International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-7, May 2020

# An Assessment of Gpu and Cpu based Convolutional Neural Network for Classification of White Blood Cells

**Babatunde R. S., Adigun J. O., Okikiola F. M**

***Abstract:*** ***White blood cell (Leukocytes) is made up of bone marrow located in the blood and lymph tissue. They are portion of the human body's immune system, thereby helping the body system to fight against infection and other related diseases. The number of leukocytes in the blood is usually part of a complete blood cell (CBC) test, which may be used to check for conditions such as infection, inflammation, allergies, and leukemia. Automation of variance count of leukocytes offers valuable information to medical pathologist to diagnose and treat of many blood based diseases. Early characterization and classification of blood sample is a major lacuna in the medical field, giving rise to lots of challenges for pathologist to adequately predict blood based disease. Several successful efforts have been made to address the aforementioned challenges with the use of machine learning generally and Convolution Neural Network in particular. However the processor configuration which can result in real time, and accurate classification of the high dimensional pattern is imminent, and a vast number of researchers are not explicit on the system configuration used to obtain the result in their report, which is the crux of this research. In this research,12,500 augment images of blood cells was obtained from the Kaggle Repository online. The leukocytes are contained in the blood smear image and categorized into five major types of their types: Neutrophil, Eosinophil, Basophil, Lymphocyte and Monocyte. The color, geometric and texture features are used by the pathologists to differentiate the leukocytes. The Simulation was done using python programing language and python libraries including Keras, pandas, sklearn, numpy, scipy and matplot for potting of graphs of results. The simulation was done on both CPU and GPU processor to compare the performance of the processors on CNNs based classification of the data. While CPU has faster clock speed GPU has more cores. Hence the evaluation metrics used which are precision, specificity, sensitivity, training accuracy and validation accuracy revealed that GPU processor outperforms CPU in terms of the stated metrics of comparison. Therefore a high configuration processor (GPU), which handles graphics better is recommended for processing image data that involves the use of machine learning techniques***

***Keywords: convolution neural network, Image data, leukocytes, machine learning.***

**Revised Manuscript Received on May 30, 2020.**
***** Correspondence Author
**Babatunde R. S.*,** Department of Computer Science, Kwara State University, Malete, Kwara. Nigeria. Email: ronke.babatunde@kwasu.edu.ng
**Adigun, J. O.**, Department of Computer Technology, Yaba College of Technology, Lagos, Nigeria.
**Okikiola F.M.**, Department of Computer Technology, Yaba College of Technology, Lagos, Nigeria.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

## I. INTRODUCTION

The white blood cell(leukocytes) are cells of the immune system that protect human body against infection and foreign diseases. The leukocytes are formed from multipotent cells in the bone marrow called hematopoietic stem cells.
Leukocytes are created from the body, which include the blood and lymphatic system. Hematologists and hematopathologist are specialists in diagnosis of diseases of the blood and bone marrow cells. Hematological tests system can help diagnose different kinds of diseases such as anemia, hemophilia, and leukemia. (Luis, Rodrigo, Alan, Flavio, Romuere, Kelson, 2017).

The detection of leucocyte based disease play an important role process achieved by medical practitioners which can be done automatically by building models resulting from analyzing datasets. These dataset composed of blood samples data, symptoms and images among others that are vital for the identification of diseases in a patient. A wrong diagnosis can be detrimental of the patient, and possibly affect the prescription of drugs that are not applicable for the treatment of such diseases. To alleviate this challenges, there exists the need to have low-cost computational systems that process the data and provide investigative support for experts at this critical stage. (Chen, Xu, Yan, Wong, Wong, Liu, 2015).

Similarly, medical organizations have always challenged with appraisal and benchmarking in employing automated detection, specifically when there are no superior models. Moreover, lack of poor selection of classification for severe diagnosis of leukemia may be expensive for health establishments (Alsalem et al., 2019).In the past decades, several medical assisted systems have been deployed in various diseases to support early detection and characterization. Such systems have helped asides other benefits, reduce the progressive symptom growth, which can result into loss of life.

