Asian Journal of Computer Science and Technology
ISSN: 2249-0701 (P) Vol.7 No.2, 2018, pp.113-117

DOI: https://doi.org/10.51983/ajcst-2018.7.2.1869

Classification of Leukemia Image Using Genetic Based K-Nearest Neighbor (G-KNN)

M. Bennet Rajesh[1] and S. Sathiamoorthy[2]
[1&2]Assistant Professor, Division of Computer and Information Science
Annamalai University, Annamalai Nagar, Tamil Nadu, India
E-Mail:benraj@gmail.com
(Received 13 July 2018; Revised 28 July 2018; Accepted 12 August 2018; Available online 18 August 2018)

Abstract **- In medical diagnostic system, classification of blood cell is more vigorous to identify the disease. The diseases which are connected with blood is alienated after the categorization of blood cell. Leukemia, a blood cancer that begins in bone marrow. Hence, it must be cured at initial stage and leads to death if left untreated. This paper introduces median filter for noise removing and Genetic based kNN for classification of Leukemia image datasets and features are extracted using gray-level co-occurrence matrix. The outcome of proposed genetic algorithm based kNN is compared with multilayer perceptron and support vector machine. The experimental outcomes evident that proposed combination performs better than the existing approach.**
Keywords: **Leukemia, K-Nearest Neighbor, Genetic Algorithm, Pre-Processing, Noise Removal, Median Filter approach**

I. INTRODUCTION

Leukemia is one type of cancer [1] that signifies the bone marrow producing well fabrication of white blood (WB) cells that inflow the bloodstream and is necessary, and is an infection-fighting section of human immune system Consequently, diagnose myelogenous leukemia has correctly analyzed by a physician based on complete blood count, bone mellow biopsy and cytogenetic analyses. Chatap *et al.,* [2] used histogram equalization and contrast brightness adjustment for improving image quality and used global threshold Ostu's algorithm for segmentation then they have extracted area, perimeter and circularity descriptors from segmented image and the features are feed to k-Nearest neighbors (kNN) algorithm to classify the image into various categories of leukaemia like Acute Lymphocytis Leukemia, Chronic Lymphocytic Leukemia, acute megakaryoblastic leukemia and Chronic myelogenous leukemia.

Regina [3] suggested a technique for automatic detection of leukaemia from microscopic images in which RGB image is transformed into CIELAB color space then the nuclei of the white blood cells are segmented based on color components using k-means cluster method. Later they have computed Hausdorff dimension as edge attribute, Local Binary Pattern and gray-level co-occurrence matrix features as texture attribute and area, perimeter, compactness, solidity, eccentricity, elongation, and form factor as shape attributes and these attributes are used to classify the microscopic images using support vector machine. In [4], adaptive mean filter and adaptive histogram equalization is employed for eliminating noise and to increase contrast of microscopy image and fuzzy c-means is employed for segmenting the nuclei, the region of interest. From the segmented nuclei, the color, texture and geometrical features are extracted and the extracted features are classified using support vector machine for identification of disease type. In [5], the microscopy images are transformed into HSV color model and they conclude that the leukocytes are more identifiable in S Component and the authors suggested Ostu's approach for choosing the threshold value and are applied for segmentation process. From the extracted region of interest, features are extracted and are classified for identification of disease.

In [6], Himali *et al.*, proposed a methodology for identifying and counting the leukemia cells in which they transformed RGB image into HSV then they found the derivation of gradient magnitude. Later they performed opening and closing morphological operations followed by watershed transform. For clustering process they incorporated k-means algorithm. In order to count the number of leukemia cells, they convert RGB image into gray scale image then applied area opening followed by hole filling. Later from the defined object boundary, they computed area, perimeter, convex hull, roundness, major axis, minor axis, standard deviation and are used as attributes to find the overlapped and non-overlapped cells. In [7], the image is transformed into grayscale image then histogram equalization is accomplished and is followed by morphological operations and then by using global threshold, segmentation operation is performed. Later the segmented objects are classified. Mohapatra *et al.,* [8] described selective filtering and unsharp masking for preprocessing stage then the preprocessed image is transformed into L*a*b* color space. K-means clustering approach has been adopted for segmentation. From segmented image they extracted fractal dimension, contour signature and shape features like area, perimeter, solidity, compactness, eccentricity, form factor and elongation. In addition to that they also extracted color and gray-level co-occurrence features and all these features are given as input to SVM for classification of leukemia. Mohapatra *et al.,* [9] performed detailed comparative study using the Fuzzy Probabilistic C Means K-Medoid, K-Means, Fuzzy C-Means techniques for segmentation process and they employed contour signature, Gustafson Kessel, Hausdorff Dimension, features for leukemia detection.

