Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

## SMART INTERNZ – ARTIFICIAL INTELLIGENCE

## PROJECT REPORT

# Deep Learning Techniques for Breast Cancer Risk Prediction


| NAME | REG NO |
|---|---|
| **RANGANATHAN S V** | 18BME0374 |
| **MOHAN RAJ A** | 18BME0355 |
| **JASWANTH S** | 18BME0209 |
| **JAHAN** | 19BAI1042 |

# DECLARATION:-

Certified that this project report titled **"Deep Learning Techniques For Breast Cancer Risk Prediction"** is the Bonafide work of **"Ranganathan S V (18BME0374), Mohan Raj A (18BME0355), Jaswanth S (18BME0209) and Jahan (19BAI1042)"** who carried out the project work under my supervision, certified further that to the best of my knowledge the work reported here does not form part of any other project / research work on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

UNDER THE GUIDANCE OF

**PRADEEPTHI DUGGARAJU**

# ACKNOWLEDGEMENT:-

Presentation inspiration and motivation have always played a key role in the success of any venture.

We express our sincere thanks to our instructor **PREDEEPTHI DUGGARAJU** to encourage us to the highest peak and provide us the opportunity to the project.

We feel to acknowledge our deep sense of gratitude to him whose valuable guidance and kind supervision given to us throughout the course which shaped the present work as its show.

I would like to thank all the technical and teaching staff of the smart internz and smart bridge, who extended directly or indirectly all support. Last but not least, I am deeply indebted to my parents who have been the greatest support while I worked day and night for the project to make it a success.

# TABLE OF CONTENTS:-

# INTRODUCTION:-

The field of deep learning has taken a dramatic twist in recent times, with the rise of the Artificial Neural Network (ANN). These biologically inspired computational models are able to far exceed the performance of previous forms of artificial intelligence in common deep learning tasks. One of the most impressive forms of ANN architecture is that of the Convolutional Neural Network (CNN). CNNs are primarily used to solve difficult image-driven pattern recognition tasks and with their precise yet simple architecture, offers a simplified method of getting started with CNNs. In this project we have built a CNN architecture to look at mammogramsand predict whether its malignant or benign.

# PROBLEM STATEMENT:-

Breast cancer is one of the main causes of cancer death worldwide. Computer-aided diagnosis systems showed the potential for improving diagnostic accuracy. But early detection and prevention can significantly reduce the chances of death. It is important to detect breast cancer as early as possible.
The goal is to classify images into two classifications of malignant and benign. As early diagnostics significantly increases the chances of correct treatment and survival. In this application, we are helping the doctors and patients to classify the Type of Tumour for the specific image given with the help of Neural Networks.

# SOLUTION:-

CNNs are applied to explore patterns in an image. This is done by convoluting over an image and looking for patterns. The network can detect lines and corners in the few front layers of CNNs. Via our neural net, however, we can then transfer these patterns down and begin to identify more complex characteristics as we get deeper. This property ensures that CNNs are very effective at detecting objects in images .

The proposed system uses CNNs to detect breast cancer from breast tissue images.The architecture of a CNN has 3 main layers, the convolutional layer, pooling layer, and fully connected layer. The first layer calculates the output of neurons which are linked with local regions. Each one is calculated by a dot product of weights and the region. For image inputs, typical filters are small in area such as $3 \times 3$.

These filters scan the image by a sliding window on the image, while learning the recurrent patterns which arise in any area of the image. The interval between filters is known as the stride. The convolution is extended to overlapping windows if the stride hyperparameter is smaller than the filter dimension. Convolutional layers bring out the features of images with precise positions. If the positions change, even a small amount for any reason, the feature maps will be different. To overcome this problem, the down sampling process must be done at the output of every convolutional layer. With convolutional layers, down samplingcan be done by changing the convolution's phase across the image. A more acceptable and common method is to use a pooling layer. Using this process, outputs will be more accurate.

## LITERATURE SURVEY:-

Breast cancer forms in breast cells and is considered as a very common type of cancer in women. Breast cancer is also a very life-threatening disease of women after lung cancer. Breast cancer is categorized into various types according to the cell's appearance through a microscope. The two main types of breast cancer are **(1)** invasive ductal carcinoma (IDC) and **(2)** ductal carcinoma in situ (DCIS). The IDC type is more dangerous, surrounding the entire breast tissue. Most breast cancer patients, approximately 80%, are in this category.

