A Transfer Learning-Based Deep CNN Approach for Classification and Diagnosis of Acute Lymphoblastic Leukemia Cells

Leo Dominick C. Magpantay
Computer Engineering Program
Batangas State University Alangilan Campus
Batangas City, Philippines
leodominick.magpantay@g.batstate-u.edu.ph

***Abstract*—Acute Lymphoblastic Leukemia (ALL) is one of the deadliest types of cancer commonly among pediatric patients. This type of cancer invades the blood and spread through neighboring organs and body systems. To classify cancer cells and non-cancer cells, specialists need to perform manual diagnosis through inspection of cell images under microscope and provide labels through annotation. However, this manual microscopic analysis is a tedious and an error-prone process. Hence, the author proposed a computer-aided diagnosis method using a transfer learning-based deep learning approach. In this study, YOLOv3 model is utilized to train a deep learning model that will classify ALL and normal cell. The model produced promising results as it has a training loss of 2.8% or a 97.2% training accuracy and a validation loss and accuracy of 2.18% and 97.82%, respectively. Based on model evaluation, the model has an mAP value of 99.8% as well. It was concluded that YOLOv3 is a good method for identifying leukemia cells from normal cells.**

Keywords—Leukemia Cell Classification, deep learning, transfer learning, yolov3

I. INTRODUCTION

Cancer is one of the deadliest diseases in the world. It is a generic term for a very large group of illness that may affect every system or part of human body [7]. Cells that are infected are also called malignant tumors or neoplasms. These cells grow beyond their usual boundaries rapidly, invading other neighboring cells and adjoining parts of the body and may spread to other inner organs. This is called metastasis and is the primary cause of death from cancer [7]. In year 2019, WHO estimated that cancer is the first or second leading cause of death before the age of 70 years in 112 of 183 countries and ranked third or fourth in a further 23 countries [11]. There are about 19.3 million new cancer cases (including non-melanoma skin cancer) and almost 10 million cancer deaths occurred in 2020 [11].

In this research paper, it focuses on Acute Lymphoblastic Leukemia (ALL). It is a type of a blood cancer that begins in white blood cells situated in the bone marrow or the soft inner part of the bones and the most common type of cancer among pediatric cancer that causes an approximately 25% of those cases. Children have a higher rate of remission than adults [3]. From an immature lymphocyte, a type of an essential white blood cell inside the immune system, this type of cancer develops. The term “acute” means it gets worse quickly and was considered a rare type of leukemia. It is very common in adults but most commonly in children [6]. The ALL invades the blood and could spread to other neighboring organs in the system such as the liver, lymph nodes, and spleen and it does not create tumors like other types of cancer. With the presence of ALL, it blocks the production of the normal cells, and the number of healthy blood cells is usually lower than normal count. To have a precise and better detection of ALL, it is necessary to inspect the white cells that were present in a blood sample. This method of diagnosis is manually done usually by experienced pathologists who also investigates the abnormalities of white cell formation under microscopic images, to determine the presence of cancer [2]. Lymphoblasts are normally present inside the bone marrow but an increased number of these could be caused by the ALL [2].

Several research proposed different methods in classifying ALL. The authors in [2] argued that the development of Computer Aided Diagnosis (CAD) systems for ALL detection is an efficient way of helping medical experts in performing the diagnosis. The said systems can automate processes using image processing and Machine Learning techniques. Their [2] study focuses on improving the dataset by applying adaptive image processing techniques to improve the sharpness of the images before proceeding to model training. They [2] also implemented parameter tuning and used a convolutional neural network (CNN) based design, VAR-PCANet. Using state-of-the-art deep CNN, they [2] processed the improved images and performed the classification. However, they [2] also argued that their method is independent of which CNN is to be used in the final classification stage. In [4], they proposed an automatic detection of ALL based-on transfer learning. This method uses a pre-trained model and only fine-tuned in the target or specified domain. They employed ShuffleNet to classify malignant and benign cells. Their model illustrated promising results with highest sensitivity, precision, specificity, accuracy, AUC, and F1 score. In [5], the authors implemented different well-known classifiers. Their system comprised of three phases: image segmentation, feature extraction, and classification. They first converted RGB images to CMYK for easier segmentation. They [5] also argued that their proposed diagnosis system detects both the nuclei and the entire membrane. After the segmentation, the features are extracted. Certain features such as the shape, color and texture are widely used as they gave good results. For classification phase, the features extracted were tested using different classifiers: k-Nearest Neighbor (k-NN), support vector machines (SVM), Naïve Bayes and Decision Trees.

