Persistence Homology for White Blood Cell Image Classification

Hamza Daruger

The University of Texas at Dallas

December 6, 2022

Background

Blood Cells - 3 types

- Red Blood Cells (erythrocytes)
- White Blood Cells (leukocytes)
  - Lymphocytes
  - Monocytes
  - Eosinophils
  - Neutrophils
  - Basophils
- Platelets (thrombocytes)

Why?

- Changes in WBCs can indicate health
- Making WBC tests more efficient

Dataset

4 classes: $\sim$3000 for each class after augmentation

- Lymphocytes - 33 images
- Monocytes - 19 images
- Neutrophils - 208 images
- Eosinophils - 86 images

Subsets used in experiment: 2 settings, 500 augmented images for each class with 80:20 train test split

- Binary - Eosinophils, Neutrophils
- Multi - Eosinophils, Neutrophils, Lymphocytes

Dataset (Cont.)

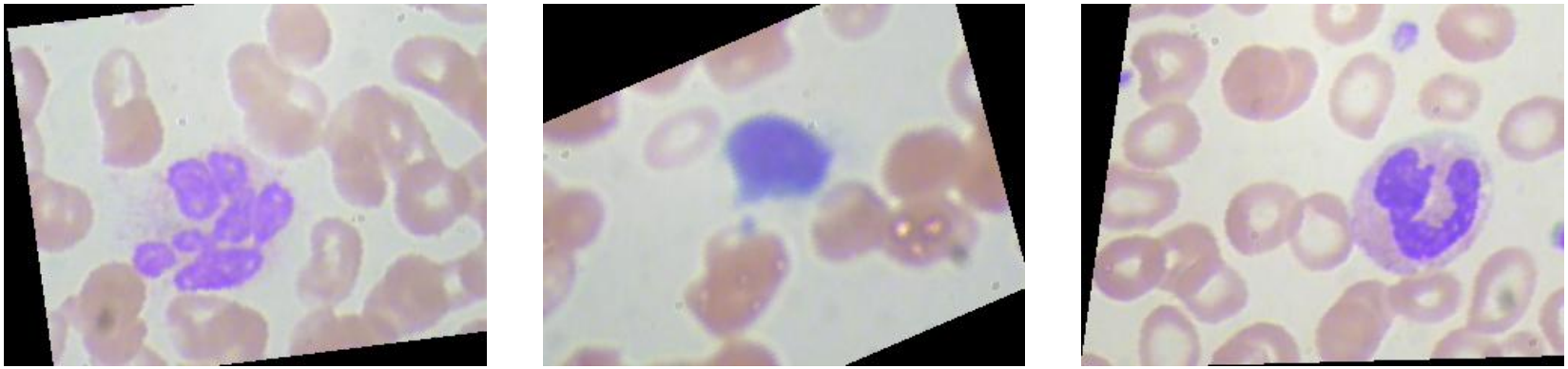

Figure: Eosinophil, Lymphocyte, Neutrophil images, respectively

Basic Methodology

1. Take RGB images and compress them into single channel images
2. Create Cubical Complexes for the images using Giotta-tda
3. Create persistence diagrams through sublevel or superlevel filtration
4. Vectorize persistence diagrams into its Betti Number for each homology dimension
5. Feed Betti Number vectors into XGBoost Classification Model for training

Betti Numbers (greyscale)

