International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-2, December 2019

WBC Analysis using Data Augment Method and Convolutional Neural Network

N. Banupriya, A. Jasmine Gilda, B. Jaison, T. Sethukarasi

Abstract: ***WBC is a White Blood Cell or White Blood Corpuscle also known as Leucocytes. The normal count of WBC ranges from 4000 to 11000/mm3. It plays a vital role in the human body. Many diseases in human start with the abnormal balance of WBC, which acquire the part of immunity. To have adequate knowledge about WBC, we have to have a clinical test like blood count test which gives the count of RBC, hemoglobin, WBC, etc. RBC is otherwise known as Erythrocyte and it does not have a nucleus, with pigment hemoglobin. Due to the presence of this pigment, blood is red in color. In RBC, O2 and CO2 are transported in and out of the tissues. Recent research explains about diseases like cancer, allergy, breast cancer, etc are caused due to lack or abnormal WBC. This comes with the solution of finding the count of WBC in two types: Manually and automated way. In our paper, we are concentrating on collecting the WBC count using the Data augmentation method and Convolutional Neural Network.The Quality of image is improved in comparison with number of augmented images. This explains that we have 12500 sample images in the dataset in which 9957 samples are trained, validated on 2487 samples and training accuracy is high with increasing epoch value.***

Keywords**: *WBC, Leucocytes, Data augmentation, CNN***

I. INTRODUCTION

The human body immunity can be explained through different biological concepts. The immune system protects the human body against various diseases. The main part of the immune system is WBC or leukocytes [18] WBC has a nucleus and there is no presence of hemoglobin. There are two types of WBC: Polymorphonuclear WBC and Mononuclear WBC. Polymorphonuclear or Granulocytes contains chromophilic granules in the cytoplasm. Its life span is less therefore one billion cells are produced in a day. Types of Granulocytes are Neutrophil (50%-70%), Eosinophil (2%-4%) and Basophil (<1%). Agranulocytes contained no granules in the cytoplasm. Types of Agranulocytes are Lymphocytes (20%-40%) and Monolcytes (2%-8%).

Revised Manuscript Received on December 30, 2019.
***** Correspondence Author
N. Banupriya, Assistant Professor, Department of CSE, R.M.K. Engineering College, Kavaraipettai. Email: nbp.cse@rmkec.ac.in.
A. Jasmine Gilda*, Assistant Professor**,** Department of CSE, R.M.K. Engineering College, Kavaraipettai. Email: ajg.cse@rmkec.ac.in.
B. Jaison, Associate Professor, Department of CSE, R.M.K. Engineering College, Kavaraipettai. Email: bjn.cse@rmkec.ac.in.
Dr. T. Sethukarasi, Professor & Head**,** Department of CSE, R.M.K. Engineering College, Kavaraipettai. Email: hod.cse@rmkec.ac.in.

A.Features of WBC

WBC (Fig.1.) contains a nucleus which exhibits amoeboid movement (crawling like).

Monocytes and neutrophils are crawling to search pathogens (a virus causes diseases). Leucocytes enter tissues in a rapid manner and it is achieved by margination (attaching to other WBC to the walls of damaged blood vessels) and diapedesis (passage of blood cells present in capillary walls into tissues). Generally, WBC posses safety from external disease. They produce chemicals that act as a defense to the body and increase immunity.

