

Deep Learning Techniques for Breast Cancer Risk Prediction using IBM Cloud

1. INTRODUCTION

1.1 Overview

Breast cancer is a dominant cancer in women worldwide and is increasing in developing countries where the majority of cases are diagnosed in late stages. The projects that have already been proposed show a comparison of machine learning algorithms with the help of different techniques like the ensemble methods, data mining algorithms or using blood analysis etc. This paper proposed now presents a comparison of six machine learning (ML) algorithms: Naive Bayes (NB), Random Forest (RT), Artificial Neural Networks (ANN), Nearest Neighbour (KNN), Support Vector Machine (SVM) and Decision Tree (DT) on the Wisconsin Diagnostic Breast Cancer (WDBC) dataset which is extracted from a digitised image of an MRI. For the implementation of the ML algorithms, the dataset was partitioned into the training phase and the testing phase. The algorithm with the best results will be used as the backend to the website and the will then classify the cancer as benign or malignant.

1.2 Purpose

Machine learning is one of the most popular models to easily train machines and create predictive models for successful decision-making. Machine learning helps with early diagnosis of breast cancer and determines the nature of the cancer by analysing the tumour size. ML methods are the leading approaches to obtain favourable outcomes among classification and prediction problems. Breast cancer research could benefit from ML techniques used to identify cancer and predict the presence or absence of tumours.

2. LITERATURE SURVEY

The role various modalities in breast imaging by Sachin Prasad N and Dana Houserkova, 2007. Mammography is the only reliable screening test proven in breast imaging.

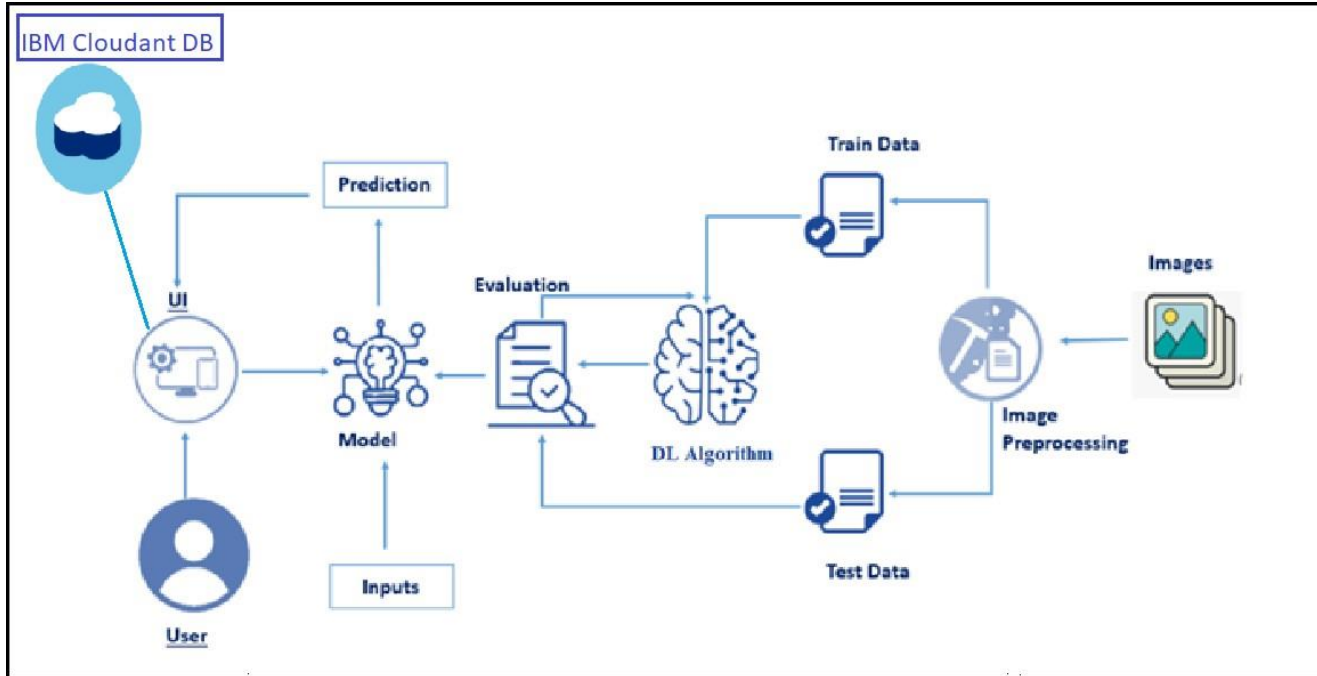
Identification of Preprocessing Technique for Enhancement of Mammogram Images by Jaya Sherma, R P Tewari and J K Rai, 2014.

