Performance Analysis of Bone Marrow Microscopic Images using Transformer for Blood Cell Cancer

Pragadeesh T
Department of Electronics and Communication Engineering
K.S.Rangasamy College of Technology, Tiruchengode
ppragapraga@gmail.com

Pragadeeswaran T
Department of Electronics and Communication Engineering
K.S.Rangasamy College of Technology, Tiruchengode
pragadeeswaranthangavel@gmail.com

Vasanth S
Department of Electronics and Communication Engineering
K.S.Rangasamy College of Technology, Tiruchengode
vasanths11403@gmail.com

Pradeep S
Department of Electronics and Communication Engineering
K.S.Rangasamy College of Technology, Tiruchengode
pradeeps@ksrct.ac.in

***Abstract*— Blood cancer develops from the excessive proliferation of white blood cells, constituting approximately one percent of total blood cells. The proposed model is Modified Swin Transformer with strided complication which performs better than ResNet50, Xception, and EfficientNetB3. The model was trained using images of white blood cells, which were initially pre-processed and stripped of best feature elements. Subsequently, the model is trained using the refined and adjusted Swin Transformer frame, leading to the final prediction of the cancer type found in the cells. As a result, the model works well for determining the precise kind of blood cell cancer seen in bone marrow microscopic images.**

Keywords— Leukemia, Blast cell identification, Deep learning, Swin Transformer, Multiple myeloma, Bone marrow images, Blood cell cancer, Microscopic imaging, Transfer learning, Clinical validation.

I. INTRODUCTION

In current times, artificial intelligence(AI) and machine literacy have boosted medical diagnostics to unprecedented heights, offering newfangled results for early spotting and accurate bracket of conditions. This transformative impact is notably visible in hematology, where such technologies exhibit immense pledge, especially in the realm of blood cell cancer opinion. Amongst these cancers, those influencing the bone gist present a unique challenge due to the intricate and complex nature of the bitsy images captured from this crucial towel. This exploration introduces a pioneering Blood Cell Cancer opinion System, harnessing the power of Swin Transformer armature for the analysis of bone gist bitsy images. The Swin Transformer, an expansion of traditional models, has shown outstanding capabilities in handling successional data, making it an ideal seeker for interpreting the intricate patterns found in medical images. The system aims to enhance the delicacy, efficiency, and speed of blood cell cancer opinion, ultimately contributing to better patient issues and more informed treatment planning. The bone gist, as a primary site for blood cell production, contains crucial individual information for various hematological malice. Traditional methods of bitsy image analysis entail manual examination by professional pathologists, a process susceptible to time constraints and human error. By integrating advanced AI technologies, such as the Swin Transformer model, this system endeavors to automate and boost the precision of bone gist cancer opinion, providing swiftly and more reliable results. The Swin Transformer's unique knack to capture long- range dependences in successional data, combined with its scalability, furnishes a sturdy foundation for dealing with the vital details present in bone gist images. The system aims to discern subtle abnormalities in blood cell morphology, pinpoint characteristic patterns linked with various types of blood cell cancers, and deliver rapid-fire and precise judgments.

II. LITERATURE

The examination of infected blood cell images is commonly categorized into three main stages: image preprocessing, feature selection, and classification. Extensive research has been conducted on various cancer types such as leukemia, carcinoma, and myeloma. Nurasyeera Rohaziat, et al put forth a method using YOLOv3 for identifying white blood cells combined with the models of CNN feature extraction. According to their research, the best mean average accuracy was obtained when YOLOv3 was used as the feature extractor in conjunction with Alexnet. The model has been tested with Alexnet, VGG16, Darknet19, and Darknet53 from YOLOv3[1]. Maryam Bukhari et al. have presented a novel framework for the detection of leukemia malignancy in small blood samples using cutting-edge methods that blend squeeze and excitation literacy with cutting-edge solutions for imaging challenges. Learning operations like squeeze and excite together have been used and that betters the performance of the model. ALL_IDB1 andALL_IDB2 are the two intimately available datasets[2]. Farkhondeh Asadi et al has proposed an assaying blood smear images plays a pivotal part in diagnosing colorful blood- related conditions. Early discovery of leukaemia, especially in its original stages, is vital for prompt treatment inauguration. Machine literacy styles in blood smear image analysis offer a rapid-fire and accurate way to diagnose early- onset leukaemia and determine subtypes easing immediate treatment[3]. Shakir Mahmood Abas et al has proposed a discovery model was trained on the" WBCs discovery dataset" over 100 ages with a minimal batch size of 16 and an original literacy rate of0.01. Each time comprised 93 duplications totaling 9300 duplications and took roughly 2.5 twinkles per time. The alternate model passed training for 50 ages with a minimal batch size of 64 and an original literacy rate of0.01. Each time comported of 34 duplications performing in an aggregate of 1700 duplications and a total training time of56.37 twinkles. The discovery model was tested on a separate dataset containing 364 images with 1557 WBCs achieving an optimum Average Precision(AP) of 96 at 9300 duplications during 100 ages[4]. Riaz Ahmad et al has proposed an automated WBC bracket enhances leukaemia opinion by replacing homemade labor. Deep neural networks though accurate face computational hurdles due to expansive features. This work suggests a channel combining transfer literacy and effective point reduction achieving high delicacy. The approach rigidity to classifying colorful blood cell types and conditions is stressed[5]. Zhana Fidakar Mohammed, Alan Anwer Abdulla has proposed an Advancements in medical