In [10], watershed algorithm and optimal thresholding is suggested for segmenting normal and abnormal lymphocytes into nucleus and cytoplasm. By suppressing the 1% of local minima in watershed algorithm, the authors reduced the error rate in segmentation process and it leads to good accuracy. In [11], k-means approach, local directional path and support vector machine has been adopted to detect the acute leukemia accurately. In [12], segmentation is achieved by using the K-means algorithm. Spatial and spectral descriptors are extracted from segmented output and the spectral features are optimized using Genetic algorithm. The computed attributes are classified by support vector machine to diagnosis the leukemia accurately. A detailed review on leukemia detection and classification is found in [13, 14].

Zeinab *et al.,* suggested automatic methodology for acute leukemia diagnosis using blood microscopic images by using the color, shape and texture features, and ensemble classifier has been used to classify the normal and unnormal leukocytes. In line with this, number of researchers proposed many methodologies for classification of leukemia. Various pre-processing, features, segmentation and classification has been discovered. However, still there is lack of technology for accurate classification of acute leukemia and it is because of lack in choosing the appropriate techniques for each and every stage of leukemia detection. In this work, we proposed a genetic based kNN for leukemia image classification and we will get significantly better results than the existing approaches.

II. PROPOSED METHODOLOGY

A. Filter based Noise Removal

The best-known request measurement filter in computerized image processing is the median filter. It is a valuable instrument for decreasing salt-and-pepper noise in a picture. The median filter assumes a key part in picture processing and vision. In median filter, pixel estimation of point p is supplanted by median of pixel estimation of 8-neighborhood of point 'p'. Median filter can be expressed as:

$$g(p) = median\{f(p), where\ p \in N_8(p)\}$$

The median filter is mainstream in view of its exhibited capacity to diminish arbitrary incautious noise without obscuring edges as much as practically identical direct low pass filter. In any case, it regularly neglects to perform well as direct filters in giving adequate smoothing of no rash noise parts, for example, added substance Gaussian noise. One of the primary detriments of the essential median filter is that it is area invariant in nature, and hence additionally has a tendency to modify the pixels not aggravated by noise.

B. Gray level co-occurrence matrix

Since Gray level co-occurrence matrix captures spatial relationship of pixels in an image. GLCM computes the frequency of a pairs of pixel with specific values horizontally (0^0), vertically (90^0) or diagonally (45^0 and 135^0) and in specified spatial relationship. From the computed GLCM, second order statistical features like correlation, contrast, energy and homogeneity are computed and they measures local variations, joint probability occurrences of specified pair of pixel, uniformity and closeness of distribution of an elements respectively.

C. K-Nearest Neighbor Approach

In image processing, k-nearest neighbor algorithm is a non-parametric strategy utilized for regression and classification. In the two cases, the input comprises of the k nearest training cases in descriptor space. kNN is a sort of instance-based learning. In k-NN classification, class membership is the output. Classification is finished by a larger part vote of neighbors. The most limited separation between any two neighbors is dependably a straight line and separation is called as Euclidean distance. The constraint of k-NN algorithm is it's sensitive to the nearby design of the information. In feature space, extraction is taken place on raw data before using k-NN algorithm. The Figure 1 portrays the steps engaged with KNN method.

D. Genetic Algorithm

Genetic Algorithm (GA) is a hunt heuristic that mirrors the procedure of natural evolution. This heuristic is regularly utilized to generate helpful answers for search and optimization issues. Genetic algorithms have a place with larger class of Evolutionary algorithms (EA), which create answers for optimization issues utilizing methods inspired by normal evolution, for example, selection, inheritance, crossover and mutation. GAs is coined by Darwin's Theory about Evolution "Survival of Fittest". GAs is versatile heuristic search in light of the evolutionary thoughts of genetics and natural selection. GAs recreates the survival of the fittest among people over successive years for taking care of an issue. Every year comprises of a populace of character strings that are analogous from the chromosome that we find in our DNA.

Every individual represents a point in hunt space and conceivable solution. The people in populace are then made to experience a process of evolution. The GA keeps up a populace of n chromosomes (arrangements) with related fitness values. Parents are chosen to mate, based on their producing offspring, fitness by means of a reproductive plan. Therefore profoundly fit arrangements are given more chances to repeat; with the goal that posterity acquires attributes from each parent. As guardians mate and deliver posterity, room must be made for fresh introductions since populace is kept at static size. People in populace kick the bucket, and are supplanted by novel arrangements, in the long run making another age once all mating chances in the old populace have been depleted.