Breast cancer starts when cells in the breast begin to grow out of control. These cells usually form a tumor that can often be seen on an x-ray or felt as a lump. The tumor is malignant (cancer) if the cells can grow into (invade) surrounding tissues or spread (metastasize) to distant areas of the body.

Breast cancer can be effectively treated through its early detection. Thus, the availability of proper screening methods is important for detecting the initial symptom of breast cancer.

Due to the intrinsic difficulties associated with an image, with meagre contrast, noise, and lack of appreciation by the eye, instruments have been prepared to make and improve image processing. Nowadays, artificial intelligence (AI), machine learning (ML), and convolutional neural network (CNN) are the quickest rising areas of healthcare industry.

AI and ML are found in the research arena that deals with and improves technological systems to resolve complex tasks through reducing necessity of human intelligence.

Deep learning (DL) which is part of machine learning family depended on artificial neural networks. DL architectures, such as DNN (deep neural networks), RNN (recurrent neural networks), DBN (deep belief networks), and CNN, are generally applied to the areas like computer vision, audio recognition, speech recognition, social network filtering, natural language processing, machine translation, drug design, bioinformatics, medical image analysis, materials scrutiny, histopathological diagnosis, and board game programs . These new technologies, in

particular DL algorithms, can be applied to improve the diagnostic accuracy and efficiency of cancer detection.

A convolution neural network (CNN) method is proposed in this study to boost the automatic identification of breast cancer by analysing hostile ductal carcinoma tissue zones in whole-slide images.

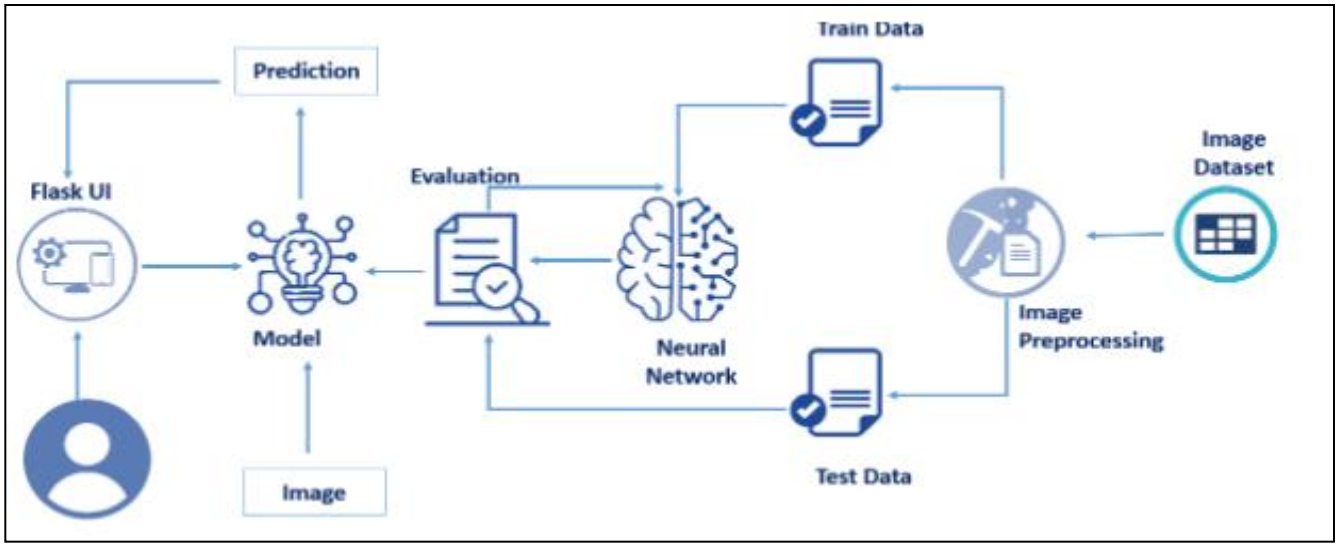
The opportunity that CNN brings to research on medical imaging is not restricted to deep CNN for extraction of the imaging feature. Indeed, a second field that can support medical research is the use of CNN for synthetic image rendering.

