International Journal of Latest Research in Engineering and Management (IJLREM)
ISSN: 2456-0766
www.ijlrem.org ǁ Volume 06 Issue 07 ǁ July 2022 ǁ PP 01-09

Blood Cells Classification using Machine Learning and Deep Learning

Muhammad Abuzar Anwar[1], Dr. Eylem Gulce Coker[2]
[1](*Electrical and Electronics Engineering, Istanbul Aydin University, Turkey)*

Abstract: Leukocytes, also known as white blood cells (WBCs), are an essential component of the immune system that protects the body against infectious illnesses, germs, and viruses. The classification of white blood cells is often used to identify disorders such as AIDS, leukemia, myeloma, and anemia. A significant number of diverse samples, including various forms of Leukocytes, related sub-types, and blood concentration, may lead to complications, making the examination susceptible to human error. In this study, one of the most common neural networks, the convolutional neural network (CNN), is used to identify several kinds of white blood cells, including eosinophil, lymphocyte, monocyte, and neutrophil. We want to identify a quick and effective classification process and collect data on the distribution of white blood cell evidence, which would eventually facilitate the diagnosis of blood-related disorders. Here, we investigate deep learning advancements in white blood cell classification, concentrating on publicly accessible microscopy image datasets of blood samples.
Keywords: White blood cells, Res2Net, deep learning, classification

1. Introduction

Deep learning utilizes many architectures but this study focuses on Convolutional Neural Networks (CNN), a deep learning model. White blood cells, also known as leukocytes, are an essential component of the body's defense against pathogenic organisms and foreign chemicals (the immune system). To appropriately protect the body, a considerable number of white blood cells must receive a message that an infectious organism or foreign material has infiltrated the body, travel to the site where they are needed, and then destroy and digest the dangerous organism or substance.

Traditionally, white blood cell counts are performed in the laboratory using a staining procedure and manual observation under the microscope. An intriguing option is the non- destructive classification method, which relies on pictures of white blood cell kinds for learning the classification issue.

1.1 Objectives of the Study

The main objectives of this thesis are as follows:

a) An end-to-end equipped deep neural network for the automatic classification of leucocytes into five categories: neutrophils, eosinophils, basophils, lymphocytes, and monocytes.
b) The exploration of a host of deep neural network systems with pre-equipped standards for enhancing the performance of classification.
c) Dataset acquisition and simulation analysis.

2.2 White Blood Cell Disorders

Like all other blood cells, white blood cells are primarily formed in the bone marrow. The number of white blood cells in a given blood volume is expressed as cells per micro liter of blood. The proportion of each of the five major types of white blood cells and the total number of cells of each type in a given volume of blood can be determined through laboratory tests.

Too few or too many white blood cells indicate a disorder. Some white blood cell disorders involve only one of the five types of white blood cells.

1. Lymphocytic leukocytosis is a nab normally high number of lymphocytes
2. Lymphocytopenia is an unusually low number of lymphocytes
3. Neutropenia is an unusually low number of neutrophils
4. Neutrophilic leukocytosis is an abnormally high number of neutrophils

Other disorders may involve a few types together or all five white blood cell types.

1.3 Leukopenia

Some of the white blood cell disorders associated with leukopenia include:

a) Aplastic anemia: A rare condition in which the body stops producing enough new blood cells.
b) Autoimmune neutropenia: A condition in which your immune system mistakenly attacks and

destroysneutrophils.

c) Congenital neutropenia: A genetic disorder in which the body doesn't make enough neutrophils
d) Cyclic neutropenia: A rare genetic disorder in which neutrophil production drops every 21 days.
e) Chronic granulomatous disease: An genetic disorder that causes specific white blood cells to malfunction and behaveabnormally.
f) Leukocyte adhesion deficiencies: A group of rare genetic disorders that affect the white blood cells' ability to fightinfection.

