Breast Cancer Diagnosis Using Deep Transfer Learning

CMPE 258 - PROJECT PRESENTATION

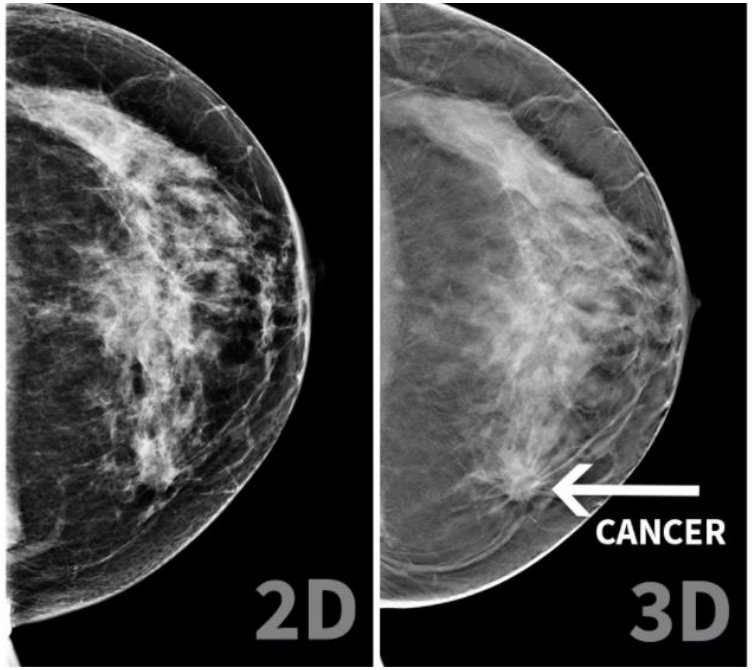
ABHISHEK BAIS, WASAE QURESHI, SUBARNA CHOWDHARY SOMA
GRADUATE STUDENT COMP SCI, SJSU

Pic Credits: PA Department of Health

The American Cancer Society's estimates for 2021

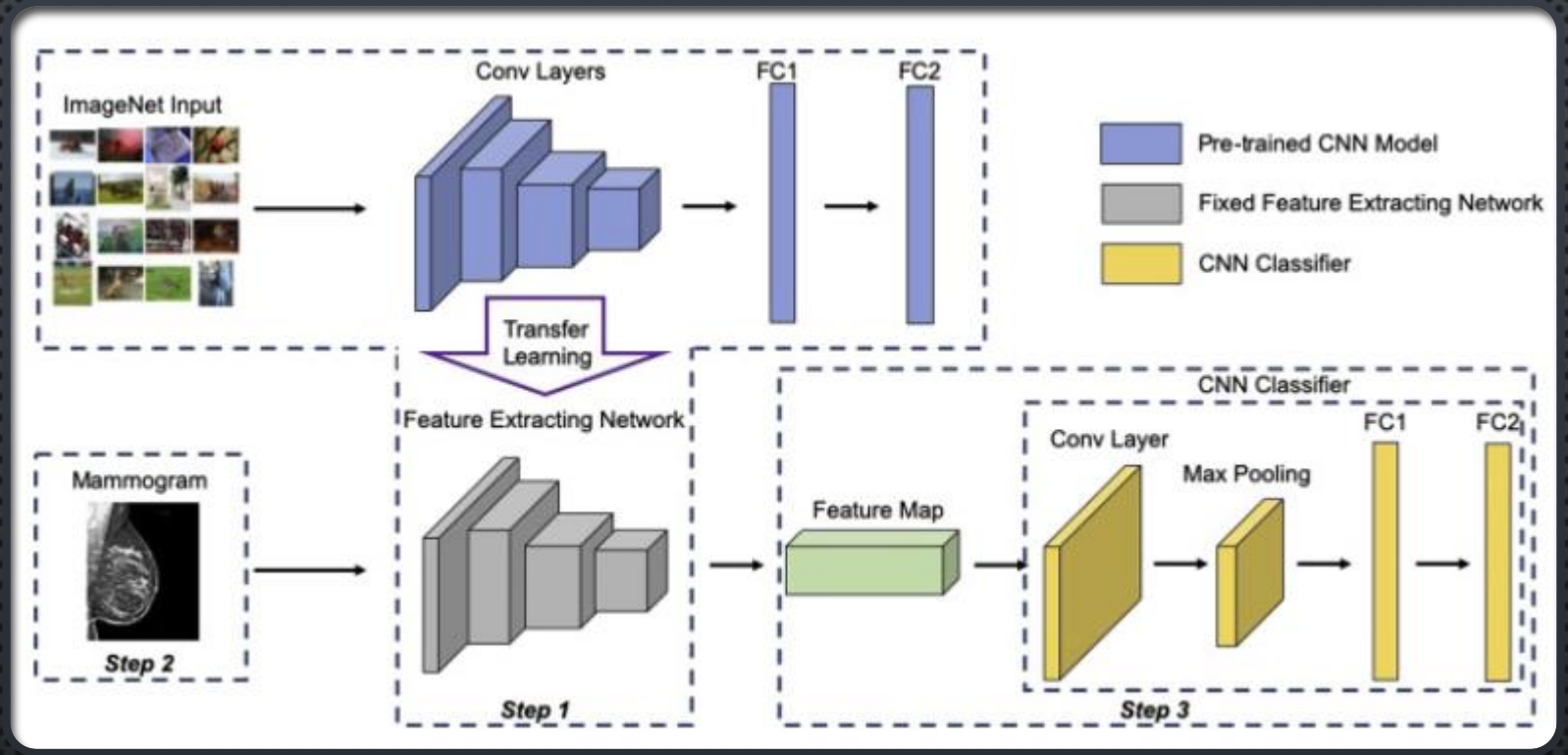
- 281,550 new cases of invasive breast cancer
- 43,600 women will die from breast cancer

