Automatic Detection and Classification of Acute Lymphoblastic Leukemia Using Convolution Neural Network

Arbab Muhammad Qasim[*], Palwasha zeb[†], Muhammad Qasim Khan[‡], Haider Ali[§]

Abstract

Acute lymphoblastic leukaemia (ALL) incited death has been recorded in the ten most perilous mortality reason for humans. The root cause of this is the slow ALL detection. Detection of ALL on time can save precious lives. Automatic detection and classification of ALL in blood smear images can improve the accuracy of ALL detection and speed up the clinical decisions of haematologist's and medical experts. It is difficult and time-consuming to detect and differentiate acute lymphoblastic leukaemia in blood smear images. As a result, both patients and medical professionals should have access to high-quality clinical decision support for acute lymphoblastic leukaemia. In this article, we developed a novel method for classifying and detecting ALL using support vector machine (SVM) and convolution neural network (CNN). In the proposed study, CNN features are initially retrieved using the Alex-Net Model after lymphocyte detection. The detected cell is then divided into two categories using SVM: normal and malignant. To demonstrate the technique's importance and effectiveness, it is compared with state of the art methods.

Keywords: Software engineering, CNN, ALL, Medical Image Analysis, Image Classification, SVM.

Introduction

Acute leukaemia develops when stem cells from the bone marrow have DNA flaws that allow them to overgrow into lymphoblast's, which then flood the body and cause a vast variety of distinct symptoms, including: bruises, nosebleeds, bleeding gums, fever, enlarged lymph nodes, joint discomfort, weakness, and infections (Brownlee, 2016). Intense leukaemia is the most well-known youth malignancy (Sonka, Hlavac, & Boyle, 2014). Acute leukaemia is further categorized into two subtypes i.e. Acute Lymphoblastic Leukaemia (ALL) and Acute Myeloid Leukaemia (AML), and is a kind of blood cancer that strikes kids between the age of 2-5 years. The lymphocyte which is affected is called

[*]Department of Computer science, Iqra National University, Peshawar, Pakistan
qasimarbab@gmail.com
[†]Department of Computer science, Iqra National University, Peshawar, Pakistan
Palwashazeb317@gmail.com
[‡]Department of Computer science, Iqra National University, Peshawar, Pakistan
mqsimkhan21@hotmail.com
[§]Department of Computer science, Islamia College Peshawar, Peshawar, Pakistan
haider6233icp@gmail.com

lymphoblast cell. The consequences of this disease can be quite serious and may cause death if not treated on time (Agaian, Madhukar, & Chronopoulos, 2018). A blood sample is very important in the medical field. Human blood cells are mainly divided into three classes, such as leucocytes, erythrocytes, and platelets. In addition, each of the cells of human blood belongs to different categories. With advent of computers and machine vision techniques, most blood-related diseases can be classified by analysing blood smear images, such as leukaemia, acute leukaemia, acute lymphoblastic leukaemia, AIDS, malaria, anima, etc. (Panetta, Sparreboom, Pui, Relling, & Evans, 2010). Medical imaging allows the medical specialists and haematologists to identify various kinds of blood cells, bone cells and soft tissue from medical imaging (Anilkumar, Manoj, & Sagi, 2022). There are two methods to diagnose ALL first one is manual and the other is automatic detection. The manual is slow and time consuming and also the haematology equipment's use to identify All is costly and may not available in all hospitals (Das & Meher, 2021).

This study aims to detect and classify ALL even in presence of poor quality or low magnification images. The purpose of this research is to deal with the challenges of ALL diagnosis outlined in the previous section by developing an automatic ALL diagnosis process. The proposed method uses blood smear images to detect ALL with accuracy and efficiency of ALL detection even in presence of poor quality or low magnification images. To develop a computerized system that can identify ALL in blood smear images Human blood cells have a few likenesses of their shapes, estimate and morphological features, because of these auxiliary similarities it is hazardous for the haematologists to distinguish and group ALL cell in blood smear images manually. ALL detection, and the detected cells should be arranged into two subtypes (i.e. normal and abnormal). The aim for detecting and classifying ALL is that the doctors diagnose ALL diseases by analysing Lymphocyte.