1.4 Leukocytosis

Some of the white blood cell disorders associated with leukocytosis include:

a) Chronic idiopathic neutrophilia: A condition in which neutrophils remain persistently elevated for no apparentreason
b) Hemolytic anemia: A disorder in which red blood cells die faster than they are made, often due to an underlying genetic or autoimmunecause
c) Idiopathic thrombocytopenia: A condition in which your immune system mistakenly attacks and destroys blood-clotting cells calledplatelets
d) Lymphoma: A group of cancers that start in cells of the lymphaticsystem
e) Lymphocytic leukemia: A type of blood cancer that begins inlymphocytes
f) Myeloproliferative disorders: Includes six types of slowing-growing cancers that cause the overproduction of white blood cells (chronic eosinophilic leukemia, chronic myelogenous leukemia, chronic neutrophilic leukemia, essential thrombocytopenia, polycythemiavera, and primary myelofibrosis)

1.5 Conventional Diagnosis of White Blood Disorders

A complete blood count is one of the most often used methods (CBC). The reference range for the total white blood cell (WBC) count might vary from laboratory to laboratory but is commonly stated as follows:

- Males: 5,000 to 10,000 cells per microliter of blood(cells/mL)
- Females: 4,500 to 11,000cells/mL
- Newborns under two weeks of age: 9,000 to 30,000cells/mL
- Children and adolescents: 5,000 to 10,000cells/mL

1.6 Problem Statement

The bone marrow sometimes produces abnormal, immature white blood cells released into the circulation. The presence of these cells causes blood diseases such as acute myeloid leukemia (AML), a malignancy of bone marrow blood-formingcells.

1.7 Significance of theStudy

Different types of leucocytes have distinct purposes. In order to determine if they are present in the right proportions, it is essential to count and identify the amount of the various leukocytes in a blood sample. This quantitative and qualitative analysis of white blood cells gives significant insight into the patient's health state. This procedure makes it possible, for instance, to evaluate individuals for leukemia, immune system anomalies, and malignancies (Shafique et al., 2018). Traditionally, identification is conducted in a laboratory, where blood cell slides are stained with specialized stains or reagents.

2. Methods

Scope of Work

The primary result of this study is the creation of CBC analysis software as a tool for medical blood testing, which allows for high-quality tests and the automatic processing of blood slide pictures to create data needed for diagnosis. We were particularly interested in the clinical classification of the five major categories of white blood cells (leukocytes) and the counting of normal red blood cells(erythrocytes).

Many medical researchers suffer from the fact that significant numbers of samples are difficult to obtain. Kaggle is a vast source of datasets, but we struggled to find a dataset representing several white blood cell disorders.

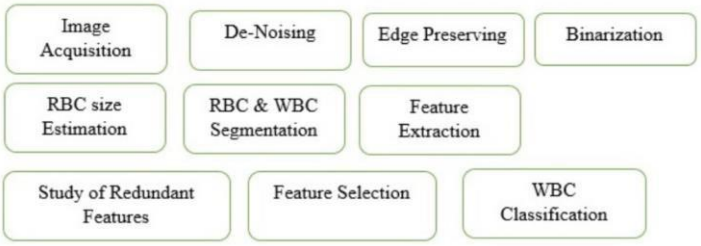


Figure 1 Workflow of WBC classification

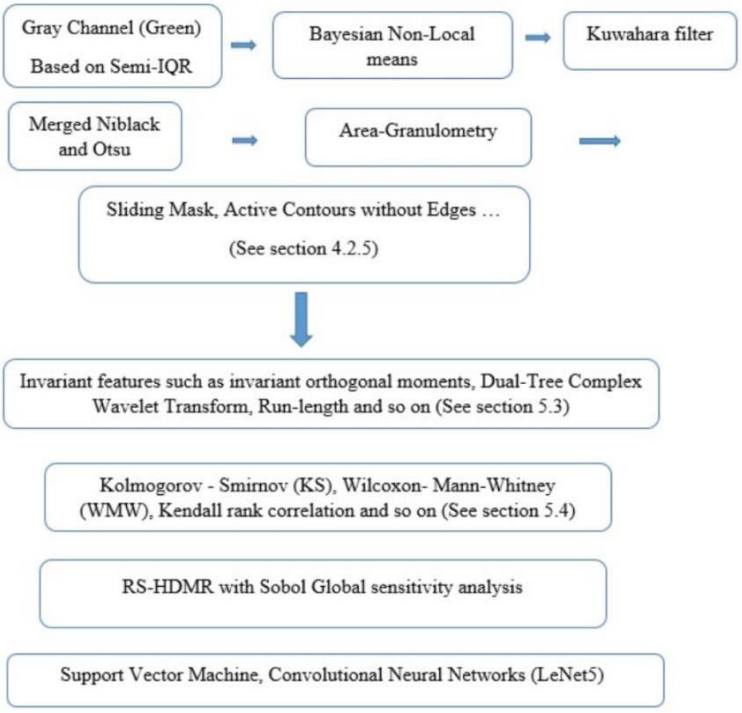


Figure 2 Framework methods of WBC classification
Source: Habibzadehet. al. (2018, April)

