gener8

momentum to market

FROM MICROSCOPES TO MACHINE LEARNING:

Identifying and Counting Blood Cells with Deep-Learning Models

Jeremy Law, PhD
Max Lenk
Lyndie Chiou
Nick Price, PhD

March 23rd, 2023

Table of Contents

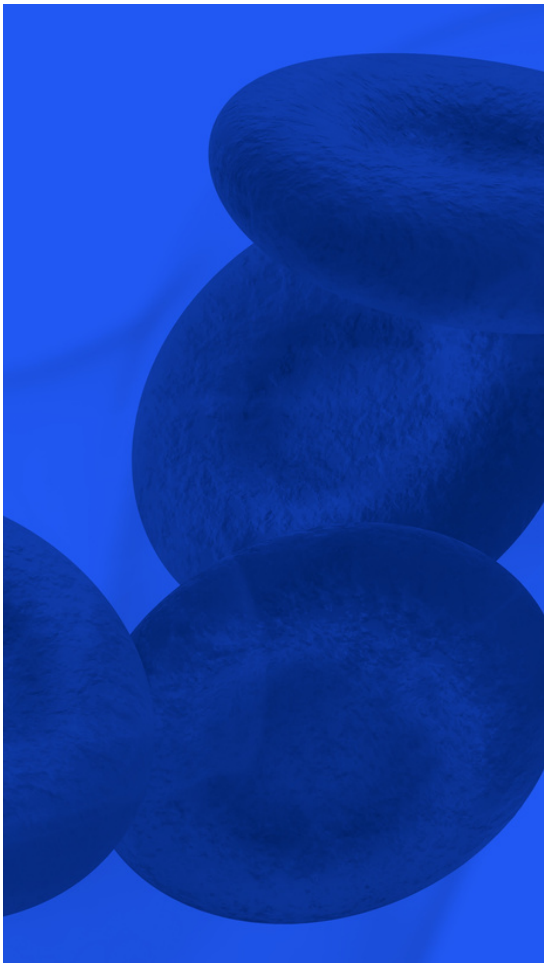
Abstract

Supervised machine learning algorithms were used to differentiate and enumerate label- and stain-free red and white blood cells.

Specifically, transfer learning was used to train three types of deep neural networks (U-net, Segmentation Transformer, and YOLO) to segment and instance the five major types of white blood cells (basophils, eosinophils, lymphocytes, monocytes, and neutrophils), white blood cell nuclei, and red blood cells.

Such artificial intelligence techniques are amenable to various medical imaging and diagnostics applications, such as hematology, pathology, oncology, and other health and life sciences.

Introduction



Our client is a medical diagnostics company developing a point-of-care (POC) blood cell analyzer, which utilizes deep-UV microscopy to create biomolecular maps for a complete blood count (CBC) with a 5-part white blood cell (WBC) differential. [1-5] The device is designed to serve oncology patients undergoing chemotherapy at risk of neutropenia, thrombocytopenia, and anemia.

To ensure timely treatment, these patients require frequent CBC testing to determine whether they are eligible for chemotherapy or need a blood transfusion and, if they are febrile, whether they must be admitted to the hospital.

The client plans to market the analyzer to oncology clinics with later expansion to primary care offices, pharmacies, and homes for in-home nurse care.

The device is a simple, rapid hematology analyzer designed for use at the point of care to improve turnaround time, streamline clinical workflows, and provide advanced testing capabilities to low-resource settings that lack a CLIA-certified laboratory, thus, enabling new care modalities such as in-home chemotherapy and remote patient monitoring.

The device does not require sample processing, reagents, complex systems, or special training to run the test or analyze the results. [4]

Preliminary work has demonstrated the ability to use machine learning algorithms, specifically deep neural networks (DNNs), to perform a 5-part WBC differential. [4] We have expanded that automation capability to include red blood cells (RBCs) and WBCs with 95% or greater accuracy. These additional functions provide the critical outputs for the client's proposed hematology analyzer. They de-risk the technology sufficiently to proceed with product development and testing to support an application for FDA clearance.

The Challenges

The preliminary work to create the 5-part WBC differential showed an accuracy of 94% across all classes. [4] To achieve this accuracy, two segmentation DNNs, two classification DNNs, and two traditional computer vision (CV) techniques were applied to the data. Furthermore, the data were entirely hand-labeled for each of the DNNs, which was time-consuming.

The primary challenges were twofold. Firstly, we had to curate and modify existing datasets to meet our needs. Secondly, we had to train and validate new algorithms to ensure they met the additional requirements of covering the expanded feature space.

The original datasets were:

1. **Segmentation:**
   a. Approximately 60,000 hand-labeled images with white blood cells masked (segmented)
   b. Approximately 60,000 hand-labeled images with white blood cell nuclei masked (segmented)
2. **Classification:**
   a. Approximately 1,300 images divided into two classes: healthy and dead WBCs
   b. Approximately 1,400 images divided into five classes: basophil, eosinophil, lymphocyte, monocyte, and neutrophil

The original networks were:

1. **White blood cell segmentation:**
   a. U-Net with approximately 24 million parameters
2. **White blood cell nuclear segmentation:**
   a. U-Net with approximately 24 million parameters
3. **White blood cell healthy/dead classifier:**
   a. RESNET18: 11 million parameters
4. **White blood cell five-class classifier:**
   a. RESNET18: 11 million parameters

The ideal implementation was an approach that allowed for rapid re-labeling of the data and combined the pixel-by-pixel segmentation and instance counting into one pass with a similar level of computational complexity with better or equivalent accuracy.

Methods

Introduction

To test various algorithms for rapid detection and develop custom models, Gener8 began by hand-labeling the initial subset of the data. We quickly transitioned to a more scalable process, detailed below. With the relabeled data, we used a combination of the gold standard and cutting-edge algorithms to compare their outputs.

