Diagnosing Leukemia Using AI

Capstone Project Three
Springboard - School of Data

David Clark
September, 2021

Outline

- Present background and business problem.
- Discuss data acquisition and cleaning.
- Exploratory data analysis.
- Modeling.
- Final results and future directions.

Background

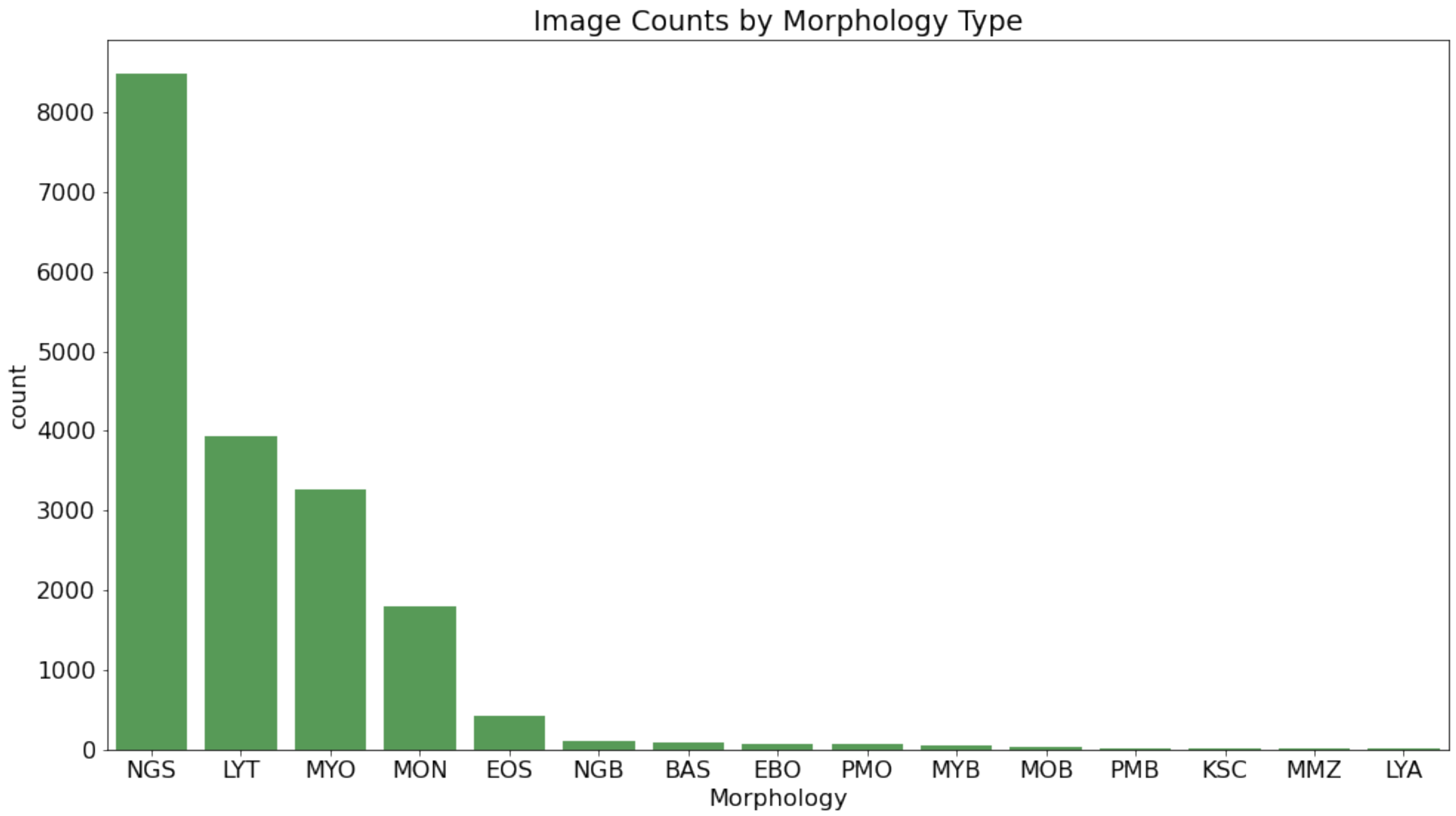
- Leukemia is a cancer of the blood often affecting young people.
- Pathology is past done by eye.
  - Time consuming.
  - Tedious.
- Automated approach would greatly reduce diagnosis time.
- Business problem: How can the doctor's at the Munich University Hospital automate the diagnosis of patients with leukemia using images from blood smears?

Data Acquisition

- Data taken from A Single-cell Morphological Dataset of Leukocytes from AML Patients and Non-malignant Controls (AML-Cytomorphology_LMU).
  - Part of Cancer Imaging Archive.
- 18,365 images.
- Each image 11Gb in size.
- 200 participants in study.

Morphology Classes

- 15 different leukocyte morphology types.
- Notice large class imbalance.



Mature vs Immature Leukocytes

- Leukocytes can be divided into two groups.
  - Mature.
  - Immature.

- Immature often associated with disease.

- Most leukocyte classes in study were mature.

