

Breast Cancer Diagnosis Using Deep Transfer Learning

CMPE 258 - PROJECT PRESENTATION

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

1 IN 8 WOMEN will be diagnosed with breast cancer in her lifetime.

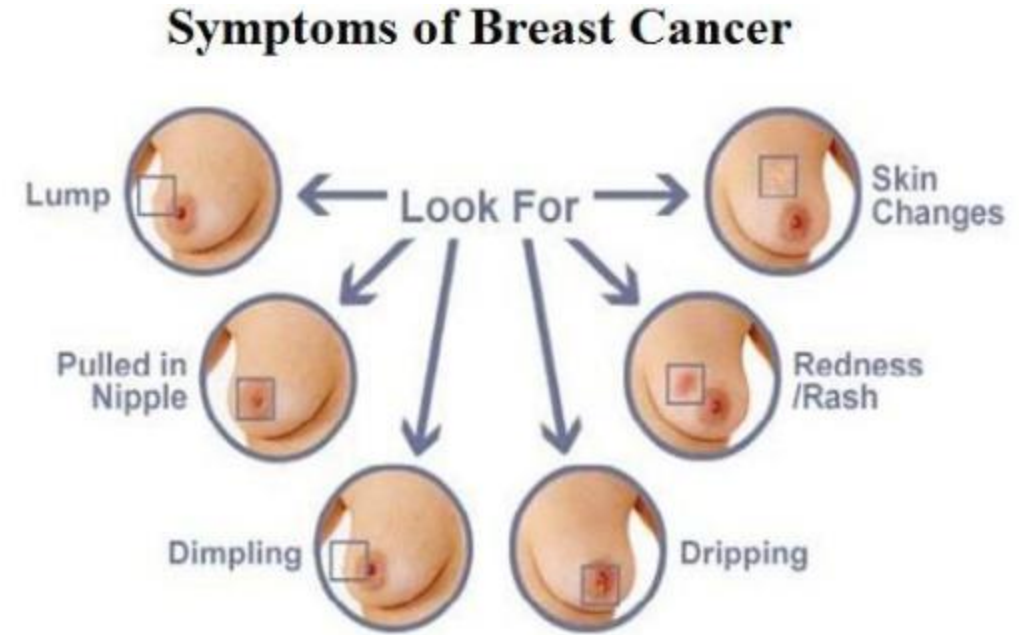
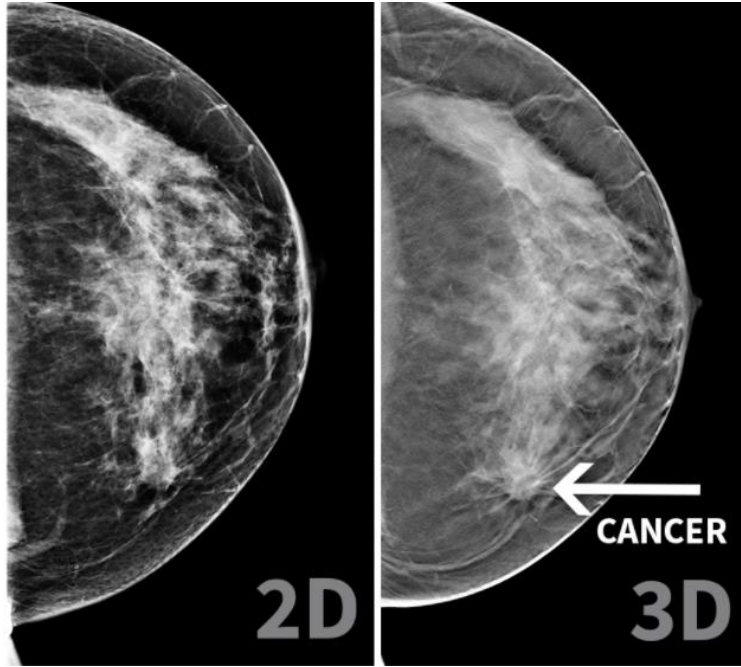


