

A Survey on Leukemia Detection using Image Processing Techniques

Mohammed Junaid Ahmed^{1,*}, Padmalaya Nayak²

¹Mtech Student, Computer Science and Engineering, GRIET, Hyderabad, Telangana, India.

²Professor, Computer Science and Engineering, GRIET, Hyderabad, Telangana, India.

Abstract. Leukemia detection and diagnosis by inspecting the blood cell images is an intriguing and dynamic exploration region in both the Artificial Intelligence and Medical research fields. There are numerous procedures created to look at blood tests to identify leukemia illness, these strategies are the customary methods and the deep learning (DL) strategy. This survey paper presents a review on the distinctive conventional strategies and Deep Learning and Machine Learning methods towards that have been utilized in leukemia illness diagnosis dependent on platelets images and to analyze between the two methodologies in nature of appraisal, exactness, cost and speed. This article covers 11 research papers, 9 of these examinations were in customary strategies which utilized image handling and AI (ML) calculations, such as, K-closest neighbor (KNN), K-means, SVM, Naïve Bayes, and 2 investigations in cutting edge procedures which utilized Deep Learning, especially Convolutional Neural Networks (CNNs) which is the most generally utilized in the field leukemia detection since it is profoundly precise, quick, and has the smallest expense. What's more, it dissects various late works that have been presented in the field including the dataset size, the pre-owned procedures, the acquired outcomes, and so forth. At last, in view of the led study, it very well may be reasoned that the proposed framework CNN was accomplishing immense triumphs in the field whether in regards to highlights extraction or classification time, precision and also a best low cost in the identification of leukemia.

1 Introduction

Now a day's Leukemia outlines a significant medical condition. In various foreign nations, a great several people are dying as a result of cancer. Generally, on the off chance that we take a proportion of the number of individuals who experience malignant growth in the course of their life, it will get 1:2 on account of men and 1:3 on account of women. Among several illnesses, the chase for getting a handle on disease, a threatening neoplastic issue is in the bleeding edge research region for some analysts like scientists, physicists[1]. Malignancy can be characterized as a few gatherings of infections each with its own pace of development, fix, and treatment. In any case, a wide range of tumors is distinguished by the unsuppressed development of strange cells inside the body which ultimately spread to various pieces of the body and can deteriorate step by step if not analyze and treated appropriately[2].

In this paper, we studied hematological neoplasia for example leukemia, myeloma, and lymphoma. These kinds of issues can influence the blood, bone marrow, and lymphatic framework. According to a review done by Leukemia and Lymphoma society, it was approximated that in 2012 an aggregate of 148,040 have been analyzed, and 54,380 kicked the bucket of leukemia, lymphoma, and myeloma in the US. In India, the complete number of people experiencing blood

disease was assessed to be roughly 104,239 out of 2010. Besides, according to the Indian Association of blood malignant growth and partnered infections, among every one of the diseases which are hazardous and can cause demise, Leukemia (white platelet disease) comprises 33% of youth disease. The attributes of leukemia are the multiplication of white platelets that are not typical in the affected bone marrow without repressing cell development.

Both ALL and AML are classes of cancer that are very crucial to diagnose. AML stands for acute myelogenous leukemia which is a blood cancer. A higher number of B lymphocytes is typically associated with ALL than T lymphocytes. They play a vital role in protecting the body from germs and contamination and destroying any cells that have been tainted. Particularly B cells are tasked with maintaining the body's immunity and eliminating germs that may contaminate it. Each was grouped into three types: L1, L2, and L3 [3].

All of these diseases display symptoms similar to other diseases, therefore, diagnosing them is difficult. A microscope is used to examine the blood as part of a leukemia determination strategy. The purpose of this test is to look for abnormally formed white blood cells, which may indicate disease. Since long back, the investigation has been performed by experienced administrators, who generally conduct two examinations: characterization and tally of cells (which is performed

* Corresponding author: mja.hyd97@gmail.com

using a cytometer today). In most cases, morphological evaluations may well be conducted by means of a single photograph rather than a blood test. Along these lines, this examination doesn't need coordination with done because of the costly expense, precision is generally something similar for various pictures, and the framework is distant screening [3].

