

Automated Hematological Analysis: A Machine Learning Approach for Blood Cell Classification

Adam Finch, Pranay Reddy Gundala, and Tyler Rowe

Area of Focus

- Blood-based diseases impact hundreds of thousands of lives each year^[1].
- A complete blood count (CBC) is one common method used to evaluate general health and diagnose various hematological conditions.
 - Key components of a CBC test include counting White Blood Cells (WBC), Red Blood Cells (RBC), and Platelets
- A blood differential test measures the amount of each subtype of WBC (Monocytes, Lymphocytes, Neutrophils, and Eosinophils).



Traditional Methods^[2]

Manual CBC Test

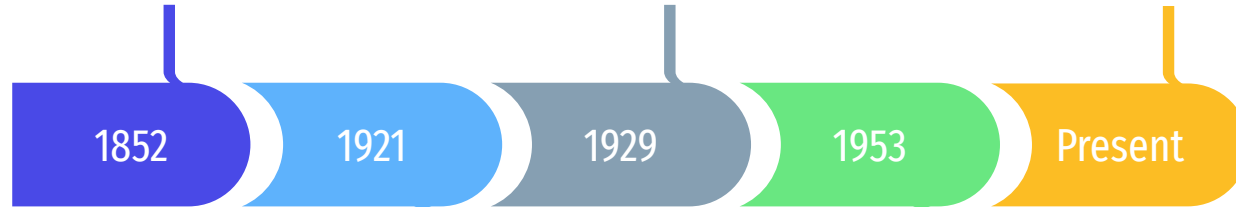
The first procedure for performing a blood count is published by Karl Vierordt.

Wintrobe-Hematocrit Method

Machinery physically separates RBC from WBC to measure volume proportions

Hematology Analyzers

Today the most common method for CBC tests is Hematology Analyzer Machines



Staining Techniques

Techniques for staining WBC and RBC were developed to make manual counting easier

Coulter Principle

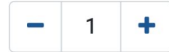
Uses electrical impedance measurements to count blood cells and determine their sizes.

Traditional Methods



**Beckman Coulter, Inc. DxH 900
Hematology Analyzer Continental US Only)**

\$ 116,548.20 ~~\$ 129,498.00~~



 **ADD TO CART**

Ask Question

 [Add to wishlist](#)

Product Information:

- **Products Model:** C23645
- **Product Brand:** [Beckman Coulter, Inc.](#)
- **Condition:** New

Problem:

- A medical grade hematology analyzer is expensive
- Time consuming process if blood samples need to be sent into a lab for counting



Proposal

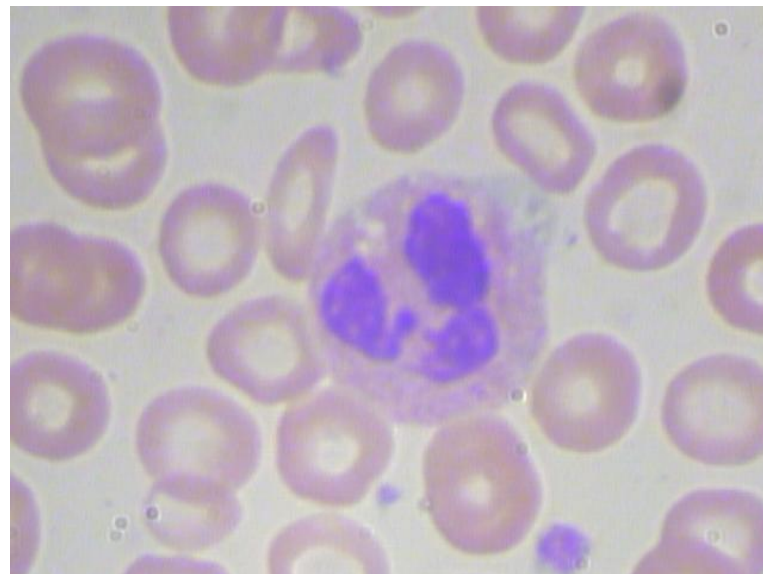
- Machine Learning Approach to CBC testing
 - **Objective 1:** Train a Convolutional Neural Network (CNN) to identify and classify between White Blood Cells, Red Blood Cells, and Platelets.
 - **Objective 2:** Classify between subtypes of White Blood Cells
 - 4 main subtypes: Monocytes, Lymphocytes, Neutrophils, and Eosinophils



