Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 4s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

A Health-Based Deep Learning System for Rapid and Precise Detection of Acute Lymphoblastic Leukemia

Dr.R. Sabin Begum[1], Dr.A. Mohamed Anwar[2], Dr.S. Anu Priya[3]

[1]Assistant Professors, Department of Computer Applications, B.S.Abdur Rahman Crescent Institute of Science and Technology,Chennai,Tamil Nadu, India.

[2]Assistant Professor, Department of CSE,Vel Tech Rangarajan Dr.Sagunthala R&D institute of Science and Technology,Avadi,Chennai, Tamil Nadu, India

[3]Assistant Professor, Department of Computer Technology, Peri Institute of Technology,Chennai, Tamil Nadu, India.

[*3]Corresponding Author:

Email ID: sabin@crescent.education, mohdsait69@gmail.com

Cite this paper as: Dr.R. Sabin Begum, Dr.A.Mohamed Anwar, Dr. S. Anu Priya, (2025) A Health-Based Deep Learning System for Rapid and Precise Detection of Acute Lymphoblastic Leukemia. *Journal of Neonatal Surgery*, 14 (4s), 87-94.

ABSTRACT

Acute Lymphoblastic Leukemia (ALL) is a prevalent and life-threatening form of cancer that requires accurate and timely diagnosis for effective treatment. Traditional diagnostic methods for ALL often involve time-consuming and subjective manual examination of blood smears, leading to potential errors and delays in diagnosis. To address these challenges, this project proposes a diagnostic system based on deep learning Convolutional Neural Networks (CNNs) and Streamlit, aimed at achieving fast and accurate classification of Acute Lymphoblastic Leukemia (ALL). The project leverages the power of deep learning CNNs to automatically learn and extract relevant features from microscopic images of blood smears. A Leukemia dataset of annotated blood smear images, consisting of Benign, Early, Precancerous and Prognosis samples, is collected and pre-processed. The images are resized,normalized, and augmented to enhance the robustness and diversity of the training data. The proposed system utilizes the popular CNN architecture VGG16 as the backbone for feature extraction. The pre-trained weights of the CNN model, learned from largescale image datasets, are utilized to initialize the model. The final layers of the CNN are modified to suit the multi-class classification task of distinguishing between Benign, Early, Precancerous and Prognosis samples. To provide an intuitive and user-friendly interface, the Streamlit framework is employed to develop the diagnostic system. The system allows users, including medical professionals, to upload blood smear images and obtain immediate predictions on the presence of Acute Lymphoblastic Leukemia. Overall this paper presents a novel and efficient diagnostic system for the classification of Acute Lymphoblastic Leukemia using deep learning CNNs and Streamlit. The integration of advanced deep learning techniques with a user-friendly interface has the potential to revolutionize the diagnostic process, enabling timely and accurate identification of ALL.

Keywords: *Convolutional Neural Network, Acute Lymphoblastic Leukemia.*

1. INTRODUCTION

Microscopic examination of blood cell images serves as a means to detect various diseases. Alterations in blood composition can indicate the onset of illnesses in an individual. When left untreated, leukemia can be fatal. Leukemia originates within the bone marrow, a delicate substance found inside bones. In the case of leukemia, the bone marrow produces atypical white blood cells. Unlike normal white blood cells, these abnormal ones fail to naturally expire, resulting in an excessive accumulation. Consequently, these irregular white blood cells disrupt the normal functioning of their counterparts.

2. LITERATURE SURVEY

In the existing literature, several researchers have contributed significantly to the development of automated systems for leukemia detection from microscopic images. Madhloom employed image arithmetic operations and thresholding techniques to identify white blood cell nuclei. Sinha distinguished between five types of leukocytes through the utilization of the k-means clustering algorithm and various classification models. Kovalev initially detected cell nuclei and subsequently applied a region-growing technique. Scotti applied threshold operations, a low-pass filter for background removal, and clustering to

segment white blood cells. Puri achieved white blood cell segmentation through edge detection and trained a neural network using morphological features for lymphoblast recognition. Halim introduced an automated system for blast counting and implemented threshold operations on the S component of the HV color space.