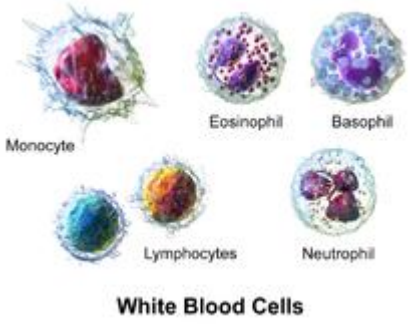

Fig 1. WBC

B. WBC types and its variations.

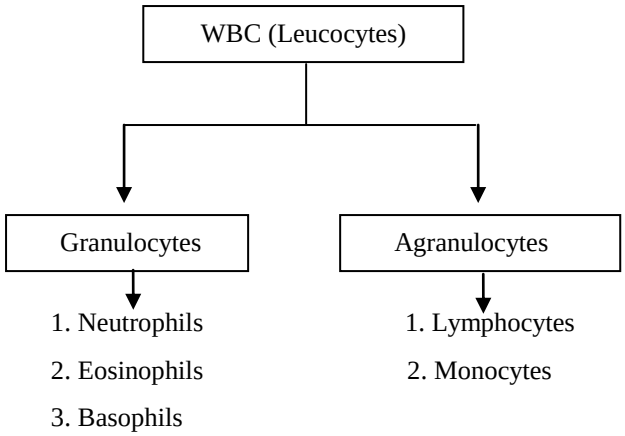

Fig - 2. Types of WBC

Fig. 2. represents the types of WBC [15]. Fluctuations of leucocyte count are from high to low. It is seen in both physiological and pathological conditions. When there is an increase in WBC count (i.e.) above 11,000 cells/mm^3, then the case is Leucocutosis, caused when emotions are high, muscular exercise, injections for adrenaline, etc.

A decrease in WBC count (i.e.) below 4,000 cells/mm^3, then the case is Leucopoenia, seen in starvation, diabetes mellitus, and pyogenic infections, all fever except typhoid and viral.

Neutrophil causes Hypomotility, which is gastrointestinal decreased and abnormal decreased motality.

Due to defective diapedesis of leucocytes, it is unable to produce O2 and inability to kill external bacteria; an increase in number causes Appendicitis, Pneumonic, Whooping cough, etc. And a decrease in number, seen in X-ray radiations, leads to Typhoid, drug toxicity etc.

Eosinophil destroys infest parasites. It produces immunity to the body in the respiratory tract, GIT and urinary tract. Increase leads to Asthma, fever, parasitic infection, Allergic, etc. Decreased count leads to ACTH injection, Cushing syndrome.

Lymphocytes produced by bone marrow, spleen, liver. Its life span varies from few days to few months (i.e.) 4 days to several months. B-lymphocytes survive for shorter periods (20%-30%), T-lymphocytes survive for longer periods (60%-70%), and NK cell's life span is not known (10%-15%). It Increases immune tolerance. An increase in number leads to all chronic diseases, Tuberculosis, Urinary tract infection, children with or without infection. A decrease in X-radiation, Cushing Syndrome, ACTH treatment, AIDS, etc.

Monocyte enters the blood stream and it is produced in the marrow. Since their movement is continuous in the blood to the tissue, the life span is not exactly counted. Produce growth factory, complement factors act as Antigen Presenting Cells (APC), immunity growth. An increase in leucocytes caused by cancer results in Leukemia. These types of WBC are randomly increasing and decreasing due to daily habitat, loss of physical exercise, food, genetics, etc.

The main objective is the types of WBC are taken and Basophil is eliminated due to its less consumption in the blood (i.e.) <1%. Neutrophil, Eosinophil, Lymphocytes, Monocytes are taken from the Blood samples which is maintained in the dataset. The samples of 9957 are trained by CNN with 350 epochs and it is trained by the augmentation method to improve accuracy. This accuracy helps to diagnose the patient's disease which is maintained in the medical reports for further treatment.

II. LITERATURE SURVEY

The problem deals with the count of WBC (i.e.) increase and decreases of Leucocytes cause various types of cancer. [1] Amber Yates explains various diseases due to the lack and rise of WBC with various symptoms like mouth ulcers, fever, etc. Classifying leucocytes is a big deal in a manual way with a microscope. In [2], the concepts are based on the segments of WBC with Nucleus and Cytoplasm. It involves automatic differential counting for the diagnosis of various diseases. Mainly it concentrates on the appearance of nucleus and cytoplasm using image processing techniques on bind with the application Self-dual Multiscale Morphological Toggle (SMMT). The technique used for nucleus segmentation is watershed transform and level set methods and techniques used for cytoplasm region are granulometric analysis and morphological transformation. The work can be carried out in the future to segment the WBC in appearance, Qualities, various properties and enhance to various new diseases with this paper.

[3]WBC is diagnosed automatically using segmentation techniques in microscopic images that extract the source from fresh blood smears. 60 microscopic images taken for testing and 92% accuracy for segmentation in the cytoplasm are obtained and 89% accuracy for segmentation in a nucleus is obtained. [4] This paper deals with counting of WBC, RBC, platelets. SVM with standard intensity and histogram features are obtained. PCA is used to extract the features and CNN tasks the images as input. Comparison of SVM and CNN are also evaluated in this paper. Blood Cell Count is accomplished by various methodologies [5]. RBC, WBC, Platelets, hemoglobin count is analyzed. Various segmentation techniques like mathematical morphology, feature extraction, enhanced threshold-based technique, contour feature point tagging are discussed.