Pic Credits: PA Department of Health

THE AMERICAN CANCER SOCIETY'S ESTIMATES FOR 2021

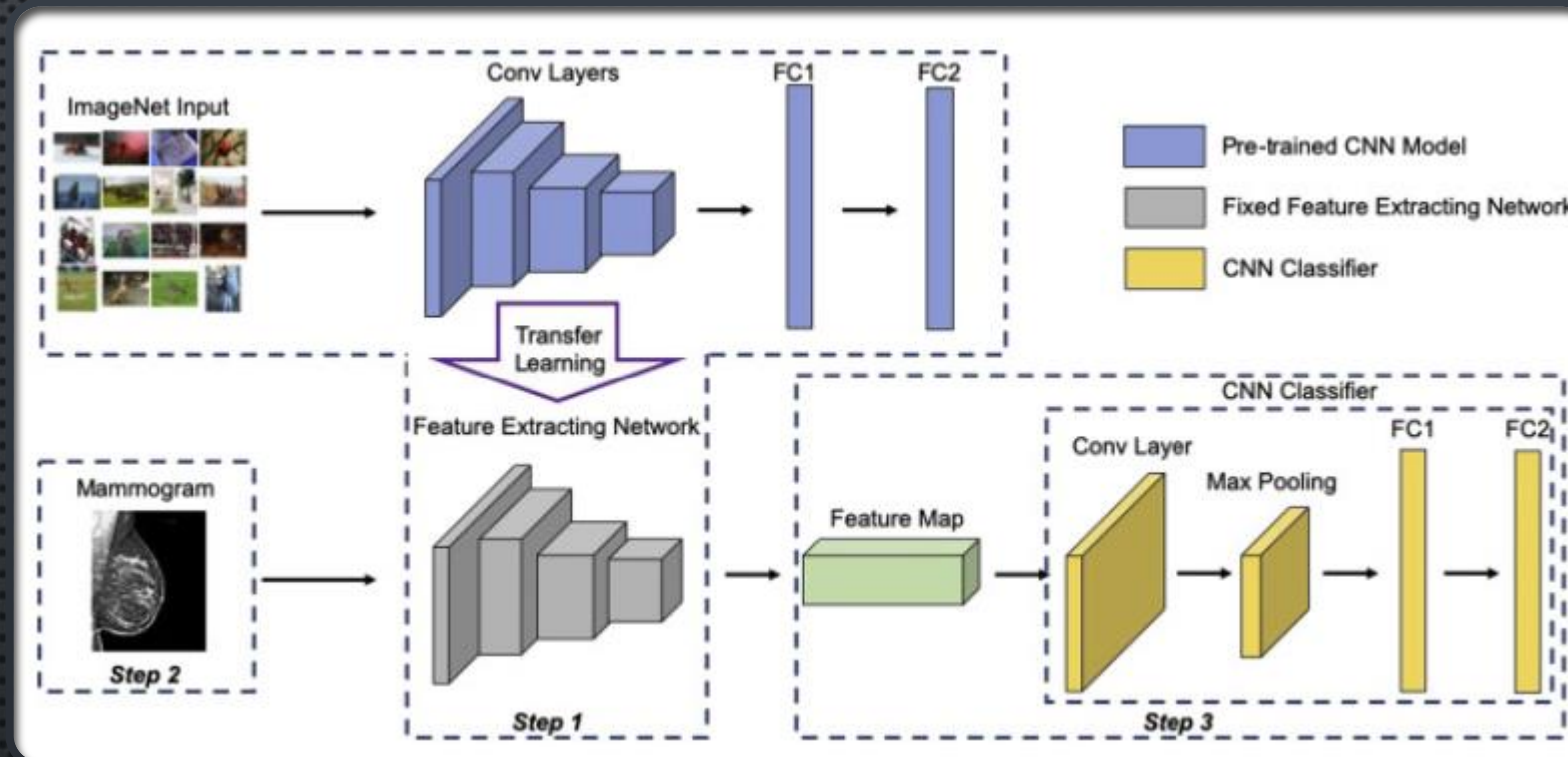
- 281,550 NEW CASES OF INVASIVE BREAST CANCER
- 43,600 WOMEN WILL DIE FROM BREAST CANCER

**EARLY, ACCURATE DETECTION OF
CANCER CAN SAVE LIVES**



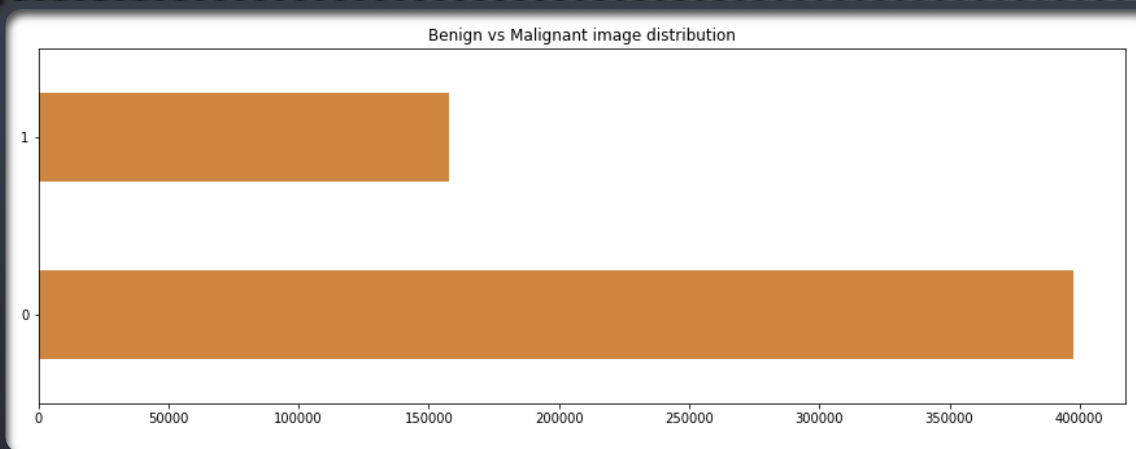
[Pic Credits: A Cancerous Word Press Report](#)