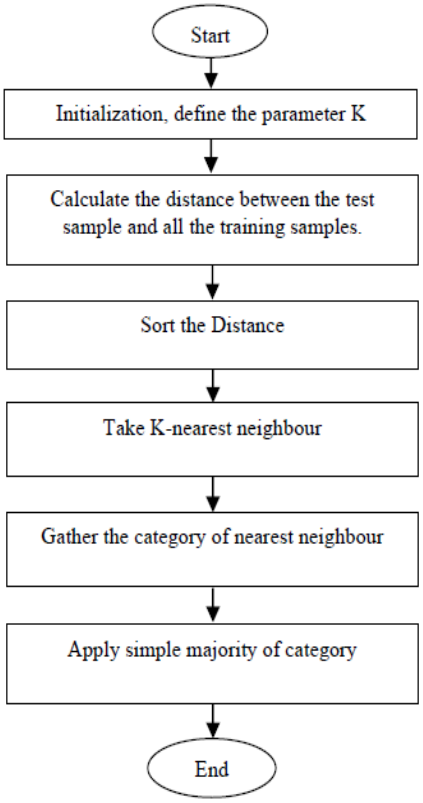

Fig. 1 KNN Classification Algorithm

Along these lines it is trusted that over progressive ages better arrangements will flourish while the minimum fit arrangements cease to exist. New ages of arrangements are delivered containing, by and large, preferable qualities over an ordinary arrangement in a previous age. In the end, once the populace has combined and isn't delivering posterity recognizably not the same as those in previous ages, the algorithm itself is said to have joined to an arrangement of answers for the current issue.

D. Proposed Genetic based KNN Framework for Leukemia Image Dataset

In the proposed approach, the gray-level co-occurrence texture features like contrast, correlation, energy and homogeneity are computed from segmented image and the global level thresholding obtained by the Ostu's algorithm is applied for segmentation and G-kNN algorithm is proposed to choose the best k-esteem with the base misclassification rate. The k-NN algorithm is truly outstanding and generally utilized for classification and characterization. In this strategy, each example esteem fit to the test dataset and it is ordered by the nearest k test in view of the preparation information. The class numbers esteems got from k test esteems, the most extreme number is resolved from class tests. The separation estimation figured by utilizing Euclidean separation measure. The separation of nearest neighbors is computed in view of the test picture utilizing separation weighted equation as.

$$\text{distance (x)} = \sqrt{\sum_{i=1}^{n} weight_i (x_i - y_i)^2}$$

where x and y are two pictures, n is the quantity of features. Therefore, the unidentified example esteem is chosen most relevant to the class from k nearest neighbor algorithm and it is utilized to locate the genuine incentive from unidentified example esteems. In this proposed technique, k esteem is acquired by Genetic Algorithm.

Pseudo code for Proposed Genetic based KNN method is as follows:

Step 1: Choose k number of tests from the preparation set to produce the underlying populace (p1).
Step 2: Calculate the separation between the preparation tests in every chromosome and the testing tests, as wellness esteem.
Step 3: Choose the chromosome with most elevated wellness esteem and store it as worldwide maximum (Wmax).
Step 3.1: For I = 1 to L do
Step 3.2: Perform Reproduction
Step 3.3: Apply the Crossover administrator
Step 3.4: Perform change and get the new populace. (p2)
Step 3.5: Calculate the nearby maximum (Nmax).
Step 3.6: If Wmax < Nmax at that point
Step 3.6.1: Wmax = Nmax;
Step 3.6.2: p1 = p2;
Step 3.6.3: Repeat
Step 4: Output: Chromosome that acquires Wmax and has optimum k-neighbors

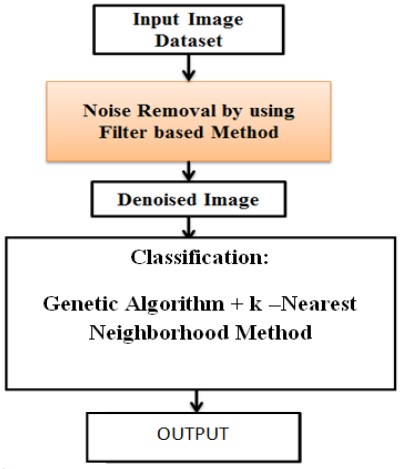

Fig. 2 Proposed framework

III. RESULT AND DISCUSSION

The acute lymphoblastic leukemia image datasource [https://homes.di.unimi.it/scotti/all/] is used in the experimental study. The efficiency of Genetic based kNN method is computed using the classification accuracy, sensitivity, specificity, processing time and finding minimum distance. The classification accuracy, specificity and sensitivity computed by applying accuracy formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

where
True Positive (TP) = accurately classified positive values
True Negative (TN) = accurately classified negative cases
False Positive (FP) = inaccurately classified negative values
False Negative (FN)=inaccurately classified positive values.

