JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Detection and Classification of Leukemia using DCNN Algorithm

[1]S. BhavanaSeshadri,[2] Dr. P. Harish, [3]Dr. N. Pushpalatha,[4]Dr. I. Suneetha

[1]M.Tech Student, Department of ECE, Annamacharya Institute of Technology and Sciences,Tirupati, Andhra Pradesh, India.

[2]Associate Professor, Department of ECE, Annamacharya Institute of Technology and Sciences,Tirupati, Andhra Pradesh, India.

[3]Professor, Department of ECE, Annamacharya Institute of Technology and Sciences, Tirupati, Andhra Pradesh, India.

[4]Professor, Department of ECE, Annamacharya Institute of Technology and Sciences, Tirupati, Andhra Pradesh, India.

Abstract: **- Leukemia is a fatal disease of white blood cells which affects the blood and bone marrow in human body. Deployed deep convolutional neural network for automated detection of acute lymphoblastic leukemia and classification of its subtypes into 4 classes, that is, L1, L2, L3, and Normal which were mostly neglected in previous literature. In contrary to the training from scratch, deployed pre-trained Online AlexNet which was fine-tuned on our data set. Last layers of the pretrained network were replaced with new layers which can classify the input images into 4 classes. To reduce overtraining, data augmentation technique was used. For acute lymphoblastic leukemia detection, achieved an acute lymphoblastic leukemia subtype classification the Accuracy was 96%, and specificity was 92.85%, our proposed method was able to achieve high accuracy without any need of microscopic image segmentation.**

Index Terms – MSVM, DCNN, Leukemia, Accuracy, Sensitivity, Specificity etc.

1. INTRODUCTION

Leukemia is produced from the bone marrow. Leukemia can cause death if treatment is not started at correct time. A thin material inside each bone is termed as bone marrow. There are three type of blood cells in every human body, they are RBC (red blood cells), WBC (white blood cells) and PLT (platelets).

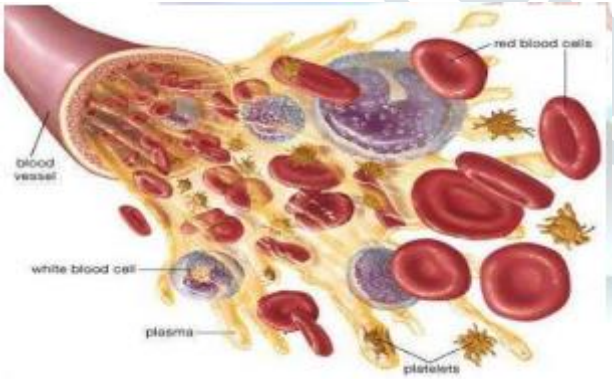


Fig.1 Blood components

The main reason of this paper is to detect leukemia occurrence. So here, concentrate only on the WBC count. Myeloid and lymphoid are the two types of stem cells. Myeloid blast which is emerged from myeloid stem cell is a cause for production of RBC, WBC and platelets. Lymphoid blast which is emerged from lymphoid stem cells is a cause for production of WBC.

Bone marrow induces abnormal white blood cells (WBCs). These abnormal cells should die after some short period of time. But actually, they do not die and they become more in count. The normal white blood cell interferes with those abnormal white blood cells in performing their normal work. And this set of circumstances is known as Leukemia. Leukemia can be divided into Chronic and Acute leukemia.

Chronic Leukemia: -Abnormal white blood cells behave like normal white blood cells with gradual increase in their count.

Acute Leukemia: -Abnormal white blood cells do not behave like a normal cells and they with rapid increase in number.

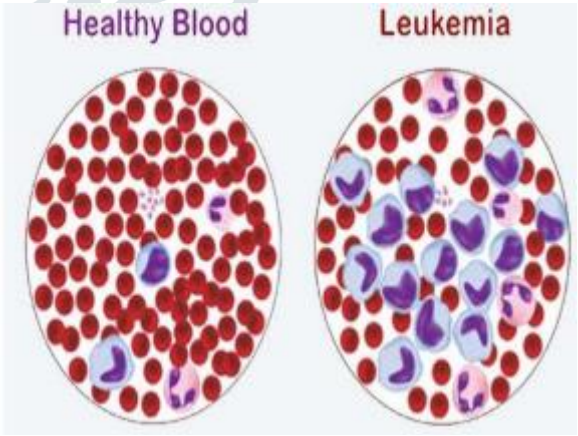


Fig.2 Healthy blood and Leukemia

The organizational framework of this study divides the research work in the different sections. The Literature survey of different papers are presented in section 2. Further, in section 3 shown Concept of Existing system is discussed and in section 4 shown the concept of proposed method, Simulation Results work is shown in section 5. Conclusion and future work are presented by last sections 6.