Mohapatra employed clustering for white blood cell segmentation and extracted diverse features, including shape, color, texture, fractal attributes, Fourier descriptors, and contour information, while training the system for leukemia recognition. DonidaLabati provided a dataset comprising blood samples from both normal individuals and leukemia patients, which proved invaluable for evaluating our proposed system. Shitong developed a technique combining threshold segmentation, fuzzy logic, and mathematical morphology, demonstrating rapid leucocyte detection, albeit with challenges in nucleus-cytoplasm separation.

Ghosh introduced a method to determine an accurate threshold for leucocyte segmentation using fuzzy divergence and various functions like Gaussian, Gamma, Cauchy, among others. This method excelled in nucleus segmentation but didn't adequately address cytoplasm extraction, a vital aspect in cancer detection. Dorini proposed a scheme for nucleus extraction utilizing the watershed transform based on the image forest transform, with cytoplasm extraction relying on size distribution information. However, this scheme faced limitations when dealing with non-round cytoplasm.

Angulo presented a system featuring a two-stage blood image segmentation algorithm, leveraging binary filtering and automatic thresholding techniques. This system performed well in extracting nuclei, cytoplasm, and nucleoli from lymphocyte images but had increased computation time due to the two-stage segmentation process. Additionally, variations in lighting conditions posed challenges in selecting optimal segmentation thresholds.

The study by Sirinukunwattana et al. represents a significant advancement in the application of deep learning to histopathology. By combining locality-sensitive descriptors with CNNs, the authors have laid the groundwork for automated and accurate analysis of complex tissue structures in cancer diagnosis. This work has inspired further research into leveraging deep learning for improving pathology workflows and advancing cancer diagnostics [8].

Esteva et al. aim to develop a deep learning model capable of classifying skin lesions at a dermatologist-level accuracy. The motivation stems from the increasing global incidence of skin cancer and the critical need for accurate and scalable diagnostic tools, especially in areas with limited access to dermatologists [2].

The authors provide a high-level introduction to deep learning, defining it as a subfield of machine learning focused on algorithms inspired by the structure and function of the brain, particularly neural networks. The paper highlights how deep learning enables computational models with multiple layers to learn data representations with various levels of abstraction [1].

3. RESEARCH OBJECTIVES

The proposed system aims to overcome the constraints of the current leukemia diagnosis system by harnessing advanced technologies, including deep learning and computer vision. Key features and advantages of this proposed system encompass:

- Accelerated Diagnosis: The system significantly expedites the leukemia diagnosis process.
- Enhanced Precision: It offers improved diagnostic accuracy.
- Objectivity and Standardization: The system provides objective and consistent results.
- Scalability: It can efficiently handle a large volume of cases.
- Error Reduction: The automated system minimizes human errors.
- Streamlined Workflow: Integration with existing laboratory systems optimizes workflow.
- Cost-Effectiveness: Despite initial infrastructure and training costs, it leads to long-term savings.

The system leverages deep learning algorithms to automate leukemia diagnosis, drastically reducing analysis time compared to manual blood smear examination. It can swiftly process numerous samples and deliver rapid results. Deep learning models, trained on extensive datasets, enhance accuracy by learning intricate patterns and features from microscopic blood smear images, ensuring reliable leukemia cell detection.

Furthermore, the proposed system assures objectivity and standardization. Deep learning models provide consistent interpretations, mitigating subjective variations among pathologists. Incorporating standardized diagnostic criteria enhances accuracy and reliability. Its scalability is advantageous in high-incidence leukemia areas or outbreak scenarios, diminishing dependence on a limited number of pathologists.

Automation minimizes human errors linked to manual examination. Precise leukemia cell detection eliminates mistakes caused by fatigue, inconsistencies, or oversights during visual examination. The system integrates seamlessly with existing laboratory systems, optimizing data management and workflow. It simplifies tracking, storage, and retrieval of patient data, promoting collaboration among healthcare professionals and overall diagnostic efficiency.

