OUTLINE OF WHAT TO DO

- EXPLANATION OF EACH MEDICAL TERM
    - We need a deeper understanding of each variable we use and on the data we have
- DATA PREPROCESSING
    - Look for notebook and use their same methods to clean data path and to generate a clearer link to each img and their respective metadata
- EDA (VISUALIZATION AND GENERAL INFORMATION ABOUT THE DATASET)
    - FOCUS ON EACH FEATURES, HOW IT DISTRIBUTE AND HOW THE TARGET FEATURE DISTRIBUTE
    - CORR MATRIX AND ANY METHOD IT CAME TO MIND TO HAVE DEEPER INSIGHT ABOUT THE DATA
- DATE ENGINEERING
    - CREATE DUMMY FOR CATEGORICAL VARIABLES
    - MAPPING FOR TARGET VARIABLE
    - ANY MORE FEATURE ENG NEEDED
- DATA AUGMENTATION
    - KEEP IT AS SIMPLE AS POSSIBLE, BUT LET'S TRY DIFF TECNIQUES
    - SIMPLE CAUSE IT MAY CAUSE OVERFITTING
- CHOOSING THE METRIC
    - (F BETA SCORE?), WE MAY TEST THIS IDK
    - EACH MODEL HAS TO SHOW ALSO THE CONFUSION MATRIX
- CNN MODEL
    - DEFINE ARCHITECTURE (Experiment diff one and use the one i made for HW2 task 5 as inspiration)
    - FOCUS DIFF TECNIQUES, IF U FIND MORE THAN THE ONE I USED BETTER
    - PLOT TRAIN AND VAL LOSS AND METRIC
- PREMADE CNN MODELS (OF UR CHOICE)
    - Just apply those using tecniques from HW2 task 5 on AlexNet
    - EXPERIMENT AT LEAST WITH 2
- CNN + BOOSTING
    - USE OF AN INTERMEDIATE LAYER
    - TUNING ON BOOSTING WITH FEATURES RETREIVED
- EMBEDDING
    - COMBINE DIFF BOOSTING MODELS
- COMPARISON OF RESULTS
    - OF ALL MODELS USED
- XAI
    - USE SHAP, LIME AND SALIENCY MAP TO UNDERSTAND HOW MODEL WORK AND ABOUT THE MISSCLASSIFIED ONE

USEFUL LINKS

Dataset:

- https://www.kaggle.com/datasets/awsaf49/cbis-ddsm-breast-cancer-image-dataset

Basic Eda:

- https://www.kaggle.com/code/awsaf49/breast-cancer-eda
- https://www.kaggle.com/code/kerneler/starter-cbis-ddsm-breast-cancer-image-eef73d6f-4#Conclusion

More advanced EDAs:

- CLEAN PREPROCESSING AND GOOD VISUALIZATIIONS: https://www.kaggle.com/code/hitheshmr/data-visiualization-cbis-ddsm

Pre made CNN:

- GoogleNet (there is also interesting EDA): https://www.kaggle.com/code/princelubisi/breast-cancer-classification-googlenet#Preprocessing
- DenseNet: https://www.kaggle.com/code/angelarentsi/breast-cancer-classification-densenet#Model:DenseNet121

Implementing CNN:

- EASY AND CLEAN https://www.kaggle.com/code/ahmedmedhat1012/cbis-ddsm-breast-cancer-image-dataset-training#Classification-Report
- MORE COMPLICATED: https://www.kaggle.com/code/joshuaampofoyentumi/breast-cancer-cnn

HW2

- https://github.com/vikavl/FDS/tree/main/Homework02

IT MAY BE INTERESTING TO MAKE CLEAN DATA FOR PREPARATION:

- https://www.kaggle.com/code/rumiyyaalili/save-images-for-gan