2. LITERATURE SURVEY

Kumar et al. [1] presented an automated detection system for acute leukemia. The system started with the pre-processing of noise and blurring in microscopic digital images. A variety of features, including color, geometric, textural, and statistical, were extracted and classified as benign or malignant. Two classification models, k-nearest neighbor (K-NN) and naïve Bayes, were used. Experiments on a dataset of 60 blood samples revealed the superiority of the K-NN classifier with its 92.8% classification accuracy.

Supardi et al. [2] introduced a classification system that differentiates between two types of acute leukemia: acute myelogenous leukemia (AML) and acute lymphocytic leukemia (ALL). Twelve features were manually extracted from image samples. Finally, a K-NN classifier was used for classification. Experiments on a dataset of 1500 images produced 86% accuracy.

Madhukar et al. [3] proposed an AML classification system that enhanced image contrasting and extracted five features. An SVM classifier performed the classification. Experiments on a dataset of 50 images produced 93.5% classification accuracy.

Setiawan et al. [4] introduced a system that could classify the cells in AML of subtypes M4, M5, and M7. Firstly, the cells were segmented by a color k-means algorithm. Then, six statistical features were extracted and input into a multi-class SVM classifier. The results produced about 87% segmentation accuracy and 92.9% classification accuracy in the best case.

Faivdullah et al. [5] proposed a three-layered framework with feature extraction, coding, and classification. Given a blood smear image of a certain patient, the objective of this framework was to decide whether a patient has leukemia and to identify which type. Dense scale-invariant feature transform was used in feature extraction. Then, the dimensionality of the extracted feature vectors was reduced in the coding layer. Finally, a multi-class SVM classifier performed the classification. Experiments on a dataset of 400 samples produced 79.38% classification accuracy.

Laosai and Chamnongthai [6] presented an AML classification system that segments nuclei by k-means and contour signature approaches. Then, feature extraction for cell size, cell color, etc., was performed via morphology. Experiments on a dataset of 100 images showed that the SVM classifier had an accuracy of up to 92%.

Patel and Mishra [7] presented an automated leukemia system based on microscopic images. The system started with noise and blurring removal during pre-processing. Then, the WBCs were segmented by k-means and Zack algorithms. Then, several features, including color, statistical, geometric, and textural, were extracted. Finally, an SVM classifier distinguished between normal and abnormal images. Experiments on a dataset of 27 images produced 93.57% accuracy.

Sajjad et al. [8] proposed a mobile-cloud-assisted framework that segments and classifies leukocytes into five classes. The framework firstly segmented white blood cells (WBCs) by a color k-means algorithm, which removed irrelevant components via morphological operations. Various types of features, including geometric, statistical, and textural, were extracted by principal component analysis. Finally, classification was performed by an ensemble multi-class support vector machine (SVM). Experiments on a dataset of 1030 blood smear WBC images produced an average accuracy of 98.6% for this framework.

Abdeldaim et al. [9]. WBCs were segmented by a combination of approaches, including histogram equalization and the Zack algorithm. Then, various features, including color, shape, and texture, were extracted and normalized. Finally, a number of classifiers were used for classification. The system was evaluated with a dataset of 260 images. With 96.01% accuracy, the best result was achieved by the K-NN classifier.

Dwivedi [10] presented a system to differentiate between ALL and AML by using the microarray gene profile and an artificial neural network for classification. The system was evaluated with a dataset of 46 samples. With 98% accuracy, the artificial neural network-based classifier achieved the best results of all the classification models. Further work to detect AML was suggested by

Finally, Sahlol et al. [11] presented an automated system to diagnose ALL. The system segmented the WBCs by histogram equalization and the Zack algorithm. Then, several features, including color, texture, shape, and hybrid features, were extracted. Then, a social spider optimization algorithm selected the most significant features. Finally, a number of classifiers were used for classification. The system was evaluated with a dataset of 260 images. With 95.67% accuracy, the K-NN classifier produced the best results.

3. EXISTING METHOD

The process for automated leukemia detection consists of 5 major modules including preprocessing, segmentation, identification and separation of grouped lymphocytes, feature extraction and classification. my overview of existing methodology was illustrated in Fig 3.

