

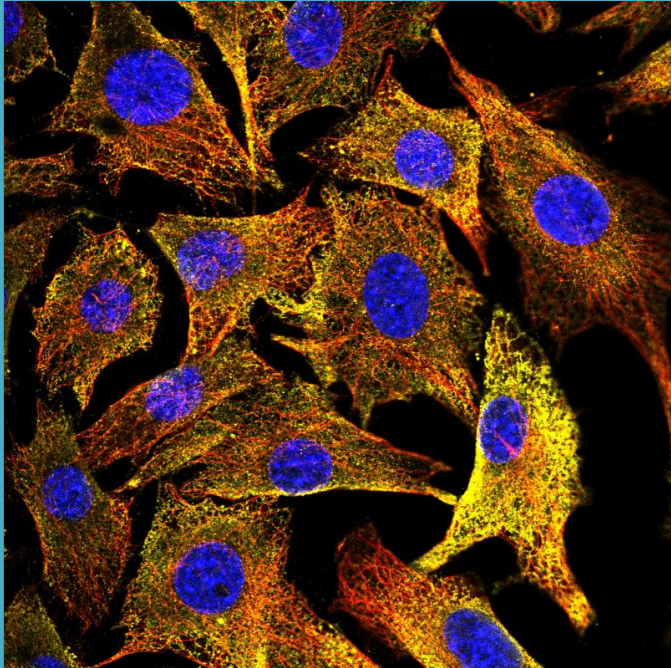


Human Protein Image Classification

Luke LaJoie
June 16, 2020

Introduction

THE HUMAN PROTEIN ATLAS 



Human Protein Classification

- Classify different proteins in human cells
- Proteins do most of the work in the cell

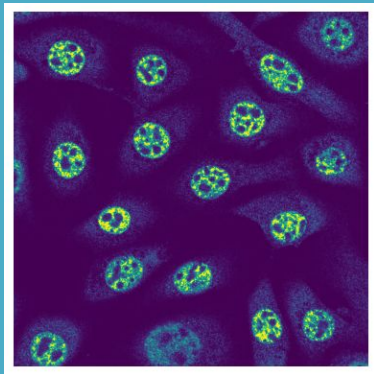
Introduction



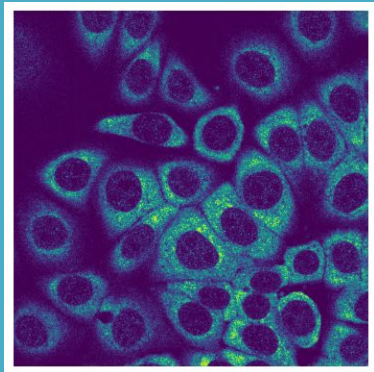
Human Protein Classification

- Classify different proteins in human cells
- Proteins do most of the work in the cell
- High throughput confocal microscope

Introduction



**Nuclear
speckles**



Cytosol

Goal

- Classify two different proteins
 - Nuclear speckles
 - Cytosol

Methodology



Gather Data

Data imagery provided by Human Protein Atlas and available via Kaggle

kaggle



Process Data

Images needed to be transformed in different ways to work with different models



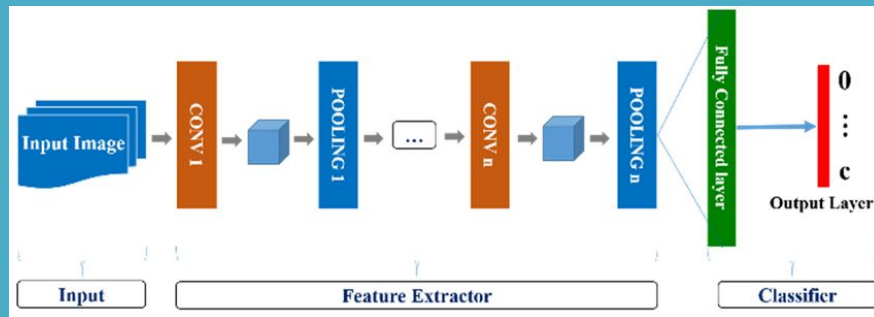
Build and Run Models

Random Forest and Convolutional Neural Network models were run and compared



Results

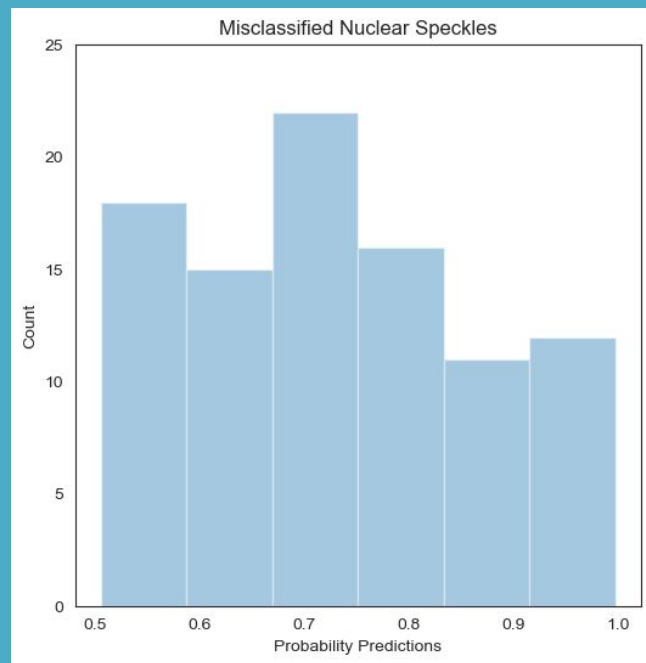
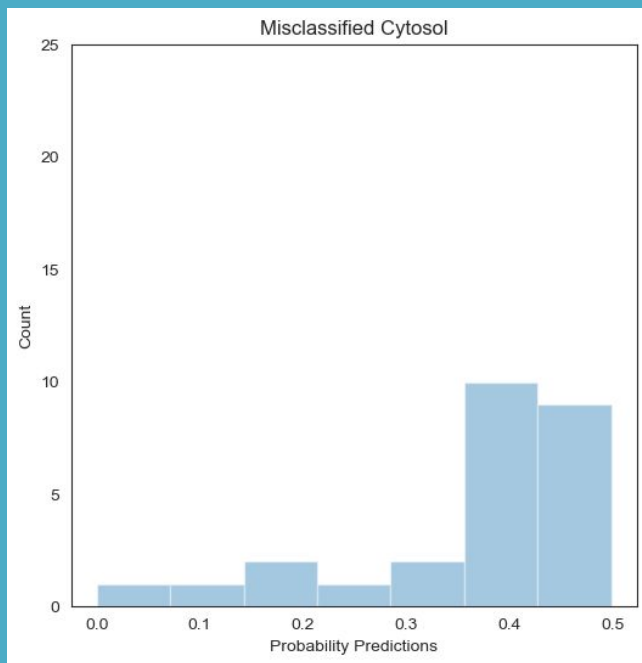
- CNN
 - 3 block VGG
 - Output: 0 to 1
 - NS: 0 to 0.5
 - C: 0.5 to 1



Accuracy Score: 0.930

Results

- Cytosol less likely to be classified incorrectly



Results

(NS)