Determining best preprocessing technique on the basis of peak signal to noise ratio for set of mammogram images.

Classification of Mammogram Images by using CNN Classifier by Ketan Sharma and Bobbin Preet, 2016. In this paper they proposed a computer aided diagnosis(CAD) system named as CNN. . They had also compared of CNN with Logistic Regression algorithm.

3. THEORITICAL ANALYSIS

3.1 Block Diagram



3.2 Hardware / Software designing

Software Requirements:

- Anaconda Navigator
- Tensor flow
- Keras
- Flask

Hardware Requirements:

- Processor : Intel Core i3
 - Hard Disk Space : Min 100 GB
 - Ram : 4 GB
 - Display : 14.1 “Color Monitor(LCD, CRT or LED)
- Clock Speed : 1.67 GHz

4. EXPERIMENTAL INVESTIGATIONS

Study shows that it provide with different test images of breast cancer images, the model detects, cost prediction of uploaded image. When we choose an image and click in to the upload it then it will shows the predicted output.


5. RESULT

New TabBreast Cancer Detection

localhost:5000

ENG IN02:23 PM10-10-2022

Detect the Tumor in the Breast



Breast cancer is one of the main causes of cancer deaths worldwide. Early diagnostics significantly increases the chances of correct treatment and survival, but this process is tedious and often leads to a disagreement between pathologists. Computer-aided diagnosis systems showed potential for improving the 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. This application is based on Convolution Neural Networks which classifies the scan as either benign or malignant tumor.


DROP THE SCAN FOR DETECTION!