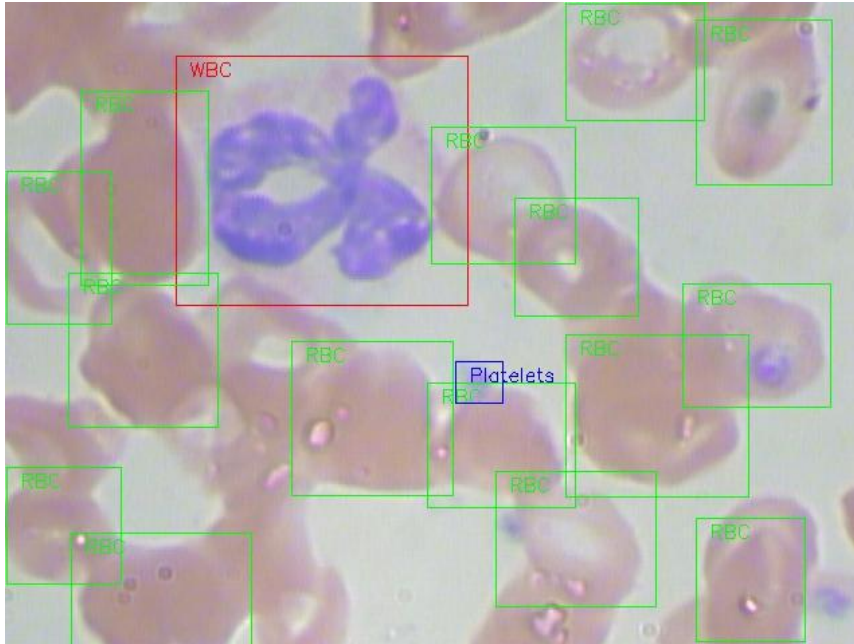


Objective 1: Data

- We utilized the publicly available dataset “Blood Cell Images”, licensed by MIT, to train our model^[3]
- The first dataset contains 410 jpg images of blood samples



Objective 1: Data

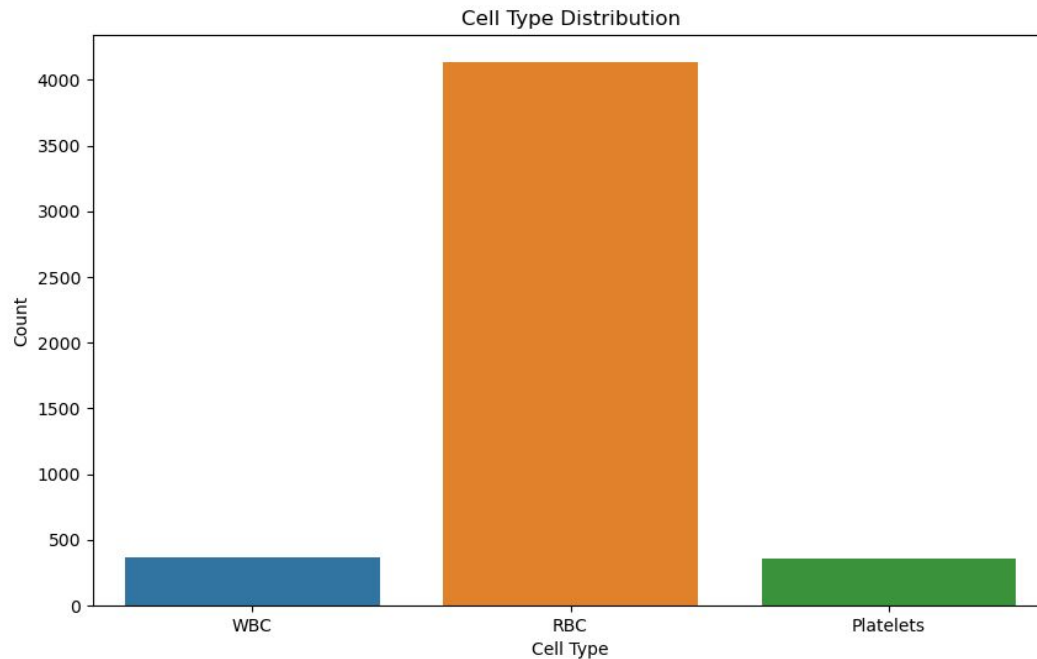


- The images are labeled with bounding box coordinates denoting where each WBC, RBC, and Platelet Cell are located.