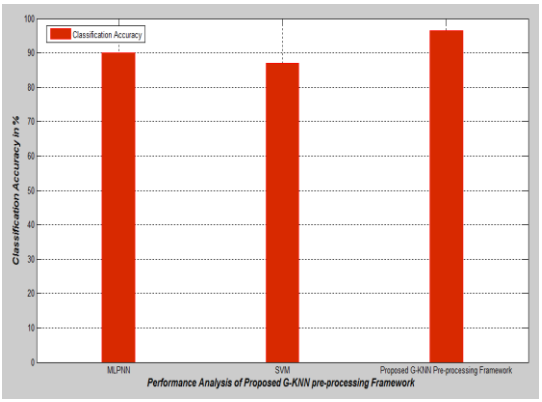

Fig. 3 Graphical illustration of efficiency analysis of suggested framework with existing methods

TABLE I EFFICIENCY ANALYSIS OF THE SUGGESTED FRAMEWORK AGAINST CLASSIFICATION ACCURACY, PROCESSING TIME AND MISCLASSIFICATION RATE

Methodology	Classification accuracy	Processing Time	Misclassification Rate
MLPNN	90%	3 sec	3%
SVM	87%	3.5 sec	1.73%
Proposed G-KNN pre-processing Framework	96.57%	2.9 sec	1.5%

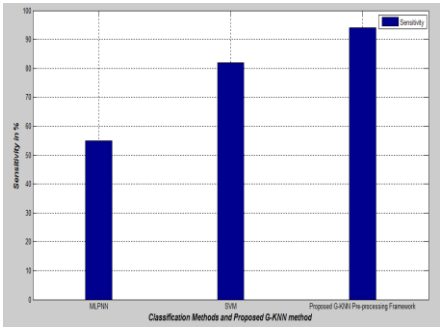

Fig. 4 Graphical illustration of efficiency analysis of suggested framework with existing methods against Sensitivity

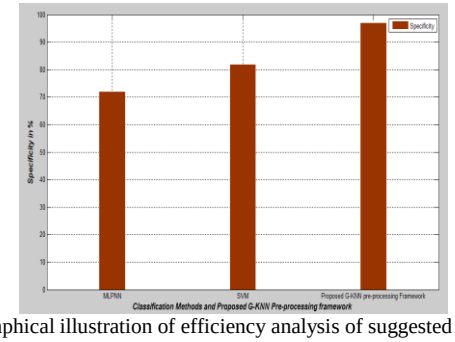

Fig. 5 Graphical illustration of efficiency analysis of suggested framework with existing methods against Specificity

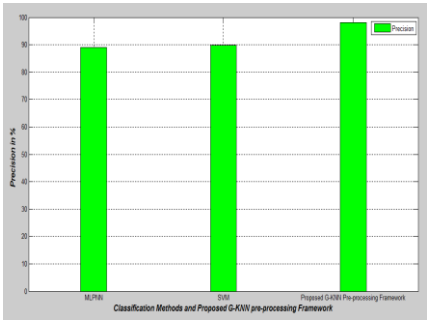

Fig. 6 Graphical illustration of efficiency analysis of suggested framework with existing methods against Precision

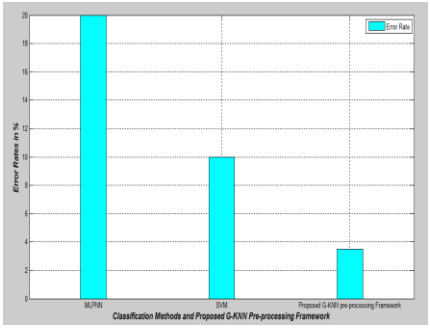

Fig. 7 Graphical illustration of efficiency analysis of suggested framework with existing methods against Error rate

TABLE II EFFICIENCY ANALYSIS OF THE SUGGESTED FRAMEWORK AGAINST ACCURACY, SENSITIVITY, SPECIFICITY, PRECISION AND ERROR RATE

Classification Methods	Efficiency Analysis in %				
	Accuracy	Sensitivity	Specificity	Precision	Error
MLPNN	90%	55	72	89	20
SVM	87%	82	82	89.9	10
Suggested G-KNN Framework	96.57%	94	97	98	3.5

From the above figures 3-7, it is clear that proposed Genetic based k-nearest neighbor framework delivers better result than existing methods like support vector machine and multi-layer perceptron neural network. The proposed combination of techniques with G-KNN framework outperforms the existing techniques.