New TabBreast Cancer Detection

localhost:5000

ENG IN02:23 PM10-10-2022

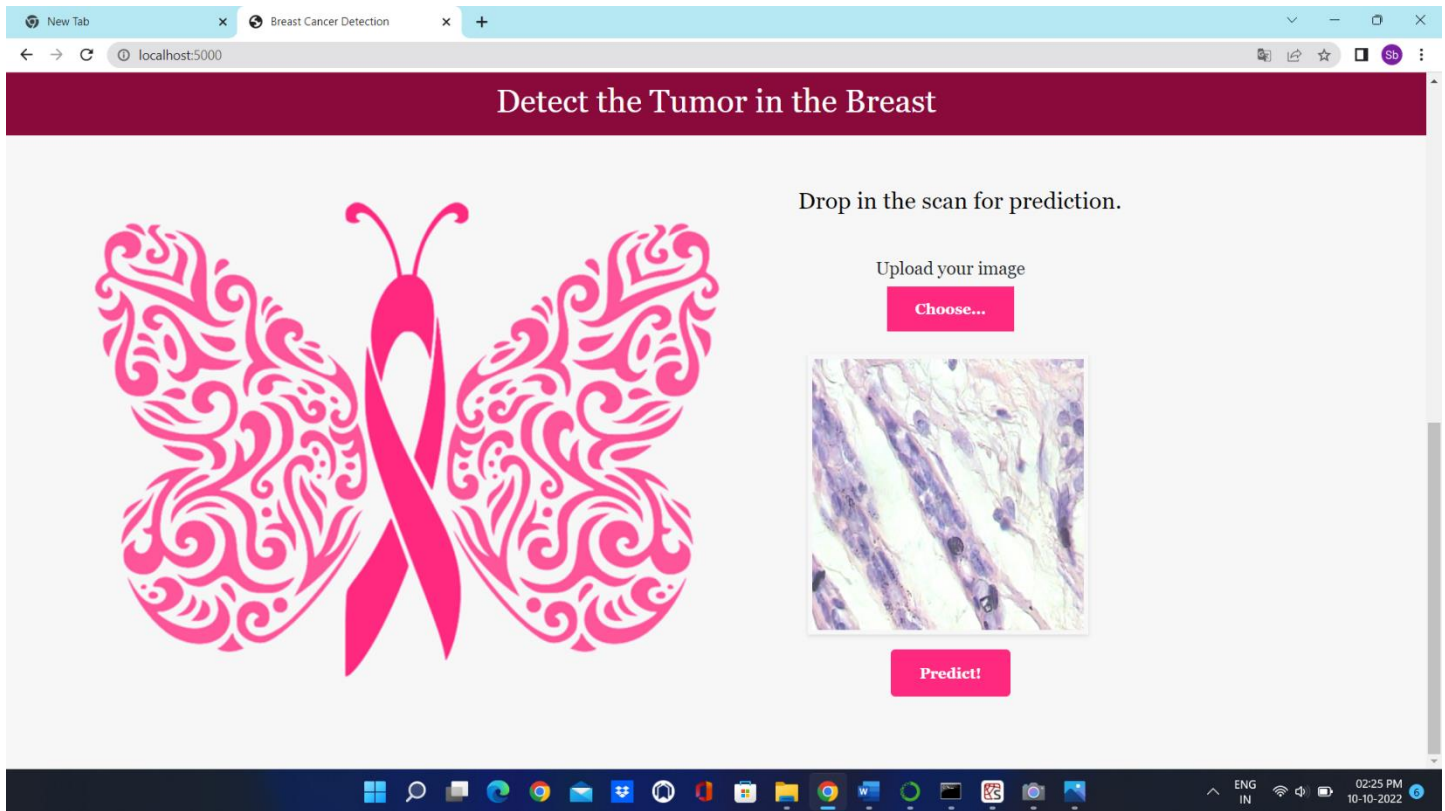
Detect the Tumor in the Breast



Drop in the scan for prediction.

Upload your image

Choose...



6. ADVANTAGES & DISADVANTAGES

Advantages:

- Increased accuracy for insurance prediction.
- Reduce the time complexity.

Disadvantages:

- Data mining techniques does not help to provide effective decision making.

7. APPLICATIONS

- Deep Learning technology is considered as one of the key technology used in breast cancer detection.
- It presents the results obtained by processing input from uploading image.