In clinical haematology, for the most part the haematologist utilizes manual strategies to distinguish and characterize leucocytes (WBCs) utilizing regular magnifying lens. Because of structure resemblances in WBCs subclasses, it is very difficult for the haematologist to recognize. The haematologist performs manual examination subject to various parameters to recognize and arrange WBCs, which is monotonous (Alagu, 2021). Manual WBCs identification also impact the health of the haematologist, for example, fatigue and effecting the eyes and back. There are a few programmed and self-loader approaches that were utilized to recognize WBCs (for example manual thresholding, dynamic shape wind,

diagram cuts etc.) and characterize into its subclass utilizing supervise learning methods (for example KNN, Linear classifiers, ANN, SVM etc.) (Sarker, 2021).

The primary theme of this work is to identify, and characterize ALL in human blood smear images. Our proposed framework comprises of three stages: 1) lymphocyte localization, 2) Feature Extraction, 3) classification. The primary contribution of this work is to show a novel Lymphocyte recognition and characterization system in blood smear images utilizing Deep Neural Network. The contributions are listed below.

1) A novel computer assisted framework is designed for identification and classification of ALL in blood smear images, helping hematic pathologists in diagnosing ALL more capably and with better precision.
2) Provides citizens with convenient access to high-quality healthcare at any time and from any location.
3) To keep an eye on anomalous discoveries and make quick judgments in order to save lives.
4) To create an accurate model for recognising and classifying immature lymphocytes, as well as computing the most important morphological traits, which will serve as the foundation for future computer-assisted leukaemia detection research.
5) To enhance the volume of images in the dataset, the data augmentation technique was used.

Literature study

In this paper the author proposes to extract the granularity features and then Support Vector Machine SVM are applied to detect WBC and then classification of Eosinophil and Basophil from WBC and then use CNN features are extract and then compared both and the proposed method has better accuracy than other state-of-the-art methods (Macawile, Quiñones, Ballado, Cruz, & Caya, 2018).

In paper (Thanh, Vununu, Atoev, Lee, & Kwon, 2018), The author proposes a CNN based method to detect Acute myeloid leukaemia (AML) is the kind of cancer that attacks on both children and adults. CNN is also used for plant disease detection (Hussain, Ahmad, Mughal, & Ali, 2018) and text classification (Ali et al., 2022). In paper (Palomo et al., 2016)The author extract granularity features with using SVM to detect WBC and then use CNN features and compared both results with each other, the author claimed that the proposed method has better accuracy. In this Paper (Mohapatra & Patra, 2010) ,The author proposed a new Technique that uses the CNN classifier to perform the ALL classification. Experimental

results demonstrate excellent performance in the differentiation of normal and abnormal cells. In this article (Mehranfar et al., 2017), The author suggests k-state grouping and histogram alignment as filtering approaches. In paper (Cabrera, Legaspi, Papa, Samonte, & Acula, 2017), The author developed a system that use various image processing techniques to detect leukaemia. Naïve Bayes classifier is used and the method ensures 90.33% accuracy. This paper (Hazlyna et al., 2018), the author proposed algorithm for leukaemia detection. The proposed method yields 98.26% accuracy. In Paper (Macawile et al., 2018) author proposed a method, to segment various types of WBCs from microscopic blood images for counting to generates more accurate results. In this paper (Gouda & Chandraprakash, 2019), The author proposes Jaya algorithm to differentiate between defective and non-defective cells. Nucleus images used to extract different features. The results show that accuracy is improved after optimizing with Jaya algorithm.

In paper (Agaian, Madhukar, & Chronopoulos, 2014) the author uses step by step process including pre-processing and nuclei segmentation and then extract features. the proposed technique managed to obtain accuracy of 98 percent. The author used k means clustering and a support vector machine to get a 98 percent accuracy(Goutam & Sailaja, 2015). In this article(Greer et al., 2013), the author designed the automated system to generalization, gridding, and extracting image data to detect cancer cases. The results show that the proposed system is able to find out the cancer database as accurate as 95.45%.In this paper (Amin, Kermani, Talebi, & Oghli, 2015) the author proposes k-means algorithm for segmentation and SVM is used to classify the cells into cancerous and non-cancerous. This study (Mohamed et al., 2018) presents an application that helps in diagnosing haematic diseases in early stages, for final decision random forest classifier is used. The proposed system achieves an average accuracy of 94%.

