



Detection of Cancer in Human Blood Cell using Classification Algorithm

C. Prabavathi¹, Mr. R. Ambikapathy²

¹Master of Computer Application, krishnasamy college of engineering & Technology, Cuddalore

²MCA., M.Phil., Assistant Professor, Master of Computer Application, krishnasamy college of Engineering & Technology, Cuddalore.

ABSTRACT

At the moment, identification of blood disorders is through visual inspection of microscopic images of blood cells. From the identification of blood disorders, it can lead to classification of certain diseases related to blood. This paper describes a preliminary study of developing a detection of leukemia types using microscopic blood sample images. Analyzing through images is very important as from images, diseases can be detected and diagnosed at earlier stage. From there, further actions like controlling, monitoring and prevention of diseases can be done. Images are used as they are cheap and do not require expensive testing and lab equipments. The system will focus on white blood cells disease, leukemia. The system will use features in microscopic images and examine changes on texture, geometry, color and statistical analysis. Changes in these features will be used as a classifier input. A literature review has been done and Reinforcement Learning is proposed to classify types of leukemia. A little discussion about issues involved by researchers also has been prepared.

Keywords: Deep learning, Navigation, Artificial intelligence, CNN.

I. INTRODUCTION

The most significant disease for human deaths is blood cancer. Cancer results in one in six deaths. Leukemia can also be known as blood cancer. The sources of cancer cells in every part of the body are only derived from the white blood cells, but in many cancer cells, the cells expand. Old cells of cancer are destroyed and your room is taken up by new cancer cells. They are producing new cancer cells that are insufficient for living and cause cancer of the blood. The blood cell growth kills other cells in your body. This leads to severe anemia, bleeding, and infection and to the spread of pain or swelling to the lymph nodes and other organs. This made them useful for diagnosing cancer at its respective stages. Imaging techniques can recognize and identify patterns in complex data sets, while effectively predicting future outcomes for blood cancer and the patient's current condition. A public data sample collection for the assessment and evaluation of segmentation and classification algorithms. A number of merit figures are provided in the cell rating for each image within the data set for comparing the performance of different algorithms. The number of blood cells is used to estimate the proportion of the blood cells. [2] Acute and chronic blood cancer is several types of blood cancer variations in it. For the treatment of acute and lymphocytic leukemia we make use of electron microscope images. The methods for enhancing contrast and segmentation are used for detection Lung cancer prediction was made taking the picture as a collection of data. The Bayesian Network and the J8 algors for the prediction of lung cancer equate various classification algorithms such as Naïve Bayes. In comparison to other algorithms, Naïve Bayes performed better. Imaging cancer, including enhancement, clustering, mathematical processing and labeling, can help in the detection of blood cancer on MATLAB. Algorithms such as SVM and KNN have provided reliable results for blood cancer detection and classification. A method for detecting calcifications with the advantages of using anisotropic diffusion has been explored. Hepatocellular carcinoma (HCC), a common type of liver cancer detected by thresholds for specific circulating blood biomarkers, has been used to try to improve the sensitivity and timeliness of detection of various types of cancer. Single layer feed forward in the hidden layer is used for classification of cancer into supervised and unsupervised methods of learning. Unsupervised learning uses fuzzy approaches and methods for non-fuzzy classification. In supervised learning, a single hidden layer of NN power is used with backward propagation training to eliminate.

II. RELATED WORKS

Abdurrazzaq et al. [1] employed a solitary value decomposition procedure. When contrasted to previous methodologies, their findings revealed an enhancement in WBC cores identification and also WBC recognition, especially for WBCs having moderate color strength. This approach has a separation efficiency of 63 percent on general. For generate ROIs for WBC fragmentation, Khomairoh et al. [2] employed the Haar cascade paradigm for WBC characteristic retrieval. The prototype was created with a sample of M4, M5, and M7 AML subgroups, with characteristics retrieved via various convolution kernels such as edges, lines, plus four-rectangle kernels. Following that, for core and mitochondrial delineation, a color-based technique was applied. Core delineation accuracy rate for M4, M5, and M7 were 87.5 percent, 90.4 percent, and 84.6 percent, correspondingly. The program, on the other hand, attained total accuracy rate of 75 percent, 71.4 percent, and 80.76 percent for mitochondrial delineation for M4, M5, and M7,

