Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.754719

A NOVEL DETECTION AND CATEGORISATION OF WHITE BLOOD CELLS USING A DUAL CONVOLUTIONAL NEURAL NETWORK APPROACH

Dr. SK. Akbar[1], I Murali Krishna[2], Vanama Kavya Sri[3], Kosuri Ruchitha Sri[4], Dhulipalla Chandra Deepika[5], Jorige Abivignay[6]

[1, 2, 3, 4, 5, 6] Department of Computer Science and Engineering, PSCMR College Of Engineering And Technology, Vijayawada, A.P.

ABSTRACT

Leukocyte detection and segmentation (WBCs) are a crucial stage in haematology imaging; a significant part in the diagnosis and management of serious illnesses is played by cell categorisation, particularly that of Leukocytes. White blood cells are also referred to as leukocytes which act as a foundation of the immune system. There are different approaches for the detection of WBC, but there are fewer papers which classify the type of WBC. The different approaches include Deep learning techniques and Support Vector Machine (SVM). The proposed methodology used deep learning strategies using the Convolutional Neural Network (CNN). Utilising blood cell imaging and the White blood cells (WBC), sequential image cropping technique and created a dataset of WBC. Then using CNN, the classification of images and type of WBC is determined as output. Through our CNN model, we can overcome the data augmentation by contemplating the blood cell images dataset and generate an accuracy of 99.191% by producing relevant results.

KEYWORDS: Convolutional Neural Network, Leukocytes, sequential image cropping technique, Deep Learning,

1. INTRODUCTION

WBCs are also addressed as Leukocytes. These are part of the immune system that is responsible for protecting our body from harmful diseases and also foreign invaders. Usually, these blood cells are brought about from multipotent cells that are present in myeloid tissue, which are also called Hematogenous stem cells. Every part of the body, including the blood, is where the lymphatic system and leukocytes are made. Only 1% of blood comprises white blood cells, although they have an immense impact. The two categories of white blood cells, granulocytes, and agranulocytes, are shown in Fig. 1. Granulocytes are nuclear leukocytes further separated into neutrophils, eosinophils, and basophils. Agranulocytes are mononuclear cells that can further be separated into two categories, monocytes and lymphocytes.

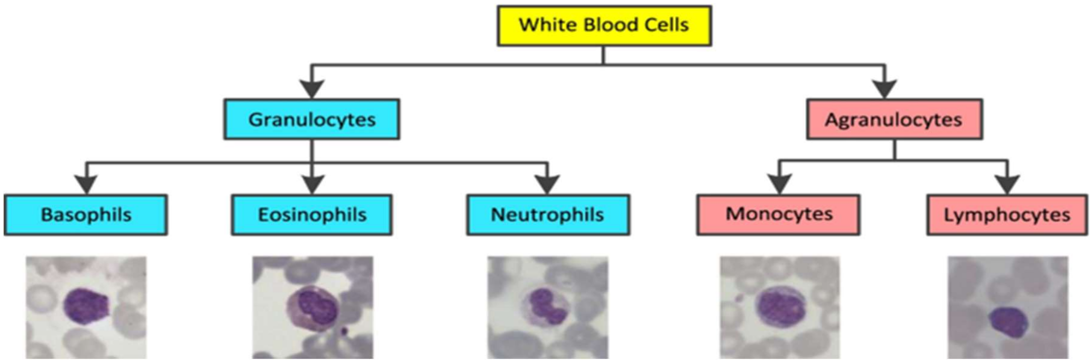

Figure 1: Classification of white blood cells (WBC)

While comparing different methods used in previous papers, they use the support vector machine (SVM) and convolution neural network (CNN). In that, the CNN gave better results for classification when compared to other techniques. So, the usage of CNN in this methodology. The previous papers took the dataset with a low no of images which can affect the results. The methodology is proposed to take a large dataset as input, which can be possible by CNN. Here the convolution neural network is the classifier to predict all the types of WBC to provide the result as classification of WBC