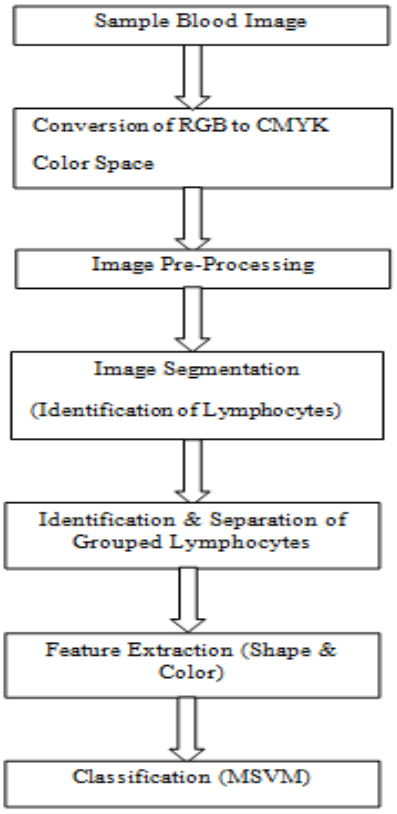


Fig.3:Existing System block diagram

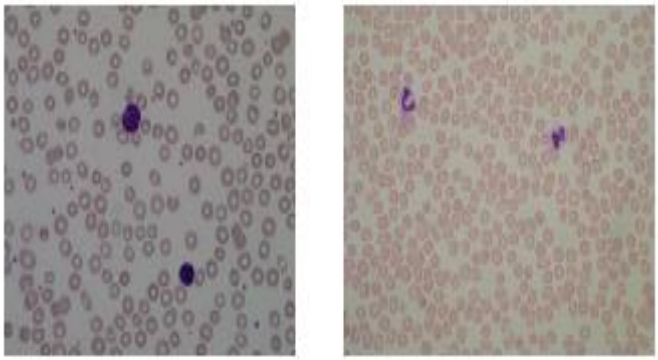
A microscopic blood image comprises of white blood cells, red blood cells and platelets. The proposed architecture first performs preprocessing over these blood smears. Then

segmentation is carried out to identify the lymphocytes. After those grouped lymphocytes are identified and separated. Then different features are extracted from the cells and classification is carried out to classify normal and blast cells.

Dataset

Blood images utilize in this research are acquired from ALL-IDB-1 dataset which is available online as a public dataset [10].

These images have 3 types of blood cells present i-e red blood cells, White blood cells and Platelets. To detect the leukemia firstly, separate white blood cells from other components of cell. Then the divided dataset needs to training samples and testing samples Blood sample images from ALL-IDB-1 dataset are shown in Fig 4

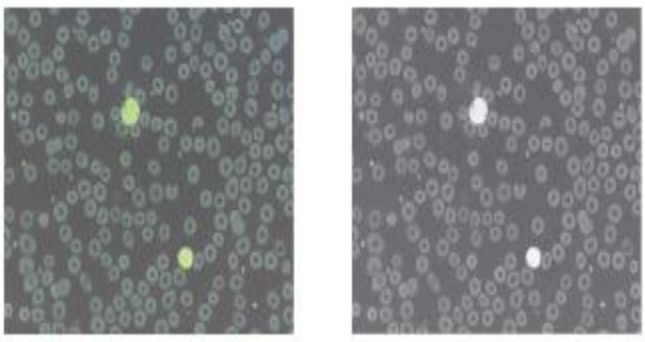


(a) Leukemia Image (b) Healthy Image

Fig. 4: Blood Sample Images from ALL-IDB-1

A. Preprocessing

In preprocessing firstly transformed the blood images from RGB (red, green, blue) to CMYK (cyan, magenta, yellow and black) color model. This highlights the WBCs from the other cells present in the blood image. Due to exposure of microscope, images have effects of unnecessary noise and blurriness. This may affect the quality of blood images and create chances of wrong diagnosis. Histogram equalization is a very common method that uses the image histogram to tune the contrast of image. [11] Therefore in preprocessing step the applied histogram equalization over the blood image to overcome different lightning effects during image grabbing. Sample Images after preprocessing are shown in Fig 5.



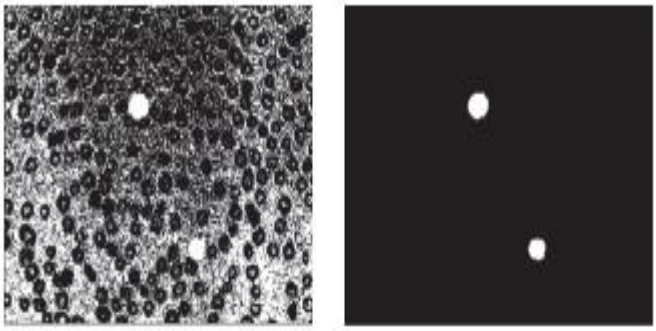
(a) Blood Image in CMYK (b) Blood Image in Gray Scale

Fig. 5: Blood Sample Image during Preprocessing

B. Image Segmentation

Segmentation is a process of obtaining region of interest by dividing the image into different regions. For the detection of lymphocytes from the blood sample need to separate WBCs from other cells and from the background. So that the applied Triangle oriented threshold method called Zack's Algorithm [12] over the blood samples to obtain required WBCs. In this algorithm a straight line is constructed between highest value of the image histogram and the lowest value of image histogram. After this an optimal threshold value is calculated and segmentation is carried out using that threshold value. Zack algorithm is found to be more effectual when the histogram shows limpid difference between the maximum and minimum value. After applying thresholding algorithm area opening function with the structuring element having circular shape is used to further clean the segmented image. After applying area opening all the objects which are smaller than the structuring element are removed from the image.

