Classification of B-Lymphoblasts against Normal Lymphocytes by Deep Convolutional Neural Network

Anilumar K. K.[1], Manoj V. J.[2], Sagi T. M.[3]

Cochin University of Science and Technology, Kuttanad campus, Alappuzha, Kerala, India[1,2]

St. Thomas College of Allied Health Sciences, Changanasserry, Kottayam, Kerala, India[3]

Corresponding author: Anilumar K. K., Email: kkanil@cusat.ac.in

Leukemia is a life-threatening ailment and is a non-tumor type of cancer which results in large number of abnormal white blood cells known as blasts. Image processing and machine learning techniques can be used to help the pathologists to detect and classify these cancer cells. The present work tries to classify the blast cells of acute leukemia (B-Lymphoblasts) against healthy lymphocytes by a custom made Deep Convolutional Neural Network (DCNN) developed by the authors and named as LeukNet using an online dataset C-NMC 2019. The LeukNet is able to classify the cancer cells against the healthy cells with an Accuracy of 86.73% without much complicated segmentation and feature extraction. Even though the accuracy is slightly less compared to the related works on the same dataset, the classification model used in the study is comparatively simple.

Keywords: Leukemia, Image processing, Deep learning, Convolutional neural network

2023. In Saroj Hiranwal & Garima Mathur (eds.), *Artificial Intelligence and Communication Technologies*, 23–28. Computing & Intelligent Systems, SCRS, India. https://doi.org/10.52458/978-81-955020-5-9-2

1 Introduction

Leukemia is usually diagnosed by the physicians from the signs and symptoms of the patients and with the help of pathologists for laboratory tests. The pathologist generally carries out a complete blood count (CBC) test of the suspected patient and may go for blood and bone marrow examination of the smear [1,2]. But these laboratory tests are error-prone and heavily depend upon the expertise of the clinical technicians and pathologists. These methods also suffer from the intra and inter-observer variabilities. Image processing methods can assist pathologists in a quick and easy diagnosis of leukemia by processing and analyzing the images of blood and bone marrow smears [3,4]. It also has an advantage that the expert can analyze the images from a remote location as the images can be easily and quickly sent over the internet [3,4,5].

Machine Learning (ML) and Deep Learning (DL) are subsets of Artificial Intelligence and they have been used widely for the Computer Aided Diagnosis (CAD) of many diseases including leukemia, for the last several years. The DCNNs have the advantage of automatically identifying and capturing the features of the images for classification compared to conventional ML techniques which depend upon user-crafted feature extraction [5]. The DL based techniques are already used by several researchers for the computer-aided detection and classification of different types of leukemia [3-12]. Rastogi et al. [3] reported the classification of leukocytes and leukemic cells with a CNN feature extractor "LeuFeatx" and they got 96.15% accuracy in classifying ALL cells against healthy cells. The studies [4-12] also reported the detection and classification of acute leukemia with CNNs using some datasets of leukemia other than the C-NMC dataset. Shubham et al. [13] tried the detection of B-cell Acute Lymphoblastic Leukemia (ALL) with the C-NMC dataset used by the present study, adopting heterogeneity loss and ensemble strategy and obtained a weighted $F1$ score of 95.26%. Oliveira et al. [14] attempted the classification of leukemic cells against normal cells on the C-NMC dataset with VGG16, VGG19 and Xception networks and achieved an $F1$ score of 92.6%. Bukhari et al. [15] carried out ALL detection using squeeze and excitation learning on ALL-IDB dataset and obtained an accuracy of 98.3%. Shafique et al. [16] tried the detection and classification of ALL on the ALL-IDB dataset using a pretrained DCNN AlexNet and they achieved an accuracy of 96.06% for the classification of leukemia. Laurence et al. [17] proposed ALL detection on the C-NMC dataset using Depthwise Separable CNNs and obtained an accuracy of 91% on the validation set. They do not reveal the accuracy obtained on the testing dataset. Zhou et al. [18] tried ALL detection on a dataset of 1732 bone marrow images collected from Shanghai Children's Medical Center using CNNs and they obtained an accuracy of 89%. D. Kumar et al. [19] proposed ALL and Multiple Myeloma detection on SN-AM dataset of bone marrow images using DCNN and obtained an overall accuracy of 97.2%.

