Leukemia Disease Detection and Classification Using Machine Learning Approaches: A Review

Astha Ratley
M.Tech (CTA) Scholar
Department of CS & Engg.
VEC, Lakhanpur, C.G., India
aastharatley02@gmail.com

Mrs. Jasmine Minj
Assistant Professor
Department of CS & Engg.
VEC, Lakhanpur, C.G., India
jasmine5feb@gmail.com

Mrs. Pooja Patre
Assistant Professor
Department of CS & Engg.
VEC, Lakhanpur, C.G., India
poojapatre.89@gmail.com

***Abstract* - Leukemia is a type of blood cancer which occurs due to abnormal increase in WBCs (white blood cells) in bone marrow of human body. Leukemia can be classified as acute leukemia and chronic leukemia, in which acute leukemia grows very fast whereas chronic leukemia grows slowly. Further both the types have two sub categories lymphocytic and myeloid. In this paper, we are going to analyze different image processing and machine learning techniques used for classification of leukemia detection and try to focus on merits and limitations of different similar researches to summarize a result which will be helpful for other researchers.**

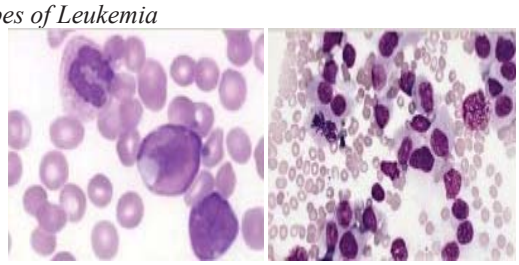
Keywords: Leukemia Disease, Deep Learning Approaches, Segmentation, Classification.

I. INTRODUCTION

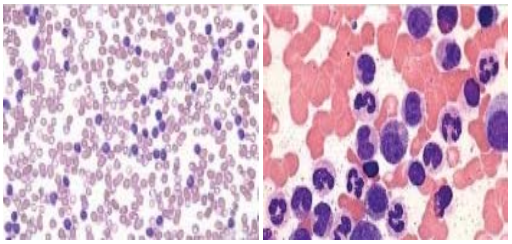
Image processing techniques are most widely used for detection of various medical diseases leukemia is one of the most interesting areas for researchers because it belongs to the category of blood cancer which can affect the persons of all ages starting from children to the old age people. The use of image processing with Computer-Based algorithm makes possible the classification of very easy. Detection of leukemia disease is when done by some expertise of the field then there may be some error present due to lack of knowledge or incorrect information present in the microscopic image. So, Computer-Based algorithm can be very beneficial in such a field to increase the detection accuracy [1]. There are two types of WBCs present in human blood and when the affected cells are monocytes type and granulocytes type, then the leukemia will be classified as myelogenous

(AML), and if they are lymphocytes, then the leukemia is classified as lymphoblastic (ALL) [2]. The blood cancer can be broadly classified into four categories i.e. ALL, AML, CLL and CML. All these type depends on the effect on WBCs (white blood cells) present in the human blood.

A. Types of Leukemia



(a) ALL (b) AML



(c) CLL (d) CML

Fig.1. Types of leukemia disease

Four types of leukemia (Figure 1) that a person can develop during their lifetime, regardless of age. These types are:

1) Acute Myeloid Leukemia (AML): This type of cancer occurs due to under development or some bad effect on bone marrow. When the WBCs rise rapidly, the working of bone marrow effected badly and causes cancer. In most of the cases early detection of such cancer may lead to successful treatment. During this type of cancer, a person may feel problem in breathing, bleeding etc. [1, 3].

2) Acute Lymphocytic Leukemia (ALL): This type of cancer generally founds in kids and the major reason of such cancer is the rapid growth in white blood cells. A number of the factors that are joined to the present sort of cancer include: radiation exposure, viral infections and transmitted diseases like Down's syndrome. Kids have a better rate of remission than older adults who are diagnosed with this kind of leukemia. ALL is further classified as L1, L2 and L3.

3) Chronic Myeloid Leukemia (CML): This form of cancer happens once the myeloid cells endure a genetic modification. When the genetic modification occurs in cells, the tradition cells could not fight properly with the infections. This type of leukemia is common among adults and is a slow growing type of cancer. The CML cancer has further 3 stages which are known as chronic phase, accelerated phase and the blast phase. In first stage, cancer is in developing stage and develops very slowly so curable at this stage. In the second phase, it becomes more effective and starts harming the blood cells and at last stage blast of cells occurs [2, 3].