## EXPERIMENTAL ANALYSIS:-

Breast cancer is one of the main causes of cancer death worldwide. Computer-aided diagnosis systems showed the potential for improving diagnostic accuracy. But early detection and prevention can significantly reduce the chances of death. It is important to detect breast cancer as early as possible.

The goal is to classify images into two classifications of malignant and benign, as early diagnostics significantly increases the chances of correct treatment and survival. In this application, we are helping the doctors and patients to classify the Type of Tumour for the specific image given with the help of Neural Networks.

### Project Architecture:-



## Software Requirements:-

- Python 3.9
- Anaconda navigator
- TensorFlow version 1.14.0
- Keras 2.2.4
- Flask
- And other python libraries like NumPy, pandas, etc..,

## Project Flow:-

1) **DATA COLLECTION**
   - **a)** In our project according to project structure, create train test folders and in them, place "0" folder images in benign and place "1" folder images in Malignant in train and test folders respectively as shown in the project structure.

2) **IMAGE PRE-PROCESSING**
   - **a)** Import image data generator library and configure it
   - **b)** Apply image data generator functionality to train and test datasets

3) **MODEL BUILDING**
   - **a)** Importing the required libraries for model building
   - **b)** Initialize the model
   - **c)** Add convolution layer
   - **d)** Add max pooling layer
   - **e)** Add flatten layer
   - **f)** Add hidden layers
   - **g)** Compile the model
   - **h)** Fit and save the model

4) **TEST THE MODEL**

**a)** Import the saved model

**b)** Load the test image, preprocess it and then predict and check for results

5) **APPLICATION BUILDING**

**a)** Build a FLASK application

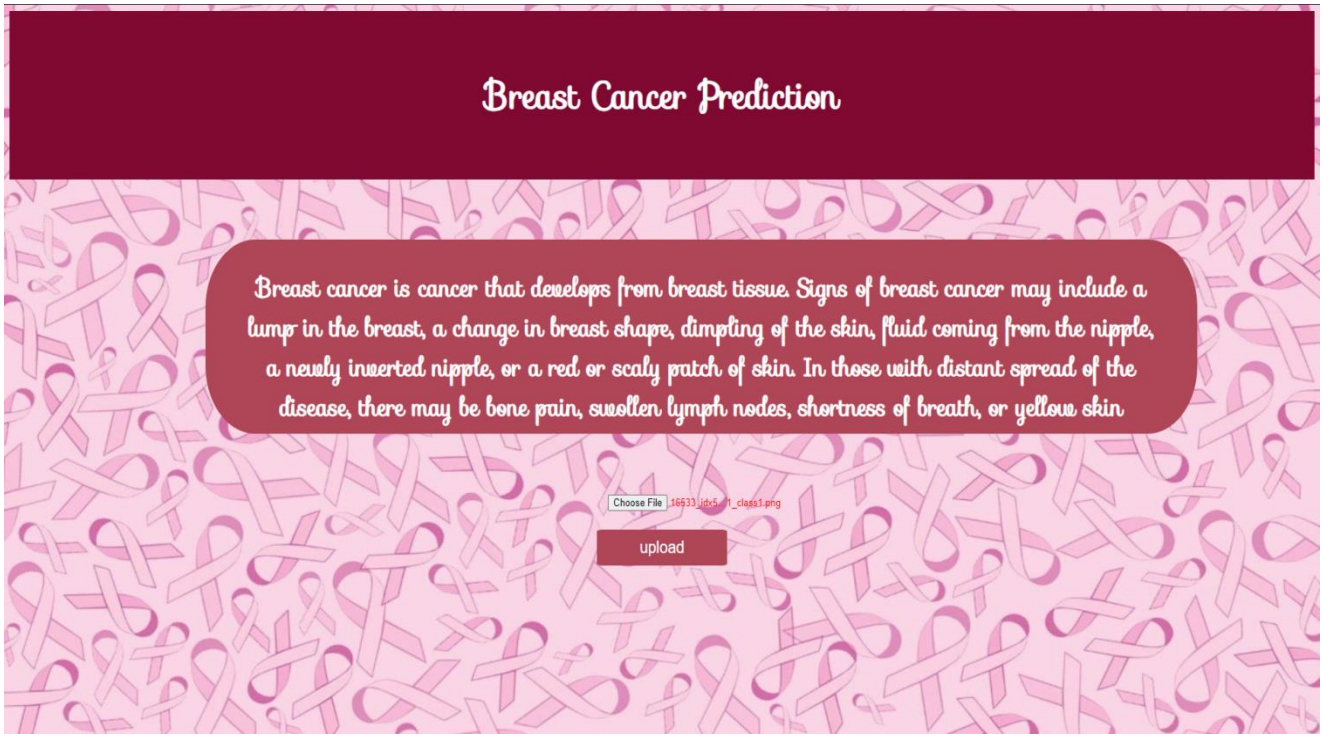
**b)** Build the HTML page and execute it