Dataset Building

Our strategy for improving the client's workflow began with expanding the scope of the models to include previously unlabeled objects and more comprehensively covering the existing labeled datasets. To add RBCs to the datasets, a new dataset correlating images to labels containing all the desired information about the RBCs was created. We reorganized the respective datasets making them parallel in structure for training models using segmentation and classification information. We used a mixed human and AI bootstrapping method to label RBCs and build a database with a size of a thousand images on the order.

As there were dozens of RBCs in each image and non-expert humans were to be able to segment them, we manually created ground-truth masks for a small number of images, as shown in Figure 1 (a). Using Cellpose's Cyto model, [6,7] which was trained on generic cell databases, we transferred the initial model and iteratively applied unlabeled images.

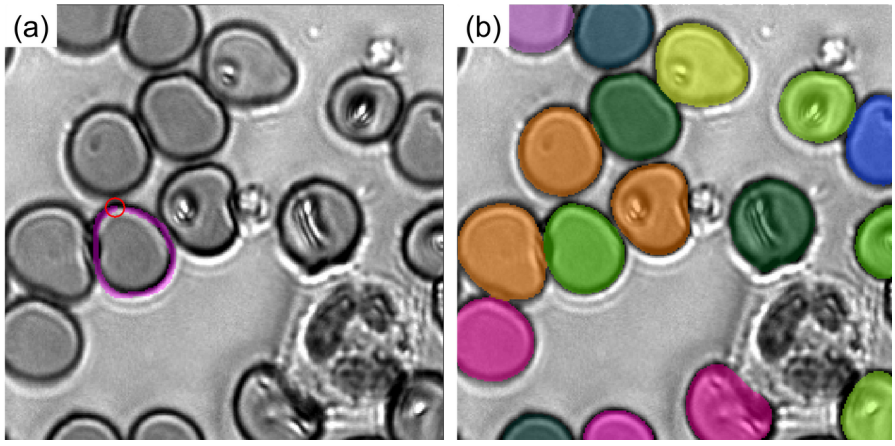


Figure 1 (a) Manually drawing an RBC mask and (b) fully segmented RBC masks as distinct instances.

We then manually curated the resulting masks to add to the dataset and continued the supervised learning process with larger quantities of images and less human intervention until the entire dataset was labeled. With the resulting specialized model, this method was easily scalable to a much larger dataset, as shown in Figure 1 (b).

Dataset Building

We utilized a CV-based matching algorithm to combine and organize the existing segmentation and classification datasets, as there was no link between the WBC images in the classification dataset and those in the segmentation dataset. Furthermore, the images in the classification dataset were slightly modified.

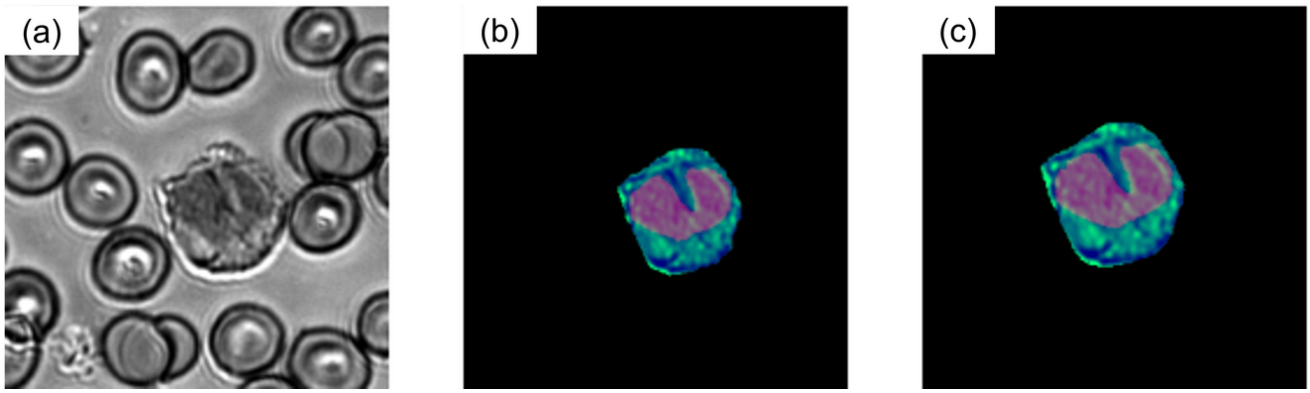


Figure 2 (a) A grayscale image from the segmentation dataset, (b) a colorized WBC generated from the grayscale image, and (c) a colorized WBC from the classification dataset.

We needed to establish a link between the images in the classification and the segmentation datasets. We established a connection between the images in the classification and the segmentation datasets by reversing the transformation process that had been applied and using a pixel-based closeness-of-fit algorithm to generate a CSV file that contained a list of correspondences between images and WBC classification. Figure 2 (a) displays an example grayscale image with a WBC, (b) a colorized WBC from the grayscale image, and (c) a colorized WBC from the classification dataset that matched (b).

Dataset Building

As a result of this groundwork, we combined the datasets in various ways to suit the needs of each model. For example, one approach involved consolidating the RBC masks with the WBC and WBC nuclei masks into a unified ground truth mask without instance information. Another method was to transform each RBC and WBC instance into a bounding box by calculating the range of each mask in the image and labeling the corresponding WBC bounding box. The datasets were reorganized to have the same structure enabling us to experiment with different model architectures and data formats. Figure 3 summarizes the original dataset and the three new datasets created for algorithmic training: the three-class dataset, the 7-class dataset, and the bounding box dataset.