**TRADITIONAL METHODS OF DETECTION SUCH AS
MAMMOGRAPHY REQUIRE EXPERTS**



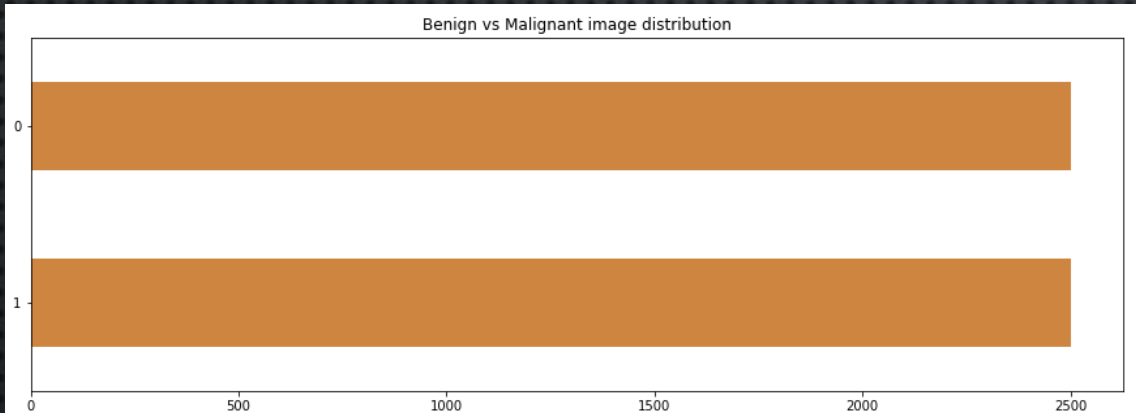
Pic Credits: Journal of American College of Radiology

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



IMBALANCED INPUT

- 397476 BENIGN CANCER IMAGES
- 157572 MALIGNANT CANCER IMAGES

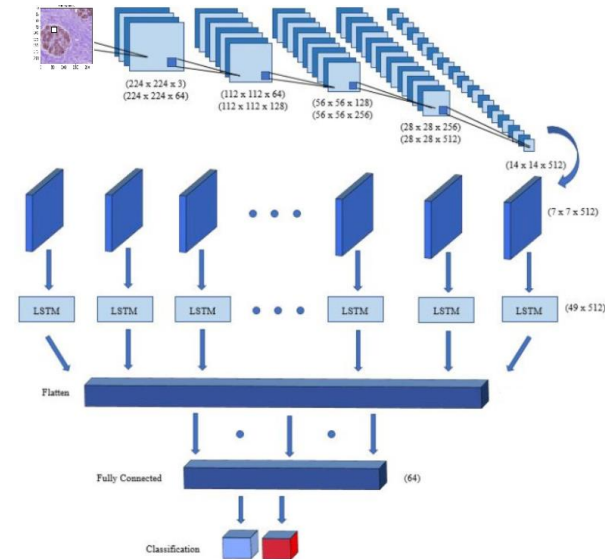
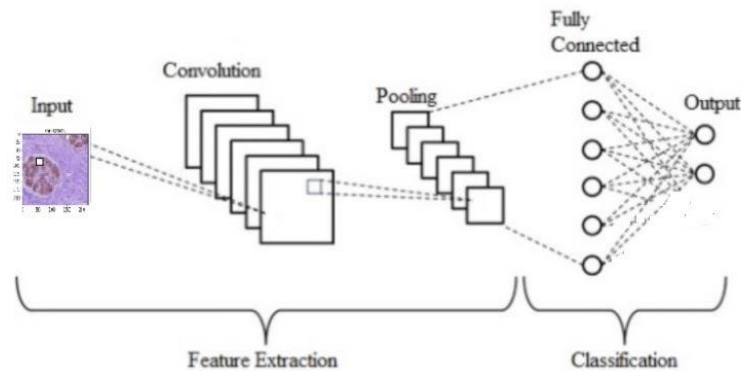