Image Acquisition

The Blood Cell Images dataset published by Paul Mooney on Kaggle is used. This collection comprises 12,500 enhanced JPEG pictures of blood cells with cell type labels (CSV). There are around 3,000 photos for four distinct cell types, organized into four separate folders (according to cell type). A digital camera is used to capture images saved on memory cards and transferred to computers as 24-bitmap images, joint photographic experts group (jpeg) images, or movies (Abbas et al., 2018).

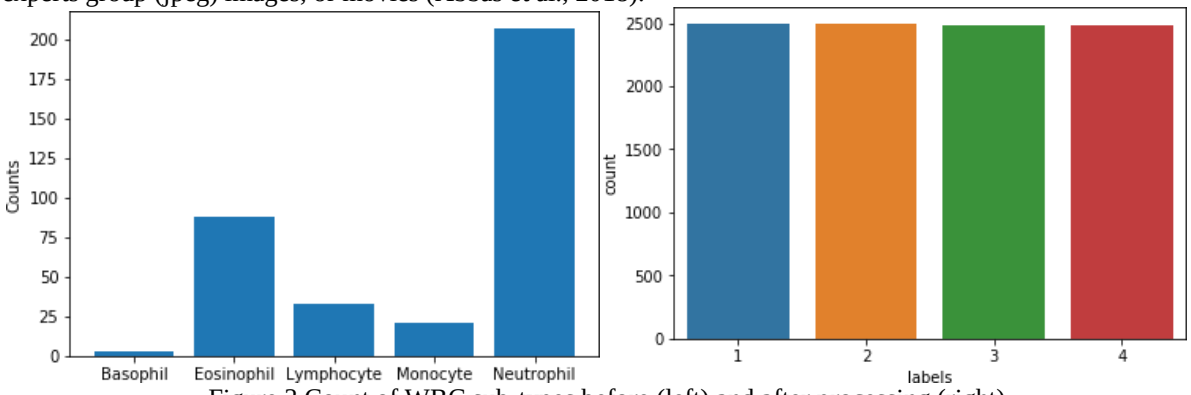


Figure 3 Count of WBC sub-types before (left) and after processing (right)
Source: P. (2018, April 12). Identify Blood Cell Subtypes From Images. Kaggle.

The staining of white blood cells approach is used to increase contrast by altering the color of particular components of the cell structure, allowing for clearer visibility of cell components. Several microscopy stains may be used, and the generic name for these stains is Romanowsky stains. The Romanowsky stain detects malarial parasites in peripheral blood using a methylene blue solution (Carleton, 1980). Jenner, Nocht, Leishman, Giemsa, Wright- Giemsa, and Leishman stains are examples of stains. Giemsa stain, Wright stain, Wright- Giemsa stain, and the Leishman stain are examples of stains in this category used to stain white blood cells. The stains may also reveal granules in certain white blood cellprotoplasm.

Image Pre-Processing

This is concerned with the enhancement or improvement of picture data to overcome undesired distortions, remove noise, or improve some features required for further evaluation in the segmentation and classification phases. This phase also includes visual geometric modifications such as rotation, scaling, and translation.

Image Segmentation

This stage involves detecting white blood cells and their nuclei and cytoplasm. It distinguishes them from red blood cells, background, and plasma in peripheral blood smear images using visual processing and signal processing techniques. Various ways for detecting and segmenting white blood cells have been proposed and integrated with other procedures. These procedures include thresholding approaches, morphological operations, scale-space assessments, edge, and boundary detection, and phase set methods using geometric active contours (GACs). Color spaces such as RGB, CMYK, and HSV with Otsu threshold are examples of modern approaches (Safuan et al., 2017).

Figure 4 Process breakdown of image segmentation
Source: Tyagi, M. (2022, January 6). Image Segmentation: Part 1 - Towards Data Science. Medium.

Isolation of Characteristics

This is an essential step in segmenting and classifying white blood cells. Other isolated attributes involve color aspects such as color distribution and histogram.

