

Blood cell type classification problem

Data used:

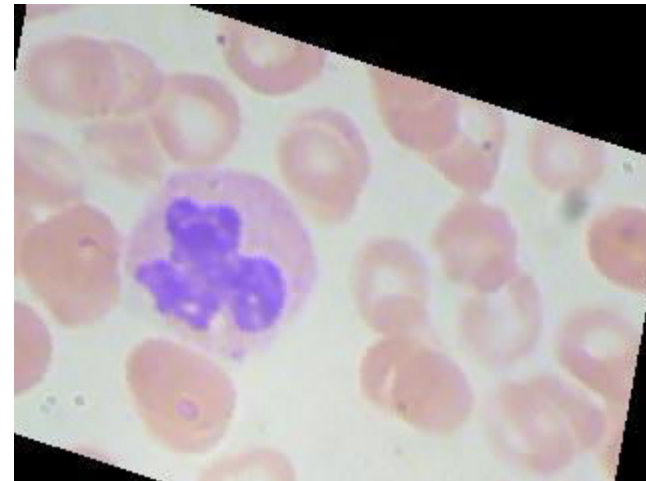
<https://www.kaggle.com/paultimothymooney/blood-cells/home>

Data description:

12500 images obtained through augmentation of 410 original images of blood cells. There are four types of cells: Eosinophil, Lymphocyte, Monocyte, and Neutrophil

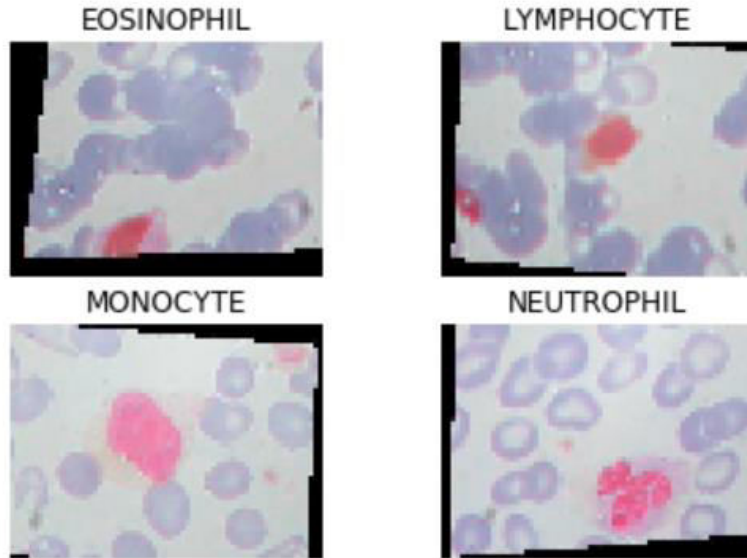
Motivation:

Automation of methods of blood cell type detection for medical applications.

