Cad Systems for Automatic Detection and Classification of Covid-19 Using Image Processing and Machine Learning

Ajesh F[1], Felix M Philip[2], Priya P Sajan[3], Robbi Rahim[4]

{ajeshf@gmail.com[1], felixphilip86@gmail.com[2], priyasajans@gmail.com[3], usurobbi85@zoho.com[4]}

Associate Professor, Department of Computer Science and Engineering, Sree Buddha College of Engineering, Alappuzha, Kerala India[1], Assistant Professor, Department of Computer Science and information Technology, JAIN (Deemed University), Kochi, Kerala, India[2], Project Engineer,C-DAC, Trivandrum, Kerala[3], Sekolah Tinggi llmu Manajemen Sukma, Medan, Imdonesia[4]

Abstract. Counting of platelets in blood cells assumes to be a significant role in the health sector. In any case, the procedure of the manual tallying of platelets in blood cells is incredibly tedious, which prompts erroneous outcomes. So as to overcome these difficulties, this research presents a fully automated software solution, enriched with image processing and machine learning techniques to distinguish and to count the number of RBC, WBC and Platelets cells in the sample blood images and to classify the different types of viruses present in it. Several problems and missing features in existing white blood cell classifiers were addressed by implementing an effusively automated method using a multiclass classifier.

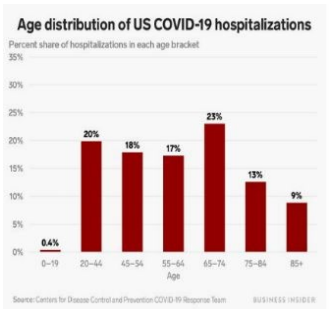
Keywords: COVID-19, Blood cells, Platelet count, CNN classifier, BoVm, SVM, ANN

1 Introduction

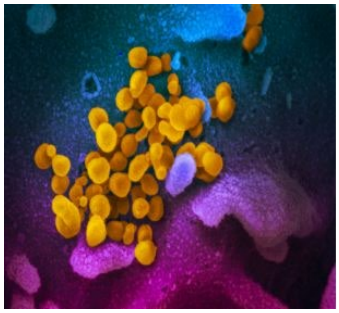
Corona viruses are a large category of viruses that can infect and cause sickness in both animals and humans. Corona viruses are spread by the air, water, and food. Corona virus infections in humans have been linked to respiratory illnesses ranging from the ordinary cold to life-threatening complications, such as the Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS) [1]. COVID19 viral infections are initiated by the most current corona virus. Individuals infected with COVID19 may develop mild respiratory distress and recover without the need for a specialised surgery. It is believed that older adults and those with unknown health issues such as cardiovascular disease, diabetes, chronic respiratory disease, and malignant development may acquire genuine sickness. More known individuals and previously ill individuals (asthma, diabetes, cardiovascular disease, etc.) offer the impression that the infection is severely sick [2]. Figure 1 (a), the static analysis report of the hospitalization of older people affected by COVID 19 and Figure 1 (b) gives the Corona virus image under an electron microscope.

B-ICSC 2022, September 20-21, Bandar Lampung, Indonesia

DOI 10.4108/eai.20-9-2022.2334191



(a)



(b)

Fig. 1. a): Statically Analysis of People Based on Age Distribution of COVID 19 Hospitalization b) Corona Viruses Image Under an Electron Microscope.

The virus is detected in the blood by obtaining blood samples from the infected person. The image of blood drop stains acquired using an electron magnifying apparatus is considered. Disease alters the physical characteristics of blood, which in turn alters the patterns of deposition of dried blood microdroplets [3]. For instance, a decrease in platelet count would result in the fading of blood, altering the physical qualities of the blood, such as its viscosity (thickness). As a result, blood stains in microdrop patterns can be used to diagnose disease [4]. The patterns of regular, healthy individuals's microdrop blood stains are easily distinguishable from the patterns of ill persons.