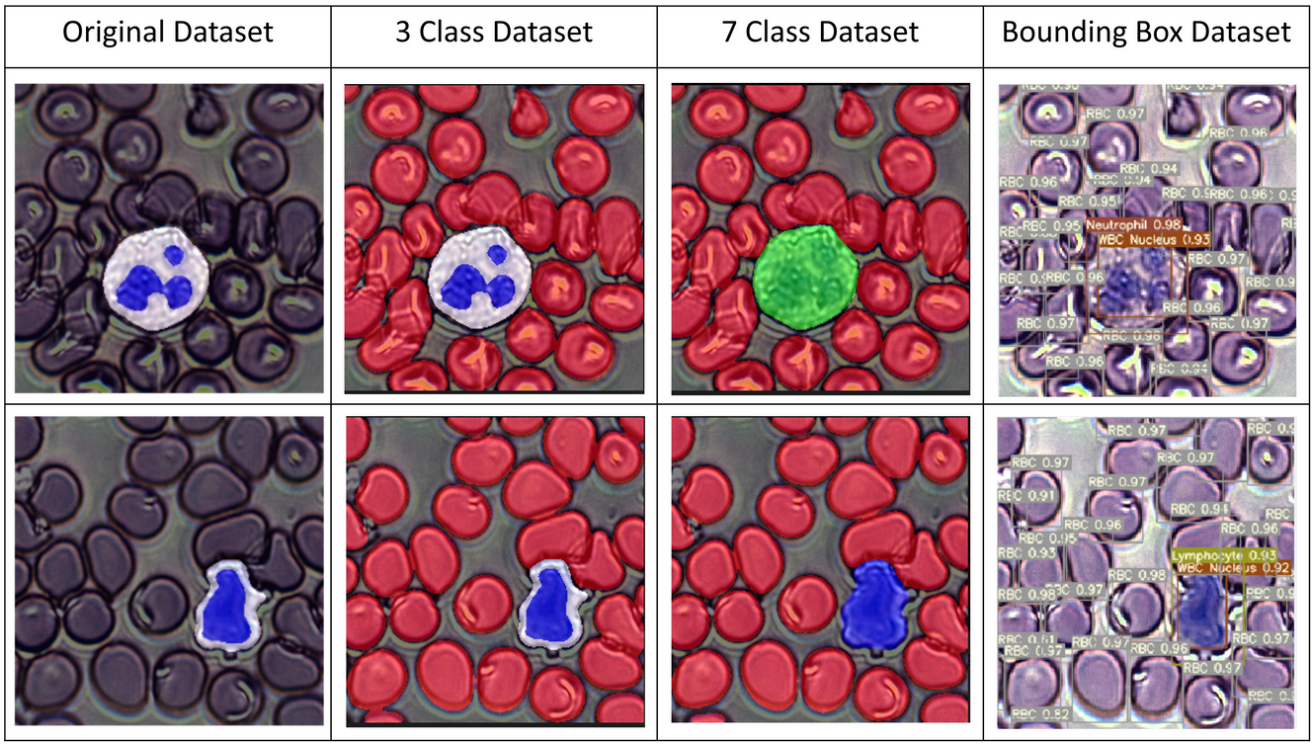


Figure 3: Two example images with overlaid labels from each dataset manifestation. The original has WBC and WBC nuclei. Three-class has RBCs in addition. 7-Class removes WBC nucleus and splits WBCs by type. Bounding Box translates pixel masks to XML bounding boxes.

Gener8 created a generic process to efficiently and semi-automatically re-label datasets. Using this process, we transformed the original four distinct datasets (two segmentation and two classification) that gave one output (WBC count) into three harmonized datasets that can each output count RBCs and WBCs. This process is scalable; it can readily be applied to more of the client's original datasets. It can also be applied to any new dataset.

Semantic Segmentation and CV Instancing

Our initial step was to train different candidate models to segment WBCs, WBC nuclei, and RBCs individually. We explored three candidate methods, whose structures are illustrated in Figure 4. One candidate was based on the U-Net architecture used by the client. To enable this model to segment multiple classes of objects simultaneously, we changed its final layer to produce multiple arrays representing the probabilities of each pixel belonging to each class. The other two candidates were larger models pre-trained on the generic image dataset, ImageNet-1k. [8] Their larger size offered a more extensive feature space that could better capture the details of each object. One of the models was ResNet-34, [9] which was a residual U-Net that included additional skip connections between layers that enabled better convergence. The other model was SegFormer MiT-B4 (SegFormer), [10] a vision transformer that combined local and global contexts more than convolutional networks. Both types of architectures are effective in the medical image processing research. [11, 12]

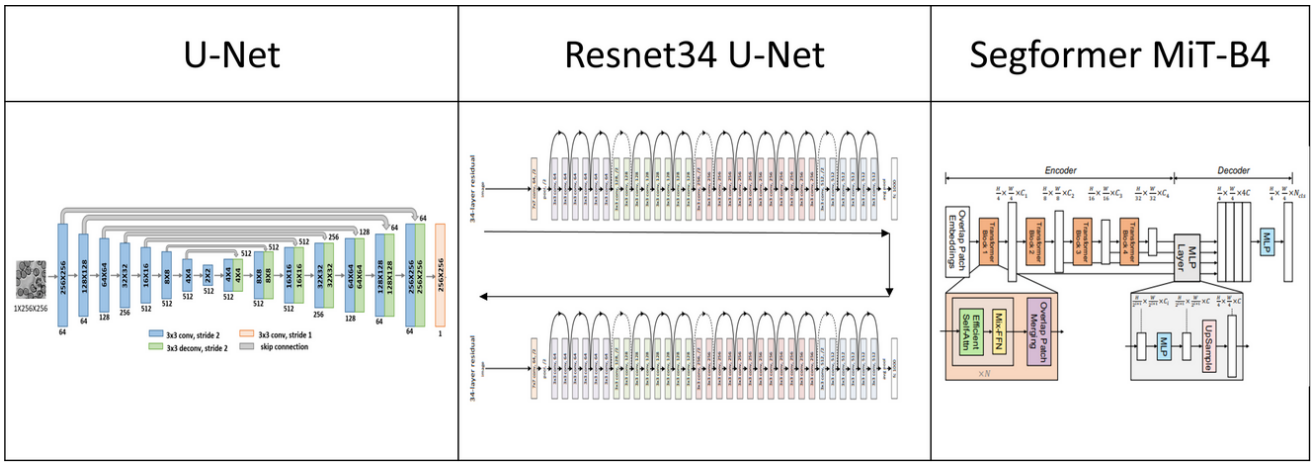


Figure 4: Illustrations of deep-learning model architectures of selected models. [4, 9, 10]

To simplify the dataflow and simultaneously segment and classify WBCs, we trained a model that represented the different types of WBCs as distinct objects in a dataset with seven classes, which included RBCs, dead WBCs, lymphocytes, neutrophils, basophils, eosinophils, and monocytes.