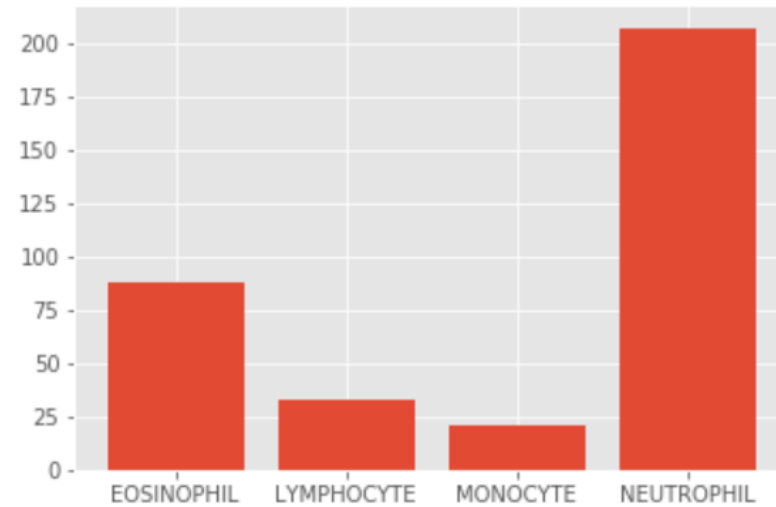


Preprocessing

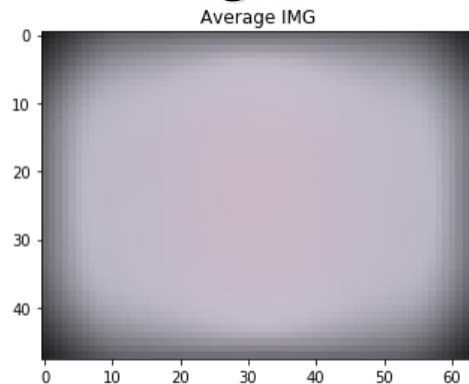
classes



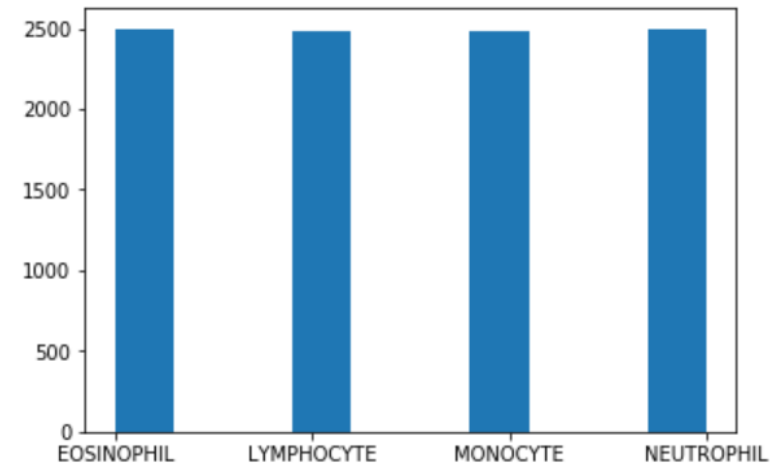
Raw samples per class(imbalanced)



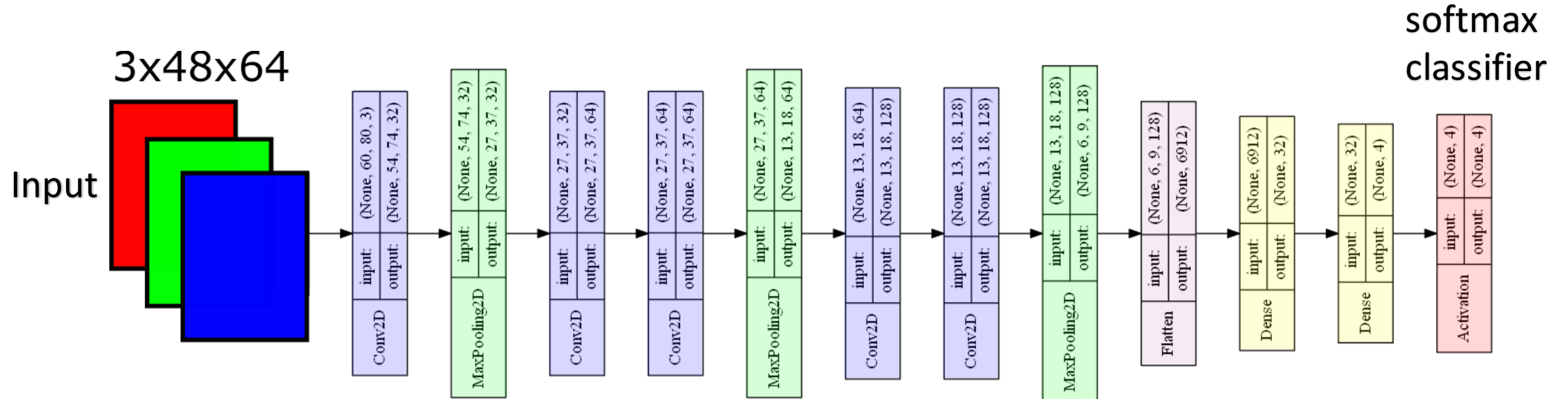
average background subtracted from each image