## II. RELATED WORK

Syed et al (2016) developed a leukocyte cell segmentation and classification system that uses a strategy to segment cell images. A Wiener filter algorithm and Curvelet transform was employed for enhancement and noise elimination of false edges of the image. The combination of entropy filter, thresholding and mathematical morphology was used for image segmentation and boundary detection.

Back-propagation neural network was used to classify leukocyte cell into its sub classes and obtained accuracy of 100%, for basophil, 96.15% for eosinophil, monocyte for 92.30%, lymphocyte for 92.30% and 96.15% for neutrophil respectively. The segmentation results obtained overwhelmed the problem of overlapping cells. Thanh et al (2018) proposed a Convolutional Neural Network (CNN) based method to distinguish normal and abnormal blood cell images. A largely augmented dataset was used in order to confirm the accuracy and reliability of the proposed CNN architecture. The method achieved an accuracy of 96.6% on a dataset containing 1188 blood cell images.

Bhukya et al., (2017) developed a leukemia detection system by proposing a new method to separate the cell Nucleus from Cytoplasm to obtain more features. The work classified the Leukocyte into healthy or unhealthy based on various features which were extracted from the cell as well as nucleus. This was achieved using the segmentation algorithm proposed in the work, thereby achieving an accuracy of 93.33% with the use of 80 images from ALL-IDB2 database. Luis et.al. (2017) developed a diagnosis system for identification of leukemia in white blood cells images using CNN, Principal Components Analysis (PCA), with group of three classifiers. In the work, the feature of the blood smear image was extracted using pre-trained CNNs to obtain the image description. PCA was used for feature selection for the final descriptor. Three classification algorithms (Support Vector Machine, Multilayer Perceptron and Random Forest) was created to classify the images. A 100% of accuracy rate was achieved with less processing time. Ramya and Rani (2019) implemented a bioinspired system that reliably extracted the features of the images using bioinspired concept. F-measure and n-fold cross validation techniques were used as evaluation metrics for the system. The results achieved by the developed system showed that classifier performed well and has comparable capability with the previous works.

Agaian, Madhukar and Chronopoulos, (2018) developed an automated classification system that contain blood images of acute lymphoblastic leukemia of multiple nuclei. The work was tested on used features in other existing systems that include cell energy and colour features. The experiment was evaluated using multiple cross-validation methods and results show that there was efficiency in the classification of acute leukemia in blood cells smear images. Manisha and Sethukaras, (2019)analyzed and appraised the existing algorithms and approaches used for detecting blood cancer. The authors reported that the application of feature extraction and dimensionality reduction techniques must be applied efficiently in order to achieve less time detection and reduction errors in diagnosis.

However, from the above reviews, the system requirement that can achieve a high performance automated detection for real time adoption were not explicitly stated by the researchers, hence the intent of this research is to carry out a comparison between the CPU and GPU based system configuration respectively, considering the machine learning algorithms performance metrics in a bid to recommend a suitable option for possible adoption.

## III. METHODOLOGY

### A. Data Acquisition, Image Loading and Labeling

The dataset used in this research was obtained from aggle online dataset repository. The dataset contains over 12000 images of both testing and training images of white blood cells samples (leukocyte) as shown in Figure 1.

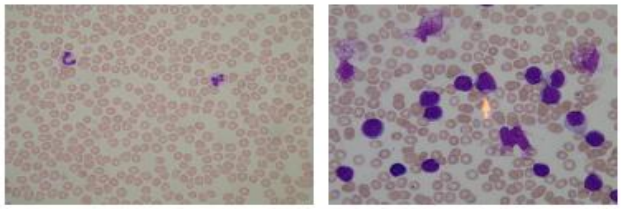

**Figure 1: Leukocyte (WBC) images**

The dataset images stored on host system was imported into python programming environment using the jupyter notebook. The python code instructions were used for image resizing and other operations are further carried out on the image data which was subsequently passed on to CNN process. A CNN consists of an input and an output layer, as well as multiple hidden layers.