2 Methodology

This approach identifies various color feature statistics with geographical measures for machine learning centered on supervised learning. The proposed CNN classifier was able to avoid the numerous drawbacks identified by existing classifier and shortcomings on manual identification. The different stages are Image acquisition and preprocessing, segmentation of blood samples followed by feature extraction and classification. The digital blood images from the microscope contain noise and can be excluded by means of a filter, which, is followed by color correction for the identification of the infected region. The WBC are extracted by region-based segmentation followed by thresholding operation. The ABCD feature extraction technique is used for feature extraction and finally CNN classifier utilized for separating infected or non-infected blood cell. Figure 3 gives the sample image of an image obtained by digital microscopic of Corona viruses and (b) gives the enhanced image of Corona viruses after adding color correction technique. A CNN is a specific sort of multilayer perceptron, however a basic neural network, in contrast to a deep learning architecture, is incapable of learning complicated characteristics. CNNs have demonstrated outstanding performance in a wide range of applications, including image classification, object identification, and medical image analysis, among others. The key concept underlying a CNN is that, in contrast to a deep learning architecture, it may extract local features from high-layer inputs and transmit them to lower layers for processing into more complex features. CNNs have demonstrated outstanding performance in a wide range of applications, including image classification, object identification, and medical image analysis, among others. The underlying concept of a CNN is that it can extract local features from high-layer inputs and transfer them to lower-layer inputs

in order to acquire more complex characteristics. A CNN is composed of three layers: convolutional, pooling, and fully connected (FC). Figure 2 depicts a typical CNN design, which includes these layers in addition to the base layer.

2.1 Image Acquisition

The blood samples are examined by digital microscope, which is interfaced to a computer so as to get a digital image for processing [5].

2.2 Pre processing

Acquired images are preprocessed to remove unwanted noise and minor substances that are not measured as blood cells.

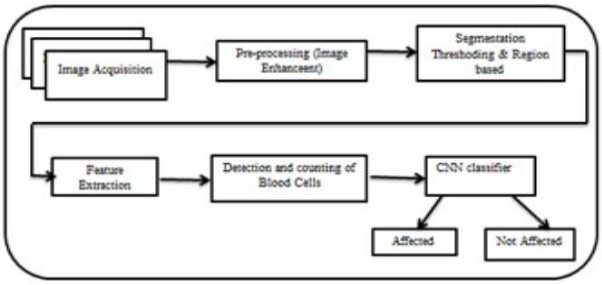


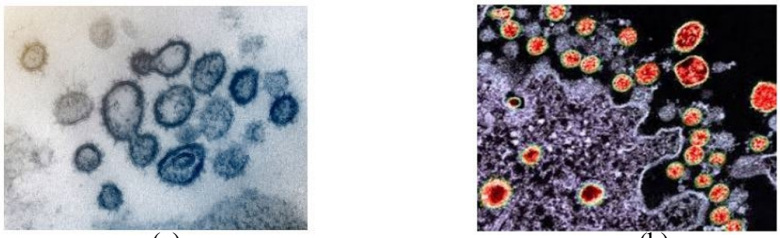
Fig. 2. Basic Block Diagram for the Detection of Corona Viruses by Using Machine Learning

2.3 Image Enhancement

The obtained digital image from preprocessing has to be improved, for better results. It can be achieved with a few image processing techniques including contrast adjustment, greyscale, edge detection, and spatial smoothing filtering.

2.4 Image Segmentation

This process includes choosing region of interest (ROI) in the image. This describes the area which comprises the blood cells. Region based segmentation [6] followed by thresholding is done in this stage.



(a) (b)

Fig. 3. (a): Digital microscopic image of Corona viruses by. (b) Enhanced image of Corona viruses after adding color correction technique.

2.5 Feature extraction

It is the most time-consuming activity in image processing since it involves tracing a large amount of redundant data into a small number of features with a smaller dimension. Here in our proposed system Texture and Shape based feature extraction technique is given importance as our blood cells also have different size and shapes. F. Detection and counting of Blood Cells The region based segmentation followed by thresholding identifies the blood cells in the image and to identify them. The function "draw circle" is used to mark circles around the detected cells and checking the quantity of cells drawn gives the complete number of blood cells. Table 1 gives the blood cell count for a normal human being and the parameters used for detection of blood cells are standard deviation, radius, roundness and it's by a given in below.

$$\text{Radius} = \frac{\text{Major Axis} + \text{Minor Axis}}{2} \quad (1)$$

$$\text{Rotation} = 4 \times \pi \times \text{area} \quad (2)$$

$$\text{Standard deviation} = \text{Convex perimeter}^2 \quad (3)$$

$$\text{Where, } X = \frac{\text{Major Axis} + \text{Minor Axis}}{2} \text{ (Medan)} \quad (4)$$

$$\text{Total Count} = C \times DF / (A \times D)$$

Where, C – Cell count of the image,

DF – Dilution Factor,

D – Depth of the height of the counting chamber

A – Input image area

Table 1: Blood Cell Count in Normal Human Being

Blood Cell Type	Women	Man
RBC	4.0-5.0 million/microliter	4.0-6.0 million/microliter
WBC	4.0-5.0 thousand/microliter	4.5-11 thousand/microliter
Platelets	4.0-5.0 thousand/microliter	150-450 thousand/microliter

