



Cancer Detection From Blood Cell Imaging Using Deep Learning: A Survey

¹Thuraka Gnana Prakash, ²Chinthala Lokesh Kumar, ³Mangilipelly Sai Kumar,

⁴Mohammed Faizan, ⁵Thumu Poorna Chander,

¹Assistant Professor, ²Student, ³Student, ⁴Student, ⁵Student

¹Dept. of Computer Science and Engineering

¹VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India.

Abstract: Leukemia, a widespread and life-threatening cancer that affects people of all ages, is a global health problem. This disorder predominantly affects White Blood Cells (WBCs), altering bone marrow and blood and causing immature lymphocyte proliferation. The accurate and prompt identification of leukemia is critical for successful treatment and increased survival rates. Currently, the diagnosis is based on manual examination of blood samples from microscopic pictures, a slow, time-consuming technique with inadequate accuracy. Furthermore, the visual resemblance between leukemic and normal cells under a microscope adds to the detection problem. Recent years have witnessed the emergence of Convolutional Neural Network (CNN)-based deep learning algorithms, setting new benchmarks in image classification. However, chances to improve their efficacy, learning processes, and overall performance remain, notably in the field of leukemia diagnosis. In this detailed analysis, we dig into several methodologies previously used in the field of blood cancer detection. In addition, we highlight the field by displaying benchmark datasets often used in leukemia detection studies. We want to clarify the intricacies and complexity within this domain through comparative study, thereby driving advancement in leukemia diagnosis and treatment.

Index Terms - Leukemia Cancer, Deep Learning, Machine Learning, Convolutional Neural Network, Microscopic images.

I. INTRODUCTION

Leukemia, a complex and pervasive medical challenge, is a type of cancer that affects the white blood cells. Blood, a vital bodily fluid, consists of plasma and three distinct cell types: White Blood Cells (WBCs), Red Blood Cells (RBCs), and Platelets, each with its own specific function. WBCs play a critical role in defending the body against infections and diseases, while RBCs transport oxygen from the lungs to the body's tissues and back, and Platelets facilitate blood clotting and bleeding control. In individuals suffering from leukemia, there is an overproduction of one specific type of blood cell at the expense of others, resulting in the presence of abnormal cells. These abnormal cells possess unique characteristics compared to their healthy counterparts and exhibit impaired functionality, particularly affecting WBCs. Moreover, these abnormalities disrupt the normal functioning of other blood cell types, notably RBCs and Platelets.

Leukemia is divided into two main forms based on the type of abnormal white blood cell that gives rise to the malignancy. Lymphoid cells give rise to lymphocytic or lymphoblastic leukemia, while myeloid cells lead to myelogenous or myeloid leukemia. Additionally, leukemia is classified into acute and chronic types, reflecting the rate at which the abnormal cells multiply. Acute leukemia is characterized by rapidly growing immature blasts, requiring immediate treatment. On the other hand, chronic leukemia involves a mixture of mature, functional cells and young cells, and progresses at a slower pace.

Leukemia comprises four primary subtypes, each distinguished by distinct characteristics and clinical behaviors:

- 1) ALL (Acute Lymphoblastic Leukemia)
- 2) AML (Acute Myelogenous Leukemia)
- 3) CLL (Chronic Lymphocytic Leukemia)
- 4) CML (Chronic Myelogenous Leukemia)

The accurate identification of leukemia often relies on the microscopic examination of blood samples conducted by trained experts. These experts rely on visually observing unique cellular features to accurately classify the specific type of cancer. However, the extensive range of cell features and occasionally ambiguous microscopic images can result in crucial data being overlooked, thereby complicating the differentiation of leukemia. In this survey, we provide an overview of the research endeavors undertaken by various scientists to address the complexities associated with the detection and classification of leukemia.

