Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN: 2581 - 611X)
VOLUME:8, ISSUE:1, FEBRUARY 2025, PAPER NUMBER:70

Comprehensive Study on Advanced Predictive Modeling fo Leukemia Blood Cancer: Integrating Deep Learning and Genomic Data

Saveetha R
Asst.Prof.Level II
Department of Computer Technology
Bannari Amman Institute of Technology
Sathyamangalam,India
saveethar@bitsathy.ac.in

Brindha R
Department of Information Technology, Bannari Amman Institute of Technology
Sathyamangalam,India
brindha.it21@bitsathy.ac.in

Jasmine Hepziba S
Department of Information Technology
Bannari Amman Institute of Technology
Sathyamangalam,India
jasminehepziba.it21@bitsathy.ac.in

Nehashree B
Department of Computer Science and Engineering
Bannari Amman Institute of Technology
Sathyamangalam,India
nehashree.cs21@bitsathy.ac.in

Dhanussh Ram K N
Department of Computer Technology
Bannari Amman Institute of Technology
Sathyamangalam,India
dhanusshram.ct21@bitsathy.ac.in

Abstract— Leukemia is a hematological disease caused by abnormal white platelets and also poses a threat to overall health. Early detection is critical to improve survival, especially for children who are more susceptible to the disease. This project combines deep learning and genomic data to conduct a comprehensive study on the process of verifying leukemia evidence. We present a deep study that includes genomic and visual data to calculate leukemia outcomes. We aim to find the best way to detect leukemia at an early stage by working on different techniques, including artificial intelligence and product analysis. Our findings exemplify the advantages and implications of many forward-thinking models and provide important information for future evaluation of this important area. This review increases accuracy and power and demonstrates the potential of genomic alignment and deep learning for the development of leukemia.

Keywords: Leukemia Identification, Profound Learning, Genomic Information, Prescient Demonstrating, Grouping, Logical artificial intelligence.

I. INTRODUCTION

Leukemia is a dangerous disease that causes abnormal white blood cells in the bones and blood. A deficiency of these white blood platelets (WBCs), called shock or shock, can lead to low levels of other platelets. The exact cause of leukemia is still unclear and can be attributed to many factors, including pollution, radiation, family genetics, and more. In 2019, approximately 61,780 cases of leukemia were reported in the United States, while approximately 9,900 new cases of leukemia were reported annually in the United States. According to the World Cancer Observatory, 42,055 cases of leukemia were reported in India in 2018 [1]. The following are normal blood cells: red blood cells are also called red platelets (RBCs); white blood cells are commonly called white platelets (WBCs) and thrombocytes (thrombocytes). Leukemia is a malignant growth of white blood cells, called myeloid or lymphoid depending on the initial cell lineage, and is severe or persistent depending on the progression of the disease.

Therefore, there are four main types of leukemia: active myeloid leukemia (AML), persistent myeloid leukemia (CML), active lymphocytic leukemia (ALL), and persistent lymphocytic leukemia (CLL). Doctors and hematologists always face problems in the early and accurate diagnosis of leukemia. Lymphatic swelling, dry skin, fever, and weight loss are common symptoms of leukemia, but they are also associated with other diseases [2]. Leukemia is initially difficult to analyze because the concept of the disease is mild. The best way to diagnose diabetes is by microscopic examination of PBS, but the best way to diagnose leukemia includes only good and destroyed bone marrow.

Over the past two decades, different studies have used techniques proposed for imaging in machine learning (ML) and PC-supported research environments, hoping to overcome the obstacles in diagnosing secondary diabetes and identifying its subpopulations. Studies have analyzed images of blood smears to detect, distinguish, and count cells from different types of leukemia. Machine learning is an important part of cognitive skills, including computation and number relationships, and is rapidly gaining recognition for diagnostic purposes. ML allows computers to adapt without knowledge and to benefit from experience. The benefits of using these techniques in processing medical records are incredible and have been well received in disease detection. Research shows that machine learning is highly effective in helping make complex medical decisions by first extracting and then analyzing features from these images.