Semantic Segmentation and CV Instancing

We implemented a traditional CV counting algorithm using a watershed method that removed all masked pixels within a certain distance of any background pixel to output the number of RBCs in an image. [13] As adjacent and overlapping RBCs were contiguous pixel-wise, this method consistently undercounted RBCs in denser images. Thus, we applied a linear correction factor to calculate the final count.

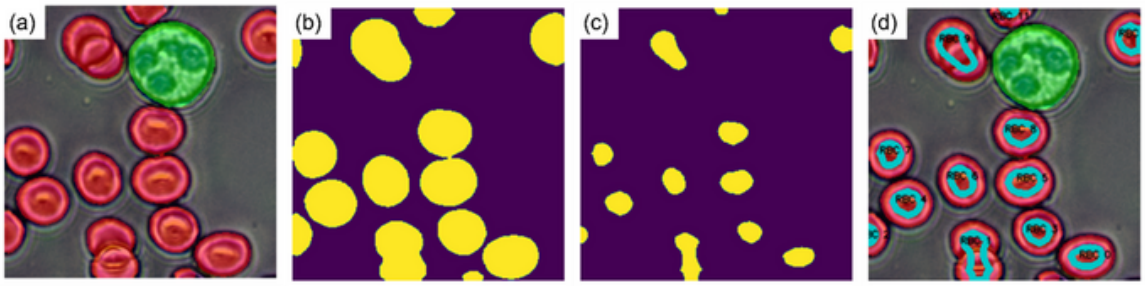


Figure 4: (a) Model output, (b) isolated RBC masks, (c) watershed algorithm mask reduction, and (d) output count of RBCs.

One-Shot Instancing and Box Segmentation

Gener8 initially tested an early version of YOLO ("You Only Look Once" [14]), an open-source, GNU-licensed algorithm that identifies all trained objects within an image in a single pass. YOLO breaks an image up into grids and uses features to mark where predefined bounding boxes should be drawn to identify detections. Its efficiency has made it a gold standard for real-time video applications. However, the early version, YOLOv3, [15] did not meet the required accuracy. Therefore, Gener8 tested YOLOv7, [16] which has a more complicated architecture that allowed better feature detection. Gener8 found that YOLOv7 demonstrated a significant leap in accuracy compared to all the previous versions, making it a viable option.

Gener8 wrote data handling tools to evaluate the algorithm, including conversion scripts that converted segmentation pixel maps into the correct format for running YOLOv3 and YOLOv7 (see Dataset Building Methods section). Once reliable detection of each cell type was achieved, Gener8 developed scripts to subsample the slides along boundaries to avoid misidentifying objects based on partial information and to reconstruct the full slide from the samples while maintaining accurate cell counts.

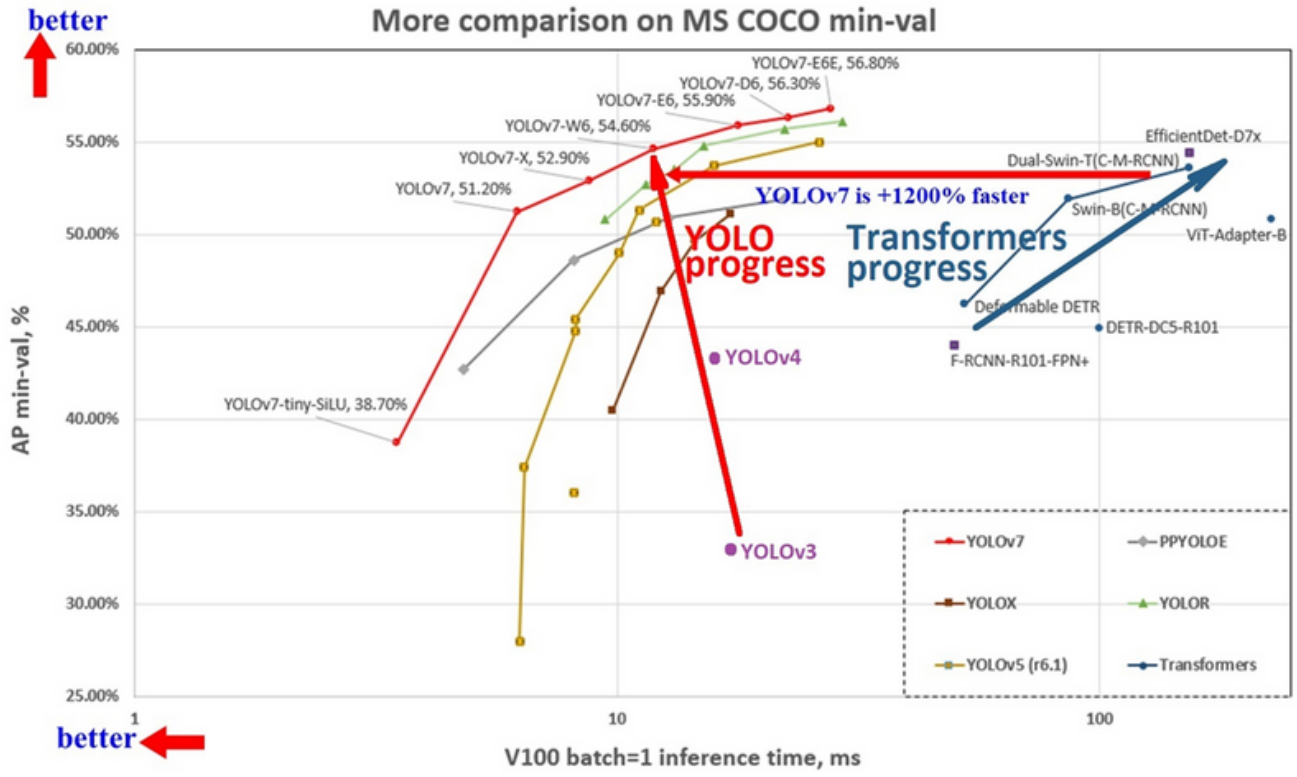


Figure 6: Plot of detection accuracy and speed for YOLOv7 [16] compared to other versions of YOLO using the COCO [17, 18] dataset. [19]

