts. These processes begin with the acquisition of images from the patients and these images are turned into a data set. In order to obtain the data set, microscopic images of the bone marrow and blood are collected from healthy and leukemia-diagnosed patients. After this process, the obtained data sets are subjected to a number of preprocessing processes to improve their quality. Finally, segmentation, feature extraction, and classification methods are applied. The data set is segmented according to the scales selected in the segmentation step [2]. In the classification step, firstly, binary classification can be made as the image has a disease or no disease, or multi-label classification can be made to determine which disease type the images are. According to the regulations made by the World Health Organization, the nomenclature of L1, L2, and L3 type leukemia, which are named as subtypes of Acute Lymphoblastic

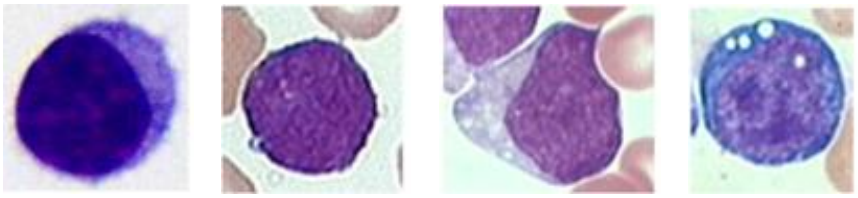
Leukemia (ALL), has not been seen to be able to distinguish the subtypes of the disease, and they have been reclassified and classified as T-ALL, B-ALL cell and Burkitt lymphoma (BL) [3]. In this study, performed binary classification.

In the literature, microscopic blood images are examined by using machine learning and deep learning methods, many studies have been conducted on the detection and classification of cancerous blood cells. When the studies are examined in terms of preprocessing techniques used in this field, it is seen that preprocessing techniques such as median filtering, adaptive segmentation, LoG (laplacian of gaussian), and image augmentation are used in [4- 11] studies. In these studies, the ASH [12], ALL-IDB1 [1], and ALL-IDB2 [1] datasets are being used. Also, classification methods are support vector machine, K-nearest neighbor, artificial neural networks, random forest, and decision tree. In this study, unlike the studies in the literature for computer-aided diagnosis of acute lymphoblastic leukemia through microscopic blood images, different preprocessing techniques and different pre-trained deep neural networks and ensemble learning techniques combined to investigate the effect of different techniques on performance.

2. MATERIALS AND METHODS

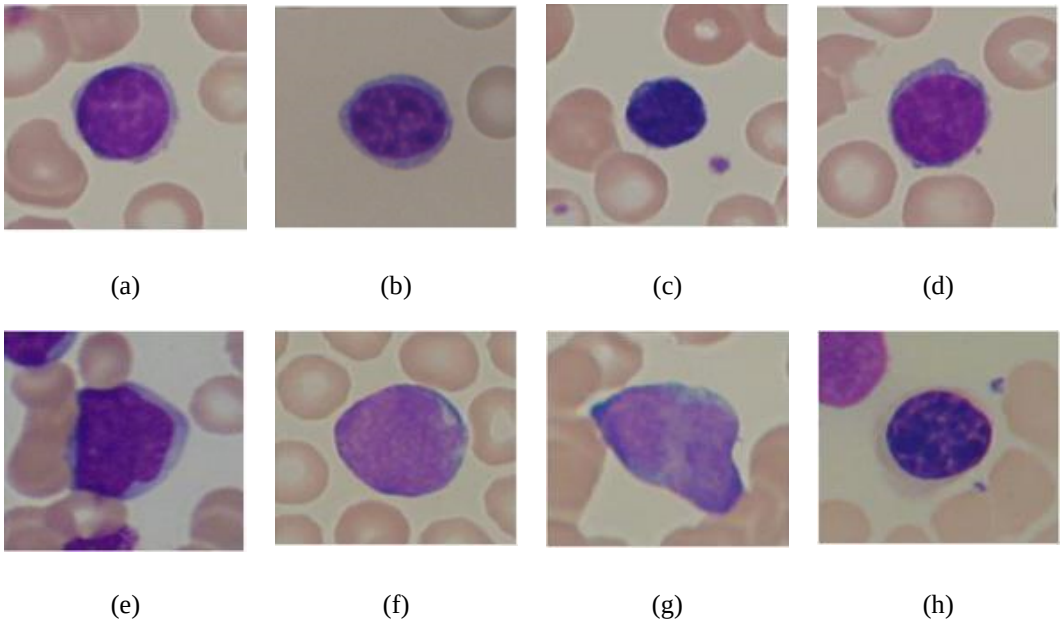
2.1 Dataset

Open-access datasets that can be used for the diagnosis of acute lymphoblastic leukemia can be listed as ASH and ALL-IDB. The dataset used in this study is the ALL-IDB [1] dataset prepared by Università degli Studi di Milano for the development of segmentation and classification algorithms in lymphoblastic leukemia diagnostic systems. All images in the content of the data set were evaluated by oncology specialists. Images were taken with optical laboratory microscopes. Images are in JPG format and have 24-bit color depth. The resolution of the images is 2592 x 1944 pixels. The dataset consists of 108 images and 39000 blood elements. The magnification of images varies between 300x and 500x [1].