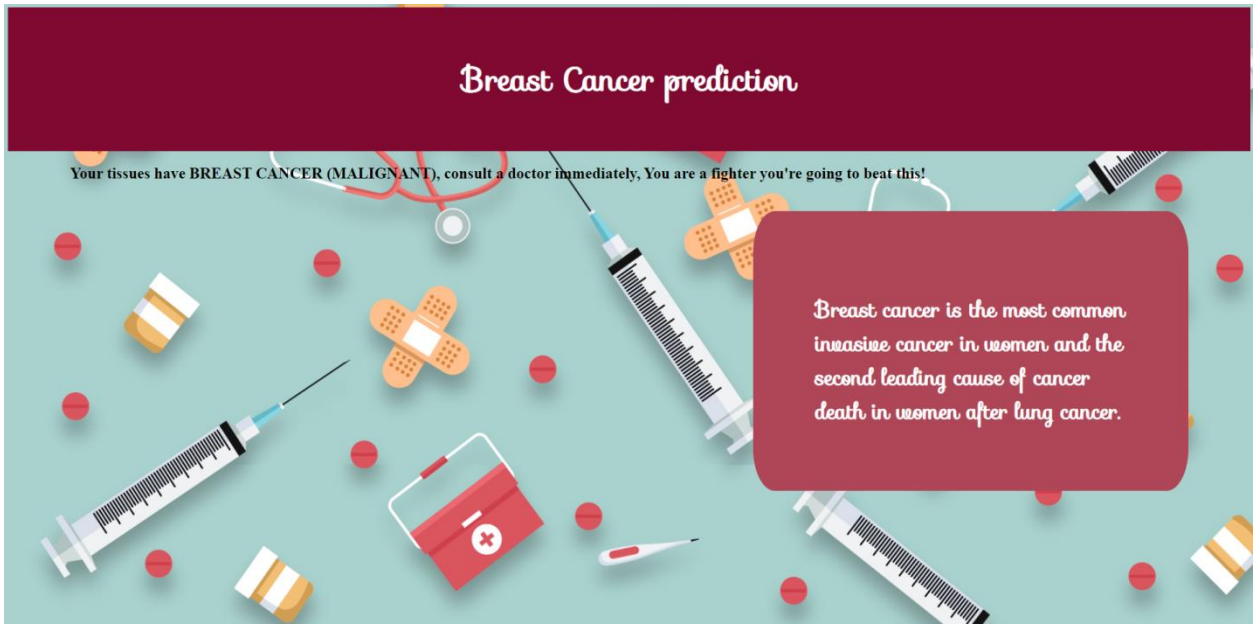
**c)** Run the app

# CONCLUSION:-

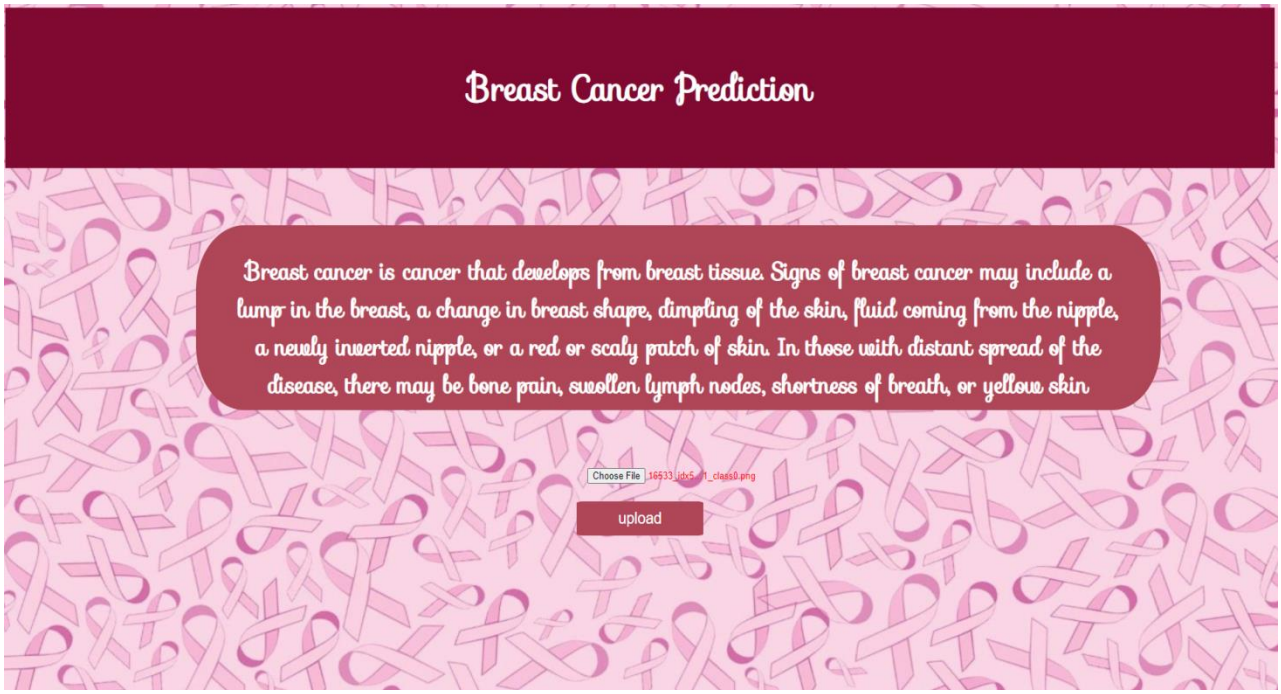
The following steps listed above are