1.1 Convolutional Neural Networks (CNN)

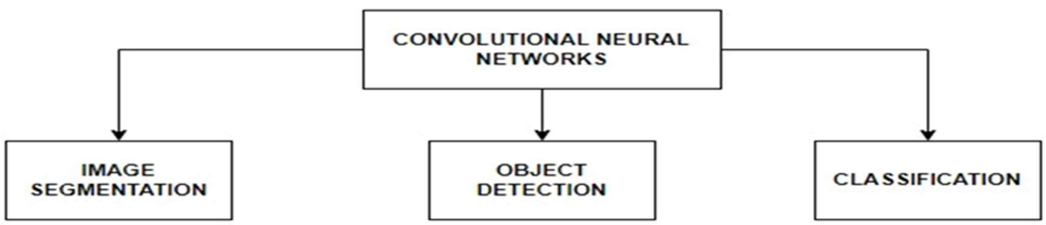

Figure 2: Different Uses of CNN

A sort of neural network is the CNN (convolutional neural network). The above Fig 2 tells the different usages of CNN. In that, the classification of images using CNN in the proposed model. This classification is possible by passing images through the different layers of CNN. The different layers are convolutional, fully connected layers, and output layers. Each layer has some specific function to perform. Every layer passes its output to the next layer as input. The convolutional layer returns a fully connected layer with the feature map. The one lengthy continuous array is passed by the completely connected layer. Finally, the classifications are done in the output layer.

2. LITERATURE SURVEY

Afrin Alma et al. [1] proposed A CNN–reinforcement learning method for acute lymphoblastic leukaemia diagnosis with automatic feature extraction. This model detects cancerous leukaemia using labelled microscopic images. To address the over-training problem, data augmentation techniques were used to increase the amount of training data in the ALL IDB. This gives 95.54% accuracy achieved using a five-fold validation procedure on 515 images. In processing raw photos features hurriedly, this model also has limitations.

Jing Chang et al. [2] Proposed Automated Classification of Leukocyte Using Neural Network The WBC categorisation job is subjected to deep learning techniques to enable fully automated execution of the work. The test obtains accurately categorises several different white blood cells in a short time. The lack of this model is to speed up categorisation; accuracy must be improved and parameters simplified.

Fatih Ozyur et al. [3] Proposed WBC detection with extreme learning machine and MRMR feature selection for the CNN model; in our research, we discovered that the feature extraction and feature selection (FS)-based CNN-MRMR-ELM approach increased accuracy. Of classification while taking more implementation time. WBC data were used to test the suggested CNN-MRMR-ELM approach. This approach yielded the maximum performance and accuracy ratio of 96.03%.

Sonali Mishra et al. [4] Proposed A GLRLM-Based feature learning for ALL detection using computer-aided design systems and SVM classifiers. They proposed a technique of leukocyte identifier in this paper by creating an autonomous system which is more trustworthy than physical labour done by operators that is conceptually cheaper. Our results show that the suggested approach can accurately categorise lymphoblasts and detect leukocytes in a blood smear with high efficiency. This has the limitation of taking more time to classify the result. The proposed approach identifies lymphoblast cells using GLRL features and has an SVM classifier accuracy of 96.97%.

Sweta et al. [5] proposed that Convolutional Neural Networks are used to Use microscopic images of the bone marrow to find white blood cancer. This research offered a reliable method for classifying multiple myeloma (MM) and ALL using the SN-AM dataset, with an accuracy of 97.2%. However, a huge scientific investigation is required to evaluate the effect on file size.

Amjad Rehman et al. [6] proposed A deep-learning classification of acute lymphoblastic leukaemia, and they provide a technique to classify ALL and bone marrow (normal) that are present in photos of discoloured bone marrow. The proposed model is trained on pictures of bone marrow utilising strong segmentation and DL approaches to provide accurate classification results using CNN. The proposed approach was accurate at 97.78%. The approach's accuracy in segmenting overlapping cells still requires enhancement.