accordingly. Saleem et al. [3] extracted WBC characteristics for sequential fragmentation and categorization using attributes merging with DarkNet-53 and ShuffleNet, as well as obtained 98.6 percent segmented efficiency. Karolinska Institute [4] Defines the microscope image as being more accurate than the 2-dimensional image in detecting and diagnosing cancer. The author of this journal article discusses new image processing techniques to detect and diagnose cancer cells by transporting tissues and reproducing a three-dimensional picture. Kumar et al [5] have computerized a method for cancer detection to minimizing human involvement. The proposed method was automated image-based acute leukemia detection systems by using incensement and filtering techniques, extract regions using a K-means clustering algorithm. The proposed system has achieved 93% accuracy in the testing phase.

III METHODOLOGY AND IMPLEMENTATION

The proposed a CNN-based framework for Blood cell Analysis towards lymphoblastic leukemia Detection. We made detailed analysis of the proposed deep learning frameworks in several aspects, e.g. the performances of networks with different depths; the influences caused by adding different components. This work provides useful guidelines for the design of deep learning network in related medical research. In this work, CNN is used as a deep learning framework for the automated detection of lymphoblastic leukemia. CNN networks benefit from a range of convolves filters. They will examine different structures in the input images. Thus, when using CNN, the input is the image itself and also the network automatically extracts the suitable aspects of the image. Our CNN model consists of multiple layers.

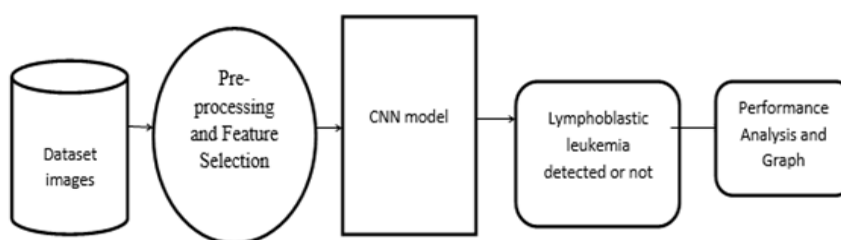


Figure 1 Proposed model

Dataset

The device entry used the "Kaggle Public Data Collection "microscopic image in the database. The results are expected to be good. This offers a solid comparative base and encompasses a wide variety of regions. It also provides a wide variety of details on blood cancer identification, it contains 360 sample data pictures obtained, and 76 data pictures of contaminated cells were taken. And then the image is pre-processed to boost color conversion image quality. The division of the image into pixels breaks the image. The segmented images are then provided to extract features inherent in a blast cell area analysis for their shape, color and texture. The microscopic image of the blood cell is created with a robotic microscope.

<https://www.kaggle.com/datasets/abdulraheem625/acute-lymphoblastic-leukemia>

Importing the necessary libraries

We will be using Python language for this. First we will import the necessary libraries such as keras for building the main model, sklearn for splitting the training and test data, PIL for converting the images into array of numbers and other libraries such as pandas, numpy ,matplotlib and tensor flow.

Retrieving the images

We will retrieve the images and their labels. Then resize the images to (224,224) as all images should have same size for recognition. Then convert the images into numpy array.

Splitting the dataset

Split the dataset into train and test. 80% train data and 20% test data.

Convolutional Neural Networks

The objectives behind the first module of the course 4 are:

- To understand the convolution operation
- To understand the pooling operation
- Remembering the vocabulary used in convolutional neural networks (padding, stride, filter, etc.)
- Building a convolutional neural network for multi-class classification in images

Computer Vision

Some of the computer vision problems which we will be solving in this article are:

1. Image classification
2. Object detection
3. Neural style transfer

One major problem with computer vision problems is that the input data can get really big. Suppose an image is of the size 68 X 68 X 3. The input feature dimension then becomes 12,288. This will be even bigger if we have larger images (say, of size 720 X 720 X 3). Now, if we pass such a big input to a neural network, the number of parameters will swell up to a HUGE number (depending on the number of hidden layers and hidden units). This will result in more computational and memory requirements – not something most of us can deal with

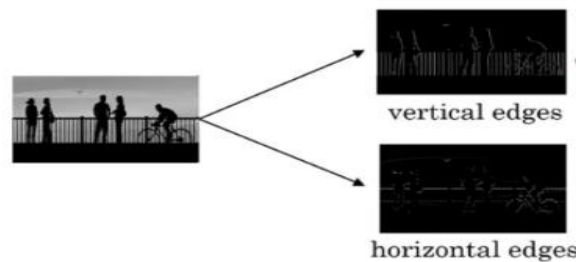
Edge Detection Example

In the previous article, we saw that the early layers of a neural network detect edges from an image. Deeper layers might be able to detect the cause of the objects and even more deeper layers might detect the cause of complete objects (like a person’s face).