BALANCED FOR STUDY

- 2500 BENIGN CANCER IMAGES
- 2500 MALIGNANT CANCER IMAGES

Breast Cancer Histopathology Images, Case Western Reserve University

DATASET DETAILS

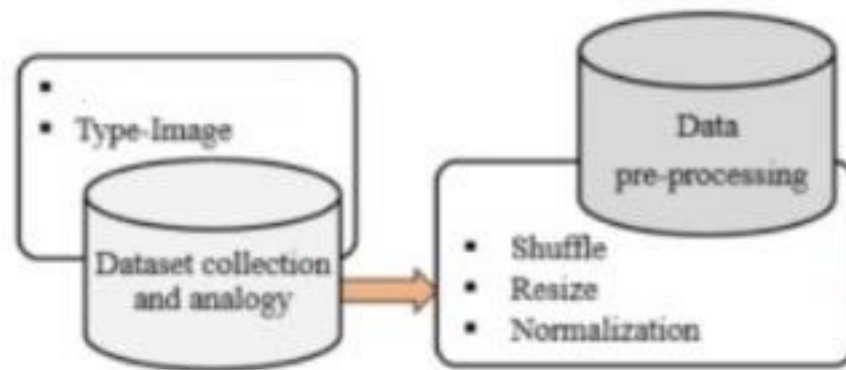


CNN + CUSTOM DNN LAYERS

CNN + LSTM

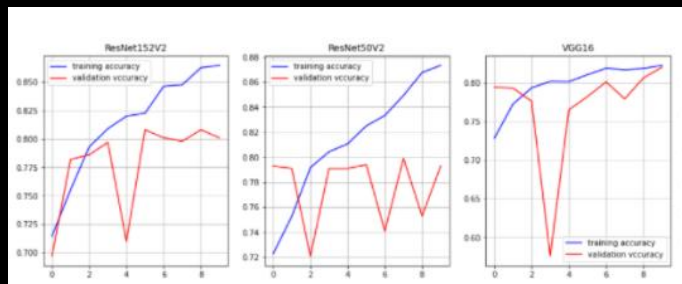
Pic Credits: Md. Zabirul Islam, Md. Milon Islam, Amanullah Asraf, Science Direct

STUDY METHODOLOGY

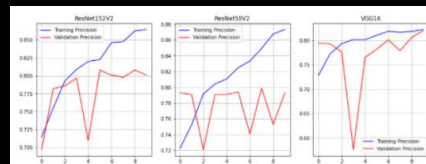


	model_name	model_params	training_accuracy	val_accuracy
2	NASNetMobile	4269716	0.830667	0.753
5	VGG16	14714688	0.822333	0.820
0	DenseNet201	18321984	0.852333	0.818
6	Xception	20861480	0.870667	0.803
4	ResNet50V2	23564800	0.873333	0.793
1	InceptionResNetV2	54336736	0.789667	0.802
3	ResNet152V2	58331648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

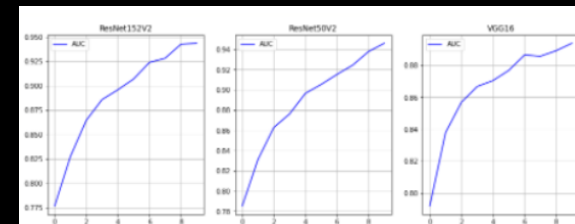
MACHINE LEARNING STAGES AND CNN MODELS



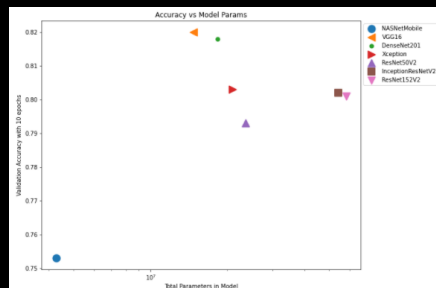
PRECISION



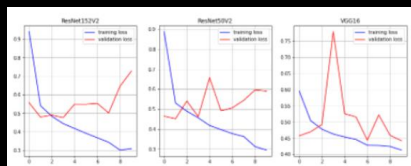
RECALL



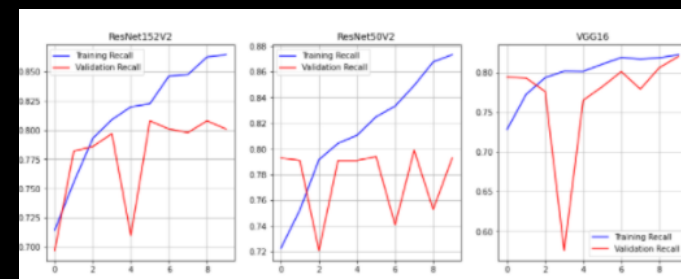
AUC



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

	model_name	model_params	training_accuracy	val_accuracy
0	VGG16	14714688	0.845667	0.8

INCREASE 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

	model_name	model_params	training_accuracy	val_accuracy
0	VGG16	14714688	0.811667	0.798