In the field of Deep Learning and Convolutional Neural Networks, the detection of different stages of leukemia using the microscopic images, in order to detect the disease stage, the image processing does six stages: 1. Image Acquisition, 2. Image Pre-processing, 3. Image Segmentation, 4. Feature Extraction, 5. Detection of cells and finally 6. Classification of cells. Some new works identified with WBCs and RBCs division depend on shading space, for example, HSI shading space, and RGB. The other strategy uses ordinary AI strategies and image processing procedures, and data processing techniques for example, K-Means clustering, Watershed Segmentation, Support Vector Machines and Fuzzy logic C-Means clustering etc.

In this paper furtherly discussed different methodologies that are used for Leukemia classification and detection by different researchers are presented in Section 2, Section 3 outlines the conclusion and Section 4 reveals the future scope and future research.

2 Literature Survey

In 2009, Abdul Nasir et al., utilized a thresholding strategy in deciding the proportion of platelets for malignant growth cell location. In this paper picture 596processing, strategies have been utilized to tally the number of platelets in the biomedical picture. With this tallied estimation of platelet, the proportion of platelet for leukemia is determined. The first picture is changed over to a grayscale picture for which an edge estimation of power is set to separate WBC to RBC (thresholding changes over a grayscale picture to a twofold picture). In the event that outcomes are not fulfilled the interaction is reshaped by setting another edge esteem. The outcomes obtained utilizing the thresholding procedure show that the proportion of RBC and WBC for a typical picture to the strange picture has an alternate scope of proportions. The proportion of a typical picture is 0 to 0.1, while the proportion of a strange picture is 0.2 to 2.5 for all and 0 to 14 for AML. The problem with this strategy is that setting appropriate limits would be extremely time-consuming and difficult[4]

In 2010, Subrajeet Mohapatra et al., Proposed a minimal effort and the proficient solution to utilize image processing analysis for quantitative assessment of stained blood microscopic scanned pictures for leukemia identification. A two-stage color segmentation system is utilized for isolating leukocytes or white platelets (WBC) from other blood segments. Discriminative highlights for example core shape, the surface is utilized for the last discovery of leukemia. In the current paper two novel shape highlights i.e., contour signature and Hausdorff measurement are executed for classifying a lymphocytic cell core. Backing Support Vector Machine (SVM) is utilized for characterization. An aggregate of

108 blood smear pictures was considered to include extraction and the last presentation assessment is approved with the aftereffects of a hematologist[5]. They have followed the method to extract the features which is image grabbing, preprocessing the image, Color conversion which is RGB to L^*a^*b color space to make the color-based clustering, Image segmentation for separating the WBC from the microscopic scanned images for this K-Nearestneighbor classification and K-means to select the sample region randomly, later Sub imaging and finally feature extraction.

They used the image size of 512 x 512for evaluation and used sequential process.The features extraction in this proposed model is SVM for training classifier and achieved accuracy of 95% was observed.

In 2010, Adnan Khashman et al.,The author proposes, in his paper, that morphological examination has been applied to leukemic photographs, from there on dividing them in two parts and obtaining improved views of the cytoplasm and core regions. To get images of cytoplasm and cores, we used bimodal thresholding, where two edge forces were given, and the border of the contaminated cell was followed, while unfavorable items were gathered from the surrounding areas. As a result, 98.33% of the generally adequate division proportion was obtained[6].

In 2011, R. D. Labati et al., come up with a novel idea of bringing the blood sample dataset which is specifically aligned for the examination and comparison of the algorithm performance for segmentation and classification. Every image in the dataset the classification of cell image is provided also specific set of merits to compare the performances of multiple algorithms. They made cancer images in JPG format and maintained the 24-bit color depth and resolution of the images is maintained which is 2592 x 1944 and separated into two versions 1. ALL-IDB1 and 2. ALL-IDB2 and they made the dataset available to the public as a freesource. They classified the dataset using the confusion matrix and evaluated using the parameters called True positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). 1.True Positive (TP) which is used to find out the number of elements classified as positive correctly, True Negative (TN) which is used to find out the number of classified as negative, False Positive (FP) which is used to find out the number of elements classified as positive but false and False Negative (FN) used to find out the number of elements classified as negative but false. Using these parameters, they found out the classification error which is 45.37% on first image set ALL-IB1 and they got a classification error of 50%[7].

In 2013, M. D. Joshi et al., was the first to propose a system to improve and divide the picture and a KNN classifier was used to characterize the shoot cells from typical lymphocytes. In the public picture dataset for leukemia investigation, 108 pictures can be used as part of the framework. Ninety-three percent of the time, this strategy has worked. Nonparametric factors are arranged using the k Nearest-Neighbors (KNN). It can also be used to separate white platelets from shoot cells using

this method, which is straightforward but extremely effective [8].