0

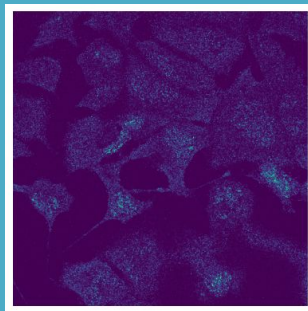
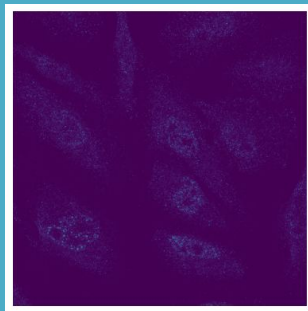
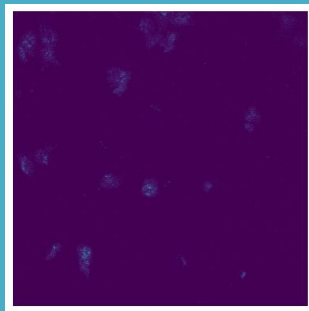
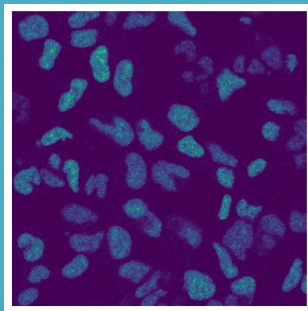
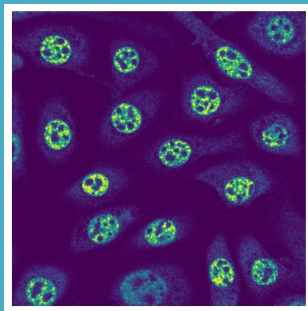
0.25

0.5

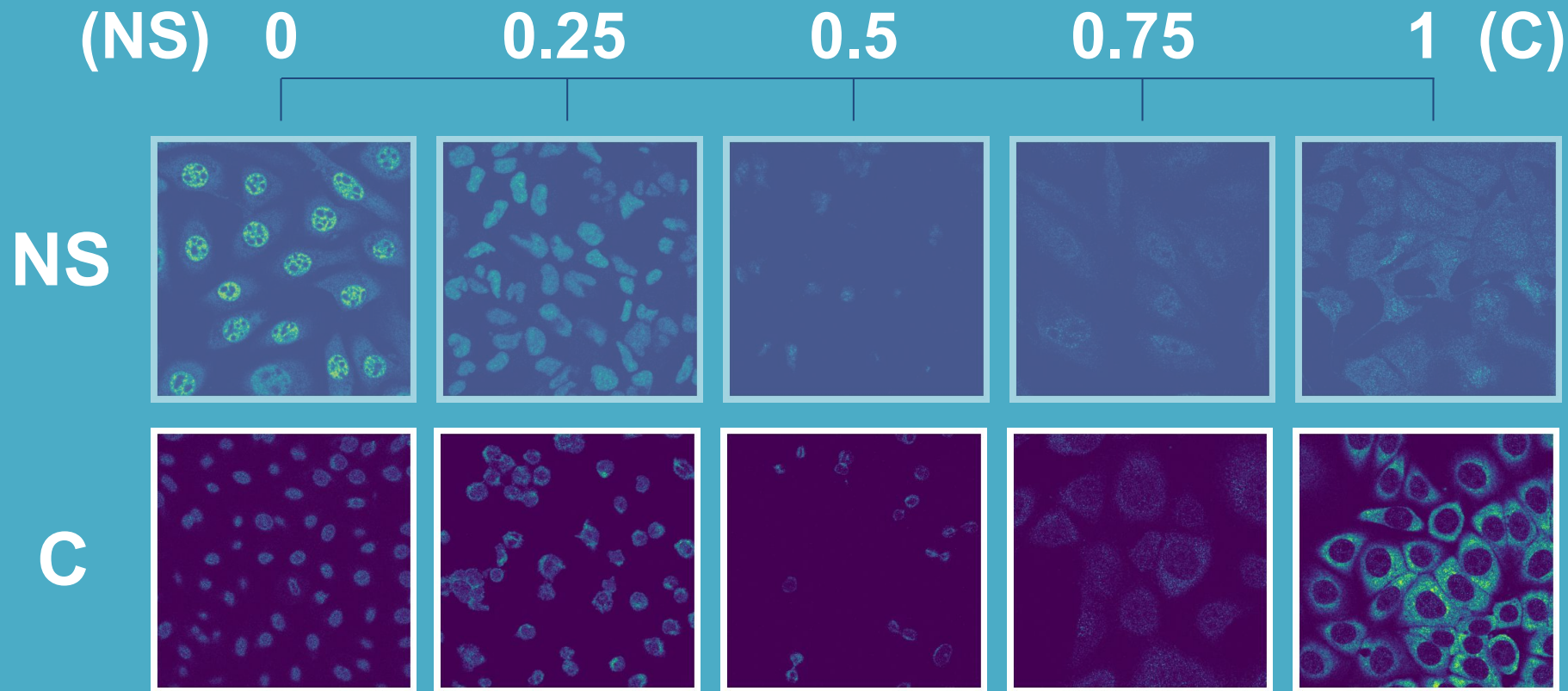
0.75

1 (C)

NS



Results





Conclusion

- CNN was a better classifier than RF for this image data
- Accuracy of 93%
- Potentially can save time and resources by automating classification



Thank you!

Luke LaJoie

[linkedin.com/in/luke-lajoie](https://www.linkedin.com/in/luke-lajoie)

github.com/lukeml

luke.m.lajoie@gmail.com

Credits

This presentation template created by [GoogleSlidesppt.com](https://www.googleSlidesppt.com)

All shapes and Icons in this template were created by [GoogleSlidesppt.com](https://www.googleSlidesppt.com)

All backgrounds in this template were created by [GoogleSlidesppt.com](https://www.googleSlidesppt.com)

Illustrated images in this template were created by [GoogleSlidesppt.com](https://www.googleSlidesppt.com)

Inserted pictures in this template were created by [pixabay.com](https://www.pixabay.com)

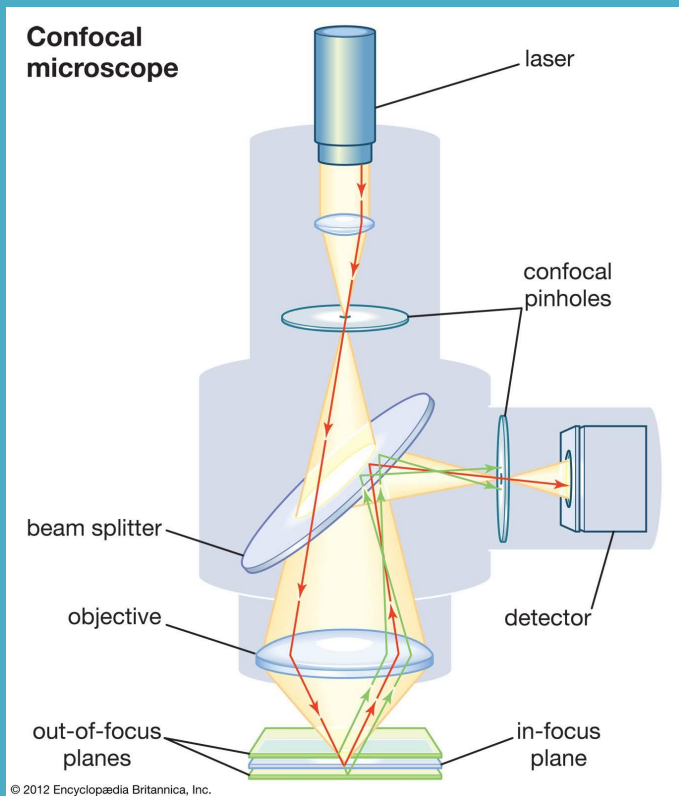
Thank you for using GoogleSlidesppt.com

Future Work

- Work with multiclass / multilabel data
- Try other pretrained models for transfer learning