2.6 Classification

This is the final phase, the features extracted is used provide the final reaction. All the features extracted are listed in the different columns with their values. Once the classified WBC images are given as an input to the proposed system, then the feature values are calculated. The values of the test image features are checked with the parameters implemented. Three parameters are being used to detect abnormalities. Thereby classifying the blood cell and infected cell. For Classification we are using convolutional neural network.

3 Discussion and Validation

To design a system for computer assisted diagnosis is a complicated process that needs knowledge about the disease's biology, information about the work flow of the hospital and insight of the technical solutions available.

3.1 Data Set Information

The real time data taken in this research work from three COVID19 designated Government hospitals from India and the sample database for training the classifier is got available from the website (http://medicalsegmentation.com/covid19/.

We make use of the Blood Cell Count Datase, which is a publicly accessible dataset of annotated blood cell pictures (BCCD). Originally, it had a total of 364 annotated smear photos, however there is a significant fault in the dataset. In fact, after dividing the dataset into two halves, training (300 files) and testing (64) files, we discovered that one annotation file in the test set did not contain any RBCs, despite the fact that the picture had RBCs. Furthermore, three annotated files have an RBC that is significantly lower than the real. We eliminate four of the incorrect files from the test set, reducing the total number of files in the test set to 60. We chose 60 training photos with annotations from a random sample for the validation set.

3.2 Software Specification

The basic software requirements needed for the executions are firstly platform or the simulator to evaluate and to execute the code of the proposed algorithms, MATLAB (2019) and the Microsoft Windows 8.1 operating system.

3.3 Performance Measures

Sensitivity, specificity, accuracy, precision and border error are statistical measures based on for variables of true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Performance metrics calculation [7].

The analysis chart for the classification of result based on different classifier is given below.

Accuracy = $\frac{TP}{TP+FN}$ x 100%

$$\text{Specificity} = \frac{TP+TN}{TP+FP+FN+TN} \text{ x } 100\%$$

$$\text{Sensitivity} = \frac{TN}{TN+FP} \text{ x } 100\%$$

3.1.1 Validation

Dengue and Corona viruses are difficult to distinguish as they share same laboratory and clinical features [8]. A person diagnosed with false positive dengue is later diagnosed to be Covid 19 positive. To diagnose Dengue the main classifier considered are BoVW, SVM, ANN and CNN classifier and its analysis chart for different classifier is given in table 2 [9]. Among the existing classifier, CNN classifier gives best result in terms of Accuracy, Sensitivity, Precision, Specificity and Border error. So in our proposed work also we suggest CNN classifier for final classification of infected blood cell [10].

3.1.2 K-fold cross-validation Algorithm

a. Define sets of values for model parameters
b. Randomize the dataset
c. K=12 groups were created from the dataset.
d. Perform the following for each parameter set:
e. for each set of resampling iterations, do
f. Provide Specific Samples
g. Fit the model to the rest of the space.
h. Make an educated guess about the holdout samples.
i. end
j. Calculate the average performance to determine whether or not the forecast will hold true.
k. end
l. Calculate the optimal parameter set
m. Fit the data to the optimum parameter set for each training parameter set.

We are able to recognise and count RBCs, WBCs, and platelets on a completely automated basis. We test our model on a test dataset consisting of 60 photos with known ground truths and find it to be accurate. Using different confidence criteria, we begin by counting the cells in the validation dataset using our model and various confidence thresholds. It is important to note that this threshold is significant in YOLO, since it is used to predict particular grid cells rather than the whole image, rather than the entire picture. In grid cells that do not contain any blood cells, the degree of confidence is minimal. As a result, by picking an appropriate confidence level, we may remove redundant and incorrect forecasts.