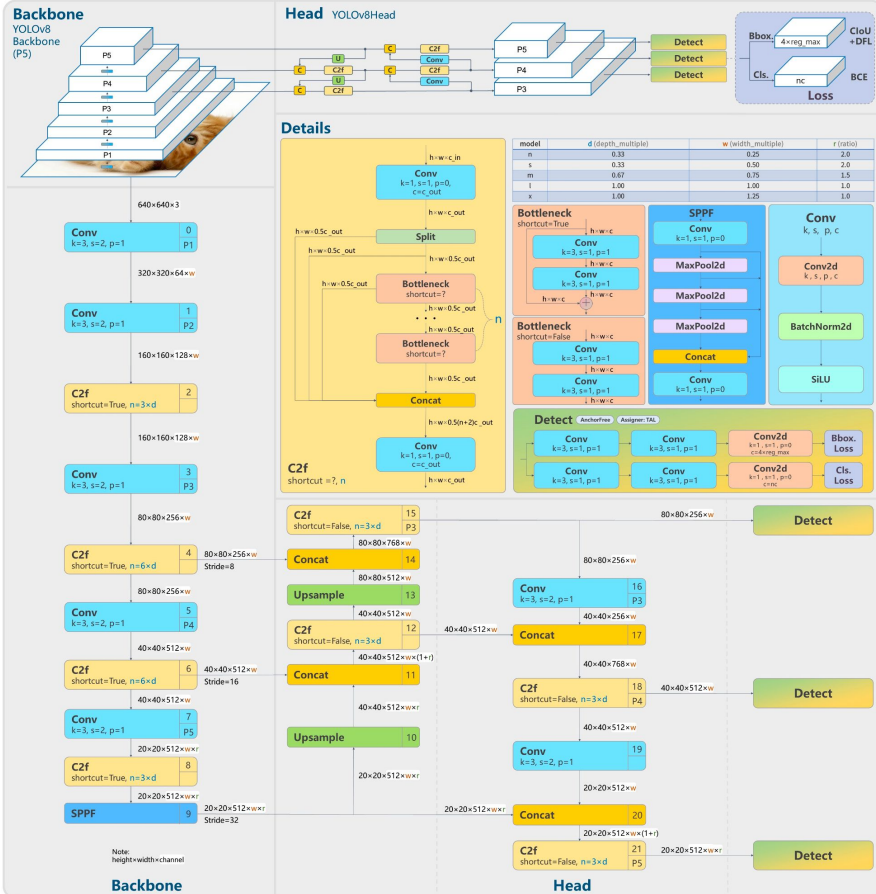
Objective 1: Data

- The dataset predictably contains far more RBC than WBC or Platelets
- In human blood samples, the typical ratio of RBC to WBC is about 600 to 1



Objective 1: Model

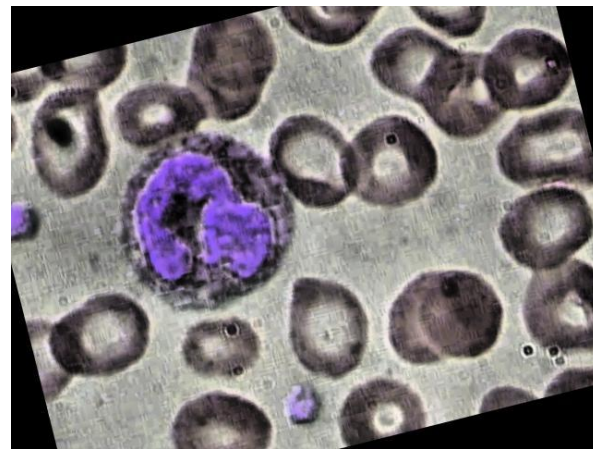
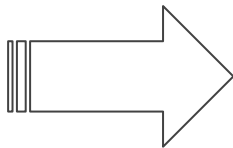
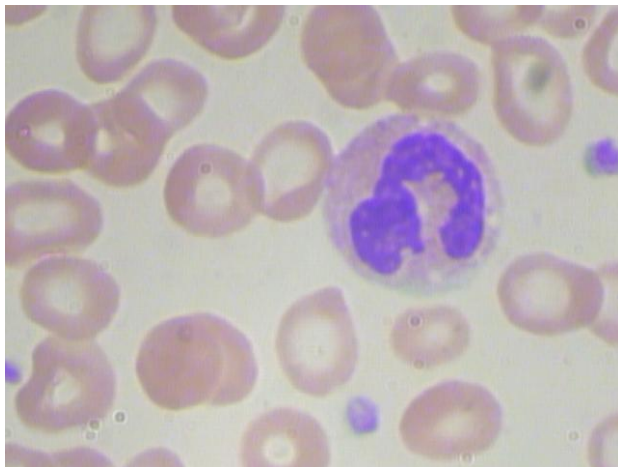
YOLOv8



- Goal: Train a bounding-box object detection model to automatically bound and classify WBC, RBC, and platelets
- To accomplish this goal, we trained our model using the YOLO v8 framework
- YOLO v8 is the most recent edition of Ultralytic's Pretrained "You Only Look Once" model^[4]

Preprocessing

- Two key components of preprocessing:
 - 1. Auto Adjust Contrast using Adaptive Equalization
 - 2. Augmentation (Rotating and flipping images in training set)