EARLY, ACCURATE DETECTION OF CANCER CAN SAVE LIVES



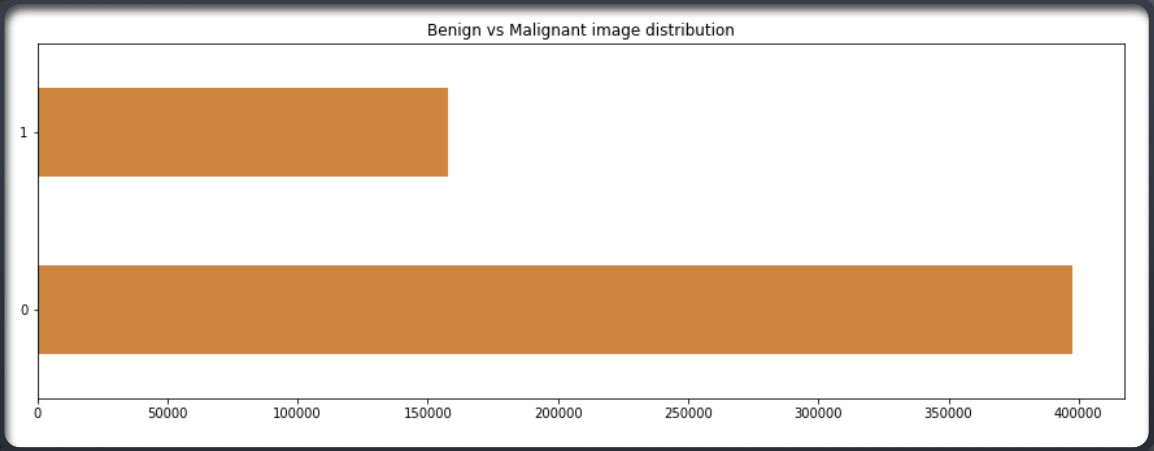
Pic Credits: A Cancerous Word Press Report

TRADITIONAL METHODS OF DETECTION SUCH AS MAMMOGRAPHY REQUIRE EXPERTS



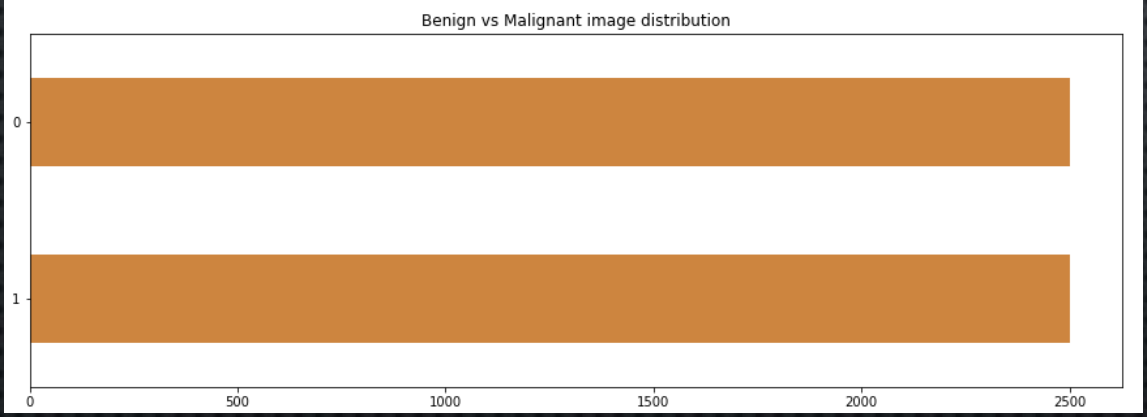
Pic Credits: Journal of American College of Radiology

