Breast Tumor Classifier

Classifies Histopathological Images as Benign or Malignant

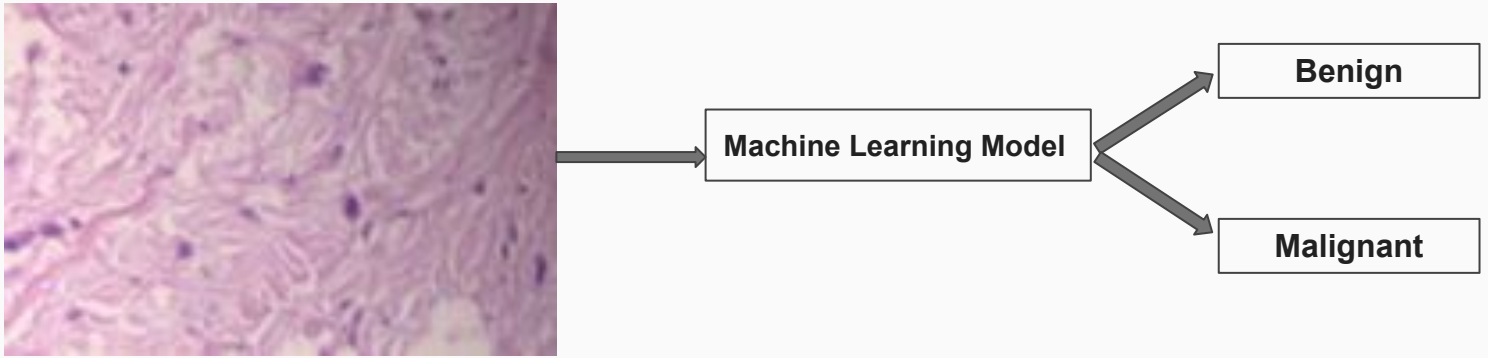
Subject Area

Intersection of Medical Imaging, Machine Learning, and Histopathology

- Medical Imaging: Utilizing histopathological images
- Machine Learning: Applying algorithms for classification
- Histopathology: Analysis of tissue samples under a microscope

Problem Statement

- Develop a Machine Learning model to classify breast tumors as Benign or Malignant



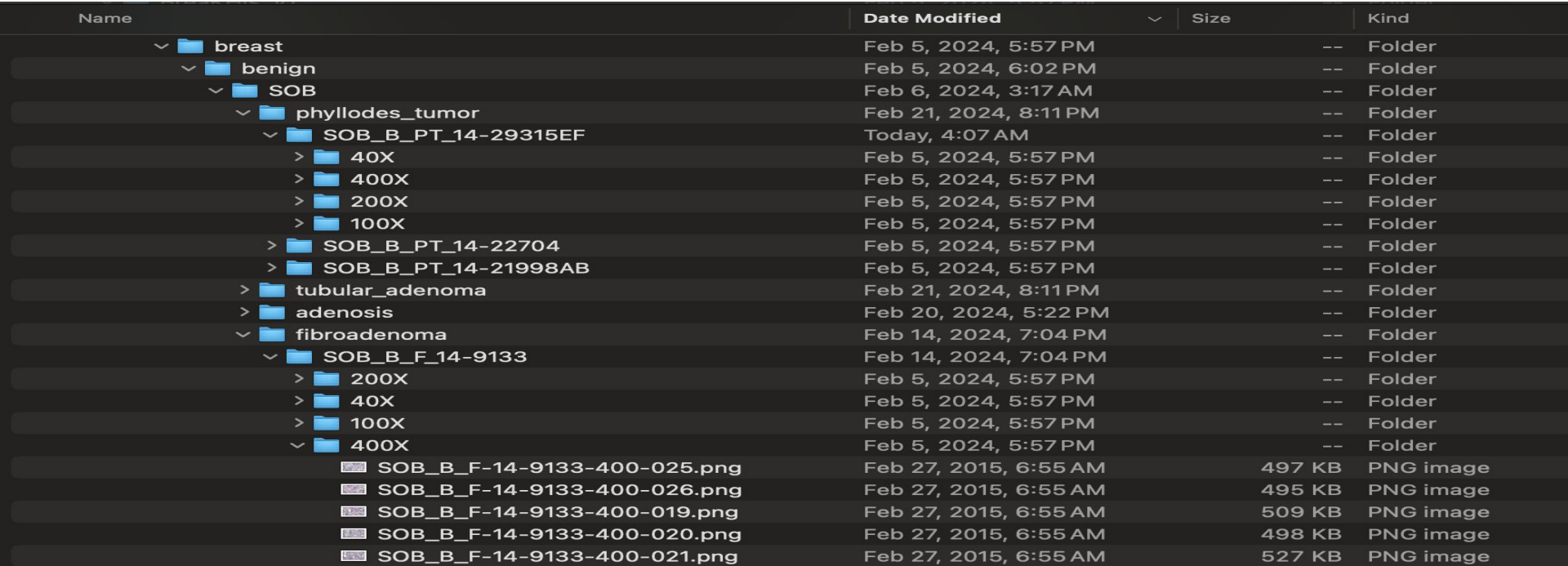
Histopathological Image

Data Source

Breast Cancer Histopathological Database (BreakHis)