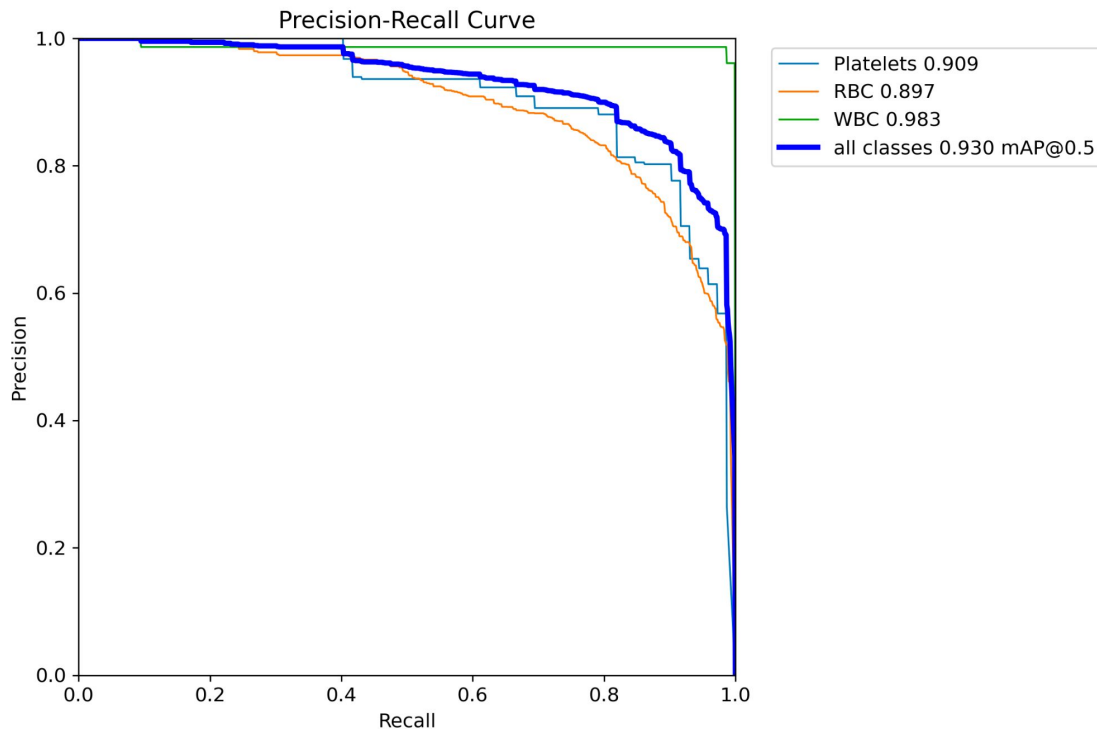
Training

- Object Detection models are computationally expensive to train, making it difficult to perform a Grid Search for hyperparameter tuning
- The best performing model we trained used these hyperparameters
- Model finished training in about 10 hours

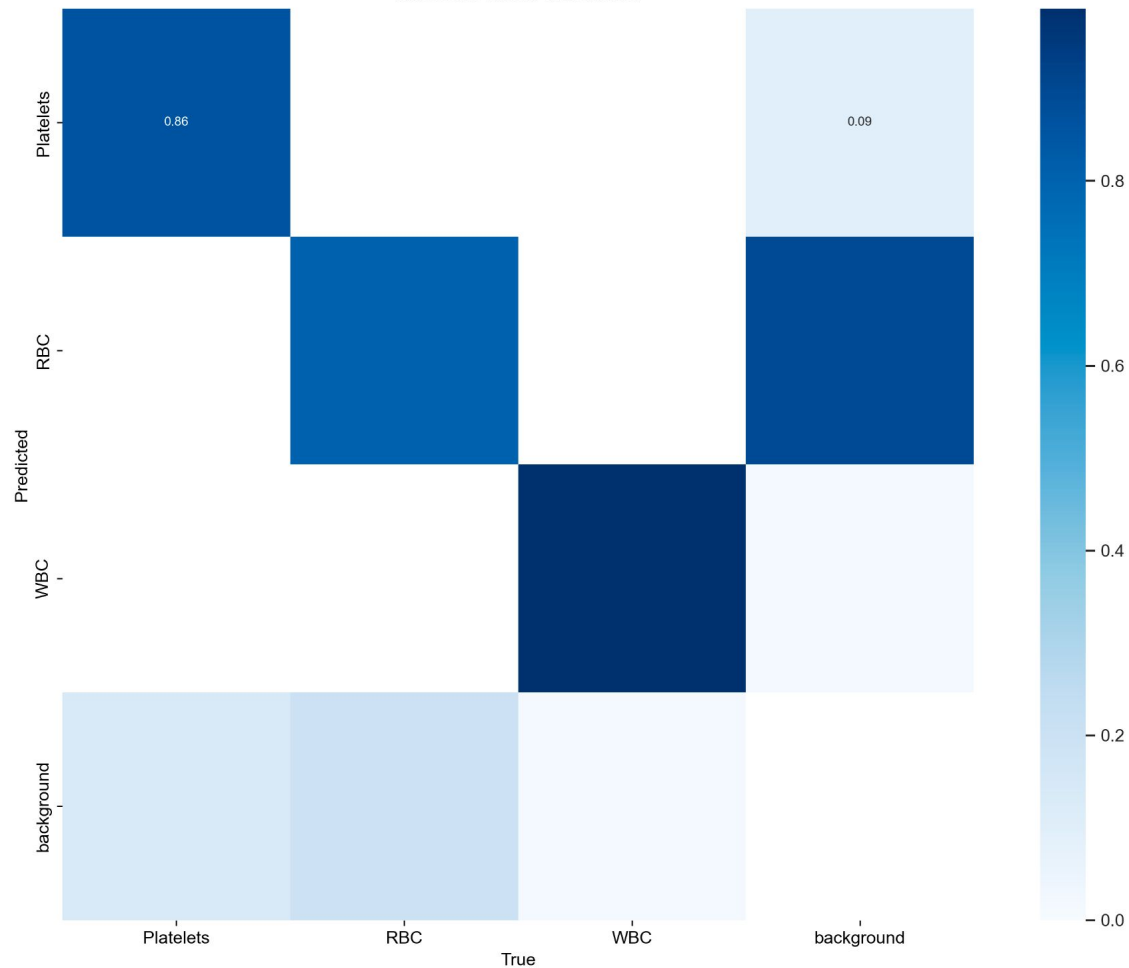
Hyper Parameter	Value
Learning Rate	0.001
Confidence Threshold	0.25
Non Maximum Suppression Threshold	0.45
Intersection Over Union Threshold	0.5
Number of Convolutional Layers	53
Epochs	25