DOMAIN TRANSFER LEARNING PROVIDES ABILITY FOR FAST, ACCURATE DETECTION REDUCING EXPERT RELIANCE



IMBALANCED INPUT

- 397476 BENIGN CANCER IMAGES
- 157572 MALIGNANT CANCER IMAGES

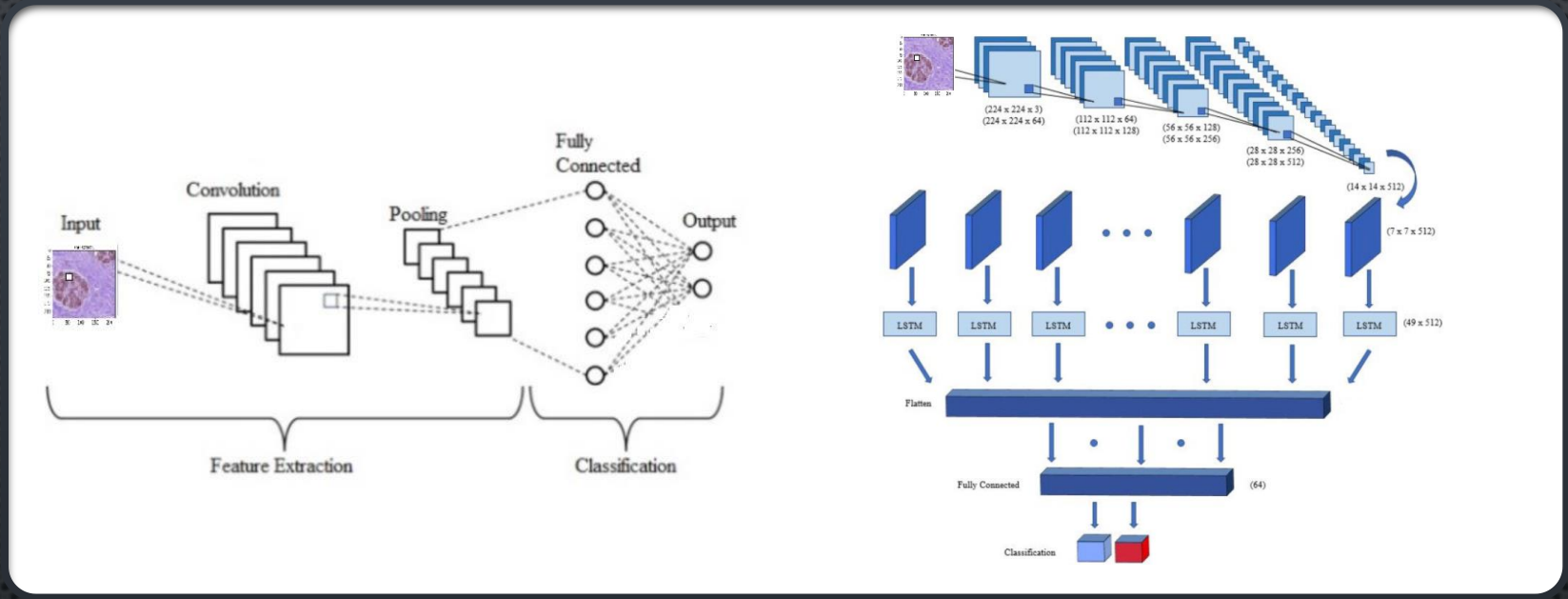


BALANCED FOR STUDY

- 2500 BENIGN CANCER IMAGES
- 2500 MALIGNANT CANCER IMAGES

Breast Cancer Histopathology Images, Case Western Reserve University

DATASET DETAILS

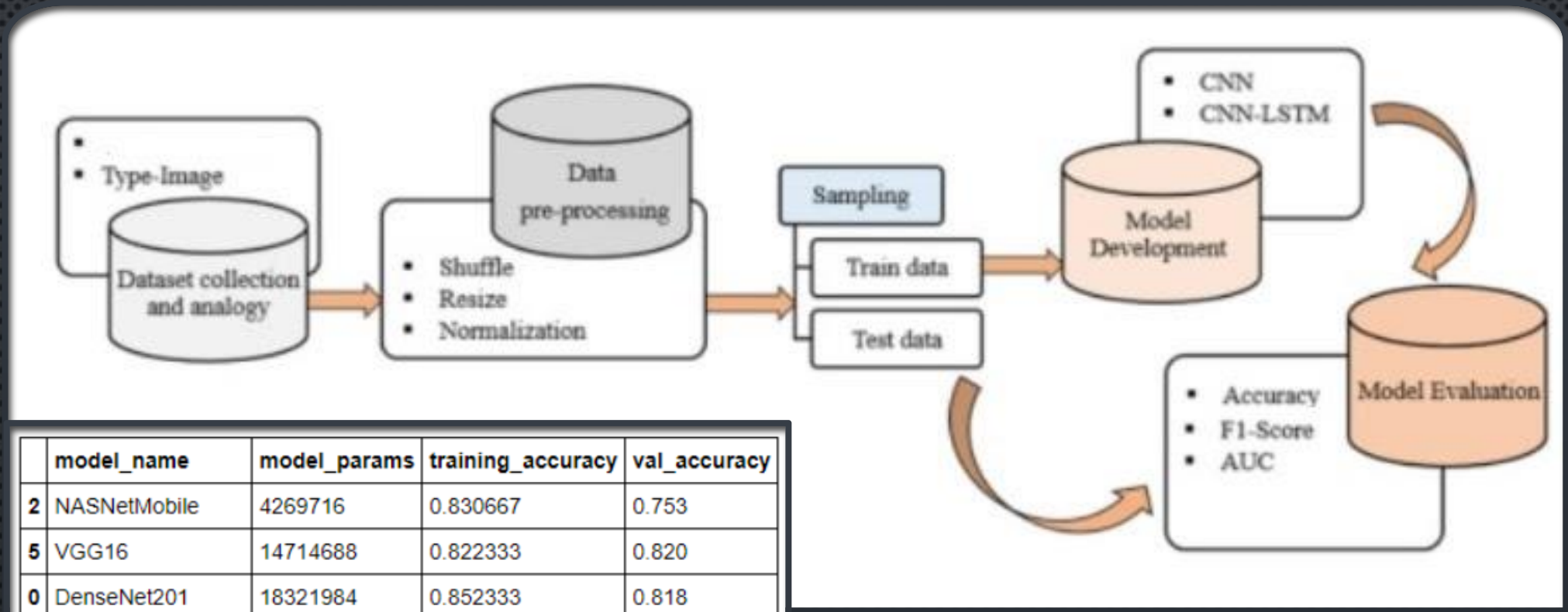


CNN + CUSTOM DNN LAYERS

CNN + LSTM

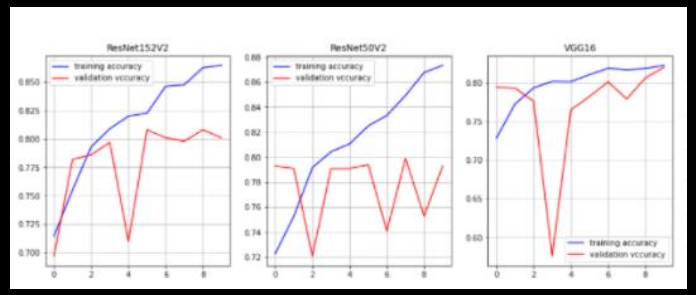
Pic Credits: Md. Zabirul Islam, Md. Milon Islam, Amanullah Asraf, Science Direct

STUDY METHODOLOGY

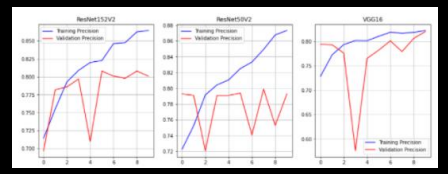


	model_name	model_params	training_accuracy	val_accuracy
2	NASNetMobile	4269716	0.830667	0.753
5	VGG16	14714688	0.822333	0.820
0	DenseNet201	18321984	0.852333	0.818
6	Xception	20861480	0.870667	0.803
4	ResNet50V2	23564800	0.873333	0.793
1	InceptionResNetV2	54336736	0.789667	0.802
3	ResNet152V2	58331648	0.864667	0.801

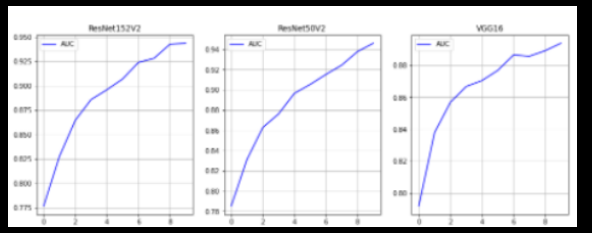
MACHINE LEARNING STAGES AND CNN MODELS



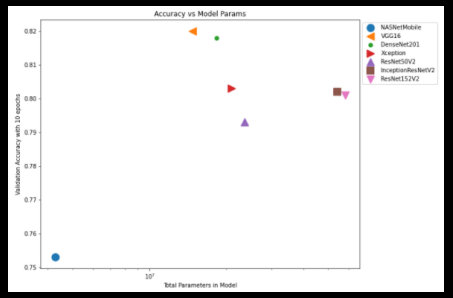
PRECISION



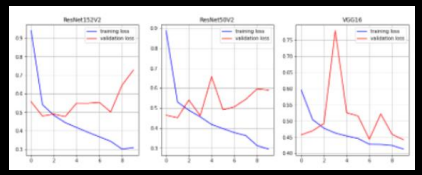
RECALL



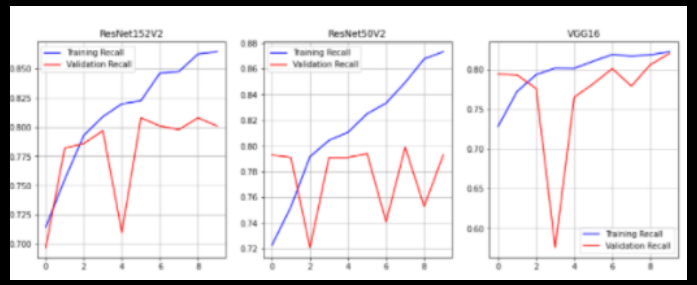
AUC



SIZE VS ACCURACY



LOSS



ACCURACY

MULLER LOOP EVALUATION

	model_name	model_params	training_accuracy	val_accuracy
2	NASNetMobile	4269716	0.830667	0.753
5	VGG16	14714688	0.822333	0.820