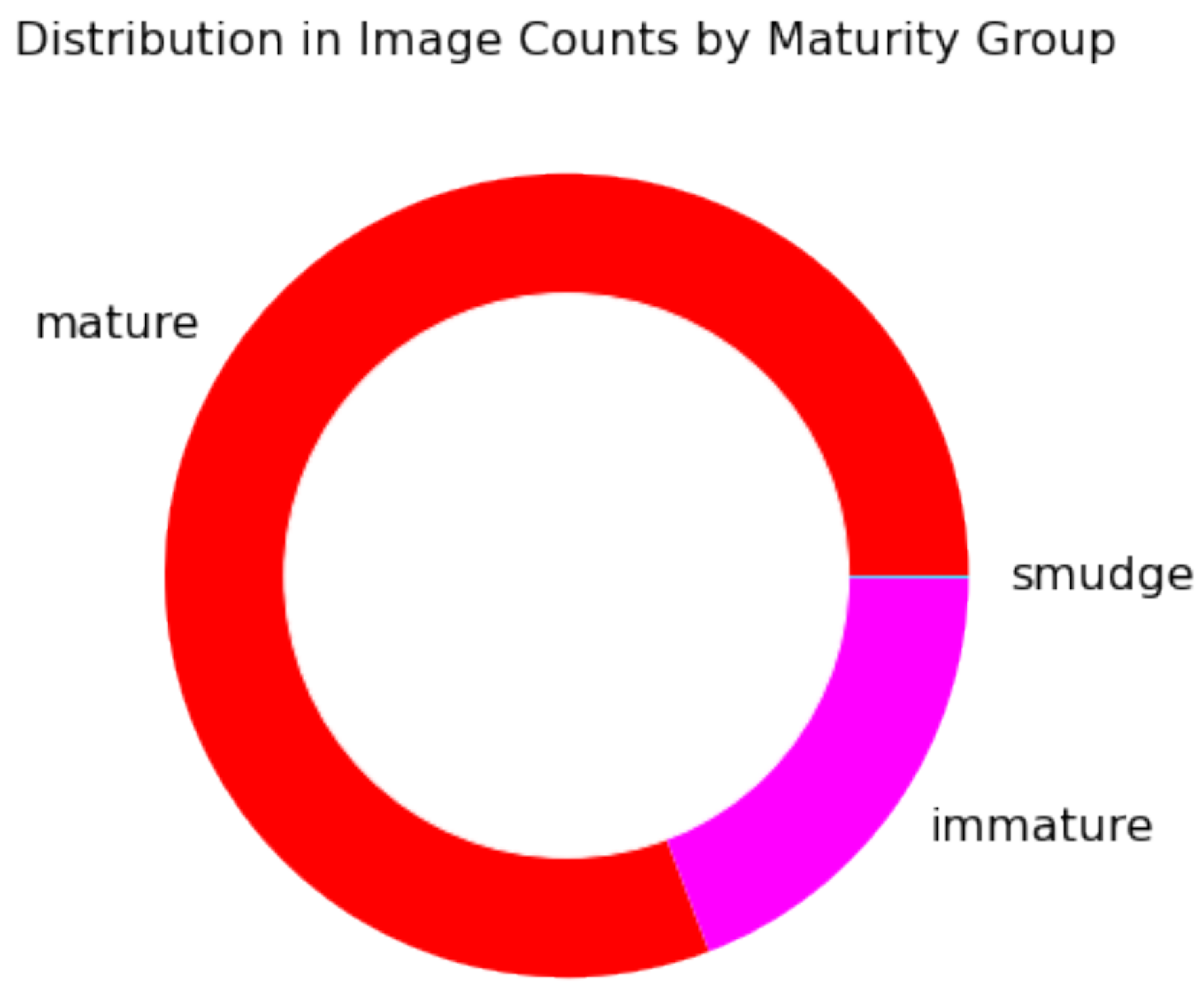


Image Examination

- Sample of images for each leukocyte morphological type.
- Leukocytes appear magenta.
- Notice varied shapes and sizes.