Augmented samples per class

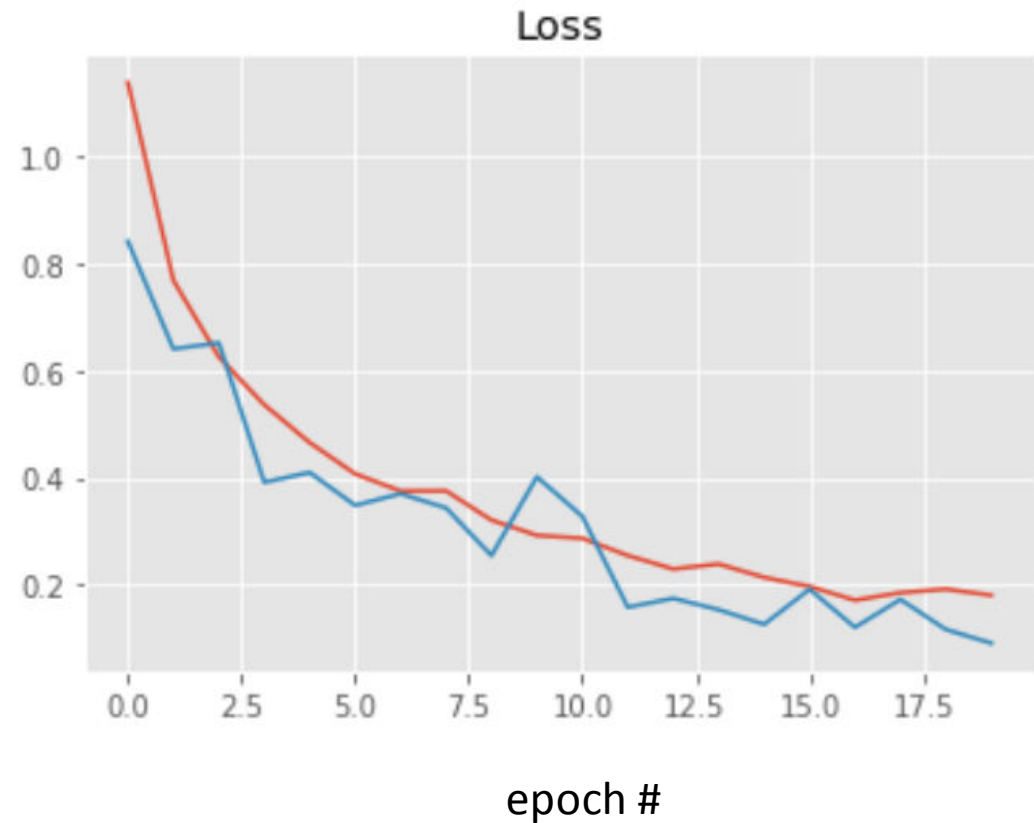
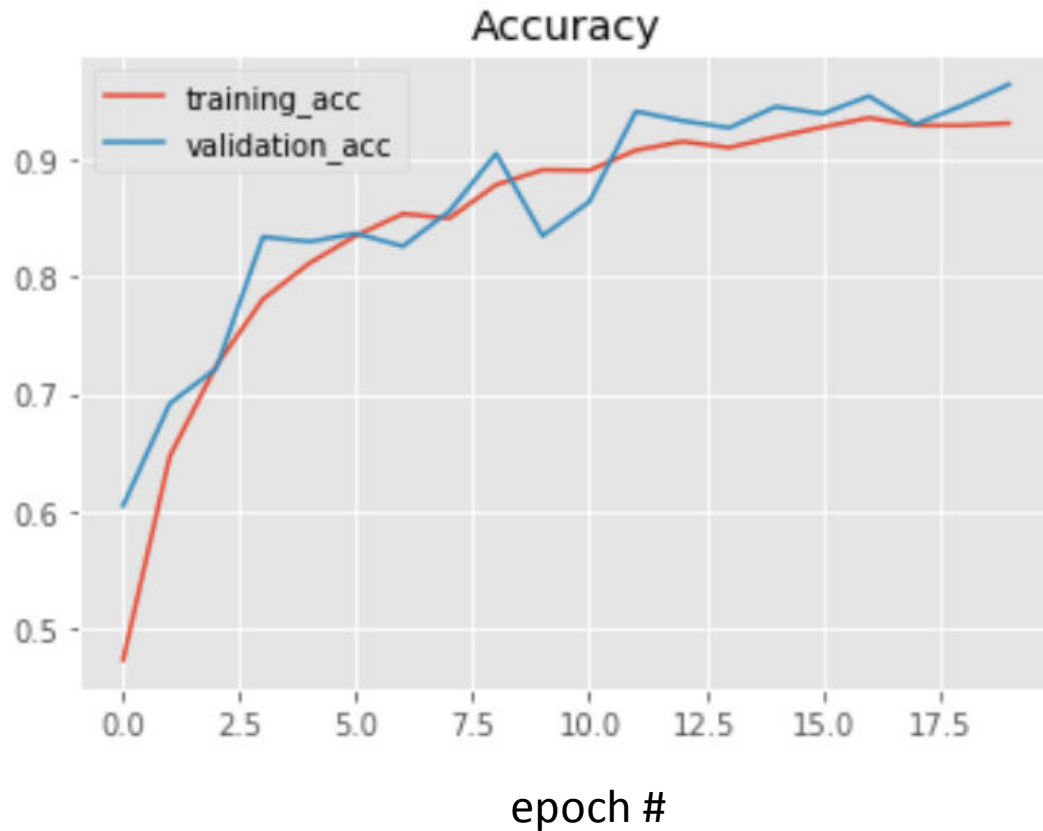