In paper [6], it implies how Deep Learning is dealing with pathology traditional problems. It provides a range of metrics towards complete blood count (CBC). The Blood count can be in an automatic and manual way. [6][9] In earlier 1950, Coulter Counter invented the automated way in which electricity is used to classify and count WBC. Later Laser flow Cytometers have emerged as an alternative both costs around tens of thousands of dollars. The manual approach is viewing through microscopic by the pathologist.

The machine learning model[16] which is used here is Deep Learning Simple CNN to classify the WBC image into Polynuclear or mononuclear.71 images are tested for accuracy with (20% of the dataset) original data's Distribution.98.6% classification task is obtained, trained for 20 epochs. Table-I Refers to the dataset used in [6] to analyze the types of WBC.

Table-I: Dataset taken for evaluation.

Details	Total taken
Dataset	352 to 10,000 images
Size	640*480
Lymphocyte	33
Monocyte	21
Neutrophil	207
Eosinophil	88
Basophil	3

High throughput microfluidic [7], the WBC separation platform achieving the priority of 48% after two consecutive processes. The two processes are separating polystyrene leads in serpentine channel and separation properties of junket cells spiked in the blood to increase the purity of WBC. In [8], pattern net fused ensemble of CNN is used to classify while blood cells 12500 with 50*50 augmented images are analyzed with this model.

With reference to paper [1], [9], a new solution is explained using the Deep learning Studio (DL Based) with a dataset of around 12,500 augmented images which is grouped into 4 different types of cells called Eosinophil, Lymphocyte, Monocyte, Neutrophil with 3000 images in each type. Simple Neural Network is built to classify into twelve classes using DL Studio.

DLS achieved 96% accuracy on validation with a given set of data involved to identify and characterizing patient blood samples. With one step forward, [10] this paper focuses on computation power and unavoidability of datasets for training problems.

Here comes the solution as Kaggle (home of Datasets), Google Colab.

Google Colab is a free cloud service that allows us to use Tesla K80 GPU where we can train our model with predefined datasets and can create a dataset using Keras, Vfjs. The dataset with 12500 images [9] are taken, image size reduced to 80*80*3 to train. CNN plays a better role when compared with a simple feed-forward network to learn the unique features.

[11] Deals with the programming language Python to calculate WBC, RBC, Platelets count. In addition to this, the Blob detection algorithm is used to detect and find the difference between RBC and WBC. This accomplishes the results with RBC-96.32%, WBC – 98.5%, Platelets – 100% with minimum epochs. A deeper Filter of CNN in paper [12] requires CD68 samples taken with 16*16 average sizes and variance. 10X Objective image is taken with 7*7 which is half of the sample size. We derive Optimized CNN Architecture. LI ot su Thershold method is used to give an accurate calculation of segmentation results. Post-process segmentation results are taken with hybrid masks. In automatic segmentation, the system gives accuracy in detecting WBC using microscopic images. LI ot su., represents microscopy image segments as a block by block.10X objective images are taken for classification.CD68 macrophages, 7*7 Convolution filter.

The concepts discussed in [13] is, 350 microscopic blood smear images are taken and the ML algorithm is compared for the classification of WBC. Multinomial Logistic Regression (MLR) algorithm gives better performance with 95% test success and uses an automatic calculation of WBC. [14] deals with skeletal injury diseases. The main cause is increment or decrement of WBC or certain proteins visualized through manual diagnosis (i.e.) through a microscope that occurs after injury. DL method used hear is Deep Quantify in the analysis of injured and normal WBC from the muscle of female.

LI ot su[12] threshold method and two-layer CNN classifier are used to obtain accuracy in the level of threshold, masking and WBC classification accuracy level is 90.64% and 89% for CD68 positive macrophages and 7/4 positive Neutrophil images with (400*400) pixel size are evaluated. 10X objective, 100X magnification images taken as source image. It gives more study in the injured area using deep quantify than deep segmentation architecture.

Table-II Shows the comparison of the Developed model using DL with samples used and results produced. Epoch's variation is obtained when we increase or decrease. In crisp, Epoch is feeding the input multiple times to reduce error. Normalization is also performed to stabilize the input for each mini batch layer. An Example can explain the epoch actual mean, 2000 samples taken with Batch size = 500 then no. of iterations will be 4 for one epoch.

