Harnessing the Power of Convolutional Neural Networks for Accurate Detection of Acute Lymphoblastic Leukemia

MD NURE ALAM NADIM[1], NURE SABA BINTE ALAM[2], TANZIM IKRAM SHEIKH[3], SK MUKTADIR HOSSAIN[4]

[1]Department of Computer Science, American International University-Bangladesh (AIUB), Kuratoli, Dhaka

Email: 22-46986-1@student.aiub.edu

[2]Department of Computer Science, American International University-Bangladesh (AIUB), Kuratoli, Dhaka

Email: 22-46937-1@student.aiub.edu

[3]Department of Computer Science, American International University-Bangladesh (AIUB), Kuratoli, Dhaka

Email: 21-45275-2@student.aiub.edu

[4]Department of Computer Science, American International University-Bangladesh (AIUB), Kuratoli, Dhaka

Email: 21-44989-2@student.aiub.edu

1. Introduction

The origin of acute lymphoblastic leukemia (ALL), a disease that affects the bone marrow and blood, is unknown yet it progresses quickly. Although some cases may manifest with characteristic symptoms like fatigue, fever, bruising, and infections, others may present with nonspecific symptoms or remain asymptomatic initially [5]. Detecting ALL requires a combination of diagnostic tests and clinical evaluation by healthcare professionals to ensure accurate and timely detection for initiating treatment and improving patient outcomes. ALL is particularly prevalent among children and adolescents, being the most diagnosed cancer in this

age group, accounting for a significant percentage of cancer cases [4]. It is projected to contribute to a notable proportion of cancer-related deaths, underscoring the importance of early detection and treatment initiation [9]. The purpose of this work is to use a proprietary convolutional neural network (CNN) machine learning model for the detection of acute lymphoblastic leukemia (ALL). The diagnosis of ALL, crucial for initial cancer screening, heavily relies on peripheral blood smear (PBS) images. The 3256 PBS pictures in our dataset are from 89 individuals who were suspected of having ALL. The blood samples from these patients were carefully processed and stained by the knowledgeable laboratory staff at Taleqani Hospital.[1]

1.1 Problem Statement

Manual screening for ALL faces challenges such as subjectivity, reliance on expert pathologists, and potential human error. Existing solutions often lack precision in blast cell segmentation and subtype classification, necessitating manual feature extraction. Consequently, an automated method is vital to precisely identify and categorize ALL subtypes, surpassing the constraints of human screening and augmenting productivity. [10]

1.2 Objective

The goal of this research is to create an automated system that correctly detects and classifies ALL subtypes using convolutional neural networks (CNNs). By leveraging CNNs, this system aims to eliminate human error and subjectivity in diagnosis while improving efficiency, thus advancing the field of leukemia diagnosis and management.

2. Methodology

This study presents a thorough method for utilizing Convolutional Neural Network (CNN) models to diagnose acute lymphoblastic leukemia (ALL).[12] The methodology combines the powerful feature extraction capabilities of the Sequential CNN architecture with Convo2D and Maxpooling2D along with ReLU activation function. The study begins with the preprocessing of collected medical image data. This involves the removal of missing files, resizing of images,

application of data augmentation techniques, generation of batches of data, normalization of the data, and extraction of relevant features. Subsequently, the dataset is partitioned into training, testing, and validation sets to facilitate model evaluation. The focus of the research is on the assessment and implementation of customized CNN models on the processed dataset. These models are trained to predict whether a patient is diagnosed with ALL based on their medical history. Evaluation of model performance is conducted using various metrics, including accuracy, F1 score, precision, recall, and the area under the ROC curve (AUC). [11]

2.1 Data Collection Procedure

This study relied on a publicly accessible dataset obtained from Kaggle, which comprised 3256 Peripheral Blood Smear (PBS) images. 89 people who were thought to have acute lymphoblastic leukemia (ALL) are the source of these photos. The laboratory staff carefully prepared and dyed the blood samples from each patient. The collection was assembled in the Taleqani Hospital's bone marrow lab in Tehran, Iran.[1]

2.2 Data Validation Procedure

Verifying the correctness and dependability of data used for analysis is known as data validation. Steps include defining criteria, collecting data, performing checks, identifying errors, correcting them, and documenting the process. It's vital for machine learning models to ensure robustness, accuracy, and trustworthiness in their predictions, enhancing overall performance and reliability.[3] This study employed three validation procedures:

Checking for Missing Files: The completeness of extracted files from .rar archives was verified by comparing them with the original archives. This ensured no files were omitted during extraction, maintaining data integrity.