4) Chronic Lymphocytic Leukemia (CLL): This form of leukemia affects the blood cells as well as the bone marrow. With this kind of cancer, the white blood cell count will increase however they are doing not work properly. If someone were to have cancer, this could be the one with the very best survival rate. It is mostly found in the case of adults and very rare in the

case of children. CLL is characterized by the clonal expansion and accumulation of leukemic cells with B-lymphocyte characteristics. It may so happen that people diagnose with CLL may lead to the case of ALL [1, 4].

II. LITERATURE SURVEY

Dharni T., et al. [5], here in this paper they use globally available data set of VIT Mumbai which contains 200 images of leukemia disease of different persons. They are supplied initial preprocessing on input image to remove the effect of light and noise then extract the features from image extracted colour feature and morphological features and finally for classification for SVM classifier has been used to detect whether glaucoma disease is present or not in the image. Lim Huey nee, et al. [6], conferred an incline scale, morphological operation, thresholding, in addition as division modification to execute segmentation of specific cells. During this paper around fifty imageries were utilized to look at the planned technique as well as the impact conferred within which the method has succeeded to achieve segmentation with qualitatively smart consequences. In [9] a repetitive thresholding algorithmic rule is employed for segmentation purpose particularly from noisy pictures. This algorithmic rule overcomes the matter of cell extraction and segmentation from serious noisy pictures. This algorithmic rule works over the adjusted threshold of pictures iteratively providing strength to the image.

In [10], an approach based on HSV conversion and morphological feature extraction is propose for the detection of AML and ALL type leukemia disease. For the classification purpose various machine learning algorithm like support vector machine decision tree are used and they claim the highest accuracy of classification near about 95%. In [11], Four classes of ALL type of leukemia has been used for the classification purpose that is L1 L2 L3 and L4. All the experiments have been performed on colour images only and also data augmentation have been done for the trip offers finally the claim the specificity was 99. 3% and accuracy was 99.5 0% to perform this task they used conventional neural network model of deep learning and all the test are performed on IDB data set which is publicly available. In [12] analysis work on an automatic Cell Count methodology is represented. an explicit methodology of segmentation for investigation white blood cells automatically is presented here. First, a straightforward thresholding approach is applied and also the algorithmic rule comes from blood smear pictures from a priori data. The labels are adjusted then so as to supply significant results. This approach uses information of the blood cell structure. This methodology is additional important as compared to ancient ways that use data of local context. It will perform correct segmentation of white blood cells although they have un-sharp boundaries.

Sigi R., et al. [13], presented a method which involves preprocessing steps like, RGB to HSV conversion. Separation of unwanted pixels, by noise removal filters and then extraction of local and global features for the purpose of classification. The final-split, comprising of genes that are extremely understanding, and it's earned by examining the frequency of look of each single factor within the subsets of dissimilar factor. a mixture of geometric distance and an increased distance remodel combining intensity gradients is employed for the watershed step in [14]. An explicit mathematical model for characteristics of cell nuclei like size and form measures is used. For every detected nucleus, a confidence score is computed by menstruation quality of nucleus within the model. Paper [15] shows the quality of an automatic morphological methodology to acknowledge the Acute Lymphocytic Leukemia (ALL) with the assistance of pictures of peripheral blood microscope. The conferred methodology individuates the leucocytes from the others blood cells, at that time it selects the lymphocyte cells (the cells causes acute leukemia), morphological indexes from those cells are evaluated then once and finally classification is performed whether or not the presence of the leukemia is there or not.

III. METHODOLOGY USED FOR LEUKEMIA DETECTION

A. Preprocessing:

The preprocessing step contains the initial noise removal present in the microscopic images. To perform noise reduction, various noise removal filters can be applied like wiener filter, median filter etc. The preprocessing steps performs the changes in pixel values of image in order to reduce random noise. Thresholding is another technique used for noise reduction where, pixel value with in a range consider as an image and other pixel values are considered as a noise.

B. Feature Extraction:

Feature extraction can be done by separating the similar gray values together. Some other feature extraction technique contains the extraction of texture features, edge and shape features and colour features.

1) Morphological features: The morphological feature contains the features extracted from binary images. the major morphological feature contains shape features, edge features, number of connected components, area and perimeters of image [18].

2) Textural features: The texture feature basically defines the inter pixel relationship between number of pixels in image. The texture feature extracts the inner information of image like smoothness, roughness, sharpness and textures of image [14].

3) GLCM Features: Gray Level Co relation Matrix features are basically texture features, and are very powerful in detection and classification in image processing. Some of the major GLCM features are contrast, homogeneity, entropy and co relation of pixels [12, 13].