https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/

Understanding the folder structures



Name	Date Modified	Size	Kind
breast	Feb 5, 2024, 5:57 PM	--	Folder
benign	Feb 5, 2024, 6:02 PM	--	Folder
SOB	Feb 6, 2024, 3:17 AM	--	Folder
phyllodes_tumor	Feb 21, 2024, 8:11 PM	--	Folder
SOB_B_PT_14-29315EF	Today, 4:07 AM	--	Folder
40X	Feb 5, 2024, 5:57 PM	--	Folder
400X	Feb 5, 2024, 5:57 PM	--	Folder
200X	Feb 5, 2024, 5:57 PM	--	Folder
100X	Feb 5, 2024, 5:57 PM	--	Folder
SOB_B_PT_14-22704	Feb 5, 2024, 5:57 PM	--	Folder
SOB_B_PT_14-21998AB	Feb 5, 2024, 5:57 PM	--	Folder
tubular_adenoma	Feb 21, 2024, 8:11 PM	--	Folder
adenosis	Feb 20, 2024, 5:22 PM	--	Folder
fibroadenoma	Feb 14, 2024, 7:04 PM	--	Folder
SOB_B_F_14-9133	Feb 14, 2024, 7:04 PM	--	Folder
200X	Feb 5, 2024, 5:57 PM	--	Folder
40X	Feb 5, 2024, 5:57 PM	--	Folder
100X	Feb 5, 2024, 5:57 PM	--	Folder
400X	Feb 5, 2024, 5:57 PM	--	Folder
SOB_B_F-14-9133-400-025.png	Feb 27, 2015, 6:55 AM	497 KB	PNG image
SOB_B_F-14-9133-400-026.png	Feb 27, 2015, 6:55 AM	495 KB	PNG image
SOB_B_F-14-9133-400-019.png	Feb 27, 2015, 6:55 AM	509 KB	PNG image
SOB_B_F-14-9133-400-020.png	Feb 27, 2015, 6:55 AM	498 KB	PNG image
SOB_B_F-14-9133-400-021.png	Feb 27, 2015, 6:55 AM	527 KB	PNG image

Insights and decision making

Benign breast tumors:

- adenosis (A)
- fibroadenoma (F)
- phyllodes tumor (PT)
- tubular adenona (TA)

Malignant tumors (breast cancer):

- carcinoma (DC)
- lobular carcinoma (LC)
- mucinous carcinoma (MC)
- papillary carcinoma (PC)

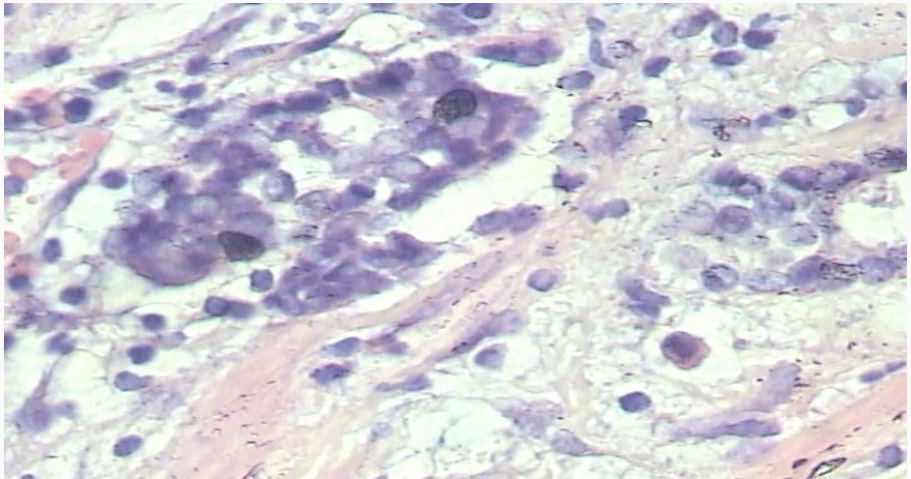
Different magnification factors: 40X, 100X, 200X, and 400X