IV. CONCLUSION

In this paper, G-kNN is proposed for classification of leukemia images and the features are extracted using the gray-level co-occurrence matrix, and for preprocessing we employed median filter. The proposed methodology outdoes in classification when compare to multilayer perceptron and support vector machine. The proposed combination of techniques is very useful for identifying the various categories of leukemia more accurately. In future we enhance this work for further improving the accuracy.

REFERENCES

[1] "Understanding Leukemia", Leukemia and Lymphoma Society Fighting Blood Cancers.

[2] Chatap, Niranjan and Sini Shibu, "Analysis of blood samples for counting leukemia cells using Support vector machine and nearest neighbour", *IOSR Journal of Computer Engineering (IOSR-JCE)*, Vol. 16, No. 5, pp. 79-87, 2014.

[3] Arputha Regina, "Detection of Leukemia with Blood Microscopic Images," IJIRCCE, Vol.3, Special Issue 3, April 2015.

[4] Tejashri G. Patil and V. B. Raskar., "Blood microscopic image segmentation & acute leukemia detection," IJERMT, Vol. 4, No. 9, Sep, 2015.

[5] C. Vidhya, P. Saravana Kumar, K. Keerthika, C. Nagalakshmi, and B.Medona devi., "Classification of acute lymphoblastic leukemia in blood microscopic images using SVM", *ICETSH-2015*.

[6] Trupti, M., A. Kulkarni-Joshi, and P. D. S. Bhosale, "A fast segmentation scheme for acute lymphoblastic leukemia detection", *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, Vol. 3, No. 2, pp. 7253-7258, 2014.

[7] Himali P. Vaghela, *et al.,* "Leukemia detection using digital image processing techniques", *Leukemia*, Vol. 10, No. 1, pp. 43-51, 2015.

[8] Deore, Sonal G., and Neeta Nemade, "Image analysis framework for automatic extraction of the progress of an infection", *International Journal of Advanced Research in Computer Science and Software*, 2013.

[9] Mohapatra, Subrajeet, Dipti Patra, and Sanghamitra Satpathy, "Unsupervised blood microscopic image segmentation and leukemia detection using color based clustering", *International Journal of Computer Information Systems and Industrial Management Applications*, Vol. 4, pp. 477-485, 2012.

[10] Emad A. Mohammed, *et al.,* "Chronic lymphocytic leukemia cell segmentation from microscopic blood images using watershed algorithm and optimal thresholding", *Electrical and Computer Engineering (CCECE), 2013 26th Annual IEEE Canadian Conference on*. IEEE, 2013.

[11] Deepshikha Goutam and Sarva Sailaja, "Classification of acute myelogenous leukemia in blood microscopic images using supervised classifier", IEEE *International Conference on Engineering and Technology* (ICETECH), 2015.

[12] P. Indira, T. R. Ganesh Babu and K. Vidhya, Detection of Leukemia in Blood Microscope Images, *IJCTA*, Vol. 9, No. 5, 2016, pp. 63-67.

[13] M. A. Alsalem, A. A. Zaidan and S. Alsyisuf, *et al.,* "A review of the automated detection and classification of acute leukemia: Coherent taxonomy, datasets, validation and performance measurements, motivation, open challenges and recommendations", *Computer Methods and Programs in Biomedicine*, 2018.

[14] M. A. Alsalem, A. A. Zaidan, K. I. Mohammed, "Systematic Review of an Automated Multiclass Detection and Classification System for Acute Leukaemia in Terms of Evaluation and Benchmarking, Open Challenges, Issues and Methodological Aspects", *Journal of Medical Systems*, 2018

[15] Zeinab Moshavash, Habibollah Danyali and Mohammad Sadegh Helfroush, "An Automatic and Robust Decision Support System for Accurate Acute Leukemia Diagnosis from Blood Microscopic Images", *Journal of Digital Imaging*, 2018.