C. Machine Learning Techniques for Classification

1) ANN (Artificial Neural Network): The neural network model is very similar to the human brain which is connected with lots of neurons to make a very complex pattern. The artificial neural network is a collection of notes connected with lots of other roads with the help of connecting strings when the information in a particular node exceeds from a limit then it passes to the next node as information. The neural network Required two phases the First phase contains the training phase to train the model and the second phase is the testing phase to check the performance of the model. Many authors propose the use of artificial neural networks in detection of leukemia disease because when the features are extracted from the MRI images

these pictures can be used to train the neural network and finally at the testing phase we can test our neural network model with the features of remaining persons. Many authors suggested an ANN(artificial neural network) as a very good option for classification of leukemia disease. In leukemia disease detection the types of disease mainly divided into four categories so we can propose the use of a neural network model which contains the same number of neurons as of features in i/p layers and two neurons in the o/p layers to categorized the deputy a disease correctly [16].

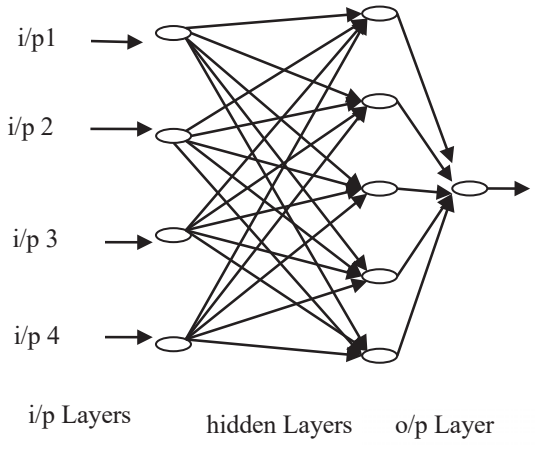


Fig.2. NNM (Neural Network Model) [16]

2) SVM (Support Vector Machine): SVM is well-known widely used classification method which is generally applied for Linear classification. It is very suitable for binary classification but now a day it becomes more popular for multiclass classification also. The support vector machine create hyper plane to divide the data into specific reason data belonging to same space are classified as the data of same class [17].

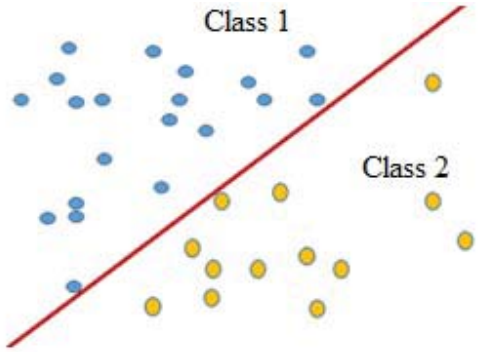


Fig.3. Support Vector Machine Classifiers [17]

3) LDA (Linear Discriminant Analysis): For the purpose of preprocessing linear discriminant analysis have been used in various researches. Linear Discriminant Analysis is "controlled" and also calculates the "linear discriminants" which will be very helpful for the segmentation in various classes. Algorithm of LDA technique is given below [19].

Step 1: Calculate discriminant matrix
Step 2: Find Eigen values.
Step 3: Sort Eigen values.
Step 4: Utilize Eigen matrix and convert to new space.
Step 5: Utilize N space and end LDA

4) Deep Learning Techniques for Leukemia Classification: Deep learning techniques are widely used nowadays for classification and identification purpose. One of the most striking feature of deep learning technique is its accuracy of classification which is always higher than the previous classifiers.

Here in leukemia disease classification we can use CNN model of deep learning for feature extraction if input is in the form of image or series of images. if we have extracted features in text or excels sheets the we can go for RNN models because they are suitable for series inputs.

Convolutional Neural Network (CNN) model contains a input layer to take image as input then is has a series of convolution layer and pooling layer for the purpose of feature extraction and feature reduction respectively. Finally, at the end, we have a fully connected layer to output layer. It also uses RELU (rectified Linear Unit) and Softmax type activation function to propagate output of lower layer to upper layer as shown in Figure 4.