Sarmad et al. [7] proposed the Identification and segmentation of leukaemia subtypes via pre-trained deep CNN. They use Data augmentation techniques were used to reduce overtraining. Methods of machine therapy were discovered to be more efficient, rapid, free, and exact than manual methods. The proposed approach was highly accurate, without microscopic image segmentation, achieving 96.06%. Sarmad failed to suppress noise from images, reducing performance.

Mohd Shafry et al. [8] White Blood Cell Feature Extraction in Acute Myeloid Leukaemia has been suggested. Images of blood smears Using CMYK-Moment Localization and DeepLearning. Utilising ROIs instead of the full image reduced noise and omitted extraneous characteristics, resulting in a more stable learning environment. This approach's major strength is that it uses traditional Machine learning strategies for feature extraction and photo editing algorithms for ROI extraction. Handmade characteristics, on the other hand, are difficult to automate, take longer to train, cannot be used to recognise different types of blood cells and cannot be generalised.

Huang et al. [9] proposed a Spectrum analyser Image Based Immune Cell Classification Using Modulated Gabor and CNN, with the aid of medical hyperspectral imaging; to identify leukocytes, an effective classification model that integrated the model Gabor. It has been demonstrated that representative features may be extracted for image classification. This model makes it conceivable that as the amount of data grows, the suggested MGCNN would require more processing power to train and test.

Reena et al. [10] proposed Leukocyte localisation and identification in blood plasma: a deep learning strategy. It identified WBCs were automatically cropped and downsized to create the database utilised for categorisation. One of the drawbacks is that the current database does not account for clustered leukocytes or contacting cells. Due to two-stage pipelines, the runtime is slightly slower than when anchor box-based detection is used.

Khamael et al. [11] proposed a New Digital Haematology Microscope Images Segmentation of White Blood Cells, Nucleus, and Cytoplasm: This describes the structure of WBCs, their features, and Hematology pictures shown in this work. The performance was evaluated utilising 30 pictures altogether. This approach has an average accuracy of 97.79%. One of the areas for improvement of our inquiry system is that omitted picture noise suppression, even though it had little effect on how our proposed process operates.

Mishra et al. [12] proposed Acute lymphoblastic leukaemia identification via texture feature-based categorisation on tiny blood smears. The suggested technique for detecting acute lymphoblastic leukaemia uses the random forest and DOST to classify blood smears based on texture features. The suggested technique outperforms existing methods in terms of accuracy (99.66%), as per simulation results derived from five k-fold stratified cross-validation runs. Due to its exact nature, which DOST removes, the S-transform has a variety of drawbacks, the main one being its increased complexity (time and space).

Abel Worku Tessema et al. [13] proposed Utilising a convolutional neural network, quantitative analysis of blood cells from microscopic pictures. The recommended method gives adequate data about Blood cells packed in the clinical examination of significant blood-related disorders because they can detect, count, and quantify morphometric features from a microscopic picture. This blood cell analysis method achieved an average mean effectiveness of 91.22%. It was tested on 1560 pictures and 2703-labelled blood platelets with various ultra; however, it needed more specific clinical data such as dimension and form.

Patil et al. [14] Proposed A WBC image classification using DL with CCA. Neural network with recurrent connections Analysis of Canonical Correlations, For blood cell categorisation, a single cell patch extraction from blood sample techniques achieved acceptable results. These techniques, however, are unable to handle the issue of numerous cells overlapping. The canonical Correlation Analysis (CCA) approach was utilised to overcome this issue95.89% accuracy is achieved using the proposed methodology.

Jianwei Zhao et al. [15] proposed that automatic convolutional neural networks were utilised to locate and categorise leukocytes. Zhao proposes automatic recognition of WBCs obtained from perfusion pictures without manual operation. The ALL-IDB database provides a better detection method than the iterative threshold method at a lower cost. Zhao's proposed method has limitations, such as not detecting all WBCs in peripheral images.

