International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 05 | May 2023 www.irjet.net p-ISSN: 2395-0072

A study on techniques to detect and classify acute lymphoblastic leukemia using deep learning.

Shaik Abdul Hameed[1], Sai Nikhil G[2], Pranav Sai Y[3], Sathya Sreekar D[4], Veerendra Reddy Y[5]

[1]*Associate Professor, Dept. of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India*
[2,3,4,5] *Student, Dept. of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India*

Abstract - Cancer *starts when cells in the body begin to grow out of control. Leukemia is a cancer that affects the body's blood-forming tissues, lymphatic system, and bone marrow. White blood cells are typically involved in Leukemia (WBC). When an individual has Leukemia, their bone marrow makes an excessive amount of dysfunctional WBC. While Acute Lymphoblastic Leukemia (ALL) predominantly affects children but is not limited to them and can also develop in adults. As a widely occurring cancer, the accurate diagnosis of ALL necessitates costly, invasive, and time-intensive diagnostic tests. The use of PBS images for diagnosing ALL plays a crucial role in the initial screening of cancer cases versus non-cancer cases. Our project aims to Automate the process of detection of Acute Lymphoblastic Leukemia (ALL) using Peripheral Blood Smear (PBS) images and provide a channel between patients and doctors for consultancy regarding the diagnosis process.*

Key Words: **Acute Lymphoblastic leukemia, peripheral Blood smear, Lymphatic system, CNN (Convolution Neural Network), CBC (Complete Blood Count), Differential Leukocyte Count (DLC)**

1. INTRODUCTION

Over 900,000 people are diagnosed with leukemia each year, sometimes known as blood cancer, however many people are unaware of the risks associated with such frequently fatal illnesses. Most of blood cancers are uncommon, life-threatening diseases that only affect small patient populations. Patients may have a sense of abandonment as a result of leukemia's rarity and find it challenging to locate the support and information they require.

If treatment for acute Leukemia is not started promptly, the patient may pass away from the condition within a few months. Any cancer must be detected early to receive prompt treatment and improve survival rates. Individuals who are ill cannot waste time because they require quick attention. We need systems that quickly use the most recent technological advancements and accurately analyze blood samples. Early detection of Acute Lymphoblastic Leukemia (ALL) symptoms in individuals can considerably improve their chances of survival.

2. RELATED WORK

In [1], Naina Sharma, Ankit Mukopadhyay, Aman Shrivastava, and Aman Garg proposed a system to help identify Leukemia through pictures of blood cells. The proposed system is supposed to be more accurate than present physicians. A system is proposed due to the scrutiny of blood or bone marrow pictures being negative for detection and very time-consuming. The authors propose using segment-stained peripheral blood smears using color-based clustering to divide a cell into a nucleus and cytoplasm. Then, SVM is used to differentiate the types of WBC and use CNN with the last layer of the CNN fed to a Random Forest algorithm to categorize the WBC.

In [2], C. T. Tchapga, M. Thomas Attia, A. T. Kouanou, T. F. Fonzin, P. K. Fogang, M. Brice Anicet, T. Daniel proposed how ML algorithms can be used with big data using Apache Spark Framework and how to classify biomedical images using the machine learning algorithms. Apache Spark overcomes the limitations of the Hadoop framework by being faster (by 100 times in memory), making data for iterations, queries, and loading, and supporting SQL Query over Map Reduce. The authors also suggest a two-step process: the first is creating an algorithm using labelled images, and the second is classification done through unlabeled images.

In [3], A. Genovese, M. S. Hosseini, V. Piuri, K. N. Plataniotis, and F. Scotti propose a system to detect ALL based on the adaptive un-sharpening of the peripheral blood smear images. The authors propose using adaptive un-sharpening using Computer Aided Diagnosis (CAD). Adaptive un-sharpening is a process used to increase the focus of an image until a certain threshold. The system used for ALL detection performs adaptive un-sharpening initially and then performs the classification. After experimenting with 260 images of WBC, the authors identified the accuracy of the detection model to be 96.84%. The only con of the model is that the model is a detection model but not a classification model.