Image Classification

It is the process of predicting or assigning a label to an input picture. The dimension with the value 3 indicates three color channels (Red, Green, and Blue).

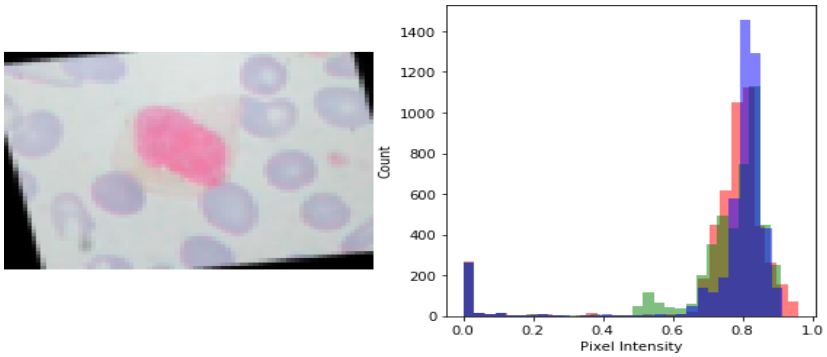


Figure 5 Pixel intensity of an image from dataset
Source: P. (2018, April 12). Identify Blood Cell Subtypes From Images. Kaggle.

$$f(x) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1\le i<j\le n} f_{ij}(x_i,x_j) + \sum_{1\le i<j<k\le n} f_{ijk}(x_i,x_j,x_k) + .. + f_{12\ldots n}(x_1,x_2,..x_n)$$

Where f_0 represents the constant mean effect and, $f_i(x_i)$ represents the influence of variable x_i independently upon the output f(x). Further, the function $f_{ij}(x_i, x_j)$ is a second-order term describing the interaction between two feature series ($x_i$and x_j) upon the output f(x). where f_0 represents the constant mean effect and $f_i(x_i)$ represents the influence of variable xi (each unique feature coefficient) on the output f(x). In addition, the function $f_{ij}(x_i, x_j)$ is a second- order term that describes the interaction between two feature series ($x_i$and x_j) and the output f(x). Higher-order terms will be zero if no interaction exists between the input feature variables; in this case, the HDMR expansion will only include terms such as f_0 order and$f_i(x_i)$.

White blood cell type	Number of instances
Neutrophils	50
Eosinophils	39
Lymphocytes	52
Monocytes	48
Basophils	53

Figure 6 Instances of white blood cell subtypes

Neural Networks:

Each hidden layer consists of a group of neurons, each wholly linked to all neurons in the layer underneath it. The last fully-connected layer is the output layer, and in classification settings, it provides class scores.

There are certain drawbacks to employing neural networks to analyze massive picture datasets, such as the need for more parameters. Consequently, we introduced Convolutional Neural Networks.

Convolutional Neural Networks:

A Convolutional Neural Network is a Deep Learning method that can take an input picture, assign value (learnable weights and biases) to various aspects/objects in the image, and differentiate between them. CNN networks employ a large number of layers.

$$G[m, n] = (f * h)[m, n] \sum_j \sum_k h[j, k] \cdot f[m - j, n - k]$$

Convolutional Layer:

In the convolution layer, it is aimed to obtain attribute maps. The convolution process takes place on the image matrix by traversing an NxN matrix [12].

Pooling Layer:

This layer is often applied after the convolution stage, with the top pooling layer serving as its primary counterpart. The pooling layer employs an NxN matrix[13].

Fully-Connected Layer:

There is a connected layer after the convolution and centering layers of a convolutional neural network. The data from the preceding layer is transformed into a one-dimensional matrix in this layer.

Rectified Linear Unit (Relu) Layer:

The activation function is a function that creates an output value that corresponds to the input to the artificial neuron cell. It requires less processing power than the sigmoid and hyperbolic tangent functions, which has made it increasingly popular in multilayer networks.

Dropout Layer:

While the network is being trained in the deeper learning layers, memorizing the network called extreme learning can take place. Some nodes that learn the network must be disabled to prevent memorizing the network. In this way, memorization of the network is controlled, which improves the network'sperformance.

Normalization:

In deep artificial neural networks, it is necessary to normalize the network to improve its performance. The normalization of input data improves the network's performance and ensures that the data used is

periodically scaled.