The size of structuring element was calculated depending on the average size of different objects present in the blood image. Fig 6. shows the result of segmentation and area opening.



(a) Image Segmentation (b) Background Removal

Fig. 6: Segmentation Result

D. Identification and Separation of Grouped Lymphocytes

In this stage one must identify and separate the grouped and ungrouped lymphocytes from the image. For this purpose, utilize roundness ratio to separate the grouped lymphocytes.

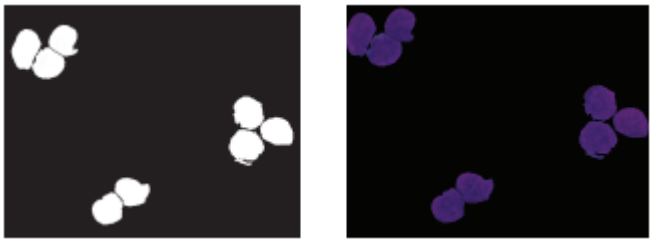
$$Roundness = \frac{4 * \pi * area}{(convex\_perimeter)^2}$$

When roundness is equal to 1 it means that the detected object is circular and if the roundness is less than 1 it means the object is not circular. Perceived that a roundness ratio of 0.85 will be good to distinguish single lymphocyte from the group of lymphocytes. So, the objects having roundness ratio greater than the provided threshold are classified as single lymphocyte and the objects having roundness ratio less than the provided threshold are classified as grouped lymphocytes which needs to be separated in the next step. The image containing grouped lymphocytes is shown in Fig 7.

Fig. 7: Grouped Lymphocytes

To separate the grouped lymphocytes, apply watershed segmentation to the distance transform. It is an image segmentation method which starts from the initial pixel called marker and consistently torrent all other adjoining pixels of that marker known as catchment basin. The advantage of this technique over other methods is that the regions develop from the resulting boundaries are closed and connected as compared to standard edge-based methods that create disconnected boundaries and usually post-process is required to form closed regions [13].

To remove the unnecessary spots and to fill the small holes in the nucleus of the cells binary morphological opening and closing operation are performed over the blood image. Size of structuring element used in morphological operation is smaller than the nucleus but have enough size to remove the unnecessary spots and fill the small holes. Final resulted images are shown in Fig 8.



(a) Watershed Result (b) Final Result

Fig. 8: Result of Watershed and Morphological Operations

E. Features Extraction

In features extraction retrieved the information from the images like their specific structure of shape, color, texture etc. By using features extraction, which can reduce the dimensions of images in such a way that they become more instructive and less redundant. This method is very useful when an algorithm must process huge dataset which can be repetitive also, then by applying this method the dataset will be reduced to minimum dimensions which contains the most relevant information of the image. This data will be given to algorithm as input which will be easier for the algorithm to interpret and compute the results. For leukemia detection the extracted shape and color features to obtain the relevant information from the blood images

Classification

For the order utilized Multi Support Vector Machine Algorithm (MSVM) classifier which is a directed learning calculation. It is extremely successful when to perform double grouping, accordingly, utilize this calculation on the grounds that to order two classes typical and impact. The given our extricated include vector as a contribution to the MSVM which then, at that point, group the cells into typical and impact contingent upon their separate properties. MSVM is utilized with the straight piece to accomplish the greatest precision and is successful in the high layered spaces. At the point when the quantity of elements is broad, other part will most likely be unable to plan with elite execution and achieved an accuracy of 93.70%.

4. PROPOSED METHOD

The purpose of our project is to develop a system that can automatically detect cancer from the blood cell images. This system uses a convolution network that inputs a blood cell images and outputs whether the cell is infected with cancer or not. The appearance of cancer in blood cell images is often vague, can overlap with other diagnoses, and can mimic many other benign abnormalities. These discrepancies cause considerable variability among medical personnel in the diagnosis of cancer. Automated detection of cancer from blood cell images at the level of expert medical personnel would not only have tremendous benefit in clinical settings, it would also be invaluable in delivery of health care to populations with inadequate access to diagnostic imaging specialists.