Appendix

Introduction



Human Protein Atlas Kaggle competition

- High throughput confocal microscopy
- Classify subcellular proteins in human cells

Introduction

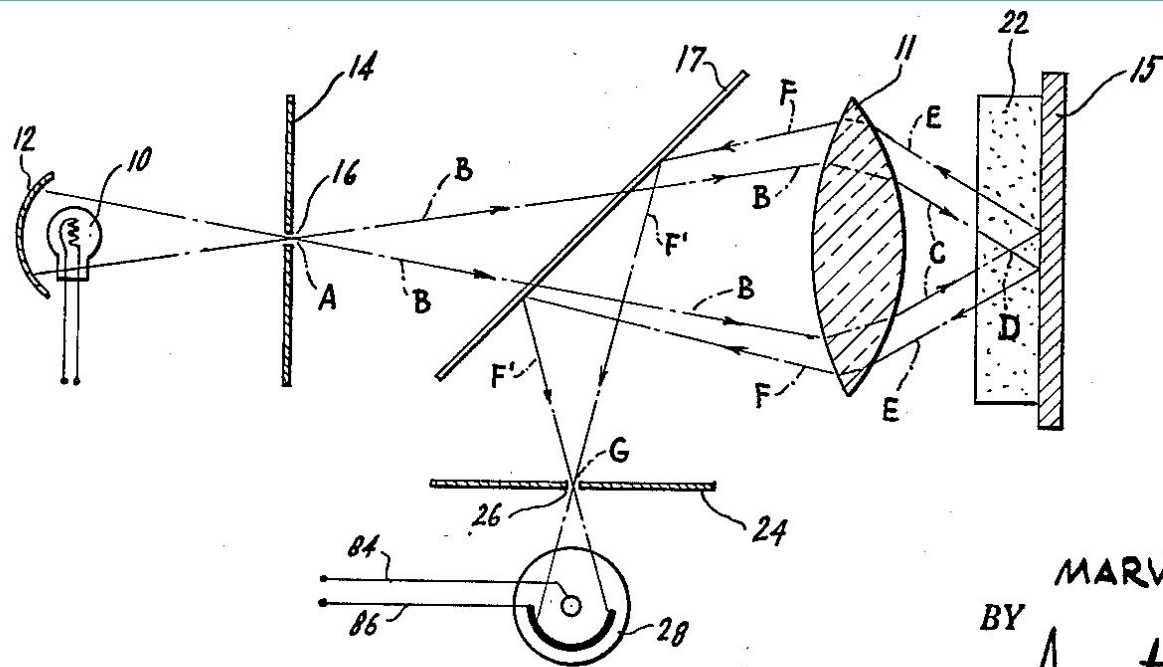


FIG. 3.

INVENTOR.

MARVIN MINSKY

BY

Ameter & Levy
ATTORNEYS

Methodology



Get data from Kaggle



Clean and process data



Build and run models



Compare and interpret

Methodology

Insert the title of your subtitle Here



Text Here

You can simply impress your audience and add a unique zing and appeal to your Presentations.



Text Here

You can simply impress your audience and add a unique zing and appeal to your Presentations.



Text Here

You can simply impress your audience and add a unique zing and appeal to your Presentations.



Text Here

You can simply impress your audience and add a unique zing and appeal to your Presentations.

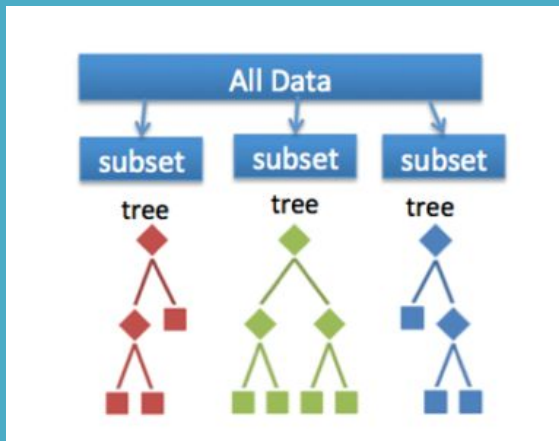


Text Here

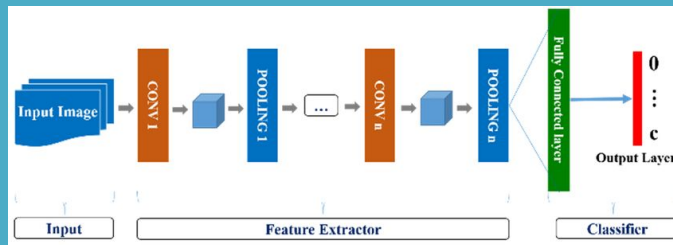
You can simply impress your audience and add a unique zing and appeal to your Presentations.