8. CONCLUSION

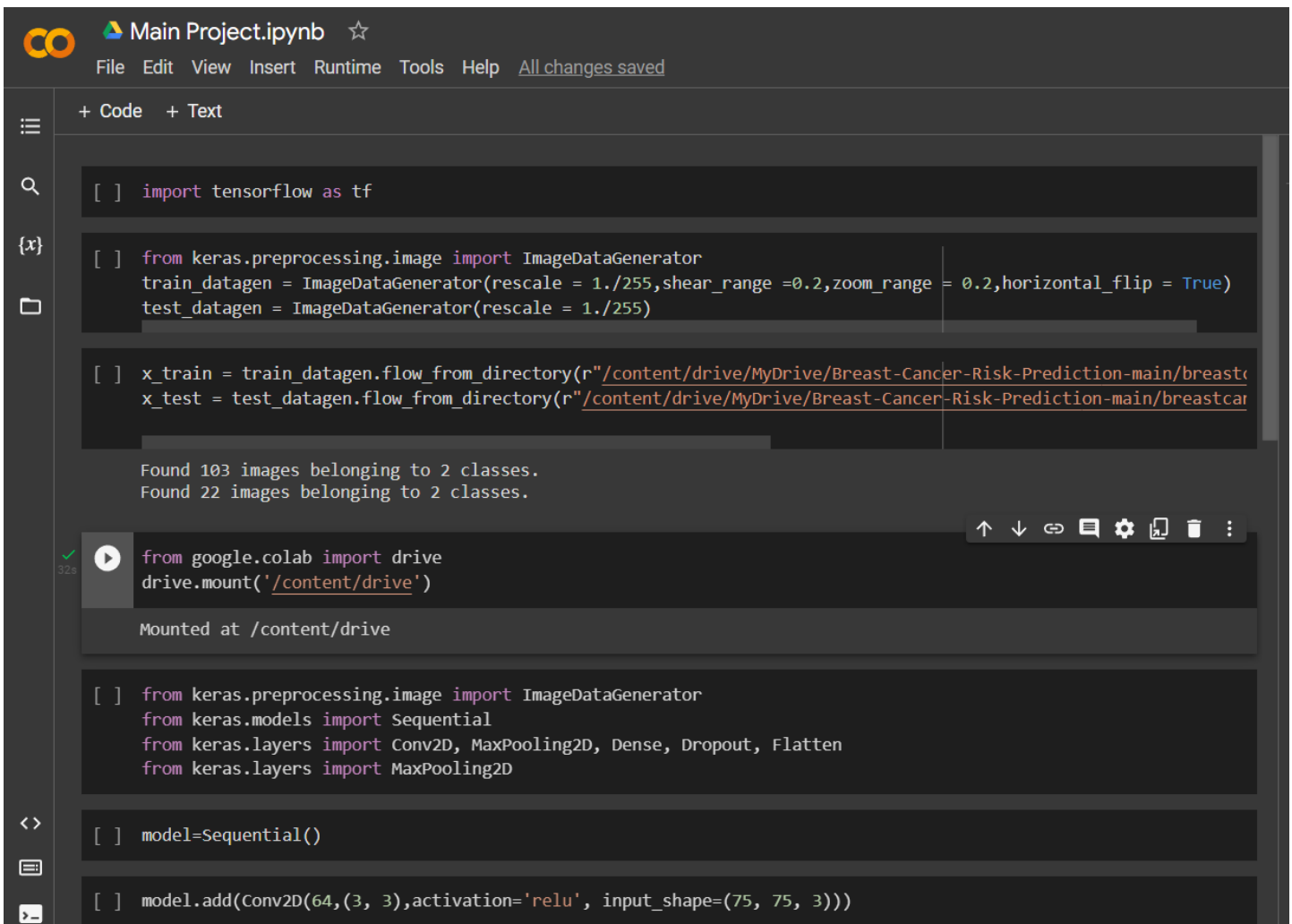
In this project, we have established the application to predict from uploaded image based on the IBM cloud application. Breast cancer prediction can only use this web app to predict the cancer.

9. FUTURE SCOPE

The analysis of the results signifies that the integration of multidimensional data along with different classification, feature selection and dimensionality reduction techniques can provide auspicious tools for inference in this domain. Further research in this field should be carried out for the better performance of the classification techniques so that it can predict on more variables. We are intending how to parametrize our classification techniques hence to achieve high accuracy. We are looking into many datasets and how further Machine Learning algorithms can be used to characterize Breast Cancer. We want to reduce the error rates with maximum accuracy