performed by our team, and herewith we attach snaps of our web page we achieved.

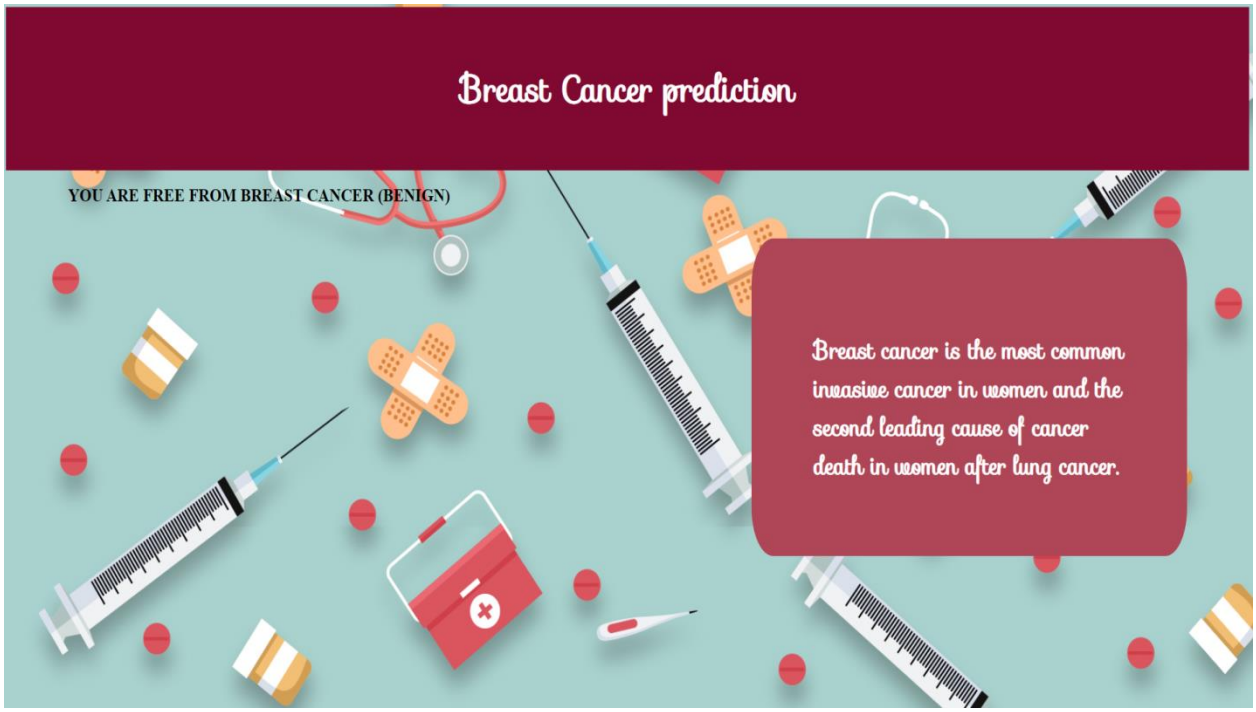
1. CLASS 1 (MALIGNANT)