(a) (b) (c) (d)

Figure 2.1 Sample images of ALL subtypes in the dataset, (a) Sample image taken from a healthy person, (b) T-Cell, (c) B-Cell, (d) Burkitt lymphoma (BL).



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 2.2 (a-d) Sample images from healthy individuals without ALL in the ALL-IDB dataset, (e-h) Sample images from patients diagnosed with ALL.

2.2 Preprocessing

Some preprocessing methods have been applied to the data set in order to achieve the goal and increase the accuracy of the model. The suggested results to achieve positive results by improving the overall performance of the system are in the list below:

Resize Images: In order to prepare the images for processing within the defined network architecture, resizing is performed to ensure they conform to the required input dimensions.

In this study, the resizing operation is applied to each image in the dataset, modifying its dimensions to 224 pixels in width and height, with 3 channels representing the RGB color space. This transformation maintains the aspect ratio of the original images while conforming to the input requirements of the defined network architecture.

Data Augmentation: In the initial preprocessing step, data augmentation is applied to triple the size of the training set in the ALL-IDB1 database. The test set isn't affected by this process in order to do a fair comparison between the results that are

published in the literature. This process involves various transformations, such as rotation in different angles and mirroring the image to generate additional training samples.

Utilizing the capabilities of the OpenCV library, the images in the training set are modified to introduce variations and increase the diversity of the data. By performing these transformations, the number of training images is tripled, resulting in a larger and more representative training set.

Data augmentation plays a crucial role in addressing the challenges posed by limited training data. By expanding the training set, the model can learn from a more comprehensive range of examples, encompassing various orientations, perspectives, and visual characteristics. This augmentation technique mitigates the risk of overfitting and enhances the model's generalization ability, enabling it to make accurate predictions on unseen data.

The augmented dataset, consisting of the original images and their transformed counterparts, facilitates the training process and improves the model's performance. With a more extensive training set, the model can capture a broader range of features and patterns associated with Acute Lymphoblastic Leukemia, leading to enhanced diagnostic accuracy.

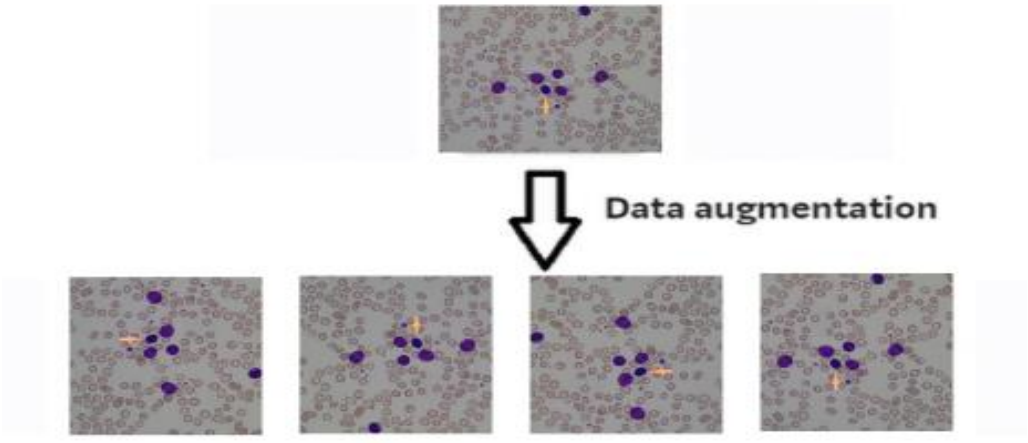


Figure 2.3 Applied preprocessing techniques to an image

2.3 k-Fold Cross Validation

In addition, to transfer learning, K-fold Cross Validation is employed as a validation method to assess the performance of the machine learning model developed in this study. This technique divides the dataset into K equal parts, or folds, and evaluates the model's performance by using each fold in turn as the test data while the remaining folds are used for training.

K-fold Cross Validation serves multiple purposes in this study. Firstly, it helps detect and mitigate the issue of overfitting, which occurs when a model performs exceptionally well on the training data but fails to generalize to unseen data. Evaluating the model on different folds, provides a more comprehensive assessment of its performance on various subsets of the data, ensuring that it can handle diverse inputs and make accurate predictions on real-world data.

Furthermore, K-fold Cross Validation improves the generalization ability of the model. By training the model on different combinations of the folds and evaluating its performance across the folds, it provides a more reliable estimate of the model's performance and its ability to generalize to unseen data. This is particularly beneficial in ensuring that the developed model is robust and effective in practical applications.

In this study, a 10-fold cross-validation approach is utilized. This means that the dataset is divided into 10 equal parts, with each part used as the test set once while the remaining nine parts are used for training. The process is repeated for each fold, and the performance metrics are calculated and aggregated to provide an overall assessment of the model's performance.

By employing 10-fold cross-validation, this study ensures a thorough evaluation of the developed model, taking into account the variability of the data and providing a more reliable estimate of its performance. This validation technique enhances the reliability and credibility of the results obtained, bolstering confidence in the model's ability to accurately diagnose Acute Lymphoblastic Leukemia.