A. DATASET

Images used in this study were obtained from ALL-Image Database (IDB) data set which is a public data set available online[12]. This data set was divided into 2 versions. Acute lymphoblastic leukemia-IDB 1 consisted of 108 images where 59 images were from healthy patients and 49 images were from patients affected with leukemia. Acute lymphoblastic leukemia-IDB 2 data set consisted of 260 images having single cell where 130 images were from patients affected by leukemia and 130 were normal images. These images had resolution of 257 × 257 with 24-bit color depth. In below Fig, shown the samples of cancerous and healthy images of ALL-IDB2. According to FAB classification, ALL was further categorized into 3 subtypes, which were L1, L2, and L3. Our main objective of this study was to classify the subtypes of ALL which were mostly neglected in the previous literature. Due to their interclass similarity and intraclass variability they were difficult to classify. Therefore,used the ALL-IDB 2 data set. For the classification of subtypes of ALL, these blasted images were labeled L1, L2, and L3 by an expert oncologist who labeled each blasted image into ALL subtype manually.

B. METHODOLOGY

The process for automated leukemia detection consists of 5 major modules including pre-processing, segmentation, identification and separation of grouped lymphocytes, feature extraction and classification.In my proposed method, used a contrast enhancement filtering to perform pre-processing of blood images. Then fuzzy c-mean was carried out for the segmentation of white blood cells. Features including shape features, color features and texture features were extracted using GLCM(Grey Level Co-occurrence Matrix). After that classification was carried out using Deep Convolutional Neural Network (DCNN) to classify normal and blast cells. The flowchart of the proposed system has been shown in below Fig 9.

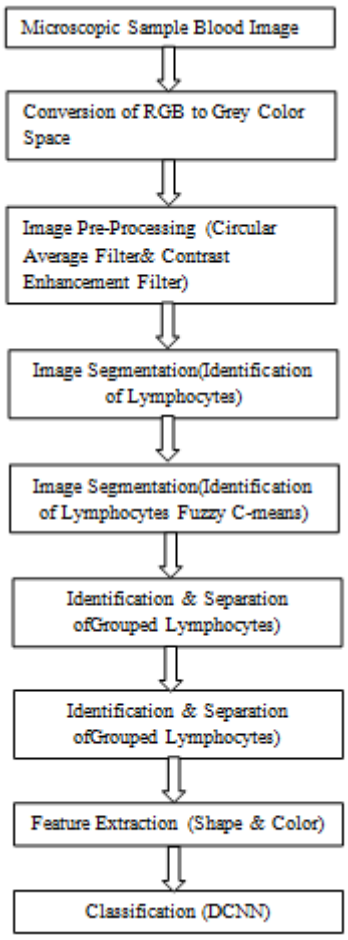


Fig.9:Proposed Block Diagram

C. Material and Methods

Data Set Images used in this study were obtained from (Acute Lymphoblastic Leukemia) ALL-Image Database (IDB) data set which is a public data set available online. This data set was divided into 2 versions. Acute lymphoblastic leukemia-IDB 1 consisted of 108 images where 59 images were from healthy patients and 49 images were from patients affected with leukemia. Acute lymphoblastic leukemia-IDB 2 data set consisted of 260 images having single cell where 130 images were from patients affected by leukemia and 130 were normal images.

D. Image Pre-Processing

The steps taken are:

1. Read image
2. Resize image
3. Remove noise (Denoise)
4. Segmentation

Average (or mean) filtering is a method of 'smoothing' images by reducing the amount of intensity variation between neighboring pixels. The average filter works by moving through the image pixel by pixel, replacing each value with the average value of neighboring pixels, including itself.

Contrast enhancement processes adjust the relative brightness and darkness of objects in the scene to improve their visibility. The contrast and tone of the image can be changed by mapping the gray levels in the image to new values through a gray-level transform.

E. Fuzzy C-Means Segmentation

The intensity image matrix is given as input to fuzzy C-means (FCM) clustering algorithm, that processes the given image and its components as fuzzy sets based on the objective function selected to the problem. The reason behind choosing FCM is while doing the segmentation it sets a global threshold value which is not only considering the region of interest, instead it is considering all the pixels of the sample image. Every pixel of the fuzzy image has a degree of belongingness to all the clusters rather than a single cluster in FCM. Which means the intensity value of sample pixel is a part of all the clusters in varying degrees lying in between 0 and 1. The pixels with the lower degree are located at the edges of clusters. This significant factor is needed to increase the sensitivity of medical image processing. The fuzzy segmentation of the blood smear image is done upon three classes namely small, middle, and large. If the cut-off value is set as 0 then the threshold value falls in between small and middle class, if it is 1 then the threshold value will be in between middle and large class.

F. Feature Extraction

GLCM is one type of feature extraction. GLCM is abbreviated as (Grey Level Co-occurrence Matrix) or Grey-Level Spatial Dependence Matrix. GLCM is a statistical process of analyzing texture which narrates the connections of spatial of every pixel. The information of texture of an image are gained by these statistics, they are contrast, correlation, energy, homogeneity.

