

## GOKARAJU RANGARAJU INSTITUTE OF ENGINEERING AND TECHNOLOGY Department of Computer Science and Engineering Major Project Phase-II Review-2

### Smart Platform for Breast Cancer Classification using Deep Learning Techniques

- Guide: G.Mallikarjuna Rao, Professor

M.BHAVITA-20241A0591 M.SAHITHI-20241A0595 N.VARSHA-20241A0599 R.NIKITHA-20241A05B1

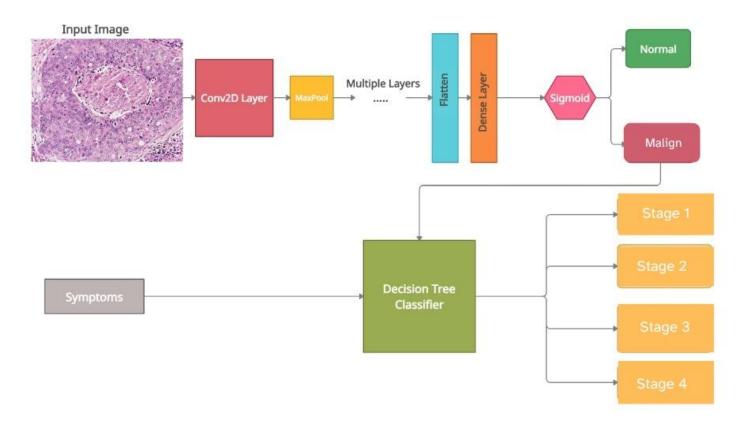


#### **Proposed Architecture:**

In our breast cancer classification project, we will utilize medical image dataset to identify intricate patterns. These datasets will be fed into various classifiers to discern whether a given instance is benign or malignant by using deep learning cnn models. Upon identifying patterns indicative of cancer, we will further classify the stage of breast cancer by using symptoms and classifying them with DecisonTreeClassifier on a symptoms dataset. Additionally, we will provide specific precautions and recommend medicines to mitigate the condition. To facilitate easy access and understanding, we will develop a user interface for online checking. However, it's important to note that this interface is meant for informational purposes only, and for optimal results, individuals should consult trained medical professionals.



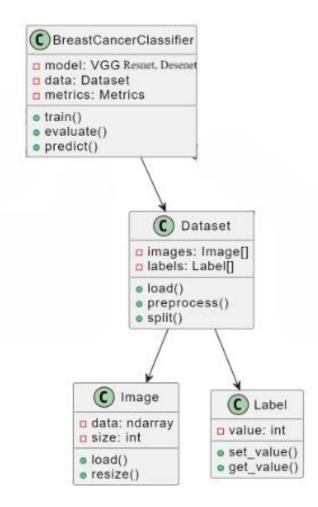
### **Proposed Architecture:**





### **Module 1: Pre-Processing**

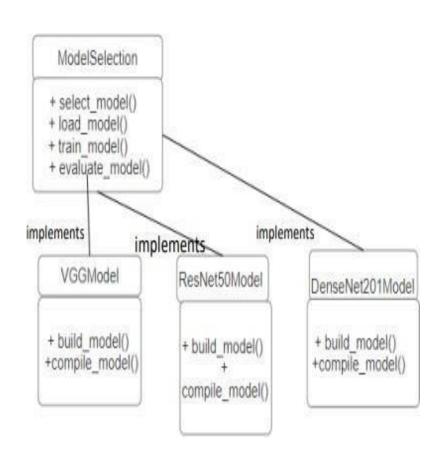
We will start by gathering the necessary data for o project. After acquiring the data, we will perform several essential data preprocessing tasl including resizing, color conversion, normalization, labeling, and overall data preparatic This crucial step is vital to ensure that the data is appropriately formatted and ready for o analytical procedures.





#### **Module 2: Model Selection**

- DenseNet201 may be a good choice if you have a large dataset and need a highly accurate model. It excels at capturing intricate image features
- ResNet50 is a solid choice for most image classification tasks. It strikes a balance between model complexity and performance and can work well with medium-sized datasets.
- VGG models, such as VGG16 or VGG19, are simpler and have fewer parameters compared to DenseNet201 and ResNet50. They may be suitable for smaller datasets or when computational resources are limited.
- DenseNet-201 achieves high accuracy while maintaining parameter efficiency, making it suitable for resource-constrained environments.