Results

- Mean Average Precision (mAP): 0.93 across all classes
- Focus class (WBC) performed especially well
 - 0.983 mAP
 - Very high in both recall and precision metrics

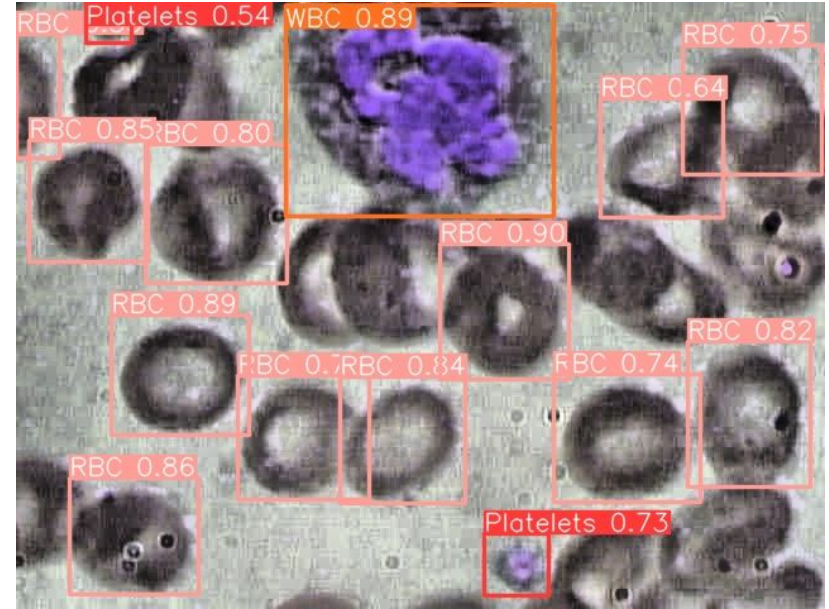
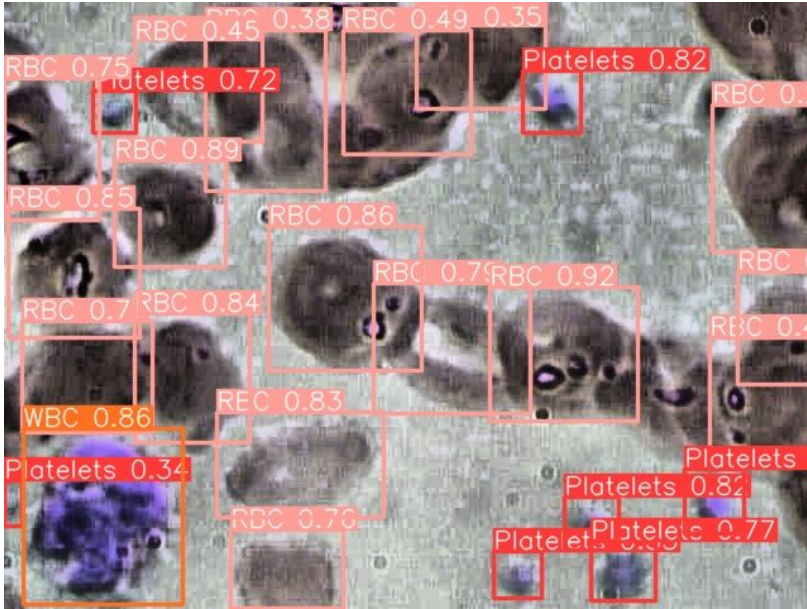


Confusion Matrix Normalized



- Most common mistake: failure to detect RBC.
- Very few false positives for any class