While initial implementation may require investment, long-term cost savings emerge from reduced manual labor reliance, faster turnaround times, and improved accuracy. Consequently, resource utilization is optimized, curtailing expenses tied to conventional diagnostic methods. Overall, the proposed system revolutionizes leukemia diagnosis with the potency of deep learning, computer vision, and automation. It streamlines diagnostics, augments precision, and bolsters efficiency, promising swift treatment initiation and enhanced patient care. Figure 1 depicts the proposed system.

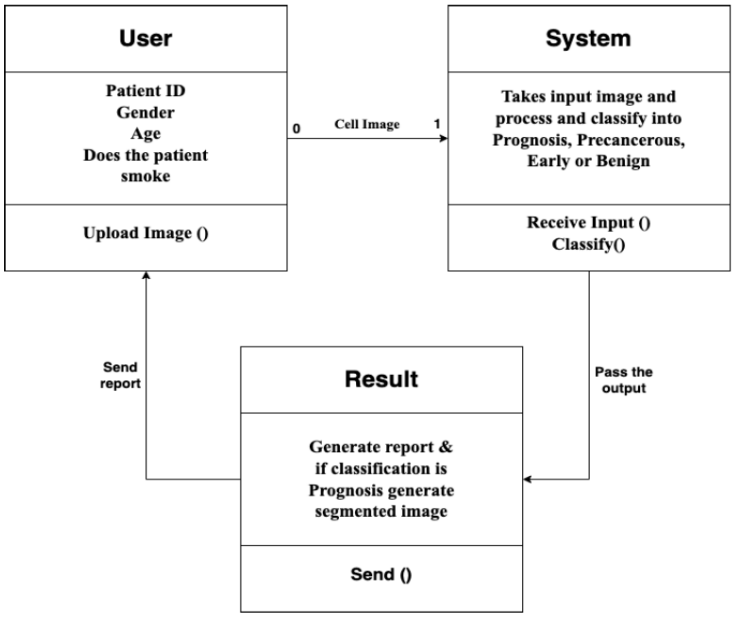

Figure 1: Proposed System

4. METHODOLOGY

The system architecture for leukemia classification employing the VGG16 model comprises various interconnected components working harmoniously to achieve effective and precise leukemia classification. Here the comprehensive overview of the system architecture and fig 2. depicts the system architecture.

4.1 Data Collection and Pre-processing:

The system commences with data pre-processing, involving the collection of a dataset comprising microscopic images of blood smears. These images are sourced from Kaggle.The ALL dataset is meticulously curated, ensuring accurate labeling and annotation of samples representing Benign, Early, Precancerous, and Prognosis cases. Pre-processing steps may encompass tasks such as image resizing, normalization, and augmentation techniques, aimed at enhancing the overall quality and diversity of the dataset.

4.2 Model Creation and Training:

The VGG16 model's pre-trained weights serve as an initial foundation, facilitating the effective extraction of meaningful features from the blood smear images.The VGG16 model is initialized with these pre-trained weights and subsequently customized to suit the specific requirements of leukemia classification. The original output layer of the VGG16 model is replaced with a new fully connected layer, having the appropriate number of nodes corresponding to the leukemia classes to be classified. The customized VGG16 model is then trained using the prepared training dataset. During training, the pre-processed images are fed into the model. The system calculates the loss function and iterates over multiple epochs, fine-tuning the model's parameters to minimize loss, enhance accuracy, and subsequently saves the trained model.

4.3 Streamlit Interface:

The system seamlessly integrates with the Streamlit framework, enabling the creation of an intuitive and user-friendly interface. Streamlit empowers the development of interactive web applications with minimal coding effort. The interface allows various users, including medical professionals, to upload their blood smear images for leukemia classification. Uploaded images undergo processing by the pre-trained VGG16 model, and the resulting classification outcomes are presented to the user in a clear and understandable manner.