2.4 Convolutional Neural Networks

Convolutional neural networks (CNNs) are a type of neural network designed specifically for image processing and computer vision tasks. They are composed of multiple layers of interconnected "neurons" that process and analyze the input data. CNN learns features directly and eliminates manual feature extraction. The main method used in this study is CNN.

As seen in Figure 2.4, the input layer represents the data. In the convolution layer, the images given as input are inserted into a number of convolutional filters, and each filter activates certain features in the image. The RELU layer is also called the activation layer. Only the features activated in this layer are carried over to the next layer. In the pooling layer, the output is simplified and the number of parameters the network has to learn is reduced. These processes are repeated over and over and different features are learned each time. In the flatten layer, the data is reduced to one dimension, and a feature vector is obtained. The resulting vector is sent as an input to the fully-connected layer, which is the last layer that performs the classification, and results are obtained according to the previously obtained features. Then, calculations are made in the SoftMax layer and results such as the classification layer and the class of the image and the error rate are obtained.

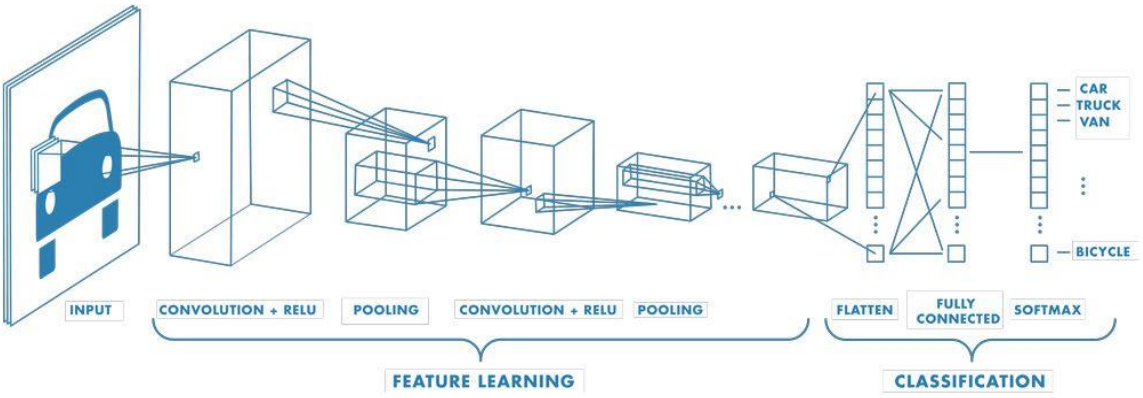


Figure 2.4 CNN Architecture [13]

2.5 Transfer Learning

Transfer Learning is a valuable technique that allows the knowledge gained from solving one task to be utilized in solving a related task. It involves leveraging a pre-trained model or learned features as a starting point and adapting them to the new task through fine-tuning. This approach can significantly enhance model performance and reduce training time.

The process of implementing transfer learning involves several key steps. Firstly, a suitable pre-trained network is selected and loaded into the system. This pre-trained network is typically trained on a large and diverse dataset, such as VGG-16, enabling it to learn general features applicable to various tasks.

Next, the last fully connected layer and the last classification layer of the pre-trained network are adjusted to align with the requirements of the new dataset. These layers are replaced with new ones, with the classification layer's output neurons matching the number of classes in the new dataset.

The dataset is then divided into training and test sets. The training set is used to fine-tune the pre-trained network, while the test set is employed to evaluate the model's performance on the new task.

To facilitate the fine-tuning process, appropriate training options are set. This includes selecting the optimization algorithm, defining the learning rate, and batch size, and determining the number of training epochs. Striking a balance between allowing the model to learn from the new dataset and retaining the valuable knowledge captured by the pre-trained weights is crucial.

The network is then fine-tuned by freezing the weights of the pre-trained layers to prevent modification, while only adjusting the weights of the newly added layers. This allows the model to adapt specifically to the new task, while still benefiting from the pre-trained layers' knowledge.

Subsequently, the modified network is trained using the training set. During this process, the weights of the new layers are updated, and the model learns to make predictions tailored to the new task.

Finally, the performance of the model is evaluated using the test set. Various evaluation metrics, such as accuracy, precision, recall, and F1-score, are calculated to assess the model's effectiveness in the new task.

By following this process, transfer learning enables the efficient utilization of existing knowledge, leading to accelerated development and deployment of high-performance models for new tasks.