• CNN Architecture



There is a BatchNormalization layer after each Conv2D layer.
There is a 40% Dropout layer after each MaxPooling2D layer.

Training over 20 epochs

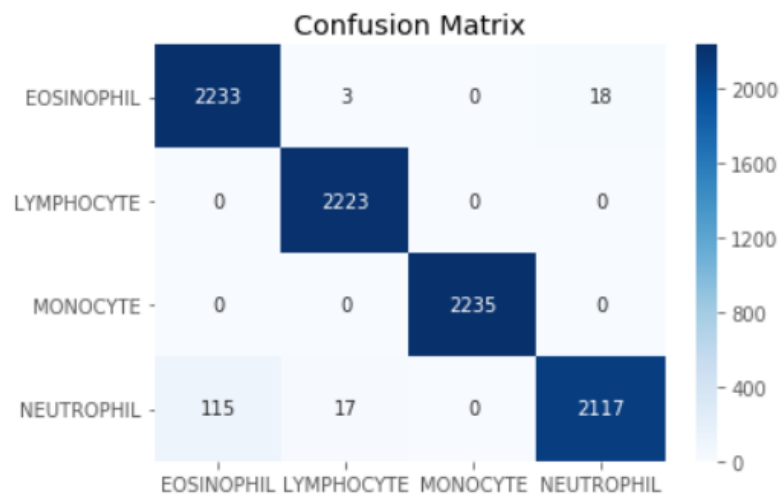


Training on CPU with ~6 minutes per epoch.