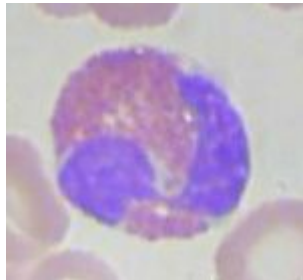
Results - Test Set Visualization



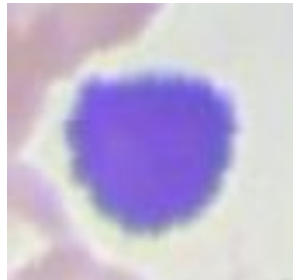
Objective 2: WBC Subtype Classification

- We wanted to go one layer deeper and train a model to classify WBC subtypes, as this is an important aspect of blood differential testing
- For this task we used a modified version of the original dataset which labels each WBC as one of four subtypes

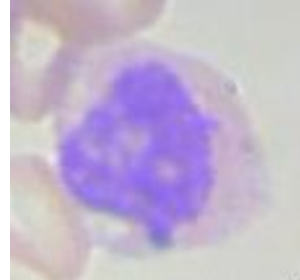
EOSINOPHIL



LYMPHOCYTE



MONOCYTE

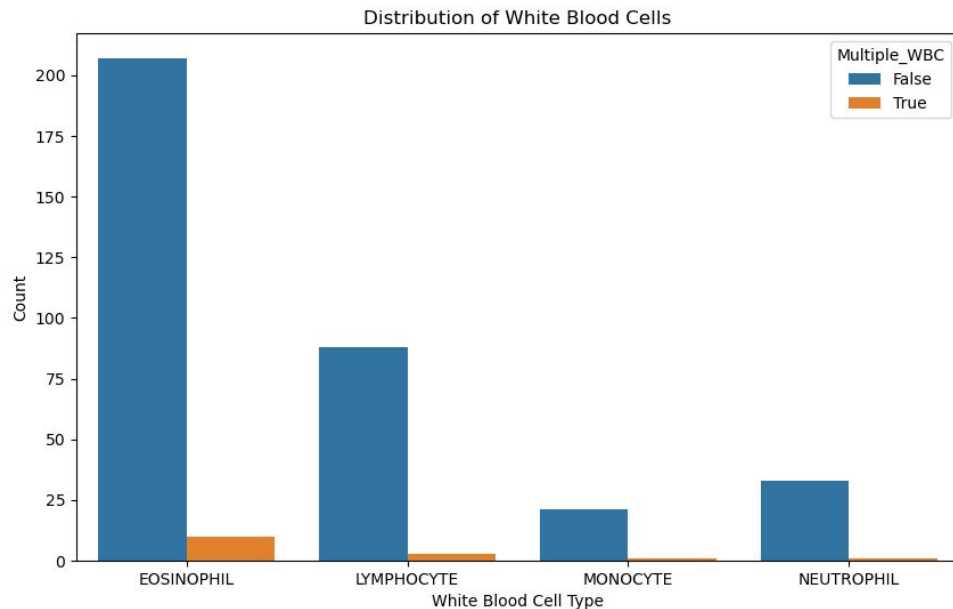


NEUTROPHIL





Objective 2: Data



- The majority of images in dataset 2 contain only one white blood cell
- The dataset is unbalanced towards Eosinophil White Blood Cell Types



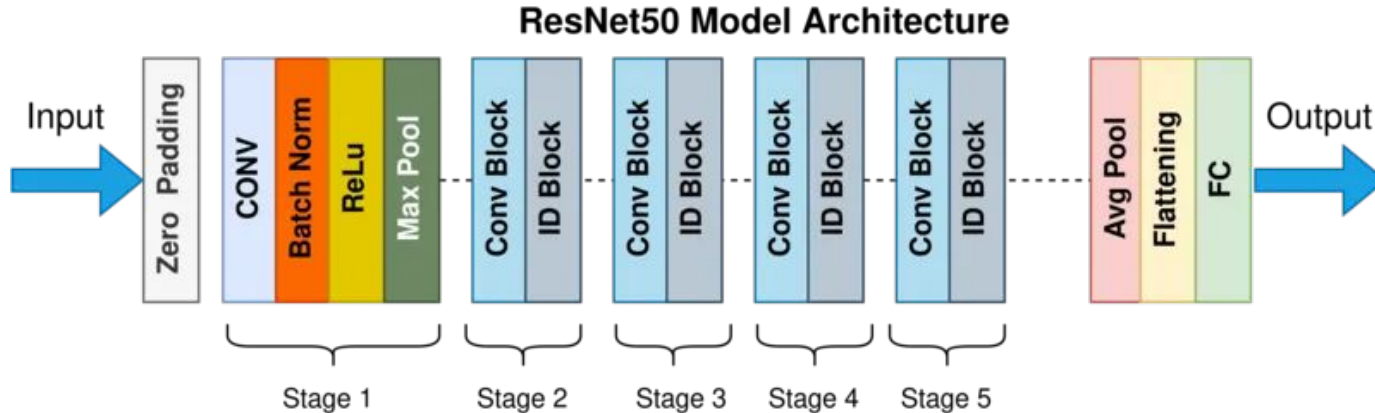
Model Selection

- Due to the visual similarity of WBC subtypes, accurate classification proved difficult
- We tested many models to compare performance
- ResNet-50 performed the best on the test set