G. Classification of DCNN

Neural networks are used in the automatic detection of cancer in blood samples. Neural network is chosen as a classification tool due to its well-known technique as a successful classifier for many real applications. The training and validation processes are among the important steps in developing an accurate process model using DCNNs. The dataset for training and validation processes consists of two parts; the training features set which are used to train the DCNN model; whilst a testing features sets are used to verify the accuracy of the trained using the feed- forward

back propagation network. In the training part, connection weights were always updated until they reached the defined iteration Number or suitable error. Neural networks are used in the automatic detection of cancer in blood samples.

Neural network is chosen as a classification tool due to its well-known technique as a successful classifier for many real applications. The training and validation processes are among the important steps in developing an accurate process model using DCNNs.

5. SIMULATIONS RESULTS

The input image in a specific format is read from the path provided. The image is resized to analyse the optimum sized images shown in Fig10.

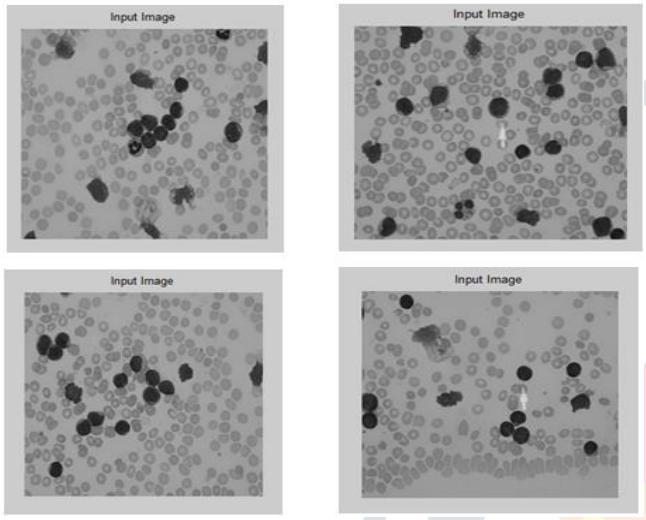


Fig. 10: Input image from data base

The input image is filtered using unsharp contrast enhancement filter and the circular averaging filter to display output as the filtered image shown in Fig 11.

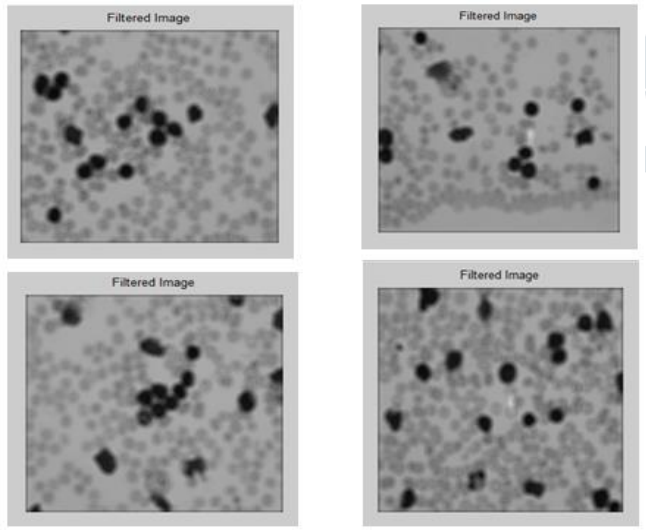


Fig 11: Filtered image

The filtered image as shown in the above Fig is converted into the binary image of black and white form with a threshold, if all the image pixels luminance is greater than threshold then 1(white) and otherwise 0(black) as BW image shown in Fig12.

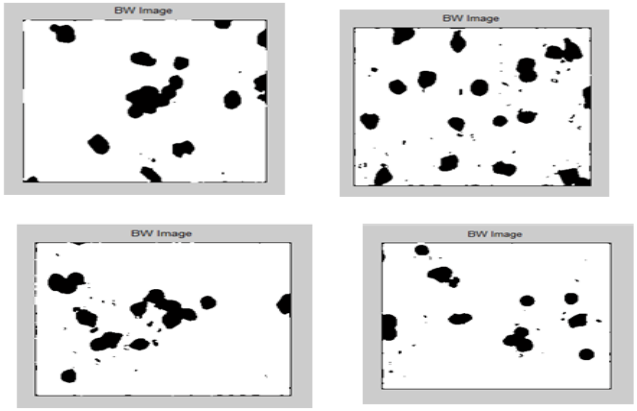


Fig. 12:Black & White image binary form

The previous image is segmented using the morphological operations and the contour is detected for the image is shown in Fig 13.