Table 2: K-Fold Cross Validation Approach and Results of Different Performance Metrics

Number of Folds	Sensitivity %	Specificity %	Precision %	F1-score %	Accuracy %
Fold-1	98.50	96.77	99.49	98.99	98.27
Fold-2	99.50	100.00	100.00	99.75	99.57
Fold-3	97.00	93.55	98.98	97.98	96.54
Fold-4	99.50	93.55	99.00	99.25	98.70
Fold-5	98.50	100.00	100.00	99.24	98.70
Fold-6	97.50	93.55	98.98	98.24	96.97
Fold-7	97.00	96.77	99.49	98.23	96.97
Fold-8	96.50	90.32	98.47	97.47	95.67
Fold-9	96.50	100.00	100.00	98.22	96.97
Fold-10'	96.00	90.32	98.46	97.22	95.24
Average	97.65	95.48	99.29	98.46	97.36

In order to calculate the average absolute error between the ground truths and the estimated number of cells in the validation dataset, we use the ground truths as the reference for the computation. It is possible to get the smallest possible average absolute error value for each cell by using a range of confidence thresholds, and we may utilise those confidence values to make judgments throughout the identification process of blood cells.

The error is calculated with the help of the following equation:

$$\varepsilon^{\text{cell}} = \frac{1}{N} \sum_{i=1}^{N} \left| \chi^{(i)}_{\text{groundtruths}} - \chi^{(i)}_{\text{estimated}} \right|$$

For each cell (RBC, WBC, or platelets), N specifies the size of the validation dataset (in our experiment, it is 60), is the total number of cells, and is the average absolute error value for that specific cell (in our experiment, it is 0.5). Table 1 shows the results of the computation of the error values. As can be seen in the chart, we may use a nominal threshold of 0.55 for counting RBCs while doing the count. The threshold for WBC and platelets, on the other hand, is shown to be significantly lower (0.35 and 0.25 in our experiments). As a result, the following are the relevant thresholds for each type of cell that have been determined:

The confidence level for RBCs is 55%, the confidence barrier for WBCs is 35%, and the confidence threshold for platelets is a quarter.

4 Result

In our proposed method, blood sample image from the electron microscope is taken into consideration. Preprocessing and morphological operation are done on the input image.

Segmentation is done so as to separate ROI based on Region Based method and thresholding operation, which is then followed by Texture and Shape based feature extraction method. Detection and Counting of Blood Cells is done further. All this information is fed to the classifier for training and testing. The classifier used is this system is CNN as it gives best results with dengue fever, whose virus has the same property and great similarity with the Corona viruses. The technique of the suggested method is clearly defined in Figure 4. In this example, (A) the input image, (B) the grey scale image, (C) the contrast adjustment, (D) histogram equalisation, (E) region based segmentation, and (F) the stimulated output are all shown. Table 3 shows the results of the examination of several classifiers performed with MATLAB 2019a.

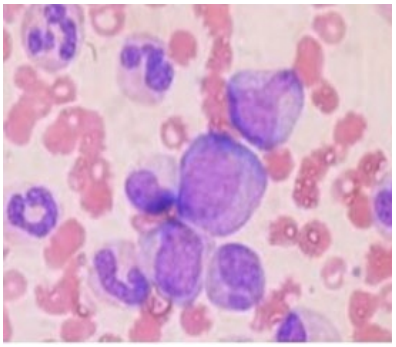
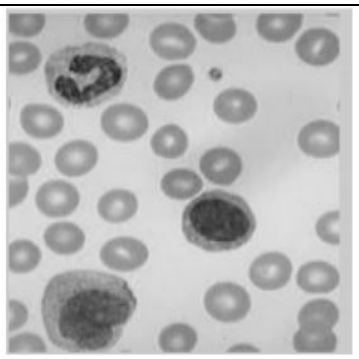
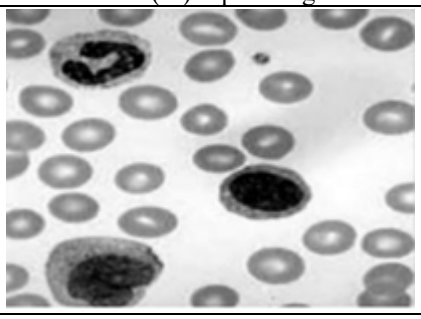
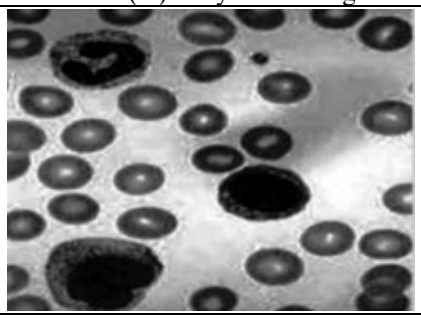
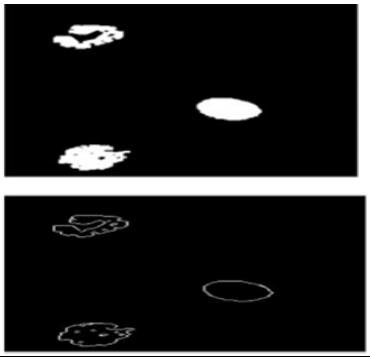
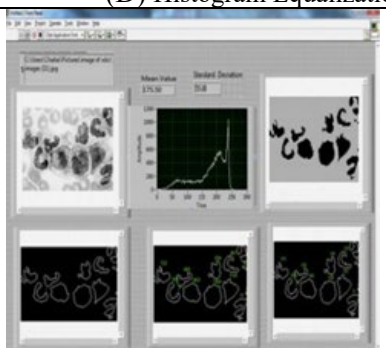
<table>
<tr><td>  </td><td>  </td></tr>
<tr><td>(A) Input image</td><td>(B) Gray Scale Image</td></tr>
<tr><td>  </td><td>  </td></tr>
<tr><td>(C) Contrast Adjustment</td><td>(D) Histogram Equalization</td></tr>
<tr><td>  </td><td>  </td></tr>
<tr><td>(E) Region Based Segmentation and Thresholding and its Edge Detection</td><td>(F) Stimulated output Using MATLAB</td></tr>
</table>