The author of this study developed an automated approach to analyse data acquired from micro-images in order to identify cancer patients. Photography, data minerals, and diagnostics are the three key components of the suggested framework. Photo generalisation, gridding, and image data extraction are all operations performed at the image analysis level. The production of information and the selection of genes are both covered by information management. The findings suggest that the proposed method can locate the cancer database with a 95.45% accuracy rate (Baghel & Sunkaria, 2018).The author proposed an automated technique to detecting leukaemia early. Currently, professionals manually examine microscopic images, which is a time-consuming

process that reduces accuracy. In this paper, the author presents some filtering approaches for the SVM classification algorithm, including k-state grouping, histogram alignment, and the zack. The proposed system has been successfully assembled in MATLAB, with a precision of 93.57 percent (Schmidhuber, 2015) .

Methodology

The purpose of this research is to process and analyse images in order to develop an automatic method to support medical research that will allow pathologists to segment and identify acute lymphocytic/lymphoblastic leukaemia (ALL). The research technique depends on computer vision techniques to analyse blood smear images for ALL detection and classification. To detect and classify ALL, we first locate lymphocytes using a blood smear image, and then compute a set of criteria that leads to enhanced reparability of classes using classifiers. Blood smear images make it simple to recognize lymphocytes because their nuclei are complete change from the background and other blood cells. Due of variances in their cell shapes, sizes, and edges, it is difficult to extract features from lymphocytes.

Lymphocytes have a regular form and a compact nucleus, as well as regular and continuous borders. Aside from that, lymphoblast's have uneven shapes and tiny nooks/holes in their cytoplasm. As a result, the suggested method identifies lymphocytes from a blood smear image, and then uses CNN Features to classify the image as normal or lymphoblastic. Numerous strategies have been developed for detecting and classifying white blood cells from human blood smear images, but each methodology has its own set of advantages and disadvantages. For this study, blood smear images were acquired from local hospitals. Figure 1 shows the suggested method for detecting and classifying All, and involve the following 2 phases: training phase, testing phase and shown in Figure 2.

The model is initially trained using blood smear images, and then tested in the testing phase with a blood smear image for ALL detection and classification. The planned method for detection and classification of ALL further involve the following steps; 1) Lymphocyte localization, 2) Feature extraction, 3) Classification of the blood cell into ALL and non-ALL cell. In the proposed work, Convolution Neural Network (CNN) method is use to localize ALL within microscopic blood smear images. The suggested work starts with the detection of lymphocytes, followed by the extraction of CNN characteristics using the Alex Net Model, and finally the classification of the identified cell into two groups, healthy and cancerous, using Support Vector Machine (SVM).