The researcher proposed a transfer-learning based deep learning approach for Acute Lymphoblastic Leukemia classification. The YOLOv3 method will be used as it can provide excellent results in terms of accuracy and minimal complexity [1]. Its structure will be discussed as well as the metric values used to evaluate the trained model. In section 2, it will cover the whole methodology, section 3 will cover the results and discussions while conclusion will be covered in section 4.

II. Methodology

In this section, the steps done in the research are discussed thoroughly. Each stage will be well-explained as well as the tools utilized. These steps include data gathering and dataset preparation, image annotation, model training and evaluation, and deployment. This is further demonstrated in in Fig. 1.

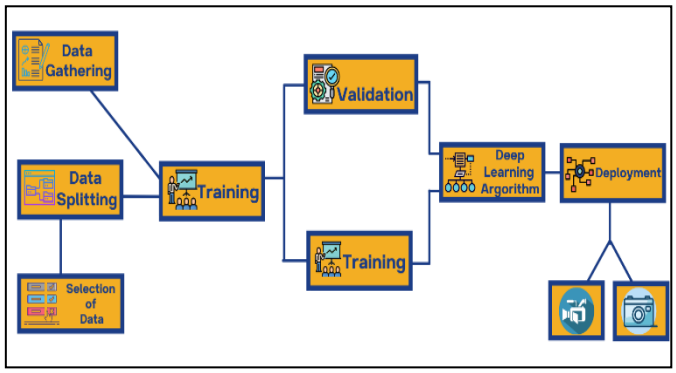

Fig. 1. Conceptual Framework

A. Data Gathering and Dataset Preparation

The dataset used in this study was acquired from a popular data science website *Kaggle.com* [13]. It was comprised of cells segmented from microscopic images and are representation of images in the real-world. It contains some staining noise and illumination errors, but these errors have largely been fixed during the acquisition phase.

In total there are 15,135 images from 118 patients with two labelled classes: Normal Cell and Leukemia Blast. Only 300 from these images are selected to be used for model training and evaluation. Seventy percent (70%) or exactly 240 of these images are used for training and the remaining thirty percent (30%) or 60 images are used for validation. Since there are two classes or labels, the selection of data must be equal. In this case, 150 are from (HEM) Normal Cell and the other half are from Leukemia Blast (ALL). The training images consist of 120 HEM-labelled (Hemocyte) images and another 120 ALL-labelled (Acute Lymphoblastic Leukemia) images. On the other hand, the validation images consist of 30 HEM-labelled images and another 30 ALL-labelled images.

Image samples from the dataset are shown in Fig. 2.

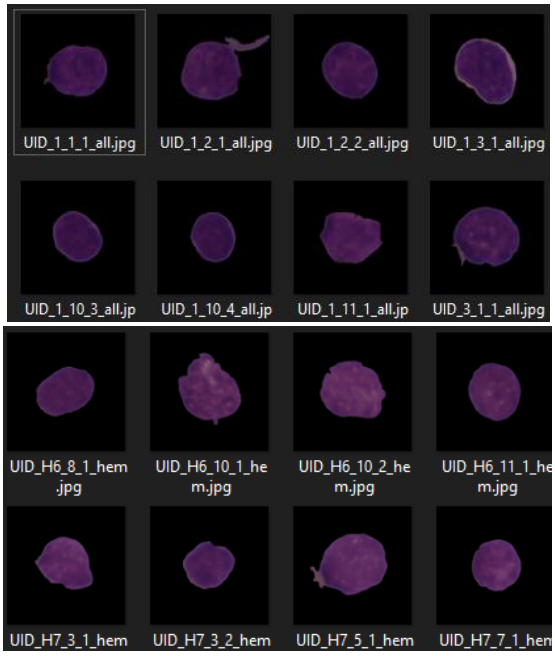

Fig. 2. Image Samples from Dataset

B. Image Annotation

Image annotation is the task of annotating images with labels. Usually, it involves manpower, and, in some cases, computer aided [8]. The labels are predetermined already and are chosen to provide the deep learning model the information it needs. This process also helps the AI model to perform well and yields good results in terms of overall precision and accuracy.

