Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

IDENTIFICATION OF CANCER CELLS IN BLOOD USING DEEP LEARNING TECHNIQUES

KV Satyanarayana, Associate Professor, Computer Science Engineering Department, Raghu Engineering College Visakhapatnam, Andhra Pradesh.

Y Hema Varshini, P Neelaveni, P Hemanth Kumar IV Year BTech Students, Computer Science Engineering Department, Raghu Engineering College Visakhapatnam, Andhra Pradesh.

[*1]vsatyanarayana.kalahasthi@gmail.com

[3]19981a05c7@raghuenggcollege.in

[2]19981a05i1@raghuenggcollege.in

[4]19981a05c5@raghuenggcollege.in

Abstract

Hematological cancer, also referred to as blood cancer, affects the lymphatic, bone marrow, and blood systems. Effective blood cancer therapy and management depend on early diagnosis of the disease. Deep learning algorithms have recently become a viable strategy for the early diagnosis of blood cancer. These techniques involve the separation of cancerous and non-cancerous samples using deep learning algorithms after convolutional neural networks (CNN) have extracted features from blood samples or photos. Various studies have suggested different deep learning-based methods, including hybrid models and feature selection, to enhance the accuracy of blood cancer detection [1]. Results indicate that deep learning techniques can achieve high accuracy in detecting blood cancer and have the potential to improve the diagnosis and treatment of this disease [2]. The paper also emphasizes the future directions of this field. However, further research is needed to develop more precise and dependable models for clinical use.

Keywords: ***Leukemia, WBCs, deep learning models, blood smear images.***

I. INTRODUCTION

The body's defense against diseases and infections caused by pathogens like bacteria, viruses, and fungi that enter the body from the outside depends on leukocytes, commonly known as white blood cells. The lymphatic and circulatory systems both carry these blood cells around, which are produced in the bone marrow.

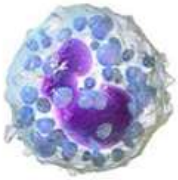

Figure 1: White blood cell

White blood cells, red blood cells, and platelets are the three different cell types that make up blood; each has a particular job to do. Blood also contains plasma [3]. Red blood cells carry oxygen from the lungs to the body's tissues and the other way around. White blood cells are used by the human body to fight diseases and infections. Blood clots and bleeding are controlled by platelets. White blood cells that are immature and proliferate in large numbers in leukemia prevent the development of remaining blood cells, typically platelets and red blood cells. It has been discovered that human body has the white blood cell ratio as 1000:1 [1]. Therefore, there are 1000 red blood cells to 1 white blood cell. Leukemia can grow in two separate types of white blood cells, which are as follows:

• Lymphocytes

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

• Myeloid tissues

Lymphocytic or lymphoblastic leukaemia is the name given to leukaemia produced by lymphoid cells, whereas myelogenous or myeloid leukaemia is the name given to leukaemia caused by myeloid cells. Acute and chronic leukemia are separated based on how quickly the cells are replicating. Young cells or immature blasts, are often the aberrant blood cells that are dysfunctional in acute leukaemia. These cells are expanding quickly. If acute leukaemia is not treated promptly, it swiftly gets worse. Chronic leukaemia produces mature functional cells in addition to young blood cells [4][5]. The growth of blasts in chronic leukaemia is sluggish. The illness takes longer to worsen.

There are four primary forms of leukemia:

1. ALL (Acute lymphoblastic leukemia)
2. AML (Acute myelogenous leukemia)
3. CLL (Chronic lymphocytic leukemia)
4. CML (Chronic myelogenous leukemia)

The organizational structure of this paper is as follows:

The section II consists of Literature survey which explains about earlier research, section III is Methodology, tells about proposed approach and workflow. Results of the suggested methodology are shown in section IV, which also compares them to those of earlier methods and approaches. The study concludes with a discussion of its potential for future development and increased dependability.

II. LITERATURE SURVEY

A. Smear Blood Images for Leukemia Detection and Classification Using Machine Learning.

This review paper by H. T. Salah et al. offers a complete and well-organized summary of the state of all known machine learning (ML)-based leukemia detection and classification models that analyze PBS images. When used to detect leukemia in PBS images, machine learning could yield incredible results, based on the average accuracy of the machine learning techniques used to analyze PBS pictures for leukemia. In terms of sensitivity and accuracy, deep learning (DL) surpassed all other machine learning (ML) algorithms in classifying different leukemia cases.

B. Machine Learning and Microscopic Image Analysis for Leukemia Detection

In this study, the Faster-RCNN machine learning method is employed to forecast the likelihood that cancer cells would develop [6]. Here, two loss functions are applied to the classifier model and the RPN (Region Convolutional Neural Network) model to identify comparable blood objects. Following object identification, calculation of the corresponding item, and detection of leukemia based on the count of the corresponding object.