In [4], N. Mahmood, S. Shahid, T. Bakhshi, S. Riaz, H. Ghufran, and M. Yaqoob proposed the significance of clinical data and phenotypic data, i.e., environmental conditions to detect Acute Lymphoblastic Leukemia. They used different models

like classification and regression trees, Random Forest, gradient boost and c5.0 decision tree algorithm. They applied ten-fold cross-validation to compare all algorithms and found that classification and regression trees (CART) have high accuracy, i.e., 80.6%. The CART model shows the significance of each feature to identify ALL. Some feature authors considered are age, gender, WBC count, platelets count, hemoglobin level, financial status and drinking water. Authors found out the importance of each variable with their importance in percentages, i.e., platelet 43%, hemoglobin 24%, white blood cells 45%. The main advantage here is that there is no need to process images to determine whether a person has ALL. The data collected is easy to get. However, using this data, authors cannot classify ALL stages

In [5], Sahana K Adyanthaya discussed different techniques used in text recognition and stages involved in text recognition like pre-processing, segmenting the text, extraction of features from text and post processing. Some findings are that Tesseract OCR performs well in extracting text. However, by dividing images into segments, the classifier recognizes text in each image. Due to this method, there is an improvement of about 20% in recognizing text. The author also discussed tasks various tasks in text recognition, i.e., noise removal using a gaussian filter or mean filter, segmentation using Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA), Chain Code (CC) and then classifying using distinct classifiers based on techniques such as ANN or SVM.

In [6], Vasundhara Acharya and Preetham Kumar developed a system to segment blood smear images accurately. This paper primarily focused on image segmentation, as most features extracted directly depend on segmentation. Cytoplasm can be extracted from the image using K-medoids and k-means. However, K-medoids show better performance, so K-medoids are used to extract cytoplasm. The nucleus is extracted, producing a binary image of cytoplasm and binary images of white blood cells. Separating the surroundings, RBCs from the WBCs gives precise results. However, the main disadvantages found were that border cells located at the extreme corner are not dealt with accurately, and sub-classification is not done

In [7], Dr. Leena Patil and Miss. Anagha M. Pawar proposed the best methods to detect ALL. As part of the experiment, various CNN models were evaluated, including R-CNN, Fast R-CNN, and YOLO, including regional-based convolutional neural networks. The obtained outcomes were the Model Mean Average Precision and Frames Per Second. The R-CNN achieved 62.4% and 0.5, the Fast R-CNN recorded 7-% and 0.5, while YOLO showed 63.4% and 45, respectively. Yolo offers a rapid static processing paradigm for real-time streaming analysis and image classification. Fast R-CNN, however, has a very high rate of accuracy.

In [8], A. Rehman, N. Abbas, T. Saba, Syed Ijaz ur Rahman, Z. Mehmood, and H. Kolivand suggested a deep learning method for classifying acute lymphoblastic leukemia. To obtain accurate classification results, the model is trained on images of bone marrow using CNNs and strong segmentation techniques. The results of the experiment were then compared to those of the Naive Bayesian, K-NN, and Support Vector Machine classifiers. Experimental results showed that the suggested method produced an accuracy of 97.79%. The findings allow pathologists to use the suggested technique to identify acute lymphoblastic leukemia and its subgroups.

In [9], Sarmad Shafique and Samabia Tehsin proposed utilizing Pretrained Deep CNNs to detect and the categorization leukemia. They have constructed a deep CNN for the automatic detection of ALL and categorization of its variants into four categories, L1, L2, L3, and Normal. To mitigate the issue of overfitting, a data augmentation method was employed. To assess the efficacy across different photos, they also analyzed data sets with distinct color models. They demonstrated 100% sensitivity, 98.11% specificity, and 99.50% accuracy in the diagnosis of ALL. The categorization of ALL subtypes had sensitivity, specificity, and accuracy scores of 96.74%, 99.03, and 96.06%, respectively.