As the number of measurement tools increases and higher quality data is produced, there is an urgent need to develop additional data mining strategies. It is not possible to analyze big data or find patterns in data with traditional methods. In the literature, studies on leukemia detection and automatic detection-based classification generally use some type of filtering, then segmentation, feature extraction and classification using various machine learning (ML) methods such as SVM and artificial neural networks. The design methods used are divided into two as supervised learning or individualized learning strategies.

Methods such as liner iteration, SVM, ANN etc. are used to find a place in a separate category for groups when segmentation strategies using k-means, fuzzy C-means (FCM) etc. are used in the care group [3]. There are some tests for computer readability and leukemia diagnosis using deep learning (DL) in the literature. There are many classification methods available in the literature, including initialization, local techniques, clustering, flow classification, dynamic images, ANN-based eight operators, etc. A study by the authors on segmentation methods used to separate small smears in leukemia detection and other tasks showed that most of the related tasks (about 41%) used image segmentation of clusters. However, all these classification methods are difficult to implement due to their heavy requirements. The strategy focuses on efforts to convert leukemia cells into healthy cells without using separate techniques.

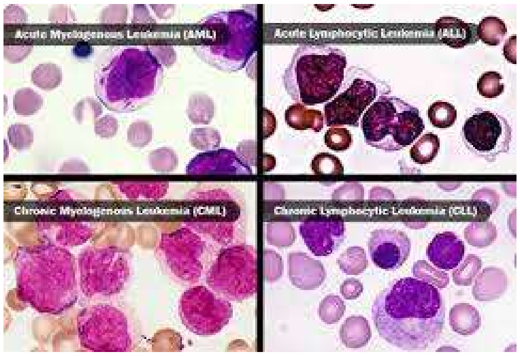

Fig.1. Types of Leukemia

II. RELATED WORKS

Experts have done many reviews in the past in this area. This article provides an overview of machine learning algorithms for leukemia detection. Due to its importance in the documentation process and legal assessment, the survey is an essential tool for professionals with an accuracy rate of 100% for Selected Trees, 97% for KNN, 91% for Gullible Bayes, and 75% for SVM. The dataset is small. Wogado et al. In the presentation, achieving 100% accuracy is an infinite challenge involving multiple perceptrons, [4] support vector machines, and indirect trees. Considering the later results of the experiment, it is tempting to think that using this technique, it is not necessary to separate the mysterious cells to find them, because it reduces the image processing time of the blood smear.

This paper proposes to achieve accurate and early detection of diseases through the integration of AI, the use of effective decision-making processes and model development, monitoring the classification accuracy of 92% compared to other classification models. The research presented by Alharbi F and Vakanski K uses AI and the best guidelines to characterize leukemia subtypes with the aim of achieving greater accuracy and addressing the use of Nausea RNA-Seq technology. Similarly, Rupapala et al. This research, which examines microcluster data clearly, identifies different mechanisms to deal with leukemia. The proposed method was also tested using T-test to assess its significance. LVTree can beat ADASYN or Chi2 by 100%. Das, Meher et al. The paper shows how iterative, convolutional brain organization and functional learning can further improve leukemia treatment.

These promising strategies show extraordinary promise in terms of accuracy and spread of leukemia in various patient populations. Summary of correlation analysis of computational localization of leukemia using DL and ML. K. Jha and H. S. Dutta applied deep learning to ALL markers in the ALL_IDB2 dataset. They performed image classification with semi-model based on mutual information (MI) and used CNN for joint classification and achieved 98.7% accuracy. L.H.S. concentrates Wogado et al. The accuracy of identifying leukemia reached 99.2% with the help of brain tissue network (CNN) calculated and analyzed by SVM. Amjad Rehman et al. Apply ALL different evidence for hidden information using HSV based initialization and pre-training network AlexNet team and guarantee 97.78% accuracy.