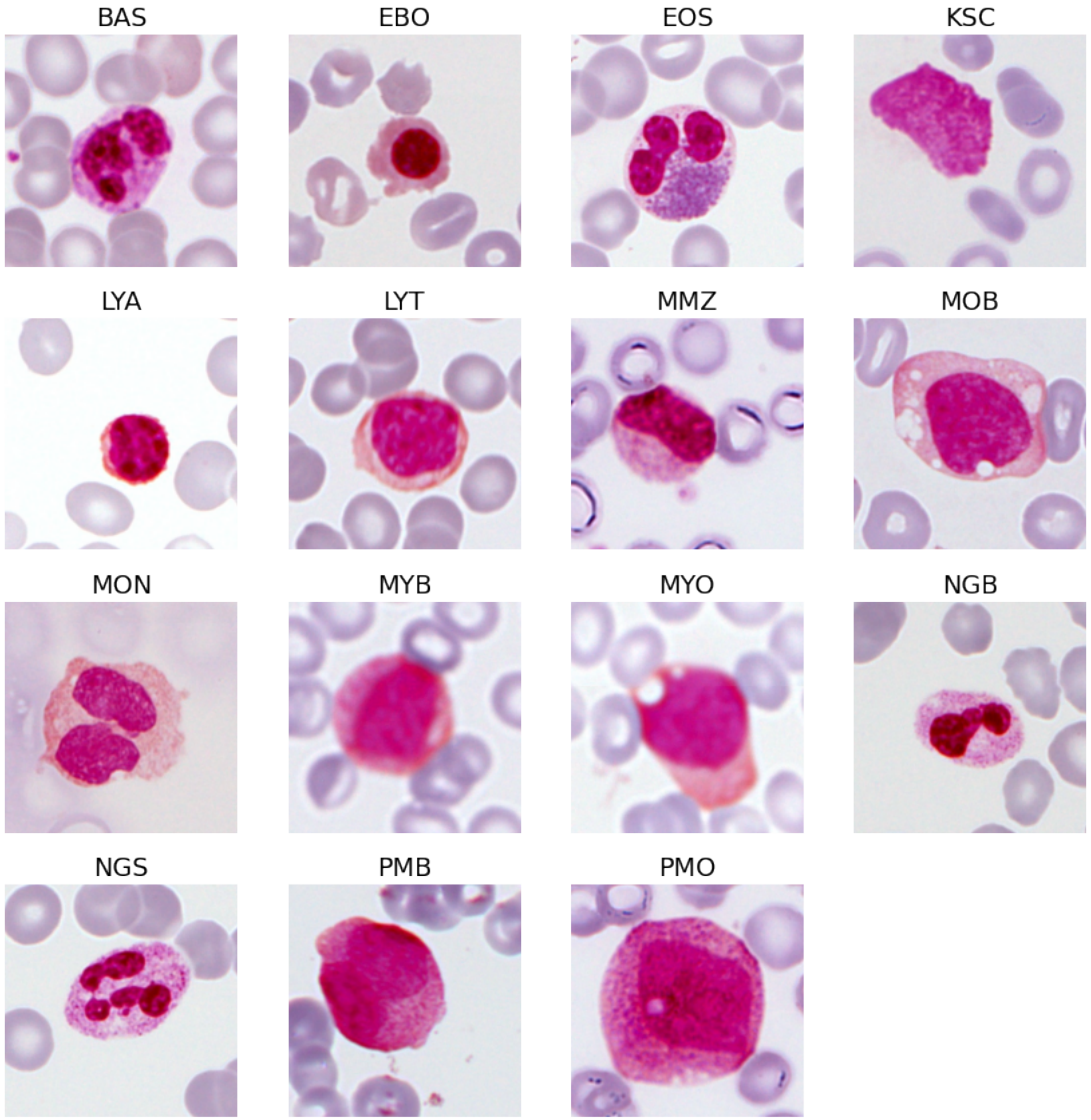


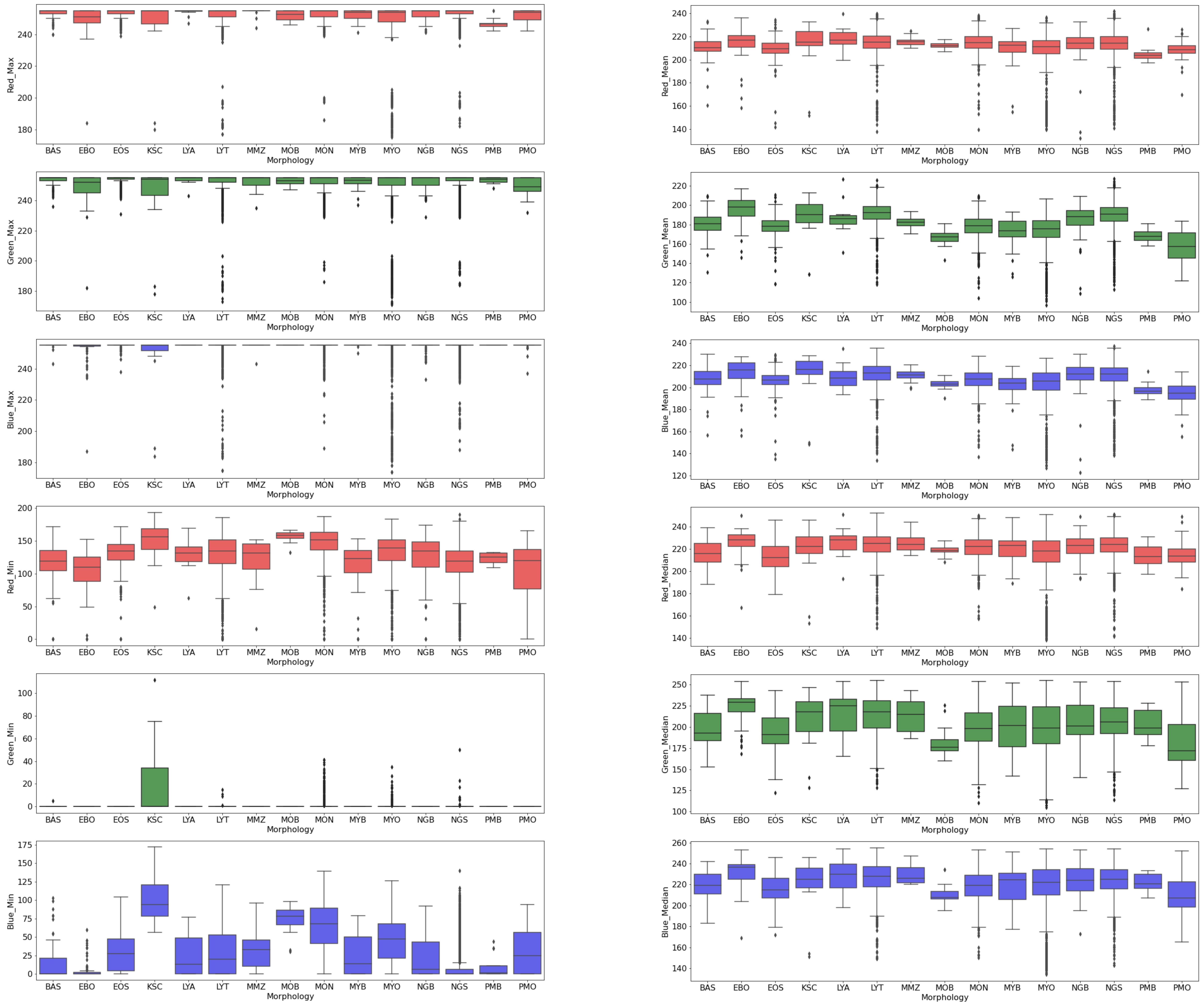
Image Examination

- Leukocyte images separated by color channel.
- Appear brightest in blue.

Image Statistics

- Image statistics, by color channel, for each leukocyte morphology type.

- Statistics include:
  - Maximum.
  - Minimum
  - Mean.
  - Median.



Baseline Models

- Picked four baseline models that are good at classification:
  - Logistic Regression.
  - Random Forest.
  - XGBoost.
  - Support Vector Classifier.

Baseline Models: Logistic Regression

- Poor F1 scores.
- F1 score drops to zero as class size drops.

Table 1: Logistic Regression Results Summary

morphology	precisio	recall	f1-score	support
NGS	0.74	0.73	0.73	1697
LYT	0.56	0.58	0.57	787
MYO	0.51	0.55	0.53	653
MON	0.27	0.27	0.27	358
PMO	0.25	0.14	0.18	14
EOS	0.03	0.02	0.03	85
BAS	0.00	0.00	0.00	16
EBO	0.00	0.00	0.00	16
KSC	0.00	0.00	0.00	3
LYA	0.00	0.00	0.00	2
MMZ	0.00	0.00	0.00	3
MOB	0.00	0.00	0.00	5
MYB	0.00	0.00	0.00	8
NGB	0.00	0.00	0.00	22
PMB	0.00	0.00	0.00	4

Baseline Models: Random Forest

- Best F1 score improves to 0.9.
- F1 score drops to zero as class size drops.

Table 2: Random Forest Results Summary				
morphology	precision	recall	f1-score	support
NGS	0.84	0.98	0.90	1697
LYT	0.90	0.75	0.82	787
MYO	0.64	0.89	0.74	653
MON	0.78	0.27	0.40	358
EBO	1.00	0.06	0.12	16
BAS	0.00	0.00	0.00	16
EOS	0.00	0.00	0.00	85
KSC	0.00	0.00	0.00	3
LYA	0.00	0.00	0.00	2
MMZ	0.00	0.00	0.00	3
MOB	0.00	0.00	0.00	5
MYB	0.00	0.00	0.00	8
NGB	0.00	0.00	0.00	22
PMB	0.00	0.00	0.00	4
PMO	0.00	0.00	0.00	14

Baseline Models: XGBoost

- Best F1 score improves more, to 0.95.
- F1 score drops to zero as class size drops.

Table 3: XGBoost Results Summary

morphology	precision	recall	f1-score	support
NGS	0.94	0.97	0.95	1697
LYT	0.89	0.89	0.89	787
MYO	0.74	0.87	0.80	653
MON	0.67	0.61	0.64	358
EOS	0.83	0.35	0.50	85
EBO	1.00	0.06	0.12	16
BAS	0.00	0.00	0.00	16
KSC	0.00	0.00	0.00	3
LYA	0.00	0.00	0.00	2
MMZ	0.00	0.00	0.00	3
MOB	0.00	0.00	0.00	5
MYB	0.00	0.00	0.00	8
NGB	0.00	0.00	0.00	22
PMB	0.00	0.00	0.00	4
PMO	0.00	0.00	0.00	14

Baseline Models: Support Vector Classifier

- No improvement in top F1 score.
- F1 score drops to zero as class size drops.

Table 4: Support Vector Classifier Results Summary

morphology	precision	recall	f1-score	support
NGS	0.94	0.97	0.95	1697
LYT	0.89	0.89	0.89	787
MYO	0.74	0.87	0.80	653
MON	0.67	0.61	0.64	358
EOS	0.83	0.35	0.50	85
EBO	1.00	0.06	0.12	16
BAS	0.00	0.00	0.00	16
KSC	0.00	0.00	0.00	3
LYA	0.00	0.00	0.00	2
MMZ	0.00	0.00	0.00	3
MOB	0.00	0.00	0.00	5
MYB	0.00	0.00	0.00	8
NGB	0.00	0.00	0.00	22
PMB	0.00	0.00	0.00	4
PMO	0.00	0.00	0.00	14

Deep Learning Architecture

- Convolutional neural networks are great at image recognition.

- Create simple CNN with two convolutional layers.