Semantic Segmentation and CV Instancing

Figure 7 plots the mean intersection over union (IOU [20]) for each cell type for the three-class dataset. Mean IOU is a ratio that measures the accuracy of pixel detection and classification by computing the overlap between predicted and ground truth pixels. It is a harsh metric that focuses on pixel-wise correctness and is chosen to compare the relative performance of different models, as it highlights their potential inaccuracies.

The three-class models achieved promising results, with exceptionally high accuracy for RBCs and improved accuracy for WBCs and WBC nuclei compared to the original models. The better performance of WBCs and WBC nuclei may be attributed to the additional context provided by the other object classes. The SegFormer model had the highest accuracy, as judged by mean IOU, possibly due to its more complex architecture. These findings support the feasibility of using pre-training on generic image datasets to enable the use of larger models, even with limited training data availability.

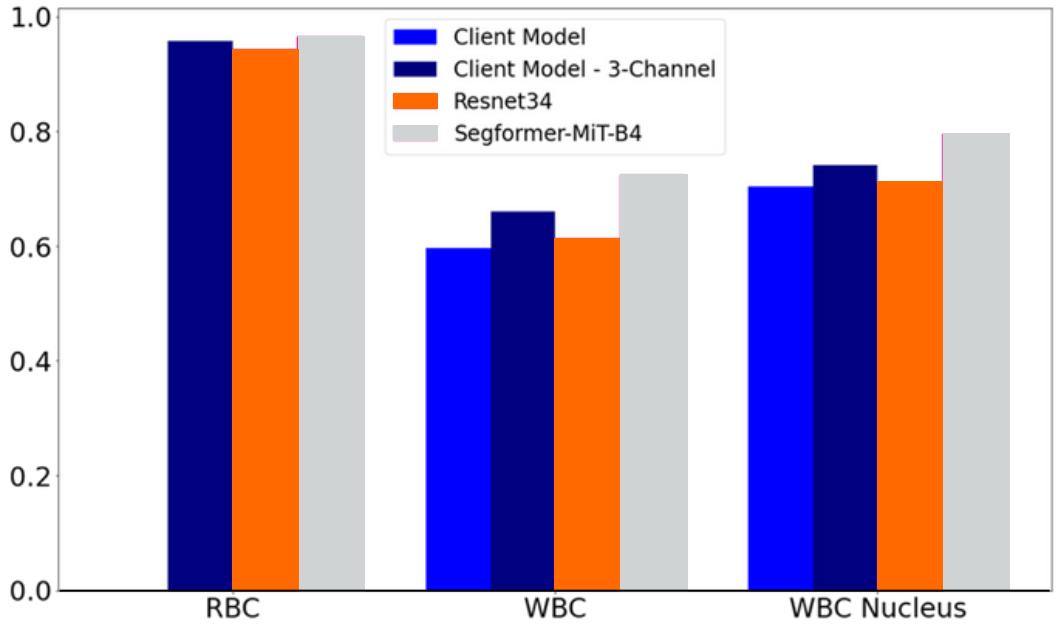


Figure 7: Comparison of model performance on the three-class dataset. Mean IOU [20] (ratio of true positives to false positives and false negatives) over each class displayed. IOU is primarily used as a loss function and is a harsher metric focusing on pixel-wise correctness rather than hematological accuracy.

Semantic Segmentation and CV Instancing

The SegFormer was successful in segmenting and classifying WBCs on the 7-class dataset. Figure 8 illustrates the distribution of IOU per image for each WBC class in the test dataset. The pixel-level accuracy of RBCs and WBCs was comparable to that of the simpler models. Neutrophils and lymphocytes constituted the majority of the WBCs in the dataset, and the SegFormer showed a high level of accuracy in distinguishing between them. The other classes were infrequently represented in the dataset, and the effectiveness of the SegFormer for these classes can only be determined with additional data. In summary, using a larger pre-trained model provided encouraging results for integrating segmentation and classification of WBCs and RBCs.