Using only 30 images per group, this study attempted to separate lymphoblasts from non-lymphoblasts with the highest accuracy of 99.13% using advanced deep learning methods. This study uses deep learning to identify B lymphocytes and B lymphocytes with an accuracy of 96.17%. This study is dedicated to identifying whole cells compared to cells and using highly reliable tools to determine how to do this with an accuracy of 84%. This analysis uses a population of CNN values developed for the group of leukemia cells compared to normal cells for both data, achieving the highest accuracy of 96.11%. Study using SE-ResNeXt50 CNN with the highest accuracy of 89.88% for group B lymphocytes and B lymphocytes.

The features of lymphoblasts (pre-B and pre-T) were obtained using CNN and AI based techniques and 81.74% accuracy was achieved with CNN classifier. This study attempted to compare AML killer cells with platelets using ResNext CNN design and achieved the highest accuracy of 94% in isolating myeloblasts. Rousey claims 99.67% detection accuracy for ALL and AML based on k-means clustering followed by SVM for extraction and classification. This study attempted to detect lymph nodes using data generated by Quicker R-CNN classifier and achieved 95% local speedup. This study included DL to combine WBCs and the highest accuracy in this category is 96.1% [9]. This analysis attempted to characterize WBC using BCCD and LISC data using the R-CNN system and achieved an accuracy of 97.52%.

III. PROPOSED WORK

3.1 Image Acquisition:

Any nearby hospital is used to obtain high-quality blood images.

3.2 Image Pre-processing:

The image may have some effects. There may be blind spots in the image that are important for our review. Use a neutral filter to remove noise in the image. Use a Wiener filter to soften the image. Maintenance image should be done. To improve the analysis, during image cleaning, everything except white blood cells in the image and white blood cells at the edges of the image should be removed. The robustness of image maintenance should be estimated [5]. First, the area and curvature of each white blood cell should be estimated, then no one else can find the benefit of the cleaning images. Patterns for dark images for further processing.

3.3 Image Segmentation:

Detecting white blood cells is our main goal at this stage. There are five types of white platelets. It was mentioned above, but our main analysis focused only on lymphocytes and myeloid cells [10]. Therefore, we only collect lymphocytes and myeloid cells, the other three types of white platelets (e.g. neutrophils, basophils and eosinophils) are removed from our images.

We used K-means clustering to identify white platelets. When we change the image to dark, the base of the white platelets becomes the blurriest area in the image. We used histogram adjustment and then Zacks calculation to classify lymphocytes and myeloid cells. We use various techniques to distinguish leukocyte aggregates. Finally, the nucleus and cytoplasm are extracted from the lymphocytes.

3.4 Feature Extraction:

In this stage, we try to extract some features from the image process. Here, we try to determine the importance of granulocyte and lymphocyte nuclei. Highlight extraction is the most used method to convert images into data, so we can see them using good patterns, and finally, we can distinguish between dangerous data and the data sheet is not damaged [8]. Diversity element: fills the middle color of the grayscale image. Mathematical concepts: boundaries, spans, area, rectangle, size, convexity, concavity, uniformity, extension, unusualness, solidity. Texture properties: obtain entropy, strength, uniformity, correlation. Definitions: obtain the slope, average, transit and inclination of the line.

3.5 Image Classification:

IIn the last step, subtraction is used to determine the final result. All the individual components and their activity are recorded in a single section. Before entering an image into the request, we first calculate the results. Use the last determined value to review the quality of the image values.