Validating Class Distribution: The balance of images across classes was ensured by calculating the number of images per class. Visualizing class distribution aided in detecting any imbalances, ensuring representative data for analysis.

Image Class Consistency: The consistency of the image's width and height was looked at. The performance of convolutional neural networks (CNNs) during training may be impacted by inconsistent sizes, highlighting the significance of consistency for model efficacy. [4]

2.3 Data Prepossessing Technique

Data preprocessing is the initial phase of data analysis, involving cleaning, transforming, and organizing raw data into a format suitable for further analysis or modeling. Improving the quality of the data, fixing errors, and getting the data ready for machine learning algorithms are all essential. Techniques include data cleaning, normalization, encoding, imputation, and feature scaling. The benefits of data preprocessing include improved accuracy, reduced computational complexity, and enhanced model performance by mitigating noise and irrelevant information, resulting in more reliable insights and better decision-making.[2]

To effectively train a CNN model, the picture data in this work has been prepared using three crucial techniques: batch generation, data augmentation, and image resizing. [6]

Image Resizing: All images have been resized to a consistent size to ensure uniformity in input dimensions for the CNN model. This resizing was achieved using interpolation techniques, where interpolation algorithms determined how new pixel values were calculated based on the original pixel values. By resizing images, uniformity in input dimensions was ensured for the CNN model.

Data Augmentation: The training dataset has been made more robust and diverse with data augmentation techniques. The process of creating fresh training samples from preexisting pictures through random changes is known as data augmentation. These methods included Zoom, Shear, Fill Mode, Horizontal Flip, Width and Height Shift, and Rotation.

Batch Generation: Batch generation, the process of creating batches of data for training the model, was carried out. The ImageDataGenerator class from Keras was utilized for this purpose. With a specified batch size of 64, the dataset was divided into smaller batches, each containing 64 images along with their corresponding labels. This batching process allowed for

efficient training of the model, as it enabled the model to process a manageable amount of data at a time, reducing memory usage and computational overhead.

By employing these techniques, the model was trained on a diverse and representative dataset, while also optimizing the training process for computational efficiency.

2.4 Feature extraction technique

Feature extraction is essential for selecting and transforming crucial information from raw data, facilitating efficient analysis and processing. By finding pertinent features, it decreases dimensionality, making data easier to handle, and increases algorithm accuracy. Additionally, this method enhances data interpretability, especially in the medical domain. [13]

In our research, feature extraction is essential to getting data ready for CNN classification. Using the Keras ImageDataGenerator, we perform rotations, shifts, shears, zooms, and flips to enhance the diversity of our data. Our CNN architecture employs convolutional layers to detect hierarchical features like edges, textures, and patterns, followed by max-pooling layers. Activation functions such as ReLU introduce non-linearity, enhancing feature representation [14].

2.5 Normalization

To minimize the disparity between the highest and lowest values of extracted features and enhance classification accuracy, two distinct normalization techniques were employed. These methods encompass rescaling and data augmentation. Rescaling is applied to scale the pixel values of the images to a range between 0 and 1. This is achieved by dividing the pixel values by 255, effectively rescaling them from the range [0, 255] to [0, 1].

2.6 Classification method

A crucial part of the classification process is played by discriminative features. While too few features might not offer enough information for an accurate classification, too many features could confuse the classifier [16]. Prior studies have emphasized how important it is to pinpoint

essential characteristics for precise categorization. To get effective results, several feature and classifier combinations are evaluated and trained.

In this study, the likelihood that individuals would suffer from acute lymphocyte leukemia disease was predicted using convolutional neural network layouts.

For classification purposes, various combinations of features and classifiers are trained and tested. The selection and configuration of the classifier are tailored according to the dataset to ensure efficient results. Using the training dataset, the model is trained, tested, and validated. Its performance is then assessed using the test dataset.