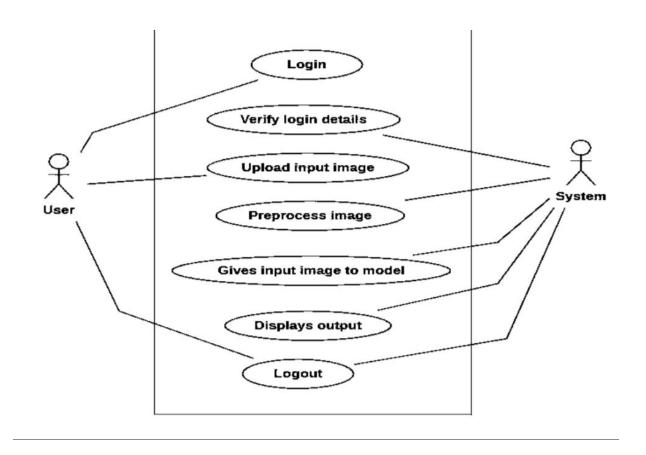




#### **Module 3: Testing and User Interface**

The data is ready for model training and evaluation. The training process typically involves monitoring performance on the validation set to optimize the model. Finally, the model's generalization is assessed by evaluating it on the shuffled test set (X\_test and Y\_test) to determine how well it performs on unseen data.

Developing a platform where individuals can upload breast cancer images for classification. A website where users can upload there breast cancer image and result of it( cancerous or not and the stage) can be directly seen on screen.





### **Experimentation:**

importing libraries

Dividing into training and testing Data:

```
JUDYTET Densenet201 Last Checkpoint: 3 hours ago
      View Run Kernel Settings Help
                                                                                                                           JupyterLab ☐ # Python 3 (ipykernel) ☐
[1]: import json
      import math
      from PIL import Image
      from keras.applications import ResNet50, MobileNet, DenseNet201, InceptionV3, NASNetLarge, InceptionResNetV2, NASNetMobile
      from keras.callbacks import Callback, ModelCheckpoint, ReduceLROnPlateau, TensorBoard
      from keras.callbacks import ReduceLROnPlateau, ModelCheckpoint
      from keras.preprocessing.image import ImageDataGenerator
      from tensorflow.keras.utils import to categorical
      from keras.models import Sequential
      from keras.optimizers import Adam
      import matplotlib.pyplot as plt
      import pandas as pd
      from sklearn.model_selection import train_test_split
      from sklearn.metrics import cohen_kappa_score, accuracy_score
      import scipy
      from tqdm import tqdm
      import tensorflow as tf
      from keras import backend as K
      from functools import partial
      from sklearn import metrics
      from collections import Counter
      import json
      import itertools
```

```
def Dataset_loader(DIR, RESIZE, sigmaX=10):
    read = lambda imname: np.asarray(Image.open(imname).convert("RGB"))
    for IMAGE_NAME in tqdm(os.listdir(DIR)):
        PATH = os.path.join(DIR, IMAGE NAME)
        _, ftype = os.path.splitext(PATH)
        if ftype == ".png":
            img = read(PATH)
            img = cv2.resize(img, (RESIZE,RESIZE))
            img = cv2.GaussianBlur(img,(5,5),cv2.BORDER_DEFAULT)
            arr=np.array(img)
            IMG.append(arr)
    return IMG
benign_train = np.array(Dataset_loader("C:\\Users\\Varsha\\Desktop\\DataSources\\BreaKHis 400X\\train\\benign", 224))
malign_train = np.array(Dataset_loader("C:\\Users\\Varsha\\Desktop\\DataSources\\BreaKHis 400X\\train\\malignant", 224))
benign_test = np.array(Dataset_loader("C:\\Users\\Varsha\\Desktop\\DataSources\\BreaKHis 400X\\test\\benign", 224))
malign_test = np.array(Dataset_loader("C:\\Users\\Varsha\\Desktop\\DataSources\\BreaKHis 400X\\test\\malignant", 224))
                                                                                        371/371 [00:09<00:00, 38.18it/s]
                                                                                        777/777 [00:20<00:00, 37.57it/s]
                                                                                        176/176 [00:04<00:00, 39.06it/s]
                                                                                        369/369 [00:08<00:00, 42.95it/s]
```