Table 1: Existing Techniques Analysis

S.No	**Author**	**Algorithm Used**	**Demerits**	**Accuracy**
1.	Afrin Alma	Convolution Neural Network With Reinforcement Learning Method	In processing raw photos features hurriedly, this model also has limitations	95.54%
2.	Jianwei Zhao	Automated Convolution Neural Network	such as not detecting all WBCs in peripheral images	86%
3.	Fatih Ozyur	Convolution Neural Network With MRMR-ELM Method	classification while taking more implementation time	96.03%
4.	Sonali Mishra	*Support Vector Machine Classifier*	taking more time to classify the result	96.97%
5.	Sweta	Convolution Neural Network	An extensive scientific investigation is required to evaluate the effect on file size.	97.2%

3. PROPOSED METHODOLOGY

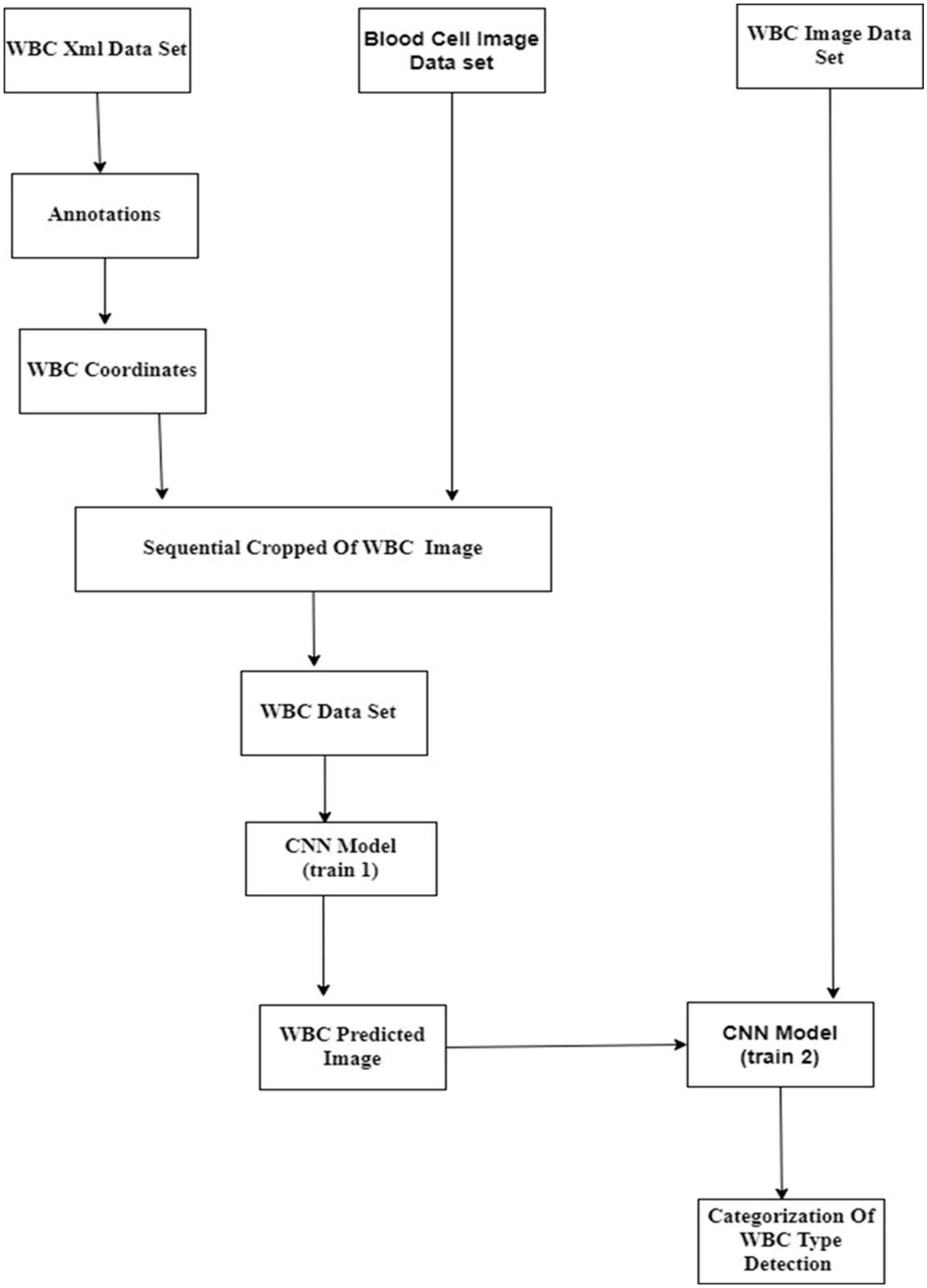

Figure 3: Proposed methodology for detection of WBC type

3.1 Dataset Used for this Proposed Model

The dataset Blood Cell Images is downloaded from Kaggle. It is classified into two types one is master dataset1 and other is master dataset2. In master dataset1, the blood cell images are in the form of JPEG format, and each image consists of one XML file which explains the details of the image. That includes the coordinates of RBC, WBC, and blood platelets. In master dataset 2, the different types of WBC images are present. The different types of WBC present are Eosinophil, Neutrophil, Monocyte, and Lymphocyte. The total dataset consists of 12,500 Blood cell images.

3.2 Creation of WBC Dataset Using Sequential Image Cropping

The Annotation folder consists of all the XML files of Blood cell images. A.CSV files with descriptions of the various blood cell types are created. The CSV file is very easy to find the details of blood cells, and then the extraction of the coordinates of WBC is taken from each image. Then it is given as input to the Sequential Image Cropping. In Sequential Image Cropping, the extraction of WBC is done using the coordinates taken from the CSV file created above. The function used in Sequential Image Cropping is
"Cropped_Image = Image.crop" ("left,top,right,bottom")

3.3 Convolutional Neural Network (CNN)