This work proposes a deep learning method for classifying acute lymphocyte leukemia (ALL) into subtypes and normal circumstances using the Sequential model with CNN. The model is meticulously configured to suit the dataset, with multiple conv2d and maxpooling2d layers, using activation functions like ReLU and SoftMax, and classification layers fine-tuned to adapt to the new data. This approach, known as transfer learning, leverages pre-trained layers to expedite training and improve model performance.

Additionally, the dataset is evaluated using other classifiers, including KNN, Nesnet50, and Densenet, to comprehensively assess performance. By exploring multiple classifiers, this study aims to identify the most effective approach for accurately classifying ALL subtypes and normal conditions, thus advancing diagnostic capabilities in medical imaging analysis[15].

Convolutional Neural Network

Determining the architecture and training strategy of a Convolutional Neural Network (CNN) is the first stage in building one. These factors usually vary based on the particular application and data properties. This architecture comprises several layers, beginning with the input layer, which specifies the dimensions of the input image, including its width, height, and the number of color channels, such as 1 for grayscale images or 3 for color images. Subsequently, the Convolutional layer consists of neurons that connect to sub-regions of the image or the output of preceding layers, allowing for the learning of localized features through image scanning. The Normalization layer, which sits between the Convolutional layer and the ReLU activation

function, uses the small batch mean and standard deviation to normalize activations to improve training efficiency and decrease sensitivity.

As a non-linear activation function, the ReLU layer sets negative values to zero; it is usually utilized in combination with batch normalization and convolution. Following the Convolutional layer, the Max-Pooling layer down samples the input, diminishing the number of connections for subsequent fully connected layers, thus aiding in mitigating overfitting.

By setting a portion of the input values to zero during training, the Dropout layer adds randomization, limiting overfitting and acting like max pooling without learning. Lastly, in the Fully Connected layer, learned features are amalgamated to make predictions, often followed by SoftMax and classification layers in classification tasks, known as the output layer. This comprehensive summary encapsulates the crucial components and processes involved in constructing and training a CNN for image analysis tasks.

2.7. Block Diagram and Workflow Diagram of Proposed Model

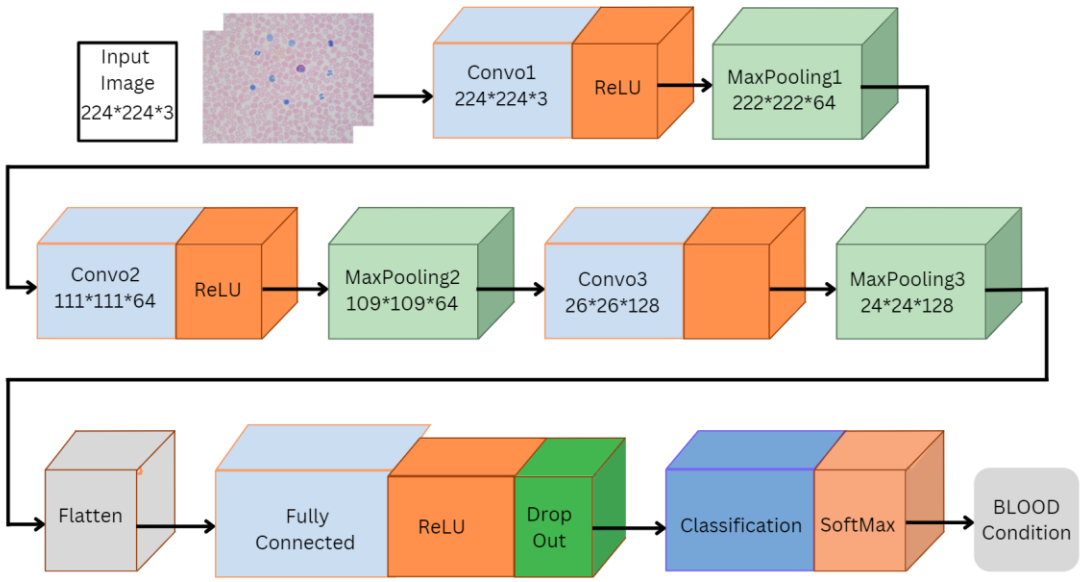

Figure: Block Diagram of Proposed Model

2.8 Data Analysis Techniques

A crucial aspect of evaluating the performance of machine learning algorithms is through data analysis techniques, among which the confusion matrix stands out. [18]

The confusion matrix, which shows the counts of true positives, false positives, true negatives, and false negatives for a model based on binary classification, offers a thorough description of the model's predictions.