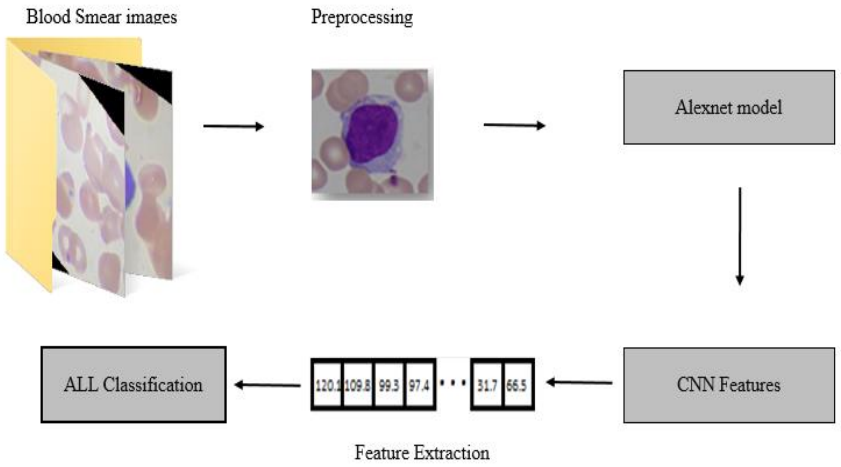


Figure 1: Proposed Framework

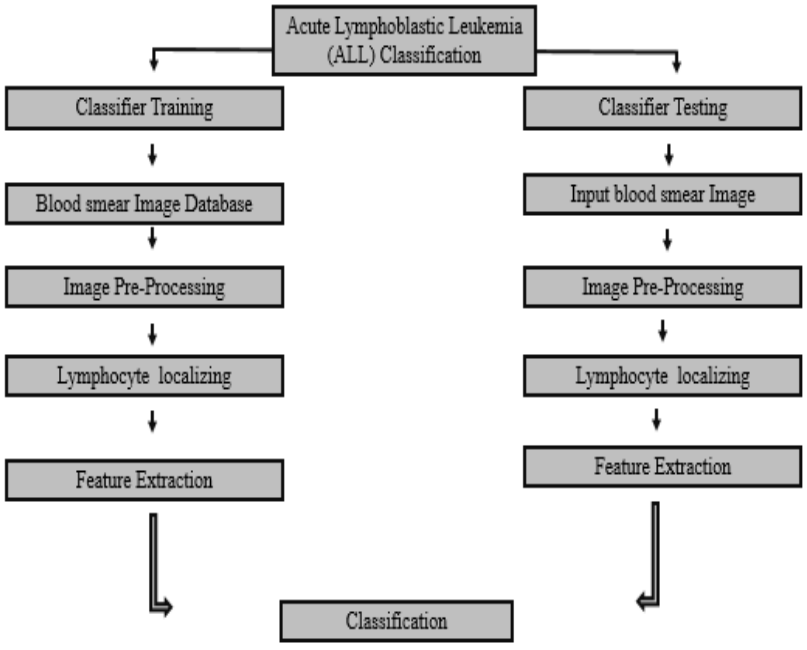


Figure 2: Training and Testing

Lymphocyte Localization

Our work is focused on WBC to locate ALL in blood smear images since leukaemia can only be recognised by analysing the WBC.Therefore, lymphocytes are detected and then classified ALL using CNN (Alex Net model) as shown in Figure 3.

Alex Net Architecture

We proposed using a CNN architecture called Alex Net [16] to extract CNN features, as shown in Figure 4. Alex Net uses specific operations like convolution, pooling, and regularisation to extract deep features from image data. Alex Net architecture consists of different layers including one input layer , convolution layers, Relu layers and fully connected layers (Alom et al., 2018).

