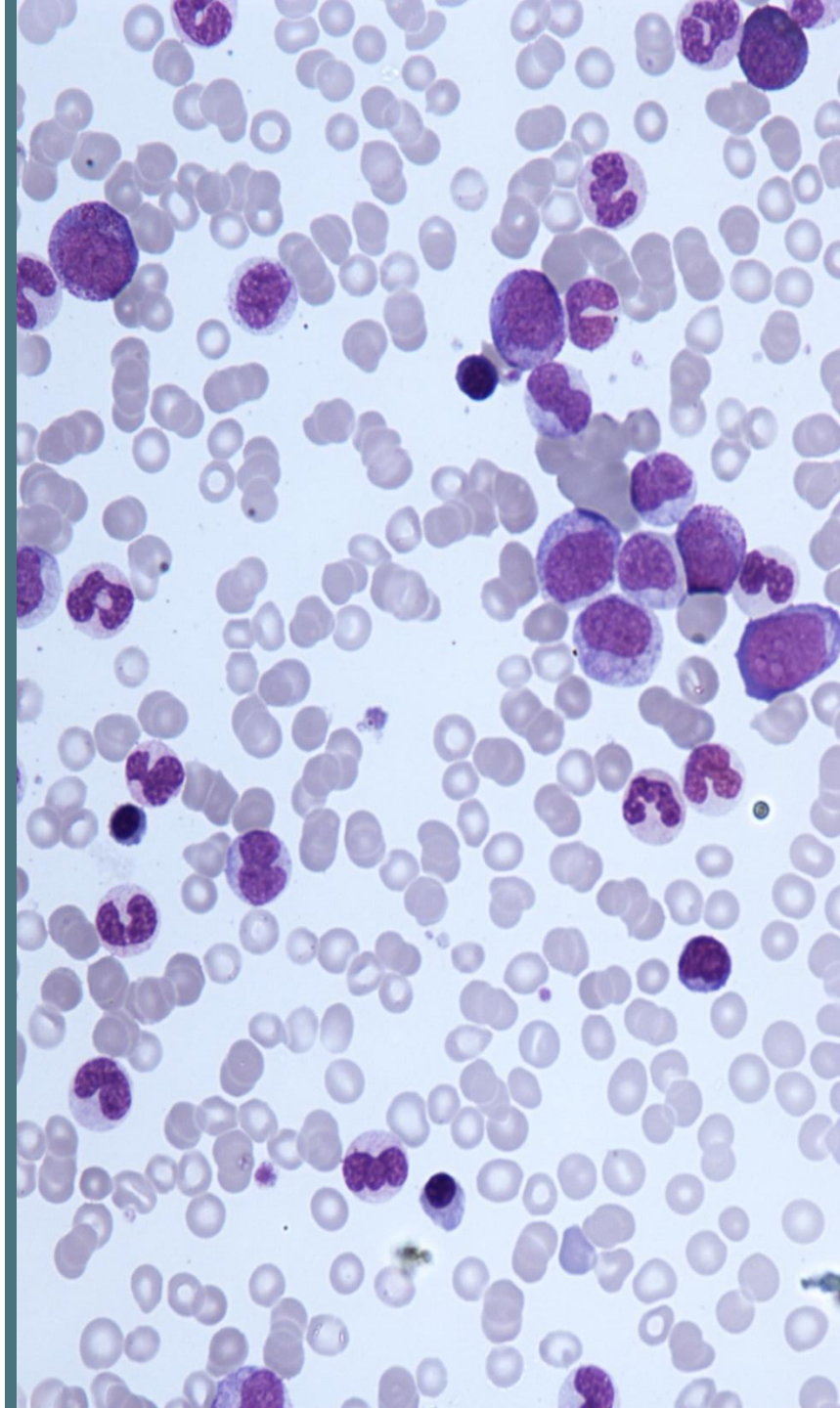


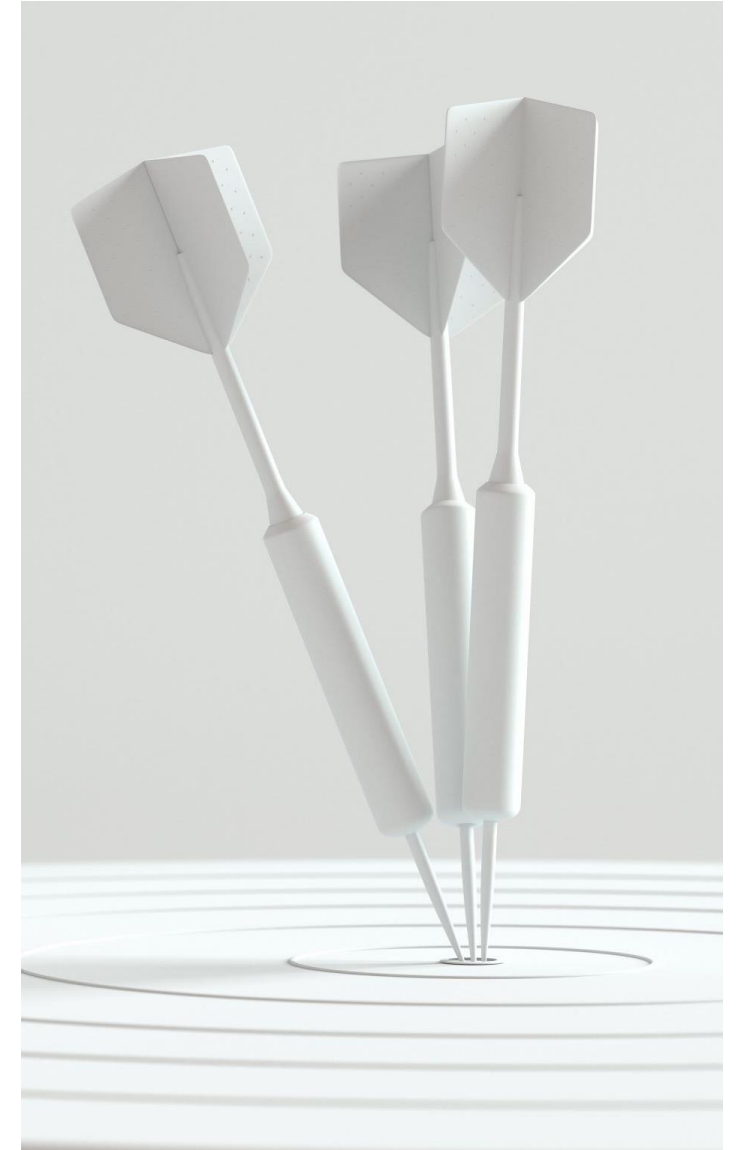
# CLASSIFICAZIONE DI GLOBULI BIANCHI CON TECNICHE DI MACHINE LEARNING

*Tommaso Mingrone*



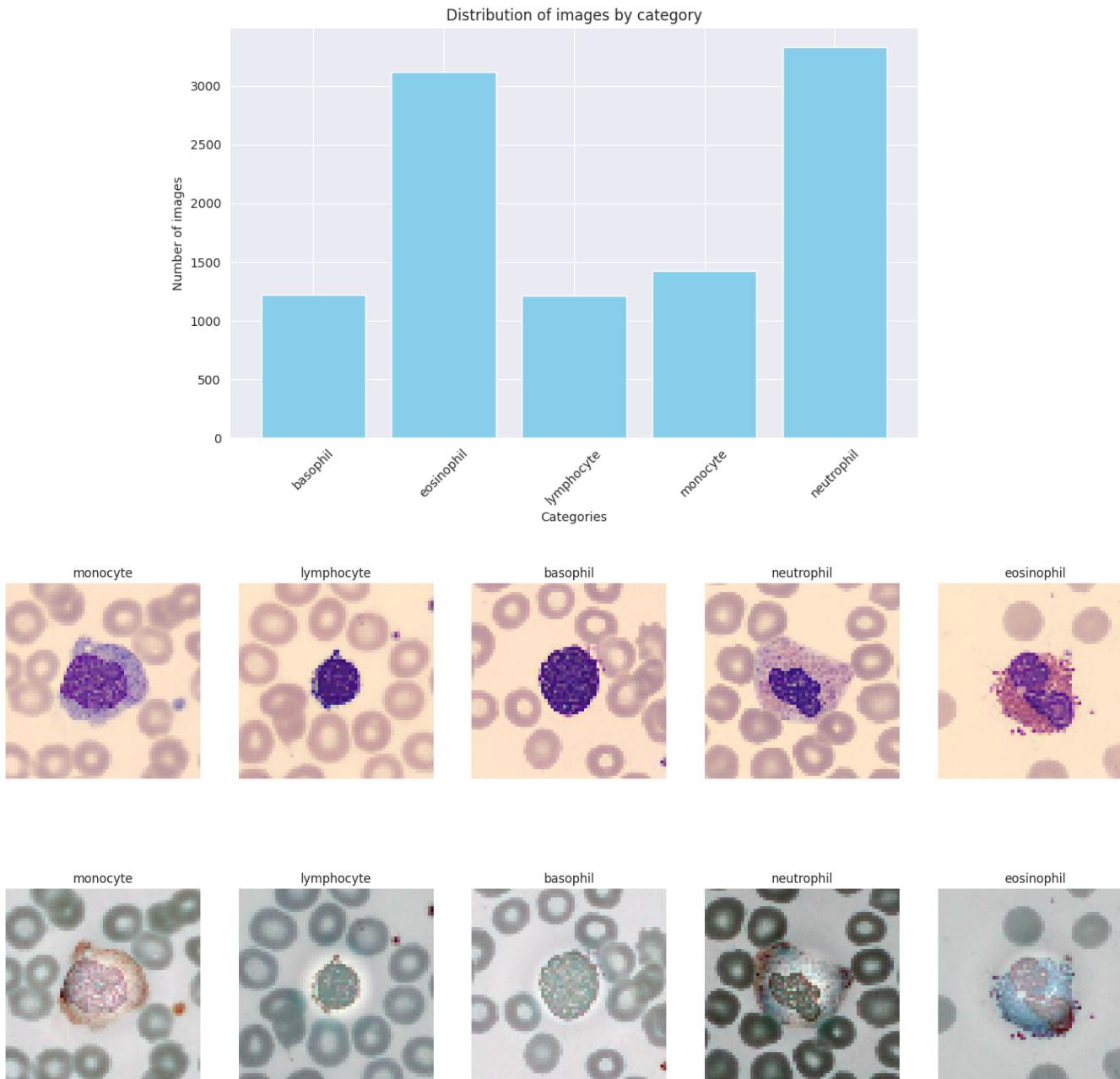
# OBIETTIVI DELLO STUDIO

- Principale:
  - Classificare immagini di globuli bianchi di individui sani
- Complementari:
  - Studiare la struttura dei dati
  - Trovare il metodo di classificazione più efficiente

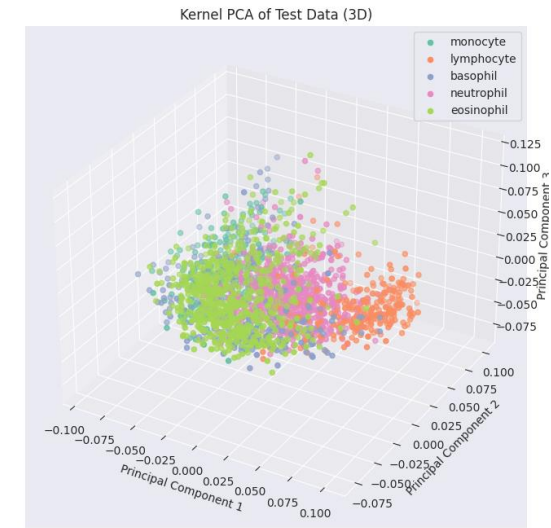


# DATA PROCESSING

- Rimosse 3 classi : ig, platelet, erythroblast
- Riduzione della risoluzione delle immagini da 360x363 a 64x64
- Normalizzato per ogni pixel in scala da 0 a 1 per ciascun canale colore (RGB)
- Standardizzato