2. CLASS 0 (BENIGN)






## APPLICATION:-

The following application could be used in a better understanding with if patients past history details are also uploaded, and the accuracy can be increased if the system is fed to a large no of and variety of dataset. With various technologies and AI(artificial intelligence) applications coming up in the near future we can add a lot of advantage to the current system and get more realistic results. We can also apply a series of techniques involving scanned images as well as other related data points together for the classification of the dataset .Artificial Intelligence (XAI) techniques could also be used to understand the different regions of the breast during classification.

## FUTURE SCOPE:-

Based on the investigation and findings of the study, the following recommendations are forwarded for further research works. We have only considered breast mass abnormality. So, the study can be extended to include macro classification of abnormalities that is not considered here. By increasing the amount of dataset, the performance of the model can be enhanced.

Finally, since the network is designed to consider multiple biological scales, the proposed system can be extended for whole-slide breast histology image classification relevant for clinical settings. Although the DL (deep learning) methods show promising improvements in breast cancer diagnosis, there are still issues of data scarcity and computational cost, which have been overcome to a significant extent by applying data augmentation and improved computational power of DL algorithms.

# CODE SNIPPETS:-

## 1) MODEL CODE:-

### 1. IMPORT LIBRARIES

In [1]:
```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Convolution2D, MaxPooling2D, Flatten
from tensorflow.keras.preprocessing.image import ImageDataGenerator
```

### 2. IMAGE PREPROCESSING

In [2]:
```
train_data = ImageDataGenerator(rescale = 1./224, shear_range = 0.2, zoom_range = 0.2, horizontal_flip = True)
test_data = ImageDataGenerator(rescale = 1./224)
```

In [3]:
```
x_train = train_data.flow_from_directory(r'DATASET\TRAIN', target_size = (32,32), batch_size = 32, class_mode = "binary")
x_test = test_data.flow_from_directory(r'DATASET\TEST', target_size = (32,32), batch_size = 32, class_mode = "binary")
```

```
Found 102394 images belonging to 2 classes.
Found 44087 images belonging to 2 classes.
```

```
HERE THE TWO CLASSES ARE BENIGN AND MALIGNANT
BENIGN - CLASS 0
MALIGNANT - CLASS 1
```

### 3. INITIALIZING THE MODEL

In [4]:
```
detector = Sequential()
```

### 4. ADDING CONVOLUTION LAYER

In [5]:
```
detector.add(Convolution2D(32,(3,3), input_shape = (32,32,3)))
# the 3 in input shape indicates it is coloured image(RGB)
# (3,3) indicates feature detector matrix size
# 32 in no of feature detectors
```

## 5. ADDING POOLING LAYER

In [6]:
```
detector.add(MaxPooling2D((2,2)))
# (2,2) is pooling matrix size
```

In [7]:
```
detector.add(Convolution2D(64,(3,3), input_shape = (32,32,3)))
detector.add(MaxPooling2D((2,2)))
```

In [8]:
```
detector.add(Convolution2D(128,(3,3), input_shape = (32,32,3)))
detector.add(MaxPooling2D((2,2)))
```