### B. Convolutional, pooling and fully connected layer

In this research, the hidden layers comprise the following convolutional, pooling, normalization and fully connected layers. The convolutional layers applied convolution operation to input image and the result is passed to the next layer. The CL emulates the response of an individual neuron to visual stimuli. After loading and labelling, the Convolutional layer (CL)used local connections with the tied weights to perform feature extraction. The input to the convolutional layer is an image of the form x*x*r, where ‘x’ is height and width of image and ‘r’ is the number of channels. The images are filtered by the size y*y*r such that ‘y’ is less than the scope initial image. A section of the image, called a filter -size-portion is extracted and the computation of the convolution resulted to a unit number of output to which a bias is added. The filter is glided on the entire whole image to compute the output. The output strides were concatenated to obtain a feature map which is a reduced form of the initial image. The boundary of the input layer is then padded with zeroes. The Pooling layer (PL) was employed to decrease the height and width of image feature map. This reduces the processing time, number of parameters and epochs used. This research utilized max pooling that has a filter size of f*f was used for the maximum operation over the image. Therefore, the application of the pooling was completed with a filter size of 2*2 using a stride of 2thus decreasing the image into half size.

The classification layers were applied to classify the features extracted. The extracted neuron is linked to the neuron of previous layer. The softmax activation function is applied to translate the class of each output of network into probability distribution. Back propagation and Gradient descent techniques were used to determine the choice of parameter for the network. The values of these set of parameters were randomized and varied for every training image to obtain a correct output.

The accuracy of the classifier is inversely proportional to the cost; hence the cost must decrease. Therefore, the training was done until the cost is constant and the output stored in a binary file named *model*. A new query image is passed on to the network as input to be classified and its probability is calculated, giving the output as the *inference* or prediction. The whole training data was distributed into groups of images called *epochs* that comprises of 800 or more images resulting into about 10 iterations to train the training dataset.
A sample CNN is shown in Figure 2.

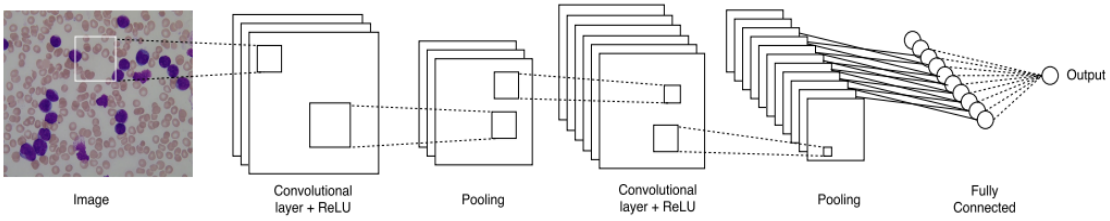


**Figure 2: A sample convolution neural network**

The process flow of the proposed classification system is shown in Figure 3. The flow begins by preprocessing the input images and subsequently passing the preprocessed image on to the CNN for classification. The classification was carried out differently on both GPU based system as well as CPU based system