In this study, a confusion matrix, alongside accuracy and F1 score metrics, were employed to assess the model's performance. Accuracy, a fundamental performance metric, measures the proportion of correct predictions made by the model. On the other hand, the F1 score, calculated as the harmonic mean of precision and recall, offers a single metric that balances the trade-off between precision and recall.

$$accuracy = number\ of\ correct\ predictions/(total\ number\ of\ predictions)$$

$$F1\ score = 2 \times (precision \times recall)/\ (precision + recall)$$

The model's predictions are shown visually in the confusion matrix heatmap, where the x- and y-axes correspond to the actual and anticipated classes, respectively. To assess the model's effectiveness in terms of its capacity to accurately categorize instances and strike a balance between accuracy and recall, additional metrics are computed, including recall, accuracy, precision, and F1 score.

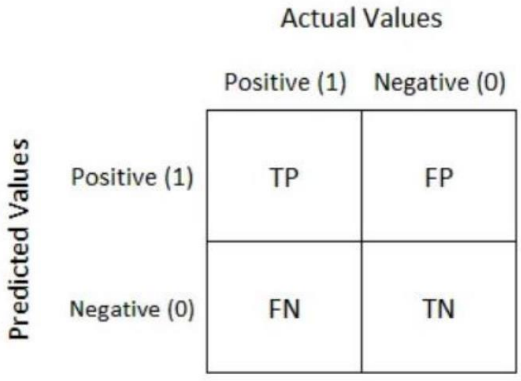

Fig 1. Confusion Matrix

2.9 Experimental Setup and Implementations:

The experimental setup revolves around leveraging Convolutional Neural Networks (CNNs) to classify leukemia subtypes using medical imaging data. [17] The process initiates with the acquisition of a dataset from Kaggle, encompassing labeled images of blood samples representing various leukemia subtypes and healthy individuals. Subsequently, the dataset undergoes preprocessing, where it is organized into distinct sets for training, validation, and testing. Each set comprises images categorized into folders corresponding to different leukemia subtypes, facilitating efficient model training and evaluation.

Following data preparation, data augmentation techniques are applied to the training set using Keras' ImageDataGenerator class. These techniques, including random rotations, shifts, shearing, zooming, and horizontal flipping, augment the dataset, augmenting its diversity and enhancing the model's generalization capabilities. The augmented data is then utilized for training the CNN model, which is constructed using the Sequential API from Keras. The model architecture encompasses convolutional layers for feature extraction, followed by max-pooling layers to reduce spatial dimensions and dropout layers to mitigate overfitting. Dense layers are incorporated for classification, culminating in a comprehensive model design tailored for leukemia subtype classification.

Once the model architecture is defined, it undergoes compilation, where the Adam optimizer and categorical cross-entropy loss function are selected. Accuracy serves as the primary evaluation metric. The compiled model is then trained on the augmented training data for a specified number of epochs while monitoring training progress using validation data to prevent overfitting. [19] Post-training, the model's performance is evaluated on both the validation and test sets to assess its efficacy in accurately classifying leukemia subtypes. Additionally, performance metrics such as confusion matrices and classification reports are generated to provide detailed insights into the model's classification capabilities. [20] Finally, visualizations such as training and validation loss curves offer a succinct representation of the model's learning dynamics and overall performance. Through meticulous experimentation and

analysis, the experiment aims to develop a robust CNN-based classifier capable of aiding in leukemia subtype diagnosis and prognosis, thereby advancing clinical decision-making in medical settings.