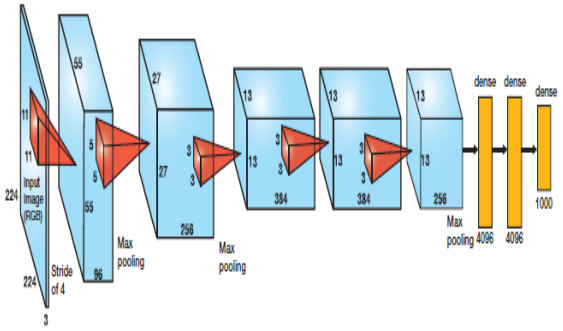


Fig.4. Alexnet architecture of CNN model

IV. PERFORMANCE ANALYSIS OF PREVIOUS APPROACH

We are analyzing the performance of different classification techniques and focusing on major techniques that can be used in future classification of disease. Table 1 summarizes the different approaches used in prediction of leukemia disease then also discussed the used datasets and results of previous methods

TABLE 1: PERFORMANCE OF DIFFERENT APPROACHES

Author and Year	Descriptions	Dataset used	Performance
Scotti F. et.al. [20]	Morphological features are extracted and for classification Linear, K-NN, classifiers are applied	Dataset collected from M. Tettamanti Research Center	They are getting Approx. 88 % accuracy

		(ALL-IDB)	
Markiewicz et al. [21]	Textural, statistical, and geometrical feature are extracted and classification done using Support Vector Machine (SVM)	Local dataset used	They claim the accuracy 90%
Gupta et al. [22]	Texture features and shape features are extracted and classification done using Support Vector Machine (SVM)	Own data set containing more than 345 images	They claim the accuracy of 93%
Escalante et al. [23]	Statistical, and morphological features are extracted and PSO technique applied for diseases classification	Dataset collected from Mexican Social Security Institute Which contains 633 images	They got the accuracy 97.68% (for binary classification) and 94.21% (multi class classification)
Putzu et al.[24]	Textural, and color features are taken for classification various classifiers are applied like SVM- L, SVM-Q, SVM P, SVM- R, K- NN	IDB1 and IDB 2 dataset used	Accuracy varying according to different classifiers
Yi et. al.[25]	In this paper an effective segmentaion method have been introdiced which can resolve the problem of over segmenation and under segmentaion using PSO.	Local dataset	NA
L.B. Dorini et al. [26]	WBC segmentation have been applied to divide the image into foregraound and back ground, then. Image processing techniques used for classification	Local dataset used	NA

Merits of existing approaches:

Uses morphological feature extraction method whereas sum of the approaches uses color texture and other features of microscope classification of leukemia disease. Well image processing method can lead to good feature extraction so that a good level of classification may achieved.

Demerits of existing approaches:

The problem which has been observed in the existing approach is related to the accuracy of classification. The classification accuracy ranges between 75% to 95% in most of the method, whereas on using deep learning models such as convolutional neural network model and other similar models this accuracy may improve up to 97% or more.

V. CONCLUSION

Here, in this paper we have discussed different preprocessing techniques, and leukemia classification techniques. Also the paper contains a brief knowledge about recent available method used for classification with results and analysis. Here, we conclude that leukemia disease can be classified using many latest machine learning algorithms. But when we have a large dataset of images then it is better to use deep learning architectures for classifications.

REFERENCES

[1] American Cancer Society, “facts spring 2014‖ Leukemia Lymphoma Society: Fighting Blood Cancer, Revised April 2014.

[2] Kalyanmoy Deb, A. Raji Reddy, ―Reliable classification of two-class cancer data using evolutionary algorithms‖, Elsevier, BioSystems , Vol.72, pp.111–129, 2003.

[3] M. Oostindjer, J. Alexander, G. V. Amdam, G. Andersen, N. S. Bryan, D. Chen, D. E. Corpet, S. De Smet, L. O. Dragsted, A. Haug et al., “The role of red and processed meat in colorectal cancer development: a perspective,” Meat science, vol. 97, no. 4, pp. 583–596, 2014.

[4] R. Takiar, D. Nadayil, and A. Nandakumar, “Projections of number of cancer cases in india (2010-2020) by cancer groups,” Asian Pac J Cancer Prev, vol. 11, no. 4, pp. 1045–1049, 2010.

[5] Dharani T,Hariprasath S., "Diagnosis of Leukemia and its types Using Digital Image Processing Techniques," in Proceedings of the International Conference on Communication and Electronics Systems (ICCES 2018) IEEE Xplore Part Number:CFP18AWO-ART; ISBN:978-1-5386-4765-3

[6] L. H. Nee, M. Y. Mashor, R. Hassan, "White Blood Cell Segmentation for Acute Leukemia Bone Marrow Images," International Conference on Biomedical Engineering (ICoBE), Penang, Malaysia,IEEE, 2012.

[7] J. M. Bennett, D. Catovsky, M.-T. Daniel, G. Flandrin, D. Galton, H. R. Gralnick, and C. Sultan, “Proposals for the classification of the acute leukaemias french-american-british (fab) co-operative group,” British journal of haematology, vol. 33, no. 4, pp. 451–458, 1976.