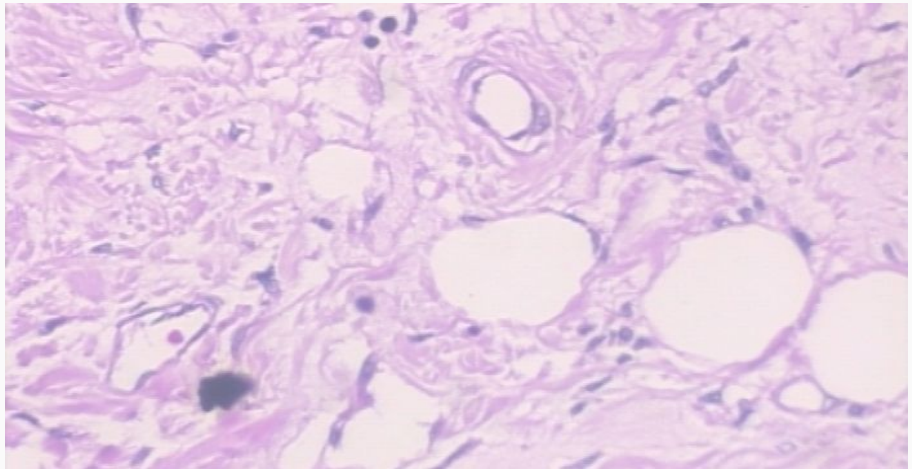
The task is to predict one of two possible outcomes - Benign or Malignant