C. Machine Learning Techniques for Leukemia Disease Detection and Classification: A Review.

To provide a summary of the results that will be helpful to other researchers, the authors of this study examine various machine learning and image processing techniques that are used to categorize leukemia diagnoses [7]. They also make an effort to concentrate on the advantages and disadvantages of other connected studies. The authors draw the conclusion that a variety of modern ML methods can be used to categorize the leukemia sickness. Nonetheless, it is preferable to employ deep learning architectures for classifications when there is a huge dataset of images.

III. METHODOLOGY

Dataset Explanation

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

The images used in this research are taken from the online public dataset from Kaggle platform. There were two sections to this dataset. There were 4961 training photos in all, 2483 of which featured healthy people, and 2478 of which featured blood cancer patients. Using 620 photos from each class, we used a total of 1000 to evaluate the model. The resolution of these pictures was 320x240. Three main steps make up the suggested strategy.

A. Preprocessing.

The deep learning model was trained using the method described in this work, which involved three steps. The subsequent actions are:

A. Convert to RGB: All leukemia pictures are converted to the RGB color model during this step.
B. Resize all to 227x227: Because each image was captured using a different device, each one has a different pixel scale [7]. With this operation, all photographs will have a fixed dimension of 227x227 pixels.
C. Data Augmentation.

B. Pre trained deep learning model and feature extraction.

Convolutional neural networks with transfer learning (AlexNet) were employed to train our model. Three different types of layers make up CNN's architecture: convolutional, pooling, and fully linked layers.The initial (convolutional)layer only learns low-level features. Increasing the layers will increase the features and the network will learn more specific training task patterns [8]. In our work, the pooling layer was employed to cut down on the number of features in the final data and to overcome the overfitting problem. Convolutional layers calculate the output by performing the activation function (rectified linear unit, ReLu).

C. Step3:Classification with deep learning models.

The CNN is created to recognize patterns in two dimensions and is composed of layers of neurons. CNN employs three different types of layers: convolutional, pooling, and fully connected. With the exception of the input layer, our network has 11 tiers. A RGB color image with individual processing for each color channel is fed into the input layer [8]. The convolution network's first six layers are known as the convolution layer. In the first two convolution layers, a picture is subjected to 16 of 3x3 filters. 32 of 3x3 filters are applied to an image by the final two layers. The below figure 2 shows the workflow of the proposed system.

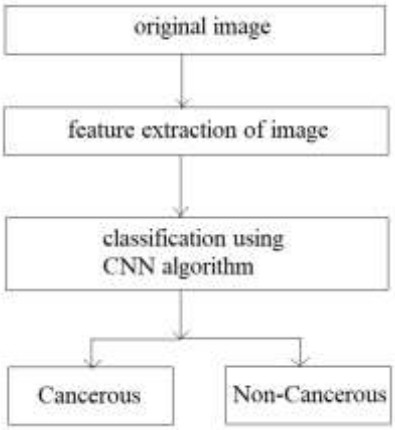

Figure 2: Workflow of CNN algorithm

And the image is subjected to 64 3x3 filters on the final two convolution layers. The ReLU activation function is utilized by the nonlinear transformation sublayer. The image is subjected to a 2*2 filter in the max pooling sublayer, which reduces the size of the image to half. A 32x32 array has been created for each of the 64 features that the convolution network has so far extracted, one for every color channel.

There is an eighth layer called flatten layer, simply concatenates the entries of the multidimensional array, the flatten layer reduces a multidimensional array to a one-dimensional array. A one-dimensional array with a size of 4800x4800 is the layer's output. The ninth layer, ANN is a fully connected with a ReLU activation function which converts 4800 input values into 64 output values. The eleventh layer down is the dropout layer [9]. Using an ANN with full links and the sigmoid activation function to convert 64 input values into two class labels, makes

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

up the eleventh and final layer.

In order to determine the weights for the suitable filters in the convolutional sublayers, the weights that result in the least amount of error in the 2 fully connected layers, we first train a convolution network by making use of the data from the training set [9]. The information from the validation set is then used to determine the cross-entropy loss and validation error of the convolution network. Until we have trained the convolution network for 10 epochs, this procedure is repeated. The data from the test set is then used to evaluate the convolution network's performance [10].

Classification

Cancer in blood samples is automatically detected using neural networks. Because the neural network technique is well known for being a promised classifier for most of the real-world scenarios, it is employed as a classification tool. For the purpose of creating a comprehensive process model utilising CNNs, the validation and training procedures are crucial. Using a feed-forward back propagation network, the CNN model is trained and the correctness of the trained model is evaluated using a set of testing features. These two sections of the dataset are used for the training and validation processes [10]. Throughout the training phase, the connection weights were continuously adjusted until they hit the desired iteration threshold or the suitable error. Cancer in blood samples is automatically detected using neural networks.