image processing, particularly in segmentation enhance disease identification accuracy. This study emphasizes WBC segmentation in microscopic blood images using a threshold-based technique. Experimental findings demonstrate enhanced WBC extraction in terms of both efficiency and segmentation quality as compared to color-k-means clustering[6]. Ms. Minal D. Joshi et al has proposed an Leukaemia disrupts normal blood cell production, causing anemia, bleeding, and infections. Early identification and classification into ALL or AML are vital for effective treatment. Diagnosis includes complete blood count bone marrow biopsy, and morphological analysis by haematologists. Automation through image processing accelerates and improves leukaemia diagnosis[7]. Lorenzo Putzu, Cecilia Di Ruberto has proposed a automating leukocyte identification and classification in microscopic images to aid in diagnosing acute lymphocytic leukaemia. The method demonstrates robust identification and accurate classification of leukocytes, showcasing potential for further development in enhancing the separation step and utilizing a multi-class classification model for different leukocyte types, including lymphoblasts. Future improvements should involve expanding the dataset to enhance training examples and improve overall accuracy[8]. AhmedT. Sahlol et al has proposed a hybrid approach improves accuracy and efficiency by refining the feature vector size. By leveraging SSA algorithm with statistical operations, critical features are selected, optimizing performance. SESSA further streamlines by reducing features while retaining essential information. Choosing appropriate model architecture, such as VGGNet, proves more effective than overly complex models. Utilizing optimization algorithms for feature selection holds promise in efficient classification, resource conservation, and enhanced understanding[9]. Mohamed Esmail Karar et al has proposed an automated acute blood cancer diagnosis achieved using GAN classifier in IoMT framework with microscopic blood smear images. GAN model showed competitive performance generating synthetic data efficiently with advantages in dealing with limited training data. Future includes automating GAN classifier design and enhancing security/privacy in the medical IoT-based system[10]. Anuj Sharma et al has proposed a method using various segmentation scenarios to solve blood cell nucleus segmentation problems, particularly for leukemia cell detection from microscopic images. The model is subsequently compared with cutting-edge methods based on execution time and accuracy, achieving close to 99% classification accuracy on the ALL-IDB dataset. Future plans involve extending the model to handle larger datasets and other types of cancer[11]. Deepika Kumar et al has proposed a deep learning model, specifically utilizing convolutional neural networks, to predict cancer type from images with 97.2% accuracy. The model surpasses Support Vector Machines, Decision Trees, Naive Bayes in performance. However, it acknowledges the need for a broader experimental study considering database size dependency[12].

III. Proposed Methodology

A. Dataset Description

The method proposed involves using a sourced from two distinct subsets found within a collection obtained from Kaggle. Initially, there are 15632 images of patients with Leukemia in one subset. This combined dataset is then used to train the Swin Transformer model, with the aim of distinguishing between Leukemia and Lymphoma cancer cells. An example of the dataset is illustrated in Figure 1.

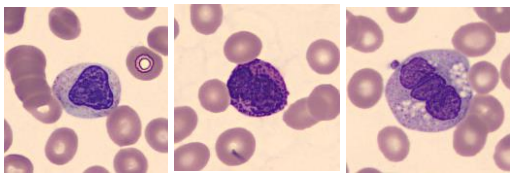

Fig. 1. Image Dataset

B. Data Pre-processing

The large dataset of white blood cell cancer is first pre-processed. The pre-processing process undergoes image resizing and augmentation. All collected images from the dataset are resized to 360 x 360 before being used in training.