Naming conventions of .png files

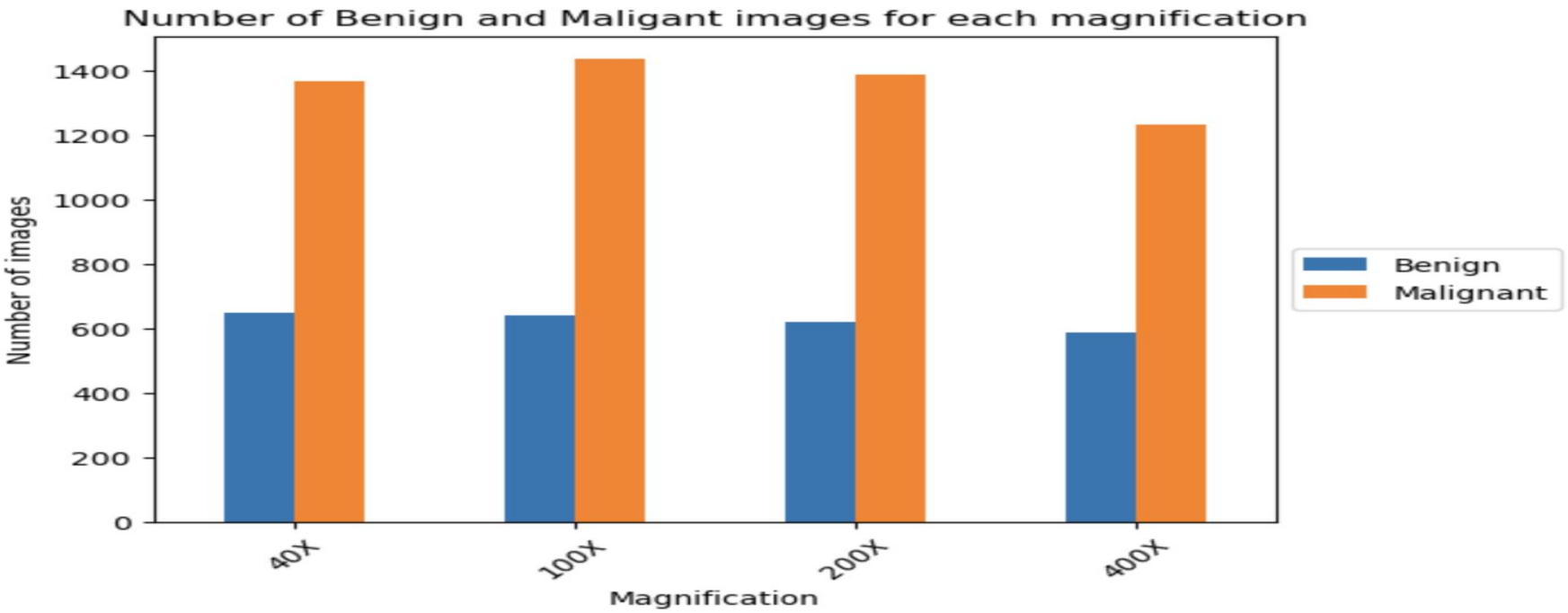
<BIOPSY_PROCEDURE>_<TUMOR_CLASS>_<TUMOR_TYPE>-<YEAR>-<SLIDE_ID>-<MAG>-<SEQ>

Example:
SOB_B_F-14-9133-400-025.png; SOB_M_DC-14-2523-400-026.png

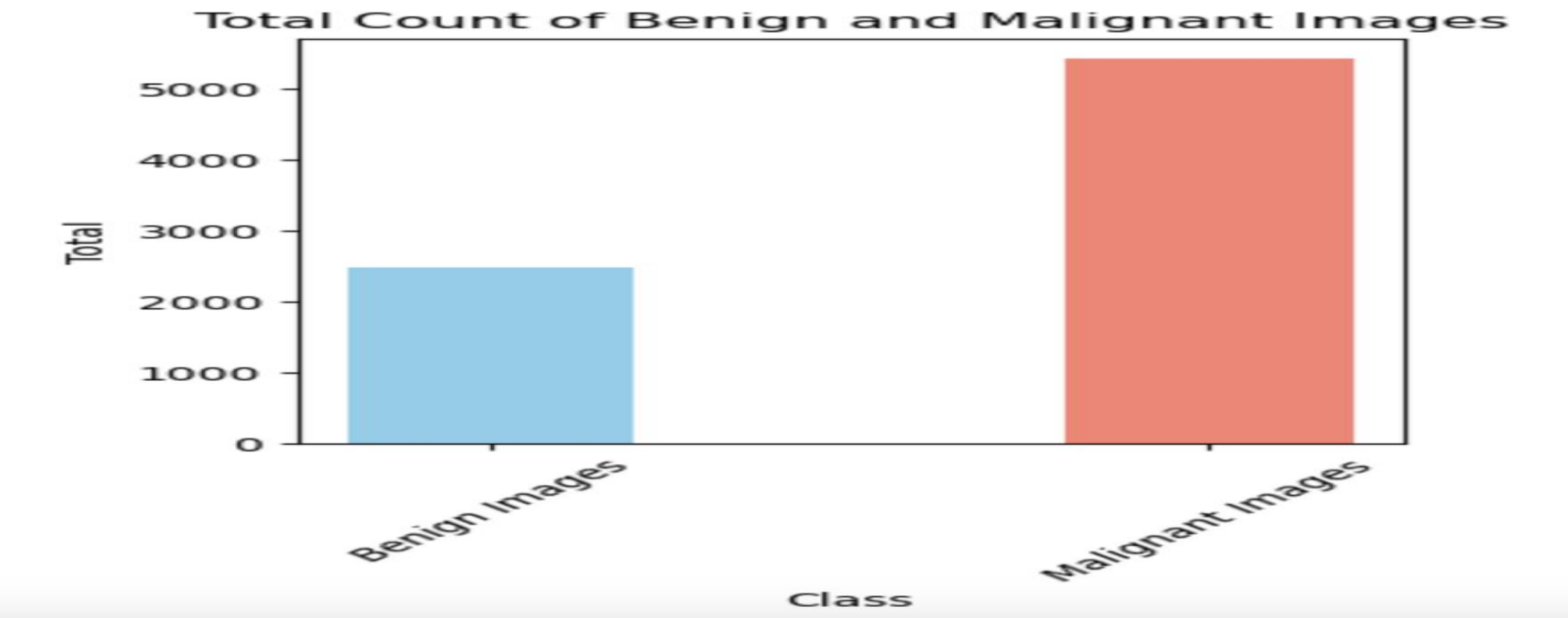