#### Building densenet Model:

```
MajorProj.ipynb
  File Edit View Insert Runtime Tools Help Last edited on September 21
+ Code + Text
                                                                         Downloading data from https://storage.googleapis.com/tensorflow/keras-applications/de
       def build_model(backbone, lr=1e-4):
                                                                         f_kernels_notop.h5
            model = Sequential()
                                                                         model.add(backbone)
                                                                         Model: 'sequential'
           model.add(layers.GlobalAveragePooling2D())
           model.add(layers.Dropout(0.5))
                                                                         Laver (type)
                                                                                             Output Shape
                                                                                                              Paran #
           model.add(layers.BatchNormalization())
           model.add(layers.Dense(2, activation='softmax'))
                                                                         densenet201 (Functional)
                                                                                            (None, 7, 7, 1928)
                                                                                                              18321984
           model.compile(
                                                                          global_average_pooling2d (G (None, 1920)
                loss='binary_crossentropy',
                                                                          lobalAveragePooling2D)
                optimizer=Adam(lr=lr),
                metrics=['accuracy']
                                                                          propout (Dropout)
                                                                                             (None, 1928)
                                                                         batch_normalization (BatchN (None, 1920)
                                                                                                              7689
            return model
                                                                          ormalization)
       K.clear_session()
       gc.collect()
                                                                          dense (Dense)
                                                                                             (None, 2)
                                                                                                              3842
       resnet = DenseNet201(
           weights='imagenet'.
           include_top=False,
                                                                         Total params: 18,333,506
           input_shape=(224,224,3)
                                                                         Trainable params: 18,100,610
                                                                         Non-trainable params: 232,896
```



#### Building Resnet and VGG16 Model:

```
return model
[ ] K.clear session()
    gc.collect()
    pretrained resnet base = tf.keras.applications.resnet v2.ResNet50V2(
        include_top=False,
        input shape=(224, 224, 3),
        weights="imagenet"
    pretrained resnet base.trainable = False
[ ] learn control = ReduceLROnPlateau(monitor='val acc', patience=5,
                                      verbose=1, factor=0.2, min lr=1e-7)
    filepath="weights.best.hdf5"
    checkpoint = ModelCheckpoint(filepath, monitor='val acc', verbose=1, save best only=True, mode='max')
pretrained resnet base = tf.keras.applications.resnet v2.ResNet50V2(
         include_top=False,
        input_shape=(224, 224, 3),
        weights="imagenet"
    pretrained_resnet_base.trainable = False
[ ] resnet50_model = build_model(pretrained_resnet_base ,lr = 1e-4)
    resnet50 model.summary()
```

```
JUDYTER VGG16 Last Checkpoint: last month
File Edit View Run Kernel Settings Help
   + % □ □ ▶ ■ C → Code ∨
                                                                                                                           JupyterLab ☐ # Python 3 (ipykernel
             return model
   [9]: K.clear session()
         gc.collect()
         pretrained vgg base = tf.keras.applications.VGG16(
             include top=False,
             input_shape=(224, 224, 3),
             weights="imagenet"
         pretrained_vgg_base.trainable = False
         WARNING:tensorflow:From C:\Users\Varsha\AppData\Local\Programs\Python\Python311\Lib\site-packages\keras\src\backend.py:277: The name tf.reset default gr
         aph is deprecated. Please use tf.compat.v1.reset_default_graph instead.
         WARNING:tensorflow:From C:\Users\Varsha\AppData\Local\Programs\Python\Python311\Lib\site-packages\keras\src\layers\pooling\max pooling2d.py:161: The nam
         e tf.nn.max_pool is deprecated. Please use tf.nn.max_pool2d instead.
         Downloading data from https://storage.googleapis.com/tensorflow/keras-applications/vgg16/vgg16_weights_tf_dim_ordering_tf_kernels_notop.h5
         58889256/58889256 [==========] - 19s @us/step
   [10]: learn_control = ReduceLROnPlateau(monitor='val_acc', patience=5,
                                          verbose=1,factor=0.2, min lr=1e-7)
         filepath="weights.best.hdf5"
         checkpoint = ModelCheckpoint(filepath, monitor='val_acc', verbose=1, save_best_only=True, mode='max')
   [11]: pretrained_vgg_base = tf.keras.applications.VGG16(
             include top=False,
             input_shape=(224, 224, 3),
             weights="imagenet"
         pretrained_vgg_base.trainable = False
   [12]: vgg_model = build_model(pretrained_vgg_base, lr=1e-4)
         vgg model.summary()
```