5. IMPLEMENTATION

Implementation includes all those activities that take place to convert from the old system to the new one. The new system may be new, replacing an existing system or it may be a major modification to the system currently put into use. In this work,

design is done by using the Python.

5.1 Data collection and pre-processing

Image-based classification dataset is collected from the Kaggle website. Preprocessing include resizing the images to a consistent size, normalizing pixel values, and applying data augmentation techniques such as rotation, flipping, or zooming to increase the diversity of the training data.

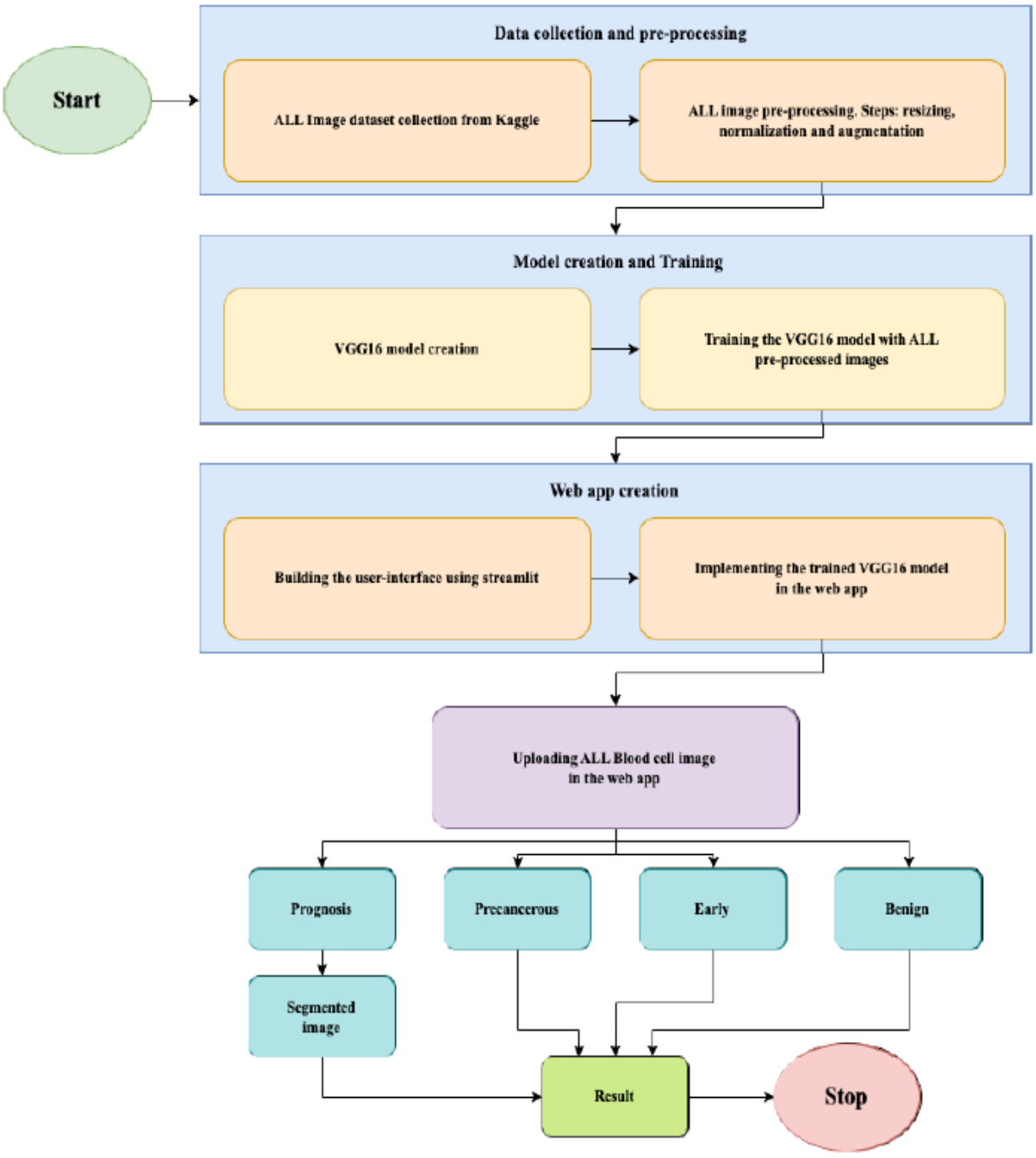

Figure 2: System Architecture

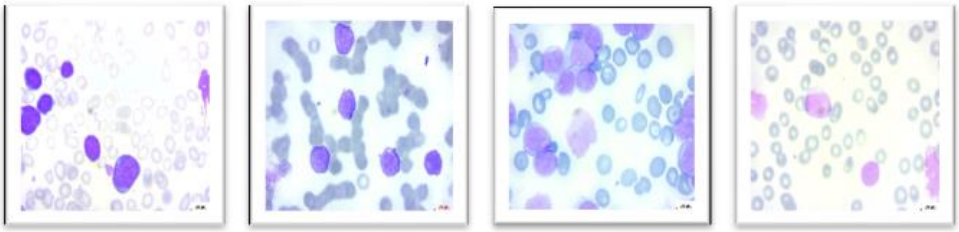

Figure 3: Data Visualization of ALL cancer images

5.2 Model Training and Evaluation

To evaluate the model's performance, it's necessary to split the dataset into two sets: the training set and the test set. The training set is used to train the model, while the test set is used to evaluate its performance. A common split is to allocate around 70% of the data for training and the remaining 30% for testing

VGG16 is a popular deep convolutional neural network (CNN) architecture for image classification. It was introduced by the Visual Geometry Group (VGG) at the University of Oxford. VGG16 is known for its simplicity and effectiveness in capturing intricate image features, making it a widely adopted model in computer vision tasks.Figure 3 shows the data visualization of ALL cancer images

5.2.1. Key Features and Advantages of VGG16 in Leukemia Cancer Classification:

VGG16, a convolutional neural network (CNN) architecture, proves highly effective for leukemia classification. Here are its notable features and advantages in this context:

Deep Network: VGG16 boasts a deep CNN structure with 16 layers, including 13 convolutional layers and 3 fully connected layers. Its depth enables it to learn intricate and hierarchical features from input images, a critical aspect for precise leukemia classification.