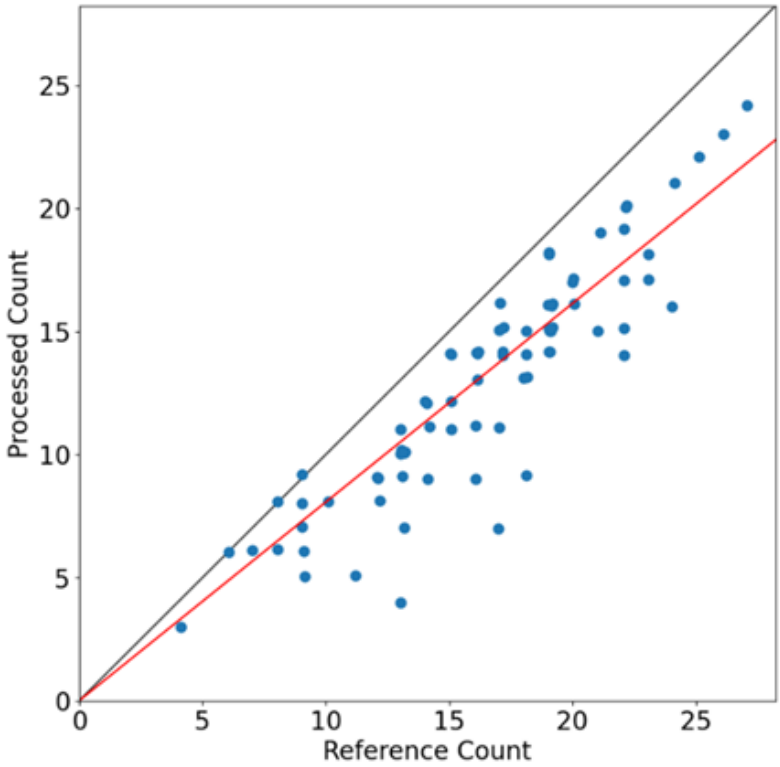


Figure 8: Distribution of IOU [20] per image for each WBC class present in the test dataset.

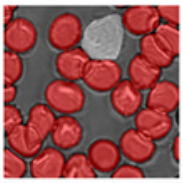
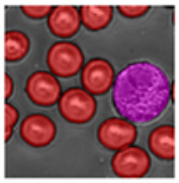
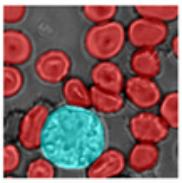
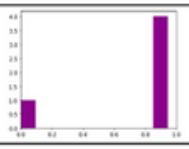
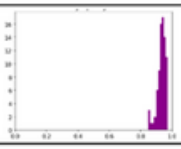
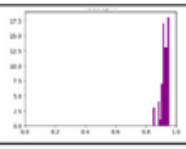
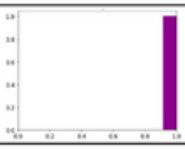
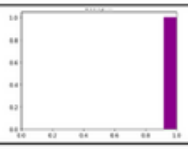
Class	Dead	Lymphocyte	Neutrophil	Eosinophil	Basophil
Example					
IOU Distribution					
# Cells in Test Set	4	80	76	1	1

Figure 9: Plot of predicted RBC count against ground truth RBC count. Target line (black) before correction and (red) after correction

The RBC counting algorithm achieved an accuracy of 80% without correction and 95% accuracy with correction when tested on an independent dataset. Figure 9 illustrates the predicted RBC count against the ground truth RBC count, with the target line shown in black before and red after correction.

Semantic Segmentation and CV Instancing

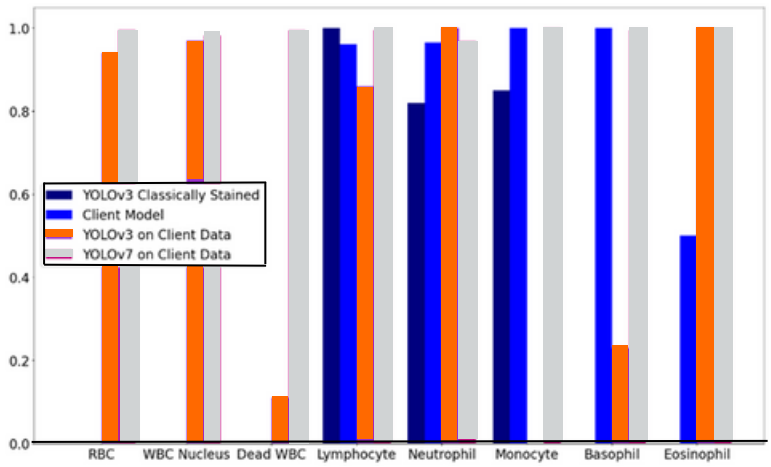
WBC counting can be achieved from these segmentation networks in one of two ways. The first way would be to apply the client's original CV and DNN classifiers or slightly modified versions of those classifiers. We expect that similar accuracy can be achieved with such results. The second way would be to apply the same watershed technique. We did not use this technique to our output as the datasets have one WBC per image. Thus the output of any CV technique would be 100% accuracy and not an accurate comparison to a real-world scenario in which we may have zero, one, or more than one WBC per image.

In summary, these pixel-by-pixel segmentation and CV techniques added the ability to count RBCs with 95% accuracy and maintained the ability to measure WBCs with 94% accuracy using the client's existing processes. Furthermore, the new network topologies improved the segmentation task by approximately 10-25% and scale to more features for future additional classes. We also showed that networks pre-trained on ImageNet data could be efficiently retrained on biological data.

One-Shot Instancing and Box Segmentation

Figure 10 (a) displays a bar chart comparing the results of YOLOv3 (in orange) and YOLOv7 (in gray), with the traditionally stained sample results presented in dark blue [21] for reference. Table 1 shows the accuracies for each class and network type displayed in Figure 10 (a). In Figure 10 (b), an example image with bounding boxes overlayed shows the class and predicted probability of class using YOLOv7. Compared to YOLOv3, YOLOv7 showed improved detection accuracy, with over 99% accuracy achieved in most categories. YOLOv7 also outperformed stained-cell images, which have greater resolution and fewer imaging artifacts, in detecting more categories of WBCs with higher precision. Ultimately, YOLOv7 achieved 97% or greater accuracy in detecting eight classes of objects, with an average accuracy of 99%.

(a).



(b).

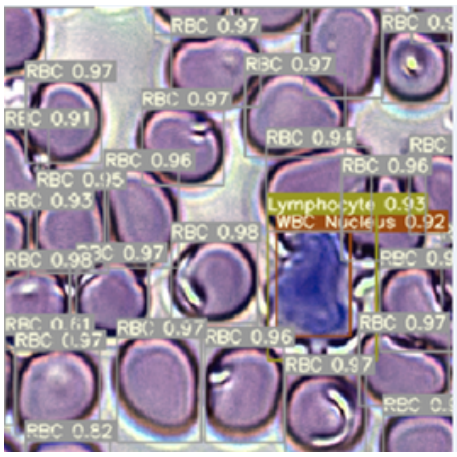


Figure 10: (a) A bar chart comparing the results of YOLOv3 [15] (orange) and YOLOv7 [16] (gray). For comparison, traditionally stained sample results are shown in dark blue and the client's model is shown in light blue. [21] (b) An example image with bounding boxes overlayed showing class and predicted probability of class using YOLOv7.