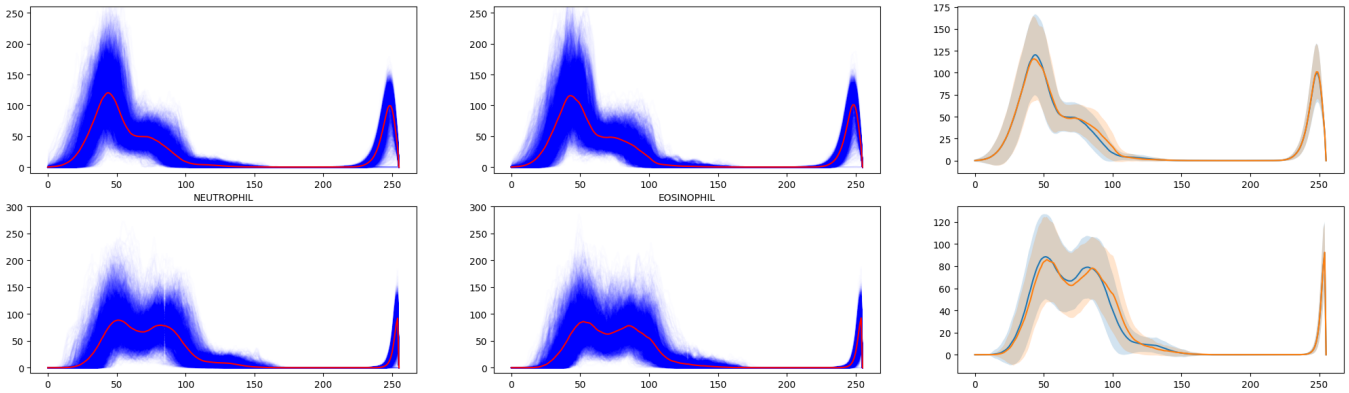

Figure: Binary

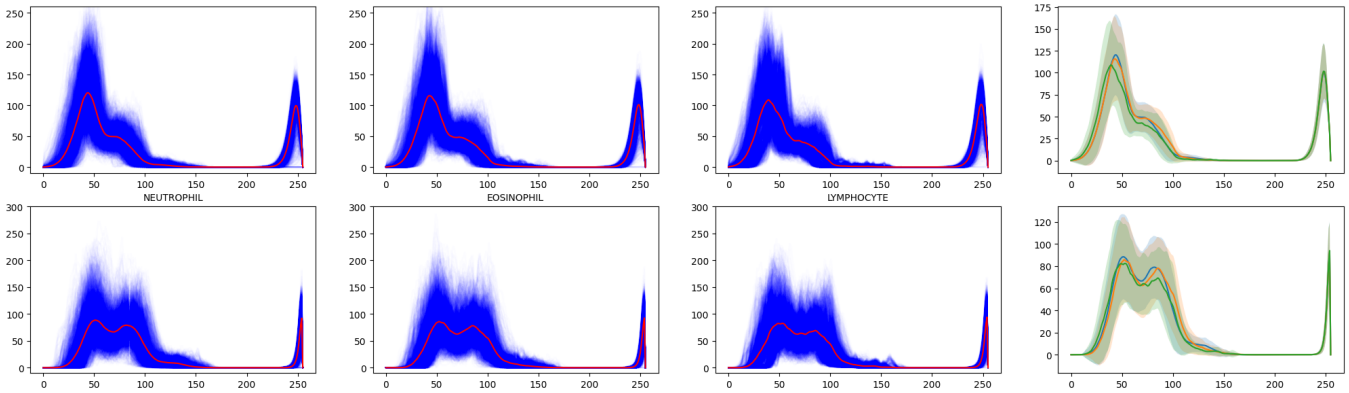

Figure: Multi

Betti Numbers(green)