In 2014, Monica Madhukar et al., Proposed a technique which automatically does the detection as well as segmentation of Acute Myelogenous Leukemia using microscopic scanned images of blood smears. This proposed approach differs from the other approaches. 1. The simplicity of the developed approach. 2. Classification of total blood smear images as opposed to sub images and 3. Algorithms to detect and segment nucleus cells. A minuscule blood picture of size 184×138 is considered for assessment. The predominance of the plan is exhibited with the assistance of an analysis. Highlight extraction with and without the LBP administrator introduced extremely fascinating outcomes. The framework built without utilizing the LBP administrator gave productivity of 93.5%. Each of the three approval techniques was consolidated into our framework. In any case, the exhibition of HD (Hausdorff dimension), specifically, subsequent to utilizing LBP (Local binary pattern) expanded the classifier execution by 4% [8]. By utilizing LBP, the edges of the cores of the myeloblasts were removed in an exceptionally articulated way. This successful edge location upgraded the HD, as the case mean AML was substantially more than the crate means non-AML pictures. To see the effect of HD in the list of capabilities, the classifier was run with HD as the solitary component. This was done twice, once with applying LBP administrator and once without LBP administrator. Every one of the boundaries for assessment was removed for the two sets. It was seen that, when LBP was not utilized, the HD execution was uniquely around 70%, though when LBP was utilized, the rate heightened to 93%. This obviously shows the impact of the LBP administrator on the framework. To see the viability of the created calculations, a preliminary was run contrasting the framework's presentation on sub images and entire pictures. The acquired outcomes further substantiated the effect of the LBP administrator.

In 2014, JakkrichLaosai et al., Developed an algorithm to analyze color pictures of stained fringe blood spreads and recognize the types of White Blood Cells (WBC) from them [9]. In the interaction, features are extracted, segmentation is performed, and classification is performed. Research for this project focuses on the characterization of Foil of Bretagne (lymphoid) and Almeida Lloyd (myeloid). In order to distinguish anomalies and guarantee the analysis, doctors must examine and distinguish abnormalities. The trial results showed that the presentation of ID leukemia utilizing our process of picture preparation could order 100 example pictures of lymphoid undifferentiated organisms as well as myeloid undeveloped cells. Using clustering techniques in K-Means, this technique has been tested. Using obvious signs of shape and surface, highlights emerge from the divided cytoplasm and core. Different classifiers have been investigated on various blends of capabilities. The outcomes introduced here depend on preliminaries led with typical cells. The best utilizing SVM was of 92% [10].

In 2014, Chaitali Raj et al., Proposed a model called nucleus segmentation which is implemented using LabVIEW and MATLAB. By using the histogram equalization method in LabVIEW environment, they segmented the nucleus from the segmented image by using statistical parameters such as MEAN and SD. By calculating the perimeter of each cell, 128 microscopic images were examined and features were extracted from them. In the test, the cells are categorized into blasts or normal cells based on cell feature values. The nucleus segmentation of images is achieved in the MATLAB environment by converting all the scans from RGB to grayscale so the nucleus part of the cell will appear the darkest part of the image. The examination shows how difference extending, for example, straight differentiation extending and histogram balance can be consolidated along with the utilization of picture math tasks to get rid of the individual pieces of the white platelet while keeping the core of the white platelet. This example also shows how filtering can make the nucleus easier to detect by a threshold by making it darker [11].

In 2015, Romel Bhattacharjee et al., come up with an idea of an automated blood smear image analysis system which is very grate tool for identifying the detection of Acute Lymphoblastic Leukemia. The fundamental way of diagnosing the cancer is divided into 6 following steps which is 1. Input the image which means sending the blood scan report as an input 2. Image grabbing and cropping 3. Pre-processing 4. Segmentation, 5. Feature extraction and finally 6. Classification

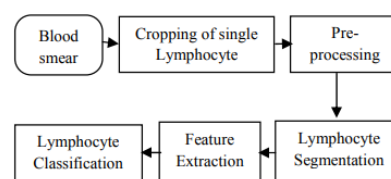


Fig.1. Automated blood smear image analysis

In this experiment the researcher done the comparison of segmentation accuracy of the proposed algorithms. Which is