Softmax Layer:

The classification layer receives input data from previous layers and uses it to classify the data.

Deep Learning:

Deep learning is one of the most up-to-date and widely used areas of machine learning that allows computers to process and learn data and understand data in terms of hierarchy. Deep learning is a sub-branch of machine learning in which algorithms are modeled on the structure and the way the brain works, called artificial neural networks [9]. Deep learning is widely used in many fields such as image and video processing, biomedical image and signal processing, robotics, chemistry, advertising, search engines, finance, natural language processing, and classification.

Densely Connected Convolutional Networks

Recent research has shown that convolutional neural networks can be deeper, more accurate, and more successful at training if the synapses between layers proximal to the input and near the output are shorter. Densenets offer various advantages, such as:

- ○ addressing the vanishing gradientproblem
- ○ promoting the spread ofcharacteristics
- ○ promote the re-use oftraits
- ○ significantly reducing the size and the number ofparameters

Densenets require fewer parameters than conventional CNNs due to the elimination of the requirement to learn recurring feature maps (Bengio et al., 2013). Densenet layers, on the other hand, are thin and offer just a few novel feature maps. This has the advantage of lowering the proportion of parameters used by boosting feature reuse, which lowers the number of recurring feature maps learnt (Ciresan et al., 2012). Within a dense block, the sizes of the characteristic maps stay constant, allowing for concatenation even when the volume changes.

Transition blocks, on the other hand, execute downsampling between dense blocks using 1x11 convolution and 2x2 pooling levels. This network model includes a hyperparameter growth rate, which affects the percentage of feature maps created by each layer and the quantity of data provided to the overall state by eachlayer

3. Results and Disscusion

We applied and compared multiple CNN-based architectures like Res2Net, and SeparableConv2D-based architectures. We found that Res2Net had the highest accuracy for test data, 0.946, whereas the other models showed accuracies less than 0.94.

This portion of the thesis presents the details of the data set as well as the experiments that were performed, coupled with the precise settings, outcomes of experiments and the discussions. The training and testing of the model were performed on a workstation using Google Colaboratory.

3.1 Hyper-parameterselection

The model's behavior is controlled through hyper-parameters. To optimize a CNN, a range of hyper-parameters might be used based on the application and data. These hyper-parameters influence several parts of the network, including the design, the optimization process, and data processing. The table below covers some of the key hyper-parameters that are tuned for the model used in the final integrated design.3.6.

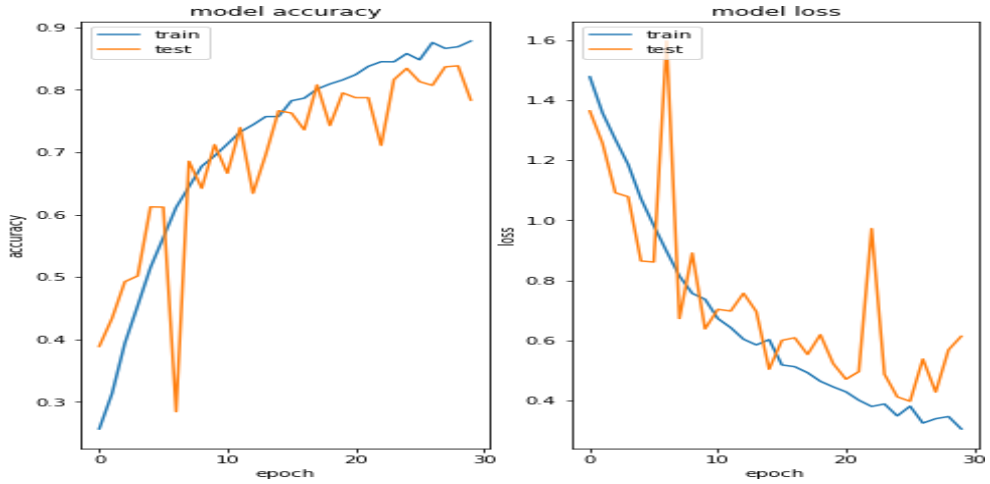
The optimal hyper-parameters are picked manually from the minimum to maximum range shown in table 2 for each of the models discussed in section 3. Each convolutional layer has its own set of hyper-parameters such as kernel size, stride, and pooling. Hyper-parameters are also chosen depending on the results of many training and validation runs. Table 2 shows the combination of hyper-parameters used for each training and validation run. The collection of hyper-parameters that produce the best results is documented and used in the final integrated model.