# UNSUPERVISED LEARNING

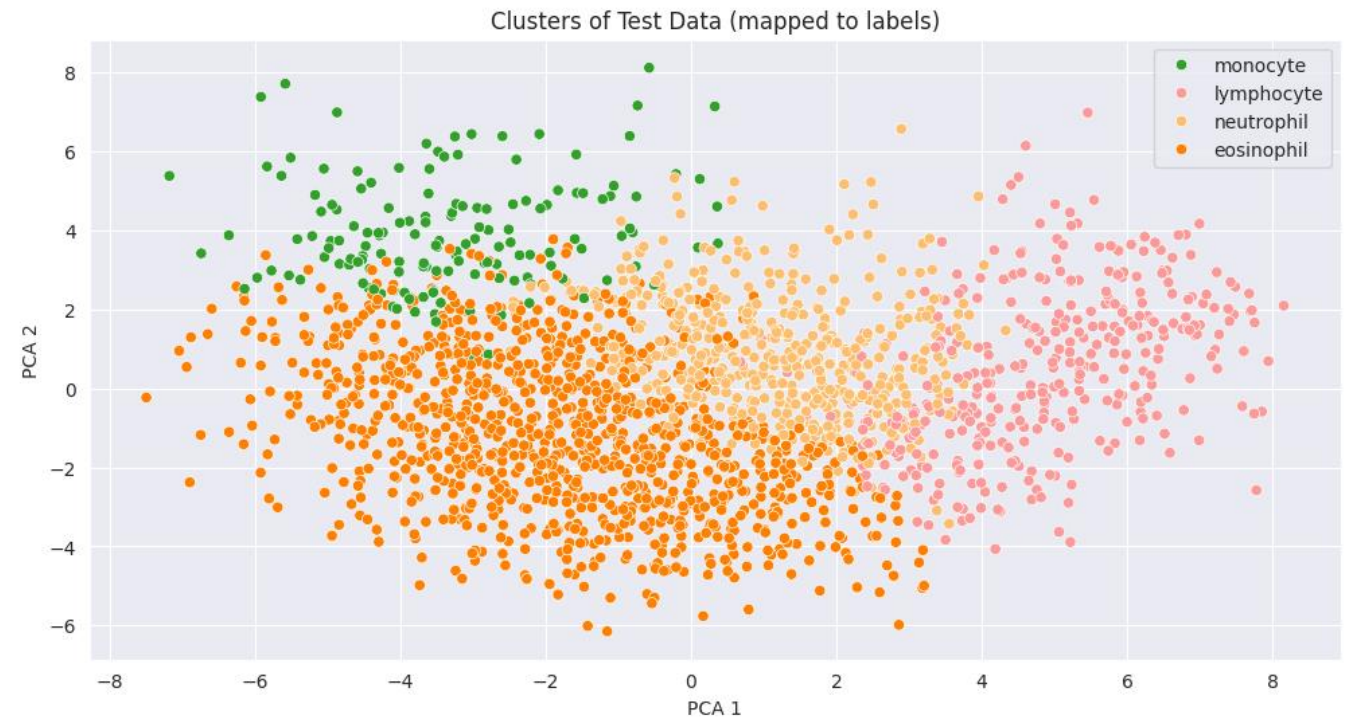


Confusion Matrix for KMeans Model

	monocyte	lymphocyte	basophil	neutrophil	eosinophil
monocyte	69	1	0	8	206
lymphocyte	3	210	0	23	7
basophil	15	36	0	15	178
neutrophil	18	100	0	399	158
eosinophil	48	17	0	64	494

True label

Predicted label

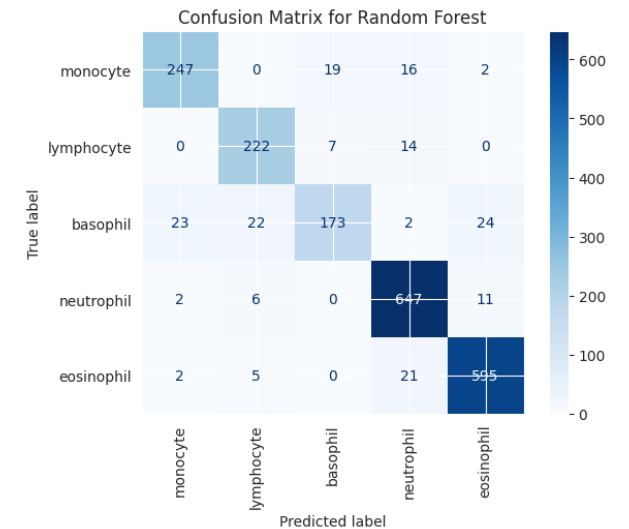
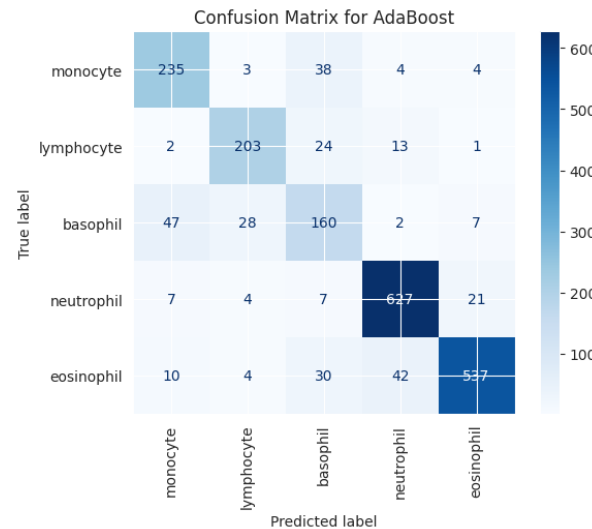
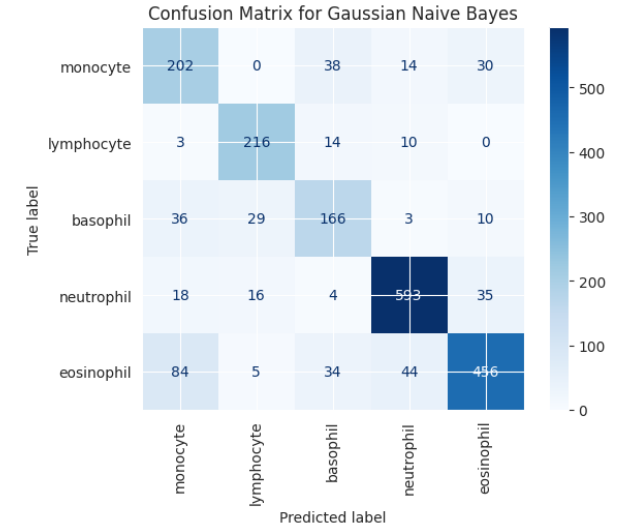
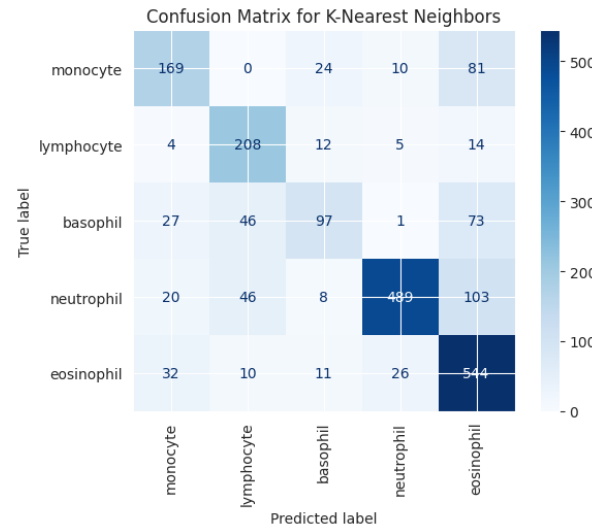