#### **DecisionTreeClassifier:**

```
import pandas as pd
df = pd.read excel('C:\\Users\\Varsha\\Desktop\\symptoms.xlsx')
print (df)
def changeYesNo(string):
    for d in df.index:
        if df.loc[d, string] == 'Yes':
            df.loc[d, string] = 1
            df.loc[d, string] = 0
changeYesNo('SmallLump')
changeYesNo('ChangesinSizeShape')
changeYesNo('SkinChanges')
changeYesNo('NippleChanges')
changeYesNo('LumpinUnderarm')
changeYesNo('Swelling')
changeYesNo('LargeLumpMass')
changeYesNo('BonePain')
changeYesNo('Cough')
changeYesNo('AbdominalPain')
changeYesNo('Headaches')
changeYesNo('WeightLossFatigue')
print(df)
feature_cols = ['SmallLump', 'ChangesinSizeShape', 'SkinChanges', 'NippleChanges',
                 'LumpinUnderarm', 'Swelling', 'LargeLumpMass', 'BonePain',
                 'Cough', 'AbdominalPain', 'Headaches', 'WeightLossFatigue']
X = df[feature cols]
Y = df['Stage']
```

```
X = df[feature cols]
Y = df['Stage']
from sklearn.model selection import train test split
X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.2,
 shuffle=False)
import numpy as np
import pandas as pd
from sklearn.metrics import confusion matrix
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification report
from joblib import dump, load
clf = DecisionTreeClassifier()
clf = clf.fit(X train,y train)
y_pred = clf.predict(X_test)
print("Expected Output Predicted Output")
for d in range(len(y test)):
    print(y_test.iloc[d], y_pred[d])
from joblib import dump, load
dump(clf, 'decisiontree.joblib')
clf = load('decisiontree.joblib')
```



#### **User Interface:**

SmallLump = changeYesNo(SmallLump)

```
Edit Selection View Go Run ···

∠ Search

app = Flask( name )
                                                                     app.py X
def changeYesNo(s):
                                                                     > Users > Varsha > Desktop > flask > 🕏 app.py > ..
    if s == 'Yes':
                                                                    55 def predict():
        return 1
    else:
                                                                             arr = np.array([[SmallLump, ChangesinSizeShape, SkinChanges, NippleChanges, LumpinUnderarm, Swelling, LargeLumpMass, BonePain, Cough, A
                                                                             print(arr)
        return 0
                                                                             clf = load('decisiontree.joblib')
@app.route('/', methods=['GET'])
                                                                             stage = clf.predict(arr)
def hello world():
                                                                             file = request.files['file1']
                                                                             file.save('static/file1.png') # Save as PNG
    return render template('index.html')
                                                                             model = load model("trained.h5")
                                              Loading...
                                                                             img path = 'static/file1.png' # File path with correct extension
@app.route('/predict', methods=['GET', 'POST'])
                                                                             img = cv2.imread(img_path)
                                                                             tempimg = image.load img(img path, target size=(224, 224))
def predict():
                                                                             img_array = image.img_to_array(tempimg)
    SmallLump = request.form['SmallLump']
                                                                             img_array = preprocess_input(img_array)
    ChangesinSizeShape = request.form['ChangesinSizeShape']
                                                                             img = np.expand dims(img array, axis=0)
    SkinChanges = request.form['SkinChanges']
    NippleChanges = request.form['NippleChanges']
                                                                             # Make the prediction
                                                                             prediction = model.predict(img)
    LumpinUnderarm = request.form['LumpinUnderarm']
    Swelling = request.form['Swelling']
                                                                             # Check if the maximum predicted probability is greater than or equal to 0.5
    LargeLumpMass = request.form['LargeLumpMass']
                                                                             if np.max(prediction) >= 0.5:
    BonePain = request.form['BonePain']
                                                                                prediction label = "Malignant"
                                                                             else:
    Cough = request.form['Cough']
                                                                                prediction_label = "Benign"
    AbdominalPain = request.form['AbdominalPain']
    Headaches = request.form['Headaches']
                                                                             classification = 'Prediction : %s' % (prediction label)
    WeightLossFatigue = request.form['WeightLossFatigue']
```



#### **User Interface:**

```
∠ Search

                                                                                                                                Edit Selection View Go Run ...
C: > Users > Varsha > Desktop > flask > ♠ app.py > ♠ predict
      def predict():
          if prediction_label == "Malignant":
              phase = ''
              for ele in stage:
                  phase += ele
              classification = "Prediction: Malignant and Stage: %s" % (phase)
              # Add recommendations based on cancer stage
              recommendations = get_recommendations_based_on_stage(phase)
              classification += "\n\nRecommendations:\n%s" % (recommendations)
119
          return render_template("/predict.html", prediction_label=classification)
      def get recommendations based on stage(stage):
          recommendations = {
               'Stage 1': 'Surgery, such as lumpectomy or mastectomy, often accompanied by radiation therapy; additional treatments like hormone
               'Stage 2': 'Combined treatment approach involving surgery, chemotherapy, radiation, and hormone therapy; specifics depend on tumor
               'Stage 3': 'Combined treatment approach involving surgery, chemotherapy, radiation, and hormone therapy; specifics depend on tumor
               'Stage 4': 'Metastatic or advanced breast cancer management with systemic therapies like chemotherapy, hormone therapy, targeted the
          return recommendations.get(stage, 'No specific recommendations available for this stage.')
      if __name__ == "__main__":
          app.run(debug=True)
```