Methodology

- PCA
- Random Forest



- CNN
 - 3 block VGG



Results

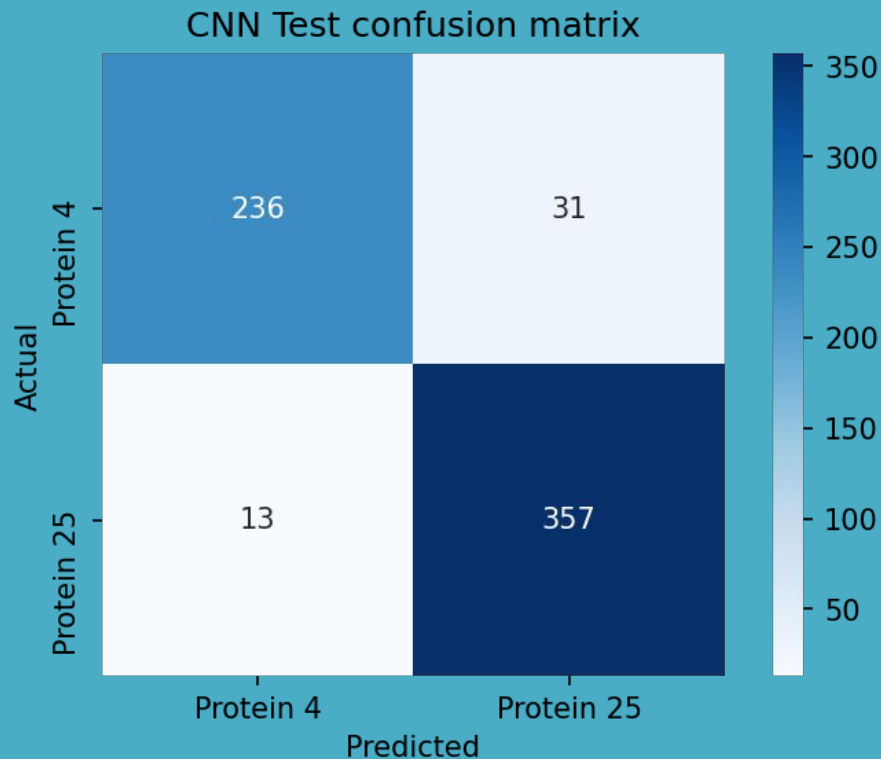
CNN

Accuracy Score: 0.930

Precision Score: 0.920

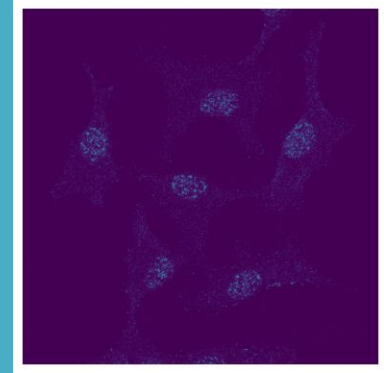
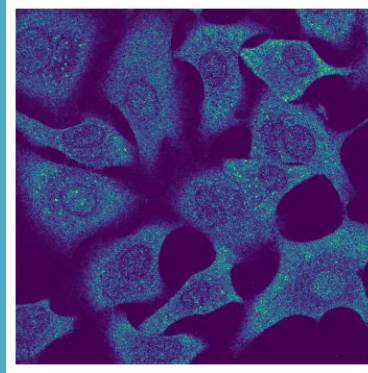
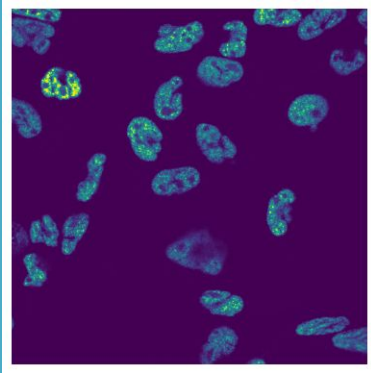
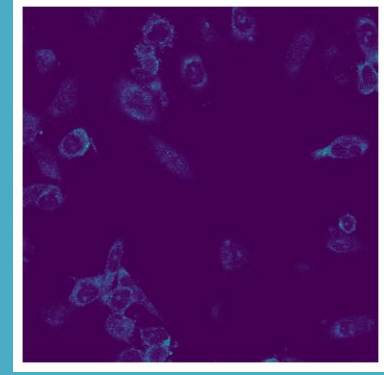
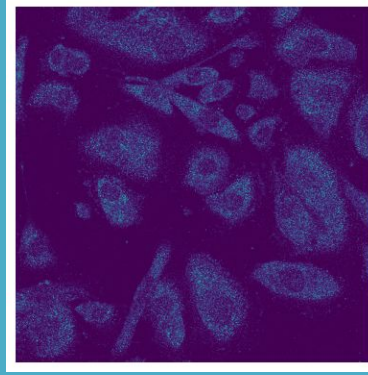
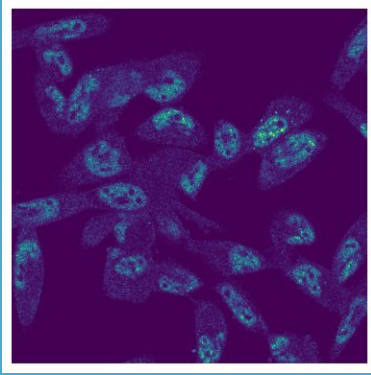
Recall Score: 0.964

F1 Score: 0.941

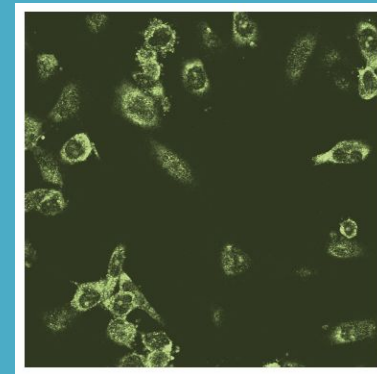
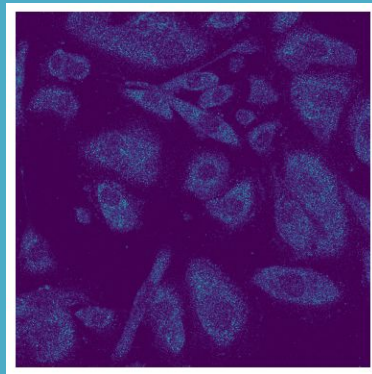
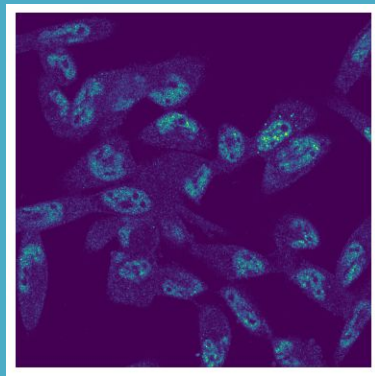


Protein 4 Protein 25

?

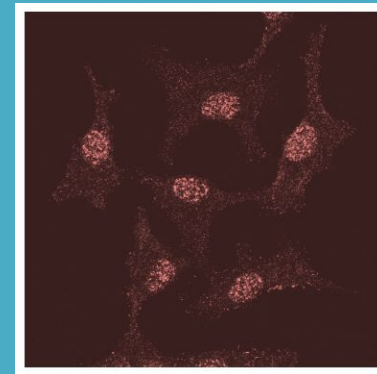
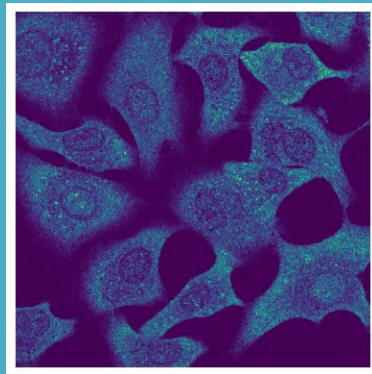
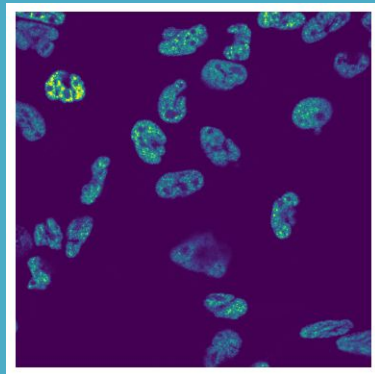