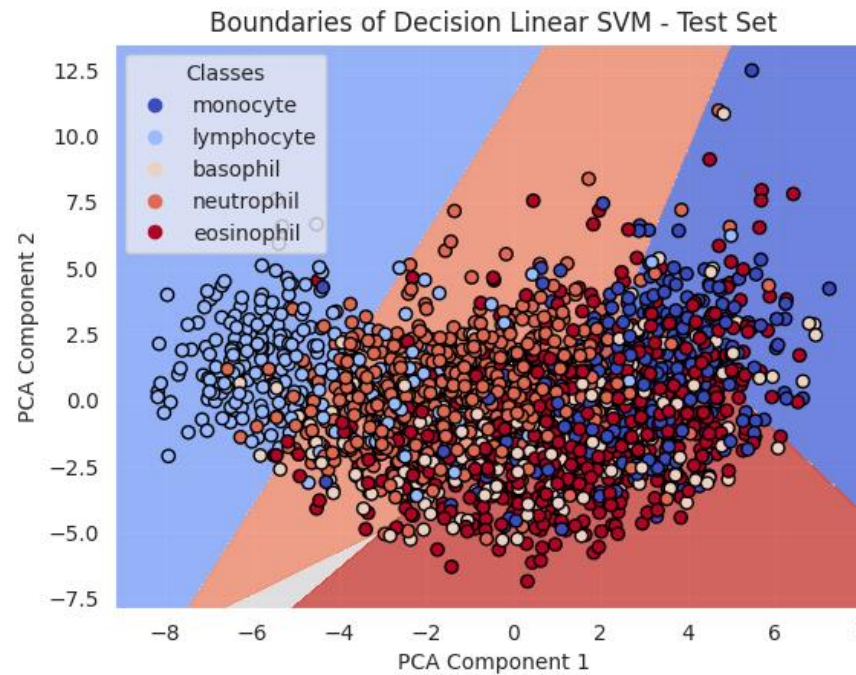
# SUPERVISED LEARNING

Test accuracy:

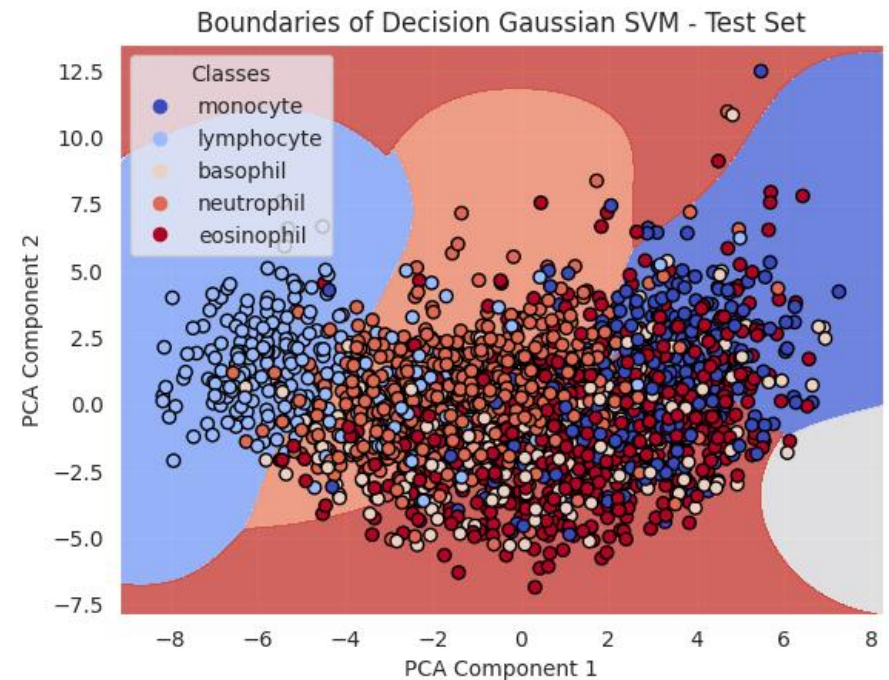
- K-Nearest Neighbors (73%)
- Naïve Bayes (79%)
- Adaboost (86%)
- Random Forest (91%)



# SUPPORT VECTOR MACHINE MULTICLASS

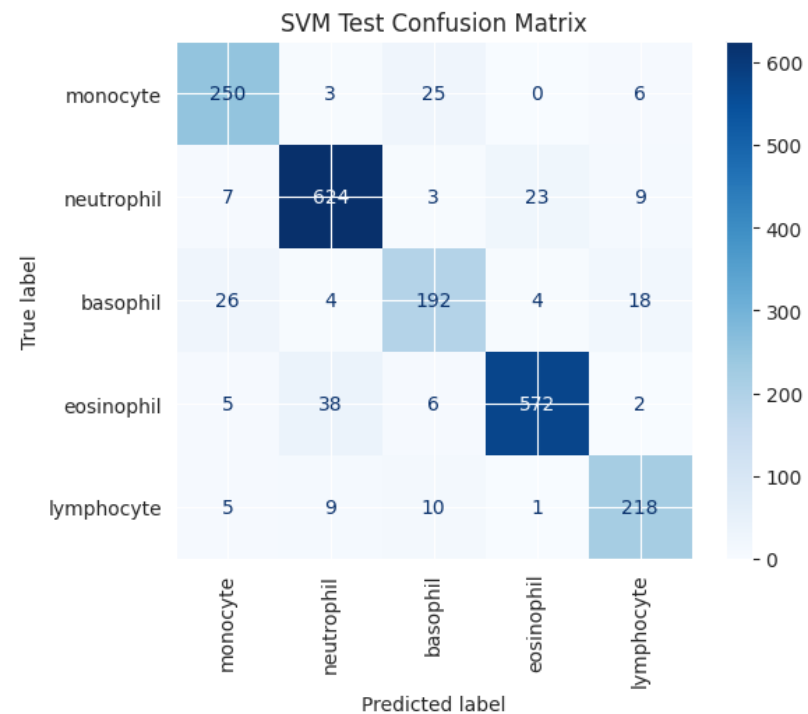


SVMM Linear Kernel  
Test Accuracy (57%)