Model	Test Set Accuracy
ResNet-50	87.62 %
VGG	41.77 %
K-Nearest Neighbors	38.0%
MobileNet	32.36 %
XGBoost	31.0 %
SVM	31.0 %
Random Forest	22.1 %

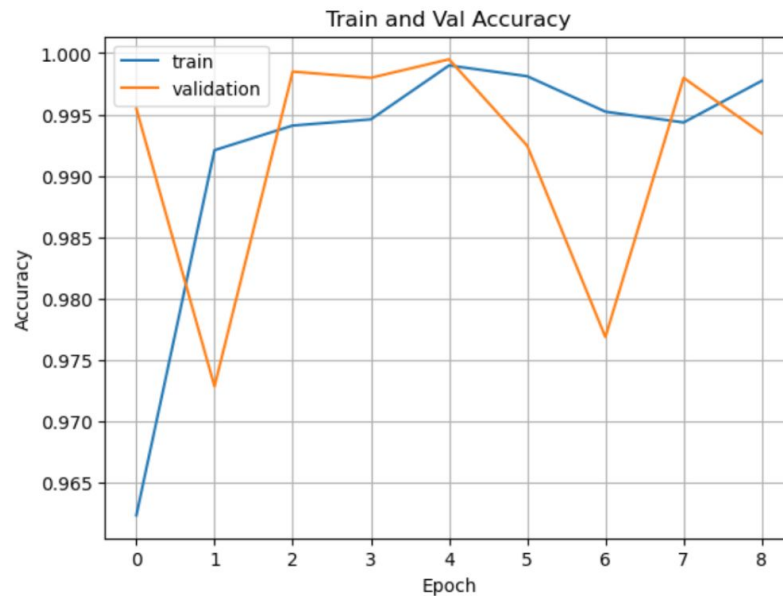
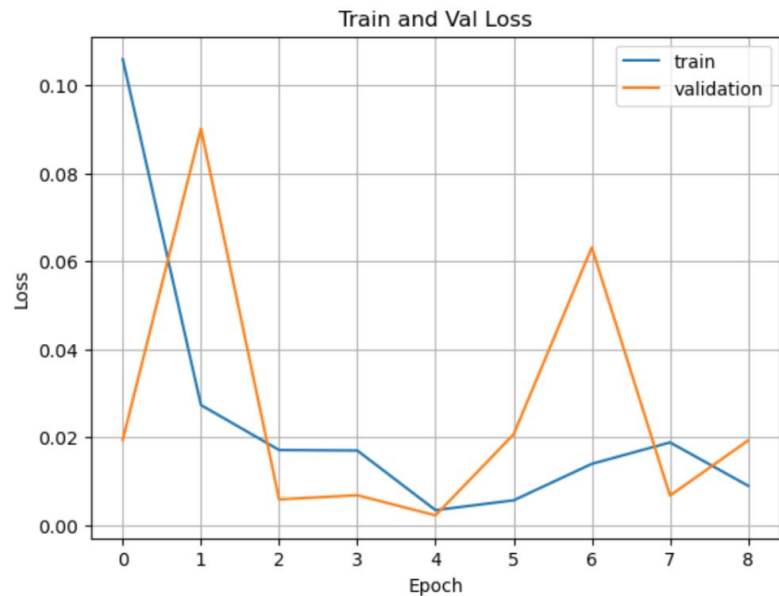
Model

- ResNet-50 is a deep convolutional neural network (CNN) developed by Microsoft Research^[5].
- Three main stages:
 - 1. Convolutional Layers: Extract features from the input images
 - ReLu activation function
 - Max Pooling and Batch Normalization
 - 2. Convolutional / Identity Block Hidden Layers: Learns Residual Functions to map the input to its label
 - 3. Fully Connected Layers: Generate the final classification



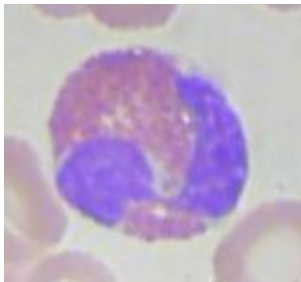


Training

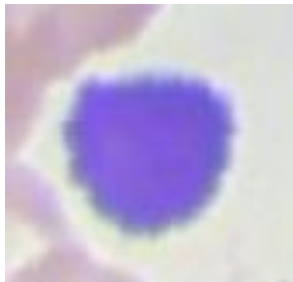


Results -- Test Set

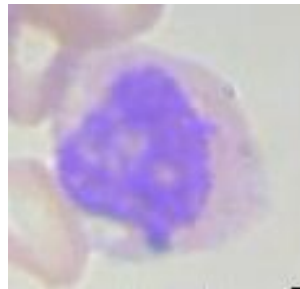
CLASS	PRECISION	RECALL	F1-SCORE
EOSINOPHIL	0.85	0.84	0.84
LYMPHOCYTE	1.00	1.00	1.00
MONOCYTE	0.99	0.93	0.85
NEUTROPHIL	0.74	0.93	0.82