In [10], L. Pan, G. Liu, F. Lin, S. Zhong, H. Xia, X. Sun, and H. Liang suggested a strategy to forecast relapse in acute lymphoblastic leukemia. On the 336 newly diagnosed ALL children's training sets, clinical variables were graded using Monte Carlo cross validation nested within 10-fold cross validation. A forward feature selection method was utilized to determine the minor list of distinguishing variables. To ensure assessment of the model's performance for new patients, an additional set of eighty-four patients, who are not part of the initial training or testing, was included for evaluation. The 14-feature Random Forest model performs well among all risk-level categories, with the standard-risk group showing the best accuracy at 0.829.

In [11], J. Wang, N. Y. Min Shen, X. Zhang, Y. Wang, MSN, Y. Liu, Z. Geng, C. Yuan, FAAN presented a Mobile App to assist carers of children with ALL. Mobile apps, which are typically practical and user-friendly, are essential m-health tools. ALL is the most common cancer-related death in children, which amounts to 26.84% of all pediatric malignancies. It affects patients under 15 the most frequently of pediatric cancer types. The frequency of ALL in young children is highest in those between the age of 2 and 4. The curability of ALL has shown significant progress in the last few decades, with a rise of over 75% due to advancements in the diagnostic process and treatment. The cancer diagnosis of a kid still has an impact on the parents and their family. The eight components in the app, them being Personal Information, Treatment Tracking, Family Care, Economic and Social Assistance, Information Center, Self-assessment, Questionnaires, Interactive Platforms, and Reminders, were

created to assist caretakers. Medical professionals and parents have run pilot tests.

In [12] Dr. J. Sasi Kiran, N. V. Kumar, N. S. Prabha, M. Kavya suggested architecture for General CRS. The system begins by performing preprocessing steps such as noise removal, thresholding, and skeletonization, which are primarily used to separate the foreground (ink) from the background paper. After preprocessing, word, character, and line segmentation utilizing techniques such as projection analysis, white space-and-pitch, and CCL follow. The architecture contains normalization, which converts the randomly sized image to the standard sized image, after segmentation. Feature Extraction, Classification (characteristics can include Bayesian classifier, KNN-classifier, RB-function, SVM), and Preprocessing are all included in the architecture's final portion (Involves grouping of symbols)

In [13], Chaitali Raje and Jyoti Rangole, using image processing, devised a way to identify leukemia in microscopic images. The technique entails cell identification, picture acquisition, preprocessing, image segmentation, and feature extraction. The first technique, nucleus segmentation with Labview, entails converting the color image to a grayscale image, improving the grayscale image with the histogram equalization approach, calculating statistical parameters, and then classifying the cell as a blast or ordinary cell. Another technique, nucleus-segmentation with Matlab, entails converting the color image to grayscale, conducting 'L' and 'H', utilizing otsu's Thresholding, and finally classifying the cell.

In [14], Atin Mathur, Ardhendu S. Tripathi, and Manohar Kuse use Leishman-stained blood stain images to classify white blood cells, which are subsequently used to discover, diagnose, and monitor hematological and non-hematological illnesses. Images are first stain adjusted wrt a target image that is chosen because there is a stark visual contrast between the WBCs. The background cells are crucial in medical image processing and extracting the White Blood Cell nucleus and cytoplasm individually, as seen in the RGB image followed by WBC segmentation. The segment that has the biggest impact on how well a classifier performs is feature extraction. The size, compactness, NCR, and properties including ANR, NOL, MCP, and roughness are all factors taken into consideration when classifying cells. An algorithm for classification is supervised.

In [15], Cartic Ramakrishnan, Abhishek Patnia, Eduard Hovy, and Gully APC Burns have developed a method to draw out text from a pdf file by breaking it up into word blocks using the GPL version of JPedal, and then methodically combining word blocks to create "chunk-blocks" by following specified rules. Section heads and subheadings must be identified as different segments from the body of their associated sections, and segments should be rectangular to assist sequence-preserving text extraction. The LA-PDFText iterates over the classified blocks in the last step of the technique, stitching together multiple contiguous portions as well as sections, sub-sections, and headings.