Convolutional neural networks (CNN) classify the various types of white blood cells by identifying similarities between images. Three convolutional layers are utilised to categorise the photos. These three convolutional layers employ max pooling and dropout layers. The max pooling layer is used to extract the most value from the feature map and to represent the highest feature. Pooling layers come in two different varieties. The dropout layer eliminates unnecessary details from the image before sending only the necessary characteristics to the following convolutional layer. The input layer and the output layer are connected using the fully connected layer (FC layer). The Flatten and dense layers are used in the FC layer. Flatten layer is used for the conversion of 2D arrays into single-dimensional long-running arrays. The dense layer helps to take input from the previous layers. Up to the FC layer, the Relu activation function is used. Finally, the output layer, the type of WBC, can be determined. The softmax activation is used in the output layer. The CNN algorithm contains two main processes known as convolution and sampling. In the convolution layer, it creates an activation map through filters. In sampling, it downsamples the output of the convolutional layer and produces the class scores through which the white blood cell is predicted.

Algorithm for CNN:

```
Begin
  for each epoch n=1,...,N:
    for each batch (X_batch, y_batch) in the training set:
      x←Conv2D (16, 3, padding=' same', activation = relu, input_shape=(80,80,3) )
      y←MaxPoolin2D ( )
      x←Conv2D (32, 3, padding =' same', activation = relu)
      y←MaxPoolin2D ( )
      z←Dropout (0.25)
      x←Conv2D (64, 3, padding = ' same', activation=   relu)
      y←MaxPoolin2D ( )
      z←Dropout (0.25)
      f←Flatten ()
      d←Dense (128, activation = "relu")
      z←Dropout(0.25)
     d←Dense(128, activation = "softmax")
```

End for
End For
End

The above CNN algorithm is used twice in the proposed methodology for the first time; the input given is the WBC data set. Using CNN, the similarities between the images are found. After finding the image similarities, the output is again given as input for the second time CNN. By the second time CNN, the classification of WBC is done. The output of the second time performing the kind of WBC will be determined by CNN.