II. RELATED WORK

In the domain of medical image analysis, the challenge of leukemia diagnosis from microscopic blood samples is formidable. This paper presents an innovative approach that integrates data acquisition, data augmentation, and a sophisticated Convolutional Neural Network (CNN) architecture with squeeze and excitation learning. The methodology begins by acquiring microscopic blood sample images, addressing the challenge of limited data through data augmentation. By artificially diversifying the training dataset, data augmentation enhances the model's resilience. At its core, the methodology employs a deep CNN architecture with squeeze and excitation learning, emphasizing the discriminability of features between leukemic and normal cells. Thorough experiments, encompassing both cropped cells and full-scale images, in conjunction with data augmentation, validate the effectiveness of the methodology. The model is implemented using Python and Keras on Google Colab, utilizing a 12GB NVIDIA Tesla K80 GPU for efficient training. Importantly, their proposed framework attains an impressive accuracy rate of 94%.

Data preprocessing techniques were utilized, such as ADASYN for dataset balancing and Chi2 for feature selection. The researchers developed a hybrid classification model named LVTrees, which merges logistic regression, support vector classifier, and extra tree classifier. The integration of ADASYN and Chi2 led to a significant enhancement in model performance, resulting in a remarkable 100% accuracy in predicting blood cancer types. The study's findings were rigorously validated through k-fold cross-validation and statistical tests, demonstrating the model's superiority over existing methods. This research highlights the potential of hybrid machine learning models and data preprocessing in improving the accuracy of blood cancer prediction.

This paper proposed solution involves preprocessing techniques such as color conversion, filtering, and histogram equalization to enhance the quality of the image. The K-means algorithm is utilized for data segmentation, while the Zack algorithm is employed for thresholding. The paper highlights the necessity of refining the identification, thresholding, and segmentation phases to ensure precise detection. Furthermore, the study emphasizes the significance of shape feature extraction in enhancing the counting of white blood cells and the overall accuracy of segmentation. The results demonstrate that the suggested method achieves accuracy rates ranging from 72.2% to 97.8% across various features, including K-means, histogram equalization, linear contrast stretching, and share-based features.

The leukemia detection process has been divided into four distinct stages: pre-processing, feature extraction, classification model building, and classifier evaluation. During pre-processing, blood cell images are refined, while feature extraction identifies crucial characteristics. The use of Gradient Boosting Decision Tree (GBDT) classifiers is prevalent, with an impressive 86% accuracy in detecting leukemia cells. The paper highlights the significance of the F1 score as an evaluation metric and compares GBDT with Support Vector Machine (SVM), with GBDT surpassing SVM with an accuracy of 86% compared to SVM's 83%. This research offers valuable insights for the advancement of leukemia cell detection methods.

The study delves into the intricacies of machine learning techniques for leukemia subtype classification, presenting noteworthy model accuracies. Notably, Support Vector Machines (SVM) proved to be highly effective indistinguishing between lymphoid and myeloid stem cells achieving an accuracy of 92% and successfully classifying Acute Lymphoblastic Leukemia and Acute Myeloid Leukemia. The k-Nearest

Neighbor (KNN) algorithm demonstrated an accuracy of 80%, while Neural Networks achieved 93%, showcasing their robust classification capabilities in this context. The Naïve Bayes Classifier also proved competent in handling leukemia classification problems, achieving an accuracy of 80%. Finally, the Deep Learning approach, utilizing Convolutional Neural Networks (CNN), outperformed the other techniques with an impressive accuracy of 97%, particularly in the subtyping of Acute Lymphoblastic Leukemia and normal cases. These results highlight the suitability of various machine learning algorithms for different aspects of leukemia classification, providing valuable insights into their respective strengths and limitations.

This paper delves into the latest advancements in medical image processing techniques used in the detection of cancer. It explores the utilization of automatic Otsu's thresholding for enhancing and segmenting images, as well as the integration of the Zack Algorithm, k-Nearest Neighbors (kNN) classifiers, kernel fuzzy clustering, and wavelet transforms. Additionally, it compares the effectiveness of K-means clustering and neural networks in image segmentation. Notably, the Zack Algorithm has achieved an impressive accuracy rate of 92% for nucleus segmentation and 78% for cytoplasm segmentation, specifically in identifying cancer infected cells. These innovative methods streamline the cancer detection process, reducing the need for manual expert observations and significantly improving diagnostic efficiency.