DROPOUT 0.5

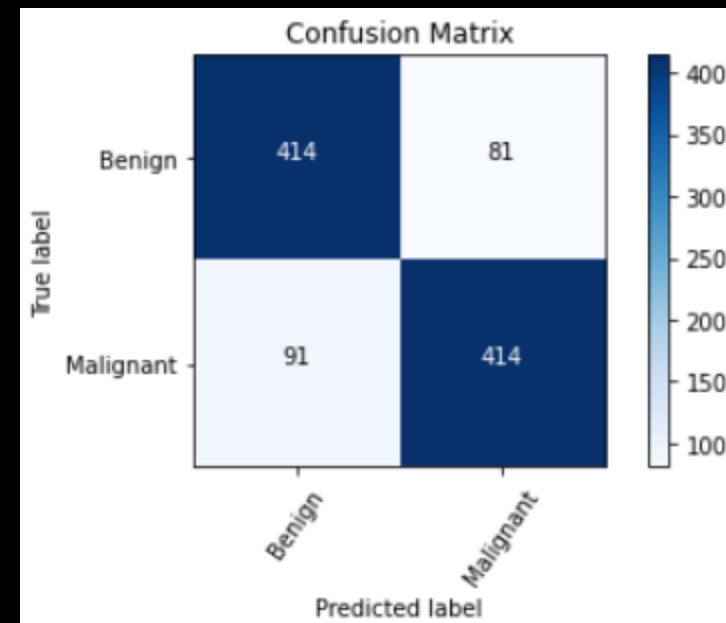
	model_name	model_params	training_accuracy	val_accuracy
0	VGG16	14714688	0.848667	0.826

**UNFREEZE 4 CNN
LAYERS**

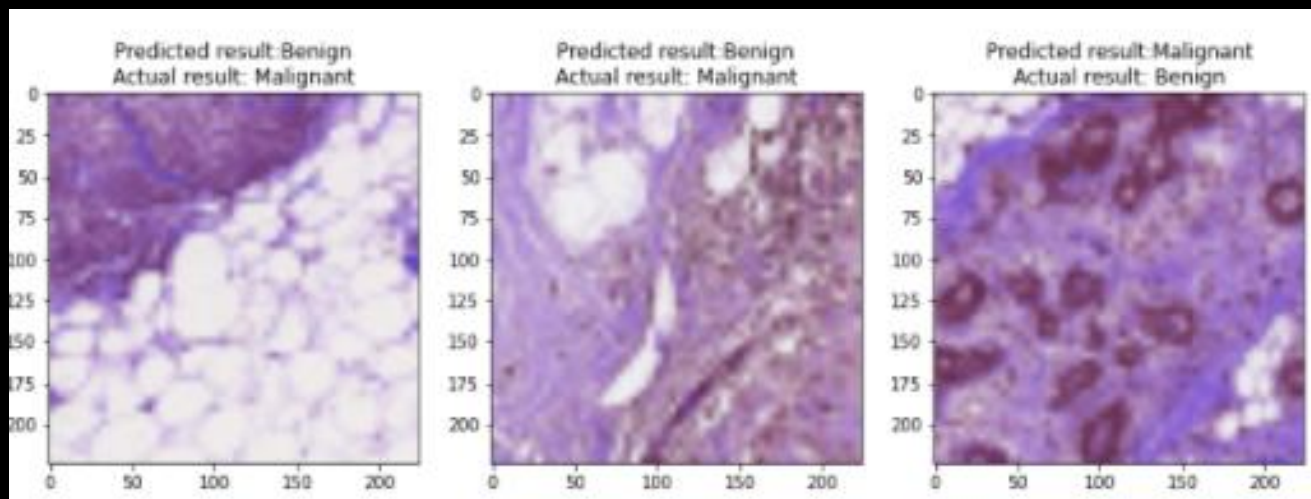
HYPER PARAMETER TUNE VGG16

	model_name	model_params	training_accuracy	val_accuracy
0	VGG16	14714688	0.848667	0.826