Table-II. The comparison of Model developed for CBC.

S.No	Model	Dealing with	No. of Samples	Accuracy results
1	Athelas (DL-Simple CNN)[17]	WBC and Neutrophil	71 images with dataset having 352 to 10000 images.	98.6% with 20 epochs
2	DL Studio(DL-Simple CNN)	Classify esophill, lymphocyte, Monocyte, Neutrophil	Some datas tested from 12500 images	96%
3	Keras and Tfjs (DL-Simple CNN)	Computational power and unavailability data set	Took data from 12,500 images and size reduced to (80*80*3) size	92.67% with 30 epochs
4	Python OpenCv Programming Language(DL)	WBC, RBC, Platelets	Each 10 images taken processed from dataset	RBC-96.32%, WBC-98.5%, Platelets-100%
5	High throughput Separation platform.	Purification of WBC	288 ml/h for diluted blood and 144ml/h undiluted blood	Purification obtained by 48%
6	PECNN	WBC	Trained with 30 epochs	99.9%

In Table-II each model is analyzed with certain dataset and the results are obtained around 96 % to 99.9%. But the input samples are taken with limited count and less than 40 epochs. To improve more accuracy with high epochs, we go for our proposed model.

III. PROPOSED MODEL

Nowadays we have many handheld smart devices to check our BP, heart rate, etc example Smart Watches, without having the assistance of a hospital or doctor. So maintaining our body physically and mentally is a big deal in the real world. Even many devices come forward to analyze our immunity through WBC, it keeps challenging in counting, visualizing the image through a microscope. Generally, some real samples are fetched from the person to research the development of bacteria is done and the immediate remedial measures are taken by drugs. In this paper, we are dealing with the count of WBC efficiently using DL and Data augmentation [19] CNN concepts. It also deals with the quality of the image or data.

A.Brief

Here we combine simple CNN with the data augmentation method. A set of input images is given to train using CNN and the trained images are given to the data augmentation model. The method involved in data augmentation is the creation of own data or random data to decrease the problem of quality of the image of data. We have two sets of batches that have 2 Conv2D and MaxPool in a single batch. Then it is normalized using Batch Normalization. These produced samples are applied to the augmented model, in which random images are generated. The importance of creating the random is to obtain the accurate value of WBC.

The word random refers to many directions of images. When samples are fed in a manual way, it cannot be always in the same direction. So when we go for the augment method, many directional images with 40^{o}, $80^{o,}$ etc can be generated randomly. This helps to train our model and give an efficient output.

There are three types of augment method in CNN. First, Data set generated, then expanding an existing dataset, finally producing the output. Second, inplace/ on-the-fly data augmentation. In this method we generate random data set completely and it is trained. Third, coding data set generation and inplace augmentation. Our model comes under this category. Fig.3 implies the architecture of proposed model.

Dense Layer can be increased to get more accuracy, this achieves hyper parameter concept. Our data set have 12500 images with four classes esophill, lymphocyte, Monocyte, Neutrophil of each 3000 images. We train all types of WBC in order to find the outcome result of disease, occurred or not. Out of this we took 9957 images to train using CNN model then the trained set is given for data augmentation. The obtained output will be checked for the accuracy later.CNN model is trained with 350 epochs and gives the accuracy of 99.86 and augment model is trained with 200 epochs which give approximately 89.81%.This values are analyzed by changing the epoch, hyper parameter etc.

B. Algorithm

- **Simple CNN Model**
  - Input is given for training to the model
  - Simple CNN is developed with 2 dense layers, 1 dropout layer, Batch1 and Batch 2.
  - In our model Batch contains 2Conv2D and one Pool layer.
  - Trained data using Simple CNN obtained in second Dense Layer and given for augmentation process.
- **Data Augmentation Model**
  - It accepts group of images used in training
  - Each image with random transformation is made from this batch of images.
  - It replaces the original batch with the obtained (new).
  - Training is made with Obtained batch.
  - Final output with more accuracy is obtained

C. Architecture

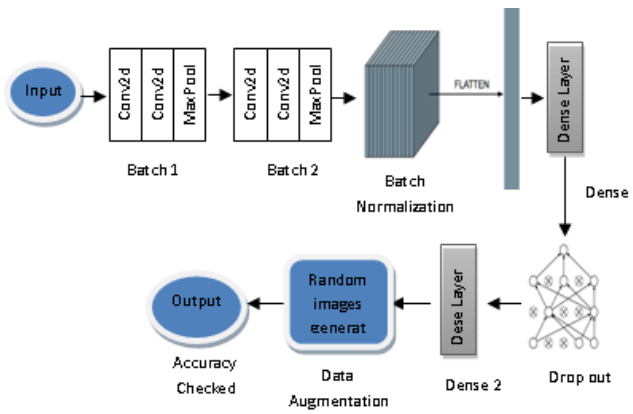

Fig.3 Architecture of Proposed model.