3.2 Accuracy

Accuracy is one of the criteria used to assess the performance of various techniques. Tables 3 and 4 illustrate the number of parameters as well as training, validation, and test accuracy percentages from various approaches examined in the WBC identification and differential classification sections. Plotting the model loss

and model accuracy of the training and validation sets is one approach of evaluating the models built using the suggested methodology. These learning curves may be used to determine if a model is over-fit, under- fit, or well-fit.

The loss and accuracy curves in figure X show that a decent fit has been achieved. The validation loss curve approaches stability and has a minor gap with the training loss. The validation training curve has similar qualities in that it climbs to the point of stability and has a tiny gap with the training accuracy. The graphs X shown above are for one of the better models used in the final design.



In multi-channel and multi-input techniques, the number of parameters rises. The smaller the gap between training and validation accuracy, the lower the danger of over-fitting. Furthermore, the OVA and 3-class classification strategy is the most accurate, with a test accuracy of 95.45 percent in cell identification and 90.49 percent in cell differentiation.

3.3 Confusion Matrix

As explained in 3.7, the confusion matrix is another way of assessing the performance of our classification methods, as shown in figure 33. This confusion matrix can be evaluated using the ’False Positives’, ’False Negatives’, and ’True Positives’ from each of the confusion matrices.

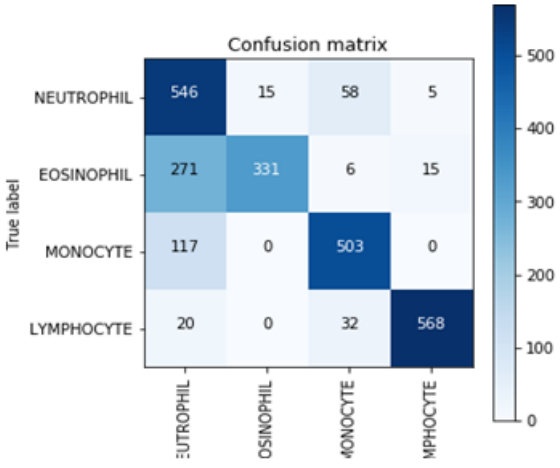


Figure 7 Confusion matrix of the model
Source: P. (2018, April 12). Identify Blood Cell Subtypes From Images. Kaggle.

3.4 Verification plot

The verification plot is used to check the final integrated design's performance. To check the performance of our solution, a graph of error % versus total count generated from the 'Sysmex XN-100' analyzer is shown. By graphing the sample count against the count of the WBC from 'Sysmex XN-100,' figure X depicts the performance of the OVA and 3 Class approaches. The error is determined by subtracting the projected value from the actual value. The company provides a reference count of the WBC for comparison. The plot is split into

two parts with error percentages ranging from 10% to the red dotted line. The blue dotted line is made by finding the error percentage line that covers 95% of the samples. A 9.2 percent mistake is shown by the blue line. This verification plot uses the OVA and 3-class classifier's final combined design.

The algorithm's performance is determined by the error percentage for 95 percent samples. According to industry requirements, at least 95 percent of the samples classified by our neural network must fall under the 10% error range. In the verification plot displayed in Figure 34, 96.7 percent of the points are within the 10% errorrange.

4. Conclusion

However, deep learning is valuable for medical diagnosis, including blood, throughout the years.

Deep learning techniques based on convolutional neural networks are now the method of choice for detecting and classifying white blood cells in medical imaging applications.

Because convolutional neural networks provide excellent results on large datasets, they require a large amount of data and processing resources for training. In some instances, the data collection is limited and may not be large enough to train a convolutional neural network from the start. In such a case, transfer learning may be used to leverage the power of convolutional neural networks while simultaneously lowering the cost of computing. The classification of white blood cells was addressed in this thesis utilizing deep convolutional neural networks. The suggested method addresses the issue of inadequate data for training white blood cell classification algorithms to achieve acceptable performance. Data augmentation techniques such as image processing operations and trained generative adversarial networks (GANs) are used to increase the quantity of training data. We use current models such as VGG, ResNet, and DenseNet for classification networks, which are either built from scratch or previously trained. The acquired findings reveal a significant increase in the performance of the various classifier networks in relation to the additional data gained using the data mentioned earlier in augmentation strategies.