The utilization of advanced image processing techniques, such as k-means clustering, marker-controlled watershed, and HSV color-based segmentation, for the purpose of detecting and categorizing leukemia. The researchers conducted their study using a dataset comprising 220 blood smear images, enabling them to differentiate between patients with leukemia and those without. To identify the various types of leukemia (ALL, AML, CML, or CLL), they employed an SVM classifier. Their innovative approach involved converting the images from RGB to grayscale and transforming them into the Lab color space to achieve precise segmentation. The distances between clusters within the images were measured using the Euclidean distance metric. Notably, this methodology yielded an impressive accuracy rate of 87%. This holds great promise for enhancing early diagnosis and treatment methods.

This survey paper explores the utilization of Convolutional Neural Networks (CNNs) in the automated identification of white blood cancer in bone marrow images. The proposed model, an optimized Dense Convolutional Neural Network (DCNN), achieves an impressive accuracy rate of 97.2%. It surpasses conventional machine learning techniques such as SVM, Decision Trees, Random Forests, and Naive Bayes. The model's resilience is demonstrated through the use of confusion matrices, highlighting its ability to accurately classify instances as True Positive and True Negative. Utilizing the analysis of microscopic images and the application of machine learning. The team of researchers collected a dataset consisting of 256 samples from patients diagnosed with leukemia, employing the Faster-RCNN algorithm to detect objects and achieving an impressive accuracy rate of over 90%. By identifying and quantifying the components of white blood cells, this approach enables the prediction of the likelihood of leukemia development in blood samples. The Faster-RCNN algorithm plays a pivotal role in this process by extracting features from images, while a Region Proposal Network streamlines the detection of objects. This groundbreaking method shows great potential in facilitating early diagnosis and monitoring of leukemia.

The application of deep learning algorithms, namely YOLOv5, YOLOv8, Faster R-CNN, and SSD, in the identification of leukemia cancer within microscopic blood sample images. It places particular emphasis on the utilization of the squeeze and excitation learning process to augment feature representation and effectively differentiate between leukemic and normal cells. The investigation encompasses various data preprocessing techniques, with YOLOv5 being employed for object recognition, fine-tuned on a specific dataset, resulting in an impressive accuracy rate of 86%.

This comprehensive paper presents an in-depth examination of machine learning-based approaches utilized in the automated detection of blood leukemia. It emphasizes the remarkable advantages of these approaches compared to traditional classifiers, particularly in terms of accuracy and speed. The utilization of Support Vector Machines (SVM) and regression models, including linear and non-linear polynomial regression, is thoroughly discussed. Significantly, the achieved accuracy of 91% demonstrates the immense potential of these techniques. As a result, this paper serves as an invaluable asset for professionals and researchers operating within this field.

The integration of machine learning with image processing is explored in this study. It encompasses preprocessing methods like noise removal filters and thresholding, with a particular focus on the effectiveness of Linear Discriminant Analysis (LDA) in enhancing images. Additionally, the research highlights the application of Convolutional Neural Networks (CNN) and various machine learning algorithms for image classification. The performance of linear classifiers, K-NN, and Support Vector Machine (SVM) classifiers is thoroughly examined. Previous research has reported accuracy rates ranging from 88% to 93% using different approaches. This study offers valuable insights for professionals and researchers in the field.

Technique for diagnosing leukemia through blood smear images is presented in this paper. The method utilizes pre-processing, segmentation, and shape feature analysis to differentiate between cancerous and healthy cells. The authors discovered that the hue channel and Moving K-mean clustering produced the most favorable outcomes. By examining shape features such as area, perimeter, and compactness, the approach allows for the early detection of cancer cells. When tested on a leukemia dataset, the method demonstrated an impressive accuracy rate of 88%.

An innovative image processing method for distinguishing red blood cells from young white blood cells is introduced in this paper. The approach utilizes techniques such as histogram leveling, contrast stretching, and morphological operations. Pre-processing steps, including color transformation, filtering, and histogram leveling, are included. By employing the k-means algorithm, the paper achieves an impressive 88% accuracy in image segmentation. The methodology involves image conversion, feature extraction, and analysis of damaged regions through region assessment. Additionally, noise reduction and image enhancement techniques, such as contrast enhancement and histogram equalization, are employed.