IV. RESULTS AND DISCUSSION

The existing model with 300 epochs gives the accuracy of training 99.6% and validation accuracy did not improve from 58.44%. Since validation is done in less efficient. In the proposed model, We increased the epoch value to 350, Training accuracy is 99.86% and validation accuracy reached 72.618%. This achieves the exact count, improved quality of image, etc to diagnose the patient in critical condition, cancer, etc.

Fig. 4 Gives the Training and Validation loss and Fig. 5 gives the Training and validation accuracy for the existing model with 300 epochs. Fig. 6 Gives the Training and Validation loss and Fig. 7 gives the Training and validation accuracy for the existing model with 350 epochs. After augmenting, the results for training and validation for the proposed model is given in Fig. 8 and Fig. 9.

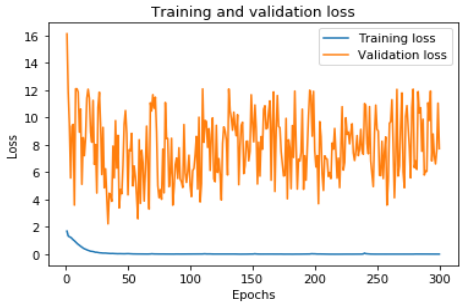

Fig. 4 Training and Validation loss (300 epochs)

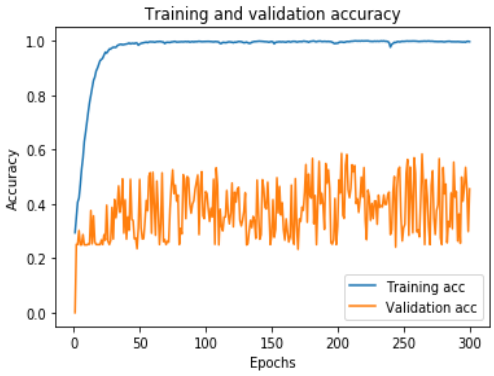

Fig. 5 Training and validation accuracy (300 epochs)

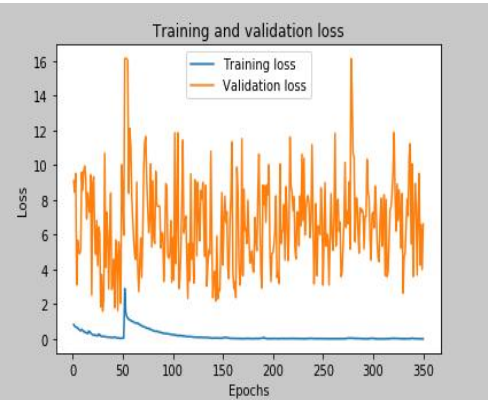

Fig. 6 Training and Validation loss (350 epochs)

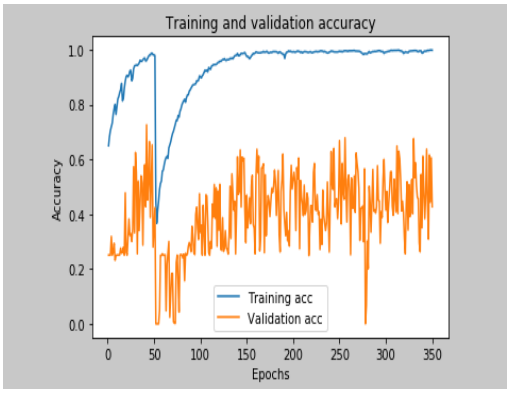

Fig 7 Training and validation accuracy (350 epochs)