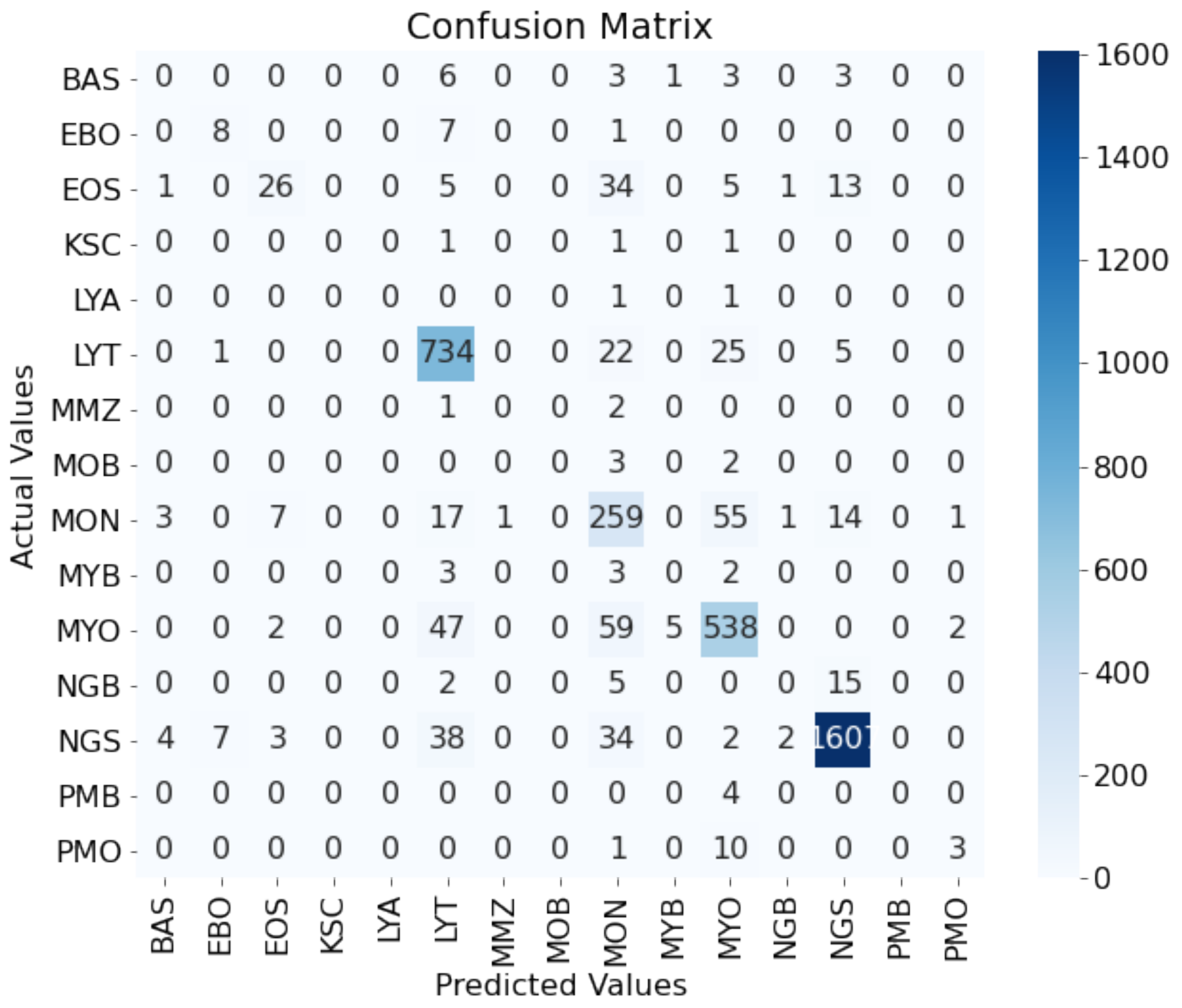
```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 48, 48, 8)         80
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 24, 24, 8)         0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 18, 18, 16)        6288
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 9, 9, 16)          0
_________________________________________________________________
flatten (Flatten)            (None, 1296)              0
_________________________________________________________________
dense (Dense)                (None, 600)               778200
_________________________________________________________________
dense_1 (Dense)              (None, 150)               90150
_________________________________________________________________
dense_2 (Dense)              (None, 38)                5738
_________________________________________________________________
dense_3 (Dense)              (None, 15)                585
=================================================================
Total params: 881,041
Trainable params: 881,041
Non-trainable params: 0
_________________________________________________________________
```