1. Fuzzy K-Means Clustering in $L^*A^*B^*$ color space in which features are not discussed and the segmentation accuracy of 92% achieved.
2. Fuzzy C means clustering technique is used to and extracted the features like Area, Perimeter, Compactness, Solidity, Homogeneity, and Entropy and achieved an segmentation accuracy of 93%.
3. Watershed transform and HSV color model and achieved an segmentation accuracy of 94.5%.
4. Arithmetical operations and Otsu's thresholding segmentation technique is used to find out the features like Area, Perimeter, Circularity and achieved a segmentation accuracy of 93%.
5. Finally Morphological operator's segmentation technique is used and extracted features like Area, Perimeter, Circularity, Axis Length, Form Factor and achieved 96.67%

Apart from these researchers Romel Bhattacharjee et al., experimented using Cross validation, KNN, SVM and AVM classifiers are trained and tested over the dataset. Despite the fact that 75% of the information is taken to prepare the classifiers independently for the exhibition checking of the classifiers by cross-approval, every one of the classifiers is likewise tried to have the most unnumber of info information for preparing and to create the ideal yield. KNN is taken care of with 6 example information, SVM with 10, ANN with 20 and K-implies Clustering is finished with 10 example information, where the classifiers produce an equivalently decent yield. kNN with 6 example information delivers a particularity of 95.23% and furthermore has the least computational intricacy. Along these lines, a Graphical User Interface (GUI) has been readied utilizing MATLAB–GUI with KNN classifier for simple, quick, and powerful ID of ordinary and impact lymphocytic cells[12].

In 2017, Shaikh Mohammed Bilal N et al., proposed the computer aided leukemia detection using the digital image processing techniques[13]. In this proposed system is a classification of a cell model utilizing different cell morphological highlights. For this reason, it is important to do a pixel-level pre-processing and the segmentation of the nucleus and cytoplasm. This methodology thinks about segmentation as a vital undertaking for the separation of platelets. Segmentation ends up being precise and has viable outcomes on such clinical analytic machine-based methodologies. Henceforth as a choice, we utilize a cell division methodology that thinks about shadings, surfaces, and different highlights even in overpopulated cell smears of blood at the pixel level picture assessment accordingly extricating the cytoplasm and core in something very similar. The cytoplasm foundation cells and core separation make this work upgrade its outcomes interestingly with different works. in order to boost the proficiency and precision (over 90%) along these lines diminishing time intricacy of each assignment in the calculation for example utilized a basic quick calculation, for example, DFT or FFT alongside methodology in MATLAB for picture handling, ANN and SVM for classification purpose WEKA as opposed to utilizing all algorithms. Ordered finding prompting sublevel detection of cancer ALL by implementing image processing techniques utilizing MATLAB followed by the arrangement and subtype grouping utilizing SVM alongside ANN.

In 2019 Supriya Mandal et al., DTS Mirafr Software Technologies Pvt Ltd Bangalore, India introduced a methodology for malignant growth cell detection by extricating significant highlights from the platelet pictures and learning various classifiers. they observed that Gradient-Boosting Decision Tree classification calculations give preferred outcomes over Support Vector Machine[14]. they also additionally inferred not many significant highlights like the presence of nearby nuclei and the proportion of inconsistency looks like a nucleus, which essentially affects malignant growth cell location. Our procedures can be utilized in a restricted figuring climate without a Graphics Processing Unit.

This recommended method to identify ALL cancer cell is classified into four segments such as 1. Pre-processing, 2. feature extraction, 3. classification model building and 4. evaluation of classifier. The researcher calculated the sensitivity, F1 score, AUC, specificity from the training dataset to evaluate the performance of all three classifiers[15]. Also, they separately estimated the result of characteristics like neighboring nuclei, area variation of the shape of nuclei of the convex framework toward the LightGBM model. Also, they have attuned the classification applying the probability assessment of determining or identifying cancer cells employing the LightGBM model. In this experiment probability threshold value of 0.52 gives the best performance of the model extracted highlights from pictures and constructed a classifier utilizing SVM and Gradient Boosting Decision Tree (GBDT) to run on different platforms. This picture classifier utilizing GBDT likewise gives cutting-edge execution for malignancy cell discovery. They investigated numerous highlights which is a unification of mathematical, surface, and analytical highlights of the cell pictures. this study shows that GDBT based classifier outflanks the Support Vector Machine model. Additionally, this model study explains that the form of nearby cell cores and region contrast essentially affects the discovery of malignancy cells. They have accomplished an 85.6% of F1 score on validation learning.

Table 1. Algorithm & Accuracy.