Protein 4 Protein 25



25

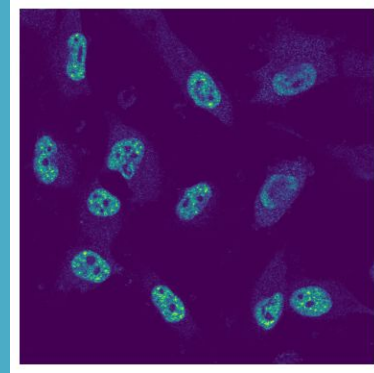
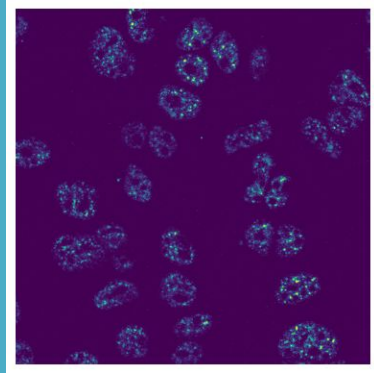
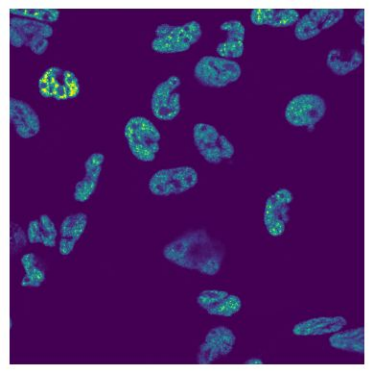
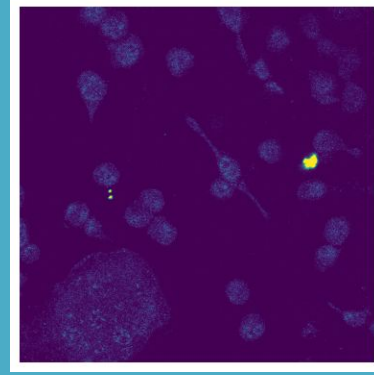
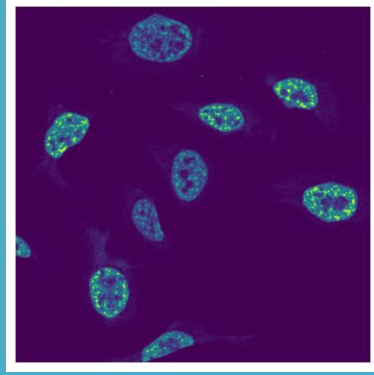
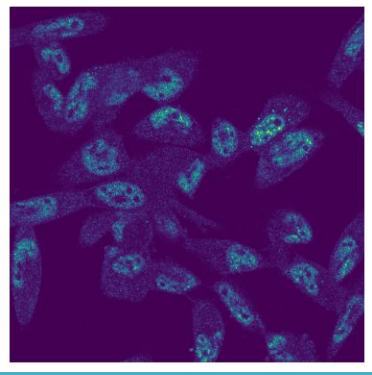
0.528



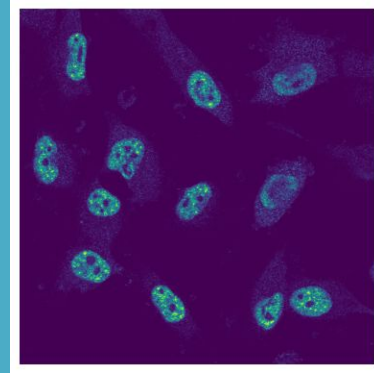
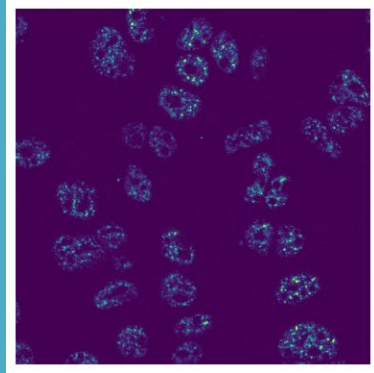
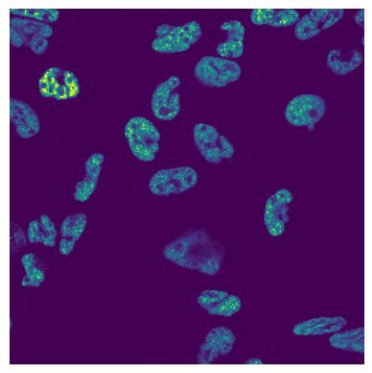
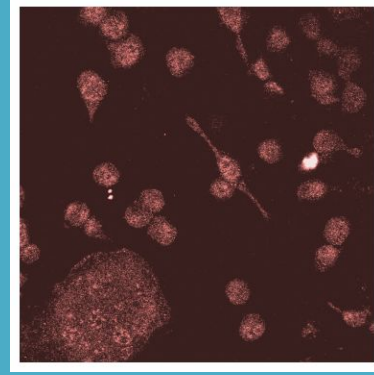
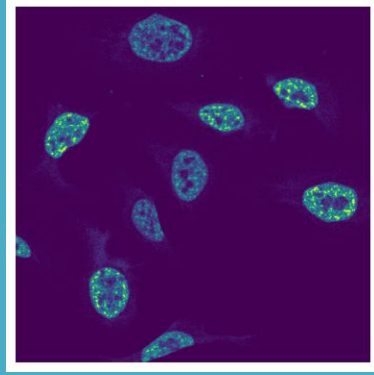
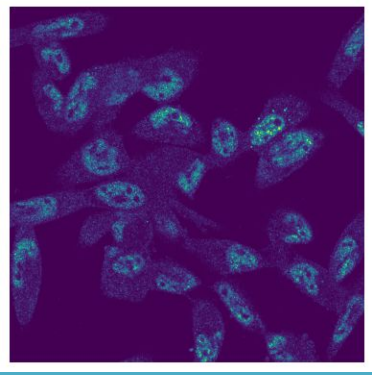
4

0.550

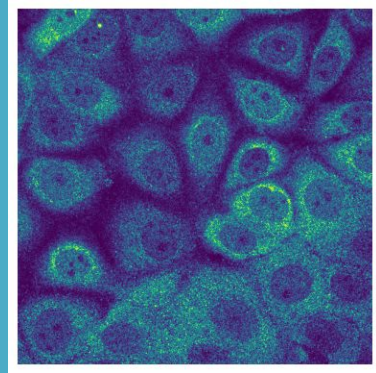
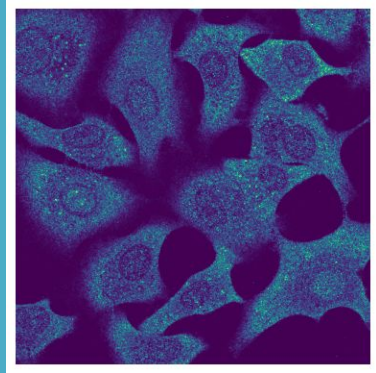
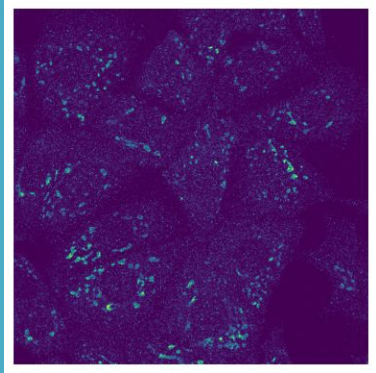
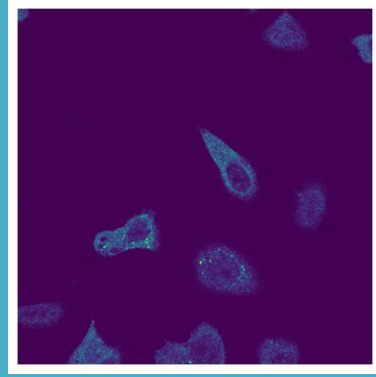
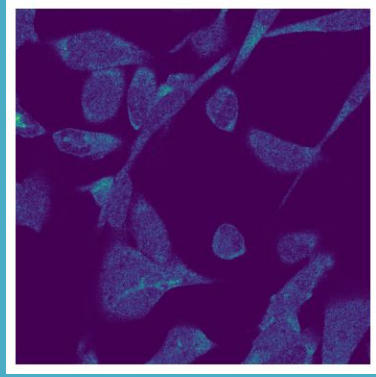
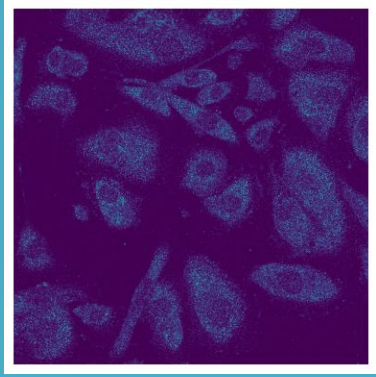
Results: Nuclear Speckles



