

Digital Image Processing and Python for Acute Lymphoblastic Leukemia (ALL) Detection using Convolution Neural Network (CNN) Algorithm

Dr. M. Sangeetha¹, S. D. Akila²

¹Professor, Department of Computer Science, Dr. N.G.P. Arts and Science College, Coimbatore-48, Tamil Nadu, India

²II M. Sc. Computer Science, Dr. N.G.P. Arts and Science College, Coimbatore-48, Tamil Nadu, India

Abstract: *The timely identification and diagnosis of leukemia, or the precise differentiation of malignant leukocytes at the lowest possible cost in the early stages of the disease, is a significant problem in the field of illness diagnostics. Despite the high prevalence of leukemia, laboratory diagnosis institutes have limited flow cytometer equipment and labor-intensive procedures. The goal of the current systematic review was to examine the literature on machine learning-based leukemia identification and classification. The promise of machine learning (ML) in the diagnosis of diseases served as the driving force behind it. Leukemia is a malignancy of the blood-forming tissues that affects the lymphatic and bone marrow systems. If caught early, treatment options are more effective. This study created a novel classification model for blood microscopic images that can identify between images with and without leukemia.*

Keywords: leukemia diagnosis, machine learning, blood microscopic images, systematic review, cost-effective

1. Introduction

Blood is made up of plasma and three distinct cell types: platelets, red blood cells, and white blood cells. Each type of cell in blood has a specific function. Oxygen is carried by red blood cells from the lungs to the body's tissues and vice versa. The body uses white blood cells to fight off infections and illnesses. Platelets aid in clotting and manage hemorrhaging. Leukemia is a blood cell malignancy characterized by an excessive growth in white blood cells, which are immature blood cells that disrupt other blood cells, primarily red blood cells and platelets. The ratio of white blood cells in our body is 1000:1. It indicates that there is one white blood cell for every 1000 red blood cells. Leukemia can develop in two different types of white blood cells, and they are: lymphoid cells, Myeloid cells. Lymphocytic or lymphoblastic leukemia is the name for leukemia caused by lymphoid cells; myeloid-caused disease is known as myelogenous or myeloid leukemia. Leukemia is classified as either acute or chronic, depending on the rate at which the cells divide into new ones. Immature blasts, or young cells that are malfunctioning, are typically the aberrant blood cells in acute leukemia. These cells are proliferating quickly. If acute leukemia is not treated right away, it will worsen swiftly. Chronic leukemia produces mature functional cells in addition to young blood cells. The growth of blasts in chronic leukemia is sluggish. The illness takes longer to worsen. There are four main types of leukemia: ALL, or Acute Lymphoblastic Leukemia, Acute Myelogenous Leukemia (AML), Chronic Myelogenous Leukemia (CML), Chronic Lymphocytic Leukemia (CLL).

2. Problem Definition

The Leukemia & Lymphoma Society estimates that 174,250 Americans will receive a diagnosis of leukemia, lymphoma, or myeloma in 2018. Blood cancer diagnoses occur in the United States about every three minutes. The National Cancer Institute estimates that there will be around 61,780 new cases of cancer in 2019, or 3.5% of all new cases. Similar to acute

leukemia, the patient would pass away in a matter of months if treatment was not received in a timely manner. Furthermore, treating this kind of cancer, or any sort of cancer, requires early cancer detection.

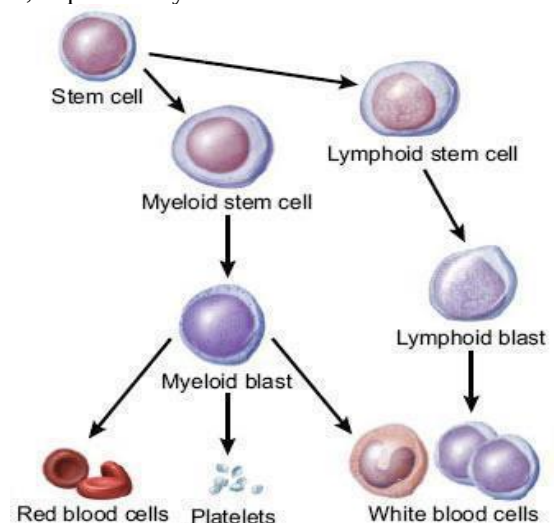


Figure 1: The formation of Myeloid and Lymphoid series of cell. (From Wikimedia Commons, the free media repository)

3. Objective

Our project's goal is to create a system that can automatically identify cancer from pictures of blood cells. This technology determines whether or not a blood cell is cancerous by using a convolution network to process the image of the cell. On blood cell pictures, the look of cancer is frequently nonspecific, can coexist with other diagnoses, and can resemble a variety of different benign abnormalities. Medical professionals' diagnostic variability in cancer diagnosis is significantly influenced by these disparities. In addition to being extremely helpful in clinical settings, automated cancer detection from blood cell images at the expert level of medical practitioners would be essential in providing healthcare to communities lacking access to diagnostic imaging specialists.