Year	Author	Algorithm/ System	Accuracy
2009	Abdul Nasir et al	Thresholding strategy to convert the image into grayscale to find out the edges of the cell	-
2010	Subrajeet Mohapatra et al.,	Hausdorff measurement, K-Nearest neighbor classification and K-means to select sample, SVM for training	95%
2010	Adnan Khashman et al.,		98.33%
2011	R. D. Labati et al.,	Classified the dataset using True positive (TP), True Negative (TN), False Positive	Error – 45.37% on ALL-IB1 Error – 50% on ALL-IB2

		(FP) and False Negative (FN).	
2013	M. D. Joshi et al.,	KNN Classifier	93%
2014	Monica Madhukar et al.,	Image Processing Hausdorff dimension (HD), Local binary pattern (LBP)	HD – 70%. LBP- 93%
2014	JakkrihLaosai et al.,	KNN and SVM	92%
2014	Chaitali Raj et al.,	LabVIEW MATLAB	-
2015	Romel Bhattacharjee et al.,	FUZZY logic	92% 93% 94.5% 93% 96.7%
2017	Shaikh Mohammed Bilal N et al.,	SVM and ANN	90%
2019	Supriya Mandal et al.,	SVM Gradient Boosting Decision Tree (GBDT) LightGBM	85.6%

3 Conclusion

Research in this work examines various methods of leukemia cancer detection based on deep learning, image processing, and machine learning. By combining deep learning and image processing, it is possible to accelerate the cycle of leukemia analysis and enhance the chances of removing leukemia patients from further harm. A technique of image processing can overcome the subjectivity of a microscopist's operating in situations where the eyewitness's involvement and health level affect the results. In future studies, analysts might consider arranging ALL cells by subtype, for instance, L1, L2, and L3, rather than only separating dangerous and non-carcinogenic cells. In addition to gaining numerous highlights and finding other proper highlights, there is the opportunity to detect the kind of acute leukemia. To get proficient and delegated details, you can apply component determination and dimension reduction methods. Further investigation may also be undertaken in the case of another leukemia type like acute myeloid leukemia or AML, and those subtypes, since the majority of these investigations actually focus on ALL exclusively. Having attempted to characterize ALL versus AML by this point, we could find out which and what ALL and AML types there are.

4 Future Work

In future research we are planning to implement the CNN model for classification and detection of ALL and AML Leukemia cells and their subtypes to determine the type of Leukemia

References

1. S. Mishra, B. Majhi, P. K. Sa, 3rd Int. Conf. Recent Adv. Inf. Technol., 460-466, (2016)
2. A. Belhekar, K. Gagare, R. Bedse, Y. Bhelkar, K. Rajeswari, M. Karthikeyan, Proceeding - IEEE Int. Conf. Comput. Commun. Autom. ICCCA, (2019).
3. R. G. Bagasjvara, I. Candradewi, S. Hartati, A. Harjoko, Proc. - 2nd Int. Conf. Sci. Technol.-Comput. ICST, (2016)
4. Abdul Nasir, Aimi Salihah & Mustafa, Nazahah, Mohd Nasir, Nashrul, Int. Conf. Comput. Intell. Man-machine syst. Cybernetic - Proc, (2010)
5. S. Mohapatra and D. Patra, Int. Conf. Syst. Med. Biol. ICSMB, 49-54, (2010)
6. Adnan Khashman, Esam Al-Zgoul, In Proceedings of the 4th WSEAS international conference on Computer engineering and applications, (2010)
7. R. D. Labati, V. Piuri, F. Scotti, 18th Int. Conf. Image Process. ICIP., (2011)
8. Joshi, M. D., A. H. Karode, Int. J. Emerg. Technol. Learn., (2013)
9. S. Agaian, M. Madhukar, A. T. Chronopoulos, IEEE Syst J, **8**, 995-1004, (2014)
10. J. Laosai, K. Chamnongthai, Int. Electr. Eng. Congr. iEECON, (2014)
11. C. Raje, J. Rangole, Proc. IEEE Int. Conf. Commun. Signal Process. ICCSP, 255-259, (2014)
12. R. Bhattacharjee, L. M. Saini, Communication and Information Technology Conference, (2015)
13. M. B. N. Shaikh, S. Deshpande, IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol. RTEICT 2017 - Proc., (2017)
14. S. Mandal, V. Daivajna, R. V, INDICON 2019 - 16th IEEE India Counc. Int. Conf., 1-4, (2019)
15. R. G. Bagasjvara, I. Candradewi, S. Hartati, A. Harjoko, Proc. - 2nd Int. Conf. Sci. Technol.-Comput. ICST, (2016)