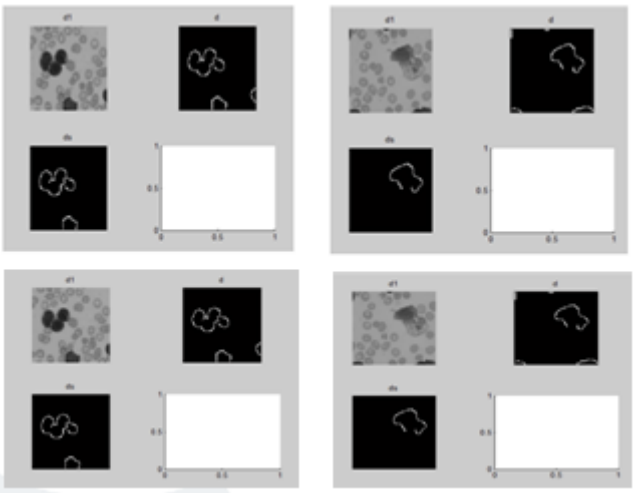


Fig. 13: Segmented images

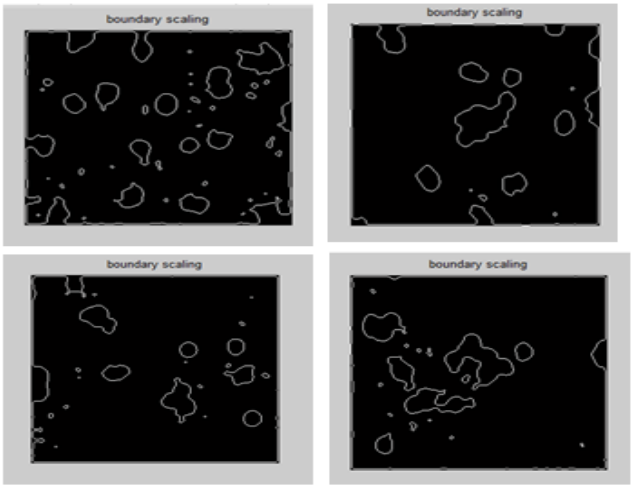


Fig. 14: boundary images

The background of the image is detected and boundary scaling of image.The Features of the images are captured as shown in the Fig 15.

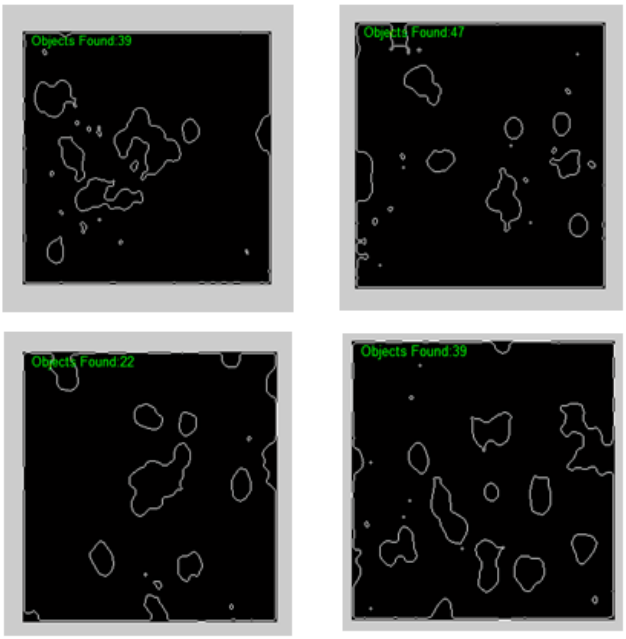


Fig.15: Feature Extraction

CONFUSION MATRIX

For 100 Images

Positive = Cancer Cell

Negative = Normal Cell

Type	Positive	Negative
Positive	True Positive	False Positive
Negative	False Negative	True Negative

Type	Positive	Negative
Positive	70	2
Negative	2	26

1. Accuracy $= \frac{Tp+Tn}{T_P+Tn+Fp+Fn}$

$= \frac{70+26}{70+26+2+2}$

$= 96\%$

2. Sensitivity $= \frac{Tp}{T_P+Fp}$

$= \frac{70}{70+2}$

$= 97.22\%$

3. Specificity $= \frac{Tn}{Tn+Fp}$

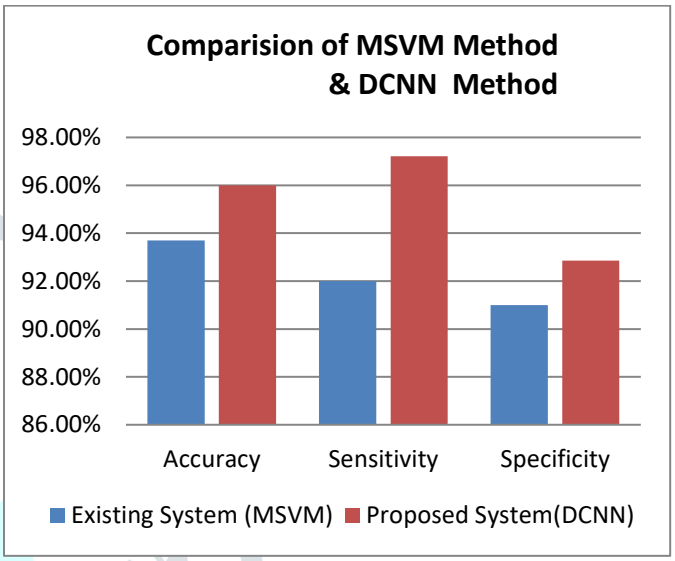
$= \frac{26}{26+2}$

$= 92.85\%$

COMPARISION TABLE

Table I: Comparison of Existing and Proposed Methods

Parameter	Existing System (MSVM)	Proposed System (DCNN)
Accuracy	93.7%	96%
Sensitivity	92%	97.22%
Specificity	91%	92.85%