The researcher used Python's LabelImg to annotate and label the dataset. A bounding box was built and applied to the relevant object, which is the body of the cell, refer to Fig. 3. The result of the annotation is then stored inside an XML file with the coordinates of the annotated pictures in PascalVOC format. This is shown in Fig. 4.

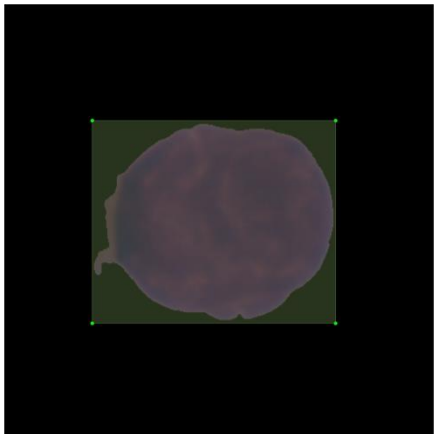

Fig. 3. Annotated Cell Image

```
<annotation>
	<folder>images</folder>
	<filename>UID_1_1_1_all.jpg</filename>
	<path>C:\Users\leodominick\Desktop\Leukemia Classification\training\images\UID_1_1_1_all.jpg</path>
	<source>
		<database>Unknown</database>
	</source>
	<size>
		<width>450</width>
		<height>450</height>
		<depth>3</depth>
	</size>
	<segmented>0</segmented>
	<object>
		<name>ALL</name>
		<pose>Unspecified</pose>
		<truncated>0</truncated>
		<difficult>0</difficult>
		<bndbox>
			<xmin>91</xmin>
			<ymin>122</ymin>
			<xmax>349</xmax>
			<ymax>335</ymax>
		</bndbox>
	</object>
</annotation>
```


Fig. 4. XML File Content

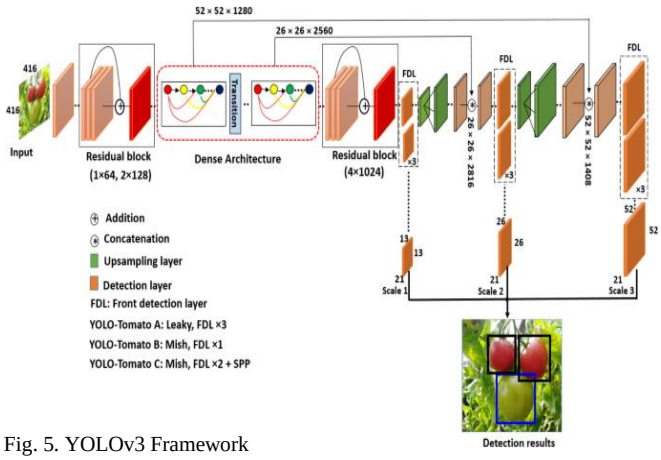

Fig. 5. YOLOv3 Framework

C. *Model Training and Evalution*

For model training, the YOLOv3 deep learning algorithm is used by the researcher. The 'You Only Look Once' v3 or YOLOv3 is among the most extensively used deep learning-based object detection methods [12]. It employs a single convolutional neural network (CNN) to predict bounding boxes in the whole image and then it divides it into multiple parts. The CNN becomes more effective because of its capacity to make several predictions at once [12].

The researcher used YOLOv3, given its framework and architecture in Fig. 3. It is easy to use with its fundamental structure and widely usage. Usually, it is suggested to perform data augmentation procedure to improve the dataset as well as the performance of the model. Different techniques may be applied such as filtering methods, cropping, HSV distortion and flipping. In this study, the author did not undergo data augmentation as the given dataset has enough number of samples and were preprocessed already.