ACCURACY



CONFUSION MATRIX

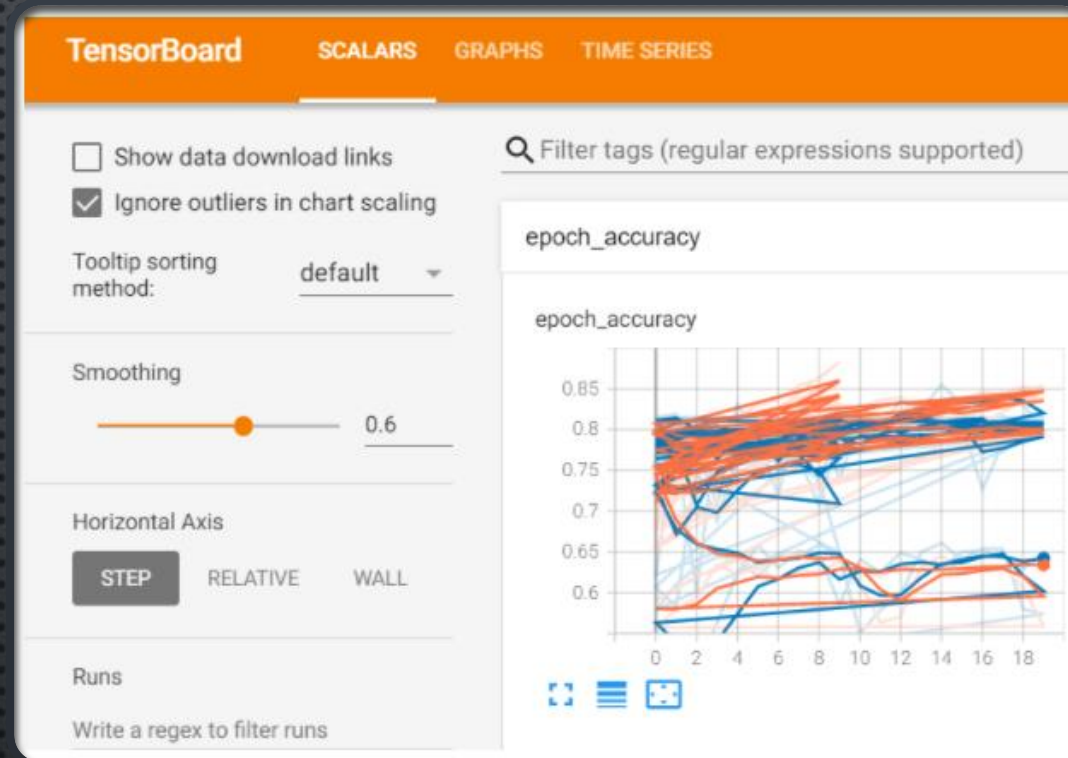


FEW COMMON ERRORS IN PREDICTION

BEST MODEL VGG16

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

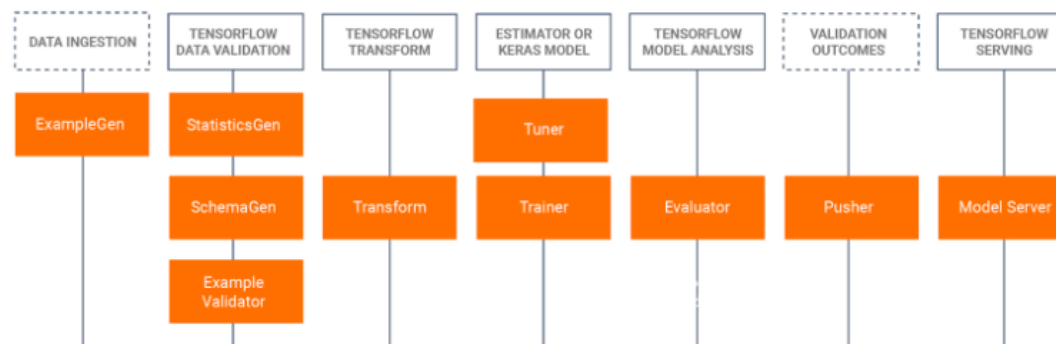


TENSORBOARD INTEGRATION

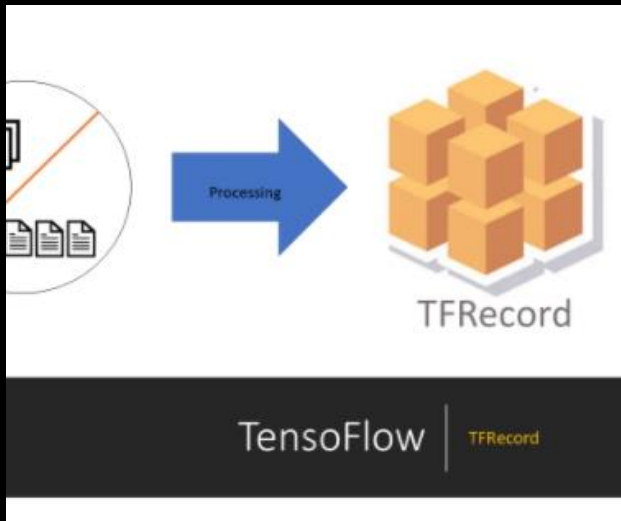
[DEMO LINK](#)



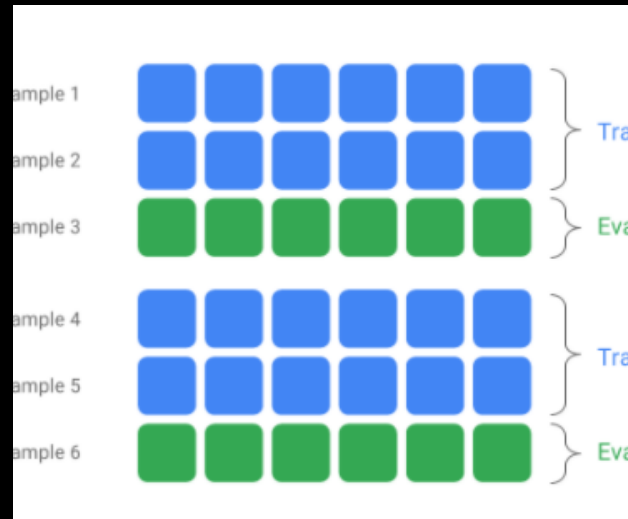
TensorFlow Extended



AUTOMATED PIPELINES WITH TFX



PRE-PROCESS



EXAMPLE GEN



STATISTICS GEN

ain' split:

sr/local/lib/python3.7/dist-pack
pd.set_option('max_colwidth', -

anomalies found.

al' split:

anomalies found.