Standard Architecture: VGG16 adheres to a standard architecture characterized by small convolutional filters (3x3) and incorporated max-pooling layers. This design choice facilitates the capture of both local and global features present in input images.

Pre-Trained Model: VGG16 is often used as a pre-trained model on extensive image datasets like ImageNet. Pre-training empowers the network to acquire generic features from a diverse array of images. These features can subsequently be fine-tuned for leukemia classification, harnessing the wealth of pre-existing knowledge encoded in the model.

Transfer Learning: VGG16's pre-trained weights can be seamlessly transferred to the leukemia classification task, even in scenarios with limited leukemia-specific data. This approach saves time and computational resources by commencing from a robust initialization point and refining the model specifically for leukemia classification. It mitigates the necessity for an extensive annotated dataset.

Effective Feature Extraction: VGG16's stacked convolutional layers systematically extract high-level features from input images. These features have the capacity to capture crucial patterns and characteristics specific to leukemia cells, thereby facilitating precise classification.

Wide Applicability: VGG16's architecture has consistently demonstrated strong performance across a broad spectrum of image classification tasks. Its versatility renders it suitable for leukemia classification, enabling it to effectively capture distinguishing features of leukemia cells in microscopic blood smear images.

Availability of Pre-Trained Models: Pre-trained VGG16 models are readily accessible in popular deep learning frameworks like TensorFlow and PyTorch. This accessibility simplifies the implementation and utilization of VGG16 for leukemia classification.

Leveraging VGG16 for leukemia classification amalgamates the benefits of deep learning, transfer learning, and pre-trained models. Its capacity to capture intricate features and capitalize on pre-existing knowledge positions it as a potent tool for achieving accurate and efficient leukemia classification.