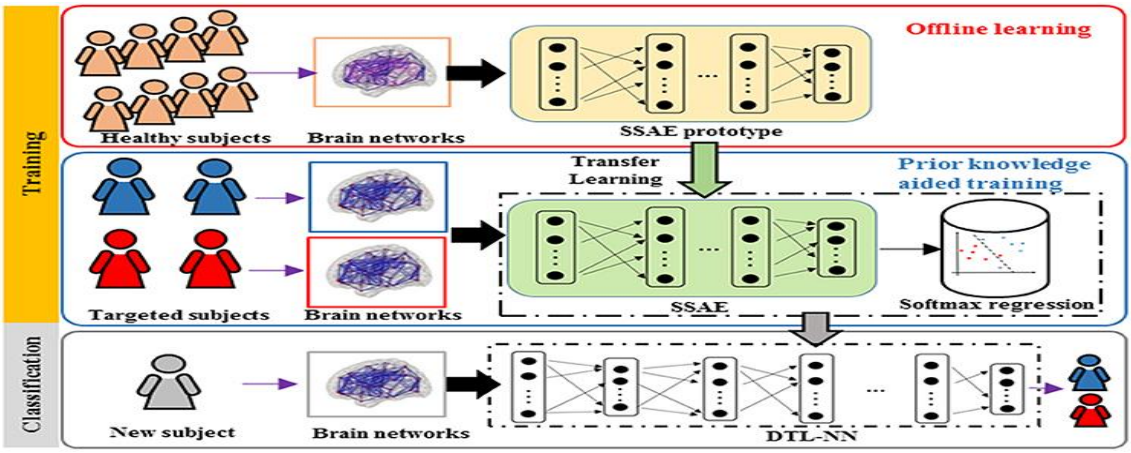


Figure 2.5 Transfer learning process

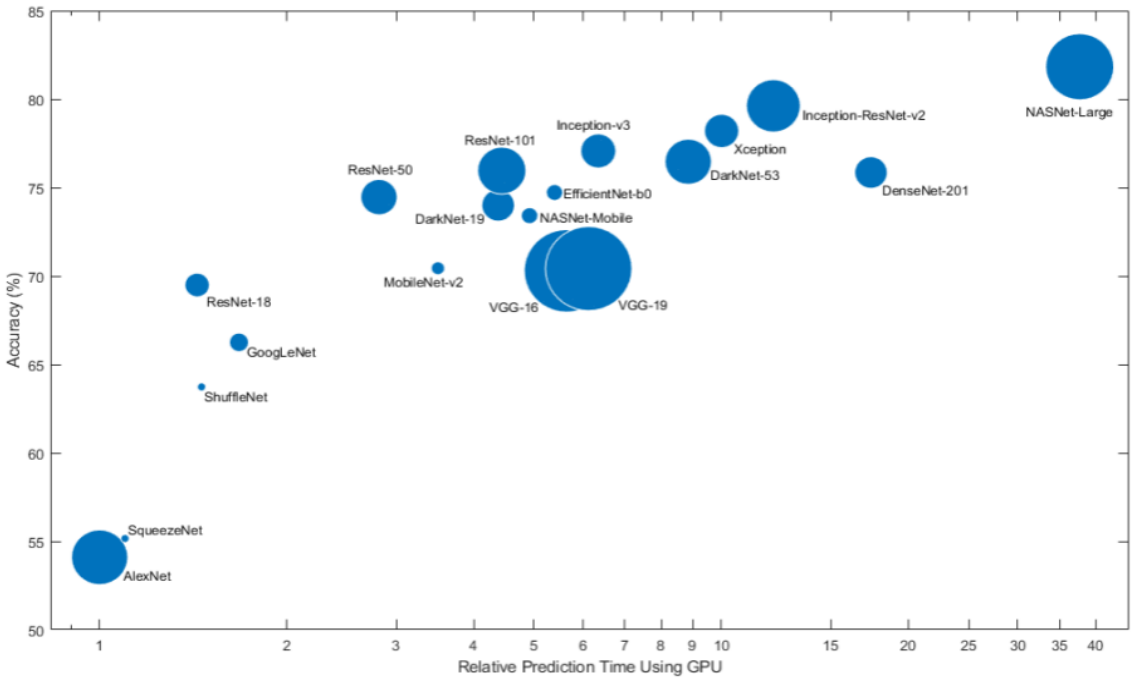


Figure 2.6 Pre-trained networks

2.6 Ensemble Learning

Ensemble learning is a machine learning technique that involves combining the predictions of multiple models to make a final prediction. The goal of ensemble learning is to improve the performance of the model by leveraging the strengths of multiple models and reducing the impact of individual model biases or errors. The primary objective of ensemble learning is to maximize the benefits of several models while minimizing the negative effects of any individual biases or errors. The goal of ensemble learning is to combine the predictions of various models to provide a final prediction that is more trustworthy, accurate, and robust than any single model could produce on its own.

In this particular study, the stacking ensemble learning method is employed. A sophisticated strategy called stacking involves training a variety of models, aggregating their predictions through a meta-model, and then combining the results. The meta-model learns how to best integrate the predictions from the basis models as input to produce a final prediction. This stacking procedure improves the ensemble's predictive accuracy by enabling it to identify intricate relationships and patterns that an individual model could have overlooked.

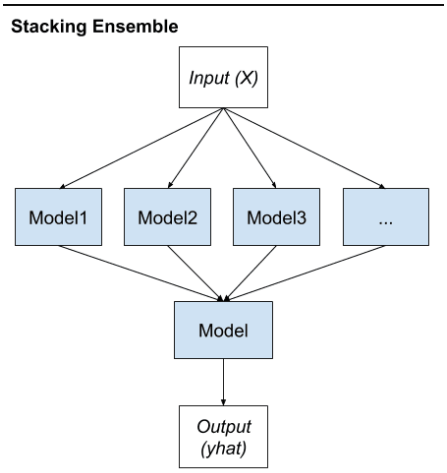


Figure 2.7 Stacking ensemble learning process

Machine learning prediction performance can be enhanced through ensemble learning, and stacking in particular. Ensemble learning enables academics and practitioners to approach complicated issues more efficiently and generate more accurate and robust predictions by leveraging the pooled expertise of several models.