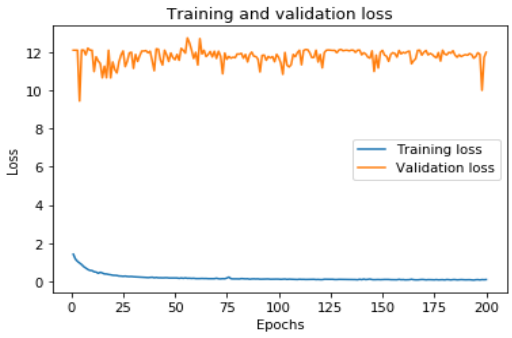

Fig. 8 Training and Validation loss after Data Augmentation

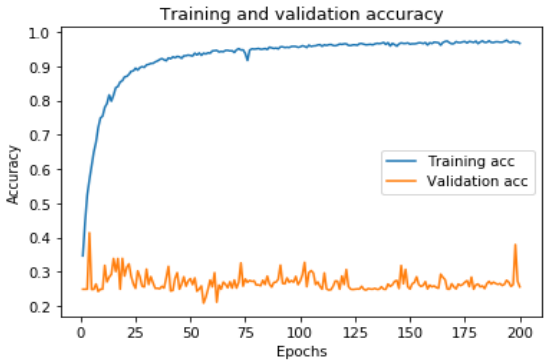

Fig. 9 Training and validation accuracy after Data Augmentation

V.CONCLUSION

We conclude, Types of WBC: Eosinophil, lymphocyte, Monocyte, Neutrophil is trained and the trained image comes out with the decision making of disease occurrence using CNN and augmentation method. First, the data set is analyzed and trained by Simple CNN, Second it is retrained by the Data augmentation method to enhance the accuracy of the image.By improving the accuracy,more number of diseases can be diagnosed and cured in right time.Our model gives the Training accuracy of about 99.86%.

VI. FUTURE WORK

More infected blood samples can be taken and scrutinized with our model. Clinical data like lab reports can also be analyzed for more samples of images This model can be extending by increasing in hyperparameter and data samples which can achieve 'n' number of solutions. It can compare with good samples to exercise the percentage of infection in the blood more accurately and can be carried out for further medical diagnosis, achieves the path for future work.

REFERENCES

1. https://www.verywellhealth.com/white-blood-cell-disorders- overview-4013280.
2. Leyza Baldo Dorini, Rodrigo Minetto and Neucimar Jeronimo Leite, "Semiautomatic White Blood Cell Segmentation Based on Multiscale Analysis", vol. 6, no. 1, January 2007.
3. Firdaus Ismail Sholeh, "White Blood Cell Segmentation for Fresh Blood Smear Images", ICACSIS 2013 ISBN: 978-979-1421-19-5.
4. Mehdi Habibzadeh, Adam Krzy˙zak, and Thomas Fevens, "White Blood Cell Differential Counts Using Convolutional Neural Networks for Low Resolution Images" ICAISC 2013, Part II, LNAI 7895, pp. 263–274, 2013.Springer-Verlag Berlin Heidelberg 2013.
5. Mr. Dixit Kolhatkar , Ms. Nisha Wankhade, "Detection and Counting of Blood Cells using Image Segmentation : A Review", 2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (WCFTR'16),978-1-4673-9214-3/16/$31.00
6. https://blog.athelas.com/classifying-white-blood-cells-with-convolutional-neural-networks-2ca6da239331.
7. Jun Zhang, Member, IEEE, Dan Yuan, Ronald Sluyter, Sheng Yan, Qianbin Zhao, Huanming Xia, Say Hwa Tan,Nam-Trung Nguyen, "High-Throughput Separation of White Blood Cells From Whole Blood Using Inertial Microfluidics", 2017.
8. Justin L. Wang, Anthony Y. Li, Michelle Huang, Ali K. Ibrahim, Hanqi Zhuang, Ali Muhamed Al, "Classification of White Blood Cells with Pattern Net-fused Ensemble of Convolutional Neural Networks (PECNN)", 978-1-5386-7568-7/18/$31.00,2018 IEEE.
9. https://medium.com/@rgrgrajat1/classifying-blood-cell-images-with-deep-learning-studio-9695cf0daa7b.
10. https://towardsdatascience.com/building-a-blood-cell-classification-model-using-keras-and-tfjs-5f7601ace931.
11. Raymond Joseph Meimban, Alexa Ray Fernando, Armil Monsura, Jonrey Ranada, J.C Apduhan, "Blood Cells Counting using Python OpenCV", ICSP 2018.
12. Yang Jiao, Barbara St. Pierre Schneider, Emma Regentova, Mei Yang, "Automated Quantification of White Blood Cells in Light Microscopy Muscle Images: Segmentation Augmented by CNN", ICVISP 2018, ACM ISBN 978-1-4503-6529-1/18/08.
13. Abdullah Elen,Kamil Turan, "Classifying White Blood Cells Using Machine Learning Algorithms", Cilt/Volume:11 Sayı/Issue:1 Ocak/January 2019.
14. Yang Jiao, Barbara St. Pierre Schneider, Emma Regentova, Mei Yang," Deep Quantify: deep learning and quantification system of white blood cells in light microscopy images of injured skeletal muscles", doi: 10.1117/1.JMI.6.2.024006.
15. Qiwei Wang, Shusheng Bi, Minglei Sun, Yuliang Wang, Di Wang, Shaobao Yang, "Deep learning approach to peripheral leukocyte recognition", https://doi.org/10.1371/journal.pone.0218808.
16. Laura Bigorra,Iciar Larriba,Ricardo Gutiérrez-Gallego,"Machine learning algorithms for the detection of spurious white blood cell differentials due to erythrocyte lysis resistance", doi:10.1136/jclinpath-2019-205820.
17. George-Gay, Beverly, and Katherine Parker. "Understanding the complete blood count with differential." Journal of Peri Anesthesia Nursing 18.2 (2003): 96-117.
18. L.Prakasam reddy, "Fundamentals of Medical Physiology",6th edition, Paras Medical Publisher.
19. Jun Ding, Bo Chen, Member, IEEE, Hongwei Liu, Member, IEEE, and Mengyuan Huang, "Convolutional Neural Network With Data Augmentation for SAR Target Recognition", http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