6. CONCLUSION

In this work, investigated an application of DCNNs in whichpre-trained Alex Net is deployed for the detection and classification of ALL and its subtypes. By performing,can achieve 96% accuracy for leukemia detection for its subtype's classification. This automated diagnosing system can help in early diagnosing of leukemia so that it can be treated effectively. This work has been implemented in MATLAB 2013a.

Future Scope

In future, one of the promising directions for researchers is to deploy different deep learning architectures for classification and detection of leukemia and compare these architectures to check which network performs well for the diagnosing of ALL. Also, can deploy deep learning models to learn from scratch with larger image data sets so that this diagnostic system can be used in everyday life and help the pathologist and oncologist to diagnose the leukemia in better way. This approach can also be further improved to a fully automated system by defining its input output parameters and can consolidate it as a part of sub module to a fully automated system. Also, another future direction for

researchers is to develop an automated detection system for AML so that all different types of blood cancer can be automated.

REFERENCES

[1] Kumar, S.; Mishra, S.; Asthana, P. Automated detection of acute leukemia using k-mean clustering algorithm. In Advances in Computer and Computational Sciences; Springer: Berlin/Heidelberg, Germany, 2018; pp. 655–670.
[2] Classification of Blasts in Acute Leukemia Blood samples Using k-Nearest Neighbour—IEEE Conference Publication. Available online: https://ieeexplore.ieee.org/abstract/document/6194769/ (accessed on 3 February 2020).

[3] Madhukar, M.; Agaian, S.; Chronopoulos, A.T. Deterministic model for acute myelogenous leukemia classification. In Proceedings of the 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Seoul, Korea, 14–17 October 2012; pp. 433–438.

[4] Setiawan, A.; Harjoko, A.; Ratnaningsih, T.; Suryani, E.; Palgunadi, S. Classification of cell types in Acute Myeloid Leukemia (AML) of M4, M5 and M7 subtypes with support vector machine classifier. In Proceedings of the 2018 International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 6–7 March 2018; pp. 45–49.

[5] Faivdullah, L.; Azahar, F.; Htike, Z.Z.; Naing, W.N. Leukemia detection from blood smears. J. Med. Bioeng. 2015, 4, 488–491. [CrossRef]

[6]. Laosai, J.; Chamnongthai, K. Acute leukemia classification by using SVM and K-Means clustering. In Proceedings of the 2014 IEEE International Electrical Engineering Congress (iEECON), Chonburi, Thailand, 19–21 March 2014; pp. 1–4.

[7] Patel, N.; Mishra, A. Automated leukaemia detection using microscopic images. Procedia Comput. Sci. 2015, 58, 635–642. [CrossRef]

[8] Sajjad, M.; Khan, S.; Jan, Z.; Muhammad, K.; Moon, H.; Kwak, J.T.; Rho, S.; Baik, S.W.; Mehmood, I. Leukocytes classification and segmentation in microscopic blood smear: A resource-aware healthcare service in smart cities. IEEE Access 2016, 5, 3475–3489. [CrossRef]

[9] Abdeldaim, A.M.; Sahlol, A.T.; Elhoseny, M.; Hassanien, A.E. Computer-aided acute lymphoblastic leukemia diagnosis system based on image analysis. In Advances in Soft Computing and Machine Learning in Image Processing; Springer: Berlin/Heidelberg, Germany, 2018; pp. 131–147.

[10] Dwivedi, A.K. Artificial neural network model for effective cancer classification using microarray gene expression data. Neural Comput. Appl. 2018, 29, 1545–1554. [CrossRef].

[11] Sahlol, A.T.; Abdeldaim, A.M.; Hassanien, A.E. Automatic acute lymphoblastic leukemia classification model using social spider optimization algorithm. Soft Comput. 2019, 23, 6345–6360. [CrossRef]

[12] F. Scotti, "Robust Segmentation and Measurements Techniques of White Cells in Blood Microscope Images", in Proc. of the 2006 IEEE Instrumentation and Measurement Technology Conf. (IMTC 2006), Sorrento, Italy, pp. 43-48, April 24-27, 2006. ISSN: 1091-5281. [DOI:10.1109/IMTC.2006.328170]