APPENDIX

Source Code



The screenshot displays a Jupyter Notebook titled "Main Project.ipynb" with a menu bar (File, Edit, View, Insert, Runtime, Tools, Help) and a status bar ("All changes saved"). The notebook contains several code cells. The first cell imports TensorFlow. The second cell imports Keras ImageDataGenerator and creates train and test datagen objects. The third cell uses flow_from_directory to load data from a local path, with output showing 103 training and 22 testing images. The fourth cell imports Google Colab drive and mounts the content drive. The fifth cell imports Keras Sequential model and various layers. The sixth cell initializes the Sequential model. The seventh cell adds a Conv2D layer to the model.

```
[ ] import tensorflow as tf

[ ] from keras.preprocessing.image import ImageDataGenerator
train_datagen = ImageDataGenerator(rescale = 1./255, shear_range = 0.2, zoom_range = 0.2, horizontal_flip = True)
test_datagen = ImageDataGenerator(rescale = 1./255)

[ ] x_train = train_datagen.flow_from_directory(r"/content/drive/MyDrive/Breast-Cancer-Risk-Prediction-main/breastcancer",
                                                class_mode='binary', shuffle=True)
x_test = test_datagen.flow_from_directory(r"/content/drive/MyDrive/Breast-Cancer-Risk-Prediction-main/breastcancer",
                                          class_mode='binary', shuffle=True)

Found 103 images belonging to 2 classes.
Found 22 images belonging to 2 classes.

from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive

[ ] from keras.preprocessing.image import ImageDataGenerator
from keras.models import Sequential
from keras.layers import Conv2D, MaxPooling2D, Dense, Dropout, Flatten
from keras.layers import MaxPooling2D

[ ] model=Sequential()

[ ] model.add(Conv2D(64,(3, 3),activation='relu', input_shape=(75, 75, 3)))
```

Files

🔍

📁

📄

🔗

🔒

{x}

..

drive

sample_data

breastcancer.h5

<>

☰

🔍

Disk: 100% available

+ Code + Text

✓ [13] model.add(Conv2D(64,(3, 3),activation='relu', input_shape=(75, 75, 3)))
0s

✓ [14] model.add(MaxPooling2D(pool_size = (2,2)))
0s

✓ [15] model.add(Flatten())
0s

✓ [16] model.add(Dense(units= 40 ,kernel_initializer='random_uniform',activation = 'relu'))
0s

✓ [17] model.add(Dense(units= 1,activation = 'softmax',kernel_initializer= 'uniform'))
0s

✓ [18] model.compile(loss='binary_crossentropy',optimizer='adam',metrics=['accuracy'])
0s

✓ [19] model.save('breastcancer.h5')
0s

✓ [20] from keras.models import load_model
from keras.preprocessing import image
import numpy as np
from tensorflow.keras.models import load_model
0s