This paper delves into sophisticated image processing methods, primarily emphasizing Convolutional Neural Networks (CNNs) for the purpose of image classification. The pre-processing stage encompasses the elimination of noise through median filtering and the conversion of RGB images to grayscale. The segmentation techniques employed involve grayscale, image binarization, and adaptive thresholding. A crucial aspect of this study is the application of supervised learning using training datasets, and the paper emphasizes the advantages of utilizing a substantial amount of training data for CNNs, rendering it a valuable asset in the realm of image and signal processing.

II. MAJOR TRENDS AND KEY FINDINGS

The review of 15 research papers on blood cancer detection, deep learning models, such as Squeeze and Excitation Learning, hybrid logistic vector trees, and Convolutional Neural Networks, are identified as crucial tools for improving disease detection accuracy from microscopic blood samples. Image enhancement techniques are vital for enhancing the quality and visibility of important details in these samples, ensuring precise diagnostics.

These papers primarily focus on using deep learning to detect various blood cancers, particularly leukemia, with the potential for early and accurate disease detection. However, challenges include the need for extensive, high-quality datasets for effective model training and the ongoing effort to make machine-based diagnoses more transparent and interpretable for medical professionals. The insights from this literature are invaluable for our project, which aims to enhance the accuracy and reliability of blood cancer detection by harnessing deep learning models, with a special emphasis on leveraging these innovative methodologies.

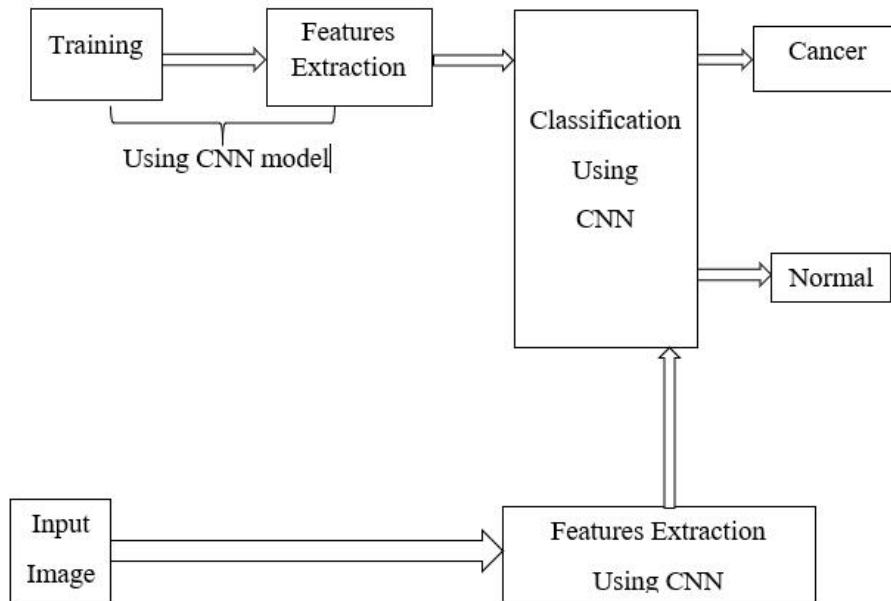
IV. PROPOSED METHODOLOGY

A. Data Collection and Preparation:

The first step in the development of AI for detecting Blood cancer involves the collection of a diverse and extensive dataset of microscopic blood cell images. It is essential to obtain data from various clinical sites to ensure that it accurately represents the overall population. Following this, the collected data undergoes preprocessing to ensure consistency and high quality. This preprocessing stage may involve tasks such as eliminating noise, standardizing images and rectifying any imperfections. Notable datasets that can be utilized for this purpose include the KAGGLE dataset and the OASIS dataset.

B. Model Development:

After preprocessing the data, a deep learning model is trained to detect blood cancer in microscopic blood images. Various deep learning architectures, such as convolutional neural networks (CNNs) and transformers, can be utilized for this purpose. The training process follows a supervised learning approach, where input



images are annotated as indicative of either cancer or normal.