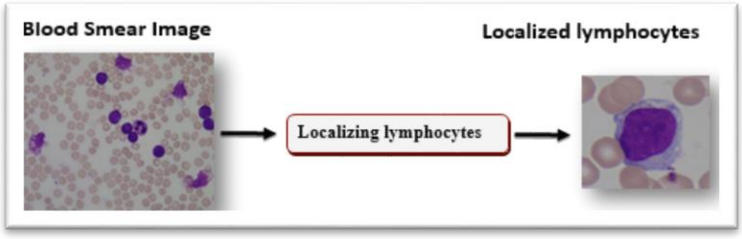


Figure 3: Lymphocyte localization

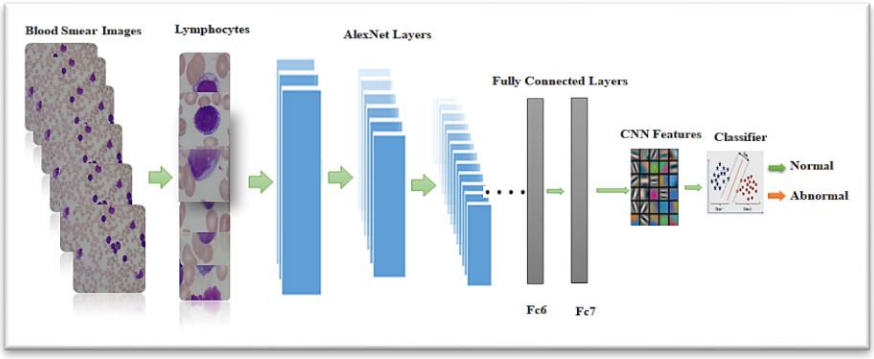


Figure 4: General overview of Alex net Architecture

Feature Extraction

Following lymphocyte localization, the next step is to extract a feature, which is one of the most crucial phases in ensuring classification accuracy. Images' characteristics show their inherent similarity. The classifier then compares various images and groups them into distinct classes using

functions and their labels. Utilizing CNN characteristics, discriminatory functions are retrieved. Feature extraction is shown in Figure 5.

Classification

After features extraction, the next and final phase is classification. During learning, most classifiers follow the same approach: a division is drawn through the data set and used to classify classes. A decision boundary is the name given to the division. To classify healthy and cancerous cells in the proposed framework, we employ an SVM classifier. Figure 4.6 shows classification.

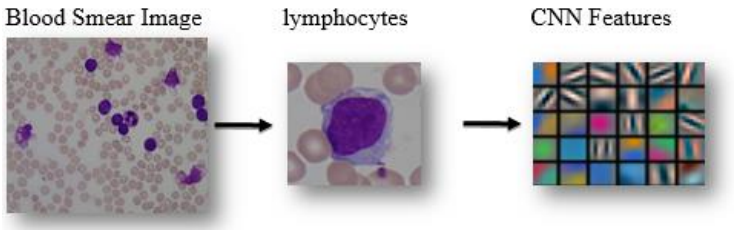


Figure 5: Feature extraction from blood smear images

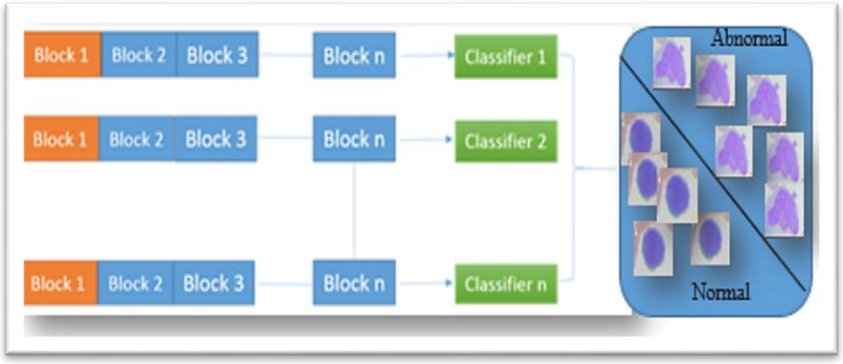


Figure 6: ALL Classification

First, the Alex-Net Model is used to train the classifier using the retrieved CNN features. The classifier then determines which cells are normal and which are malignant. The classifier will identify whether the input blood smear image is healthy or malignant by comparing the extracted cell features to stored features of normal and abnormal cells. Images of blood smears obtained from the nearby hospital. MATLAB is used to implement the planned system. We used a deep convolutional

neural network in our research Using an SVM classifier, classify human blood cells into subnormal and malignant groups. Table 5.2 shows the experimental results, which show 98 % of accuracy.

Dataset

Four thousand blood smear lymphocyte samples were collected from the Hayatabad Medical Complex Peshawar, Pakistan, for this study. The dataset contains 4000 images, with 2000 normal lymphocytes and 2000 abnormal cells. Each image was saved as a JPG file with a resolution of 960 x 1080 pixels. During processing in MATLB, the images' dimensions were changed to 227 x 227 pixels. The details of dataset are given in Table1.