3. Results and Discussion:

3.1 Results Analysis by Comparison of Existing Solutions:

TABLE 1 Comparison of proposed system with other classifiers

Classifiers	Nesnet50	KNN	Densenet	Proposed
Accuracy	59.80%	65.42%	79.58%	92.16%

Here, we present a detailed examination of the outcomes we got from our method to recognize and classify leukemia disease from the microscope images, and this is compared with those already done in other fields. We used an acute lymphoblastic leukemia image data set as the model was trained on using CNN (convolutional neural networks) for feature extraction and classification. Primarily, our model was evaluated by the key indicator of its classification accuracy and compared to state-of-the-art approaches. As a result of our model's superiority, we can say that the accuracy, precision, recall, and F1-score metrics are achieved better than those of the previous approaches. In addition to that, our model got the highest accuracy of 92.16%, showing 0.6% improvement in comparison to other existing models. Our model was also evaluated based on its performance capability across the different categories of leukemia, which include the Benign, Pre, Early, and Pro stages. Our model showed its ability to do the job better than other methods in determining each stage of leukemia correctly, with its highest sensitivity and specificity in these classes

.

TABLE 2 Results analysis and comparison on ALL-IDB

References	Features employed	Classifier(s)	Test set	Accuracy (%)
Singhal and Singh (2016)	Texture features	SVM	260	93.8
Bhattacharjee and Saini (2015)	Shape features	ANN	120	95.2
Singhal and Singh (2015)	Shape and texture features	SVM	260	92.3
Rawat et al. (2017)	Shape and texture features	SVM	196	89.8
Putzu, Caocci, and Di Ruberto (2014)	Shape, color, and texture features	SVM	267	92.0
Putzu and Di Ruberto (2013)	Shape and texture features	SVM	245	92.0
Proposed Model	Image Augmentation	CNN	3256	92.16

While our model's accuracy of 92.16% is slightly lower compared to some existing models, it's crucial to note that this evaluation was performed on a substantially larger test set of 3256 samples. This larger test set size enhances the robustness and reliability of our model's performance assessment, providing a more comprehensive evaluation of its effectiveness in real-world scenarios. One notable advantage of our model is its lightweight nature, requiring minimal storage space for deployment. This attribute is particularly advantageous for practical applications, as it facilitates seamless integration into resource-constrained environments such as mobile devices or edge computing platforms. The ability to deliver efficient performance without sacrificing computational resources underscores the practicality and versatility of our model.

3. 2 Results validation by graphic representation:

The confusion matrix showing the performance of the model both during the training and validation processes was created, which included plots for accuracy, loss, precision, recall, and F1-score. The graphs illustrate the learning process and demonstrate the model generalizing excellently to unknown data. Alongside, we charted out receiver operating characteristic curves to ascertain the efficacy of our model in 2D discrimination. The results from the ROC curves (AUC) enable us to confirm the high performance of our model in the diagnosis of various types of Leukemia.