Classification reports for Training/Validation

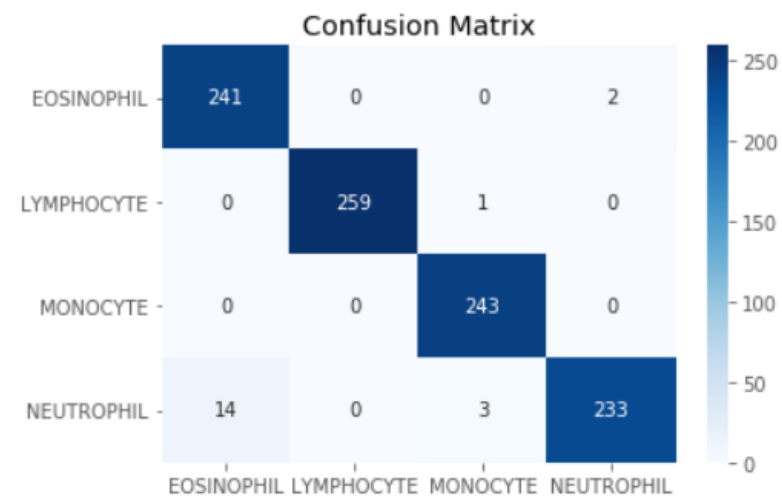
Training set

Classification Report				
	precision	recall	f1-score	support
EOSINOPHIL	0.95	0.99	0.97	2254
LYMPHOCYTE	0.99	1.00	1.00	2223
MONOCYTE	1.00	1.00	1.00	2235
NEUTROPHIL	0.99	0.94	0.97	2249
micro avg	0.98	0.98	0.98	8961
macro avg	0.98	0.98	0.98	8961
weighted avg	0.98	0.98	0.98	8961



Validation set

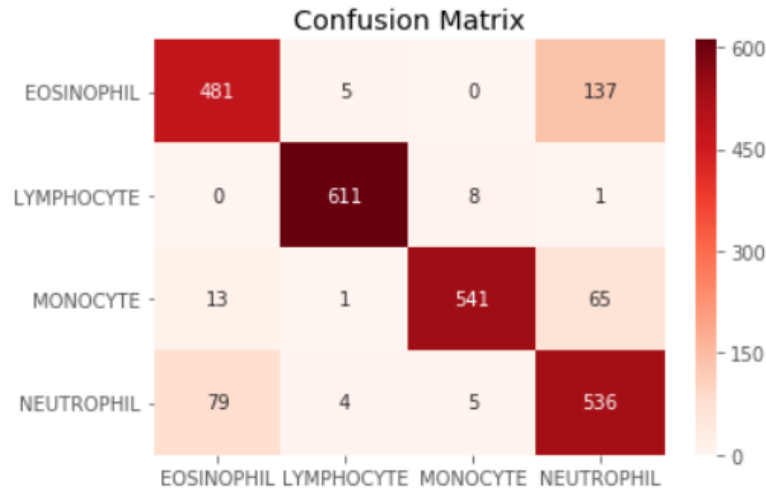
Classification Report				
	precision	recall	f1-score	support
EOSINOPHIL	0.95	0.99	0.97	243
LYMPHOCYTE	1.00	1.00	1.00	260
MONOCYTE	0.98	1.00	0.99	243
NEUTROPHIL	0.99	0.93	0.96	250
micro avg	0.98	0.98	0.98	996
macro avg	0.98	0.98	0.98	996
weighted avg	0.98	0.98	0.98	996



Ignore high scores for validation set. Original images were initially augmented. There is a lot of overlap between training and validation sets

Classification report for Test set

Classification Report				
	precision	recall	f1-score	support
EOSINOPHIL	0.84	0.77	0.80	623
LYMPHOCYTE	0.98	0.99	0.98	620
MONOCYTE	0.98	0.87	0.92	620
NEUTROPHIL	0.73	0.86	0.79	624
micro avg	0.87	0.87	0.87	2487
macro avg	0.88	0.87	0.87	2487
weighted avg	0.88	0.87	0.87	2487



- Accuracy: 87%
- Some EOSINOPHILS and MONOCYTES are incorrectly predicted as NEUTROPHILS. Due to imbalance of the raw set (200N, 80E, 20L, 30M) there is a bias towards the majority (N) class.
- LYMPHOCYTES are almost perfectly predicted due to being very different from other 3 classes. Their cell lacks any distinguishing structure, its just an isotropic blob.
- We need more images of E and M