Table 1

Dataset of ALL

S.No	Human Blood Cell	No of Images
1	Normal Lymphocytes	2000
2	Abnormal Lymphocytes	2000
Total		4000

In first step we localize lymphocytes from blood smear images then use Alex Net model for extracting CNN features. SVM classifier was also used to differentiate human blood cells into subnormal and pathological categories. Table 2 shows the experimental findings, which reveal that the accuracy is 98 %. The aggregate accuracy and loss score, which are generated on the test data set, can be used to determine the model's performance. The analysis of the proposed Technique was done using the following, equations:

$$\mathrm{FPR} = \left(\frac{\mathrm{FP}}{\mathrm{TN}+\mathrm{FP}}\right) \quad (1)$$

$$\mathrm{FNR} = \left(\frac{\mathrm{FN}}{\mathrm{TP}+\mathrm{FN}}\right) \quad (2)$$

$$\mathrm{F-Measure} = 2 * (\mathrm{FPR} * \mathrm{FNR})/(\mathrm{FPR} * \mathrm{FNR}) \quad (3)$$

$$\mathrm{Accurcy} = \left(\frac{\mathrm{TP}}{\mathrm{Total\ test\ Images}}\right) * 100 \quad (4)$$

Where true positive are data sets that are accurately categorized as relating to the intended category. False positive are data sets that are wrongly identified as relating to the intended category. True negative are data sets that are correctly classified as not relating to the intended category. False negative are data sets that are wrongly categorized as not

relating to the intended category. The analysis of the ALL classification technique is given in Table No 2.

Table 2
Performance parameters of the proposed study

Metrics	SVM	ANN	Proposed Technique
FPR	0.006	0.003	0.001
FNR	0.094	0.094	0.098
F-Measure	0.094	0.096	0.098
Accuracy	94.3	96.4	98.9

It is often shown that F-measure of the proposed method for 4000 microscopic images of effected and unaffected ALL classification in blood smear samples have 98% Accuracy that shows better classification results. For this purpose, 2 parameters were observed as follows: shape and size. Figure 7 demonstrates the visual classification results of the planned technique. The performance of the planned technique is compared with two recent approaches. It may be seen that the planned technique achieves a high overall accuracy for ALL Classification. Figure No. 8 show that the technique achieves high accuracy.

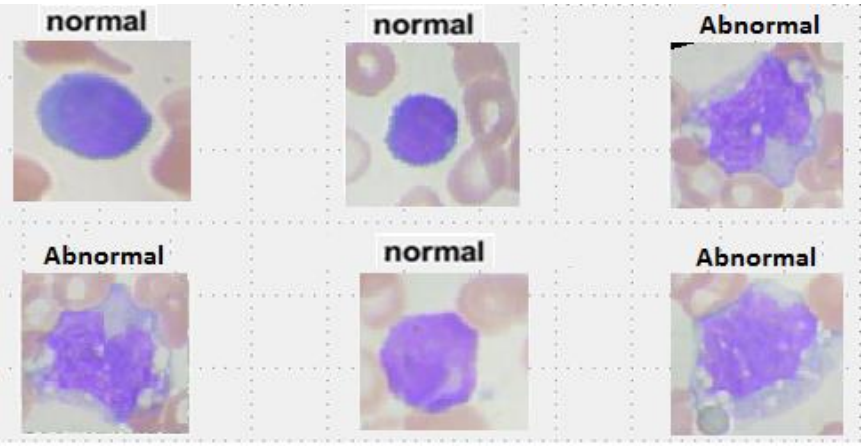


Figure 7: Visual Classification

Conclusion

The goal of this study is to develop an automated system that can detect ALL in blood smear images with better accuracy. The proposed method has the potential to save many lives. The work's major focus is on locating lymphocytes, extracting CNN characteristics, and classifying them. Firstly, a sub class of WBCs i.e. lymphocytes is localized in blood smear

image. The lymphocyte image is then used to extract a series of CNN features.