BASELINE
- 10 epochs
- Patience 20
- Early stop
- rmsProp optimizer
- Dropout 0.2
- Adaptive learning, minlr 1e-4

	model_name	model_params	training_accuracy	val_accuracy
0	VGG16	14714688	0.845667	0.8

INCREASE EPOCHS
- 20 epochs

	model_name	model_params	training_accuracy	val_accuracy
0	VGG16	14714688	0.839333	0.733

BATCH NORMALIZATION

	model_name	model_params	training_accuracy	val_accuracy
0	VGG16	14714688	0.814	0.793

AUGMENT IMAGES
- Zoom 10%
- Rotate 10% random
- Horizontal flip
- Vertical flip

	model_name	model_params	training_accuracy	val_accuracy
0	VGG16	14714688	0.811667	0.798

DROPOUT 0.5

	model_name	model_params	training_accuracy	val_accuracy
0	VGG16	14714688	0.804	0.789

ADAM OPTIMIZER

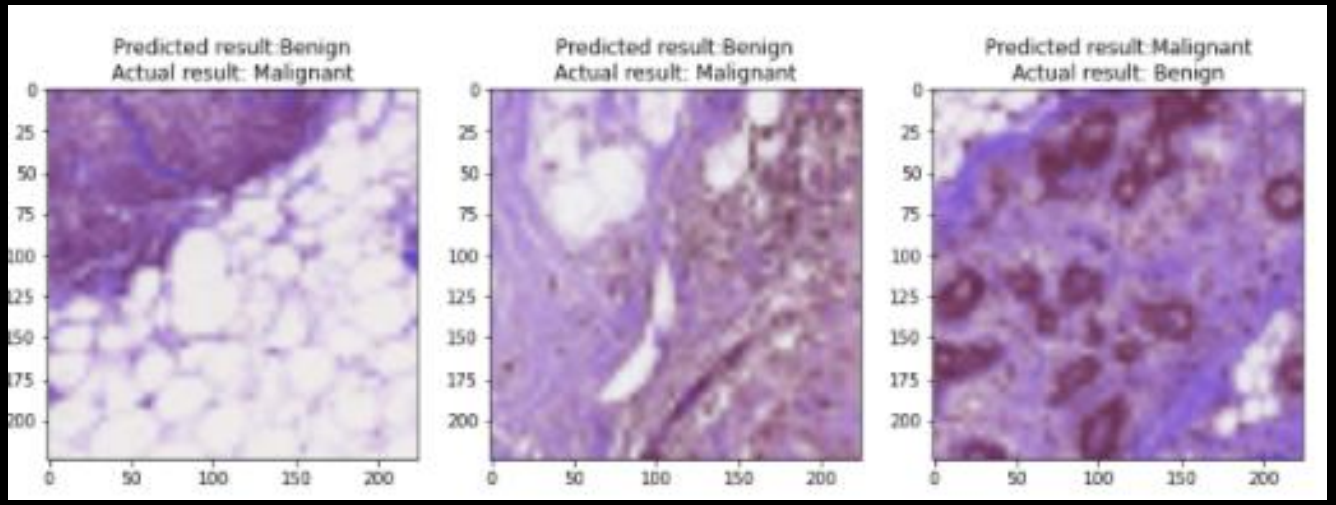
	model_name	model_params	training_accuracy	val_accuracy
0	VGG16	14714688	0.848667	0.826

UNFREEZE last 4 CNN LAYERS

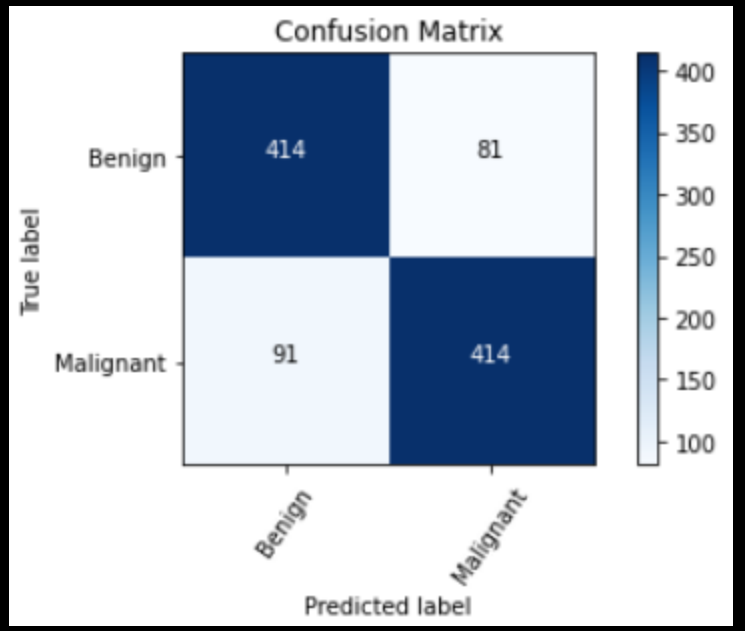
HYPER PARAMETER TUNE VGG16 ENSEMBLE

	model_name	model_params	training_accuracy	val_accuracy
0	VGG16	14714688	0.848667	0.826