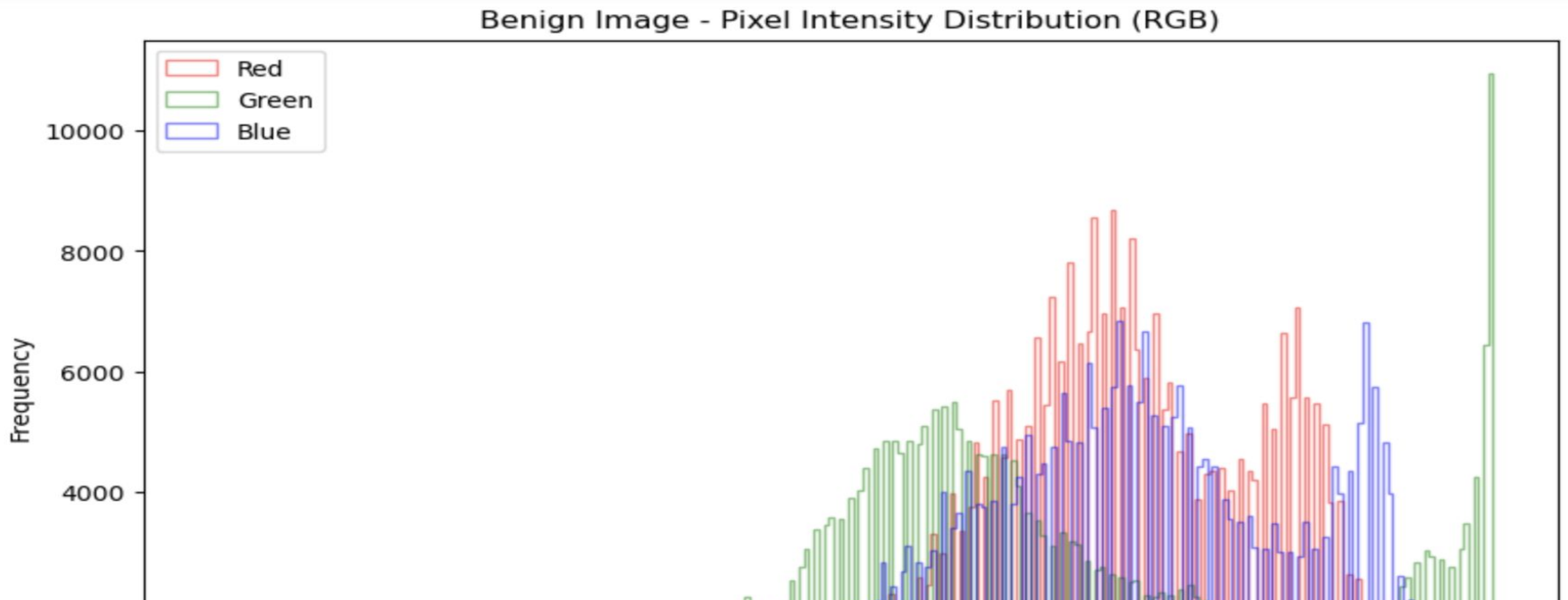
Exploratory Data Analysis



Class imbalance

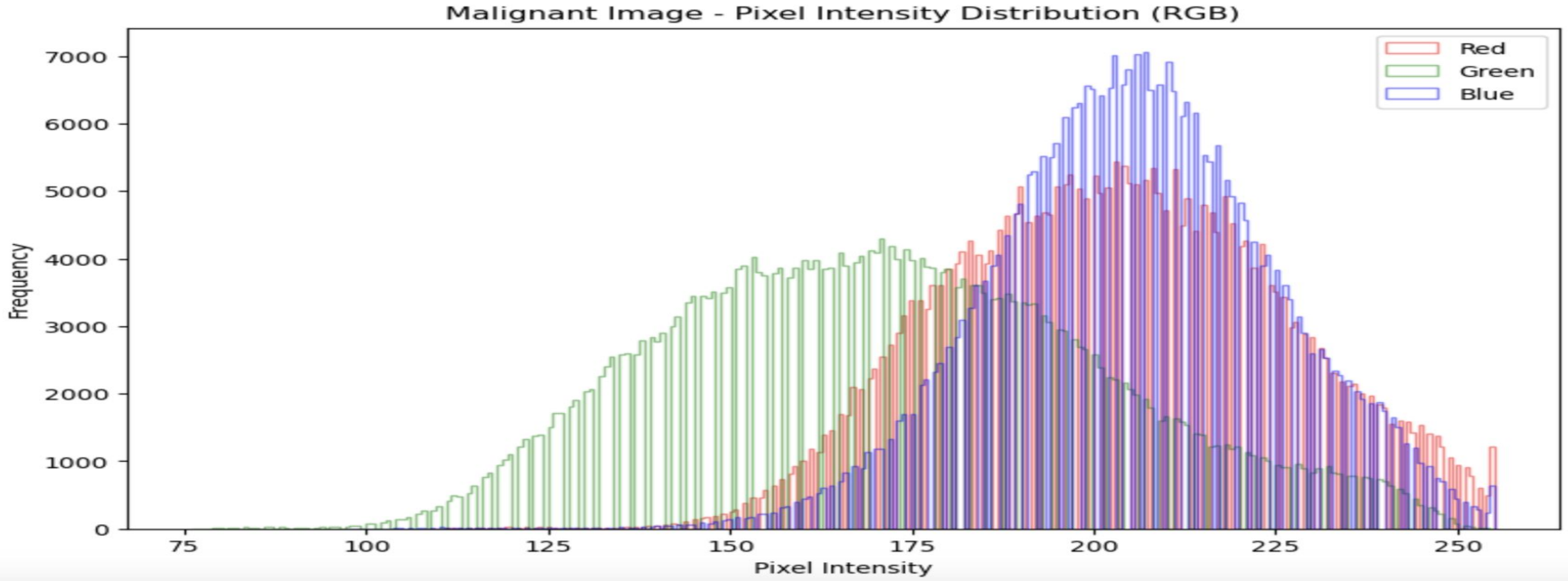


More EDA


Benign Image - Pixel Intensity Distribution (RGB)
Red
Green
Blue
Frequency
10000
8000
6000
4000