The model training was done in Google Colab. It is a product from Google Research and allows everyone to write and execute python code, hosting Jupyter notebook, through their browser. It is well-suited specially for machine learning or data analysis. It also provides free access to computing resources including graphical processing units [citation].

After the training period, the best-trained model was chosen for deployment. To evaluate the best performing model, mean Average Precision (mAP) (1) is used to compare the trained models. It is the average precision (AP) of all the performed tests [9]. The O is the total number of queries in the set and AP stands for average precision (AP) for a specific query, o [1].

It is a popular evaluation metric used especially for object detection i.e., localization and classification. In classification, it tells the human what is inside the image (e.g., dog or cat). On the other hand, localization determines the location of an instance (e.g., bounding box coordinates) [10]. The accuracy of the models will improve as mAP increases.

On the other hand, to reduce the effect of the curve wiggles, the accuracy must be interpolated at the consecutive recall stages until AP (2) is defined [1]. The area under the interpolated curve is the AP, which can be calculated using the given equations. The interpolated precision P_{interp} (3) determines a certain degree of recall value *r* as the maximum level of accuracy discovered at any stage of recall value $r' \geq r$ [1].

$$(1)\quad mAP = \frac{\sum_{i=1}^{O} AP_i}{O}$$

$$(2)\quad AP = \sum_{i=1}^{n-1}(r_{i+1} - r_i)(P_{interp})(r_{i+1})$$

$$(3)\quad P_{interp}(r_{i+1}) = \max_{r' \geq r} p(r')$$

About the given formula in (1), for a given query *o*, the corresponding AP is calculated and the mean of all those AP scores will be the mAP. It quantifies how good the trained model is.

D. *Deployment*

In this study, the author made a graphical user interface (GUI) using different functional tools like Anaconda IDE, PyQt5, and the ImageAI detection library. It has three main functions which are image detection using image file inputs, video detection using video file inputs, and live feed detection which will use a computer's camera device. It utilizes the chosen deep learning model h5 file, accompanied by the JSON configuration file.

New data which are not included in the training dataset must be used to ensure that there will be no biases that may affect the accuracy (4) of the model. For live feed detection, the cell image was shown using a printed materials as well as for video detection.

$$(4)\quad Accuracy = \frac{No.\ of\ detected\ objects}{Total\ no.of\ objects}\ x\ 100$$

III. Results and Discussions

In this section, the training, validation and testing or deployment results will be discussed.

A. *Training and Validation Findings*

The training and validation results are shown in Fig. 6. Initially, the training level is set at 25 and the required number of epochs to be acquired are 25. However, during training, some model (h5) files are missing, hence the author employed additional training levels to acquire a total of 25 epochs, resulting to a total of 32 training levels.