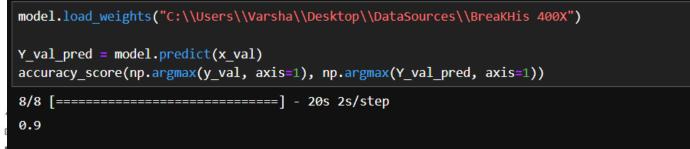


7.5

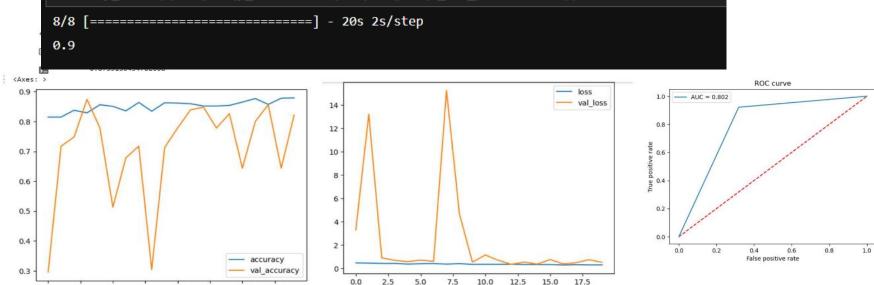
10.0

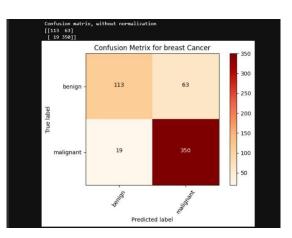
### GOKARAJU RANGARAJU INSTITUTE OF ENGINEERING AND TECHNOLOGY Department of Computer Science and Engineering

#### **Results and Discussions:**

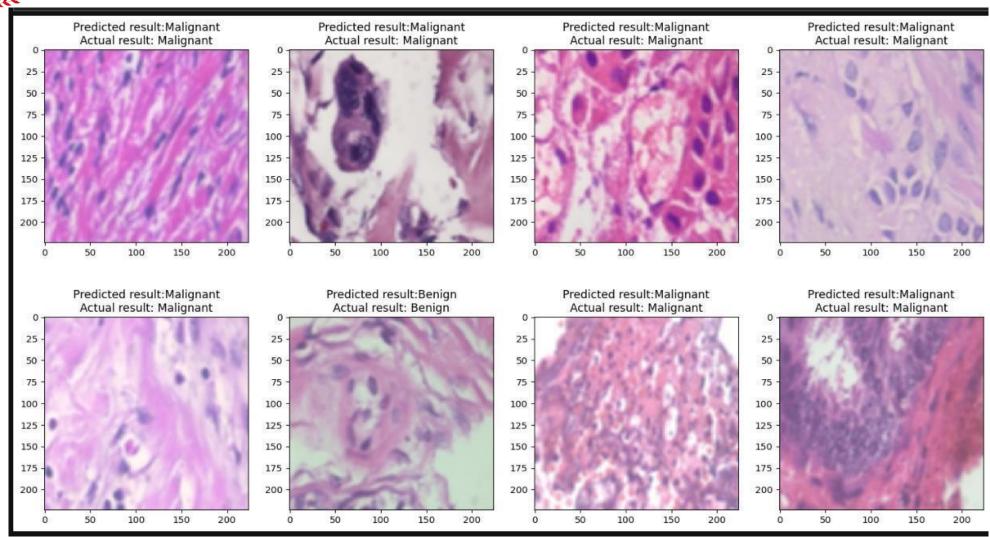


Densenet201:



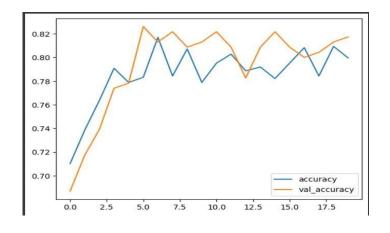






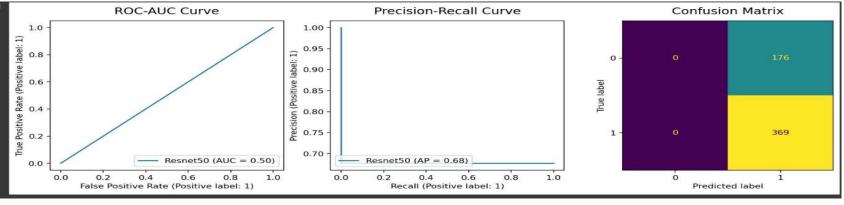


#### Resnet50



# 0.65 - loss val\_loss val\_loss val\_loss val\_loss val\_loss val\_loss

#### **Accuracy:**



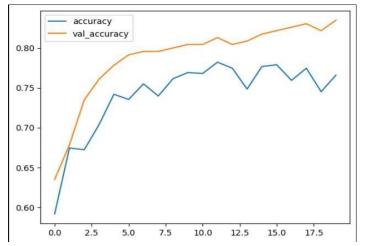


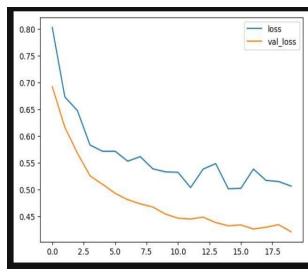
#### vgg16

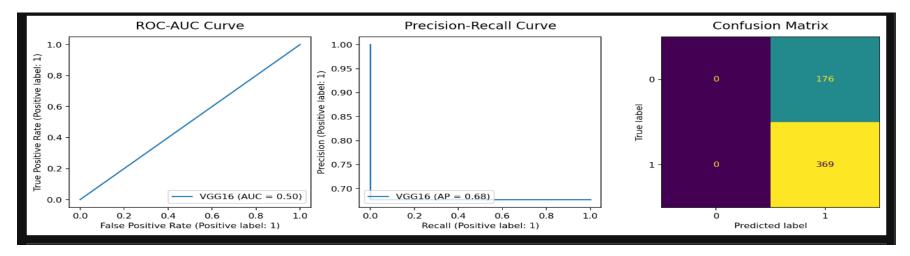
#### Accuracy:

```
Y_pred = vgg_model.predict(X_test)
accuracy_score(np.argmax(Y_test, axis=1), np.argmax(Y_pred, axis=1))

18/18 [========] - 77s 4s/step
: 0.8330275229357799
```









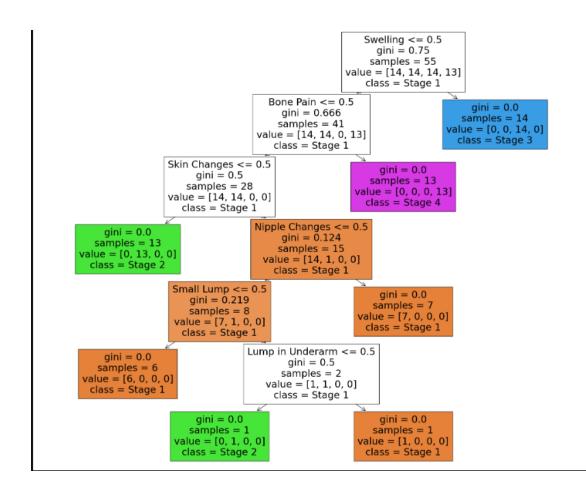
#### Results on diverse Datasets:

Type	Densenet201	Resnet50	vgg16
Mammography	1/1 [===================================	1/1 [===================================	VGG16 Accuracy: 0.5714285969734192
MRI	9/9 [======= Validation Accuracy: 0.525	ResNet50 Accuracy: 0.4250000	VGG16 Accuracy: 0.487500



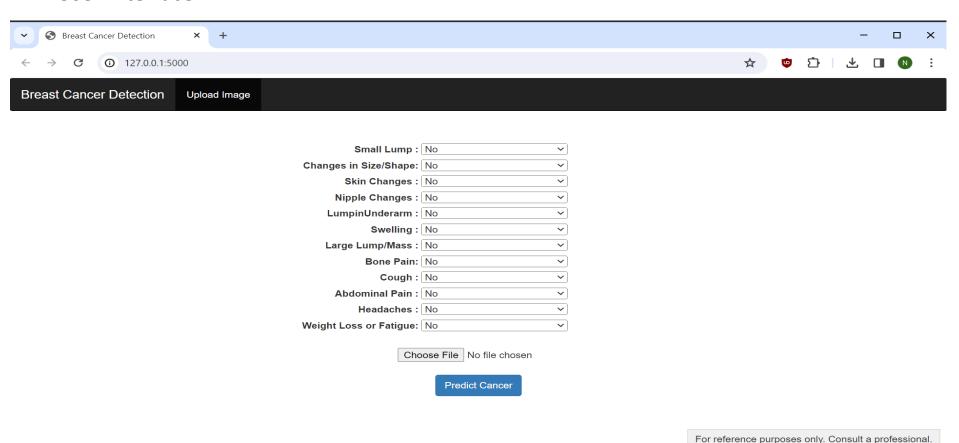
#### Results for DecisonTreeClassifier:

Ļ						
Expected Outpu	ut   Predicted Output					
+	+					
Stage 4	Stage 4					
Stage 3	Stage 3					
Stage 1	Stage 1					
Stage 2	Stage 2					
Stage 4	Stage 4					
Stage 3	Stage 3					
Stage 1	Stage 1					
Stage 2	Stage 2					
Stage 4	Stage 4					
Stage 3	Stage 3					
Stage 1	Stage 1					
Stage 2	Stage 2					
Stage 4	Stage 4					
Stage 3	Stage 3					
+	+					
Accuracy: 1.0						





#### User Interface:





### **Discussion**

#### **Integration of Heterogeneous Data Sets:**

Heterogeneous data sources, including mammograms and mri images, were integrated by standardizing formats, addressing missing values, and normalizing data for consistency. Feature extraction techniques such as texture analysis and edge detection were employed across modalities, followed by fusion methods to create a unified representation. Challenges included aligning disparate data formats, selecting relevant features, and ensuring consistent normalization.

#### **Stage Detection and Classification:**

Stages 0-IV indicate the severity of cancer progression based on symptoms, crucial for determining appropriate interventions like surgery, chemotherapy, or radiation therapy. Understanding cancer stages aids in prognosis and treatment efficacy assessment.

#### **User Interface Development:**

The user-friendly interface allowed users to upload mammograms and ultrasound images, select classification models, and visualize results including predicted cancer stages and feature importance.

#### **Patient Education:**

Integrated educational resources provided patients with information on cancer stages, treatment options, and preventive measures, empowering informed decision-making for their healthcare journey.



	Test Cases	Expected Result	Actual Result	Test Result (Pass/Fail)
1	Check for Breast Cancer using Malign Dataset in Model	Presence of Breast Cancer	Breast Cancer is present	Pass
2	Check for Breast Cancer using Benign dataset in Model	Breast Cancer should not be present	Breast cancer is not present	Pass
3	Check for Breast Cancer by uploading in UI (1st Malign image)	Breast Cancer should be present And correct stage is given	Breast cancer is present and correct stage is given acc to symptoms	Pass
4	Check for Breast Cancer by uploading in UI (2nd Malign image)	Breast Cancer should be present And stage is given	Breast cancer is present and correct stage is given acc to symptoms	Pass



Testcases 1 &2:

In CNN MODEL

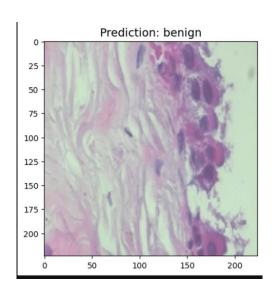


Fig. Sample Output of Normal Breast

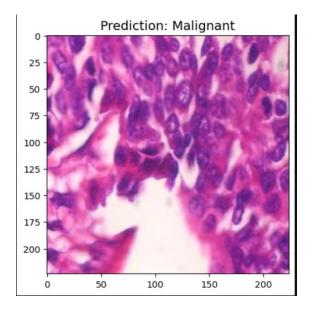
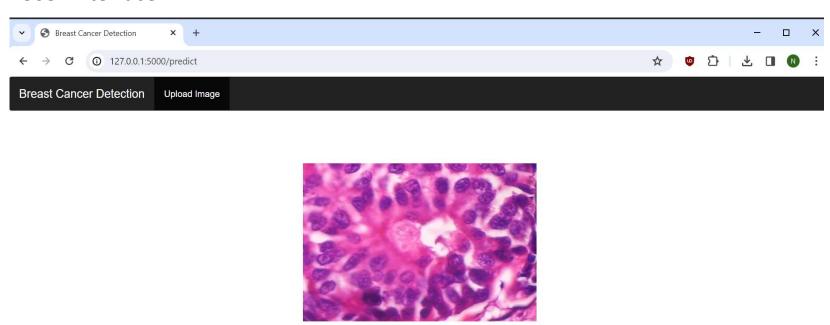