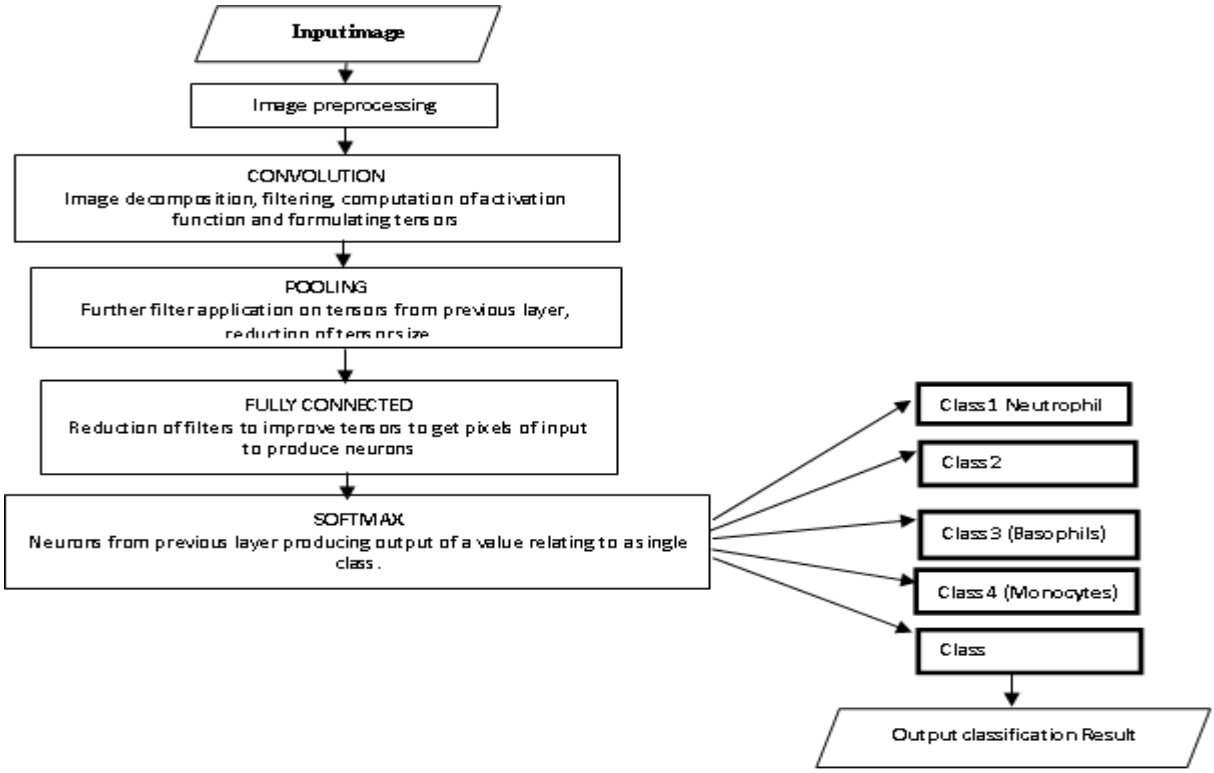


**Figure 3: Process Flow of CNN classifier**

## IV. RESULT AND DISCUSSION

The simulation was carried out on both a CPU and a GPU processor in other to compare the performance of such processors in classifying Leukocytes. The result of the classification process was evaluated with metrics including training accuracy, sensitivity, specificity and confusion matrix

### A. Classification using CPU based Processor

Dataset used on CPU based Processor comprises of a total of 803 of both training and testing set of white blood cell images. The system configuration is shown in the table 2 below. CNN as a classifier require a high hardware configuration in order to result in a more optimized classification outcome.

**Table 1: Hardware Configuration/Requirement used**

| Hardware Component | Configuration/Requirement |
|---|---|
| Processor | Intel Dual core 1.73Ghz |
| Processor Type | CPU |
| RAM | 3.0 GB |

### B. Training Accuracy and Validation Accuracy

The Training accuracy is the accuracy of the CNN classifier on the training images.

The Validation accuracy is the accuracy obtained from classifying the test images on the CNN Classification model. Figure 4depicts the training and validation loss. It can be deduced from this result that the training loss is high and at a constant rate with increase in the number of epochs. However, the Validation loss is low and almost at a constant rate with increase in number of epoch. Therefore increase in epoch has no significant effect on validation loss. Similarly, figure 5 shows the accuracy and correct classification of the training and test images

```
plt.plot(epochs, val_loss, label='Validation loss')
plt.title('Training and validation loss')
plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.legend()

plt.show()
```

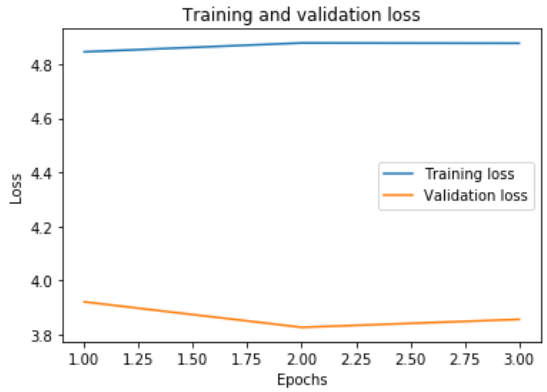


**Figure 4: Training and Validation Loss from CPU**

```
plt.plot(epochs, val_acc, label='Validation acc')
plt.title('Training and validation accuracy')
plt.xlabel('Epochs')
plt.ylabel('Accuracy')
plt.legend()

plt.show()
```