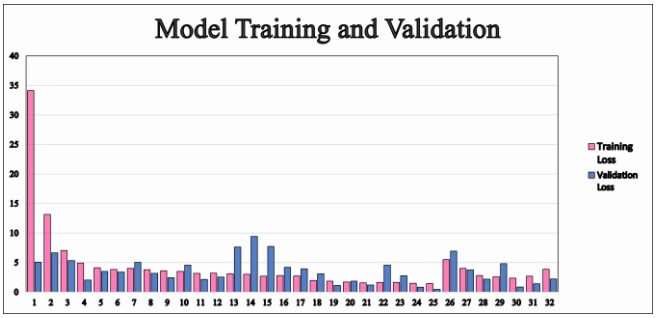

Fig. 6. Training Results

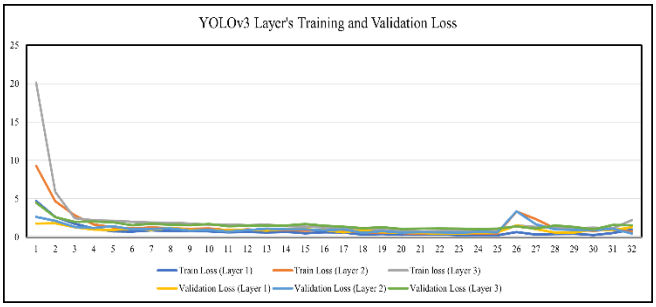

Fig. 7. YOLOv3 layers' training results

During the initial training stage, it yields a train loss (represented by pink bars) of approximately 34.14% and a validation loss (represented by blue bars) of approximately 5.07%. However, at the last training stage, it had a train loss of only 3.87% and a validation loss of 2.22%. As demonstrated in the Fig. 6, at 25th training stage, it yields the lowest train and validation loss values of 1.44% and 0.47% correspondingly. On the other hand, it was the 1st training stage that yields the highest train and validation loss values.

It was also concluded, as shown in Fig. 8 that the average runtime is 607.69 seconds, the average train loss is 4.37%, and the average validation loss is 3.65%.

TABLE I. AVERAGES

Averages	
Runtime	607.69 seconds
Train loss	4.37%
Validation loss	3.65%

B. *Evaluation Findings*

For evaluation, the mAP represents the precision of the dataset. It is expressed as a percentage and the validity value is close to its greatest attainable value since the mAP equates to a value of 1 or 100% [1]. The mAP of each model is shown in Fig. 8.

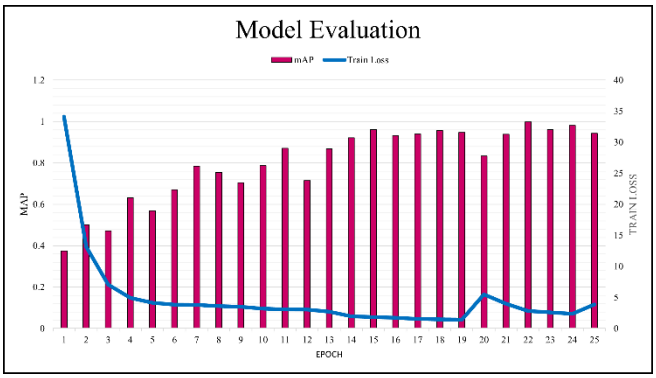

Fig. 8. Evaluation Results

It was concluded that epoch no. 22 obtained an mAP value of 0.998 (99.8%) and was considered the highest performing model. Its training loss value is 2.8% and the validation loss is 2.18% as shown in Fig. 6 (no. 28).

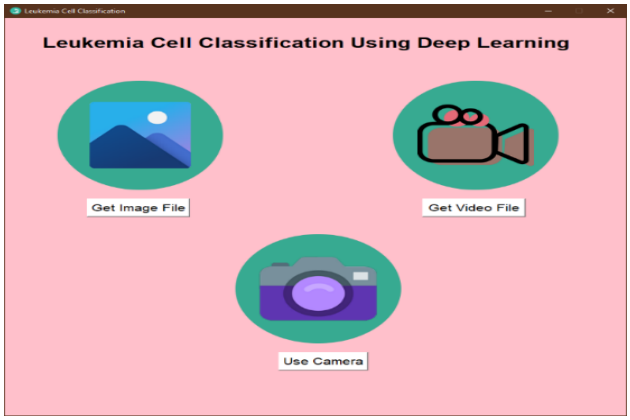

Fig. 9. Graphical User Inter Design

C. *Deployment Findings*

The created simple GUI design is shown in Fig. 9. It has three main functions: To get an image file input, a video file input and to use the camera device. The results are stored in a CSV file.