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

```
data_root = os.path.join(_data_root, 'data')
labels_path = os.path.join(_data_root, 'labels.txt')

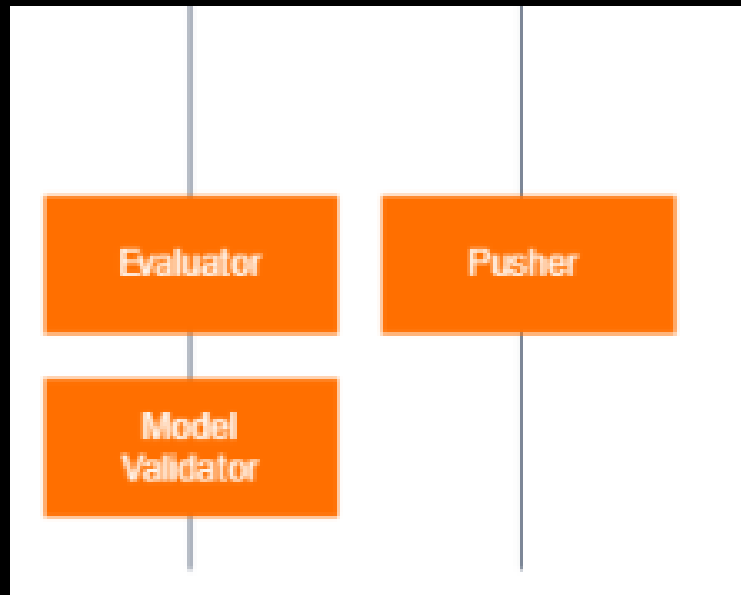
trainer = Trainer(
  module_file=module_file,
  ston_executor_spec=executor_spec.ExecutorClassSpec(GenericExecutor),
  samples=transform.outputs['transformed_examples'],
  schema=schema_gen.outputs['schema'],
  transform_graph=transform.outputs['transform_graph'],
  train_args=trainer_pb2.TrainArgs(num_steps=150),
  eval_args=trainer_pb2.EvalArgs(num_steps=1),
  ston_config={'labels_path': labels_path})

run(trainer)

data from https://storage.googleapis.com/tensorflow/keras-applications/vp
889256 [-----] - 0s 0us/step

-----] - 34s 3s/step - loss: 0.3885 - sparse_categorical
-----] - 7s 2s/step - loss: 0.3509 - sparse_categorical
-----] - 7s 2s/step - loss: 0.3335 - sparse_categorical
```

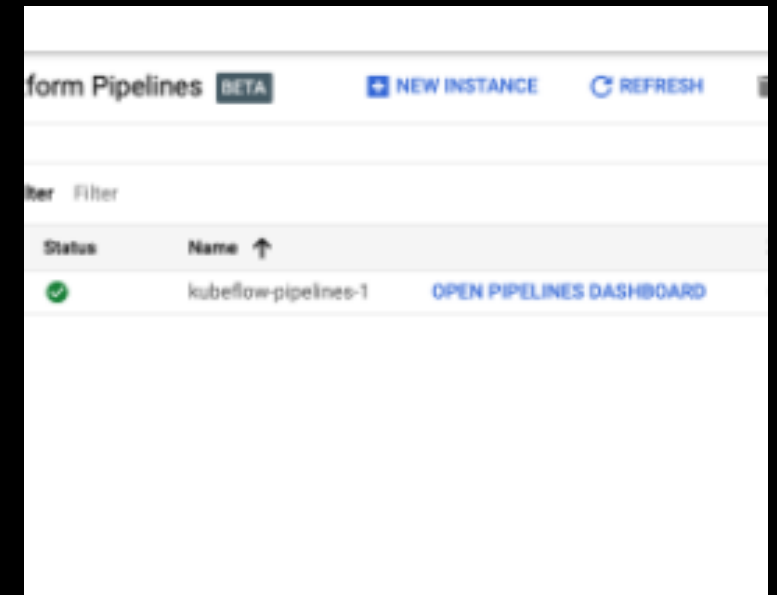
TRAINER



TFX EVALUATOR, PUSHER



CLOUD
PLATFORM



DEPLOYED ON GCP AI
PLATFORM

_pipeline	Uploaded
ne_20210508020039	5/7/2021,
ne_20210508015527	5/7/2021,
ne_20210508012943	5/7/2021,
ne_20210508010546	5/7/2021,
ne_20210508002646	5/7/2021,
ne_20210507070146	5/7/2021,
ne_20210507064315	5/6/2021,
ne	5/6/2021,