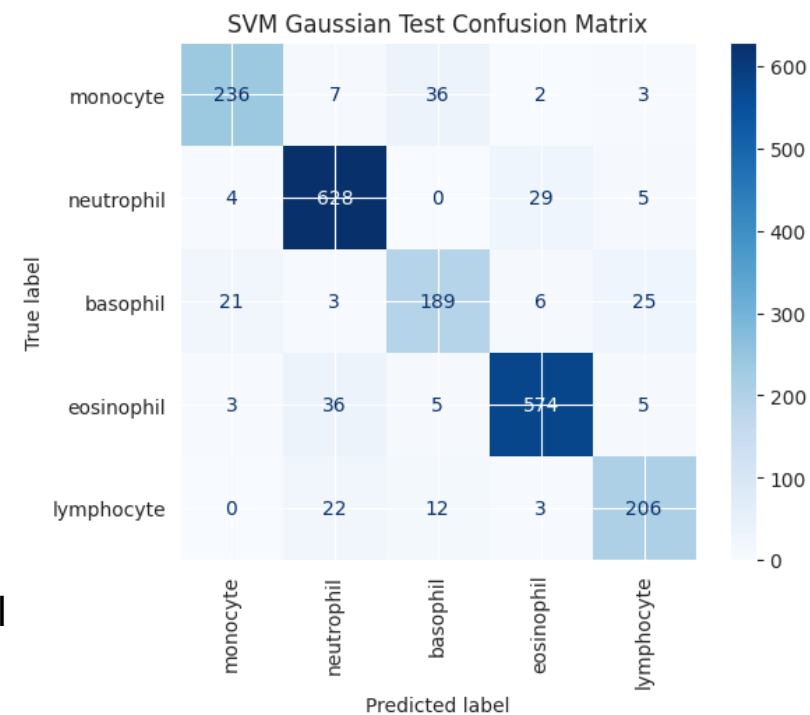


SVMM Gaussian Kernel  
Test Accuracy (58%)

# SVMM ALTA DIMENSIONALITÀ



SVMM Linear Kernel  
Test Accuracy (90%)

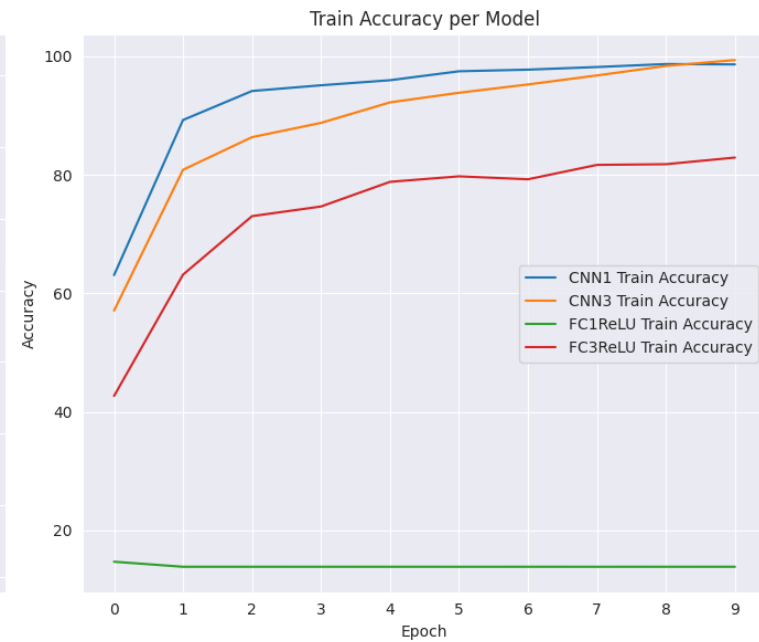
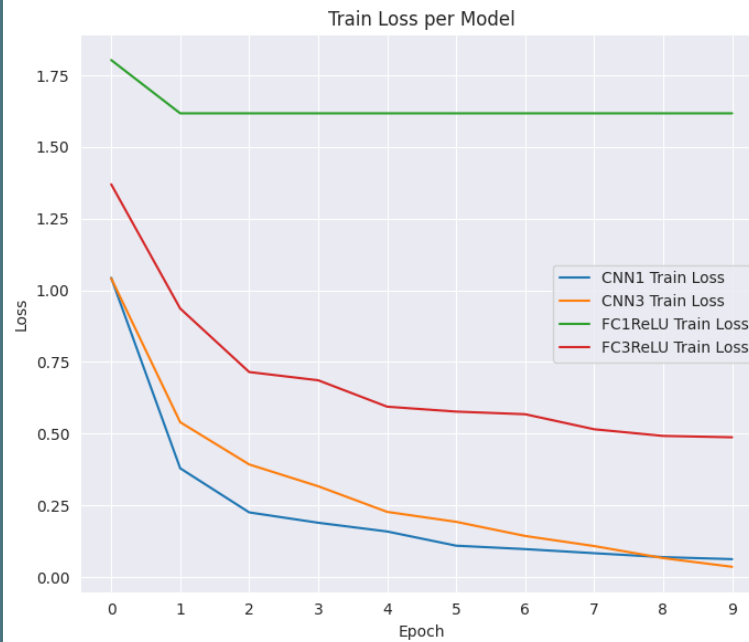


SVMM Gaussian Kernel  
Test Accuracy (89%)