2.7 Performance Evaluation Parameters

Precision, sensitivity, accuracy, and F1 score are the four measures used in this work to assess performance. Using performance metrics based on four essential findings for assessing the classifier, such as True positive (TP), False positive (FP), True negative (TN), and False negative (FN), the created DR classification and efficacy are assessed. Below is an explanation of every measure that uses the formula. The confusion matrix contains the evaluation findings. Consider, for instance, a "CxC" matrix with Pij as its elements (i, j = 1, 2, 3..., number of classes). Let J stand for the True Positives (TP) count, K for the False Negatives (FN) count, M for the False Positives (FP) count, and N for the True Negatives (TN) count in this matrix. While FP and FN represent the erroneously classified information, TP and TN show the correctly classified data. The TPs and FPs for each actual class i in multiclass classification may be obtained by taking the p anticipated classes. The performance of the model may then be examined by deriving several evaluation metrics from the confusion matrices. The equation (3.4.1) shows the formulas for TP, FP, TN, and FN.

$$\begin{aligned} &\text{TPs, } Ji = Pii \\ &\text{FNs, } Ki = \Sigma p\ j=1 Pij - Ji \\ &\text{FPs, } Mi = \Sigma p\ j=1 Pji - Ji \\ &\text{TNs, } Ni = \Sigma p\ j=1 \Sigma n\ k=1 Pik - Ji - Mi - Ki \end{aligned} \qquad (2.8)$$

Accuracy (A) system's efficiency is determined by the accuracy measure. It is shown in equation (2).

$$A = \frac{TP + TN}{TP + TN + FP + FN} \qquad (2.9)$$

F1-score (f1) the ensemble mean of sensitivity and precision is determined for evaluating the F1-score.

$$f1 = \frac{2TP}{2TP + FP + FN} \tag{2.10}$$

Precision (P) the proportion of predicted positives which are genuine positive is determined using precision metric.

$$P = \frac{TP}{TP + FP} \tag{2.11}$$

Recall (R) or Sensitivity: The proportion of actual positives that are correctly identified.

$$R = \frac{TP}{TP + FN} \tag{2.12}$$

3. EXPERIMENTAL RESULTS

Experimental studies up till this point carried out only using the ALL-IDB1 dataset which consists of 108 microscopic blood images in total. The dataset is divided into training and test sets as %70 and %30 respectively. First, the original images and then the augmented images are tested using K-Fold Cross Validation with Transfer Learning. After this step is complete an Ensemble Learning Model is created using the best-performing three models via the stacking approach.

Pre-trained networks which are displayed in Figure 2.6 are tested with the original dataset and the best-performing four of them are chosen to evaluate the rest of the experiments. ResNet-50, VGG-16, Xception, and InseptionResNetV2 were qualified pre-trained candidates for this classification task.

In order to examine the difference between the original and augmented dataset as well as the effects of the ensemble learning model on the accuracy of the classification process three major testing phases have been done. These first two uses 10-Fold Cross Validation and Transfer Learning to evaluate the model the only difference is one of them uses the original dataset and the other one uses augmented dataset. By using this approach relationship between accuracy and dataset size is clearly examined. In the continuation, the ensemble learning model is created by using the models that have been trained with augmented dataset to enhance our results.

3.1 Results on Original Dataset

The first testing phase started with the original dataset by using 10-Fold Cross Validation and Transfer Learning individually utilizing the pre-trained networks called ResNet-50, VGG-16, Xception, and InseptionResNetV2. The accuracy rates range from 49,69% to 88,18%. The results are shown in Table 3.1.