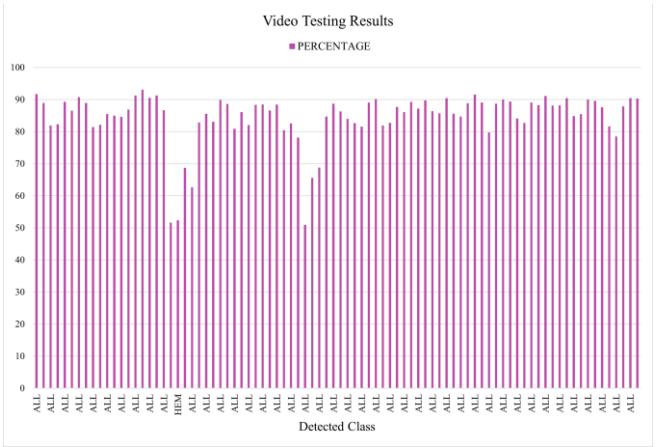

Fig. 10. Video Testing Results

Fig. 11. Video Frames Testing Result

To test the trained deep learning model, the researcher used image and video file inputs and a camera device for live feed detection. The live feed output is shown in Fig. 10, and it is evident that the detection accuracy is changing or varying over time. For cell image detection, the researcher used printed materials that will serve as the test input for the deployment. The images used were fresh and not included in the training dataset. However, its ground truths were pre-determined. In Fig. 11, it shows that the output for an ALL-labelled image was 89.873 and 54.947 for an HEM-labelled image.

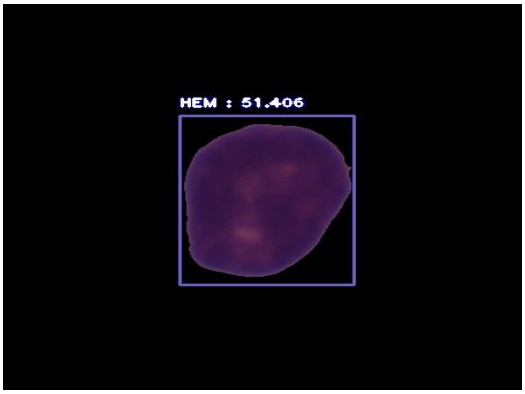

Fig. 11. Image Testing Result

IV. Conclusion

Cancer is one of the most death causing diseases in the world and millions become victim every year. One of its types is the Acute Lymphoblastic Leukemia (ALL) and regarded as the most common type of childhood cancer that also accounts for approximately 25% of the pediatric cancers. Primarily, the task of identifying immature and malignant leukemic cells from normal cells under the microscope is quite challenging as morphological or structural similarities is a great factor that affects it. Henceforth, oncologists and experts provide the ground truth and labels through annotation. To provide a novel solution, the researcher proposed a method of identifying leukemic blasts cells from normal cells using deep learning and convolutional neural network (CNN).

With small number of datasets, the proponent was able to train deep learning models that use the YOLOv3 algorithm and utilize CNN as well. The trained models were assessed using a popular evaluation metric for object detection - mAP. The best model was selected yielding an mAP score of 0.998. During model inference or deployment, the selected model performed and yielded promising accuracy results.

It was then concluded that the YOLOv3 is a good method for identifying leukemia cells from normal cells and might be a big help in oncology medicine.

Acknowledgement

The author would like to express his heartfelt gratitude to the Institution and College he was connected to. Furthermore, he would also like to thank Dr. Alvin S. Alon for his utmost guidance and for sharing his valuable knowledge especially in the field of Machine and Deep Learning and that helped the author finished this academic paper.

References

[1] P. M. B. Melo *et al.*, “Indoor Human Fall Detection Using Data Augmentation-Assisted Transfer Learning in an Aging Population for Smart Homecare: A Deep Convolutional Neural Network Approach,” vol. 2017, pp. 1–6, 2021.

[2] A. Genovese∗, V. P., Mahdi S. Hosseini†, and F. S., Konstantinos N. Plataniotis‡, “Acute Lymphoblastic Leukemia Detection Based on Adaptive Unsharpening and Deep Learning Angelo Genovese ∗, Mahdi S. Hosseini †, Vincenzo Piuri ∗, Konstantinos N. Plataniotis ‡, Fabio Scotti ∗ Department of Computer Science, Universit ` a degli Studi,” pp. 1205–1209, 2021.

[3] W. Daelemans, “Machine Learning Approaches,” pp. 285–304, 1999, doi: 10.1007/978-94-015-9273-4_17.