3. CONCLUSIONS

From the above literature survey, we can say that most of the papers deal with preprocessing, i.e., ways to extract features from images by different techniques un-sharpening algorithms or converting images into binary images or PCA for segmentation, thus says that preprocessing images is the first step to get good results. Then some papers worked with CNN models to classify the images, and most models are pre-trained with modifications helping in detecting and classifying Acute Lymphoblastic Leukemia. Some papers show models that reach over 98% accuracy and more, whereas some papers discussed text recognition techniques like OCR, Layout-Aware PDF Text Extraction systems with reasonable accuracies. Each paper has its limitations where detection fails with a noisy image or cannot detect border cells; however, the classification can become more accurate by overcoming them.

REFERENCES

[1] Naina Sharma, Ankit Mukopadhyay Aman Shrivastava Aman Garg (2021), A Brief Survey on Leukemia Detection Systems, international Research Journal of Engineering Technology (IRJET).

[2] Tchapga, C. T., Thomas Attia, M. A., Kouanou, A. T., Fonzin, T. F., Fogang, P. K., Brice, M. A., & Daniel, T. (2021). Biomedical image classification in a big data architecture using machine learning algorithms. Journal of Healthcare Engineering, 2021.

[3] Angelo Genovese, Mahdi Hosseini, S., Vincenzo Piuri, Konstantinos, N. P., & Fabio, S. (2021, June). Acute Lymphoblastic Leukemia detection based on adaptive un-sharpening and Deep Learning. In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (pp. 1205-1209). IEEE

[4] Nasir, M., Saman, S., Taimur, B., Sehar, R., Hafiz, G., & Muhammad, Y. (2020). Identification of significant risks in pediatric acute lymphoblastic leukemia (ALL) through machine learning (ML) approach. Medical & Biological Engineering & Computing, 58(11), 2631-264

[5] Adyanthaya, S.K. (2020). Text Recognition from Images: A Study.

[6] Vasundhara Acharya, & Preetham Kumar (2019). Detection of acute lymphoblastic leukemia using image segmentation and data mining algorithms. Medical & biological engineering & computing, 57(8), 1783-1811

[7] Dr. Leena Patil, Miss. Anagha M. Pawar.A Literature Survey on Classification of Images using State of the Art Machine Learning Techniques

[8] Ajmad, R., Naveed, A., Tanzila, S., Syed, I. U. R., Zahid, M., & Hoshang, K. (2018). Classification of acute lymphoblastic leukemia using deep learning. Microscopy Research and Technique, 81(11), 1310-1317.

[9] Sarmad Shafique, & Samabia Tehsin, (2018). Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks. Technology in cancer research & treatment, 17, 1533033818802789.

[10] Liyan, P., Guangjian, L., Fangqin, L., Shuling, Z., Huimin, X., Xin, S., & Huiying, L. (2017). Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia. Scientific reports, 7(1), 1-9.

[11] Jinting, W., Nengliang, Y. S. M., Xiaoyan, Z., Yuanyuan, W., Yanyan, L., MSN & Changrong, Y. (2016). Supporting caregivers of children with acute lymphoblastic leukemia via a smartphone app: a pilot study of usability and effectiveness. CIN: Computers, Informatics, Nursing, 34(11), 520-527.

[12] Kiran, J. S., Kumar, N. V., Sashi, N. P., & Kavya, M. (2015). A literature survey on digital image processing techniques in character recognition of Indian languages. International Journal of Computer Science and Information Technologies, 6(3), 2065-2069.

[13] Chaitali Raje, & Jyothi Rangole (2014, April). Detection of Leukemia in microscopic images using image processing. In 2014 International Conference on Communication and Signal Processing (pp. 255-259). IEEE.

[14] Atin Mathur, Ardhendu Tripathi, S., & Manohar Kuse (2013). Scalable system for classification of white blood cells from Leishman-stained blood stain images. Journal of pathology informatics, 4(2), 15.

[15] Cartic Ramakrishnan, Abhishek Patnia, Eduard Hovy, & Gully APC Burns (2012). Layout-aware text extraction from full-text PDF of scientific articles. Source code for biology and medicine, 7(1), 1-10.