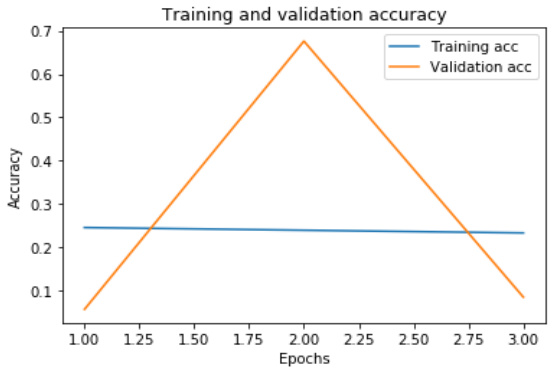


**Figure 5: Training and Validation Accuracy from CPU**

The training accuracy is low and constant with increase in epochs as shown in the figure, while the validation accuracy increases drastically but drops as the epochs was increased. Therefore the epoch at the point where the accuracy increases can be noted and adopted with a CPU system.

### C. Confusion Matrix for CNN classifier using CPU processor

The leukocyte images consist of five different types giving rise to five classes, each class possesses certain features and therefore are been classified by the CNN classification model. The actual and misclassifications are show in the confusion matrix in figure 6 in percentage. The vertical representation depicts known classes while the horizontal representation depicts the predicted classes. The diagonal boxes of the confusion matrix show the actual and correct classification of the different leukocyte according to the correct class. The other boxes show the misclassification which is the classes wrongly classified.

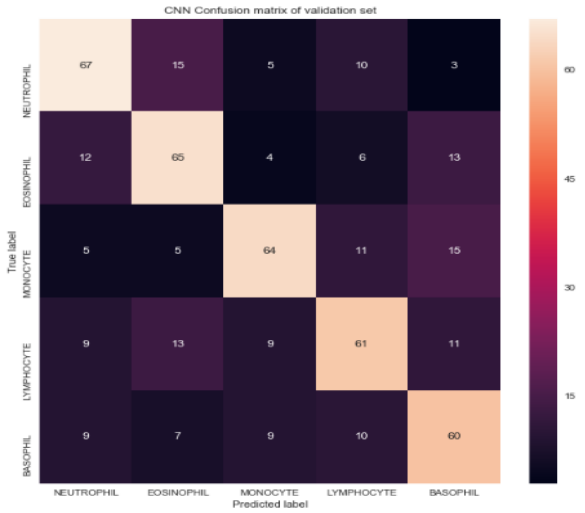


**Figure 6: Confusion matrix on CPU system**

### D. Precision, Sensitivity and Specificity Evaluation

(i) Precision: is computed as TP/ (TP+FP). Therefore, the `five class precision will be calculated as

Precision for class A = **TP**a/ (**TP**a+**E**ba+**E**ca+**E**da+**E**ea)

where:

**A** is the selected class (A…An)

**TP**a is the True Positive (correct classification)

**E**ba…. **E**ea are the errors of misclassification to the other classes know as FP (False Positive).

From the Simulation carried out using the CPU system, the result for the precision is shown in table 2.

**Table 2: Precision results for each classes of leukocyte**

| *CLASS* | *PRECISION* |
|---|---|
| NEUTROPHIL | 0.66 |
| EOSINOPHIL | 0.62 |
| MONOCYTE | 0.70 |
| LYMPHOCYTE | 0.62 |
| BASOPHIL | 0.59 |

(ii) Sensitivity corresponds to the True Positive Rate of the considered class. The sensitivity is calculated as TP/(TP+FN). The Five class sensitivity will be calculated as

Sensitivity for class A =**TP**a/ (**TP**a+**E**ab+**E**ac+**E**ad+**E**ae)

where:

**TP**a is the True Positive (correct classification)

**E**ab…. Eae are the errors of misclassification to the other classes known as **FN** (False Negative).

From the Simulation carried out using the CPU system, the following result shown in table 3 was obtained for the Sensitivity.