2 Methods

The classification of B-Lymphoblasts against healthy lymphocytes is carried out by a custom made DCNN developed on MALAB R2021a platform and the experiments were done on single CPU based computer system having 12GB RAM. We have used a simple classification model without involving any image segmentation and image enhancements as shown in Fig. 1. The images in the dataset in their raw form are given to the LeukNet for classification and training is initiated. The CNN then learns the features of the images automatically and classifies them into Lymphocyte or B-Lymphoblast.

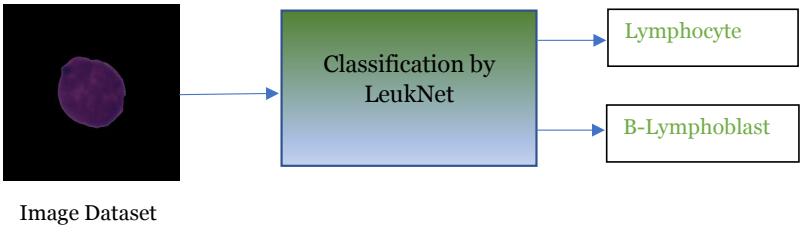

Fig. 1 Classification framework

2.1 Image Dataset

The images used in the study were obtained from the C-NMC challenge dataset of ISBI 2019 which consists of several images of leukemic (ALL) and healthy cells (HEM). We downloaded 2397 images of ALL and 1130 HEM images and since the dataset is unbalanced, we balanced the images in the two class equal to 1130 by randomly deleting 1267 ALL images. Sample images of C-NMC dataset is shown in **Fig. 2**. The images were in the bmp format having a size of 450 × 450 pixels which are then resized to 224 × 224 to suit the image input size of the DCNN, LeukNet using data augmentation by DL methods.

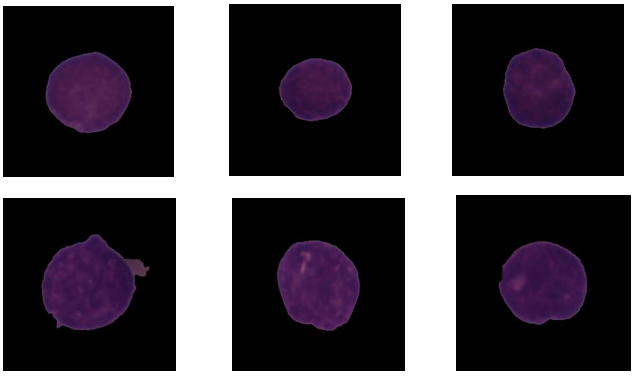

Fig. 2 Sample ALL and HEM images (Row 1: HEM, Row 2: ALL)

2.2 The LeukNet

The classifications in the present study were performed by LeukNet which is designed as a series DCNN having a depth of 5 and 17 layers as shown in **Fig. 3**. The combination of Convolution, ReLU and Max Pooling layers is repeated 3 times and it uses 2 Fully Connected layers also which makes the depth of LeukNet to 5. All the 3 convolution layers use a filter size of 3 × 3 and the number of filters in the first, second and third convolution layers are 64, 128 and 256, respectively. The cross-channel normalization layer performs a channel-wise local normalization, that is, it replaces each element with a normalized value obtained from the elements of a certain number of (5 in this case) neighboring channels. The pool size, which determines the rectangular pooling region, used in all the three max pooling layers is 3 × 3. The Dropout layer sets the probability (0.5 in this case) for dropping out input elements and it helps to prevent the network from overfitting.

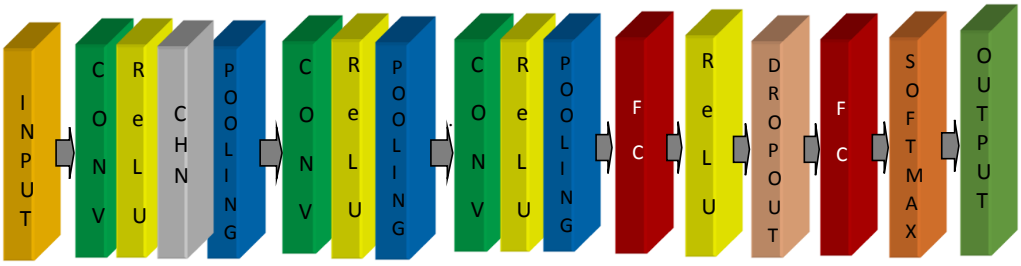

Fig. 3 Layers of LeukNet

3 Results

Training of the DCNN is initiated by splitting the images in the dataset randomly such that 70% images are used for training, 20% for validation and the remaining 10% for testing. The optimization algorithm used in the study is Adaptive Moment estimation (ADAM) [19,20] which adjusts the network parameters to minimize the loss function. We have used an initial learning rate of 0.0001, 10 Epochs,