More EDA

```
plt.show()
```



Dataset Folds.csv

Out[29]:

	fold	mag	grp	filename
0	1	100	train	BreaKHis_v1/histology_slides/breast/benign/SOB/adenosis/SOB_B_A_14-22549AB/100X/SOB_B_A-14-22549AB-100-001.png
1	1	100	train	BreaKHis_v1/histology_slides/breast/benign/SOB/adenosis/SOB_B_A_14-22549AB/100X/SOB_B_A-14-22549AB-100-002.png
2	1	100	train	BreaKHis_v1/histology_slides/breast/benign/SOB/adenosis/SOB_B_A_14-22549AB/100X/SOB_B_A-14-22549AB-100-003.png
3	1	100	train	BreaKHis_v1/histology_slides/breast/benign/SOB/adenosis/SOB_B_A_14-22549AB/100X/SOB_B_A-14-22549AB-100-004.png
4	1	100	train	BreaKHis_v1/histology_slides/breast/benign/SOB/adenosis/SOB_B_A_14-22549AB/100X/SOB_B_A-14-22549AB-100-005.png
...	...	...	...	...
39540	5	400	test	BreaKHis_v1/histology_slides/breast/malignant/SOB/papillary_carcinoma/SOB_M_PC_15-190EF/400X/SOB_M_PC-15-190EF-400-011.png
39541	5	400	test	BreaKHis_v1/histology_slides/breast/malignant/SOB/papillary_carcinoma/SOB_M_PC_15-190EF/400X/SOB_M_PC-15-190EF-400-012.png
39542	5	400	test	BreaKHis_v1/histology_slides/breast/malignant/SOB/papillary_carcinoma/SOB_M_PC_15-190EF/400X/SOB_M_PC-15-190EF-400-013.png
39543	5	400	test	BreaKHis_v1/histology_slides/breast/malignant/SOB/papillary_carcinoma/SOB_M_PC_15-190EF/400X/SOB_M_PC-15-190EF-400-014.png
39544	5	400	test	BreaKHis_v1/histology_slides/breast/malignant/SOB/papillary_carcinoma/SOB_M_PC_15-190EF/400X/SOB_M_PC-15-190EF-400-015.png