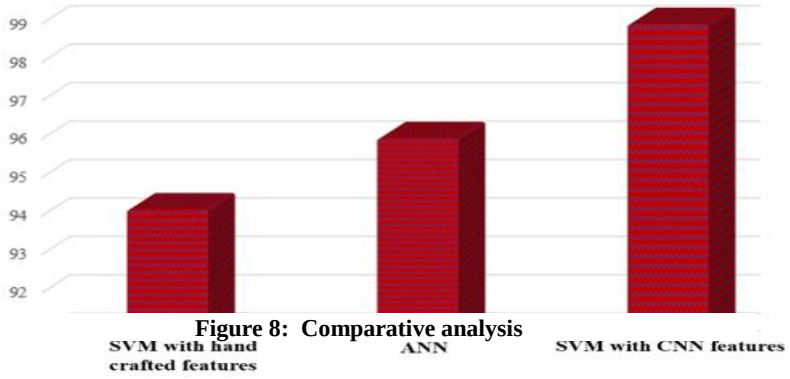
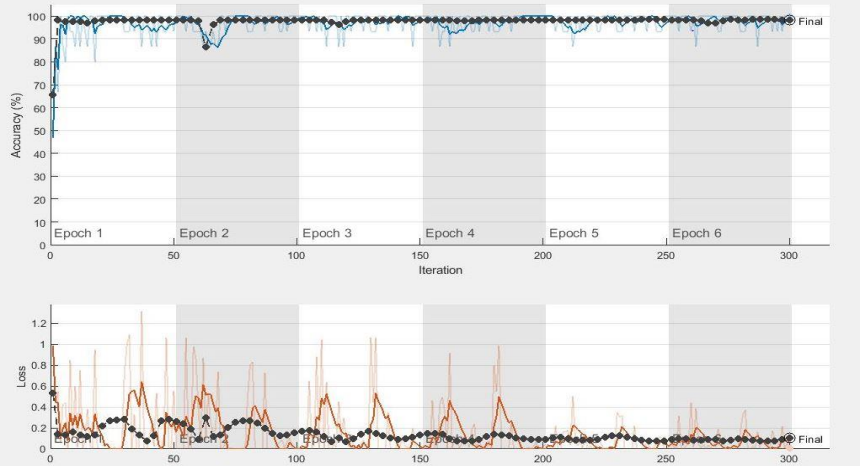


Figure 8: Comparative analysis

Figure 9: Final plot of Alex net with accuracy and loss



The SVM classifier is trained for classification based on the retrieved features. We observed through experimental results that the recommended technique leads other state-of-the art classification methods in terms of classification performance. The proposed approach has a 98 % accuracy rate. As a result, the proposed method is useful in detecting ALL early on without the necessity for a costly investigation.

In future work we can detect and classify Acute Myeloid leukaemia(AML) From blood smear images. Acute Myeloid leukaemia is the most common type of blood cancer in adults in which monocytes cells

are effected. Different deep learning techniques can be utilized to extract features and classify Acute Myeloid leukaemia.

References

Agaian, S., Madhukar, M., & Chronopoulos, A. T. (2014). Automated screening system for acute myelogenous leukemia detection in blood microscopic images. *IEEE Systems journal, 8*(3), 995-1004.

Agaian, S., Madhukar, M., & Chronopoulos, A. T. (2018). A new acute leukaemia-automated classification system. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 6*(3), 303-314.

Alagu, S. (2021). Automatic Detection of Acute Lymphoblastic Leukemia Using UNET Based Segmentation and Statistical Analysis of Fused Deep Features. *Applied Artificial Intelligence, 35*(15), 1952-1969.

Ali, H., Farman, H., Yar, H., Khan, Z., Habib, S., & Ammar, A. (2022). Deep learning-based election results prediction using Twitter activity. *Soft Computing, 26*(16), 7535-7543.

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Hasan, M., Van Esesn, B. C., . . . Asari, V. K. (2018). The history began from alexnet: A comprehensive survey on deep learning approaches. *arXiv preprint arXiv:1803.01164*.

Amin, M. M., Kermani, S., Talebi, A., & Oghli, M. G. (2015). Recognition of acute lymphoblastic leukemia cells in microscopic images using k-means clustering and support vector machine classifier. *Journal of medical signals and sensors, 5*(1), 49.

Anilkumar, K., Manoj, V., & Sagi, T. (2022). Automated detection of b cell and t cell acute lymphoblastic leukaemia using deep learning. *Irbm, 43*(5), 405-413.