4. Scope and Application

We create a technology that can identify cancer from blood cell images more accurately than licensed medical professionals. In regions of the world where access to qualified medical staff is scarce, this technology can enhance the delivery of healthcare and expand access to medical imaging knowledge.

5. Literature Review

Researchers have devised a number of methods to identify leukemia. Convolution Neural Networks (CNNs), which are based on computer vision, are among the most widely utilized techniques nowadays. Image pre-processing, clustering, morphological filtering, segmentation, feature selection or extraction, classification, and evaluation are the main phases that make up the conventional algorithm for this approach.

6. Existing Methods for Diagnosis

Medical history and physical examination: A list of the patient's prior medical issues as well as current symptoms. A person's family medical history can also be helpful in the diagnosis of leukemia. **Complete blood count (CBC):** RBC, WBC, and platelet counts are measured under a microscope after blood is drawn. **Aspiration of bone marrow:** A needle is used to remove bone marrow from the breastbone. Under a microscope, the extracted sample is examined to search for aberrant cells. **Cytogenetic analysis:** To detect individual chromosomes, a cytogenetic test uses bone marrow or blood. It displays chromosomal aberrations that aid in leukemia diagnosis and type identification. Usually, results are ready in three weeks. **Immunohistochemistry:** In immunohistochemistry, blood samples of cells are treated with certain antibodies. The color shift can be detected under a microscope. It aids in identifying the many kinds of cells that are there.

7. CNN Techniques

We have suggested a segmentation technique to extract the cytoplasm and nucleus regions from stained blood smear pictures using color-based clustering. SVM classifiers are used with pertinent features and produce good outcomes. have suggested using peripheral blood pictures to automatically identify white blood cells (WBCs) and classify them into five categories: eosinophil, basophil, neutrophil, monocyte, and lymphocyte. SVM with a granularity feature is used to first classify eosinophil and basophil from other WBCs. Convolutional neural networks are then used to identify the other three categories and extract features; random forests then use these features to classify those WBCs.

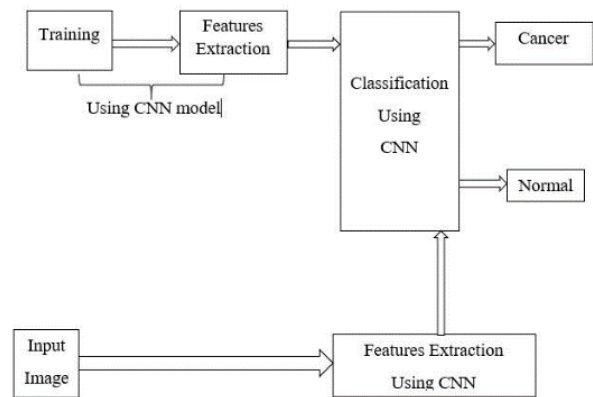


Figure 2: Block Diagram (from Deep-CNN Model for Acute Lymphocytic Leukemia (ALL) Classification Using Microscopic Blood Images: Global Research Prasanna Ranjith Christodoss, Rajesh Natarajan)

7.1. CNN Over Other Algorithms

Before CNN gained popularity, many other algorithms were employed for picture classification. In the past, characteristics were extracted from photos and fed into algorithms for categorization such as Support Vector Machines (SVM). A few algorithms further employed the image pixel level values as a feature vector. As an illustration, an SVM may be trained with 784 features, each of which represents a pixel value from a 28 x 28 image. CNNs can be compared to automatic picture feature extractors. While utilizing a pixel vector approach results in a significant loss of spatial interaction between pixels, a CNN efficiently down samples the image via convolution and subsequently employs a prediction layer at the end. Yann Le Cun initially introduced this idea in 1998 while using a single convolution layer for digit categorization. Later, Alex Net popularized it in 2012 by achieving state-of-the-art on image net by using many convolution layers. They are now the preferred algorithm for picture classification problems as a result.