Fig. Sample Output of Cancerous Breast



#### Test Case 3 and 4:

#### User Interface:



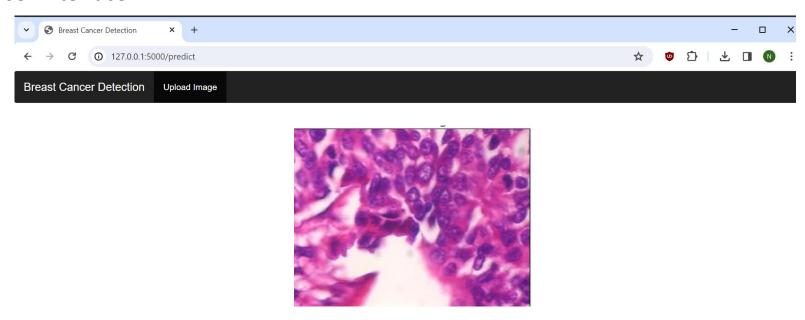
Prediction: Malignant and Stage: Stage 2 Recommendations: Combined treatment approach involving surgery, chemotherapy, radiation, and hormone therapy; specifics depend on tumor size and lymph node involvement.

Fig. Sample Output showing Malign and its stage as Prediction with recommendations.



#### Test Case 3 and 4:

#### User Interface:



Prediction: Malignant and Stage: Stage 3 Recommendations: Combined treatment approach involving surgery, chemotherapy, radiation, and hormone therapy; specifics depend on tumor size and lymph node involvement.

Fig. Sample Output showing Malign and its stage as Prediction with recommendations.



### **Proposed Main Paper:-**

This comprehensive review paper delves into the analysis of breast cancer diagnosis, focusing on key approaches such as convolutional neural networks and multi-modal integration. The examination encompasses modalities comparison, classifier evaluations, and advancements in deep learning. Within this context, the paper highlights both the progress made, and the challenges faced in the field. Addressing issues such as model interpretability and dataset limitations, the review aims to enhance accuracy for early breast cancer diagnosis. The exploration extends to advanced neural network architectures, transfer learning techniques, and the integration of multimodal data, providing insights into potential solutions. The synthesis of existing knowledge within the paper is strategically positioned to foster collaboration and innovation in breast cancer diagnosis. By bringing together insights into advancements, challenges, and emerging opportunities, the review aims to inspire collective efforts within the healthcare community.



#### **Conclusion:**

In conclusion, our project on breast cancer classification utilizing deep learning methodologies marks a significant advancement in the realm of medical imaging and machine learning. Through the integration of diverse datasets encompassing biopsy images, mammograms, and mri data, alongside the development of sophisticated deep learning models to achieve optimal performance. The culmination of our efforts is manifested in a user-friendly interface empowering clinicians and patients to leverage insights for informed decision-making and proactive health management. The suggested model accurately predicts whether a given sample of image has cancer or is normal using a number of convolutional layers. This is extremely useful in the medical field for detecting Cancer in patients early and accurately. Early diagnosis is critical for saving a person's life by ensuring that the patient receives effective and timely treatment. Detection of Cancer phase/stage using Decision Tree also helps in the diagnosis and treatment effectively.



#### References:

- [1] Murtaza, Ghulam, et al. "Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges." Artificial Intelligence Review 53.3 (2020): 1655-1720.
- [2] Vijayakumar, K., Vinod J. Kadam, and Sudhir Kumar Sharma. "Breast cancer diagnosis using multiple activation deep neural network." Concurrent Engineering 29.3 (2021): 275-284.ent care.
- [3] Yang, Zhanbo, et al. "EMS-Net: Ensemble of multiscale convolutional neural networks for classification of breast cancer histology images." Neurocomputing 366 (2019): 46-53.
- [4] Yala, Adam, et al. "A deep learning mammography-based model for improved breast cancer risk prediction." Radiology 292.1 (2019): 60-66.
- [5] Amrane, Meriem, et al. "Breast cancer classification using machine learning." 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT). IEEE, 2018.
- [6] Akselrod-Ballin, Ayelet, et al. "Predicting breast cancer by applying deep learning to linked health records and mammograms." Radiology 292.2 (2019): 331-342.
- [7] Krithiga, R., and P. Geetha. "Breast cancer detection, segmentation and classification on histopathology images analysis: a systematic review." Archives of Computational Methods in Engineering 28.4 (2021): 2607-2619.
- [8] Hu, Chuhan, et al. "Classification of breast cancer histopathological image with deep residual learning." International Journal of Imaging Systems and Technology 31.3 (2021): 1583-1594.
- [9] Sarosa, Syam Julio A., Fitri Utaminingrum, and Fitra A. Bachtiar." Mammogram breast cancer classification using gray-level co-occurrence matrix and support vector machine." (SIET). IEEE, 2018.
- [10] Buda, Mateusz, et al. "A data set and deep learning algorithm for the detection of masses and architectural distortions in digital breast tomosynthesis images." JAMA network open 4.8 (2021): e2119100-e2119100.



### THANK YOU!