Fig. 1 Proposed CNN model for classification of cancerous cells

To improve the performance and robustness of the model, a variety of techniques can be used, such as:

- **Data augmentation:** To improve the model's performance and robustness, several techniques are implemented. Data augmentation is employed to artificially expand the training dataset, reducing overfitting and enhancing generalization to unseen data. This involves applying diverse transformations, such as cropping, flipping, and rotating, to the training images.
- **Transfer learning:** Transfer learning is employed to leverage the knowledge gained from pre-trained models, thereby enhancing the performance of the deep learning model. This is particularly advantageous when working with limited training datasets.
- **Learning rate adjustment:** Optimization and adjustment of the learning rate are crucial for refining the model. Learning rate adjustment involves fine-tuning the magnitude of steps taken during the model learning process, similar to tuning an instrument for optimal performance. This meticulous adjustment contributes to optimizing the overall learning process.

C. Model Evaluation:

To evaluate the performance of the deep learning model, an assessment is conducted on a held-out test set to gauge its accuracy, robustness, and interpretability. The accuracy of the model is determined by calculating the percentage of correct predictions made on the test set. To assess the robustness of the model, its performance is analyzed using a range of adversarial examples specifically designed to deceive the model. Furthermore, the interpretability of the model is evaluated based on the quality and comprehensibility of the explanations generated through explainability techniques.

D. Clinical Validation:

This deep learning model is evaluated on a test set before deployed in a clinical context to validate its efficacy and practicality. This procedure includes gathering data from real patients and assessing the model's precision in identifying blood cancer. In addition, medical professionals such as clinicians assess the usability of the model to ensure that it is clearly understood and easy to use. By following this methodology, a sophisticated AI-powered blood cancer detection system can be developed, capable of accurately and consistently detecting

blood cancer. This system holds the promise of improving patient care by providing clinicians with valuable insights and assisting in the diagnosis and treatment procedures.

V. CONCLUSION

This paper represents a comprehensive examination of the existing literature on the diagnosis of leukemia disease, with a specific focus on the use of CNN and Deep Learning techniques to improve image classification and segmentation in the field of biomedical imaging. Our proposed approach offers an automated method for detecting cancer in blood cell images, utilizing state-of-the-art Convolutional Neural Networks (CNNs) and deep learning technology. Through a meticulous analysis of various image attributes such as texture, geometry, color, and statistical characteristics, our main objective is to enhance the accuracy and effectiveness of leukemia detection. Furthermore, we highlight the significance of benchmark datasets commonly employed in leukemia detection studies, providing a valuable resource for researchers involved in this field of study. Through this research, our ultimate goal is to advance the diagnosis and treatment of leukemia by introducing a robust and dependable system characterized by high efficacy and accuracy, reduced processing time, cost-effectiveness, and improved resilience.

VI. ACKNOWLEDGMENT

We would like to express our deep gratitude to T. Gnana Prakash, our team mentor, for his unwavering technical support and guidance throughout the project, which has been instrumental in our progress thus far.

REFERENCES

- [1] Maryam Bukhari, Sadaf Yasmin, Saima Sammad, and Ahmed A. Abd El-Latif. 2022. "A Deep Learning Framework for Leukemia Cancer Detection in Microscopic Blood Samples Using Squeeze and Excitation Learning", Hindawi, Article ID 2801227.
- [2] Vaibhav Rupapara, Furqan Rustam, Wajdi Aljedaani, Hina Fatima Shahzad, Ernesto Lee, Imran Ashraf. 2010. "Blood cancer prediction using leukemia microarray gene data and hybrid logistic vector trees model, Scientific Reports.
- [3] Christo Ananth, P. Tamilselvi, S. Agnes Joshy, and T. Ananth Kumar. 2022. "Blood Cancer Detection with Microscopic Images Using Machine Learning, ResearchGate.
- [4] Supriya Mandal, Vani Daivajna, Rajagopalan V. 2019. "Machine Learning based System for Automatic Detection of Leukemia Cancer Cell", IEEE.
- [5] Patil Babaso S, S.K. Mishra, Aparna Junnarkar. 2020. "Leukemia Diagnosis Based on Machine Learning Algorithms", IEEE International Conference for Innovation in Technology (INOCON).
- [6] M. Saritha, Prakash. B, Sukesh.K, Shrinivas B. 2016. "Detection of Blood Cancer in Microscopic Images of Human Blood Samples: A Review", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT).
- [7] Preeti Jagadev, Dr. H.G. Virani 2017. "Detection of Leukemia and its Types using Image Processing and Machine Learning", International Conference on Trends in Electronics and Informatics (ICTEI).