validation frequency of 5 and learning rate schedule to drop the learning rate by a factor of 0.5 after 7 Epochs. A total training time of 45 minutes and 34 seconds has taken and the classification resulted in a validation accuracy of 84.51%. Testing is then carried out with 10% of the images which were not used in training and validation and obtained an accuracy of 86.73%. **Fig. 4** shows few classified images in the testing phase with the associated probability of falling to that particular class (ALL or HEM). The confusion matrix is then plotted and is shown in **Fig. 5** which revels that 97 out of 113 ALL images (B-lymphoblasts) are correctly classified and 99 out of 113 healthy lymphocytes images (HEM) are also correctly classified.

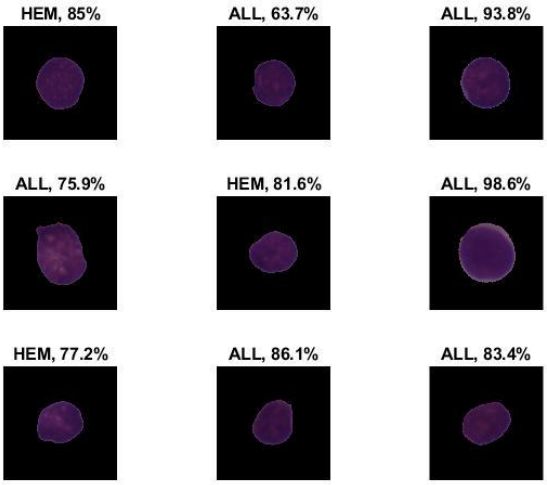

Fig. 4 Few classified images by LeukNet (testing phase)

Fig. 5 Confusion Matrix

4 Discussion

The DL based studies for hematological image analysis witnessed promising growth from the year 2018 onwards and many researchers attempted the classification of leukemic cells against normal cells. The present work tried the classification of ALL cells against HEM (normal) cells with the C-NMC dataset. The accuracy of 86.73% in the testing phase is slightly less compared to the other works [12,13,16] which used the C-NMC dataset but the present study has the advantage of using a simple classification model without the use of user defined feature extraction and image enhancement techniques. The dataset used is also small compared to the other methods which used the entire C-NMC dataset and it can also be attributed to the slight fall in accuracy. Large dataset is avoided in the study since it uses a single CPU system to perform the classification and the training time will be exorbitantly high if the dataset is very large.

5 Conclusion

A deep learning-based classification of B-lymphoblasts against healthy lymphocytes by the DCNN LeukNet using the dataset of C-NMC is detailed in the present work. The classification resulted in 86.73% accuracy in the testing phase which has 113 images in either class (ALL and HEM). The proposed work has the merit of using a simple classification framework compared to the related works which used complex feature extraction and image enhancements. The study considered the detection of ALL with the C-NMC dataset alone and future research can be done by experimenting on other datasets so that classification of leukemia subtypes can be done.

References

[1] John Bernard Henry (1989). Clinical Diagnosis and Management by Laboratory Methods. Philadelphia: W.B. Saunders Company: 17th Edition.

[2] Frank Firkin, Colin Chesterman, David Penington, Bryan Rush (1989). de Gruchy's Clinical Haematology in Medical Practice. London: Blackwell Sciences Ltd., 5th Edn.

[3] Priyanka Rastogi, Kavita Khanna, Vijendra Singh (2022). LeuFeatx: Deep learning–based feature extractor for the diagnosis of acute leukemia from microscopic images of peripheral blood smear, Computers in Biology and Medicine, Vol. 142, 105236.

[4] K. K. Anilkumar, V. J. Manoj & T. M. Sagi (2021). Efficacy of CIELAB and CMYK color spaces in Leukemia Image Analysis: A comparison by statistical techniques. *Biomedical Engineering: Applications basis and comuunications,* Vol 33, No.6, 2150042.

[5] K.K. Anilkumar; V.J. Manoj; T.M. Sagi (2021). Automated detection of leukemia by pretrained deep neural networks and transfer learning: A comparison. *Medical Engineering & Physics*. Volume 98, Pages 8-19. https://doi.org/10.1016/j.medengphy.2021.10.006.