4. RESULT ANALYSIS

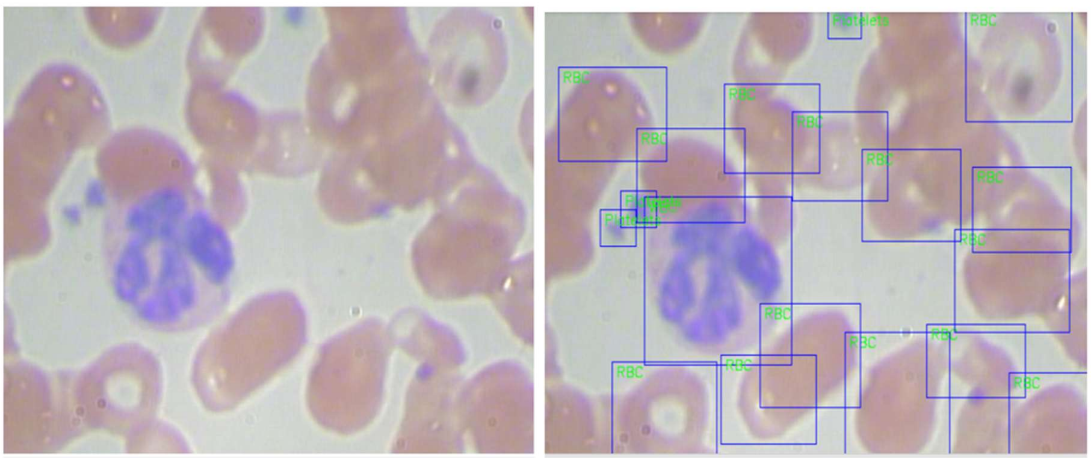

Fig a.Before segmentation **Fig b.After segmentation**

Figure 4: Before and after segmentation of the WBC image using the box function

Fig (a) is the original image of the WBC in Fig 4. After that, using the box function for the original image to detect the white blood cell in the form of a rectangle box not only detects the WBC but Also detects the red blood cells and platelets, it shows in Fig (b) in Fig 4. These cropping results can be observed clearly in the below images. It can explain the different types of blood cells present in the above image Fig a. Before segmentation.

Fig a. RBC **Fig b.** WBC **Fig c.** Platelets

Figure 5:cropped images of RBC, WBC, and Platelets

The above images fig a. RBC, fig b.WBC, fig c.Platelets in Fig 5 are cropped from the above image; as a result to describe types. This cropping can be done using the coordinates given in the XML of each image. Those coordinates are given as input to the rectangular box model and extracted from the images. From that images, only WBC images are collected and created a new dataset which is used for the classification of the type of WBC using the CNN.

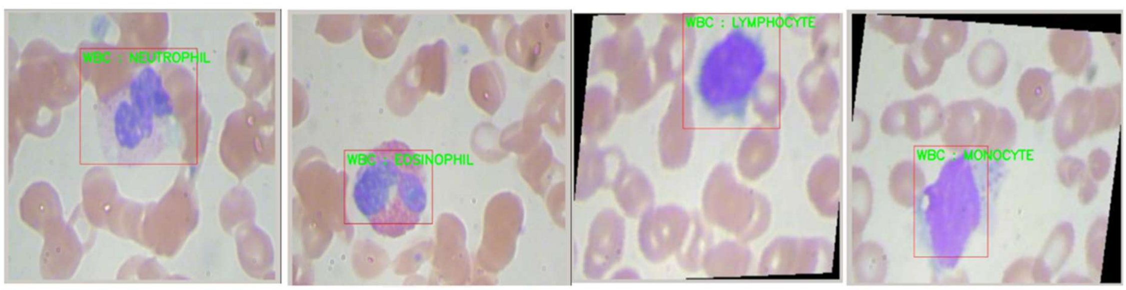

Fig a: Neutrophil **Fig b:** Eosinophil **Fig c**: Lymphocyte **Fig d:** Monocyte