ACCURACY



FEW COMMON ERRORS IN PREDICTION

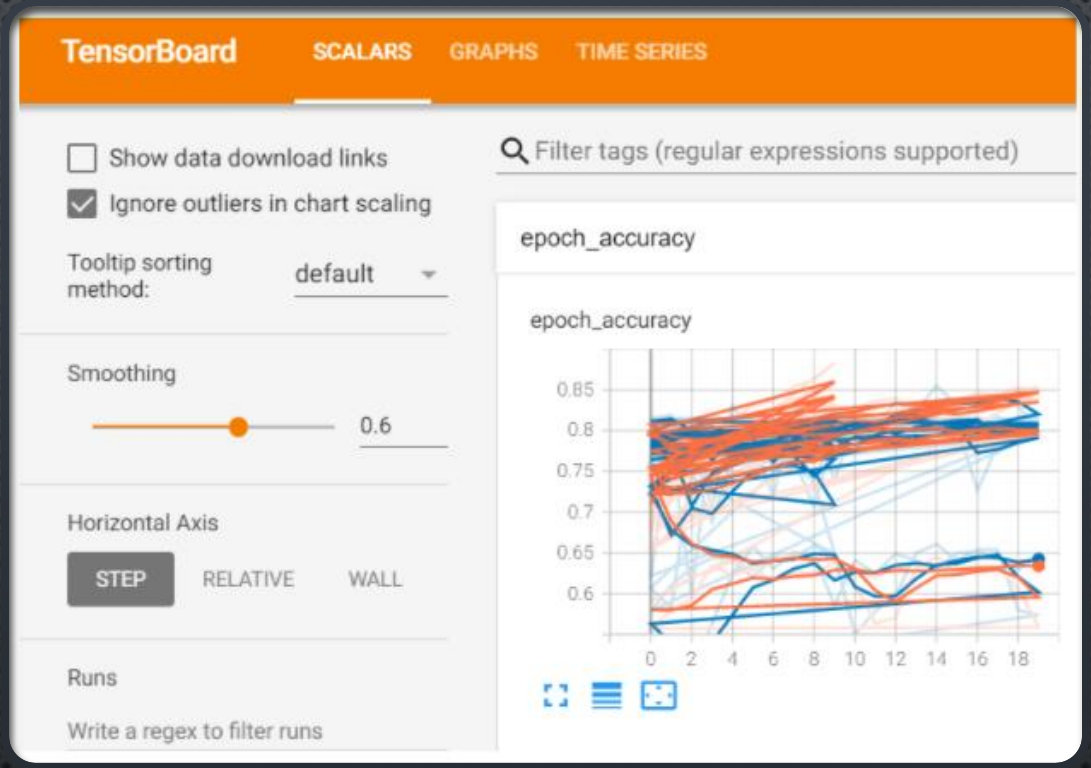


CONFUSION MATRIX

SUPPORT

	precision	recall	f1-score	support
negative	0.82	0.84	0.83	495
positive	0.84	0.82	0.83	505
accuracy			0.83	1000
macro avg	0.83	0.83	0.83	1000
weighted avg	0.83	0.83	0.83	1000