Confusion Matrix

The effectiveness of a categorization model is evaluated using a confusion matrix, a table. It is a matrix that shows the proportion of right predictions and incorrect predictions made by the model.

The confusion matrix is frequently used in supervised learning scenarios, where the model is trained on labelled data and then used to predictions on fresh, unlabelled data. The model's performance on these fresh data can be assessed using the confusion matrix.

The true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) numbers make up the confusion matrix. If the model accurately or erroneously predicted the positive or negative class determines these values.

The confusion matrix can be used to calculate various performance metrics such as accuracy, precision, recall, and F1-score, which can further help to evaluate the model's performance.

For 1000 samples our model made the following predictions:

No.of predictions=1000	Actual: Yes	Actual:No
Predicted:Yes	40(TP)	5(FN)
Predicted:No	5(FP)	950(TN)

Table 1: Confusion Matrix

Accuracy = (TP+TN) / (TP+TN+FP+FN)

= (40+950) / (40+5+5+950)

= 0.98 or 98%

Precision = (TP) / (TP + FP)

= (40) / (40+5)

= 0.88 or 88%

Recall = (TP) / (TP + FN)

= (40) / (40+5)

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

= 0.88 or 88%

F1-score = 2 * (Precision * Recall) / (Precision + Recall)

= 2*(0.88 * 0.88) / (0.88 + 0.88)

=0.89

IV. RESULTS AND ANALYSIS

The application of deep learning and machine learning models for the identification of blood cancer has been the subject of numerous studies.

Although there is a lot of promise for the application of machine learning for blood cancer detection, there are still issues that must be resolved, including the lack of readily available labelled data and the requirement for more reliable and understandable models.

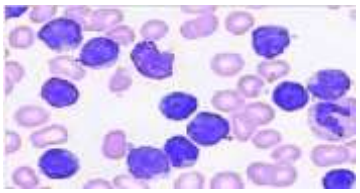

Figure 3.1: Color image

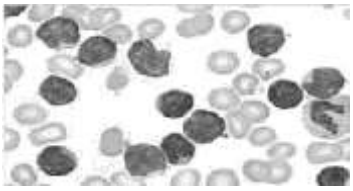

Figure 3.2 Gray scale image

Convolutional neural networks (CNNs), in particular, have demonstrated promising outcomes in the detection of blood cancer. These models may identify characteristics in medical images (such as microscopic images of blood) and utilize those characteristics to categorize the images as cancerous or non-cancerous. In several studies, CNNs have outperformed traditional machine learning models. Using CNN model provides the efficient results with accuracy of 98%.

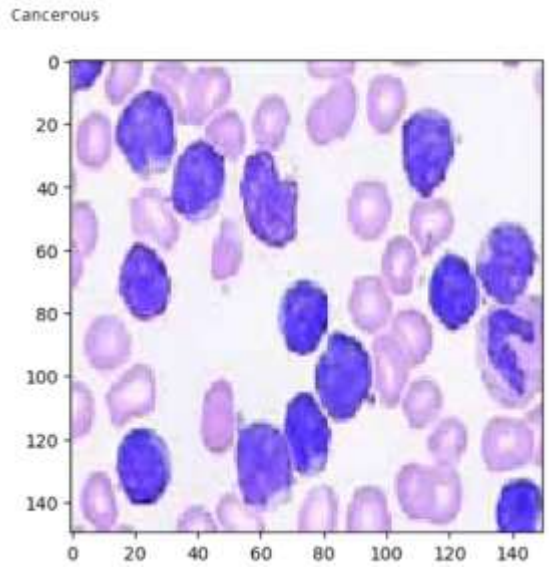

Figure 4: Cancerous Cell

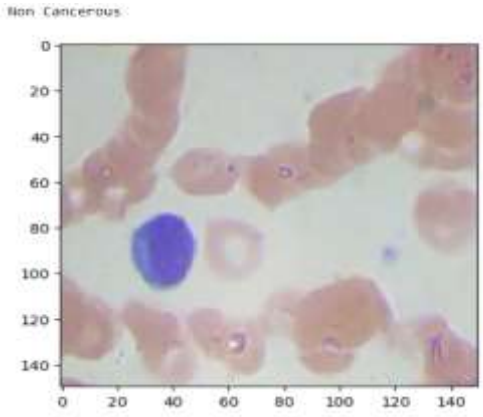

Figure 5: Non-Cancerous cell

Algorithm	Accuracy	Precision	Recall	F1-score
SVM	0.7777	0.7971	0.7016	0.8461
K-means clustering	0.7888	0.7837	0.7508	0.8592

CNN	0.9877	0.8820	0.8789	0.8804

Table 2: Comparison table for different approaches

The above table shows the comparison of different features using different algorithms and approaches.