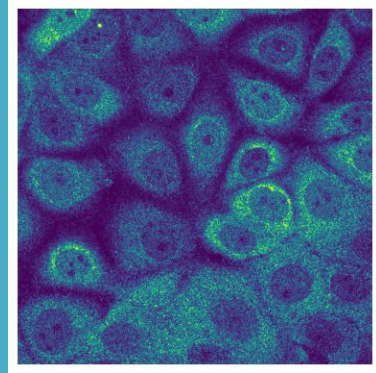
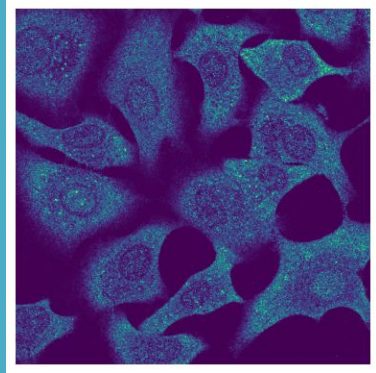
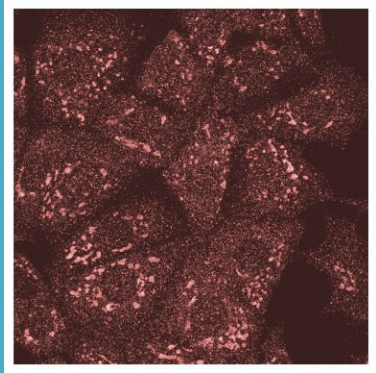
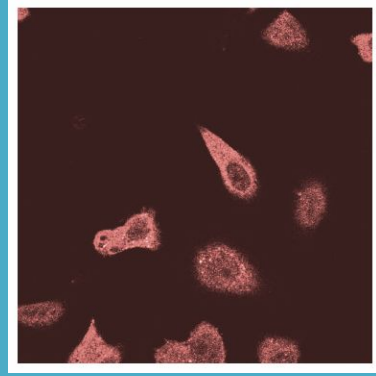
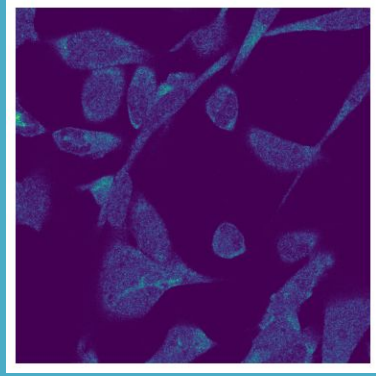
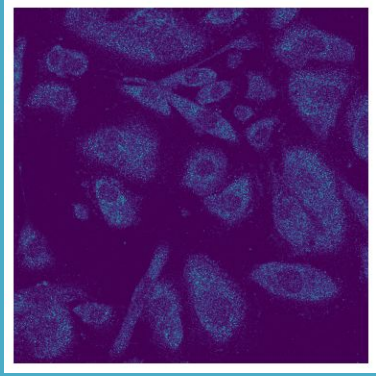
Results: Nuclear Speckles



Results: Cytosol



Results: Cytosol



Results

0

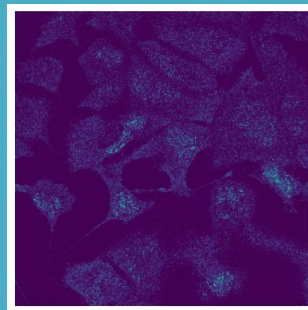
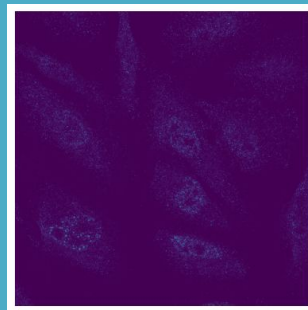
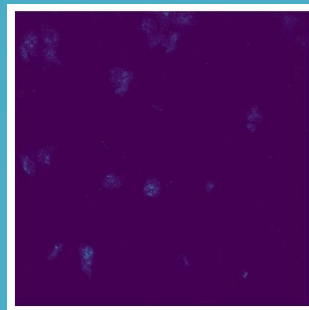
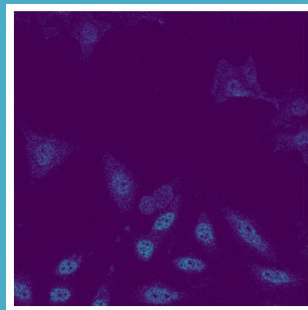
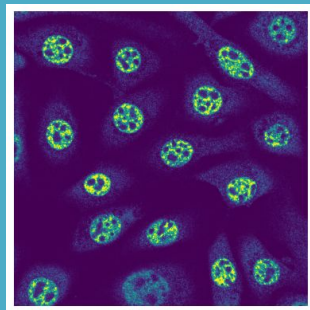
0.25

0.5

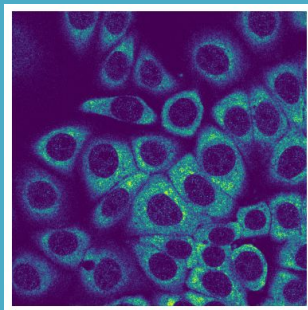
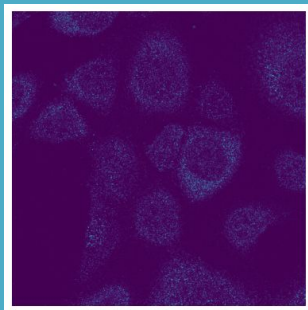
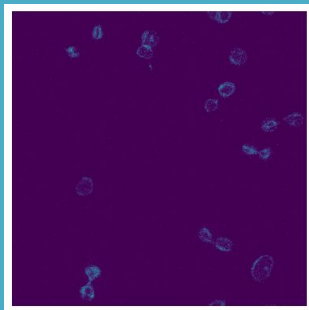
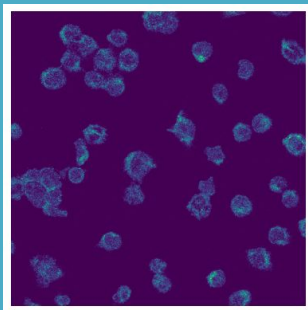
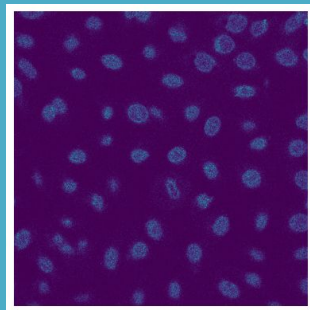
0.75