[4] P. K. Das and S. Meher, “Transfer learning-based automatic detection of acute lymphocytic leukemia,” 2021 Natl. Conf. Commun. NCC 2021, pp. 5–10, 2021, doi: 10.1109/NCC52529.2021.9530010.

[5] A. M. Abdeldaim, A. T. Sahlol, M. Elhoseny, and A. E. Hassanien, “Computer-aided acute lymphoblastic leukemia diagnosis system based on image analysis,” Stud. Comput. Intell., vol. 730, pp. 131–147, 2018, doi: 10.1007/978-3-319-63754-9_7.

[6] A. Stuart, “Acute lymphoblastic leukemia (ALL): Symptoms, diagnosis, treatment, prognosis, and survival rate,” WebMD. [Online]. Available: https://www.webmd.com/cancer/lymphoma/acute-lymphoblastic-leukemia. [Accessed: 15-Dec-2021].

[7] “Cancer,” World Health Organization. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer. [Accessed: 15-Dec-2021].

[8] "Introduction to image annotation for Machine Learning and ai," Introduction to Image Annotation for Machine Learning and AI. [Online]. Available: https://labelbox.com/image-annotation-overview. [Accessed: 15-Dec-2021].

[9] R. J. Tan, "Breaking down mean average precision (MAP)," Medium, 04-Jan-2021. [Online]. Available: https://towardsdatascience.com/breaking-down-mean-average-precision-map-ae462f623a52. [Accessed: 15-Dec-2021].

[10] S. Yohanandan, "Map (mean average precision) might confuse you!" Medium, 09-Jun-2020. [Online]. Available: https://towardsdatascience.com/map-mean-average-precision-might-confuse-you-5956f1bfa9e2. [Accessed: 15-Dec-2021].

[11] H. Sung et al., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," CA. Cancer J. Clin., vol. 71, no. 3, pp. 209–249, 2021, doi: 10.3322/caac.21660.

[12] Q. Zheng, L. Wang, and F. Wang, "Object detection algorithm based on feature enhancement," Meas. Sci. Technol., vol. 32, no. 8, 2021, doi: 10.1088/1361-6501/abe740.

[13] Gupta, A., & Gupta, R. (2019). ALL Challenge dataset of ISBI 2019 [Data set]. The Cancer Imaging Archive. https://doi.org/10.7937/tcia.2019.dc64i46r

[14] Anubha Gupta, Rahul Duggal, Ritu Gupta, Lalit Kumar, Nisarg Thakkar, and Devprakash Satpathy, "GCTI-SN: Geometry-Inspired Chemical and Tissue Invariant Stain Normalization of Microscopic Medical Images,", under review.

[15] Ritu Gupta, Pramit Mallick, Rahul Duggal, Anubha Gupta, and Ojaswa Sharma, "Stain Color Normalization and Segmentation of Plasma Cells in Microscopic Images as a Prelude to Development of Computer Assisted Automated Disease Diagnostic Tool in Multiple Myeloma," 16th International Myeloma Workshop (IMW), India, March 2017.

[16] Rahul Duggal, Anubha Gupta, Ritu Gupta, Manya Wadhwa, and Chirag Ahuja, "Overlapping Cell Nuclei Segmentation in Microscopic Images Using Deep Belief Networks," Indian Conference on Computer Vision, Graphics, and Image Processing (ICVGIP), India, December 2016.

[17] Rahul Duggal, Anubha Gupta, and Ritu Gupta, "Segmentation of overlapping/touching white blood cell nuclei using artificial neural networks," CME Series on Hemato-Oncopathology, All India Institute of Medical Sciences (AIIMS), New Delhi, India, July 2016.

[18] Rahul Duggal, Anubha Gupta, Ritu Gupta, and Pramit Mallick, "SD-Layer: Stain Deconvolutional Layer for CNNs in Medical Microscopic Imaging," In: Descoteaux M., Maier-Hein L., Franz A., Jannin P., Collins D., Duchesne S. (eds) Medical Image Computing and Computer-Assisted Intervention − MICCAI 2017, MICCAI 2017. Lecture Notes in Computer Science, Part III, LNCS 10435, pp. 435–443. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-66179-7_50.