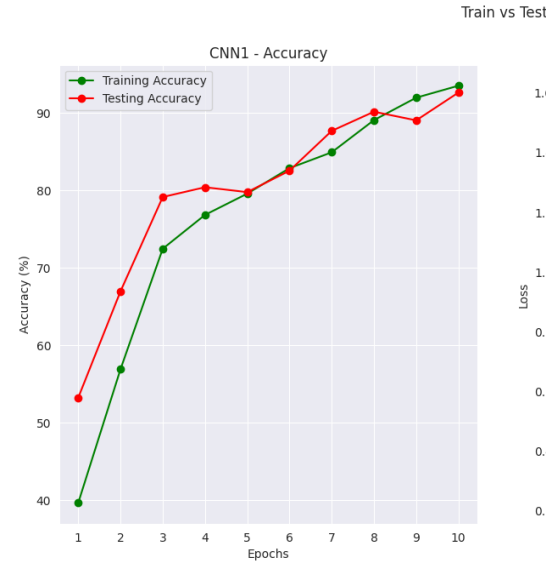
# DEEP LEARNING

Modelli:

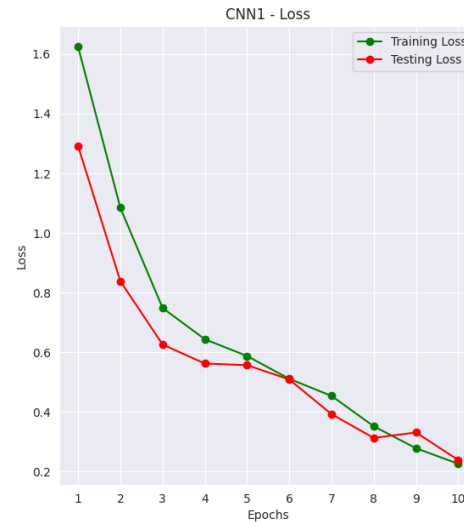
- FC1ReLU
- FC3ReLU
- CNN1
- CNN3



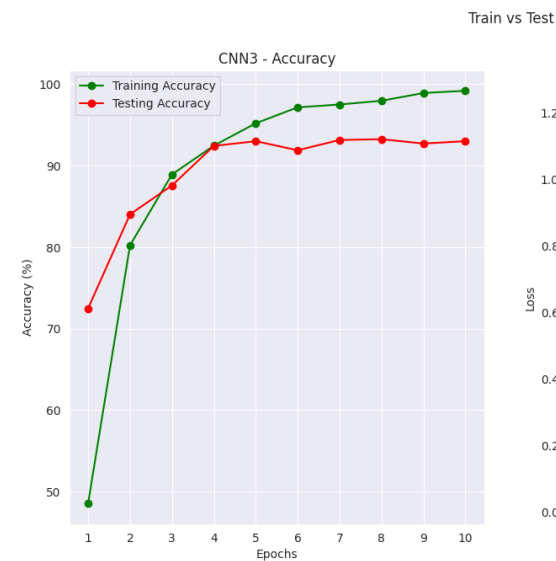




Train Accuracy: 98.99%  
Test Accuracy: 95.67%



Train Loss: 0.0561  
Test Loss: 0.1465



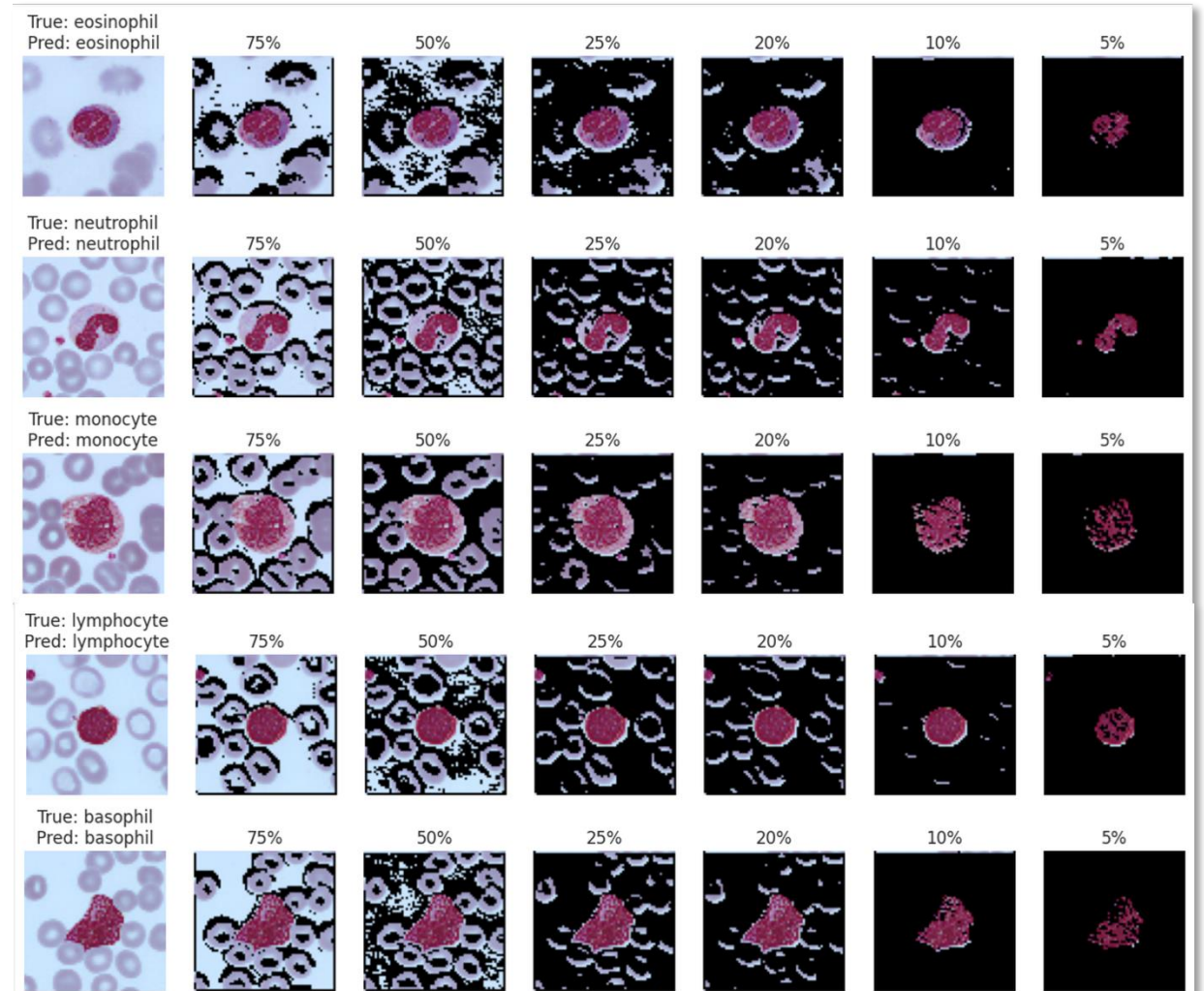
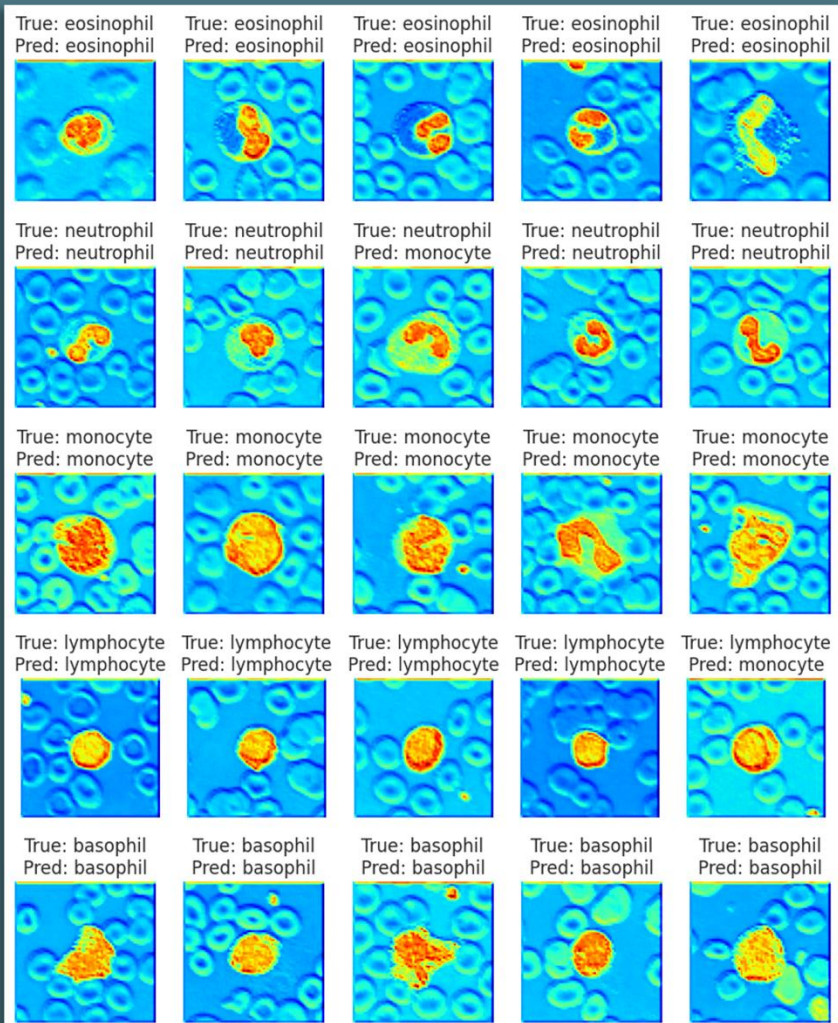
Train Accuracy: 99.35%  
Test Accuracy: 93.82%