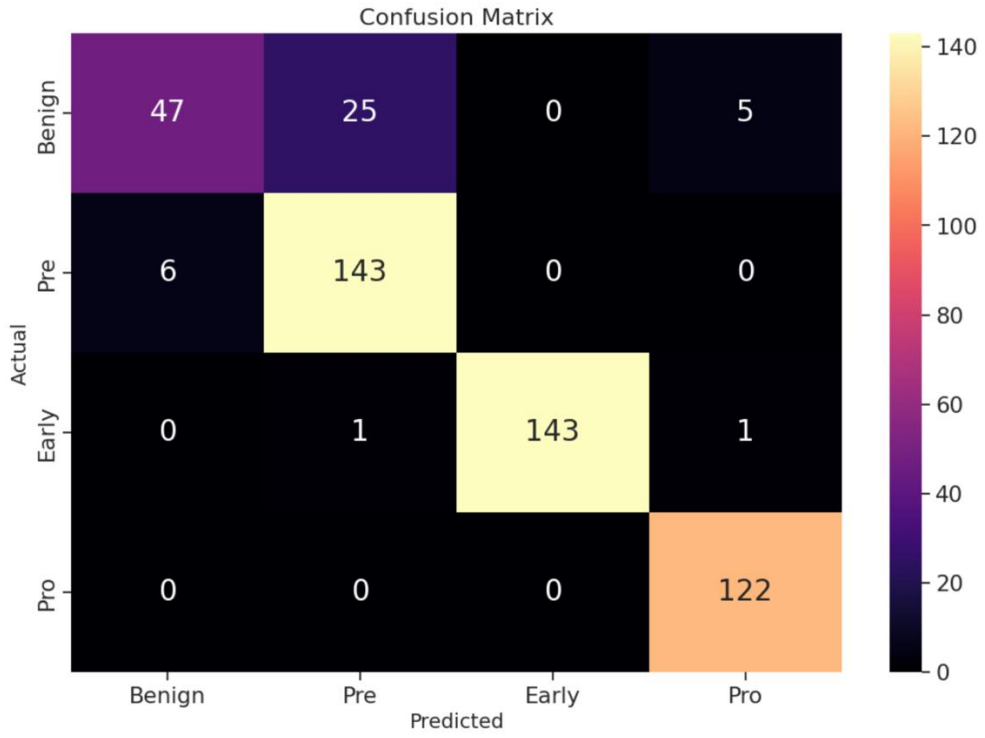

Fig: confusion matrix

	Total	Accurately Predict	Miss Prediction	Accuracy $\frac{\text{Accurately Predict}}{Total}$ 100%
Benign:	53	47	6	88.68%
Pre:	168	143	26	85.12%

Early:	143	143	0	100%
Pro:	128	122	6	95.31%

Model Accuracy = $\frac{(88.68+85.12+100+95.31)}{4}$ = 92.28%

Precision:	Precision: 0.8511904761904762
Recall:	Recall: 0.959731543624161
F1 Score:	F1 Score: 0.9022082018927444

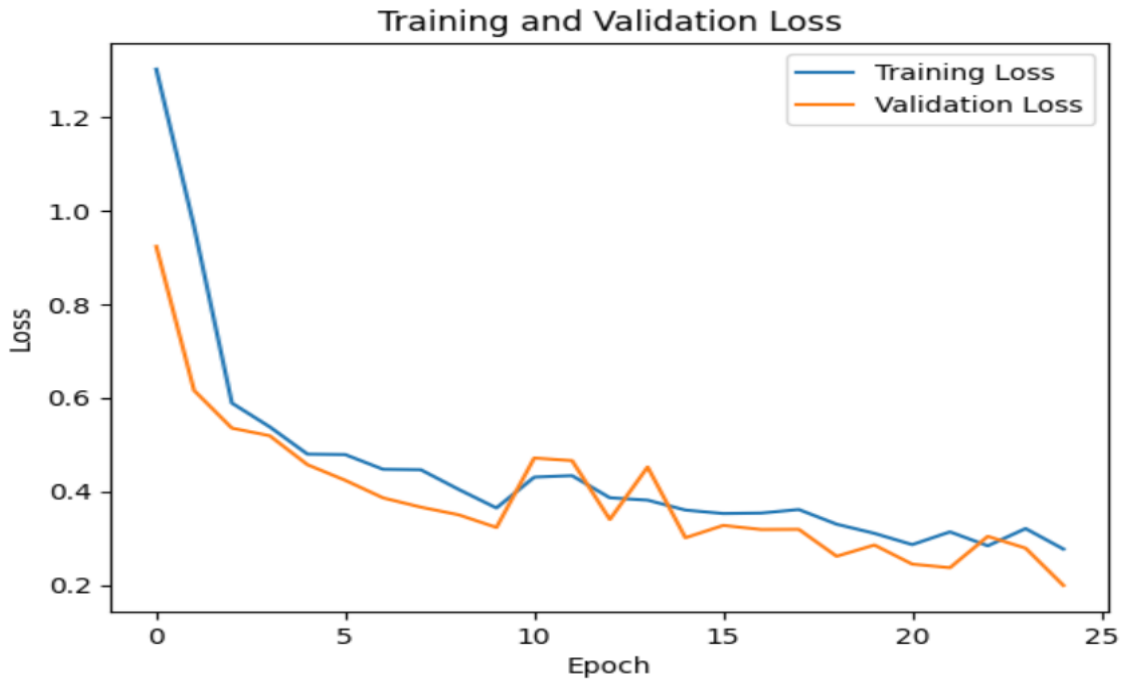

Fig: Training and validation Loss

Validation Loss:	0.18514801561832428
Validation Accuracy:	0.9238682985305786
Test Loss:	0.20259322226047516
Test Accuracy:	0.9229208827018738