One-Shot Instancing and Box Segmentation

The subsampling process yielded identification and results comparable to or better than human detection. YOLOv7 was even more accurate than humans in detecting objects that are challenging to identify, such as WBCs missing their nuclei. Figure 11 displays a slide with algorithmically stained cells, RBCs, and several neutrophils and their nuclei identified by bounding boxes utilizing the YOLOv7 algorithm.

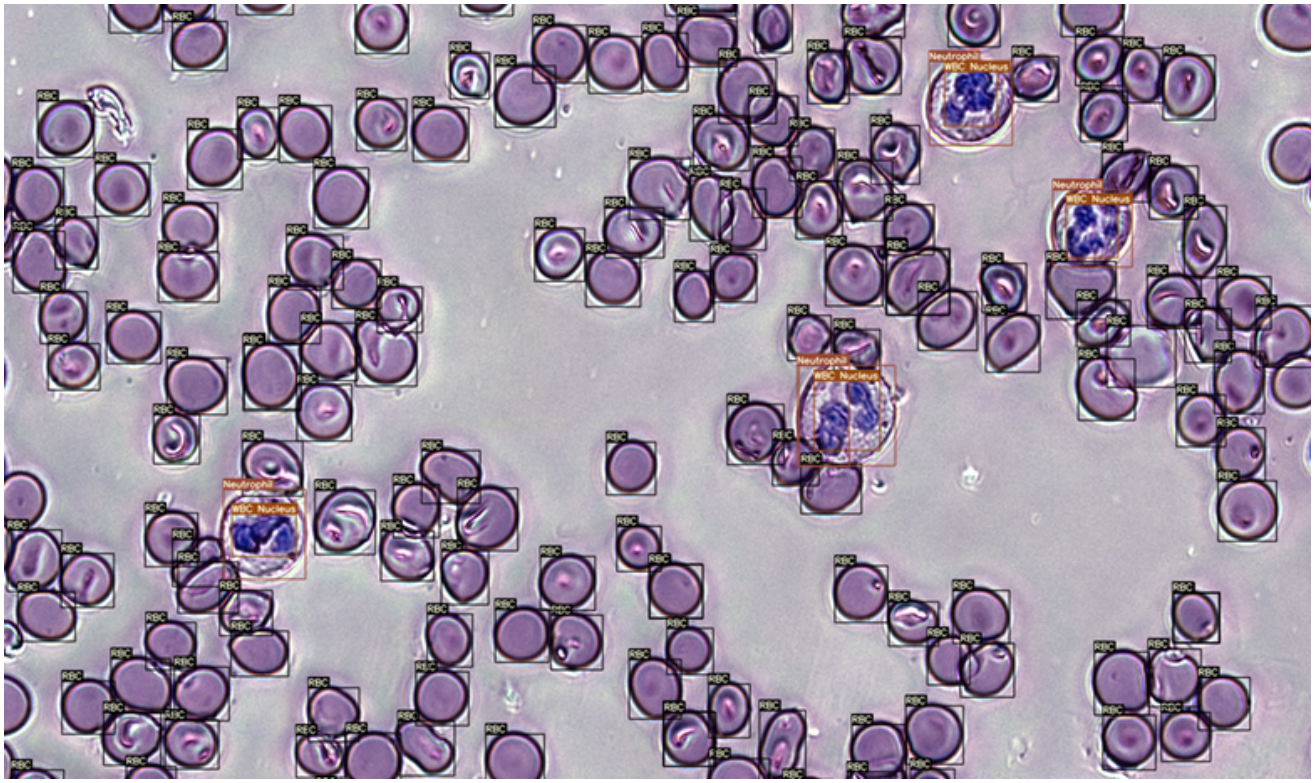


Figure 11: A slide showing algorithmically stained cells. The YOLOv7 [16] algorithm created the bounding boxes to identify RBCs, neutrophils, and their nuclei. The slide has been reconstructed from smaller samples.

One-Shot Instancing and Box Segmentation

	YOLOv3 Classically Stained	Client Results	YOLOv3 Grayscale	YOLOv3 Colorized	YOLOv7 Colorized
Basophil	N/A	1	0.75	0.23	0.995
Eosinophil	N/A	0.5	0.40	1	0.995
Lymphocyte	1	0.962	0.81	0.86	0.996
Monocyte	0.85	1	N/A	N/A	1
Neutrophil	0.82	0.965	0.93	1	0.969
RBC	N/A	N/A	0.94	0.94	0.996
WBC Nucleus	N/A	N/A	0.87	0.97	0.982
Dead WBC	N/A	N/A	0.97	0.11	0.995

Table 1: Table comparing the accuracy of five algorithms (YOLOv3 [15] on classically stained, grayscale UV, and colorized UV data; YOLOv7 [16] on colorized UV data; and the original compound network from the client on grayscale UV) for eight classes (basophil, eosinophil, lymphocyte, monocyte, neutrophil, RBC, WBC nucleus, and dead WBC).

Conclusions

Gener8 met the challenges laid out by the client. We created an extensible and efficient process to label datasets that used humans supervising a transfer learning process. Using this process, Gener8 added three additional datasets that were previously inaccessible. These datasets facilitated the investigation of other deep-learning algorithms. Gener8 showed that once trained on generic datasets, large neural networks can be re-trained on domain-specific hematological data; similar transfer learning approaches will work for various datasets and applications. The new algorithms we applied increased the accuracy and added entirely new capabilities when compared to the client's existing architecture. In the case of panoptic segmentation, we were able to add the ability to segment RBCs and improve the segmentation of WBCs by 10-25%. In the case of instance classification and box segmentation, we achieved an accuracy of more than 96%, which met the client's product requirements. Such data handling processes and machine learning techniques offer the potential to transform and innovate various medical imaging and diagnostics applications.