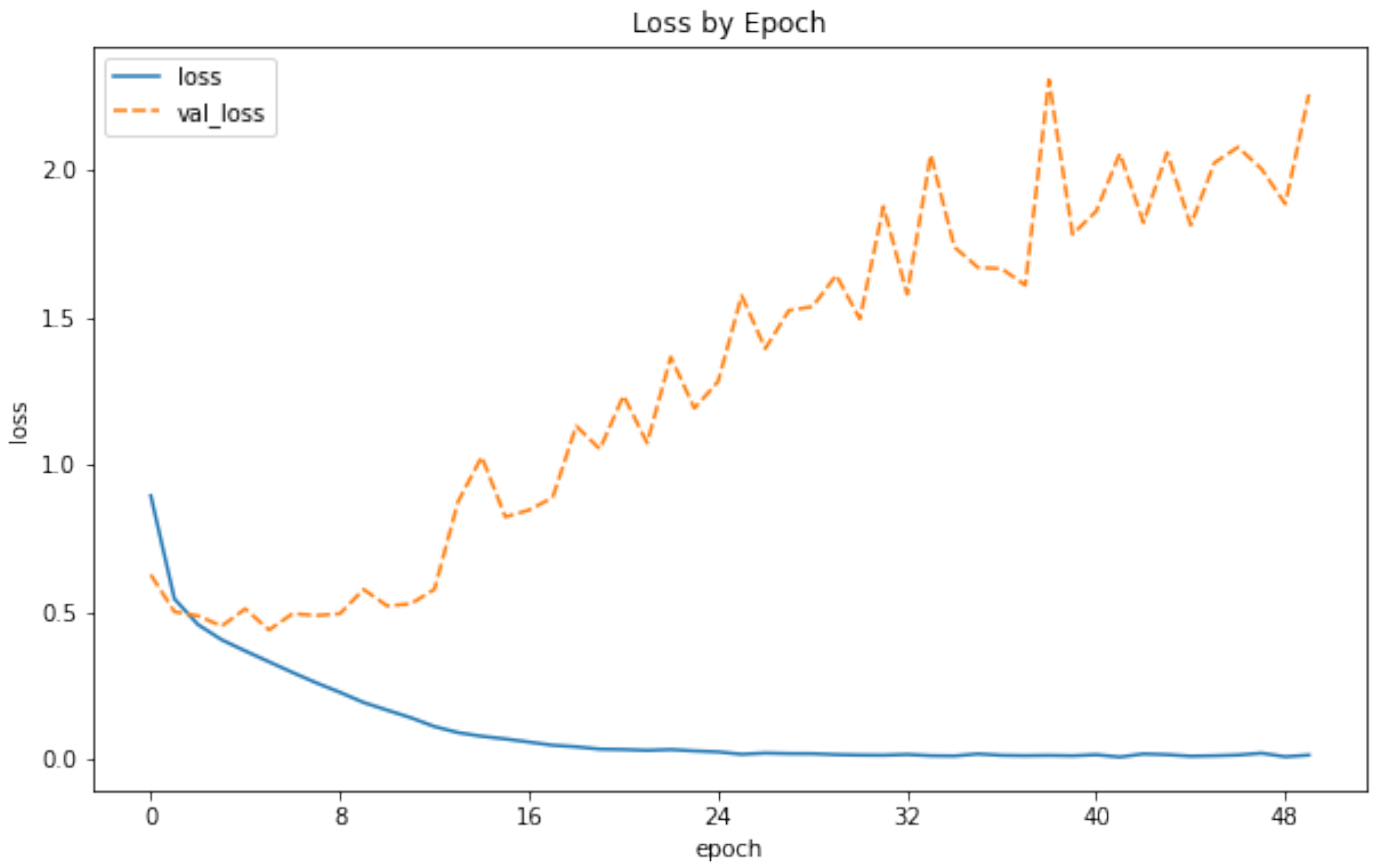
CNN Model Evaluation

- Evaluated CNN models using three strategies:
  - Learning Curve.
  - Classification Report.
  - Confusion Matrix.

CNN Model 1A

Table 5: Classification Report for Model 1a				
morphology	**precision**	**recall**	**f1-score**	**support**
NGS	0.95	0.95	0.95	1697
LYT	0.88	0.91	0.90	787
MYO	0.80	0.83	0.81	653
MON	0.59	0.68	0.63	358
EBO	0.45	0.31	0.37	16
EOS	0.40	0.27	0.32	85
PMO	0.12	0.07	0.09	14
BAS	0.00	0.00	0.00	16
KSC	0.00	0.00	0.00	3
LYA	0.00	0.00	0.00	2
MMZ	0.00	0.00	0.00	3
MOB	0.00	0.00	0.00	5
MYB	0.00	0.00	0.00	8
NGB	0.00	0.00	0.00	22
PMB	0.00	0.00	0.00	4