**Table 3: Sensitivity results for each classes of leukocyte**

| *CLASS* | *SENSITIVITY* |
|---|---|
| NEUTROPHIL | 0.67 |
| EOSINOPHIL | 0.65 |
| MONOCYTE | 0.64 |
| LYMPHOCYTE | 0.59 |
| BASOPHIL | 0.63 |

**(**iii**)** Specificity corresponds to the True Negative Rate. The specificity is calculated as TN/(TN+FP). Therefore, the five class specificity will be calculated as

Specificity for class A= TNa/ (TNa+Eba+Eca+Eda+Eea) where:

A is the selected class (A…An)

TNa is the True Negative

**E**ba…. **E**ea are the errors of misclassification to the other classes know as FP (False Positive). Table 4 shows the result.

**Table 4: Specificity results for each classes of leukocyte**

| *CLASS* | *SPECIFICITY* |
|---|---|
| *NEUTROPHIL* | 0.91 |
| *EOSINOPHIL* | 0.90 |
| *MONOCYTE* | 0.93 |
| *LYMPHOCYTE* | 0.91 |
| *BASOPHIL* | 0.90 |

## E. Classification using GPU based Processor

The GPU processor uses a remote/virtual GPU environment, hosted by Kaggle. The result from the classification was based on the system configuration in the table 5 below. Using a GPU processor, the classification process yielded an optimized result compared to CPU as shown in the result presented below.

**Table 5: Hardware Configuration/Requirement GPU used**

| ***Hardware Component*** | ***Configuration/Requirement*** |
|---|---|
| *Processor* | GPU core processor |
| *Processor Type* | GPU |
| *RAM* | 13gb |
| *System type* | Remote/Virtual Operating System |

## F. Validation Accuracy and training Accuracy

Figures 7 and 8 respectively show the Training and Validation accuracy, which shows an improved and faster result over CPU processor as seen in the previous result.

It can be seen from figure 7 that the training loss is low and constant with increase in epochs, hence the increase in epoch has no significant effect on the training loss.

However the validation loss at first was very high and then drops down and remain constant with increase in epochs. This might suggest that a low validation loss can be achieved with reduced epoch.

The training and validation accuracy achieved with increase in epoch with the GPU processor is high and at a steady rate as shown in figure 8.

```
plt.plot(epochs, val_loss, label='Validation loss')
plt.title('Training and validation loss')
plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.legend()

plt.show()
```

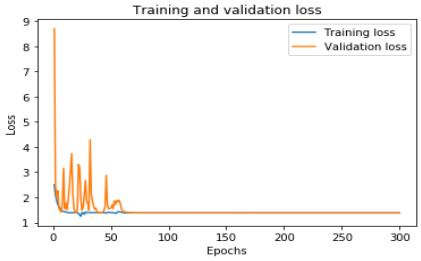


**Figure 7: Training and Validation Loss from GPU processor**

```
plt.plot(epochs, val_acc, label='Validation acc')
plt.title('Training and validation accuracy')
plt.xlabel('Epochs')
plt.ylabel('Accuracy')
plt.legend()

plt.show()
```

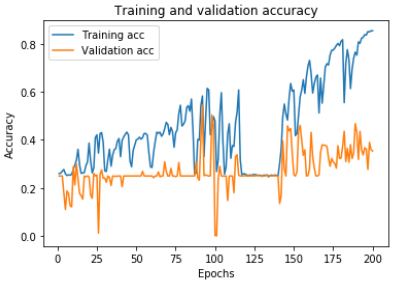


**Figure8:Training &Validation Accuracy from GPU**

## G. Confusion Matrix for CNN classifier using GPU

The confusion matrix, shown in Figure 9 is obtained from a high end system configuration and specification compared to the CPU, it produces a higher level of accuracy in terms of correct classification, reduced miss classification and shortest run time.