Train Loss: 0.0320  
Test Loss: 0.2214

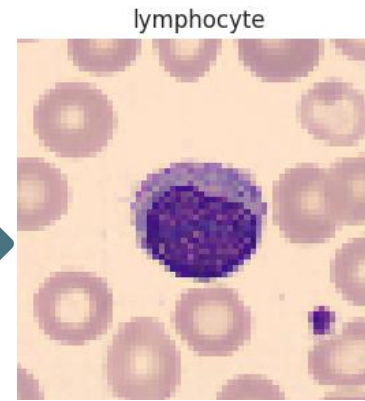
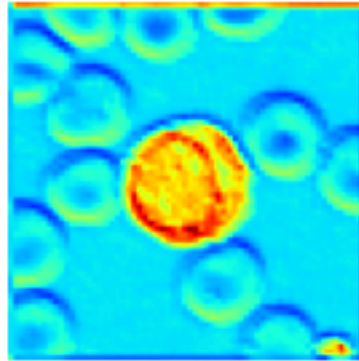
# TRAIN VS TEST DEI MODELLI MIGLIORI

# ATTENTION MAP



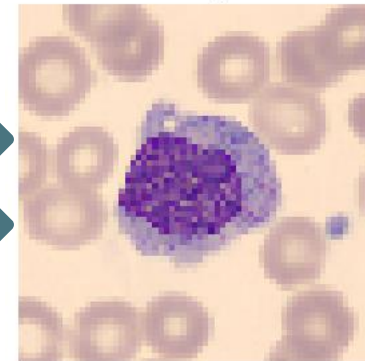
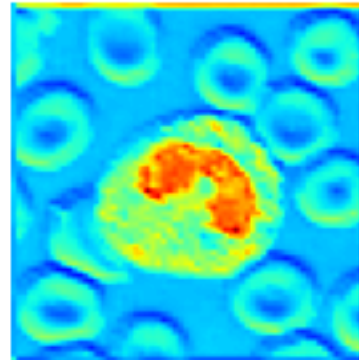
# CASI DI INTERESSE