C. Architectures

1) Resnet50

It is a 50-layer version of the widely used ResNet architecture. The innovation of ResNet-50 is its ability to train deep neural networks with more than 150 layers without running into the vanishing gradient issue123! By using a bottleneck design for its building blocks, it can reduce parameters and matrix multiplications, resulting in speedier training. ResNet-50 drives computer vision applications such as image identification and search, and is commonly utilized in these tasks.

2) EfficientNetB3

EfficientNetB3 represents a convolutional neural network (CNN) structure crafted to achieve both efficiency and precision in the task of image classification. It belongs to the EfficientNet family, characterized by compound scaling to balance model depth, width, and resolution. B3 offers improved performance with fewer parameters, making it well-suited for resource-constrained applications without compromising on accuracy.

3) Xception

Xception stands out as a sophisticated convolutional neural network (CNN) structure recognized for its outstanding performance in classifying images. Developed by Google, this model adopts depth wise separable convolutions to boost efficiency and decrease the number of parameters required. Xception's architecture enables precise extraction of features and has proven to be highly effective across various computer vision applications.

D. Proposed Model

The modified Swin transformer is the proposed model. The Swin transformer uses a strided attention mechanism. In the Swin Transformer architecture, strided attention is a key component used to efficiently gather information from an input image that is both local and global. The Swin Transformer introduces a hierarchical design that processes the input image at multiple scales, enabling it to handle images of arbitrary sizes effectively. Strided focus is essential in decreasing computational complexity while still enabling the model to grasp long-distance connections effectively. Partitioning the image into distinct, non-overlapping

sections, each viewed as a token, facilitates processing these sections with the transformer layers.

Hierarchical Processing: Instead of directly applying self-attention across all patches, the Swin Transformer employs a hierarchical processing strategy. It divides the patches into stages, with each stage processing patches at a different spatial resolution. The structured hierarchy of this model efficiently captures both local and global contexts.

Strided Attention within Stages: Within each stage, strided attention is applied to reduce computational complexity. Rather than attending to every patch, the model selectively attends to a subset of patches with a fixed stride, effectively skipping over some patches. This allows the model to capture long-range dependencies while reducing the overall computational cost.

Local and Global Context: By incorporating strided attention at different stages of processing, the Swin Transformer is able to capture both local and global context effectively. Local context is captured within each stage, while global context is captured through interactions between different stages!

Efficient Parallelization: The hierarchical processing and strided attention mechanism also enable efficient parallelization of computation, making it easier to scale the model to handle larger input images or batches.

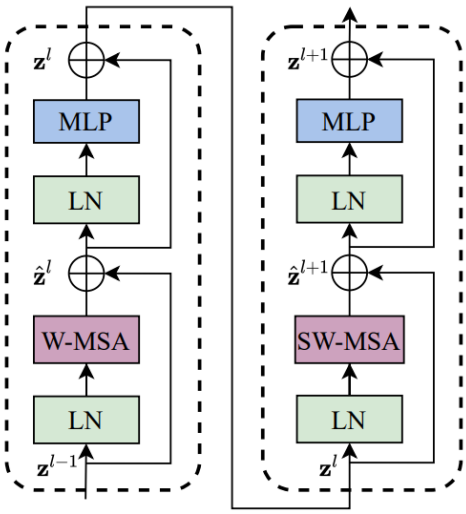

Figure 2: Swin Transformer Block

Steps for Training and Validation:

Step 1: Load the bone marrow microscopic leukemia images dataset.
Step2: Preprocess the images (360x360)
Step 3: Split the dataset into training, validation, and test sets.
Step 4: Implement the Swin Transformer architecture with strided convolution layers.
Step 5: Define the loss function, typically cross-entropy loss for classification tasks.
Step 6: Choose an optimizer (e.g., Adam) and learning rate scheduler.
Step 7: Train the model on the training dataset, monitoring performance on the validation set.
Step 8: Once training is complete, evaluate the trained model on the test set.
Step 9: Compute predictions for each image in the test set.
Step 10: Performance Metrics

Steps for Swin transformer with strided convolution:

Step 1: Replace Standard Convolution with Strided Depthwise Convolution layers.
Step 2: Pointwise convolution is used to aggregate the outputs after depthwise convolution applies a distinct convolutional filter to each input channel.
Step 3: Set the stride parameter of the depthwise convolution layer to a value greater than 1 to achieve strided convolution. This will reduce the spatial resolution of the feature maps, capturing global information while skipping some spatial locations.
Step 4: Ensure that the overall structure of the Swin Transformer block is maintained, including the use of residual connections, layer normalization, and activation functions (ReLU).
Step 5: Adjust the attention mechanism's parameters or introduce additional modifications to accommodate the changes in spatial information.
Step 6: Experiment with different configurations of strided depthwise convolution, such as varying the stride parameter.