## 6. ADDING FLATTEN LAYER

In [9]:
```
detector.add(Flatten())  #gives input layer
```

## 7. ADDING HIDDEN LAYER

In [10]:
```
detector.add(Dense(units = 512, kernel_initializer = "random_uniform", activation = "relu"))
```

## 8. ADDING OUTPUT LAYER ¶

In [11]:
```
detector.add(Dense(units = 1, kernel_initializer = "random_uniform", activation = "sigmoid"))
```

## 9. MODEL TRAINING

In [12]:
```
detector.compile(optimizer = "adam", loss = "binary_crossentropy", metrics = ["accuracy"])
```

In [13]:
```
history=detector.fit(x_train, steps_per_epoch = 102394//32, epochs = 10, validation_data = x_test, validation_steps = 44087//32)
```

```
Epoch 1/10
3199/3199 [==============================] - 2426s 758ms/step - loss: 0.4927 - accuracy: 0.7766 - val_loss: 0.4982 - val_accura
cy: 0.7649
Epoch 2/10
3199/3199 [==============================] - 291s 91ms/step - loss: 0.4156 - accuracy: 0.8187 - val_loss: 0.4397 - val_accurac
y: 0.8155
Epoch 3/10
3199/3199 [==============================] - 292s 91ms/step - loss: 0.3933 - accuracy: 0.8295 - val_loss: 0.4018 - val_accurac
y: 0.8236
Epoch 4/10
3199/3199 [==============================] - 297s 93ms/step - loss: 0.3823 - accuracy: 0.8360 - val_loss: 0.3949 - val_accurac
y: 0.8235
Epoch 5/10
3199/3199 [==============================] - 264s 83ms/step - loss: 0.3746 - accuracy: 0.8394 - val_loss: 0.3887 - val_accurac
y: 0.8348
Epoch 6/10
3199/3199 [==============================] - 254s 79ms/step - loss: 0.3655 - accuracy: 0.8426 - val_loss: 0.4426 - val_accurac
y: 0.8162
Epoch 7/10
3199/3199 [==============================] - 252s 79ms/step - loss: 0.3633 - accuracy: 0.8461 - val_loss: 0.3934 - val_accurac
y: 0.8331
Epoch 8/10
3199/3199 [==============================] - 254s 79ms/step - loss: 0.3596 - accuracy: 0.8466 - val_loss: 0.4352 - val_accurac
y: 0.8170
Epoch 9/10
3199/3199 [==============================] - 265s 83ms/step - loss: 0.3564 - accuracy: 0.8489 - val_loss: 0.4093 - val_accurac
y: 0.8279
Epoch 10/10
3199/3199 [==============================] - 265s 83ms/step - loss: 0.3517 - accuracy: 0.8506 - val_loss: 0.4245 - val_accurac
y: 0.8207
```

## 10. SAVING THE MODEL

In [14]:
```
detector.save("breast_cancer_classifier.h5")
```

## 2) HTML CODE OF INDEX:-

```
<html>
<head><link href='https://fonts.googleapis.com/css?family=Sofia' rel='stylesheet'></head>
<body>
    <style>

        body {
        background-image: url("static/background1.jpeg");
        background-repeat: no-repeat;
        background-attachment: fixed;
        background-size: 100% 100%;

      }
      .bg-dark {
          background-color: #42678c!important;
      }

      #result {
          color: #0a1c4ed1;
      }
      .col-sm{
          text-align: center;
      }
      .badge-badge-danger{
        /* color:red; */
        /* border: 2px solid black; */
        padding: 2%;
        margin: 0px 0px 0px 0px;
        background-color: rgb(128, 9, 49);
        color: white;
        font-size: 100%;
        font-family: 'Sofia';
      }
      .right{

        margin-top: 4%;
      }
      .inp{
        background: rgb(175, 70, 87);
        border: none;
        border-radius: 5%;
        width:10%;
        height:5%;
        font-size: 90%;
        color: white;

      }
      .img{
        width:10%;
        height: 20%;
        border: 0.1em solid black;
        margin-left: 45%;

      }
      .inps{
        border: none;
        color: red;
        margin-left: 5%;

      }
      .text{
        background-color: rgb(175, 70, 87);
        border-radius: 5%/35%;
        padding-left: 3% ;
        padding-right: 3%;
        padding-top: 0.5%;
```

```
        padding-bottom: 0.5%;
        margin-left: 15%;
        margin-right: 15%;
        color: white;
        height: 25%;
        width: 70%;
        font-size: 90%;
        font-family: 'Sofia';
      }
      </style>
 <div class="col-sm">
      <div class="badge-badge-danger"><h1>Breast Cancer Prediction</h1>
    </div>
    <div class="right">
      <div class="text"><h2>Breast cancer is cancer that develops from breast tissue. Signs of breast cancer may include
a lump in the breast, a change in breast shape, dimpling of the skin, fluid coming from the nipple, a newly inverted nipp
le, or a red or scaly patch of skin. In those with distant spread of the disease, there may be bone pain, swollen lymph n
odes, shortness of breath, or yellow skin</h2></div>
      <br><br>
   <form action = "/predict" method = "post" enctype = "multipart/form-data">
    <p> <input class="inps"  type = "file" name = "file" /></p>
    <p> <input class="inp"  type = "submit" value = "upload"/></p>
   </form>
</div>
</div>
</body>
</html>
```