Baghel, L. K., & Sunkaria, R. K. (2018). *High density fixed valued impulse noise removal using improved decision based hybrid median filter and its application on medical images.* Paper presented at the 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT).

Brownlee, J. (2016). *Deep learning with Python: develop deep learning models on Theano and TensorFlow using Keras*: Machine Learning Mastery.

Cabrera, R. J. A., Legaspi, C. A. P., Papa, E. J. G., Samonte, R. D., & Acula, D. D. (2017). *HeMatic: An automated leukemia detector with separation of overlapping blood cells through Image*

Processing and Genetic Algorithm. Paper presented at the 2017 International Conference on Applied System Innovation (ICASI).

Das, P. K., & Meher, S. (2021). An efficient deep convolutional neural network based detection and classification of acute lymphoblastic leukemia. *Expert Systems with Applications*, *183*, 115311.

Gouda, R., & Chandraprakash, V. (2019). Optimization Driven Constraints Handling in Combinatorial Interaction Testing. *International Journal of Open Source Software and Processes (IJOSSP), 10*(3), 19-37.

Goutam, D., & Sailaja, S. (2015). *Classification of acute myelogenous leukemia in blood microscopic images using supervised classifier*. Paper presented at the 2015 IEEE International Conference on Engineering and Technology (ICETECH).

Greer, J. P., Arber, D. A., Glader, B., List, A. F., Means, R. T., Paraskevas, F., & Rodgers, G. M. (2013). *Wintrobe's clinical hematology*: Lippincott Williams & Wilkins.

Hazlyna, H., Aini, H., Fadzilah, S., Nadia, K., Mashor, M., & Hassan, R. (2018). Fusion 7× 7 median filter and seeded region growing area extraction algorithms for effective detection of acute leukemia based on blood images. *Journal of Fundamental and Applied Sciences, 10*(3S).

Hussain, A., Ahmad, M., Mughal, I. A., & Ali, H. (2018). *Automatic disease detection in wheat crop using convolution neural network.* Paper presented at the The 4th International Conference on Next Generation Computing.

Macawile, M. J., Quiñones, V. V., Ballado, A., Cruz, J. D., & Caya, M. V. (2018). *White blood cell classification and counting using convolutional neural network.* Paper presented at the 2018 3rd International conference on control and robotics engineering (ICCRE).

Mehranfar, S., Zeinali, S., Hosseini, R., Mohammadian, M., Akbarzadeh, A., & Feizi, A. H. P. (2017). History of leukemia: diagnosis and treatment from beginning to now. *Galen Medical Journal*, *6*(1), 12-22.

Mohamed, H., Omar, R., Saeed, N., Essam, A., Ayman, N., Mohiy, T., & AbdelRaouf, A. (2018). *Automated detection of white blood cells cancer diseases.* Paper presented at the 2018 First International Workshop on Deep and Representation Learning (IWDRL).

Mohapatra, S., & Patra, D. (2010). *Automated cell nucleus segmentation and acute leukemia detection in blood microscopic images.* Paper

presented at the 2010 International Conference on Systems in Medicine and Biology.

Palomo, L., Garcia, O., Arnan, M., Xicoy, B., Fuster, F., Cabezón, M., . . . Jiménez, M.-J. (2016). Targeted deep sequencing improves outcome stratification in chronic myelomonocytic leukemia with low risk cytogenetic features. *Oncotarget, 7*(35), 57021.

Panetta, J. C., Sparreboom, A., Pui, C.-H., Relling, M. V., & Evans, W. E. (2010). Modeling mechanisms of in vivo variability in methotrexate accumulation and folate pathway inhibition in acute lymphoblastic leukemia cells. *PLoS computational biology, 6*(12), e1001019.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, *2*(3), 1-21.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks, 61*, 85-117.

Sonka, M., Hlavac, V., & Boyle, R. (2014). *Image processing, analysis, and machine vision*: Cengage Learning.

Thanh, T., Vununu, C., Atoev, S., Lee, S.-H., & Kwon, K.-R. (2018). Leukemia blood cell image classification using convolutional neural network. *International Journal of Computer Theory and Engineering, 10*(2), 54-58.