Table 3.1 Results of original dataset

ALL-IDB1					
Fold No	**Pre-trained Networks**	**Acuracy**	**Precision**	**Recall**	**F1-score**
1 Fold	InceptionResnetV2	0,5757	1,0000	0,2222	0,3636
	Xception	0,6666	1,0000	0,3888	0,5600
	ResNet-50	0,8787	1,0000	0,7777	0,8750
	VGG-16	1,0000	1,0000	1,0000	1,0000
2 Fold	InceptionResnetV2	0,6060	0,6923	0,5000	0,5806
	Xception	0,6060	0,7777	0,3888	0,5185
	ResNet-50	0,7878	1,0000	0,6100	0,7586
	VGG-16	0,9696	1,0000	0,9444	0,9714
3 Fold	InceptionResnetV2	0,4848	1,0000	0,5556	0,1206
	Xception	0,5151	0,5833	0,3888	0,4666
	ResNet-50	0,9696	1,0000	0,9444	0,9714
	VGG-16	0,9696	1,0000	0,9444	0,9714
4 Fold	InceptionResnetV2	0,6060	0,5862	0,9844	0,7234
	Xception	0,5454	0,5483	0,9444	0,6938
	ResNet-50	0,8484	1,0000	0,7222	0,8387
	VGG-16	0,9696	1,0000	0,9444	0,9714
5 Fold	InceptionResnetV2	0,5757	1,0000	0,2222	0,3636
	Xception	0,6363	0,6875	0,6111	0,6470
	ResNet-50	0,6363	1,0000	0,3333	0,5000
	VGG-16	0,9696	1,0000	0,9444	0,9714
6 Fold	InceptionResnetV2	0,6363	0,5666	0,9444	0,7083
	Xception	0,4848	1,0000	0,0555	0,1052
	ResNet-50	1,0000	1,0000	1,0000	1,0000
	VGG-16	1,0000	1,0000	1,0000	1,0000
7 Fold	InceptionResnetV2	0,9115	0,5000	1,0000	0,3333
	Xception	0,0000	0,8779	0,8301	0,8832
	ResNet-50	0,9090	1,0000	0,8333	0,9090
	VGG-16	0,9696	1,0000	0,9444	0,9714
8 Fold	InceptionResnetV2	0,5151	0,2000	1,0000	0,1111
	Xception	0,4848	1,0000	0,0555	0,1052
	ResNet-50	1,0000	1,0000	1,0000	1,0000
	VGG-16	1,0000	1,0000	1,0000	1,0000
9 Fold	InceptionResnetV2	0,5151	0,2000	1,0000	0,1111
	Xception	0,4848	1,0000	0,0555	0,1052
	ResNet-50	0,8484	1,0000	0,7222	0,8387
	VGG-16	0,4848	1,0000	0,0555	0,1052
10 Fold	InceptionResnetV2	0,4848	1,0000	0,5556	0,1206
	Xception	0,5454	1,0000	0,1666	0,2857
	ResNet-50	0,9393	1,0000	0,8888	0,9411
	VGG-16	0,4848	1,0000	0,0555	0,1052

Overall	Acuracy	Precision	Recall	F1-score
InceptionResnetV2	0,5911	0,6745	0,6984	0,3536
Xception	0,4969	0,8475	0,3885	0,4370
ResNet-50	0,8818	1,0000	0,7832	0,8633
VGG-16	0,8818	1,0000	0,7833	0,8067

3.2 Results on Augmented Dataset

The second testing phase started with the augmented dataset by using 10-Fold Cross Validation and Transfer Learning individually utilizing the pre-trained networks called ResNet-50, VGG-16, Xception, and InseptionResNetV2. The accuracy rates range from 82,11% to 99,09%. The results are shown in Table 3.2.

Table 3.2 Results of augmented dataset

ALL-IDB1 with Data Augmentation					
Fold No	**Pre-trained Networks**	**Acuracy**	**Precision**	**Recall**	**F1-score**
1 Fold	InceptionResnetV2	0,8484	1,0000	0,7222	0,8387
	Xception	0,8181	0,9285	0,7222	0,8125
	ResNet-50	0,9697	0,9473	1,0000	0,9729
	VGG-16	1,0000	1,0000	1,0000	1,0000
2 Fold	InceptionResnetV2	0,9697	1,0000	0,9444	0,9715
	Xception	0,8484	0,8095	0,9444	0,8717
	ResNet-50	0,9393	1,0000	0,8888	0,9411
	VGG-16	1,0000	1,0000	1,0000	1,0000
3 Fold	InceptionResnetV2	0,9697	1,0000	0,9444	0,9715
	Xception	0,9090	0,8571	1,0000	0,9230
	ResNet-50	0,9697	1,0000	0,9444	0,9714
	VGG-16	1,0000	1,0000	1,0000	1,0000
4 Fold	InceptionResnetV2	1,0000	1,0000	1,0000	1,0000
	Xception	0,7878	0,7391	0,9444	0,8292
	ResNet-50	1,0000	1,0000	1,0000	1,0000
	VGG-16	1,0000	1,0000	1,0000	1,0000
5 Fold	InceptionResnetV2	1,0000	1,0000	1,0000	1,0000
	Xception	0,8181	0,7727	0,9444	0,8500
	ResNet-50	1,0000	1,0000	1,0000	1,0000
	VGG-16	1,0000	1,0000	1,0000	1,0000
6 Fold	InceptionResnetV2	0,9697	1,0000	0,9444	0,9715
	Xception	0,8181	0,8750	0,7777	0,8235
	ResNet-50	1,0000	1,0000	1,0000	1,0000
	VGG-16	0,9393	1,0000	0,8888	0,9411
7 Fold	InceptionResnetV2	1,0000	1,0000	1,0000	1,0000
	Xception	0,7878	0,7391	0,9444	0,8292
	ResNet-50	1,0000	1,0000	1,0000	1,0000
	VGG-16	0,9697	1,0000	0,9444	0,9714
8 Fold	InceptionResnetV2	0,9393	1,0000	0,8888	0,9411
	Xception	0,7878	0,7391	0,9444	0,8292
	ResNet-50	1,0000	1,0000	1,0000	1,0000
	VGG-16	1,0000	1,0000	1,0000	1,0000
9 Fold	InceptionResnetV2	0,9697	1,0000	0,9444	0,9715
	Xception	0,7878	1,0000	0,6111	0,7586
	ResNet-50	1,0000	1,0000	1,0000	1,0000
	VGG-16	1,0000	1,0000	1,0000	1,0000
10 Fold	InceptionResnetV2	0,9393	1,0000	0,8888	0,9411
	Xception	0,8484	0,8095	0,9444	0,8717
	ResNet-50	1,0000	1,0000	1,0000	1,0000
	VGG-16	1,0000	1,0000	1,0000	1,0000