Expertise Employed

Listed below are the different areas of expertise utilized in this project:

- PyTorch
- TensorFlow
- OpenCV
- YOLO
- Segmentation transformer (SegFormer)
- RESNET
- Deep learning
- Artificial Intelligence (AI)
- Deep neural network
- Data bootstrapping a large, labeled dataset from an initial set with limited labeling
- Transfer learning
- Hematology
- Medical image processing

References

[1] A. Ojaghi, G. Carrazana, C. Caruso, A. Abbas, D.R. Myers, W.A. Lam, F.E. Robles, “Label-free Hematology Analysis Using Deep-Ultraviolet Microscopy,” PNAS 117 (26), 2020, https://doi.org/10.1073/pnas.2001404117.

[2] N. Kaza, A. Ojaghi, F.E. Robles, “Hemoglobin quantification in red blood cells via dry mass mapping based on UV absorption,” J. of Biomedical Optics 26 (8), 086501, 2021, https://doi.org/10.1117/1.JBO.26.8.086501.

[3] A. Ojaghi, P. Casteleiro Costa†, C. Caruso, W.A. Lam, F.E. Robles, “Label-free automated neutropenia detection and grading using deep-ultraviolet microscopy,” Biomedical Optics Express 12 (10), 6115-6128, 2021, https://doi.org/10.1364/BOE.434465.

[4] N. Kaza, A. Ojaghi, F.E. Robles, “Virtual Staining, Segmentation and Classification of Blood Smears for Label-free Hematology Analysis,” BME Frontiers, 9853606, 2022, https://doi.org/10.34133/2022/9853606.

[5] A. Ojaghi, E.K. Williams, N. Kaza, V. Gorti, H. Choi, J. Torey, T. Wiley, B. Turner, S. Jackson, S. Park, W. Lam, F.E. Robles, “Label-free deep-UV microscopy detection and grading of neutropenia using a passive microfluidic device,” Optics Letters 47 (22), 2022, https://doi.org/10.1364/OL.472691.

[6] C. Stringer, T. Wang, M. Michaelos, M. Pachitariu, “Cellpose: a generalist algorithm for cellular segmentation,” Nat. Methods, 18, 2021, https://doi.org/10.1038/s41592-020-01018-x.

[7] M. Pachitariu, C. Stringer, “Cellpose 2.0: how to train your own model,” Nat. Methods, 19, 1634—1641, 2022, https://doi.org/10.1038/s41592-022-01663-4.

[8] J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 248-255, https://doi.org/10.1109/CVPR.2009.5206848.

[9] H. Kaiming, X. Zhang, S., Ren, J. Sun, “Deep Residual Learning for Image Recognition,” arXiv:1512.03385 [cs.CV], Dec 2015, https://doi.org/10.48550/arXiv.1512.03385.

[10] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, P. Luo, “SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers,” arXiv:2105.15203 [cs.CV], Oct. 2021, https://doi.org/10.48550/arXiv.2105.15203.

[11] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, V. K. Asari, “Recurrent residual U-Net for medical image segmentation,” J. Med. Imaging 6 (1), 2019, https://doi.org/10.1117/1.jmi.6.1.014006.

[12] E. U. Henry, O. Emebob, C. A. Omonhinmin, “Vision Transformers in Medical Imaging: A Review,” arXiv:2211.10043 [cs.CV], Nov. 2022, https://doi.org/10.48550/arXiv.2211.10043.

[13] Bradski, G. (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools, https://docs.opencv.org/4.x/d3/d47/group__imgproc__segmentation.html#ga3267243e4d3f95165d55a618c65ac6e1.

[14] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, “You only look once: Unified, real-time object detection,” in CVPR, 2016, pp. 779-788.
[15] J. Redmon, A. Farhadi, “YOLOv3: An Incremental Improvement,” arXiv: 1804.02767[cs.CV], Apr. 2018, https://doi.org/10.48550/arXiv.1804.02767.
[16] C.-Y. Wang, A. Bochkovskiy, H.-Y. M. Liao, “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” arXiv: 2207.02696 [cs.CV], Jul. 2022, https://doi.org/10.48550/arXiv.2207.02696.
[17] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramana, P. Dollár, C. L. Zitnick, “Microsoft COCO: Common Objects in Context,” Computer Vision — ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8693. Springer, Cham. https://doi.org/10.1007/978-3-319-10602-1_48.
[18] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramana, P. Dollár, C. L. Zitnick, “Microsoft COCO: Common Objects in Context,” arXiv: 1405.0312 [cs.CV], Feb. 2015, https://doi.org/10.48550/arXiv.1405.0312.
[19] A. Bochkovskiy [@alexeyab84], (2022, Sept. 9). Find "Novelty" and "Incremental improvement" in this chart YOLO: 8 Apr 2018 - YOLOv3 6 Jul 2022 - YOLOv7 https://arxiv.org/abs/2207.02696 Transformers:26 May 2020 - Detr8 Oct 2020 - Deformable Detr 25 Mar 2021 — Swin 29 Jul 2021 - Dual Swin 17 May 2022 - ViT-Adapter [Twitter] https://twitter.com/alexeyab84/status/1568334188113440768
[20] D. Müller, I. Soto-Rey, F. Kramer, “Towards a guideline for evaluation metrics in medical image segmentation,” BMC Res Notes 15 pp. 210, 2022, https://doi.org/10.1186/s13104-022-06096-y.
[21] N. Praveen, N. S. Punn, S. K. Sonbhadra, S. Agarwal, “White blood cell subtype detection and classification,” 2021 8th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), 2021, pp. 203-207, https://doi.org/10.23919/EECSI53397.2021.9624268.

gener8

momentum to market