Table 3. Analysis chart based on different classifier based on the result

Classifier	BoVW	SVM	ANN	CNN
Accuracy	96.5%	97.4 %	94.3 %	**98.5 %**
Error	3.5%	2.6%	2.9 %	**1.5%**
Sensitivity	93%	92%	92%	**95%**
Specificity	100%	99%	98%	**100%**
Precision	93.5%	93.9 %	93.4 %	**95.5 %**

5 Conclusion and Future Scope

The objective of this work is to build up an automated system to detect corona virus using Image processing and machine learning technique. Corona viruses are the major cause of death globally in the present scenario and the early detection of this disease is important. This computer aided detection system helps the physician to diagnosis the disease more precisely and accurately by reducing time. The steps involved in proposed work are preprocessing, segmentation, feature extraction and classification. After preprocessing blood images are segmented by region based segmentation and thresholding. Features are extracted and final classification is done by CNN classifier. Our main objective is to accomplish 100% accuracy by employing an optimization technique along with CNN classifier. Future aim is to successfully implement the proposed work with the real time corona viruses infected image and to get 100 % exact outcome keeping up accuracy and time utilization.

References

[1] Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020 382 : 72733.

[2] Kumar M, Pandey D, Shrivastva P. Effect of GSR Biofeedback Relaxation Training on Blood Glucose and Anxiety Level of Type 2 Diabetic Patients. International Journal of Indian Psychology. 2016 4, (1) No. 82

[3] S.K. Lau, K.S. Li, A.K. Tsang, C.S. Lam, S. Ahmed, H. Chen," Genetic characterization of Betacoronavirus lineage C viruses in bats reveals marked sequence divergence in the spike protein of pipistrellus bat coronavirus HKU5 in Japanese pipistrelle: implications for the origin of the novel Middle East respiratory syndrome coronavirus", J Virol, 87 (15) (2013), pp. 86388650

[4] Wo¨lfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Mu¨ller MA, et al. Virological assessment of hospitalized patients with COVID2019. Nature 2020 581 :

[5] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 395 : 497506
[6] Gupta N, Agrawal S, Ish P, Mishra S, Gaind R, Usha G, et al. Clinical and epidemiologic profile of the initial COVID19 patients at a tertiary care centre in India. Monaldi Arch Chest Dis 2020 10 : 90
[7] LagunasRangel FA. Neutrophiltolymphocyte ratio and lymphocytetoCreactive protein ratio in patients with severe coronavirus disease 2019 (COVID19): A metaanalysis. J Med Virol 2020; 10.1002/jmv.25819.
[8] World Health Organization. Coronavirus disease (COVID19): Situation Report 169. Geneva: WHO 2020.
[9] S.U.Aswathy, Ria Mathew,"CAD system for Automatic Detection and Classification of Covid 19 in Nano lung image by Using Machine Learning Technique"International Journal of Pharmaceutical Research 2 (12), 18651870,2020
[10] World Health Organization . Laboratory testing for coronavirus disease 2019 (COVID19) in suspected human cases. Interim guidance, 2 March, 2020. Available from: https://apps.who.int/iris/handle/10665/33132 9, accessed on May 12, 2020.