[6] KK Anilkumar, VJ Manoj, TM Sagi (2021). Automated Detection of B Cell and T Cell Acute Lymphoblastic Leukaemia Using Deep Learning. IRBM. https://doi.org/10.1016/j.irbm.2021.05.005

[7] Kirshna Kumar Jha, Himadri Sekhar Dutta (2019). Mutual Information based hybrid model and deepp learning for Acute Lymphoblastic Leukemia detection in single cell blood smear images. *COMPUT METH PROG BIO*. 179:1-12.

[8] Amjad Rahman, Naveed Abbas, Tanzila Saba, Syed Ijaz ur Rahman, Zahid Mehmood, Hoshang Kolivand (2018). Classification of acute lymphoblastic leukemia using deep learning. *MICRO SC RES TECHNIQ*. 81: 1310-1317.

[9] Luis H.S Vogado, Rodrigo M.S. Veras, Flavio H.D. Araujo, Romure R.V. Silva, Kelson R.T. Aires (2018). Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification. *ENG APPL ARTIF INTEL*. 72:415-422. https://doi.org/10.1016/j.engappai.2018.04.024.

[10] A. I. Shahin, Yanhui Guo, K. M. Amin, Amr A. Sharawi (2019). White bloodcell identification system based on convolutional deep neural learning networks. *COMPUT METH PROG BIO*. 168:69-80.

[11] Feiwei Qin, Nannan Gao, Yong Peng, Zizhao Wu, Shuying Shen, Artur Grudtsin (2018). Fine-grained leukocyte classification with deep residual learning for microscopic images. *COMPUT METH PROG BIO*. 162:243-252.

[12] Jyoti Rawat, Annapurna Singh, Bhadauria HS, Jitendra Virmani, Jagtar Singh Devgun (2017). Computer assisted classification framework for prediction of acute lymphoblastic acute myeloblastic leukemia. *BIOCYBERN BIOMED ENG*. 37:637-654.

[13] Shubham Goswami, Suril Mehta, Dhruva Sahrawat, Anubha Gupta, Ritu Gupta (2020). Heterogeneity Loss to handle Intersubject and Intrasubject variability in Cancer. arXiv:2003.03295v2 [cs.CV].

[14] Maurício de Oliveira, J. and Dantas, D (2021). Classification of Normal versus Leukemic Cells with Data Augmentation and Convolutional Neural Networks. DOI: 10.5220/0010257406850692

[15] Maryam Bukhari, Sadaf Yasmin, Saima Sammad, and Ahmed A. Abd El-Latif (2022). A Deep Learning Framework for Leukemia Cancer Detection in Microscopic Blood Samples Using Squeeze and Excitation Learning. Mathematical Problems in Engineering, Article ID 2801227. https://doi.org/10.1155/2022/2801227

[16] Sarmad Shafique and Samabia Tehsin (2018). Acute Lymphoblastic Leukemia Detection and Classification of Its Subtypes Using Pretrained Deep Convolutional Neural Networks. Technology in Cancer Research &Treatment, Volume 17: 1-7.

[17] Clinton, Laurence P. Jr., Somes, Karen M., Chu, Yongjun and Javed, Faizan (2020). Acute Lymphoblastic Leukemia Detection Using Depthwise Separable Convolutional Neural Networks, SMU Data Science Review: Vol. 3: No. 2, Article 4.

[18] Zhou M, Wu K, Yu L, Xu M, Yang J, Shen Q, Liu B, Shi L, Wu S, Dong B, Wang H, Yuan J, Shen S and Zhao L (2021). Development and Evaluation of a Leukemia Diagnosis System Using Deep Learning in Real Clinical Scenarios. Front. Pediatr. 9:693676. doi:.3389/fped.2021.693676

[19] Deepika Kumar, Nikita Jain, Aayush Khurana, Sweta Mittal, Suresh Chandra Satapathy, Roman Senkerik, and Jude D. Hemanth (2020). Automatic Detection of White Blood Cancer From Bone Marrow Microscopic Images Using Convolutional Neural Networks. IEEE Access, Vol. 8.

[20] Y. Lecun; L. Bottou; Y. Bengio; P. Haffner (1998). Gradient-based learning applied to document recognition. Proc. of the IEE. 1998, 86 (11): 2278-2324. https://doi.org/10.1109/5.726791.

[21] Kingma, Diederik, Jimmy Ba (2014). Adam: A method for stochastic optimization, arXiv preprint arXiv. 1412.6980.