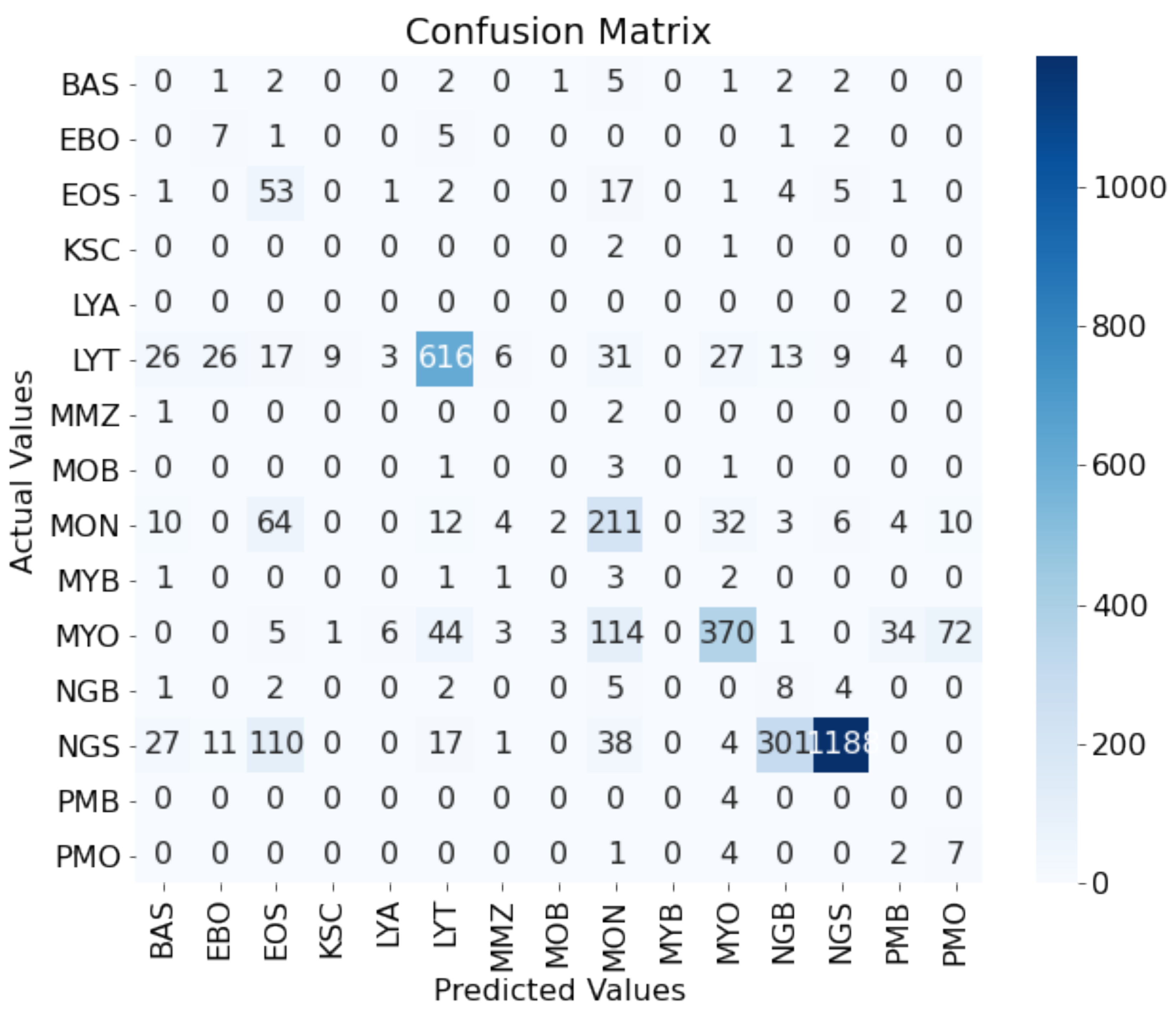


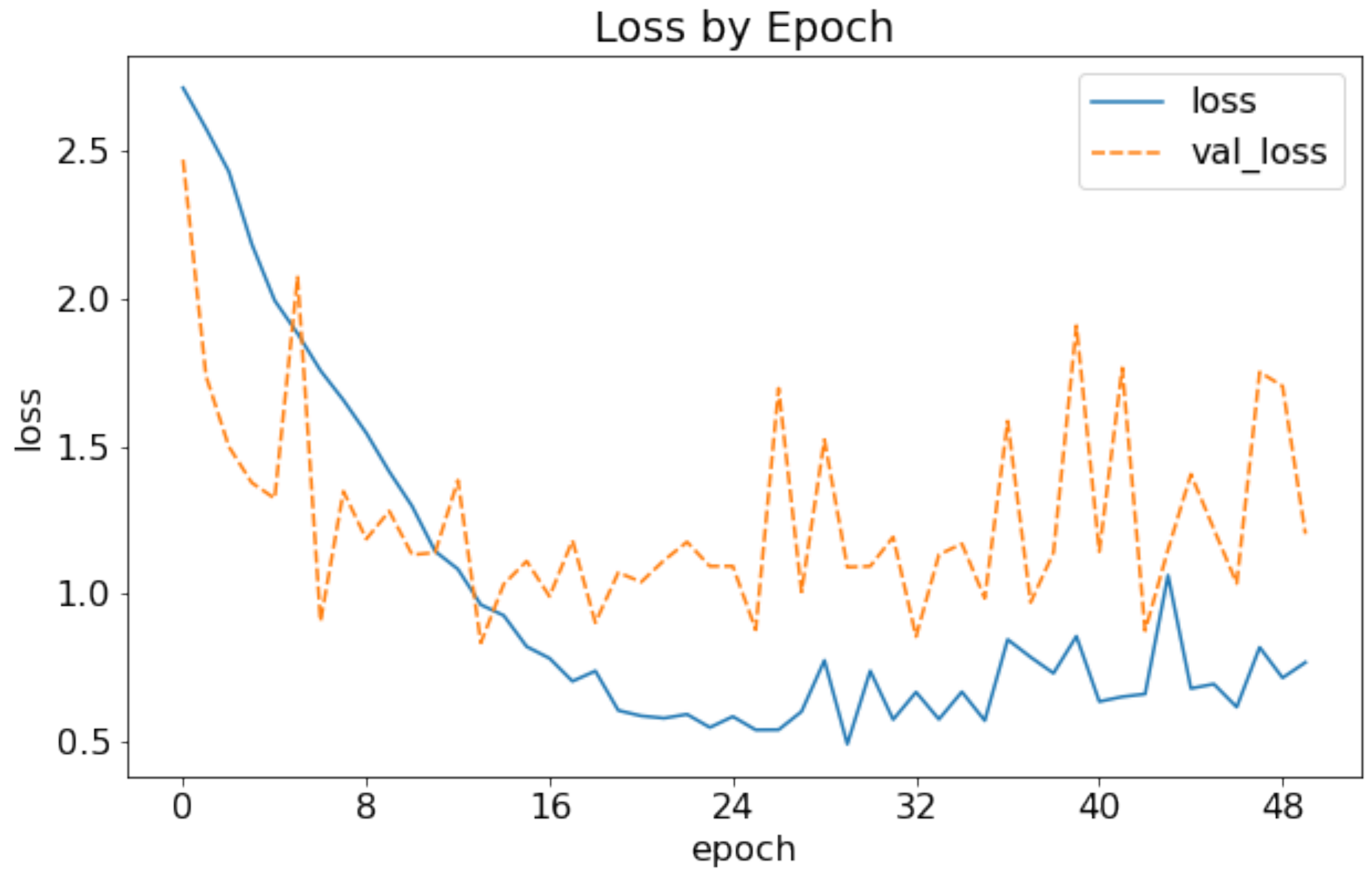


- Multi-classification problem.
- No account for class imbalance.

CNN Model 1B

Table 6: Classification Report for Model 1b				
morphology	**precision**	**recall**	**f1-score**	**support**
LYT	0.17	0.12	0.14	16
MYB	1.00	0.04	0.08	1697
PMB	0.03	0.29	0.06	14
KSC	0.56	0.03	0.05	358
MYO	0.02	0.04	0.03	85
EOS	1.00	0.02	0.03	787
NGB	0.02	0.25	0.03	4
NGS	0.00	0.44	0.01	16
LYA	0.01	0.25	0.01	8
MON	0.00	0.00	0.00	3
EBO	0.00	0.50	0.00	2
PMO	0.00	0.00	0.00	3
BAS	0.00	0.00	0.00	5
MMZ	1.00	0.00	0.00	653
MOB	0.00	0.00	0.00	22