7.2 Working of CNN

For the purpose of extracting features and classifying blood samples, CNN has been deployed.

A CNN is a type of multilayered neural network designed to identify intricate elements in data through a unique architecture. CNNs have been applied to image recognition, robot vision, text recognition in images, and autonomous cars. The CNN is designed for two-dimensional pattern recognition and consists of layers of neurons. Convolutional, pooling, and fully linked layers are the three types of layers found in CNN. With the exception of the input layer, our network has 11 layers. An RGB color image that has each color channel processed independently is fed into the input layer.

Convolution layer comprises the first six layers of a convolution network. An image in the layer is subjected to 16 3*3 filters by the first two convolution layers. An image is subjected to 32 3*3 filters by the other two layers. Additionally, an image is subjected to 64 3*3 filters by the final two convolutional layers. The ReLU activation function is used by the nonlinear transformation sublayer. The image is subjected to a 2*2 filter applied by the max pooling

sublayer, causing the image size to be reduced by half. At this stage, the convolution network extracts 64 features, one for each color channel represented by a 32*32 array.

The layer that flattens is the eighth layer. By simply concatenating the multidimensional array's entries together, the flatten layer converts a multidimensional array into a one-dimensional array. This flatten layer produces a 4800-size one-dimensional array as its output. The fully linked ANN with the ReLU activation function, which maps 4800 input values to 64 output values, is the ninth layer. The dropout layer is the tenth layer. To lessen the issue of overfitting, 50% of the input values entering the layer are set to zero. A fully linked ANN with a sigmoid activation function, which transfers 64 input values to two class labels, makes up the eleventh and final layer.

Using the training set data, we first train the convolution network to determine the weights of the suitable filters in each of the three convolutional sublayers and the weights that produce the least amount of error in the two fully connected layers. The validation error and cross-entropy loss are then obtained by evaluating the convolution network with the data from the validation set. Until we have completed ten epochs, we repeat the convolution network training process in this manner. Finally, we use the test set data to assess the convolution network's performance.

7.3 Classification

Blood samples are automatically screened for cancer using neural networks. Because neural networks are well-known for being effective classifiers in a wide range of real-world applications, they are frequently used as classification tools. One of the key phases in creating a precise process model with CNNs is training and validating the model. The training features set is used to train the CNN model, and the testing features set is used to confirm the accuracy of the learned model using the feed-forward back propagation network. These two portions make up the dataset used for training and validation operations.

Connection weights were constantly changed during the training phase until they reached the predetermined iteration number or an appropriate mistake. Blood samples are automatically screened for cancer using neural networks. Because neural networks are well-known for being effective classifiers in a wide range of real-world applications, they are frequently used as classification tools. One of the key phases in creating a precise process model with CNNs is training and validating the model.

8. Result and Analysis

Project ultimate goal is to accurately diagnose cancer using the confusion matrix, loss and accuracy graph, and iterations that were collected. We've run twenty iteration series, and from these it's evident that the loss is getting smaller every time. Loss is a function of model accuracy. In order to achieve our goal of minimizing the loss function, our model has flawlessly dropped the loss value from the beginning, and with each iteration, we have gotten closer to the minimum.