In this section, we will focus on how the edges can be detected from an image. Suppose we are given the below image:



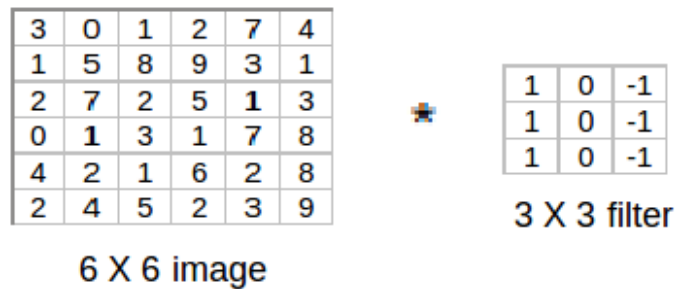
As you can see, there are many vertical and horizontal edges in the image. The first thing to do is to detect these edges:



But how do we detect these edges? To illustrate this, let’s take a 6 X 6 grayscale image (i.e. only one channel):

3	0	1	2	7	4
1	5	8	9	3	1
2	7	2	5	1	3
0	1	3	1	7	8
4	2	1	6	2	8
2	4	5	2	3	9

Next, we convolve this 6 X 6 matrix with a 3 X 3 filter:



After the convolution, we will get a 4 X 4 image. The first element of the 4 X 4 matrix will be calculated as:

3 ¹	0 ⁰	1 ⁻¹
1 ¹	5 ⁰	8 ⁻¹
2 ¹	7 ⁰	2 ⁻¹

So, we take the first 3 X 3 matrix from the 6 X 6 image and multiply it with the filter. Now, the first element of the 4 X 4 output will be the sum of the element-wise product of these values, i.e. $3*1 + 0 + 1*-1 + 1*1 + 5*0 + 8*-1 + 2*1 + 7*0 + 2*-1 = -5$. To calculate the second element of the 4 X 4 output, we will shift our filter one step towards the right and again get the sum of the element-wise product:

0 ¹	1 ⁰	2 ⁻¹
5 ¹	8 ⁰	9 ⁻¹
7 ¹	2 ⁰	5 ⁻¹

Similarly, we will convolve over the entire image and get a 4 X 4 output:

-5	-4	0	8
-10	-2	2	3
0	-2	-4	-7
-3	-2	-3	-16

So, convolving a 6 X 6 input with a 3 X 3 filter gave us an output of 4 X 4. Consider one more example:

Note: Higher pixel values represent the brighter portion of the image and the lower pixel values represent the darker portions. This is how we can detect a vertical edge in an image.

Padding

We have seen that convolving an input of 6 X 6 dimension with a 3 X 3 filter results in 4 X 4 output. We can generalize it and say that if the input is $n \times n$ and the filter size is $f \times f$, then the output size will be $(n-f+1) \times (n-f+1)$:

- **Input:** $n \times n$
- **Filter size:** $f \times f$
- **Output:** $(n-f+1) \times (n-f+1)$

There are primarily two disadvantages here:

1. Every time we apply a convolutional operation, the size of the image shrinks
2. Pixels present in the corner of the image are used only a few number of times during convolution as compared to the central pixels. Hence, we do not focus too much on the corners since that can lead to information loss

To overcome these issues, we can pad the image with an additional border, i.e., we add one pixel all around the edges. This means that the input will be an 8 X 8 matrix (instead of a 6 X 6 matrix). Applying convolution of 3 X 3 on it will result in a 6 X 6 matrix which is the original shape of the image. This is where padding comes to the fore:

- **Input:** $n \times n$
- **Padding:** p
- **Filter size:** $f \times f$
- **Output:** $(n+2p-f+1) \times (n+2p-f+1)$

There are two common choices for padding:

1. **Valid:** It means no padding. If we are using valid padding, the output will be $(n-f+1) \times (n-f+1)$
2. **Same:** Here, we apply padding so that the output size is the same as the input size, i.e., $n+2p-f+1 = n$. So, $p = (f-1)/2$

We now know how to use padded convolution. This way we don't lose a lot of information and the image does not shrink either. Next, we will look at how to implement strided convolutions.