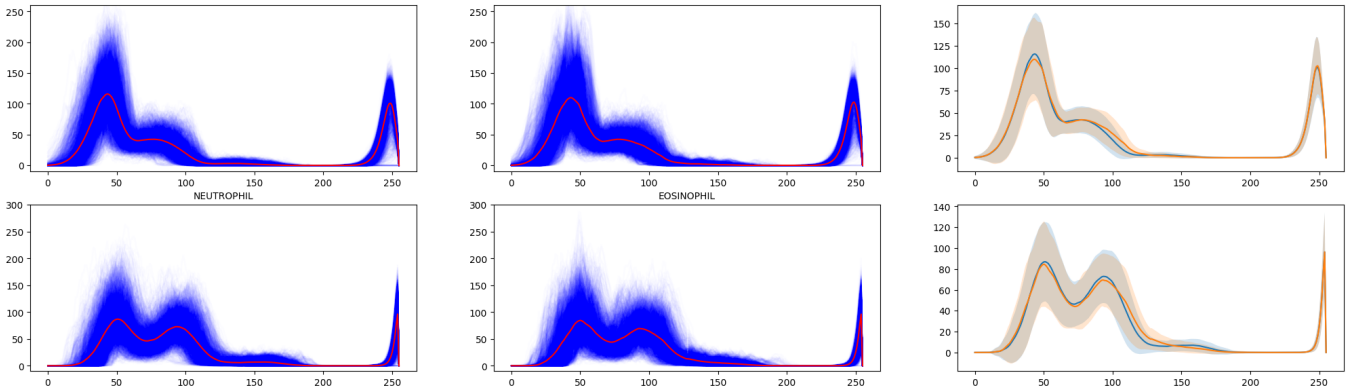

Figure: Binary

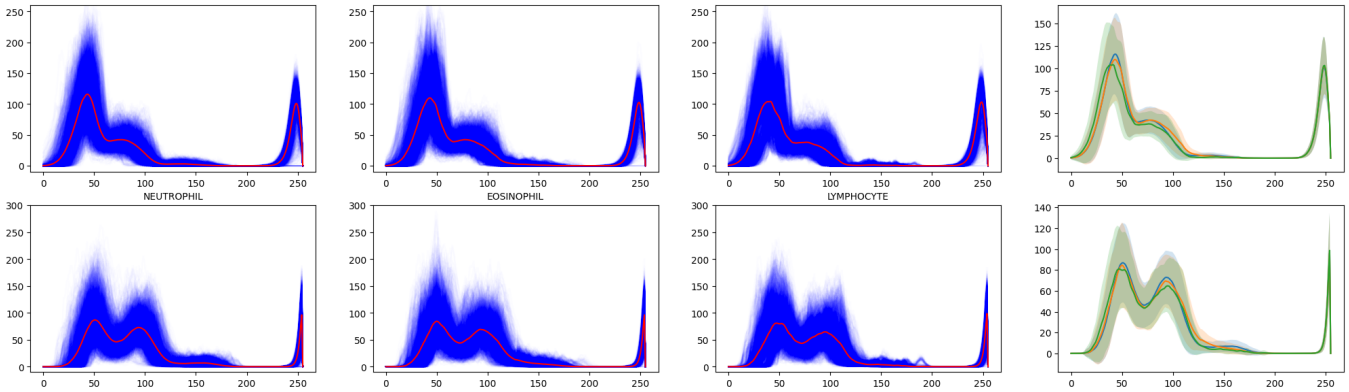

Figure: Multi

Parameter Tuning

Betti Curve:

- n_bins: 380 (green channel images), 420 (greyscale images)

XGBoost Classifier

- max_depth: 4
- gamma: 1
- learning_rate: 0.05 (binary), 0.08 (multi)
- n_estimators: 70 (binary), 90 (multi)

Results

Binary: Eosinophil, Neutrophil

96.5% accuracy, superlevel filtration, green image channel
95.4% accuracy, superlevel filtration, greyscale images

Binary: Eosinophil, Neutrophil, Lymphocyte

96.0% accuracy, superlevel filtration, green image channel
90.7% accuracy, superlevel filtration, greyscale images

Comparison

Reference	Method	Accuracy
(Habibzadeh et al., 2018)	Inception and ResNet	99%
(Bani-Hani et al., 2018)	GA-optimized CNN	91%
(Liang et al., 2018)	Hybrid CNN, Transfer learning, RNN, and LSTM	90.8%
(Sharma et al., 2019)	CNN and Transfer learning	87%
(Diouf et al., 2019)	ANN and CNN	97.7%
(Ghosh & Bhattacharya, 2020)	CNN and FCN on noise-free cell images	98.4%
(Ma et al., 2020)	DC-GAN, ResNet, and CNN	91.7%
(Banik et al., 2020)	CNN	96%

Table: Comparison of DL models trained on BCCD

from: Leukocytes Image Classification Using Optimized Convolutional Neural Networks (2022)