5. Acknowledgements

The scope of using deep learning techniques in hematology has a very wide range. Deep learning techniques can also be used to analyze other components of the blood or analyze all components in the blood to get the metrics of each component in the blood sample.

Narrowing down the future works specifically in this project, I would like to stress the classification of basophils. Though they are less in number, upon getting little data we can employ rigorous data augmentation techniques and can improve the number of basophils data for training and then use it for classification. Also, we have a strong belief that the proposed CNN architectural design methods could give convincing results if it is trained using different images of a WBC blood smear rather than training with grayscale images of WBC varying in focal length.

With the advancement in the field of hematology, it is possible to get access to numerous images related to blood and its components. Convolutional neural networks and different design techniques in CNNs are powerful tools at our disposal. By making good use of these images and we can make wonders in the field of hematology or the healthcare sector on the whole.

6. Future Work

The scope of using deep learning techniques in hematology has an extensive range. Deep learning techniques can also be used to analyze other blood components or analyze all elements in the blood to get the metrics for each element in the blood sample.

Narrowing down the future works specifically in this project, I would like to stress the classification of basophils. Though they are less in number, upon getting little data, we can employ rigorous data augmentation techniques and can improve the number of basophils data for training and then use it for classification. Also, we firmly believe that the proposed CNN architectural design methods could give convincing results if it is trained using different images of a WBC blood smear rather than training with grayscale images of WBC varying in focal length.

With the advancement in hematology, it is possible to get access to numerous images related to blood and its components. Convolutional neural networks and different design techniques in CNNs are potent tools at our disposal. By making good use of these images, we can do wonders in hematology or the healthcare sector.

References

[1] Su, M. C., Cheng, C. Y., & Wang, P. C. (2014). A neural-network-based approach to white blood cell classification. The scientific world journal, 2014.
[2] Cheuque, C., Querales, M., León, R., Salas, R., & Torres, R. (2022). An Efficient Multi-Level Convolutional Neural Network Approach for White Blood Cells Classification. Diagnostics, 12(2), 248.
[3] Habibzadeh, M., Jannesari, M., Rezaei, Z., Baharvand, H., & Totonchi, M. (2018, April). Automatic white blood cell classification using pre-trained deep learning models: ResNet and Inception. In Tenth international conference on machine vision (ICMV 2017) (Vol. 10696, p. 1069612). International Society for Optics and Photonics.
[4] Yildirim, M., & Cinar, A. C. (2019). Classification of White Blood Cells by Deep Learning Methods for Diagnosing Disease. Rev. d' Intelligence Artif., 33(5), 335-340.
[5] Nassar, M., Doan, M., Filby, A., Wolkenhauer, O., Fogg, D. K., Piasecka, J., ... & Hennig, H. (2019). Label-free identification of white blood cells using machine learning. Cytometry Part A, 95(8), 836-842.
[6] H Mohamed, E., H El-Behaidy, W., Khoriba, G., & Li, J. (2020). Improved white blood cells classification based on pre-trained deep learning models. Journal of Communications Software and Systems, 16(1), 37-45.
[7] Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., ... & Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell, 172(5), 1122-1131.
[8] Hegde, R. B., Prasad, K., Hebbar, H., & Singh, B. M. K. (2019). Comparison of traditional image processing and deep learning approaches for classification of white blood cells in peripheral blood smear images. Biocybernetics and Biomedical Engineering, 39(2), 382- 392.
[9] Jiang, M., Cheng, L., Qin, F., Du, L., & Zhang, M. (2018). White blood cells classification with deep convolutional neural networks. International Journal of Pattern Recognition and Artificial Intelligence, 32(09), 1857006.
[10] Patil, A. M., Patil, M. D., & Birajdar, G. K. (2021). White blood cells image classification using deep learning with canonical correlation analysis. IRBM, 42(5), 378-389.
[11] Habibzadeh, M., Jannesari, M., Rezaei, Z., Baharvand, H., & Totonchi, M. (2018, April). Automatic white blood cell classification using pre-trained deep learning models: Resnet and inception.
[12] In Tenth international conference on machine vision (ICMV 2017) (Vol. 10696, pp. 274- 281). SPIE.
[13] P. (2018, April 12). Identify Blood Cell Subtypes From Images. Kaggle.