Figure 6: Types of WBCs

Fig a: Neutrophil, Fig b: Eosinophil, Fig c: Lymphocyte, Fig d: Monocyte in Fig 6 Types of WBCs are the CNN classification results. It describes the type of WBC, and it can be further extended in future to detect cancer in blood cells.

Different Techniques are previously used in the existing models. In those techniques, some valuable techniques are chosen in our technique. But there are some difficulties they faced, like low accuracy, and this difficulty can be achieved by this CNN Segmentation. This model aims to escalate the accuracy of the segmented image. Finally, this is achieved by our technique, and it is observed in the following Table 2 and Fig 7 below.

Table 2: Comparison of the Existing Model with Our Model

S.No	Author	Accuracy
1	Convolution Neural Network With Reinforcement Learning Method	95.54%

2	Automated Convolution Neural Network	86%
3	Convolution Neural Network With MRMR-ELM Method	96%
4	*Support Vector Machine Classifier*	96.97%
5	Convolution Neural Network	97.2%
6	Convolution Neural Network	99.191%

In Fig 7, the CNN Model achieved greater accuracy than other existing techniques. The methodology is represented by the x-axis, and accuracy is represented by the y-axis. This bar graph can explain the benefits of our model. The CNN model is trained with many images with suitable class labels. These are given as input to the neural network. Now each image is trained with a random value that is assigned and makes a comparison with class labels of the input image and achieves an accuracy of 99.325%. The outputs from the neural network are tested against our collection of objectives with a 99.729% accuracy rate. As the culmination of our project, the suggested method aids in determining the type of white blood cell with an accuracy of 99.191%.