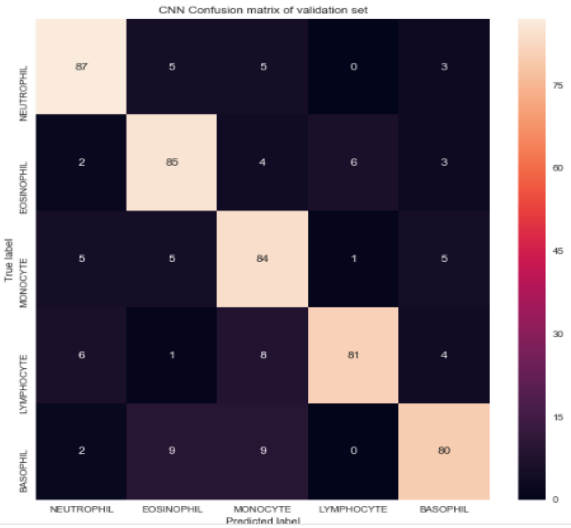


**Figure 9: Confusion matrix GPU**

### H. Precision, Sensitivity and Specificity Evaluation

The result obtained from the simulation, is presented in table 7, showing the Precision, Sensitivity and Specificity obtained from a high configuration processor (GPU).The number of epoch and batch size are 300 and 600 respectively.

**Table 7: Result of precision, sensitivity and specificity for GPU system**

| *CLASS* | *PRECISION* | *SENSITIVITY* | *SPECIFICITY* |
|---|---|---|---|
| *NEUTROPHIL* | 0.85 | 0.87 | 0.96 |
| *EOSINOPHIL* | 0.81 | 0.85 | 0.95 |
| *MONOCYTE* | 0.76 | 0.84 | 0.94 |
| *LYMPHOCYTE* | 0.92 | 0.81 | 0.98 |
| *BASOPHIL* | 0.84 | 0.8 | 0.96 |

## V. CONCLUSION AND RECOMMENDATION

This work compares the performance of CPU and GPU processor on classification of Leucocytes using CNN. Experimental results indicated that a GPU processor based system results in an improved recognition accuracy for CNN when large image samples and multiple classes are used. Therefore, a high configuration processor will be more appropriate and suitable for classification using convolutional neural network most especially while considering a multiclass classification.

The classification accuracy will be further improved upon by the extension of data set size in order to circumvent misperception between Basophil and Lymphocyte cells. This is because their shapes are very similar in small images and can be achieved by improving the CNN structure for better performance. In the future the researchers intend to experiment on the use of other variants of Artificial Neural Networks such as Capsule Network for classification of Leukocytes.