Overall	Acuracy	Precision	Recall	F1-score
InceptionResnetV2	0,9606	1,0000	0,9277	0,9607
Xception	0,8211	0,8270	0,8777	0,8399
ResNet-50	0,9879	0,9947	0,9833	0,9885
VGG-16	0,9909	1,0000	0,9833	0,9913

3.3 Ensemble Learning Results

In the final test phase best three models that have been trained with the augmented dataset chosen InceptionResnetV2, ResNet50, and VGG-16 were the best performing models and by using them an ensemble learning model is created via the stacking approach. The accuracy of the ensemble learning model is 100%. Also, 100% accuracy rate is consistent among the repeated tests.

Table 3.3 Comparison table with related works

Reference	Method	Dataset	Accuracy
(Das, 2020)	SVM	ALL-IDB1	96%
(Dese, 2021)	MCSVM	ALL-IDB	100,00%
(Rahman, 2018)	SVM,KNN	ALL-IDB	93,60%
(Das, 2021)	Random Forest	ALL-IDB1	98,46%
(Anilkumar, 2021)	SGDM	ASH	94.12%
(Rehman, 2018)	CNN,Naive Bayesian	MIAS	97.78%
(Chand, 2022)	CNN	Not specified	98,17%
(Pałczy´nski, 2021)	FC, XGBOOST, Random Forest,Decision tree	ALL-IDB	97.4%
(Abhishek, 2022)	CNN,SVM,LBP,HOG	Private	96%
(Devi, 2023)	GBHSV-Leuk(novel)	ALL-IDB1	95.41%
Our Work	**CNN,Transfer Learning,Ensemble Learning**	**ALL-IDB1**	**100%**