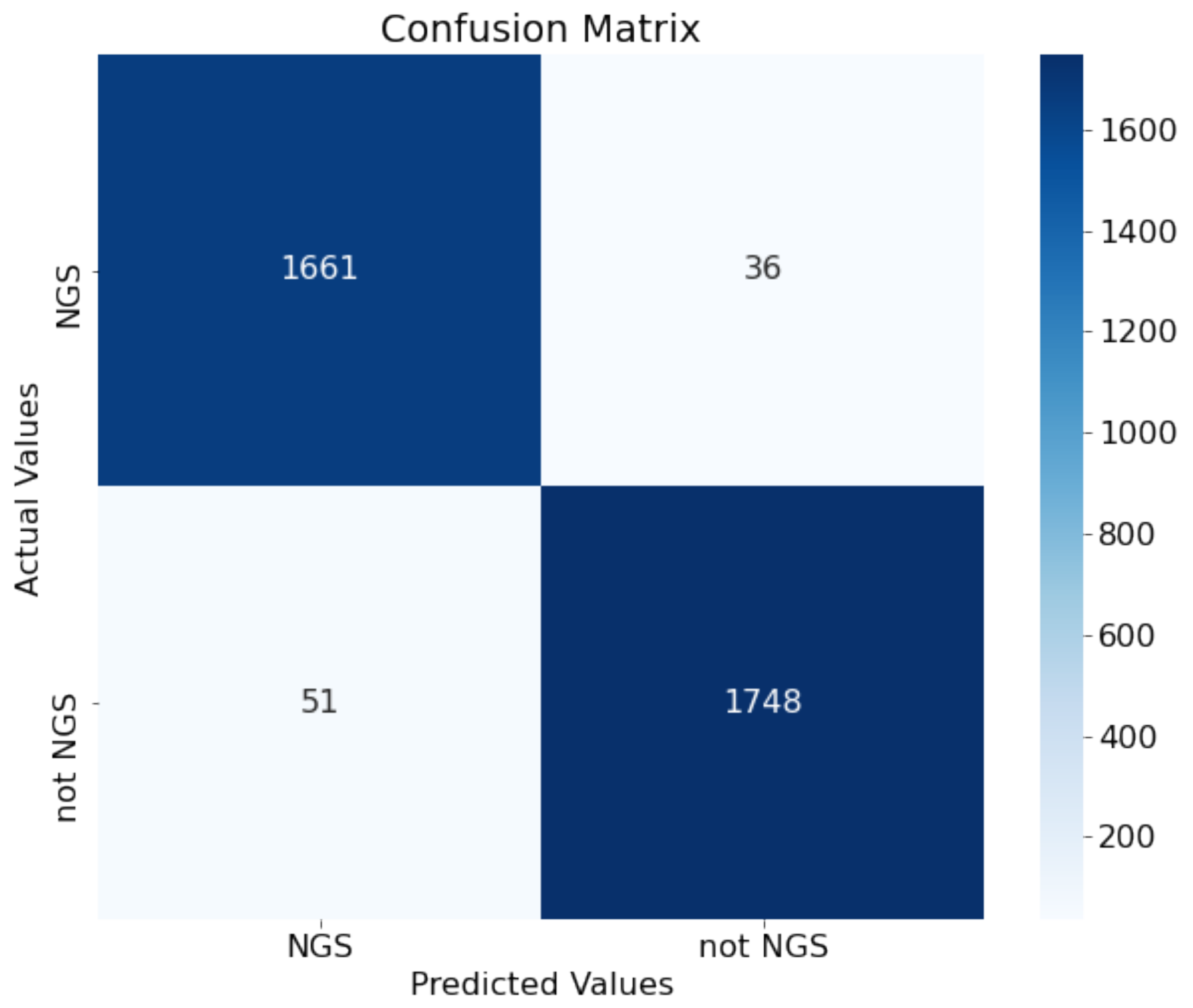


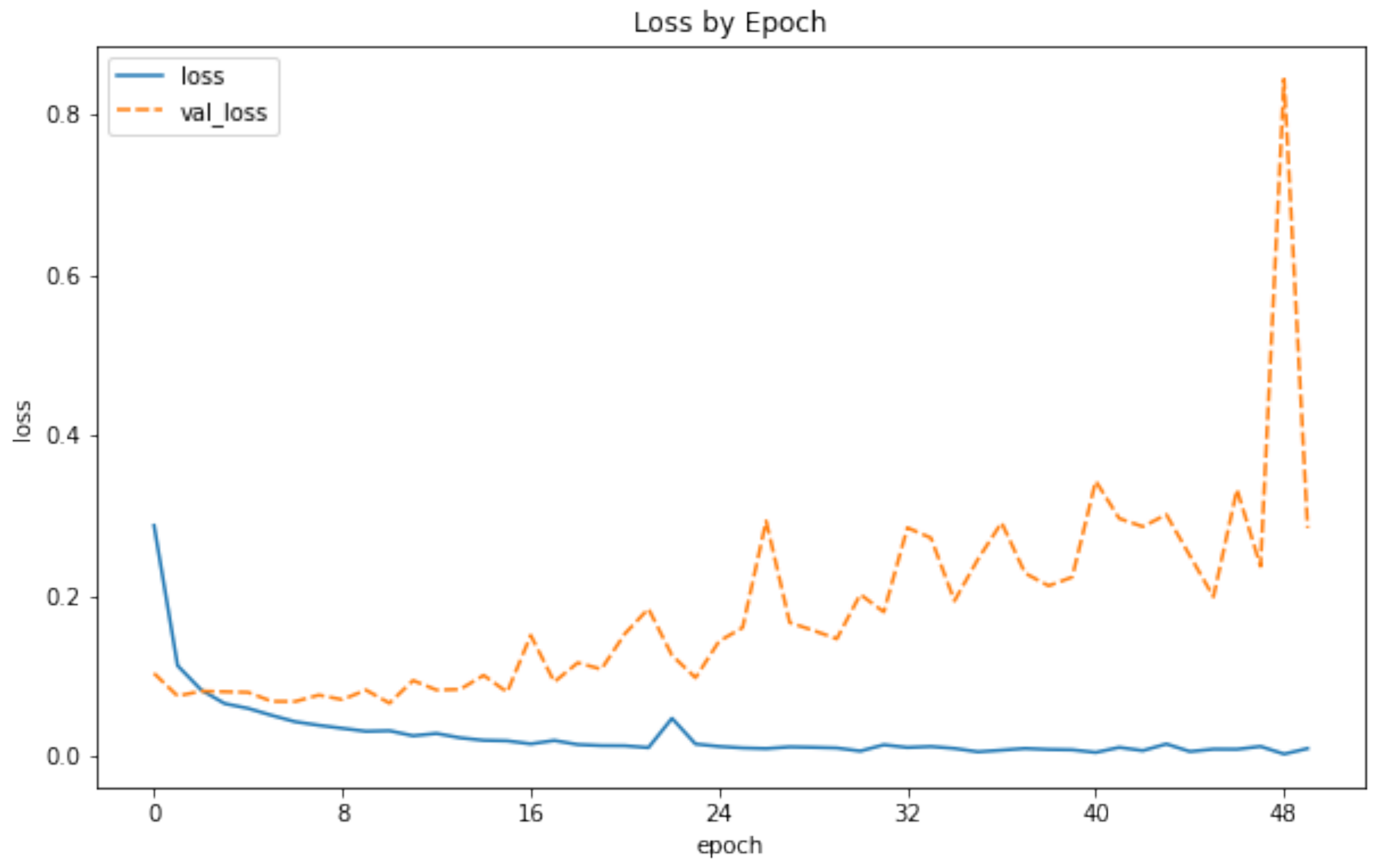


- Multi-classification problem.
- Weighted classes by class count to counter class imbalance.

CNN Model 2A

Table 7: Classification Report for Model 2a				
morphology	**precision**	**recall**	**f1-score**	**support**
NGC	0.97	0.98	0.97	1697
not NGC	0.98	0.97	0.98	1799