	Status	Duration	Experiment	Pipeline Version
pipeline	✓	0:19:20	cancer_cls_pipeline	cancer_cls_pipeline_202105080
pipeline	✓	0:20:22	cancer_cls_pipeline	cancer_cls_pipeline_202105080
pipeline	✓	0:20:09	cancer_cls_pipeline	cancer_cls_pipeline_202105080
pipeline	✗	0:00:22	cancer_cls_pipeline	cancer_cls_pipeline_202105080
pipeline	✓	0:19:17	cancer_cls_pipeline	cancer_cls_pipeline_202105080
pipeline	✓	0:00:16	cancer_cls_pipeline	cancer_cls_pipeline_202105070
pipeline	✓	0:22:29	cancer_cls_pipeline	cancer_cls_pipeline_202105070
pipeline	✗	0:13:08	cancer_cls_pipeline	cancer_cls_pipeline_202105070
pipeline	✗	0:04:56	cancer_cls_pipeline	cancer_cls_pipeline

AI Platform

Dashboard

AI Hub

Data Labeling
Notebooks
Pipelines
Jobs

Models

Model Details

NEW VERSION

Name

saved_model

Default version

v1620334911

VERSIONS

EVALUATION

BETA

Filter Filter by prefix...

<input type="checkbox"/>	<input checked="" type="radio"/>	Name	Create time	Last used	Evaluation	N
<input type="checkbox"/>	<input checked="" type="radio"/>	v1620334911 (default)	May 6, 2021, 2:01:52 PM		N/A	N

AI Platform

Dashboard

AI Hub

Data Labeling
Notebooks
Pipelines
Jobs

Models

Version Details

v1620334911

Description

Model

Model location

Creation time

Last use time

Python version

Framework

Framework version

Runtime version

Machine type

PERFORMANCE

RESOURCE USAGE

EVALUATION

BETA

TEST & USE

RUNNING & DEPLOYING KUBEFLOW PIPELINES, VERSIONING

Choose File

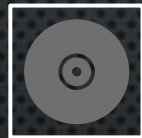
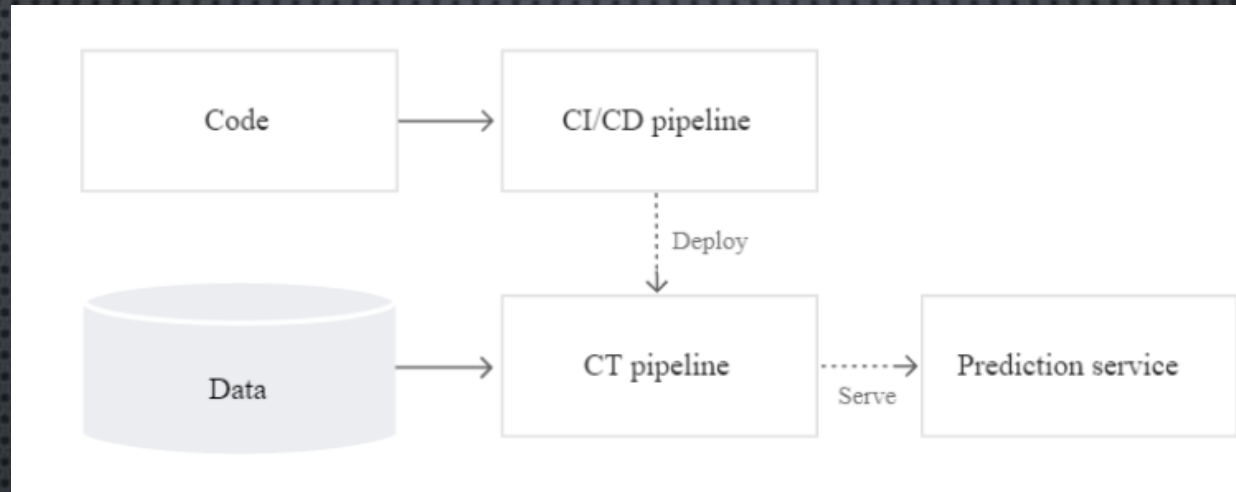
No file chosen

Submit

FLASK WEB APP HOSTED EC2

MAKES GOOGLEAPI CALLS TO MODEL (LATEST) DEPLOYED ON GCP

[DEMO LINK](#)



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

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

TEAM MEMBER CONTRIBUTIONS



**DEEP TRANSFER LEARNING CAN PROVIDE
EARLY, ACCURATE DETECTION**