Accuracy on test set

We got a accuracy of 71.3% on test set

Saving the Trained Model

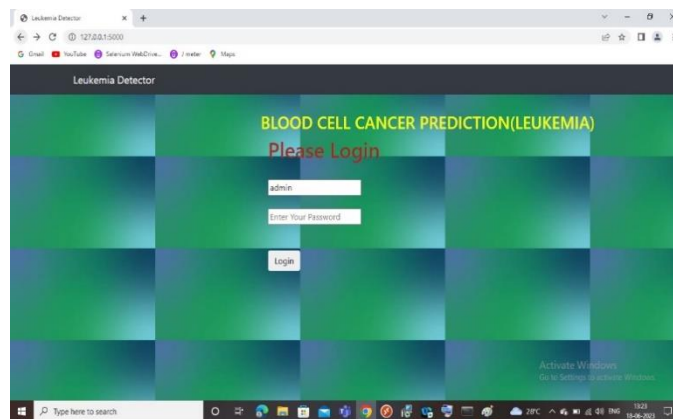
Once you're confident enough to take your trained and tested model into the production-ready environment, the first step is to save it into a .h5 or .pkl file using a library like pickle .

Make sure you have pickle installed in your environment

Next, let's import the module and dump the model into.pkl file

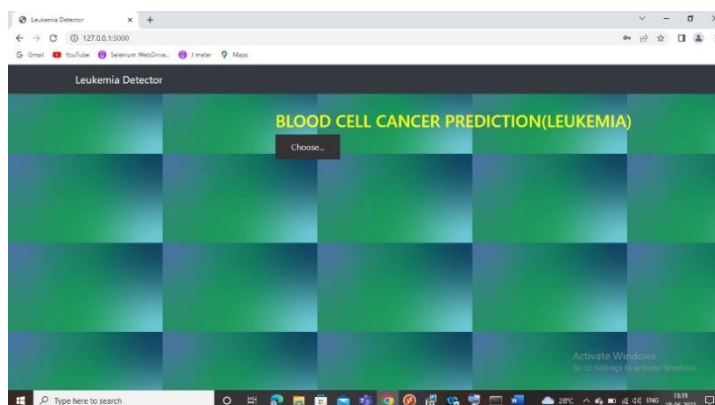
IV. RESULTS AND DISCUSSION

The used methodology of implemented machine learning that uses preprocessing techniques, classification with a CNN to automatically process a microscope image of blood cells, and identify blood cancer cells and classify them under a certain stage, are very useful method, as it resulted in a clear output that shows the validity of the used method used, and I have discussed these results in the discussion section.