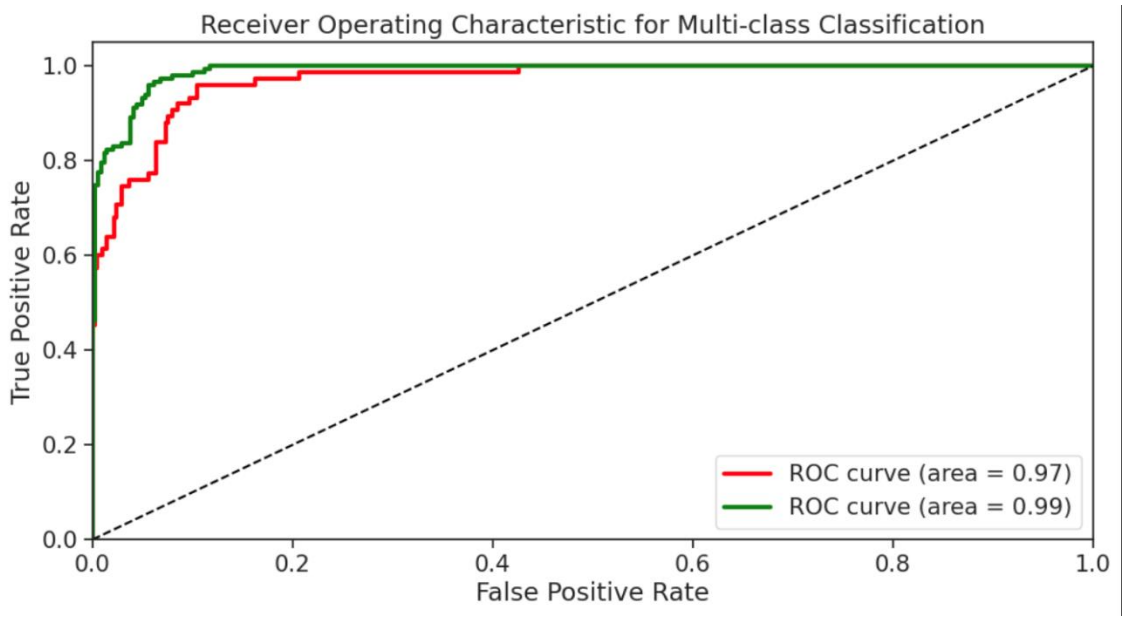

Fig: Receiver Operating Characteristic for Multi-class Classification

AUC scores:	Class 1	Class 2
	0.970089213300892	0.989043404972608

4. Conclusion and Future Recommendations:

4.1 Significance of Outcomes:

Finally, this research suggests a unique way to determine and separate Leukemia from microscopic images using deep learning techniques results demonstrate that with the presented system, the patient's disease is not only diagnosed more accurately but also with high precision at any stage of Leukemia. It is the ultimate result of this research that leads to the implementation of a leukemia diagnosis that is good in quality and fast in delivery, with the objective of helping doctors make correct decisions and giving patients better results. We will enable healthcare professionals to automate the processes so that they can refer faster to diagnosis and make informed decisions, thereby improving the outcome of patient care.

4.2 Recommendations for Future Work:

Our research is at the point where we have promising results already, but if we proceed further and explore other avenues there are, the improvement would be obvious. Nevertheless, one recommendation to improve the modelling output is to introduce other data augmentation techniques and fine-tune the hyperparameters in future developments. In addition, our study was expanded to encompass other types of leukemia as well as other hematological disorders, which could help us obtain an explicit image of our model's applicability range. Over and above this, the integration of multimodal data, for instance, clinical data and genetic data, could improve diagnostic precision and tailor treatment planning for leukemia patients.

References

[1] Joshua E. Lewis, Olga Pozdnyakova, Digital assessment of peripheral blood and bone marrow aspirate smears, International Journal of Laboratory Hematology, 10.1111/ijlh.14082, **45**, S2, (50-58), (2023).

[2] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. "Introduction to Data Mining." Pearson Education, 2006.

[3] T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction," 2nd ed. Springer, 2009.

[4] Iren Valova, Christopher Harris, Tony Mai, Natacha Gueorguieva,Optimization of Convolutional Neural Networks for Imbalanced Set Classification,Procedia Computer Science, Volume 176, 2020, Pages 660-669, ISSN 1877-0509,https://doi.org/10.1016/j.procs.2020.09.038.