E. Loss function

The training loss used in the provided code is specified during the model compilation stage. In the code snippet, the training loss is set to the categorical crossentropy loss. The categorical crossentropy is a commonly used loss function for multi-class classification problems. The equation for categorical crossentropy loss is as follows:

$$L(y,\hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij}.\log(\hat{y}_{ij})$$

The final updated parameter matrix obtained through the optimization process(Adamax) is shown in equation 2,

$$\Theta_{i,new} = \Theta_{i,old} - \alpha\frac{\partial J(\Theta_{old})}{\partial \Theta_i}$$

This process is iteratively performed across all parameters until convergence, where the loss is minimized or reaches a satisfactory level.

Where:

- The numbers N and C stand for the total number of samples and classes,
- y_{ij} is a binary indicator of whether class j is the correct classification for sample i
- $\hat{y}_{ij}$ represents the predicted probability of sample i belonging to class j,
- Θ encompasses the model parameters,
- J(Θ) is the loss function,
- α signifies the learning rate employed.

IV. Result

These plots offer insights into the model's learning trajectory, helping to identify optimal epochs for training and potential signs of overfitting or underfitting. This iterative process of training and evaluation allows for the refinement of

the model architecture and hyperparameters. Once the training is complete, the model's performance is evaluated on multiple fronts. Metrics such as accuracy and loss are computed for the training, validation, and test sets. This comprehensive assessment gives an all-encompassing see of the model's execution over diverse datasets, empowering a careful understanding of its qualities and restrictions. To pick up more profound experiences into the model's classification execution, a perplexity lattice is produced. The perplexity network outwardly speaks to the genuine positive, genuine negative, wrong positive, and wrong negative classifications, advertising a nitty gritty breakdown of the model's forecasts. This visualization aids in identifying any patterns or challenges the model may encounter, providing valuable information for further refinement. The code also includes a comparison between three different models: Resnet50, EfficientnetB3, Xception, Swin Transformer. Comparative analysis of these models can inform decisions on the most suitable architecture for the given task, considering factors such as computational efficiency and classification accuracy. As a final step, the trained model is saved for future use. This is a crucial aspect, particularly if the model is intended for deployment or further analysis. The saved model can be loaded at a later time for making predictions on new data, ensuring the preservation of the trained knowledge for practical applications. In conclusion, the code presented uses the Modified Swin Transformer deep learning architecture to give a thorough and organized method of diagnosing white blood cell malignancy.

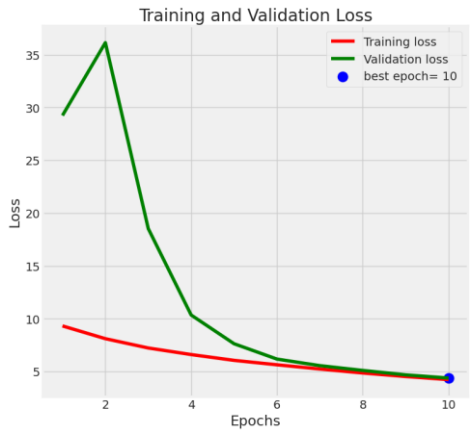

Fig. 3. Training and Validation Loss of ResNet50

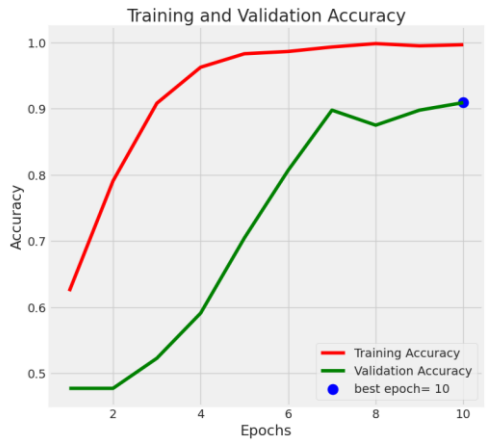

Fig. 4. Training and Validation Accuracy of ResNet50

The provided data showcases the training and validation performance of a ResNet50 model over 10 epochs. The model starts with relatively high loss but rapidly improves in both loss reduction and accuracy. However, there are signs of overfitting, with the training loss significantly lower than the validation loss. Hence, we got a training accuracy of 1.0 and validation accuracy of 0.937 for the ResNet50 model.