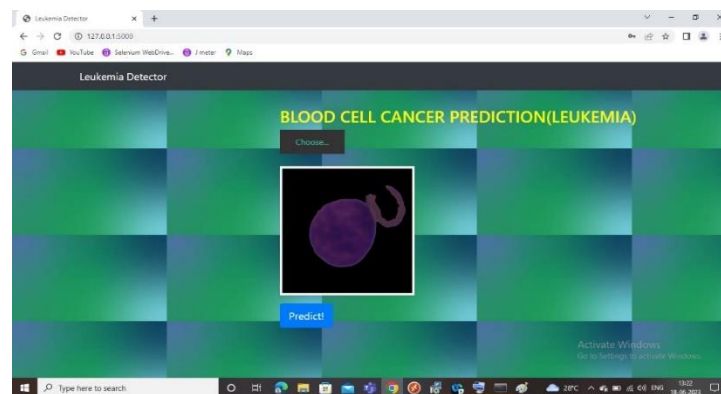
This is Login Page

Upload Page



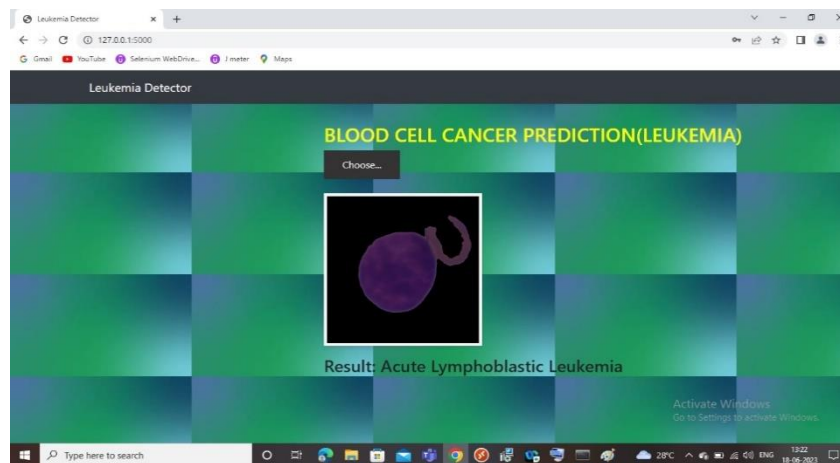
This is Upload page where blood cell dataset image is upload

Preview Page



This is preview page where blood cell data set image to view

Prediction



This is page show the prediction result.

V.CONCLUSION

A crucial step in the diagnosis of blood-related disorders is the detection of blood cancer cells. Early diagnosis of the illness has the potential to improve treatment and life expectancy. This paper suggested an approach for classifying blood cancer cells from a sample of human blood utilizing pre-processing, feature extraction techniques, and modelbuilding with the CNN algorithm. The accuracy achieved by the suggested approach was up to 85.33%. This suggested system has the ability to identify between normal and diseased blood cells

REFERENCES

- [1] Blood Cancer Detection Using Image processing shivkumar Chatarwad,Pratik Bansode, Amar Burade, Prof.T.S Chaware
- [2] R. Adollah, M.Y. Mashor, N.F.M. Nasir, H. Rosline, H. Mahsin & H. Adilah (2008), "Blood Cell Image Segmentation: A Review", Proceedings of Biomed, Pp. 141–144.
- [3] N. Ritter & J. Cooper (2007), "Segmentation and Border Identification of Cells in Images of Peripheral Blood Smear Slides", 30th Australasian Computer Science Conference, Conference in Research and Practice in Information Technology, Vol. 62, Pp. 161–169.
- [4] R. D.Labati, V. Piuri, F. Scotti, "All-Idb: the acute lymphoblastic leukemia image database for image processing",2011.
- [5] Bhagyashri G. Patil, Prof. Sanjeev N. Jain," Cancer Cells Detection Using Digital Image Processing Methods", International Journal of Latest Trends in Engineering and Technology (IJLTET),vol.3 ISSN: 2278-621X, 2014d for early detection of cancer in blood cells.
- [6] Christopher, T., & Jamera, J. (2016). Study of Classification Algorithm for Lung Cancer Prediction. A. IJSET - International Journal of Innovative Science, Engineering & Technology.

-
- [7] Abdul Nasir, N. Mustafa & Mohd Nasir (2009), "Application of Thresholding in Determining Ratio of Blood Cells for Leukemia Detection", Proceedings of International Conference on Man- Machine system (ICoMMS).
- [8] Hu.Y., Ashenayi K., Veltri, R., O'Dowd G., Miller G., Hurst R.Bonner. A comparison of neural network and fuzzy c-means methods in bladder cancer cell classification; In Proceedings of IEEE World Congress on Computational Intelligence; 6; 1994; p. 3461 – 3466.
- [9] Hiro Takahashi, Yasuyuki Murase , Takeshi Kobayashi , Hiroyuki Honda. New cancer diagnosis modeling using boosting and projective adaptive resonance theory with improved reliable index; Biochemical Engineering Journal 33; 2007; p. 100– 109.
- [10] Yuchun Tang, Yan-Qing Zhang, Zhen Huang, Xiaohua Hu, Yichuan Zhao. Recursive Fuzzy Granulation for Gene Subsets Extraction and Cancer Classification; IEEE Transactions on Information Technology in Biomedicine; 2(6); 2008; p. 723 – 730.
- [11]. R. B. Hegde, K. Prasad, H. Hebbar, B. M. K. Singh, and I. Sandhya, "Automated decision support system for detection of leukemia from peripheral blood smear images," Journal of Digital Imaging, vol. 33, pp. 361–374, 2019.