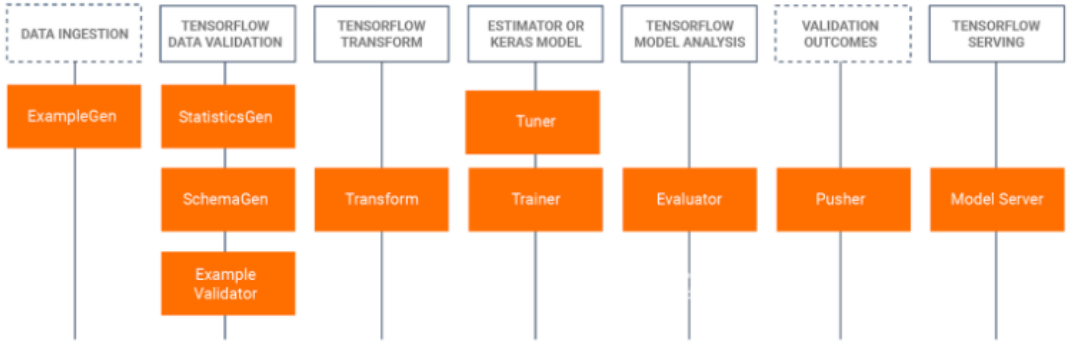
BEST MODEL VGG16



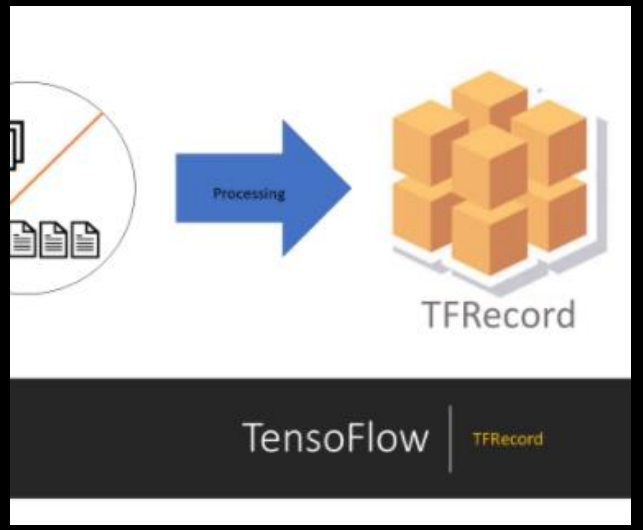
TENSORBOARD INTEGRATION

DEMO LINK - GIT

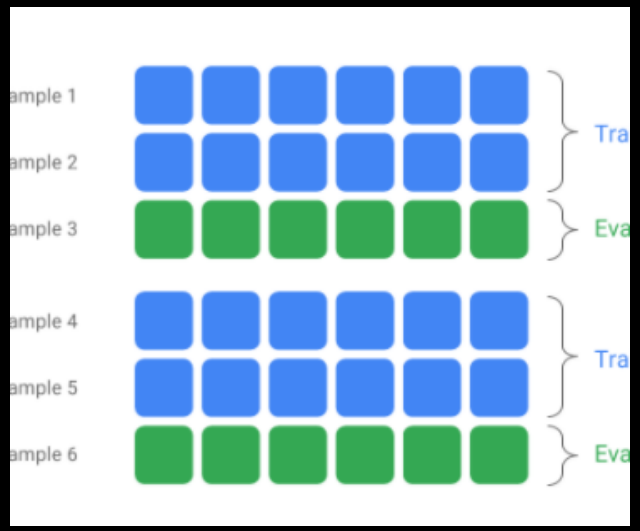
DEMO LINK - COLAB



AUTOMATED PIPELINES WITH TFX

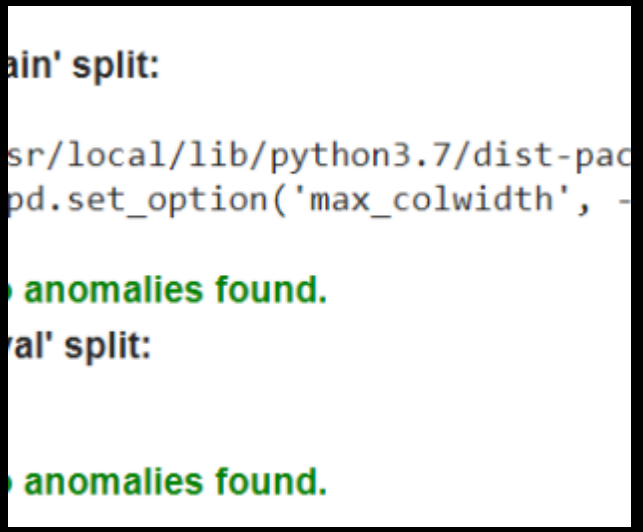


PRE-PROCESS

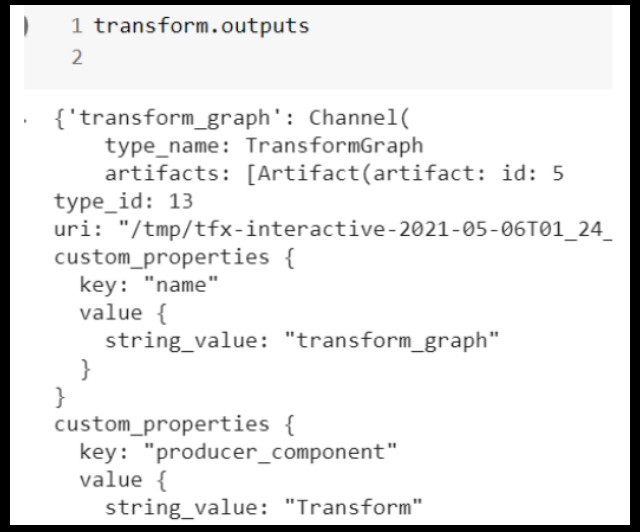


EXAMPLE GEN

STATISTICS GEN



EXAMPLES VALIDATOR

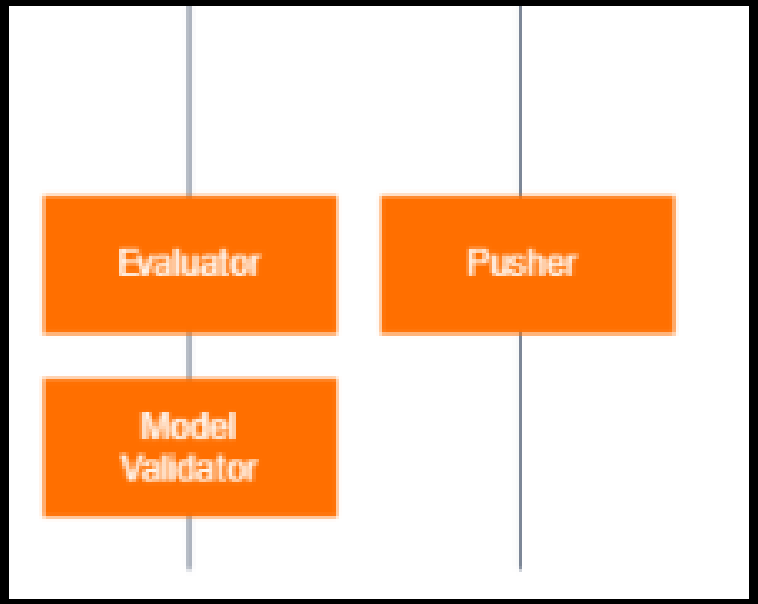


```
1 transform.outputs
2

{'transform_graph': Channel(
    type_name: TransformGraph
    artifacts: [Artifact(artifact: id: 5
type_id: 13
uri: "/tmp/tfx-interactive-2021-05-06T01_24_
custom_properties {
  key: "name"
  value {
    string_value: "transform_graph"
  }
}
custom_properties {
  key: "producer_component"
  value {
    string_value: "Transform"
```