Fig. 5. Training and Validation Loss of Xception

The provided data showcases the training and validation performance of a Xception model over 10 epochs. The model starts with relatively high loss but rapidly improves in both loss reduction and accuracy. However, there are signs of overfitting, with the training loss significantly lower than the validation loss.

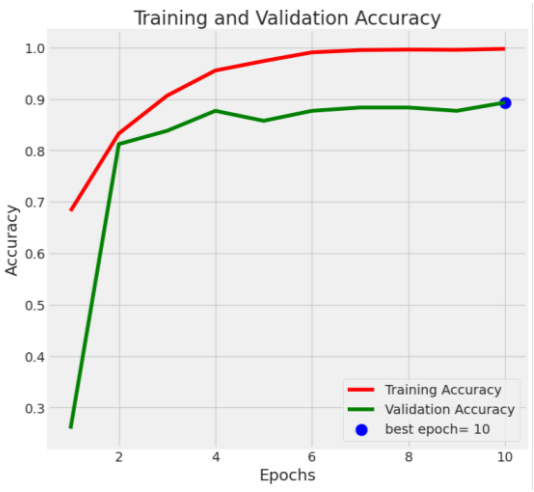

Fig. 6. Training and Validation Accuracy of Xception

The Xception also has a perfect fitting. Hence, we got a training accuracy of 1.0 and validation accuracy of 0.895 for the Xception model.

But, its not enough for the leukemia prediction, so we gone through some other architectures.

Fig. 7. Training and Validation Loss of EfficientNetB3

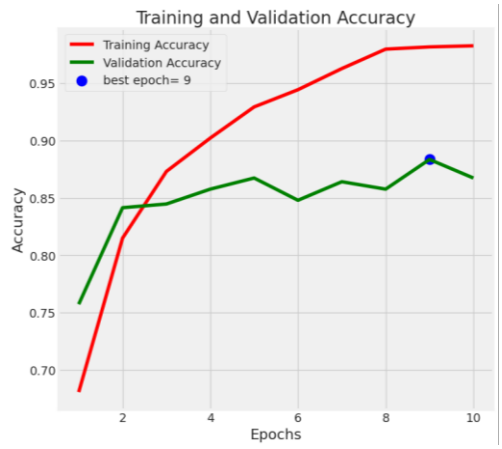

Fig. 8. Training and Validation Accuracy of EfficientNetB3

The provided data showcases the training and validation performance of a EfficientNetB3 model over 10 epochs. The model starts with relatively high loss but rapidly improves in both loss reduction and accuracy. However, there are signs of overfitting, with the training loss significantly lower than the validation loss. Hence, we got a training accuracy of 1.0 and validation accuracy of 0.867 for the EfficientNetB3 model.

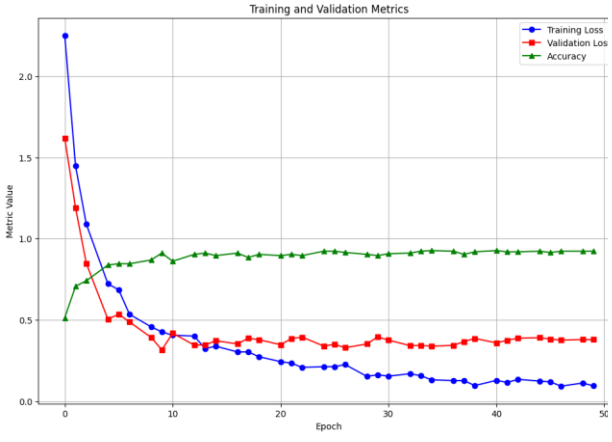

Fig. 9. Training and Validation Metrics of Swin Transformer

The given data presents a Swin Transformer model's training and validation results during a 50-epoch period. The model has a pretty significant loss at first, but it quickly reduces that loss and gains accuracy. Nonetheless, overfitting appears to be present, as the training loss is considerably less than the validation loss. As a result, the Swin Transformer model's training accuracy was 1.0 and its validation accuracy was 0.956.