## 3) HTML CODE OF PREDICT:-

```
<html>
  <head><link href='https://fonts.googleapis.com/css?family=Sofia' rel='stylesheet'></head>
<body>
  <style>
    body {
    background-image: url("static/background2.jpg");
    background-repeat: no-repeat;
    background-attachment: fixed;
    background-size: cover;
  }
  .bg-dark {
      background-color: #42678c!important;
  }
  .col-sm{
    text-align: center;
  }
  .badge-badge-secondary{
    padding: 2%;
        margin: 0px 0px 0px 0px;
        background-color: rgb(128, 9, 49);
        color: white;
        font-size: 100%;
        font-family: 'Sofia';
        text-align: center;
  }
  .right{
    margin-left: 60%;
    margin-right: 5%;
        margin-top: 1%;
```

```
        padding: 5%;
        padding-top: 5%;
        color: white;
        height: 25%;
        width: 25%;
        font-size: 90%;
        font-family: 'Sofia';
        background-color: rgb(175, 70, 87);
        border-radius: 5%/15%;
        display: inline-block;
      }
  .left{
    margin-left: 5%;
    margin-top: 1%;
    display: inline-block;
  }
  #result {
      color: #0a1c4ed1;
  }
  </style>
<div class="col-sm">
  <div class="badge-badge-secondary"><h1>Breast Cancer prediction</h1>
    </div>
    </div>
    <p class="left">{{prediction}}</p>
    <div class="right">
      <div ><h2>Breast cancer is the most common invasive cancer in women and the second leading cause of cancer death in
 women after lung cancer.</h2></div>
</div>
</body>
</html>
```

## 4) APP.PY CODE:-

```
from flask import Flask, render_template, request
import pandas as pd
import numpy as np
from tensorflow.keras.models import load_model
from tensorflow.keras.preprocessing import image
import os

app = Flask(__name__)
app.config["MAX_CONTENT_LENGTH"] = 10*1024*1024

def read_image(filename):
    img = image.load_img(filename , target_size = (32,32))
    img = image.img_to_array(img)
    img = img.reshape(1,32,32,3)
    return img

@app.route("/", methods = ['GET','POST'])
def home():
        return render_template('index.html')

@app.route("/predict", methods = ['GET','POST'])
def predict():
    if request.method == 'POST':
```

```
        file = request.files['file']

        filename = file.filename
        file_path = os.path.join('static' , filename)
        file.save(file_path)
        img = read_image(file_path)
        model1 = load_model('breast_cancer_classifier.h5')
        class_prediction = model1.predict_classes(img)
        if class_prediction[0] == 0:  #BENIGN
            result = "YOU ARE FREE FROM BREAST CANCER (BENIGN)"
        else:                         #MALIGNANT
            result = "Your tissues have BREAST CANCER (MALIGNANT), consult a doctor immediately, \nYou are a fighter you'
re going to beat this!"
    return render_template('predict.html', prediction = result)


if __name__ == "__main__":
    app.run(debug =True)
```

# BIBLIOGRAPHY:-


- https://smartbridge.teachable.com/courses/1450164/lectures
- https://www.hindawi.com/journals/jhe/2021/5528622/
- https://www.w3schools.com