Model: "model"

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 224, 224, 3)]	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808

Figure 4 VGG16 CNN Layers

5.3 Image Recognition:

VGG16 can be applied to tasks such as object recognition, scene classification, and image segmentation, where accurate classification or localization of objects within images is required.

The selected model is then trained using the labelled images from the training set. During training, the model learns to extract relevant features from the images and make predictions based on those features. This process involves optimizing the model's internal parameters by minimizing a loss function that quantifies the difference between the predicted class probabilities and the true labels of the training images.

Once the model training is complete, its performance is evaluated using the test set.The test set contains images that the model has not seen during training, providing an unbiased assessment of its generalization ability.

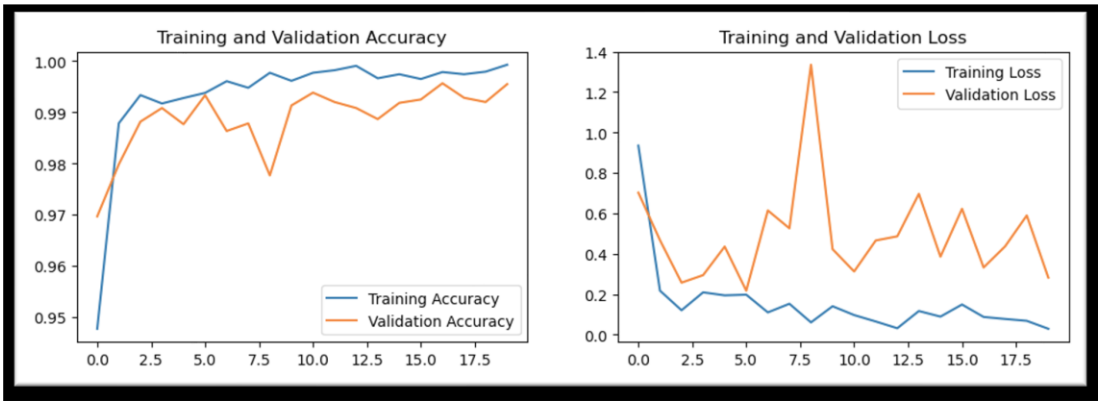

Figure 5 This graphs shows Training and Validation Accuracy and Loss

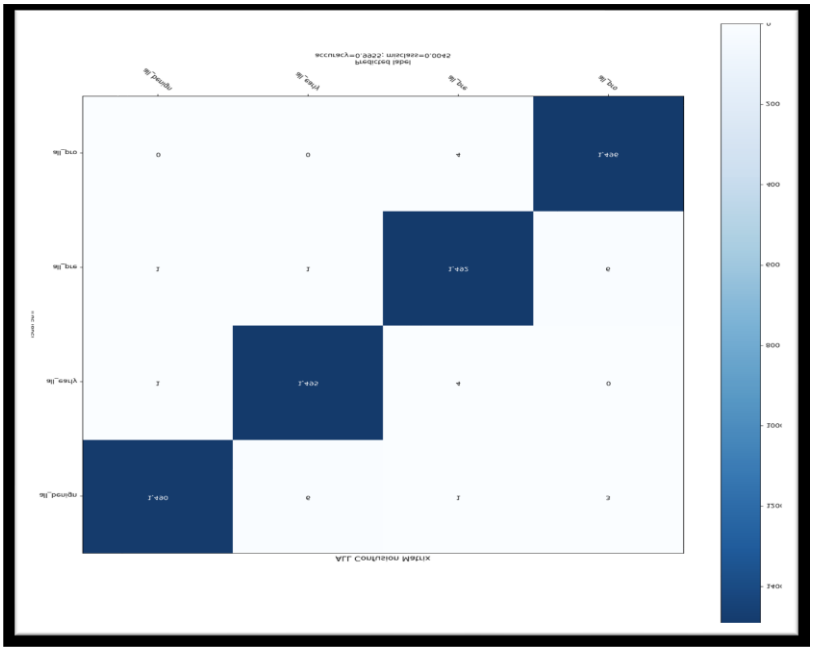

Figure5.7 This confusion matrix shows the accuracy and misclass of predicted

6. CONCLUSION

Leukemia classification using deep learning techniques, such as convolutional neural networks (CNNs), holds great promise for accurate and efficient diagnosis. The application of CNNs enables the extraction of intricate features from microscopic images of blood smears, allowing for robust discrimination between Benign, Early, Precancerous and Prognosis cases. By leveraging large datasets, pre-trained models, and transfer learning, theperformance of CNN-based models, like VGG16, can be significantly enhanced. The use of pre-trained models facilitates the learning of high-level representations from diverse image data, enabling the model to capture subtle patterns and improve classification accuracy.The development of a web app using Streamlit further enhances the practicality and accessibility of leukemia classification. The web app provides a user-friendly interface for medical professionals to upload blood smear images and obtain rapid and reliable predictions. The integration of the classification model with Streamlit simplifies the deployment and usage of the model, making it accessible to a wider range of users.Through the proposed system, the accuracy and speed of leukemia classification are improved, potentially leading to earlier detection and more effective treatment. The automated classification process reduces human subjectivity and assists medical professionals in making accurate and timely diagnoses, aiding in better patient outcomes.