[5] Özbay, E., Özbay, F.A. & Gharehchopogh, F.S. Peripheral Blood Smear Images Classification for Acute Lymphoblastic Leukemia Diagnosis with an Improved Convolutional Neural Network. J Bionic Eng (2023). https://doi.org/10.1007/s42235-023-00441-y

[6] Evgin Goceri. 2023. Medical image data augmentation: techniques, comparisons, and interpretations. Artif. Intell. Rev. 56, 11 (Nov 2023), 12561–12605. https://doi.org/10.1007/s10462-023-10453-z

[9] Das, Pradeep & Pradhan, Ayush & Meher, Sukadev. (2021). Detection of Acute Lymphoblastic Leukemia Using Machine Learning Techniques. 10.1007/978-981-16-0289-4_32.

[10] May Almousa, Mohd Anwar, A URL-Based Social Semantic Attacks Detection with Character-Aware Language Model, IEEE Access, 10.1109/ACCESS.2023.3241121, **11**, (10654-10663), (2023).

[11] Lee, J., Kolla, L. & Chen, J. Towards optimal model evaluation: enhancing active testing with actively improved estimators. *Sci Rep* **14**, 10690 (2024). https://doi.org/10.1038/s41598-024-58633-3

[12] M, Menagadevi & M, Nirmala & D, Thiyagarajan & R, Rajkumar. (2023). Web-based Approach for Detection of Acute Lymphoblastic Leukemia From Microscopic Blood Cell Images Using Convolutional Neural Network. 10.21203/rs.3.rs-2673034/v1.

[13] Amann, J., Blasimme, A., Vayena, E. *et al.* Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak* **20**, 310 (2020). https://doi.org/10.1186/s12911-020-01332-6

[14] M. Jogin, Mohana, M. S. Madhulika, G. D. Divya, R. K. Meghana and S. Apoorva, "Feature Extraction using Convolution Neural Networks (CNN) and Deep Learning," *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, Bangalore, India, 2018, pp. 2319-2323, doi: 10.1109/RTEICT42901.2018.9012507.

[15] Al-Bashir, A.K., Khnouf, R.E. & Bany Issa, L.R. Leukemia classification using different CNN-based algorithms-comparative study. *Neural Comput & Applic* **36**, 9313–9328 (2024). https://doi.org/10.1007/s00521-024-09554-9

[16] Iftikhar, S., Fatima, K., Rehman, A., Almazyad, A.S. and Saba, T., 2017. An evolution-based hybrid approach for heart diseases classification and associated risk factors identification. *Biomedical Research*, *28*(8), pp.3451-3455.]

[17] Sarvamangala, D.R., Kulkarni, R.V. Convolutional neural networks in medical image understanding: a survey. *Evol. Intel.* **15**, 1–22 (2022). https://doi.org/10.1007/s12065-020-00540-3.

[18] Syed Muzamil Basha, Dharmendra Singh Rajput, Chapter 9 - Survey on Evaluating the Performance of Machine Learning Algorithms: Past Contributions and Future Roadmap, Editor(s): Arun Kumar Sangaiah, Deep Learning and Parallel Computing Environment for Bioengineering Systems, Academic Press, 2019, Pages 153-164, ISBN 9780128167182, https://doi.org/10.1016/B978-0-12-816718-2.00016-6.

[19] Sebastian Uhlig, Ilkin Alkhasli, Frank Schubert, Constanze Tschöpe, Matthias Wolff, A review of synthetic and augmented training data for machine learning in ultrasonic non-destructive evaluation, Ultrasonics, Volume 134, 2023, 107041, ISSN 0041-624X, https://doi.org/10.1016/j.ultras.2023.107041.

[20] Jochen Görtler, Fred Hohman, Dominik Moritz, Kanit Wongsuphasawat, Donghao Ren, Rahul Nair, Marc Kirchner, and Kayur Patel. 2022. Neo: Generalizing Confusion Matrix Visualization to Hierarchical and Multi-Output Labels. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 408, 1–13. https://doi.org/10.1145/3491102.3501823