TABLE I. PREFORMANCE METRICS

PARAMETERS	EPOCHS	VALIDATION ACCURACY	F1 SCORE	PRECISION	RECALL
RESNET50	10	0.937	0.861	0.886	0.866
EFFICIENTNETB3	10	0.867	0.858	0.867	0.858
XCEPTION	10	0.895	0.870	0.869	0.875
PROPOSED - MODIFIED SWIN TRANSFORMER	50	0.956	0.88	0.89	0.881

REFERENCES

[1] Nurasyeera Rohaziat, Mohd Razali Md Tomari, Wan Nurshazwani Wan Zakaria and Nurmiza Othman, “White Blood Cells Detection using YOLOv3 with CNN Feature Extraction Models” International Journal of Advanced Computer Science and Applications(IJACSA), 11(10), 2020. http://dx.doi.org/10.14569/IJACSA.2020.0111058

[2] Maryam Bukhari, Sadaf Yasmin, Saima Sammad, Ahmed A. Abd El-Latif, "A Deep Learning Framework for Leukemia Cancer Detection in Microscopic Blood Samples Using Squeeze and Excitation Learning", Mathematical Problems in Engineering, vol. 2022, Article ID 2801227, 18 pages, 2022. https://doi.org/10.1155/2022/2801227

[3] Mustafa Ghaderzadeh, Farkhondeh Asadi, Azamossadat Hosseini, Davood Bashash, Hassan Abolghasemi, Arash Roshanpour, "Machine Learning in Detection and Classification of Leukemia Using Smear Blood Images: A Systematic Review", Scientific Programming, vol. 2021, Article ID 9933481, 14 pages, 2021. https://doi.org/10.1155/2021/9933481

[4] Abass, Shakir & Mohsin Abdulazeez, Adnan & Zeebaree, Diyar. (2021). A YOLO and convolutional neural network for the detection and classification of leukocytes in leukemia. Indonesian Journal of Electrical Engineering and Computer Science. 25. 10.11591/ijeecs.v25.i1.pp200-213.

[5] Ahmad, R.; Awais, M.; Kausar, N.; Akram, T. White Blood Cells Classification Using Entropy-Controlled Deep Features Optimization. Diagnostics 2023, 13, 352. https://doi.org/10.3390/diagnostics13030352

[6] Mohammed, Zhana & Abdulla, Alan. (2020). Thresholding-based White Blood Cells Segmentation from Microscopic Blood Images. UHD Journal of Science and Technology. 4. 9. 10.21928/uhdjst.v4n1y2020.pp9-17.

[7] Alharbi AH, Aravinda CV, Lin M, Venugopala PS, Reddicherla P, Shah MA. Segmentation and Classification of White Blood Cells Using the UNet. Contrast Media Mol Imaging. 2022 Jul 11; 2022:5913905. doi: 10.1155/2022/5913905. PMID: 35919503; PMCID: PMC9293541.

[8] Di Ruberto, Cecilia & Putzu, Lorenzo. (2013). White Blood Cells Identification and Classification from Leukemic Blood Image. Proceedings of the IWBBIO International Work-Conference on Bioinformatics and Biomedical Engineering.

[9] Sahlol, A.T., Kollmannsberger, P. & Ewees, A.A. Efficient Classification of White Blood Cell Leukemia with Improved Swarm Optimization of Deep Features. Sci Rep 10, 2536 (2020). https://doi.org/10.1038/s41598-020-59215-9
[10] Karar ME, Alotaibi B, Alotaibi M. Intelligent Medical IoT-Enabled Automated Microscopic Image Diagnosis of Acute Blood Cancers. Sensors (Basel). 2022 Mar 18;22(6):2348. doi: 10.3390/s22062348. PMID: 35336523; PMCID: PMC8949784.
[11] A. Sharma, D. Prashar, A. Ahmad Khan, F. Ahmed Khan and S. Poochaya, "Automatic leukaemia segmentation approach for blood cancer classification using microscopic images," Computers, Materials & Continua, vol. 73, no.2, pp. 3629–3648, 2022. https://github.com/sidhomj/DeepAPL/tree/master/Data
[12] Kumar, Deepika & Jain, Nikita & Khurana, Aayush & Mittal, Sweta & Satapathy, Suresh & Senkerik, Roman & D, Jude. (2020). Automatic Detection of White Blood Cancer from Bone Marrow Microscopic Images Using Convolutional Neural Networks. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.3012292.