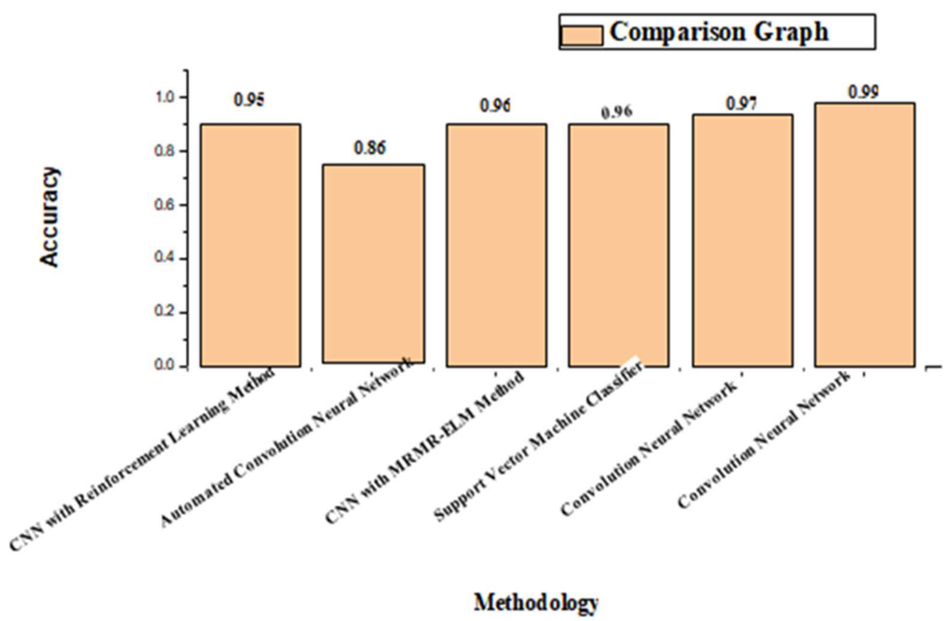

Figure 7: Graph for comparison of existing systems and proposed system

5. CONCLUSION AND FUTURE SCOPE

The classification of the many forms of WBC is the paper's primary subject. To identify the type of WBC, scientists performed several processes on a collection of white blood cell pictures. The following picture cropping technique created a dataset of WBC later using CNN to classify the type of WBC trough from the data set. One of the significant benefits of our proposed model is that it reduces complexity and improves efficiency. Finally, the model accuracy after construction is 99.191%. The future scope involves the real-time implementation of our proposed model.

6. REFERENCES

- Afrin Alma proposed A convolutional neural network–based learning approach to acute lymphoblastic leukaemia detection with automated feature extraction, 2020, http://doi.org/10.1007/s11517-020-02282-x.
- Jing Chang proposed an Automated Classification Of Leukocyte Using Deep Neural Networks, 2017, http://10.1109/ASICON.2017.8252657
- Fatih Ozyur Proposed A fused CNN model for WBC detection with MRMR feature selection and extreme learning machine,2020, http://10.1007/s00500-019-04383-8
- Sonali Mishra Proposed A GLRLM-Based Feature Extraction for Acute Lymphoblastic Leukemia (ALL) Detection Using CAD Systems and SVM classifiers, 2017, http://hdl.handle.net/2080/2727
- Sweta et al. proposed that Convolutional Neural Networks are used to detect white blood cancer from bone marrow microscopic images. 2020, http://10.1109/ACCESS.2020.3012292
- Amjad Rehman Proposed a Classification of acute lymphoblastic leukaemia using deep learning, 2018, http:///10.1002/jemt.23139
- Sarmad Shafiqu proposed Acute Lymphoblastic Leukemia Detection and Classification of Its Subtypes Using Pre-trained Deep Convolutional Neural Networks, 2018, http://10.1177/1533033818802789
- Mohd Shafry proposed a Feature Extraction of White Blood Cells Using CMYK-Moment Localization and Deep Learning in Acute Myeloid Leukemia Blood Smear Microscopic Images, 2022, http://10.1109/ACCESS.2022.3149637
- Huang proposed a Blood Cell Classification Based on HyperspectralImaging with Modulated Gabor and CNN, 2019, http://10.1109/JBHI.2019.2905623
- Reena proposed Localization and recognition of leukocytes in peripheral blood: A deep learning technique,2020,http://10.1016/j.compbiomed.2020.104034
- Khamael Proposed a Segmentation of White Blood Cell, Nucleus and Cytoplasm in Digital Hematology Microscope Images, 2020, http://10.1109/RBME.2020.3004639
- Mishra proposed Texture feature based classification on microscopic blood smear for acute lymphoblastic leukaemia detection, 2019, http://10.1016/j.bspc.2018.08.012
- Abel Worku Tessema, Mohammed, proposed Quantitative analysis of blood cells from microscopic images using convolutional neural network, 2021, http://10.1007/s11517-020-02291-w
- Patil Proposed a White Blood Cells Image Classification Using Deep Learning with Canonical Correlation Analysis, 2020,http://10.1016/j.irbm.2020.08.005
- Jianwei Zhao proposed A Automatic detection and classification of leukocytes using convolutional neural networks, 2017, http://10.1007/s11517-016-1590-x.
- Famila, S., Arulmurugan, A., Mahendar, A., Kalaiyarasan, R., Supriya, N. and Maridu, B., 2024. Transfer learning based feature extraction with metaheuristic optimization algorithm for detecting gastric cancer using optoelectronic sensor in endoscope. Optical and Quantum Electronics, 56(4), p.509.

- Mahendar, A., and Dr K. Shahu Chatrapati. "Detection and Prevention of Cyber Attacks on Cloud-Based Data Centers using Machine Learning." International Journal of Computing and Digital Systems 12.1 (2022): 1063-1070.