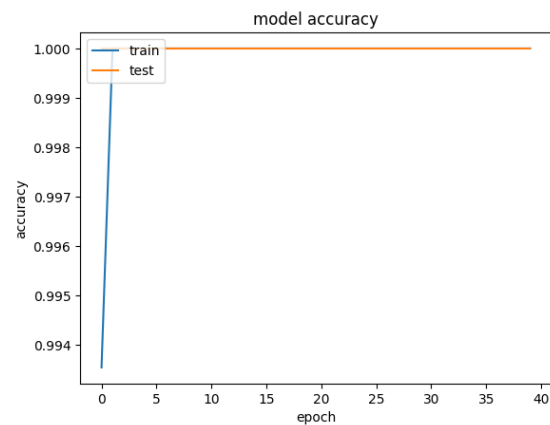


Figure 3: Epoch vs Model Accuracy

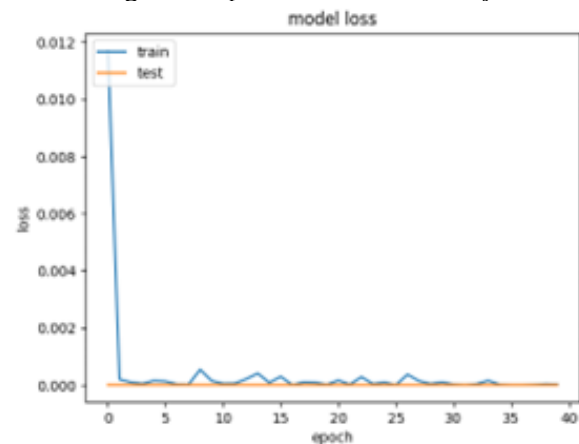


Figure 4: Epoch vs Model Loss

We then calculated the accuracy and loss curve to get our model's best outcome. These learning curves, which include a loss and accuracy curve, illustrate how well our model performs on training and validation sets in relation to the quantity of training iterations.

Our loss curve is showing that loss is diminishing and providing the best result, since it is lowering with each iteration. However, our accuracy curve shows that our model is improving with every iteration, indicating that it is becoming a more sophisticated learner.

9. Conclusion

In summary, a potential method for the identification of acute lymphoblastic leukemia (ALL) is the combination of Python-based Convolutional Neural Network (CNN) algorithms with digital image processing techniques. Medical diagnostics has advanced significantly with the use of CNNs, which are able to recognize complex patterns in images, in conjunction with Python's flexibility and large image processing library.

Digital image processing can be used to accurately and efficiently extract several aspects from microscopic blood cell images that are suggestive of ALL. Python's large ecosystem of deep learning frameworks, including TensorFlow and PyTorch, together with its ease of use and robust community support make it a perfect programming language for building CNNs.

Early diagnosis and intervention are made easier by the automated and accurate detection of ALL made possible by the use of CNN algorithms and digital image processing techniques. By facilitating prompt treatment, this not only increases the effectiveness of medical personnel but also has the potential to improve patient outcomes.

Additionally, Python's adaptability makes it possible to integrate seamlessly with current healthcare systems, opening the door for scalable and easily accessible diagnostic solutions. To guarantee these algorithms' dependability and efficacy in actual clinical situations, it is crucial to continuously improve and optimize them through stringent validation and testing.

Essentially, the integration of Python-based CNN algorithms with digital image processing is a novel method to ALL detection, providing a road towards more effective, precise, and easily accessible diagnostic instruments in the medical imaging and healthcare domains.

with improved swarm optimization of deep features. Sci Rep 10(1):2536

References

- [1] S. Mohapatra, D. Patra, and S. Satpathy, "Unsupervised blood microscopic image segmentation and unsupervised blood microscopic image segmentation and leukemia detection using color based clustering," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 4, pp. 477–485, 2012.
- [2] J. Zhao, M. Zhang, Z. Zhou, J. Chu, and F. Cao, "Automatic detection and classification of leukocytes using convolutional neural networks," *Medical & Biological Engineering & Computing*, vol. 55, no. 8, pp. 1287–1301, 2016.
- [3] Joshi MD, Karode AH, Suralkar S. White blood cells segmentation and classification to detect acute leukemia. *Int J Emerging Trends Technol Computer Sci (IJETICS)*. 2013;2(3):147–151.
- [4] Abhishek A, Jha RK, Sinha R, Jha K (2022) Automated classification of acute leukemia on a heterogeneous dataset using machine learning and deep learning techniques. *Biomed Signal Process Control* 72:103341
- [5] Abhishek A, Santhanam N, Jha RK, Sinha R, Jha K (2022) Multi Class Classification of Acute Leukemia using Transfer Learning. In 2022 International Conference for Advancement in Technology (ICONAT). IEEE
- [6] Su J, Han J, Song J (2021) A benchmark bone marrow aspirate smear dataset and a multi-scale cell detection model for the diagnosis of hematological disorders. *Comput Med Imaging Graph* 90:101912
- [7] Vogado Rodrigo LHS, Veras Flavio MS, Araujo Romuere HD, Silva Kelson RV, Aires RT (2018) Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification. *Eng Appl Artif Intell* 72:415–422
- [8] Kalaiselvi TC, Kumar DS, Subhashri KS, Siddharth SM (2020) Classification of leukemia using convolution neural network. *Eur J Mol Clin Med* 7(4):1286–1293
- [9] Sahlol AT, Kollmannsberger P, Ewees AA (2020) Efficient classification of white blood cell leukemia