39545 rows × 4 columns

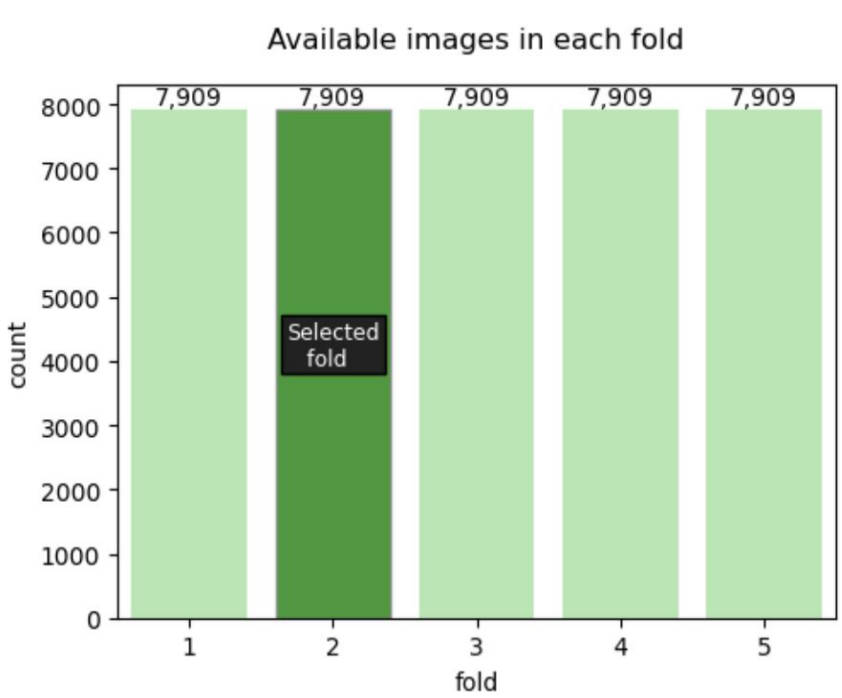
Create label column

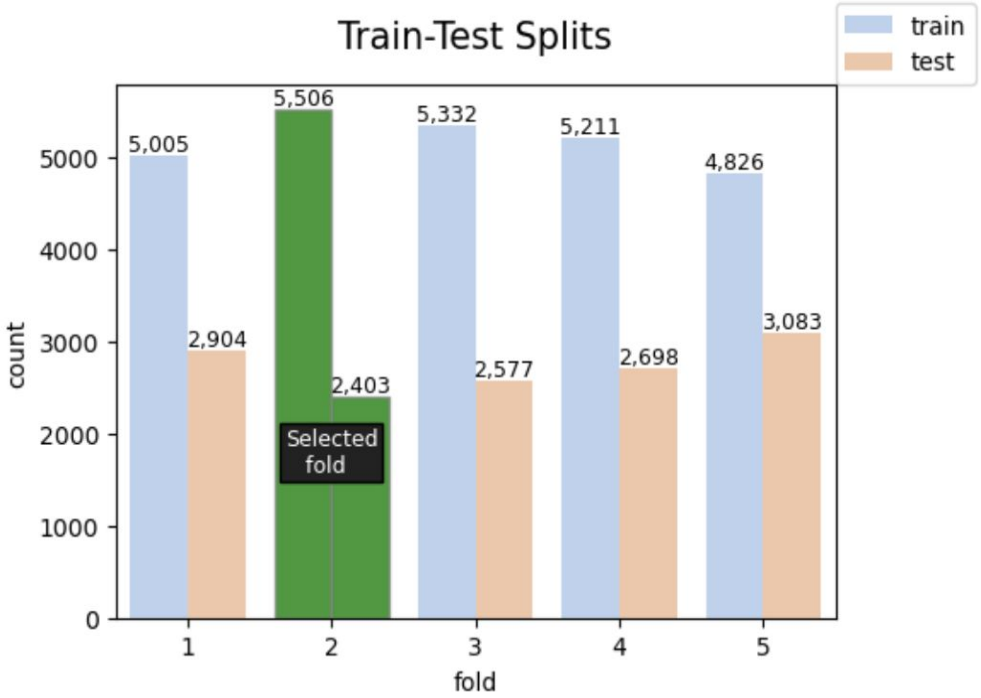
```
# Create a new column named label to label the images

fold_info["label"] = fold_info["filename"].str.extract("(malignant|benign)")
fold_info.head()
```

	fold	mag	grp	filename	label
0	1	100	train	BreaKHis_v1/histology_slides/breast/benign/SOB/adenosis/SOB_B_A_14-22549AB/100X/SOB_B_A-14-22549AB-100-001.png	benign
1	1	100	train	BreaKHis_v1/histology_slides/breast/benign/SOB/adenosis/SOB_B_A_14-22549AB/100X/SOB_B_A-14-22549AB-100-002.png	benign
2	1	100	train	BreaKHis_v1/histology_slides/breast/benign/SOB/adenosis/SOB_B_A_14-22549AB/100X/SOB_B_A-14-22549AB-100-003.png	benign
3	1	100	train	BreaKHis_v1/histology_slides/breast/benign/SOB/adenosis/SOB_B_A_14-22549AB/100X/SOB_B_A-14-22549AB-100-004.png	benign
4	1	100	train	BreaKHis_v1/histology_slides/breast/benign/SOB/adenosis/SOB_B_A_14-22549AB/100X/SOB_B_A-14-22549AB-100-005.png	benign