True: lymphocyte  
Pred: monocyte

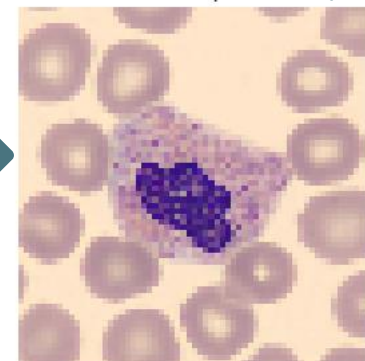


lymphocyte

True: neutrophil  
Pred: monocyte



monocyte

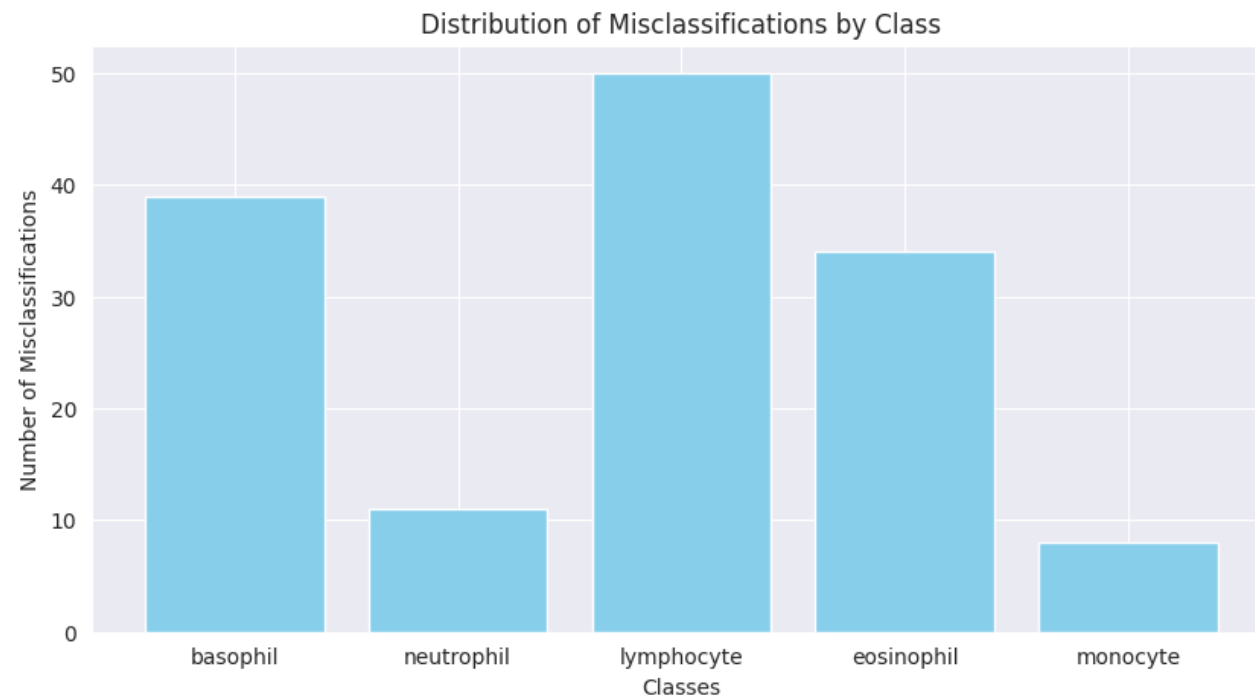


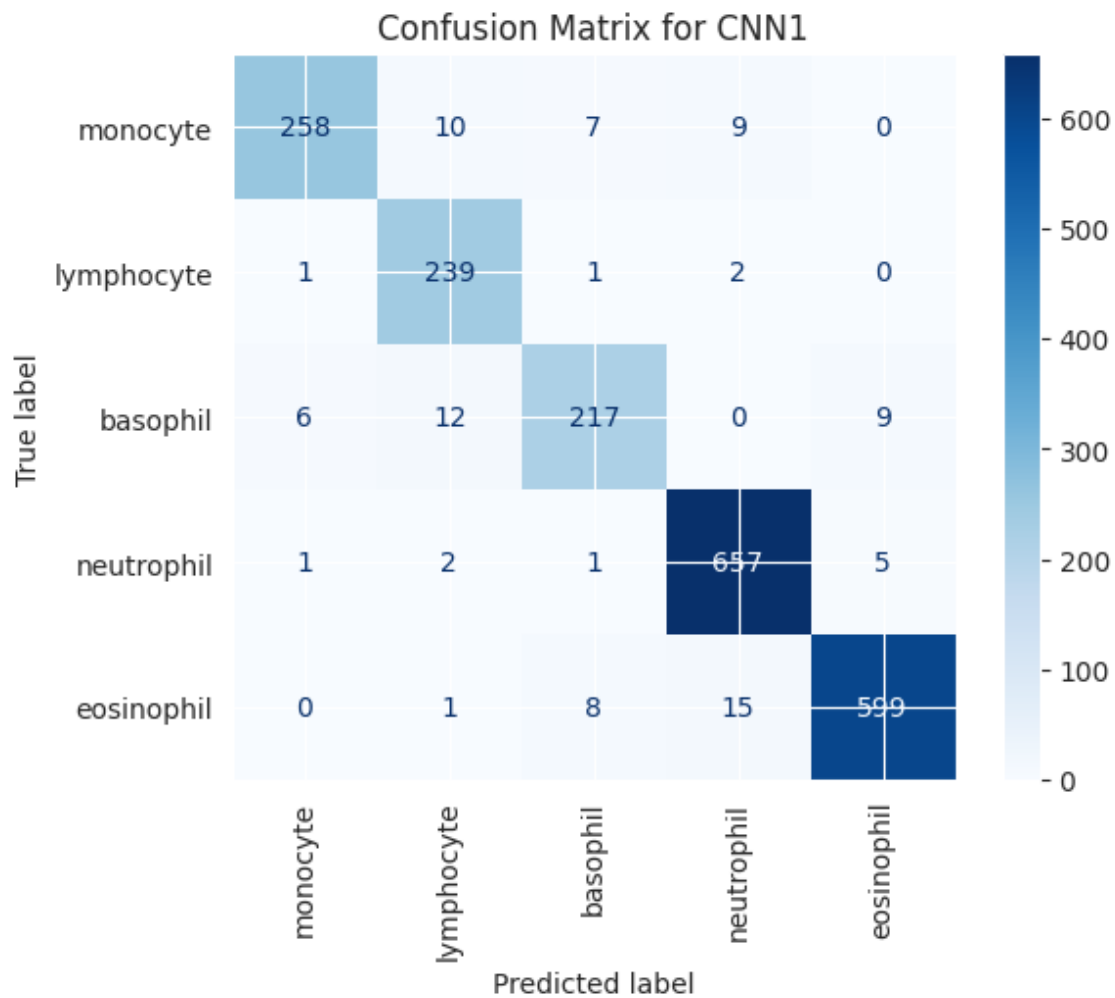
neutrophil

A volte simili per rapporto  
nucleo/citoplasma e  
nucleo monolobato

Possibile confusione  
dovuta ad artefatti del  
vetrino e sezioni non  
ottimali per l'analisi

SI POSSONO  
GENERALIZZARE  
QUESTE  
MOTIVAZIONI?





# CONCLUSIONE

- La geometria dei dati offre informazioni che sono efficaci per comprendere il problema
- Il modello più efficiente è CNN1. Tuttavia, è importante menzionare anche i risultati delle tecniche di Supervised come Random Forest e SVM ad alta dimensionalità