EOSINOPHIL



LYMPHOCYTE



MONOCYTE

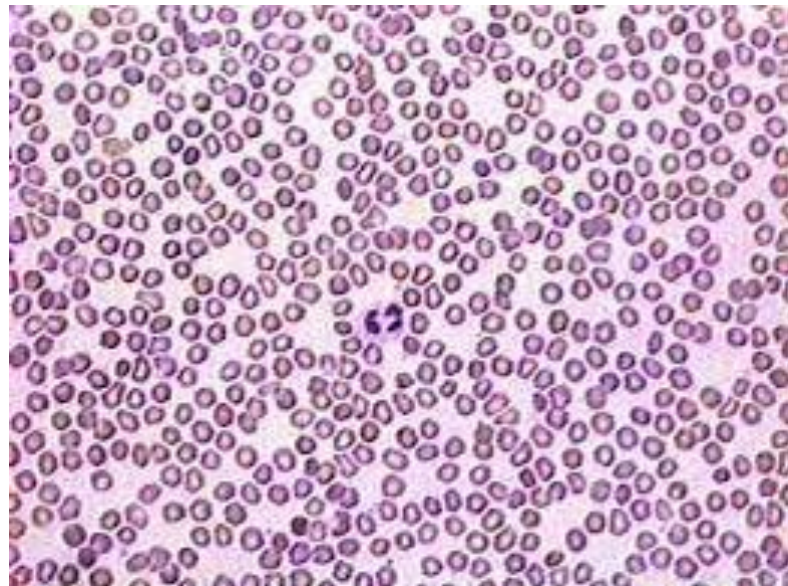


NEUTROPHIL

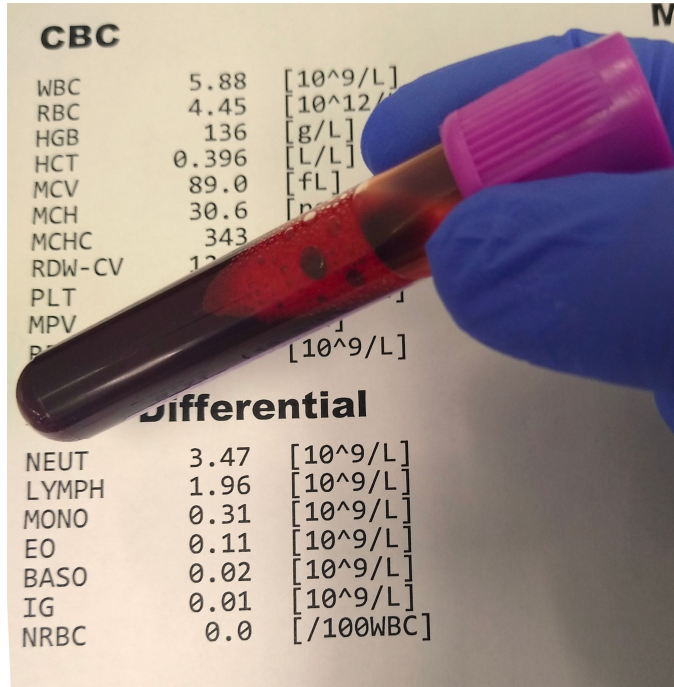


Limitations

- Increasing the number of labeled images would likely improve performance
- More time / computing power would allow for intensive hyperparameter tuning (i.e. through Grid Search Cross Validation)
- Although machine learning has the potential to reduce costs for CBC tests, imaging technology would still be necessary in order to capture images of blood samples to feed into the cell segmentation and classification models.



Conclusion



- Machine Learning models offer an innovative and cost effective method for completing CBC and Blood Differential tests
- Further exploration and experimentation in this field could lead to enhanced diagnostic accuracy, streamlined healthcare processes, and improved patient outcomes.



References

- [1] Leukemia and Lymphoma Society (2023). *Blood Cancer Statistics*. Retrieved on October 6, 2023 from: <https://www.lls.org/facts-and-statistics/facts-and-statistics-overview>
- [2] Dondelinger, Robert M. (2009). Hematology analyzers. *Biomed Instrumentation & Technology*. 2009 Jul; 43(4), 300-304. doi: 10.2345/0899-8205-43.4.300.
- [3] Mooney, P. (2018). *Blood Cell Images*. Version 6. Retrieved on September 20, 2023 from <https://www.kaggle.com/datasets/paultimothymooney/blood-cells/data>
- [4] Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLOv8 (Version 8.0.0) [Software]. AGPL-3.0. <https://github.com/ultralytics/ultralytics>
- [5] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. CoRR, abs/1512.03385. Retrieved from <http://arxiv.org/abs/1512.03385>

****Note:** Full Bibliography will be included in the Final Project Report