Train - Test Split



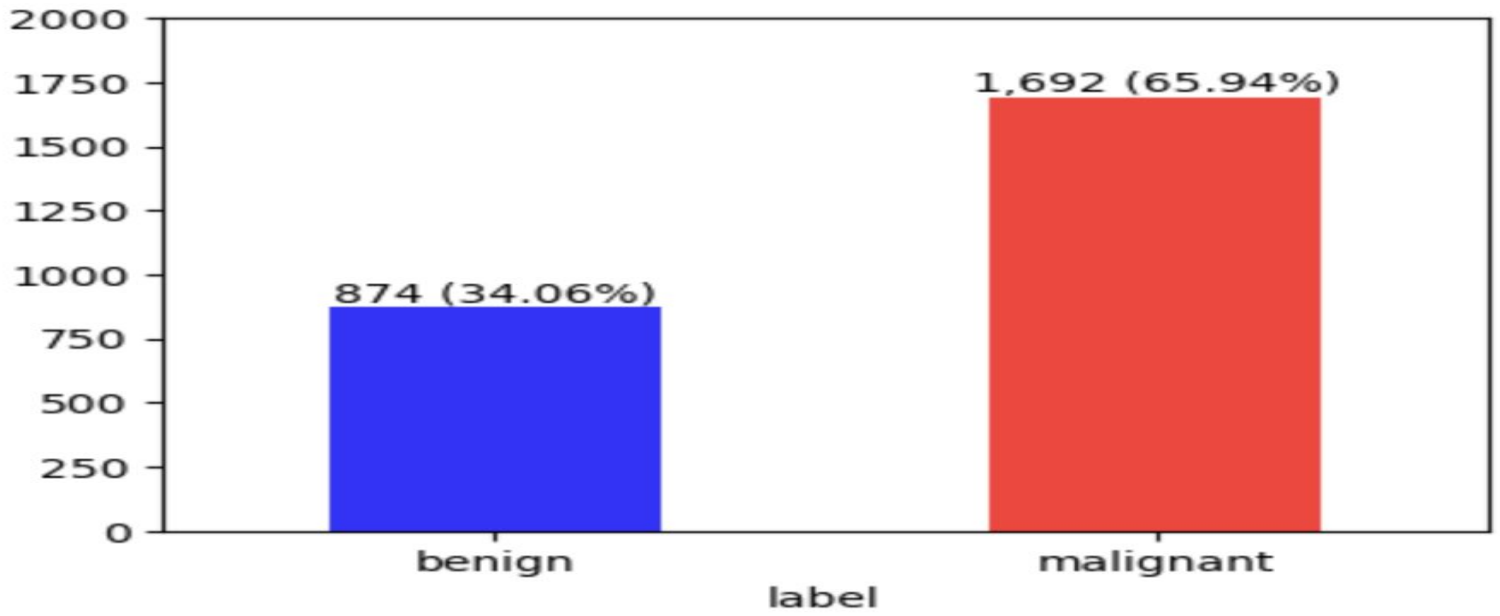


Transfer Learning

Leveraging knowledge from pre-trained models and adapting it to this dataset.

- MobileNetV3Small
- EfficientNetV2B1
- VGG16
- ResNet50V2

Target distribution

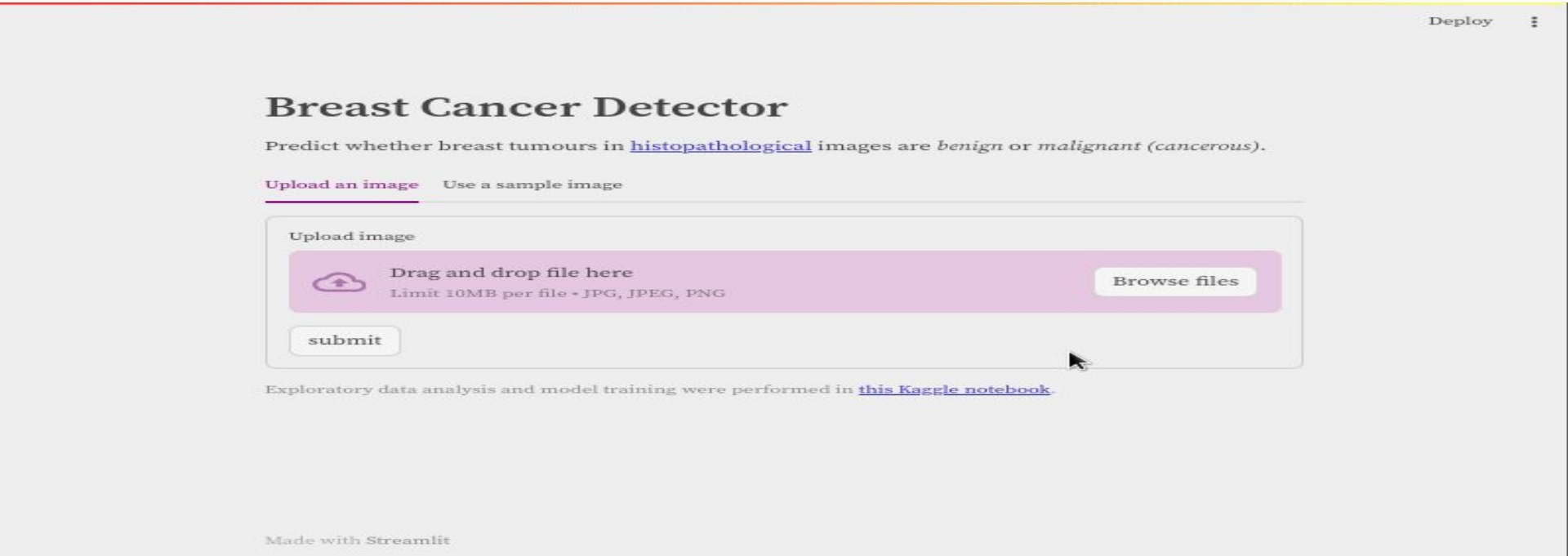


Model Evaluation

	MobileNetV3	EfficientNetB1	VGG16	ResNet50V2
Loss				
ROC-AUC				
Accuracy				

Merge & Center

Productizing the work



Thank you!