AUTHORS PROFILE

Ms. N. Banupriya, B.Tech, M.E, Assistant Professor (Computer Science and Engineering Department), R.M.K. Engineering College has been the mentor for the project entitled "Web portal for single window clearance system for providing online approval for film shooting permission " conducted in Delhi (Smart India Hackathon-2018).She has also assisted several other mini projects like "Man-Animal conflict" and "Business analyzer for food kart". Area of interest includes Machine Learning and Data Analytics. Published paper in "International Journal of Computer Aided Engineering and Technology" titled "A close scrutiny of dApps and developing an E-voting dApp using Ethereum Blockchain" during September 5& 6, 2018. Published Paper in "International Journal of Engineering Research and Technology" titled "Survey on Ranking Concepts and Text Mining Algorithms" during September 2019(Volume 8,Issue 9,September 2019).

Ms. A. Jasmine Gilda is working as Assistant Professor in the department of Computer Science and Engineering at R.M.K. Engineering College, Chennai. She received hers Masters degree in Computer Science and Engineering from Anna University, Chennai in 2012. Her research interest includes Medical Imaging, Cloud Computing and Blockchain. She is a certified as Java Programmer through the Digital Skills Readiness Program by Wipro. Acted as mentor in the Smart India Hackathon - 2018 held at Udaipur.

Dr. B. Jaison received his Masters degree in Computer Science and Engineering from Anna University, India in 2007. He completed his Ph.D degreein Computer Science and Engineering from Anna University in the year 2015 and currently working as Associate professor in the Department of Computer Science and Engineering, RMK Engineering College, Chennai. He has published more than 20 papers in international journals and conferences. He is the recognized Research Supervisor of Anna University, Chennai and currently guiding 7 Ph.D scholars. He has produced one Ph.D. scholar under his guidance. His research interests include Data mining, Bioinformatics and Knowledge discovery. He is a life member in IAENG, IACSIT and ISTE.

Dr. T.Sethukarasi is currently working as a Professor and Head in the department of Computer Science & Engineering, R.M.K. engineering College, Chennai, TamilNadu, India. She received her Ph.D., from the Faculty of Information & Communication Enginereering , Anna University , Chennai, TamilNadu, India. Her area of research includes Data Mining, Soft Computing, Big data Analytics, Sensor Networks and Network Security. She has published more than 25 papers in the international journals and conferences. She is the recognized Research Supervisor of Anna University, Chennai and currently guiding 9 Ph.D scholars'. She produced 2 Ph.D., scholars. She is a life member in IAENG, IACSIT, IEI and ISTE.

.