## REFERENCES


1. Adigun, Okikiola, Aigbokhan and Rufai, "Automated System For Grading Apples Using Convolutional Neural Network", International Journal of Innovative Technology and Exploring Engineering (IJITEE) 9(1), 2019, pp. 1456-1464
2. AgaianSos, Madhukar Monica and Chronopoulos Anthony, "A new acute leukaemia-automated classification system, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization", vol. 6. Issue 3, 2018, pp. 303-314
3. Alsalem, M., Zaidan, A.A., Zaidan, B.B., Albahri, O.S., Alamoodi, A.H., Albahri, A.S., Mohsin, A.H., & Mohammed, K.I., "Multiclass Benchmarking Framework for Automated Acute Leukaemia Detection and Classification Based on BWM and Group-VIKOR". Journal of Medical Systems, 43, 2019, pp. 1-32
4. Bhukya Raju, Prasanth B., Vihari V., Ajay. Y., "Detection of acute lymphoblastic leukemia using microscopic images of blood". International Journal of Advanced and Applied Sciences, vol. 4 No 8, 2017, pp.74-78
5. Chen X., Xu Y., Yan S., Wong D.W.K., Wong T.Y., Liu J., "Automatic Feature Learning for Glaucoma Detection Based on Deep Learning". International Conference on Medical Image Computing and Computer-Assisted Intervention MICCAI 2015: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 pp 669-677
6. Luis H. S. Vogado, Rodrigo de M. S. Veras, Alan R. Andrade, Flavio H. D. de Araujo, Romuere R. V. e Silva, Kelson R. T. Aires, "Diagnosing Leukemia in Blood Smear Images Using an Ensemble of Classifiers and Pre-trained Convolutional Neural Networks". Conference Paper, 2017. DOI: 10.1109/SIBGRAPI.2017.55
7. Madhukar M., Agaian S., and Chronopoulos A. T., "Automated screening system for acute myelogenous leukemia detection in blood microscopic images," IEEE Systems Journal, vol. 8, no. 3, 2014, pp. 995–1004.
8. Manisha. S, Sethukaras T., "Automated Detection Of White Blood Cell Cancer Diseases Using Classification Techniques" - A Survey. International Research Journal of Engineering and Technology (IRJET) Volume: 06 Issue: 02, 2019, pp. 682-687 e-ISSN: 2395-0056
9. Mehdi Habibzadeh, Adam Krzy´zak, Thomas Fevens. "Comparative study of shape, intensity and texture features and support vector machine for white blood cell classification". Journal of Theoretical and Applied Computer Science Vol. 7, No. 1, 2013, pp. 20–35. ISSN 2299-2634 http://www.jtacs.org
10. Ramya, B. Rani. V. U., "Acute Myelogenous Leukemia Detection using Circumventing Ant Colony Optimization based Convolutional Neural Network". International Journal of Recent Technology and Engineering. ISSN: 2277-3878, Volume-8 Issue-4, 2019, DOI:10.35940/ijrte.C6732.118419
11. Syed H. Shirazi, Arif Iqbal Umar, SaeedaNaz and Muhammad I. Razzak "Efficient leukocyte segmentation and recognition in peripheral blood image" IOS Press. DOI 10.3233/THC-161133. Technology and Health Care Vol. 24, 2016, pp. 335–347.
12. Thanh T. P, Caleb Vununu, SukhrobAtoev, Suk-Hwan Lee, and Ki-Ryong Kwon, "Leukemia Blood Cell Image Classification Using Convolutional Neural Network". International Journal of Computer Theory and Engineering, Vol. 10, No. 2, 2018, DOI: 10.7763/IJCTE.2018.V10.1198


## AUTHORS PROFILE

**Ronke Seyi Babatunde** holds a Bachelor of Science (B. Sc) degree and Master of Science (M.Scs.) degrees in Computer Science both from University of Ilorin, Nigeria. She obtained her Ph.D. in Computer Science from Ladoke Akintola University of Technology, Ogbomoso, Oyo State, Nigeria. She is a Lecturer I staff at the Department of Computer Science, College of ICT, Kwara State University, Malete. Nigeria. Her research areas include among others, Pattern Recognition, Artificial Intelligence, Computational Intelligence, Big Data Analytics and Data Science. She is a member of the Nigerian Computer Society (NCS) and Computer Professional Registration Council of Nigeria (CPRN).

**Johnson Oyeranmi Adigun** obtained his PhD Computer Science from Ladoke Akintola University of Technology Ogbomoso, Oyo State, and Master's degree in Computer Science from University of Lagos, Nigeria. He is a Lecturer in the Department of Computer Technology, Yaba College of Technology, Lagos. He is a member of Nigeria Computer Society of Nigeria and Computer Professionals Council of Nigeria. His current research interests span the fields of artificial intelligence, pattern recognition, machine learning and medical image analysis. He has published more than 30 research papers in the leading International Journals and Conference proceedings. Dr. Adigun has written /edited 10 books. He is a reviewer of many International Journals

**Okikiola Folasade Mercy** is a Lecturer in Computer Technology Department,Yaba College of Technology, Yaba, Lagos, Nigeria. She holds a Bachelor of Science (B.Sc.) degree in Computer science from University of Ado, Ekiti State and a master (M.Tech.) degree in Computer Science from Federal University of Technology, Akure, Ondo State. She is a member of Computer Professionals of Nigeria (CPN) and Nigeria Computer Society (NCS)

.