V. CONCLUSION AND FUTURE SCOPE

The classification of is made possible by the use of computer-based algorithms for picture processing. Leukemia should be reliably classified by the system using blood smear pictures. Modern living, pollution, and other factors are making cancer a more prevalent disease. Since cancer is a rapidly spreading disease, traditional techniques of cancer detection need a lot of time-consuming transport of sample tissue (biopsy) to a facility for cancer diagnosis. Yet, early cancer treatment is likely to boost the chances of survival for cancer patients. Results obtained with the aid of deep learning models are quick and precise, assisting in early detection and enabling low-cost diagnosis.

VI. REFERENCES

[1] M. Madhukar, S. Agaian, and A. T. Chronopoulos, "Deterministic model for Acute Myelogenous Leukemia classification," in 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Seoul, Korea (South), Oct. 2012, pp. 433–438. doi: 10.1109/ICSMC.2012.6377762.

[2] M. M. Olsen, L. J. Zitella, and Oncology Nursing Society, Eds., Hematologic malignancies in adults. Pittsburgh, Pa: Oncology Nursing Society, 2013.

[3] S. Shafique and S. Tehsin, "Acute Lymphoblastic Leukemia Detection and Classification of Its Subtypes Using Pretrained Deep Convolutional Neural Networks," Technol. Cancer Res. Treat., vol. 17, p. 153303381880278, Jan. 2018, doi: 10.1177/1533033818802789.

[4] A. A. Ferrando and C. López-Otín, "Clonal evolution in leukemia," Nat. Med., vol. 23, no. 10, pp. 1135–1145, Oct. 2017, doi: 10.1038/nm.4410.

[5] H. T. Salah, I. N. Muhsen, M. E. Salama, T. Owaidah, and S.K. Hashmi, "Machine learning applications in the diagnosis of leukemia: Current trends and future directions," Int. J. Lab. Hematol., vol. 41, no. 6, pp. 717–725, Dec. 2019, doi: 10.1111/ijlh.13089.

[6] M. S. Lasya, A. Mahalakshmi, and A. Suma, "Detection and Classification of Leukemia using Convolutional Neural Network and Updation in IoT," Int. Res. J. Eng. Technol., vol. 07, no. 03, pp. 2395–2399, Mar. 2020.

[7] A. Genovese, M. S. Hosseini, V. Piuri, K. N. Plataniotis, and F. Scotti, "Histopathological Transfer Learning for Acute Lymphoblastic Leukemia Detection," in 2021 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), Hong Kong, China, Jun,2021, pp.1–6. doi: 10.1109/CIVEMSA52099.2021.9493677.

[8] A. Rehman, N. Abbas, T. Saba, S. I. ur Rahman, Z. Mehmood, and H. Kolivand, "Classification of acute lymphoblastic leukemia using deep learning," Microsc. Res. Tech., vol. 81, no. 11, pp. 1310–1317, Nov. 2018, doi: 10.1002/jemt.23139.

[9] L. Boldú, A. Merino, S. Alférez, A. Molina, A. Acevedo, and J. Rodellar, "Automatic recognition of different types of acute leukaemia in peripheral blood by image analysis," J. Clin. Pathol., vol. 72, no. 11, pp. 755–761, Nov. 2019, doi: 10.1136/jclinpath-2019-205949.

[10] C. Di Ruberto, A. Loddo, and G. Puglisi, "Blob Detection and Deep Learning for Leukemic Blood Image Analysis," Appl. Sci., vol. 10, no. 3, p. 1176, Feb. 2020, doi: 10.3390/app10031176.

[11] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "DeTrac: Transfer Learning of Class Decomposed Medical Images in Convolutional Neural Networks," IEEE Access, vol. 8, pp.

74901–74913, 2020, doi: 10.1109/ACCESS.2020.2989273.

[12] L. H. S. Vogado, R. D. M. S. Veras, A. R. Andrade, F. H. D. de Araujo, R. R. V. Silva, and K. R. T. Aires, "Diagnosing Leukemia in Blood Smear Images Using an Ensemble of Classifiers and Pre-Trained Convolutional Neural Networks,"in 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Niteroi, Oct. 2017, pp. 367–373. doi: 10.1109/SIBGRAPI.2017.55.

[13] C. Di Ruberto, A. Loddo, and G. Puglisi, "Blob Detection and Deep Learning for Leukemic Blood Image Analysis," Appl. Sci., vol. 10, no. 3, p. 1176, Feb. 2020, doi: 10.3390/app10031176.