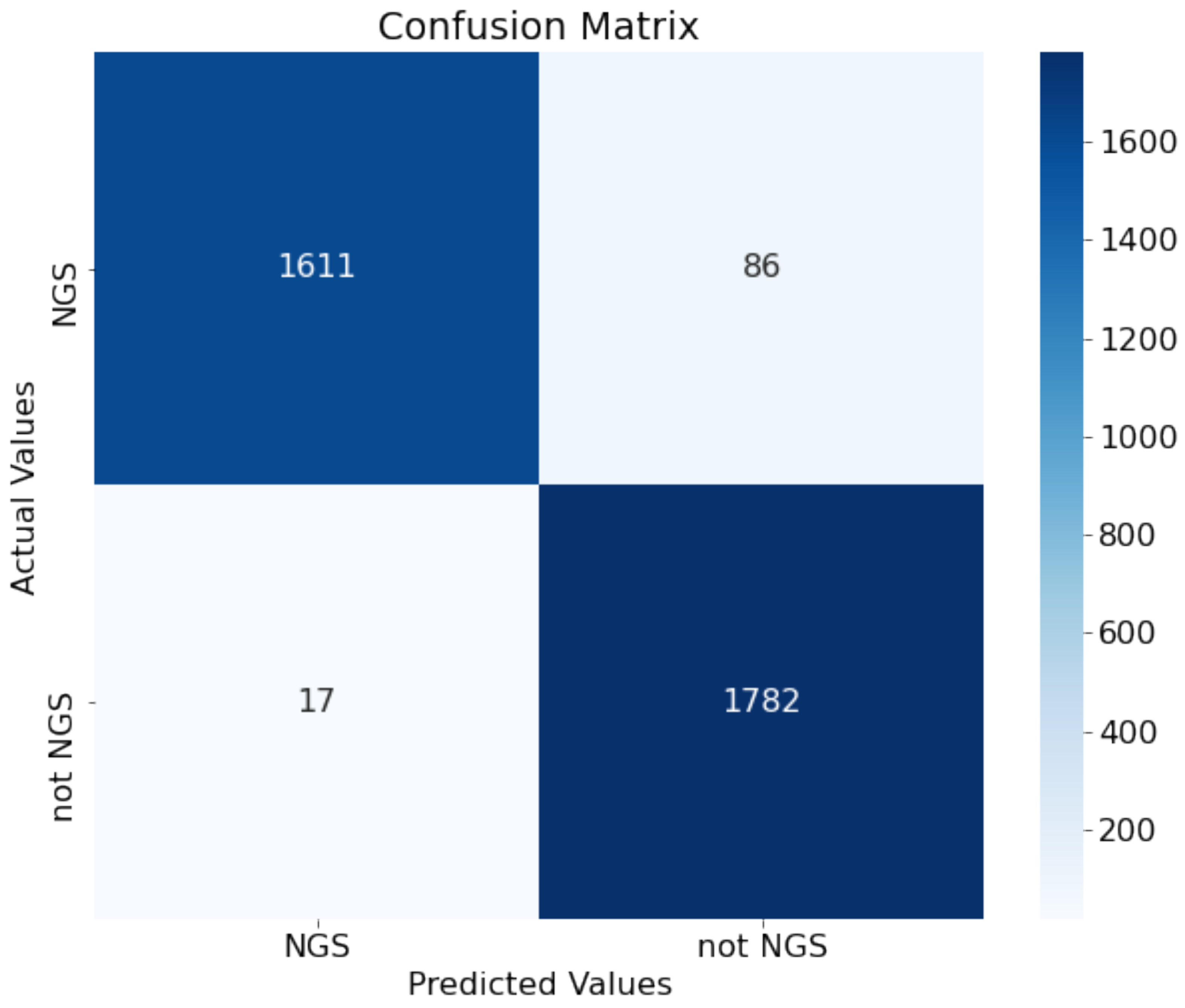


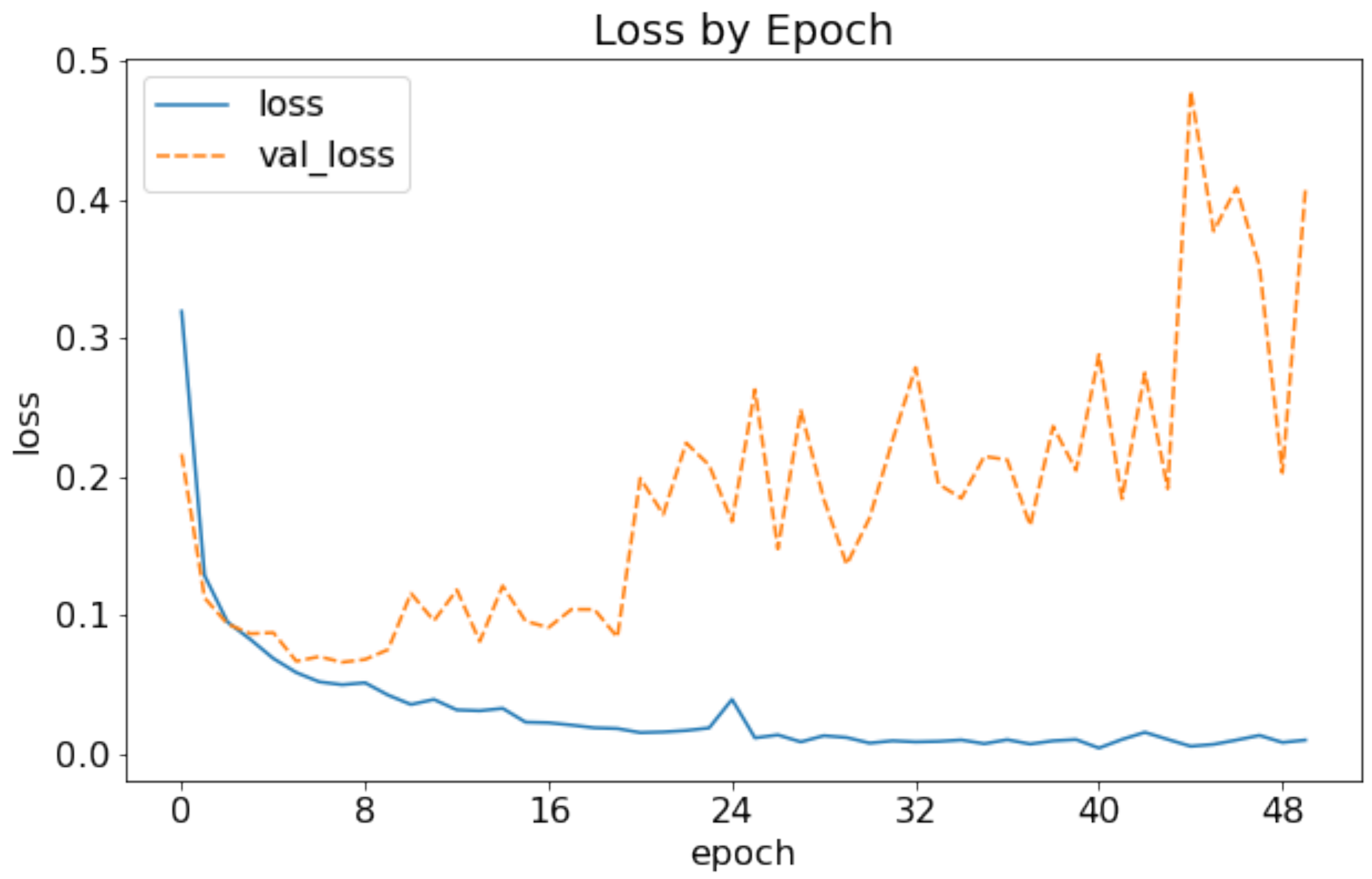


- Binary classification problem.
  - Top four class split into two classes, combining last three classes to the single class: not NGC.

CNN Model 2B

Table 8: Classification Report for Model 2b				
morphology	**precision**	**recall**	**f1-score**	**support**
NGC	0.99	0.95	0.97	1697
not NGC	0.95	0.99	0.97	1799





- Binary classification problem using weighted classes.

Conclusions

- Learning curves indicated CNNs suffered from:
  - Under-fitting.
  - Unrepresentative training set.
- However, binary classifier showed high performance.

Table 9: CNN Model Performance Summary

model	F1 Score	
	macro average	weighted average
model 1a	0.30	0.86
model 1b	0.24	0.73
model 2a	0.98	0.98
model 2b	0.97	0.97

Future Work

- Explore class imbalance remedies further by using image augmentation.
- Invest in higher performance resources from cloud computing services to train the full dataset without the need for rescaling images.
- Develop a more complex neural network architecture that will not suffer from under fitting.

Recommendations to Client

- Phase 1:
  - Use binary classifier to evaluate time saving strategies for leukocyte identification.
- Phase 2:
  - Develop CNN architecture to adequately classify all 15 leukocyte types.
  - Deliver model to client for use in pathology.

References

- Data Citation:
  - "Matek, C., Schwarz, S., Marr, C., & Spiekermann, K. (2019). A Single-cell Morphological Dataset of Leukocytes from AML Patients and Non-malignant Controls [Data set]. The Cancer Imaging Archive. https://doi.org/10.7937/tcia.2019.36f5o9ld".

- A Single-cell Morphological Dataset of Leukocytes from AML Patients and Non-malignant Controls (AML-Cytomorphology_LMU).

- Human-level recognition of blast cells in acute myeloid leukemia with convolutional neural networks

- How to use Learning Curves to Diagnose Machine Learning Model Performance