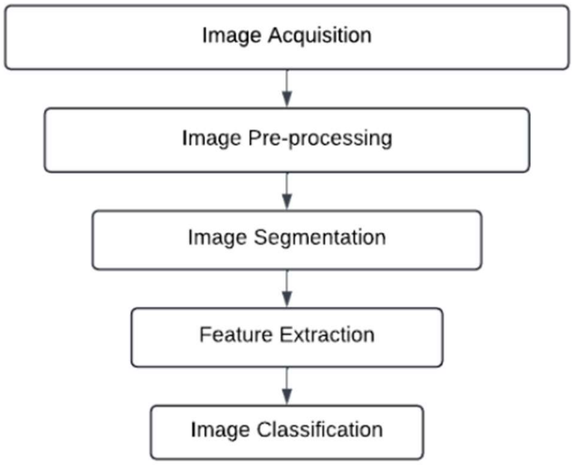

Fig.1. Proposed Method of Leukemia Detection

IV. METHODOLOGY

The concept model demonstrates the advances in data entry, the use of image processing methods, data augmentation methods, and the use of wipes for better understanding of the work. Nasal tissue allows us to use the tissue for leukemia diagnosis.

1. Input Data:

The research is based on a specific data source or field. The total number of cells in the ISBI 2019 C-NMC ALL Challenge dataset is 10,661. An image of well-segmented cells pre-marked for the dataset. This file contains the same image with dimensions 200 x 200 x 3, pre-processed to eliminate all sub-pixels outside the objects of interest. Its image (Tint, Immersion, Force) is created by modifying the custom image. There are three pieces of color data in this color space, so we know what the colors look like.

2. Data Augmentation:

Knowledge Expansion Convolutional Brain Organization (CNN) relies on knowledge expansion to avoid testing limited knowledge and improve inference capabilities. There are many ways to improve data in CNN computation, including translation, rotation, scaling, decoding, changing brightness and difference, adding Gaussian perturbation, smoothing, buffering, and truncation [6]. The use of increased knowledge helps to increase the accuracy of CNN computations, prevent overload and power, allowing them to be used for different computational tasks.

3. Feature Selection:

An important step is to make the diagnosis by extracting the important points from the blood test images. The importance of these images is separated by the use of various methods to analyze the quality and structure of the cell; usually surface studies such as integration in ensemble or Gabor channels are used [11]. Quality features include parameters such as area, roundness and eccentricity that describe the shape of the cell. By counting the cells, we can help doctors know how many cells are healthy and how many are unhealthy. Various images can be converted to high contrast using RGB to grayscale conversion. The color of the image changes; initially dark areas become lighter and brighter areas become darker. Gaussian fog and median fog are masking methods.

4. Deep Learning Techniques for Leukemia Classification:

Deep learning techniques, known for their high accuracy, are gradually being used in leukemia diagnosis. Convolutional neural networks (CNNs) are particularly powerful in image processing [7]. CNN models usually include data processing to obtain images, then several convolutional layers to perform the extraction process, pooling layers to reduce important points, and joint processing to complete the result. The model uses drivers such as ReLU (linear optimization) and Softmax to propagate data from one layer to another. For continuous messages such as calculations or texts in the form of mathematical expressions, a recurrent neural network (RNN) may be more suitable.

V. PERFORMANCE ANALYSIS OF PREVIOUS APPROACHES

In this section we review the presentation of different strategies used to predict leukemia, focusing on the main mechanisms that could be used to predict the disease in the future. Datasets and results from previous studies are shown in Table 1, which shows the various methods used to predict leukemia.

Author Name	**Key Features & Method**	**Dataset**	**Reported Accuracy**
1.Markiewicz et al.	Support vector machine (SVM) is used to extract and classify geometric shapes, statistics, and labels.	Local dataset	Reported accuracy: 90%
2.Gupta et al.	Extract and classify texture and image quality using SVM	Dataset containing 345+ images	Reported accuracy: 93%
3.Escalante et al.	PSO (Molecular Multiple Simplification) is used for the disease where the real and morphological details are eliminated	Mexican Social Security Institute dataset (633 images)	97.68% accuracy (binary classification) and 94.21% (multi-class classification)
4.Putzu et al.	Texture and color features are classified using SVM (multiple classifiers: SVM-L, SVM-Q, SVM-P, SVM-R, K-NN)	IDB1 and IDB2 datasets	Accuracy varies across classifiers
5.Yi et al.	Introduce a partitioning method that uses PSO to solve over- and under-partitioning problems	Local dataset	Not available
6.L.B. Dorini et al.	Separation of white platelets into solids/bases and subsequent characterization	Local dataset	Not available
7. Li et al. (2020)	Deep CNN for automatic classification of leukemia from microscopic images.	ALL-IDB dataset	Reported accuracy: 98.5%