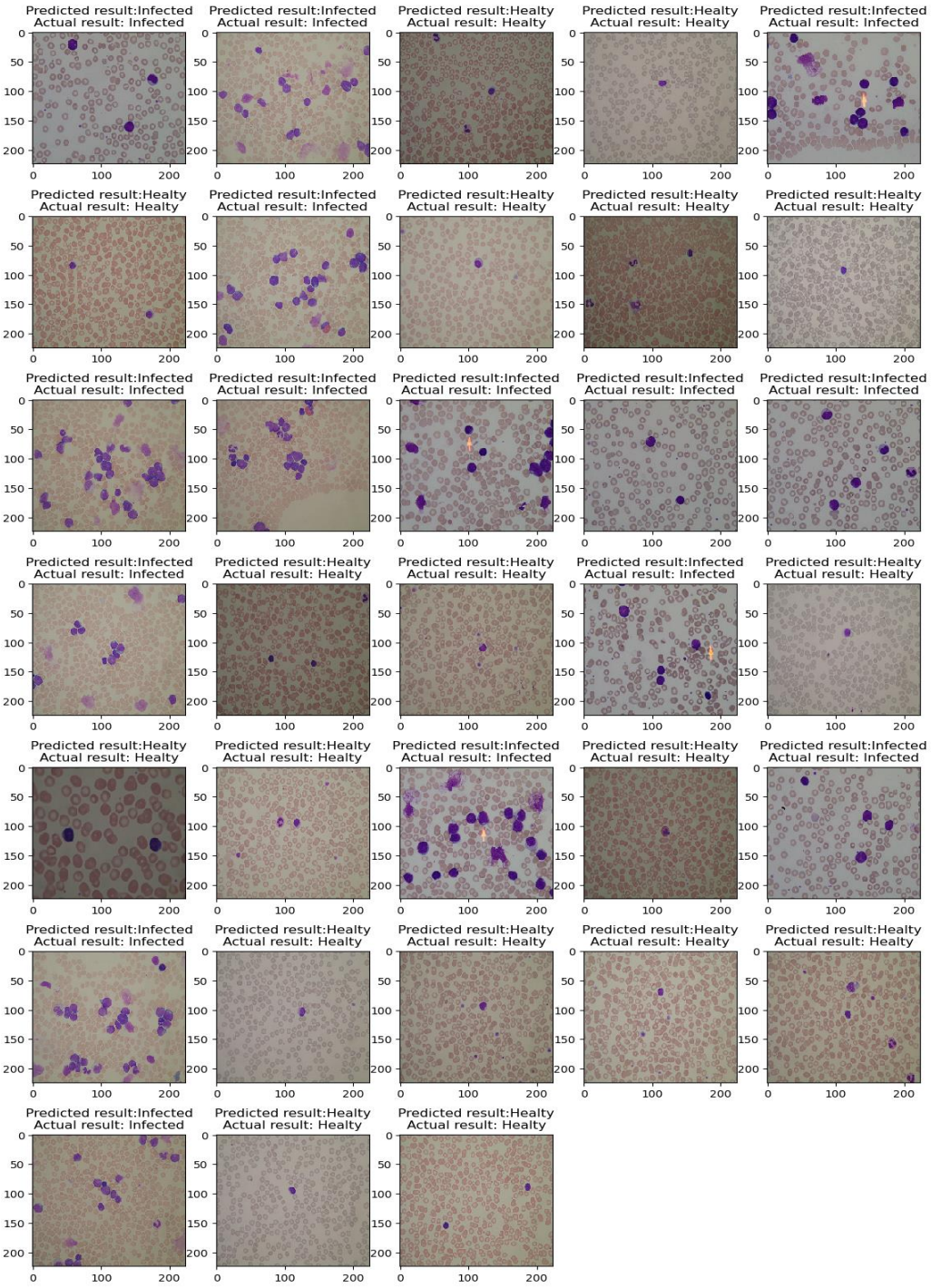


Figure 3.1 Ensemble model predictions on test set

4. CONCLUSIONS

As a result, in this study a classification and diagnostic model for accurately diagnosing Acute Lymphoblastic Leukemia (ALL) using a variety of methods has been successfully developed. The primary objective was to provide healthcare professionals with a valuable second opinion to improve the accuracy of ALL detection. Notably, the model achieved an outstanding accuracy of 100% by leveraging ensemble learning techniques.

In future work, the study aims to expand its scope by incorporating subclasses of leukemia, including T-cell Acute Lymphoblastic Leukemia (T-ALL), B-cell Acute Lymphoblastic Leukemia (B-ALL), and Burkitt lymphoma (BL), into the dataset. This expansion will enable the model to distinguish between specific types of ALL, facilitating more precise and personalized diagnoses.

Furthermore, the next phase involves documenting the outcomes and findings of this project in a comprehensive article. The article will undergo a rigorous review process before being submitted to a reputable scientific journal. Sharing the research results in this manner will contribute to the existing knowledge base in leukemia diagnosis, fostering advancements and progress within the medical community.

REFERENCES

[1] Scotti, F. (2005, September). All-IDB. ALL-IDB Acute Lymphoblastic Leukemia Image Database for Image Processing. From https://scotti.di.unimi.it/all/.

[2] A. Shah, S. S. Naqvi, K. Naveed, N. Salem, M. A. U. Khan and K. S. Alimgeer, "Automated Diagnosis of Leukemia: A Comprehensive Review," in IEEE Access, vol. 9, pp. 132097-132124, 2021, doi: 10.1109/ACCESS.2021.3114059.

[3] Daniel A. Arber, Attilio Orazi, Robert Hasserjian, Jürgen Thiele, Michael J. Borowitz, Michelle M. Le Beau, Clara D. Bloomfield, Mario Cazzola, James W. Vardiman; The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 2016; 127 (20): 2391–2405. doi: 10.1182/blood-2016-03-643544.

[4] Das, P. K., Jadoun, P., & Meher, S. (2020, September). Detection and classification of acute lymphocytic leukemia. In 2020 IEEE-HYDCON (pp. 1-5).

[5] Chand, S., Vishwakarma, V.P. A novel Deep Learning Framework (DLF) for classification of Acute Lymphoblastic Leukemia. Multimed Tools Appl 81, 37243–37262 (2022).

[6] P. K. Das, D. V. A, S. Meher, R. Panda and A. Abraham, "A Systematic Review on Recent Advancements in Deep and Machine Learning Based Detection and Classification of Acute Lymphoblastic Leukemia," in IEEE Access, vol. 10, pp. 81741-81763, 2022, doi: 10.1109/ACCESS.2022.3196037.

[7] P. K. Das and S. Meher, "Transfer Learning-Based Automatic Detection of Acute Lymphocytic Leukemia," 2021 National Conference on Communications (NCC), 2021, pp. 1-6, doi: 10.1109/NCC52529.2021.9530010.

[8] Pałczyński K, Śmigiel S, Gackowska M, Ledziński D, Bujnowski S, Lutowski Z. IoT Application of Transfer Learning in Hybrid Artificial Intelligence Systems for Acute Lymphoblastic Leukemia Classification. Sensors. 2021; 21(23):8025. doi: 10.3390/s21238025.

[9] Arjun Abhishek, Rajib Kumar Jha, Ruchi Sinha, Kamlesh Jha,"Automated classification of acute leukemia on a heterogeneous dataset using machine learning and deep learning techniques",Biomedical Signal Processing and Control,Volume 72, Part B,2022,103341,ISSN 1746-8094,doi:10.1016/j.bspc.2021.103341.

[10] L. D. C. Magpantay, H. D. Alon, Y. D. Austria, M. P. Melegrito and G. J. O. Fernando, "A Transfer Learning-Based Deep CNN Approach for Classification and Diagnosis of Acute Lymphocytic Leukemia Cells," 2022 International Conference on Decision Aid Sciences and Applications (DASA), 2022, pp. 280-284, doi: 10.1109/DASA54658.2022.9765000.

[11] A. Negi, J. Rawat, C. Gupta, S. Joshi and M. Pathak, "Ensemble CAD System for Acute Lymphoblastic Leukaemia Classification," 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), 2022, pp. 687-692, doi: 10.1109/ICIEM54221.2022.9853051.

[12] American Society of Hematology. (2001). ASH Image Bank. ImageBank. From https://imagebank.hematology.org/.

[13] S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186.