TRANSFORM

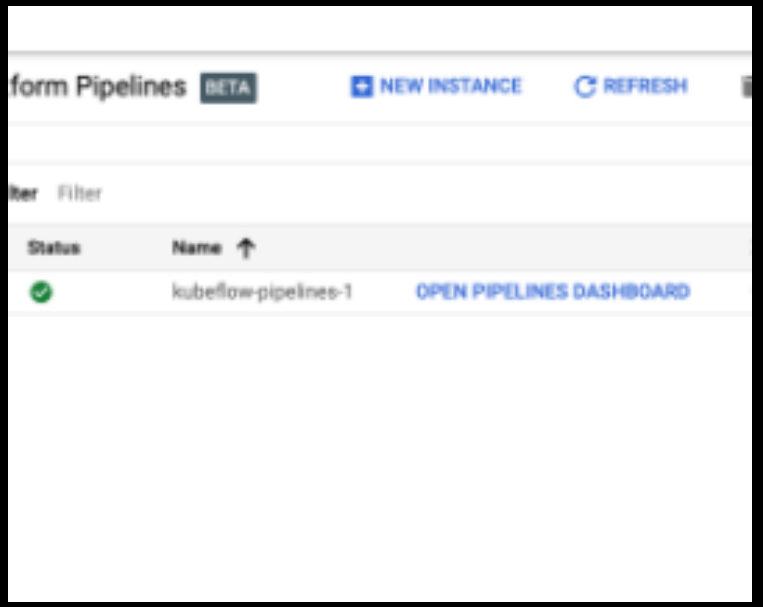
TRAINER



TFX EVALUATOR, PUSHER

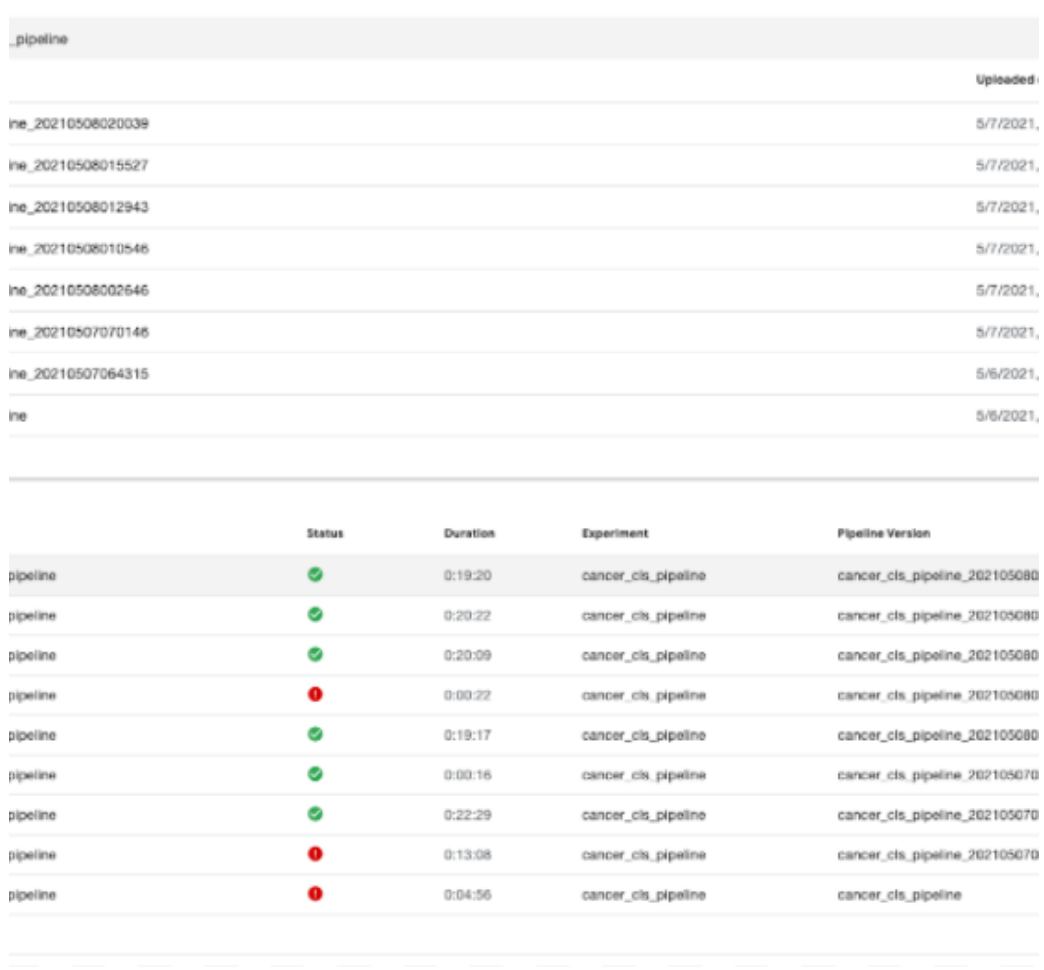
Kubeflow

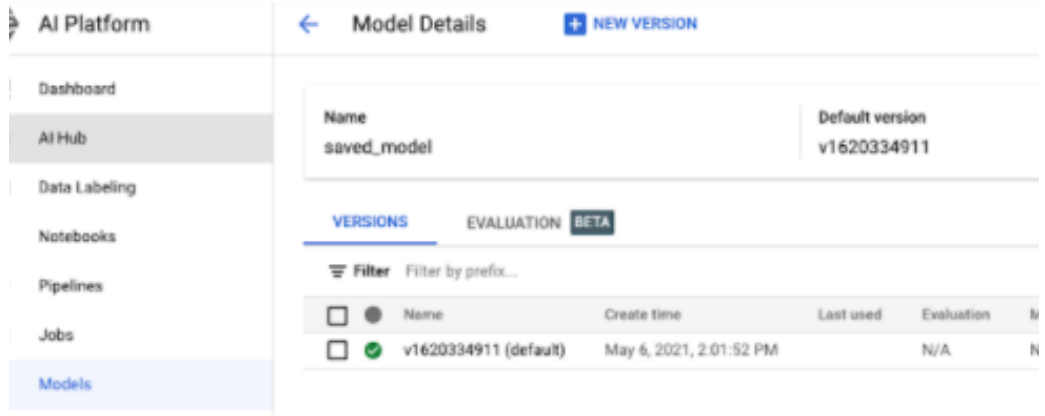
CLOUD PLATFORM

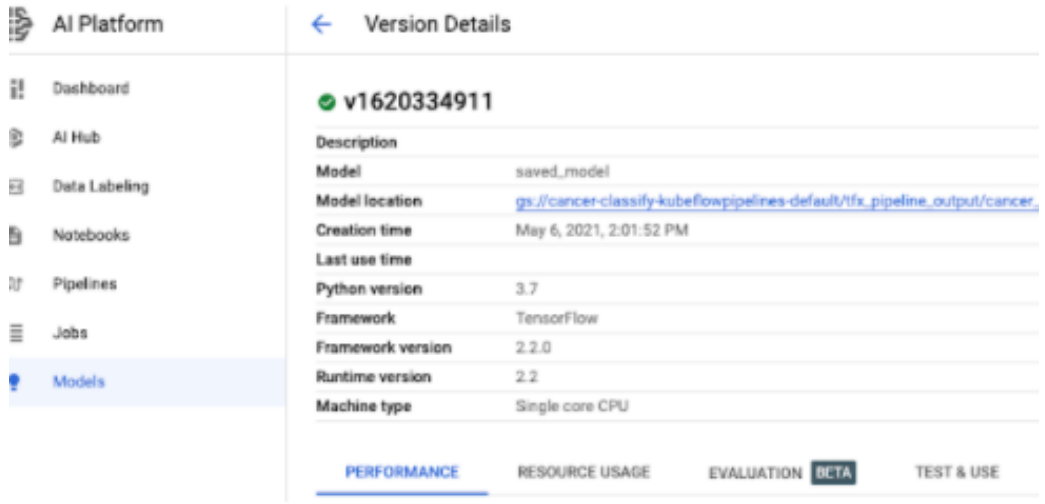


DEPLOYED ON GCP AI PLATFORM

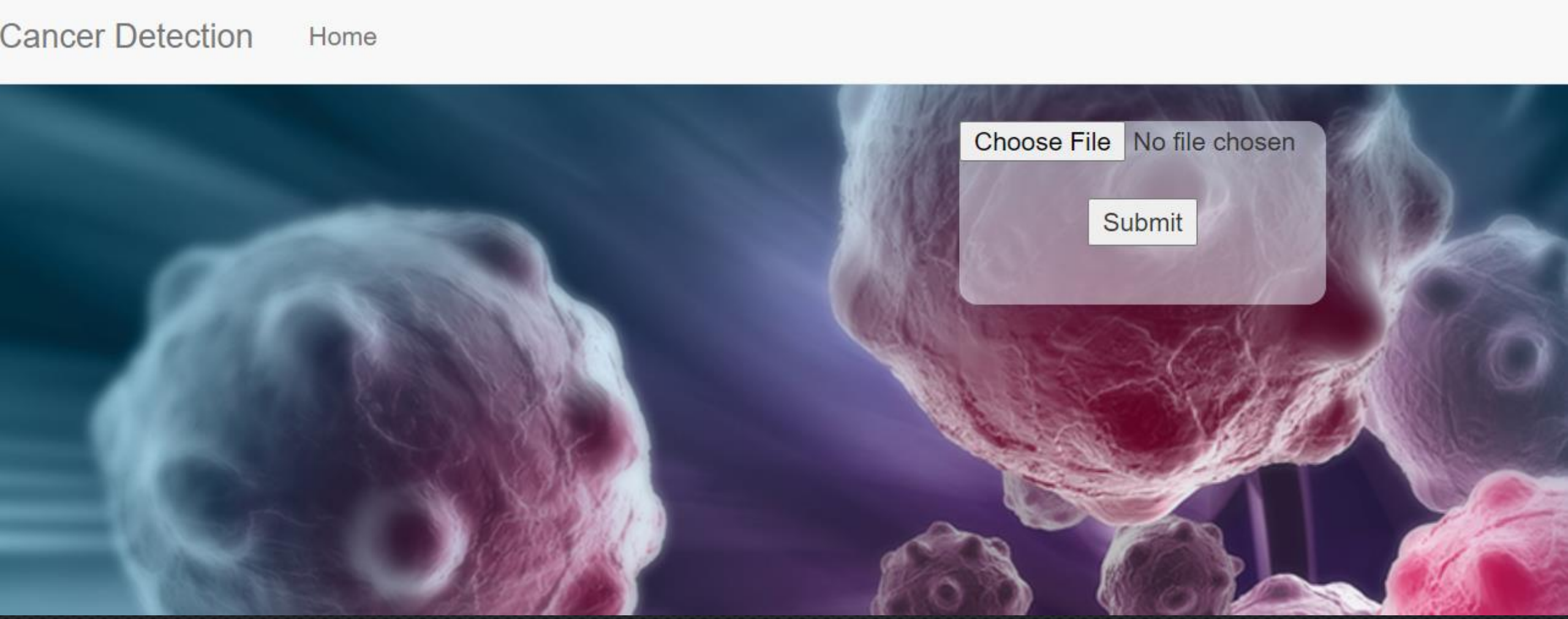
DEMO LINK – TFX, KUBEFLOW







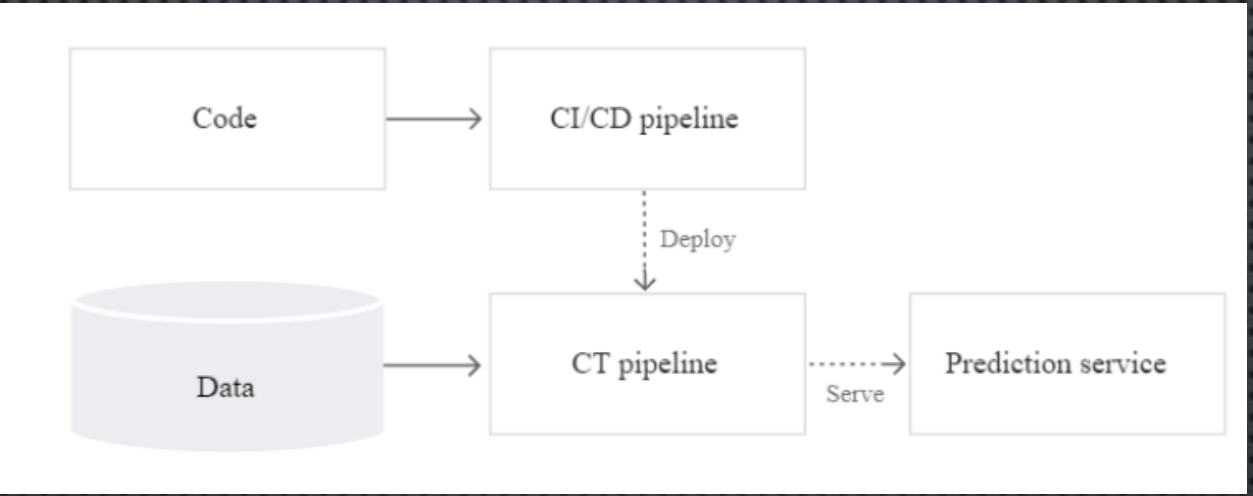
RUNNING & DEPLOYING KUBEFLOW PIPELINES, VERSIONING



FLASK WEB APP HOSTED EC2

MAKES GOOGLEAPI CALLS TO MODEL (LATEST) DEPLOYED ON GCP

DEMO LINK – FLASK EC2



CI/CD PIPELINE VIA GITHUB, TFX, KUBEFLOWS

PIPELINE, MODEL VERSIONING VIA GOOGLE AI PLATFORM

REAL TIME PREDICTION VIA TENSORFLOW SERVING, GOOGLE API & FLASK WEB APP

HOSTED ON EC2

BIG PICTURE

TEAM MEMBER CONTRIBUTIONS

Tasks	Contributor(s)
Project Discussion and Literature Review	All
Data Loading	All
Data Preprocessing (Label balancing + Augmentation)	Abhishek, Subarna
EDA	Abhishek, Subarna
Model Configuration	Abhishek
Visualization	Abhishek, Subarna
Training + Hyperparameter Tuning + Tensorboard	Abhishek
Evaluation (Prediction/Confusion matrix/Classification report/Show top errors)	Subarna

Tasks	Contributor(s)
Bug Fixing (Preprocessing + LSTM)	Wasae
End2End demo and full tfx mlops ci/cd level-2 (Brainstorming)	All
TFX Pipeline	Subarna
Kubeflow	Wasae, Subarna (assisted)
Web App	Wasae
Github Setup + Report + PPT	All

DEEP TRANSFER LEARNING CAN PROVIDE EARLY, ACCURATE DETECTION