VI. CONCLUSION

Deep CNN for automatic classification of leukemias with microscopy In this study, we investigated different prioritization methods and methods for treating leukemia. Furthermore, the results and analysis of this paper also shed light on the classification process in the state. From this research, we conclude that although many modern learning machines can correctly identify leukemias, deep learning is more effective in processing large files as it provides greater accuracy and efficiency in classifying complex tasks. Integrating deep learning with genomic data provides a promising approach to improve leukemia models. Result.pic image.

VII. REFERENCES

[1] American sickness Society. " Malignant boom Realities and Figures, Spring 2014." Leukemia Lymphoma Society: struggling with Blood ailment, Overhauled April 2014.

[2] Kalyanmoy Deb, and Raji Reddy. " reliable Order of -elegance disorder data making use of Transformative Calculations." * BioSystems*, vol. seventy two, pp. 111–129,

2003.

[3] Oostindjer, M., Alexander, J., Amdam, G. V., Andersen, G., Bryan, N. S., Chen, D., Corpet, D. E., De Smet, S., Dragsted, L. O., Haug, A., et al. " The Contribution of pink and Processed Meat to the development of Colorectal cancer: A factor of View" Meat technological know-how*, vol. ninety seven, no. 4, pp. 583–596, 2014.

[4] Takiar, R., Nadayil, D., and Nandakumar, A. "Projections of variety of Malignant boom cases in India (2010-2020) through disorder Gatherings." * Asian Pacific Diary of ailment Prevention*, vol. 11, no. 4, pp. 1045–1049, 2010.

[5] Dharani, T., and Hariprasath, S. "finding of Leukemia and Its sorts utilising computerized picture handling strategies." *IEEE Xplore part variety: complaints of the 2018 global conference on communication and Electronics structures (ICCES) CFP18AWO-art; ISBN: 978-1-5386-4765-3.

[6] Nee, L. H., Mashor, M. Y., and Hassan, R. "White Platelet division for extreme Leukemia Bone Marrow images." * IEEE held the 2012 worldwide convention on Biomedical Engineering (ICoBE) in Penang, Malaysia.

[7] "Proposals for the class of the extreme Leukaemias:" by means of Bennett, J. M., Catovsky, D., Daniel, M.-T., Flandrin, G., Galton, D., Gralnick, H. R., and Sultan, C. Cooperative organization of the French-American-British (FAB) kingdom English Diary of Haematology*, vol. 33, no. 4, pp. 451–458, 1976.

[8] Vardiman, J. W., Thiele, J., Arber, D. A., Brunning, R. D., Borowitz, M. J., Porwit, A., Harris, N. L., Le Playmate, M. M., Hellström-Lindberg, E., Tefferi, A., et al. " The 2008 Correction of the sector well being association (WHO) Grouping of Myeloid Neoplasms and extreme Leukemia: Justification and extensive modifications" Blood*, vol. 114, no. 5, pp. 937–951, 2009.

[9] Wu, H. S., Barba, J., and Gil, J. "Iterative Thresholding for division of Cells from Boisterous images." * Diary of Microscopy*, vol. 197, no. 3, pp. 296–304, 2000.

[10] Jagadev, P., and Virani, H. G. "popularity of Leukemia and Its sorts using picture coping with and AI." * IEEE, ISBN: international convention on tendencies in Electronics and Informatics* 978-1-5090-4257-nine.

[11] Shafique, S., and Tehsin, S. "severe Lymphoblastic Leukemia recognition and Order of Its Subtypes using Pretrained Profound Convolutional mind corporations." * Innovation in Malignant growth Exploration and remedy*, vol. 17, pp. 1-7, 2018.