1

NS

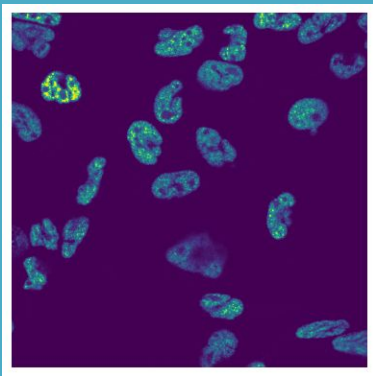


C

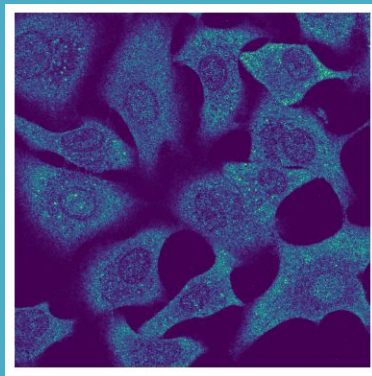


Results

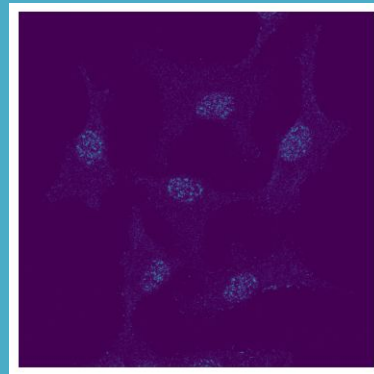
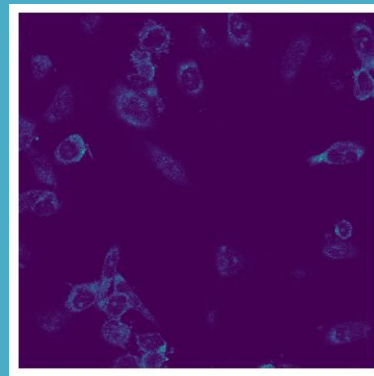
Nuclear
speckles



Cytosol

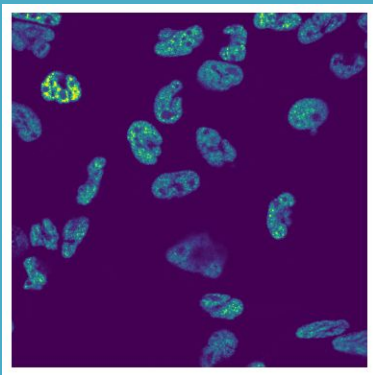


?

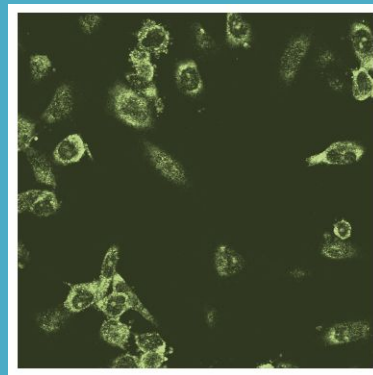
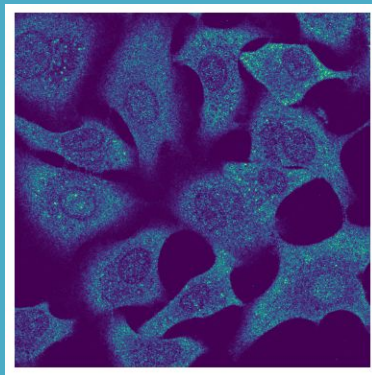