However, it is important to acknowledge that the proposed system is not without limitations. Challenges may arise due to the availability and quality of training data,potential biases in the dataset, and the need for continuous model evaluation and improvement. Collaboration with medical experts, rigorous testing, and ongoing research are essential for addressing these challenges and ensuring the reliability and validity of the classification system.Overall, the development of a leukemia classification system using CNNs and Streamlit provides a valuable tool for medical professionals, supporting faster and more accurate diagnosis, ultimately improving patient care and outcomes in the battle against leukemia.

REFERENCES

[1] LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep learning." Nature, 521(7553), 436-444.

[2] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). "Dermatologist-level classification of skin cancer with deep neural networks." Nature, 542(7639), 115-118.

[3] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sanchez, C. I. (2017). "A survey on deep learning in medical image analysis." Medical image analysis, 42, 60-88.

[4] Cruz-Roa, A., Gilmore, H., Basavanhally, A., Feldman, M., Ganesan, S., Shih, N., ... & Madabhushi, A. (2014). "Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent." Scientific reports, 7(1), 46450.

[5] Mobadersany, P., Yousefi, S., Amgad, M., Gutman, D. A., Barnholtz-Sloan, J. S., Velázquez Vega, J. E., ... & Brat, D. J. (2018). "Predicting cancer outcomes from histology and genomics using convolutional networks." Proceedings of the National Academy of Sciences, 115(13), E2970-E2979.

[6] Wang, D., Khosla, A., Gargeya, R., Irshad, H., & Beck, A. H. (2016). "Deep learning for identifying metastatic

breast cancer." arXiv preprint arXiv:1606.05718.

[7] Lakhani, P., & Sundaram, B. (2017). "Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks." Radiology, 284(2), 574-582.

[8] Sirinukunwattana, K., Raza, S. E., Tsang, Y. W., Snead, D. R., Cree, I. A., & Rajpoot, N. M. (2016). "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images." IEEE transactions on medical imaging, 35(5), 1196-1206.

[9] Bodzas A. (2019). Diagnosis of Malignant Haematopoietic Diseases based on the Automation of Blood Microscopic Image Analysis. Master's thesis, Technical University of Ostrava, Ostrava, CZ.

[10] Chen J., Ying H., Liu X., Gu J., Feng R., Chen T., et al. (2020). A transfer learning based super-resolution microscopy for biopsy slice images: the joint methods perspective. IEEE/ACM Trans. Comput. Biol. Bioinform.

[11] Gao W., Zhu Y., Zhang W., Zhang K., Gao H. (2019). A hierarchical recurrent approach to predict scene graphs from a visualion-oriented perspective. Comput. Intellig. 35 496–516. 10.1111/coin.12202.

[12] Shafique S., Thesin S. (2018). Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks. Technol. Cancer Res. Treatment.