[8] J. W. Vardiman, J. Thiele, D. A. Arber, R. D. Brunning, M. J. Borowitz, A. Porwit, N. L. Harris, M. M. Le Beau, E. Hellstr¨ om-Lindberg, A. Tefferi et al., “The 2008 revision of the world health organization (who) classification of myeloid neoplasms and acute leukemia: rationale and important changes,” Blood, vol. 114, no. 5, pp. 937–951, 2009.

[9] H. S. Wu, J. Barba, and J. Gil, “Iterative thresholding for segmentation of cells from noisy images” Journal of Microscopy, vol. 197, no. 3, pp. 296–304, 2000.

[10] Jagadev P.,virani H.G., Detection of Leukemia and its Types using Image Processing and Machine LearningInternational Conference on Trends in Electronics and Informatics, ICEI 2017978-1-5090-4257-9/17/$31.00 ©2017 IEEE

[11] shafique S. Tehsin S., “Acute Lymphoblastic Leukemia Detection and Classification of Its Subtypes Using Pretrained Deep Convolutional Neural Networks”, Technology in Cancer Research & Treatment Volume 17: 1-7, 2018

[12] Q. Liao and Y. Deng, “An accurate segmentation method for white blood cell images,” in Proceedings of the IEEE International Symposium on Biomedical Imaging, pp. 245–258, 2002.

[13] Sigi R. ,Bachtiar, M.M. fikri, Md. , "Identification Of Leukemia Diseases Based On Microscopic Human Blood Cells Using Image Processing978-1-5386-8066-7/18/$31.00 ©2018 IEEE, 2018

[14] G. Lin, U. Adiga, K. Olson, J. F. Guzowski, C. A. Barnes, and B. Roysam, “A hybrid 3Dwatershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks,” Cytometry Part A, vol. 56, no. 1, pp. 23–36, 2003.

[15] Fabio Scotti University of Milan, Department of Information Technologies, via Bramante 65, 26013 "Automatic Morphological Analysis for Acute Leukemia Identification in Peripheral Microscope Images" IEEE International Conferenceon Computational Intelligence For Measurement Systems and Giardini Naxos, Italy, 20-22 July 2005.

[16] Mohapatra S., Patra D. and Satpathy S., ―Unsupervised Blood Microscopic Image Segmentation and Leukemia Detection using Color based Clustering‖, International Journal of Computer Information Systems and Industrial Management Applications.ISSN 2150-7988 Vol. 4, pp. 477-485, 2012.

[17] S. Mahapatra et.al, ―An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images‖, ACM, vo. 24, pp. 1887- 1904, 2014.

[18] O. L. Mangasarian, W. N. Street, and W. H. Wolberg, "Breast cancer diagnosis and prognosis via linear programming," Operations Research, vol. 43, no. 4, pp. 570–577, 1995

[19] Rao, R. C, "The utilization of multiple measurements in problems of biological classification". Journal of the Royal Statistical Society, Series B, Vol.10, pp. 159–203, 1984.

[20] F. Scotti, "Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images," in 2005 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, 2005, pp. 96–101.

[21] T. Markiewicz, S. Osowski, B. Marianska, and L. Moszczy ´nski, "Automatic recognition of the blood cells of myelogenous leukemia using svm," in Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on, vol. 4. IEEE, 2005, pp. 2496–2501.

[22] L. Gupta, S. Jayavanth, and A. Ramaiah, "Identification of different types of lymphoblasts in acute lymphoblastic leukemia using relevance vector machines." in Conference proceedings:... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference, vol. 2009, 2008, pp. 6675–6678.

[23] H. J. Escalante, M. Montes-y G´omez, J. A. Gonz´alez, P. G´omez-Gil, L. Altamirano, C. A. Reyes, C. Reta, and A. Rosales, "Acute leukemia classification by ensemble particle swarm model selection, " Artificial intelligence in medicine, vol. 55, no. 3, pp. 163–175, 2012.

[24] L. Putzu, G. Caocci, and C. Di Ruberto, "Leucocyte classification for leukaemia detection using image processing techniques," Artificial intelligence in medicine, vol. 62, no. 3, pp. 179–191, 2014.

[25] L.B. Dorini, R. Minetto, and N.J. Leite. White blood cell segmentation using morphological operators and scale-space analysis. InProceedings of the Brazilian Symposium on Computer Graphics and Image Processing, pages 294–304, October 2007.

[26] B. C. Ko, J.-W. Gim, J.-Y. Nam, Automatic white blood cell segmentation using stepwise merging rules and gradient vector flow snake (2011).