Results

Nuclear
speckles

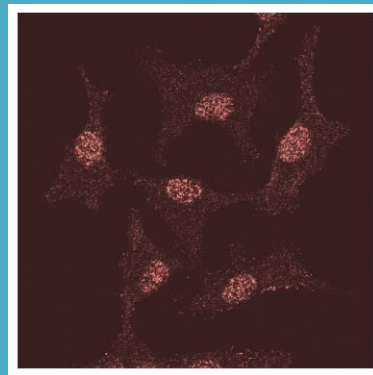


Cytosol



C

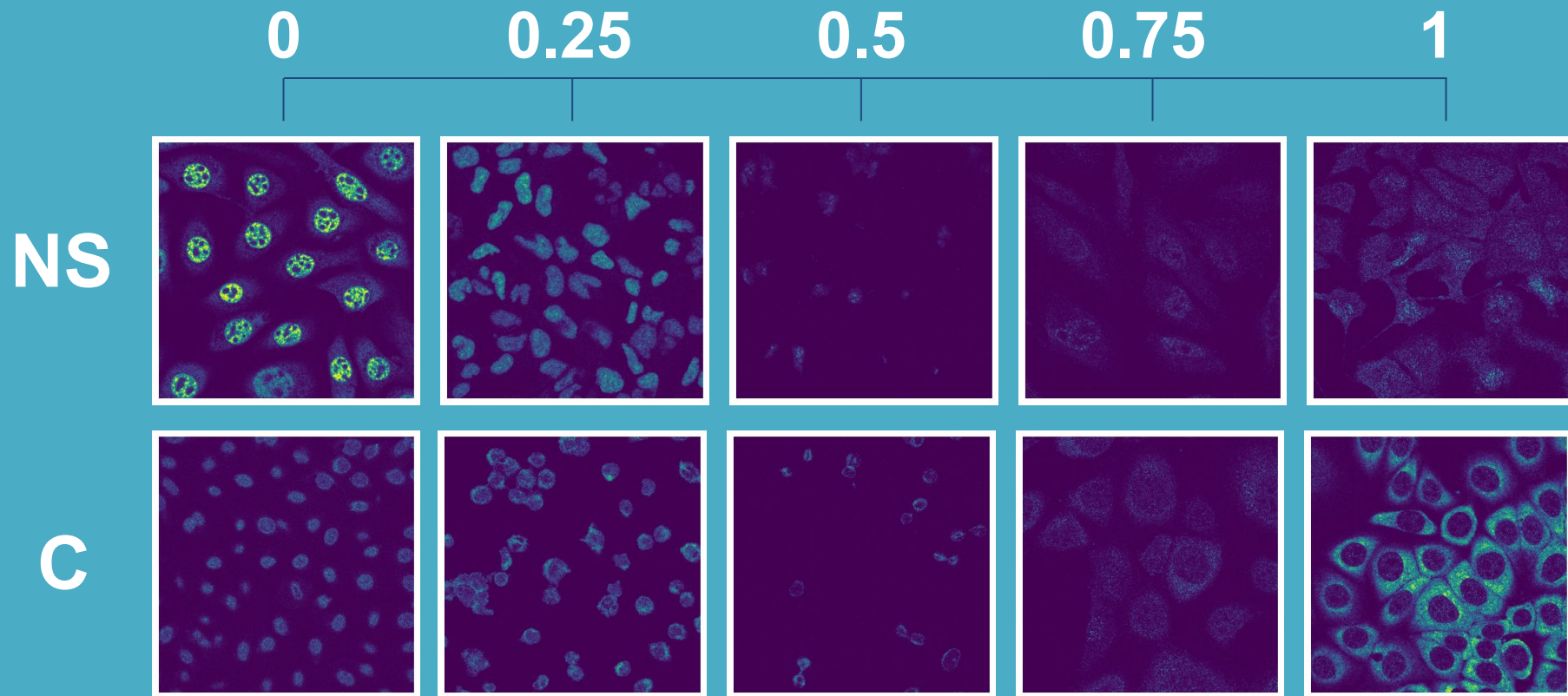
0.528



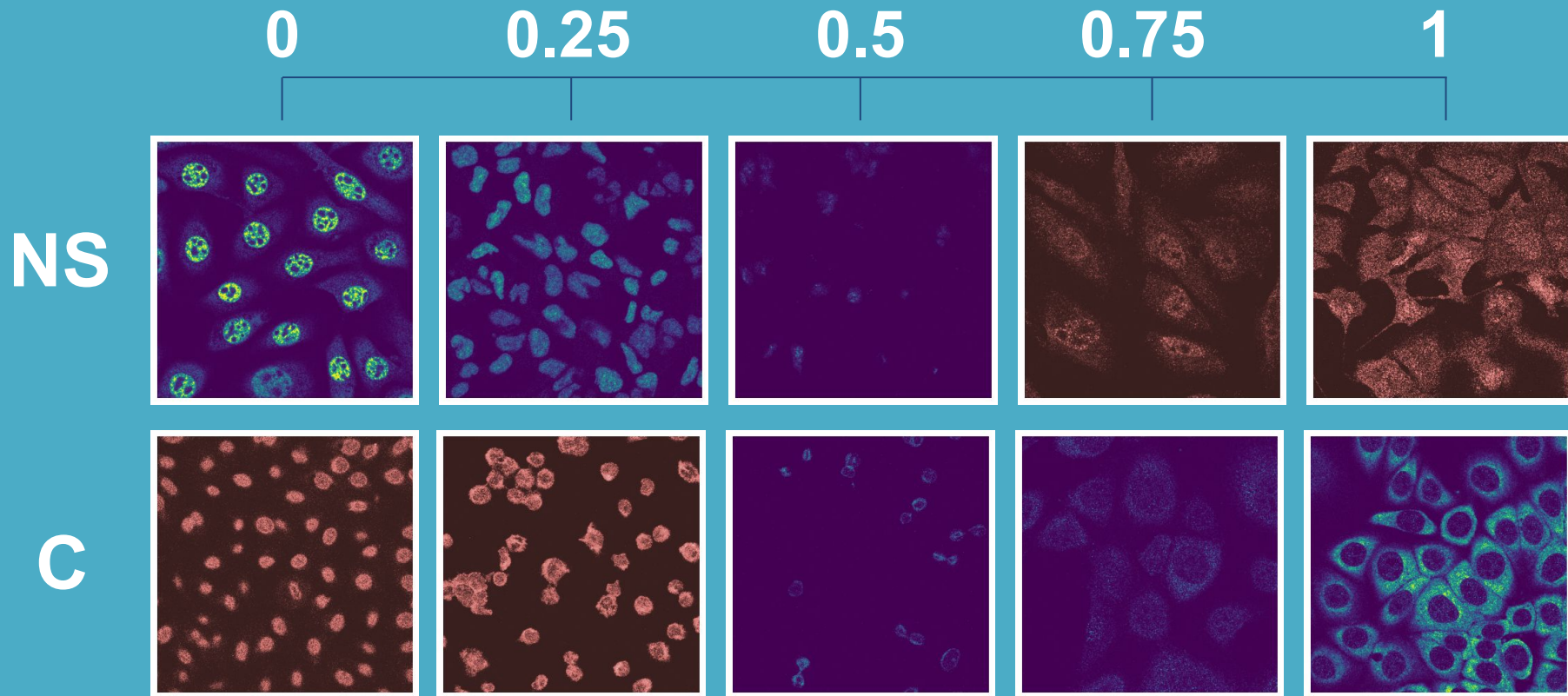
NS

0.550

Results



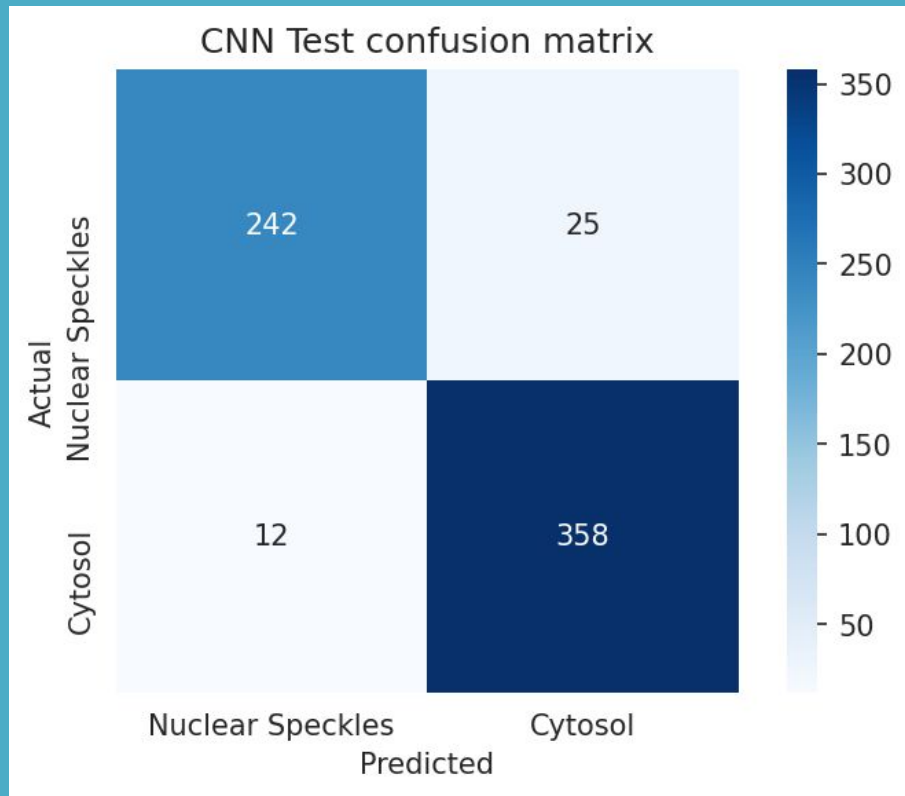
Results



Results

- CNN
 - 3 block VGG
 - Output: 0 to 1
 - Nuc Spec: < 0.5
 - Cytosol: > 0.5

Accuracy Score: 0.930





Conclusion

- CNN was a better classifier than RF for this image data
- Instances of CNN model “uncertainty” may require human intervention



Section Break

Insert the title of your subtitle Here

Results

- Cytosol less likely to be classified incorrectly