✓ [21] model = load_model("breastcancer.h5")
0s

✓ [21]
0s

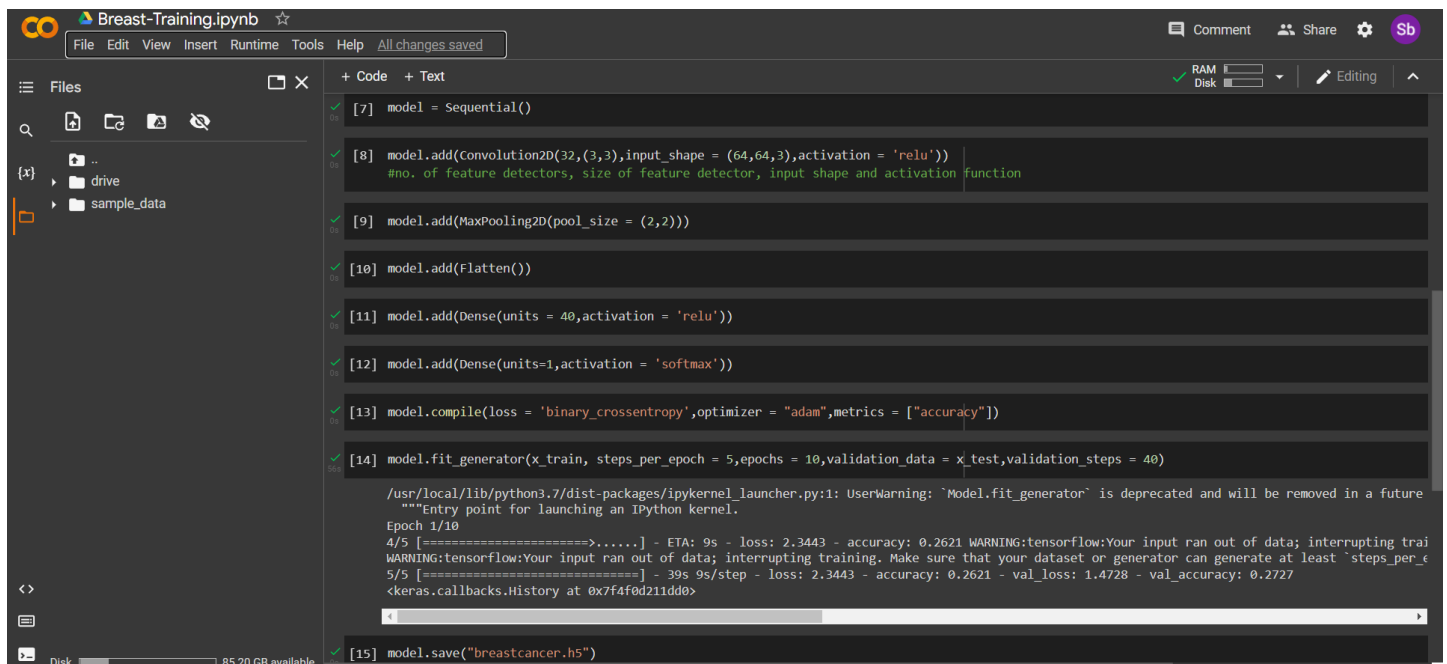
✓ [21]
0s

Training Model:

The screenshot displays a Jupyter Notebook environment with the following components:

- Top Bar:** Title 'Breast-Training.ipynb', menu items (File, Edit, View, Insert, Runtime, Tools, Help), and user profile 'Sb'.
- Left Panel:** File explorer showing a directory structure with 'sample_data'.
- Main Area:** Code cells for training a model.
 - Cell [2]:** Imports necessary Keras modules: `from keras.preprocessing.image import ImageDataGenerator`, `from keras.models import Sequential`, `from keras.layers import Dense`, `from keras.layers import Convolution2D`, `from keras.layers import MaxPooling2D`, and `from keras.layers import Flatten`.
 - Cell [3]:** Creates data generators: `train_datagen = ImageDataGenerator(rescale = 1./255, shear_range = 0.2, zoom_range = 0.2, horizontal_flip = True)` and `test_datagen = ImageDataGenerator(rescale = 1)`.
 - Cell [4]:** Flows data from directories: `x_train = train_datagen.flow_from_directory(r"/content/drive/MyDrive/Breast-Cancer-Risk-Prediction-main/breastcancerdataset/train", target_size = (180, 180))` and `x_test = test_datagen.flow_from_directory(r"/content/drive/MyDrive/Breast-Cancer-Risk-Prediction-main/breastcancerdataset/test", target_size = (180, 180))`. The output shows: 'Found 103 images belonging to 2 classes.' and 'Found 22 images belonging to 2 classes.'
 - Cell [5]:** Mounts Google Drive: `from google.colab import drive` and `drive.mount('/content/drive')`. The output indicates: 'Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).'
 - Cell [6]:** Prints class indices: `print(x_train.class_indices)`. The output is: `{'benign': 0, 'malignant': 1}`.
 - Cell [7]:** Initializes the model: `model = Sequential()`.

At the bottom, a disk usage indicator shows '85.25 GB available'.



```
[7] model = Sequential()

[8] model.add(Convolution2D(32,(3,3),input_shape = (64,64,3),activation = 'relu'))
#no. of feature detectors, size of feature detector, input shape and activation function

[9] model.add(MaxPooling2D(pool_size = (2,2)))

[10] model.add(Flatten())

[11] model.add(Dense(units = 40,activation = 'relu'))

[12] model.add(Dense(units=1,activation = 'softmax'))

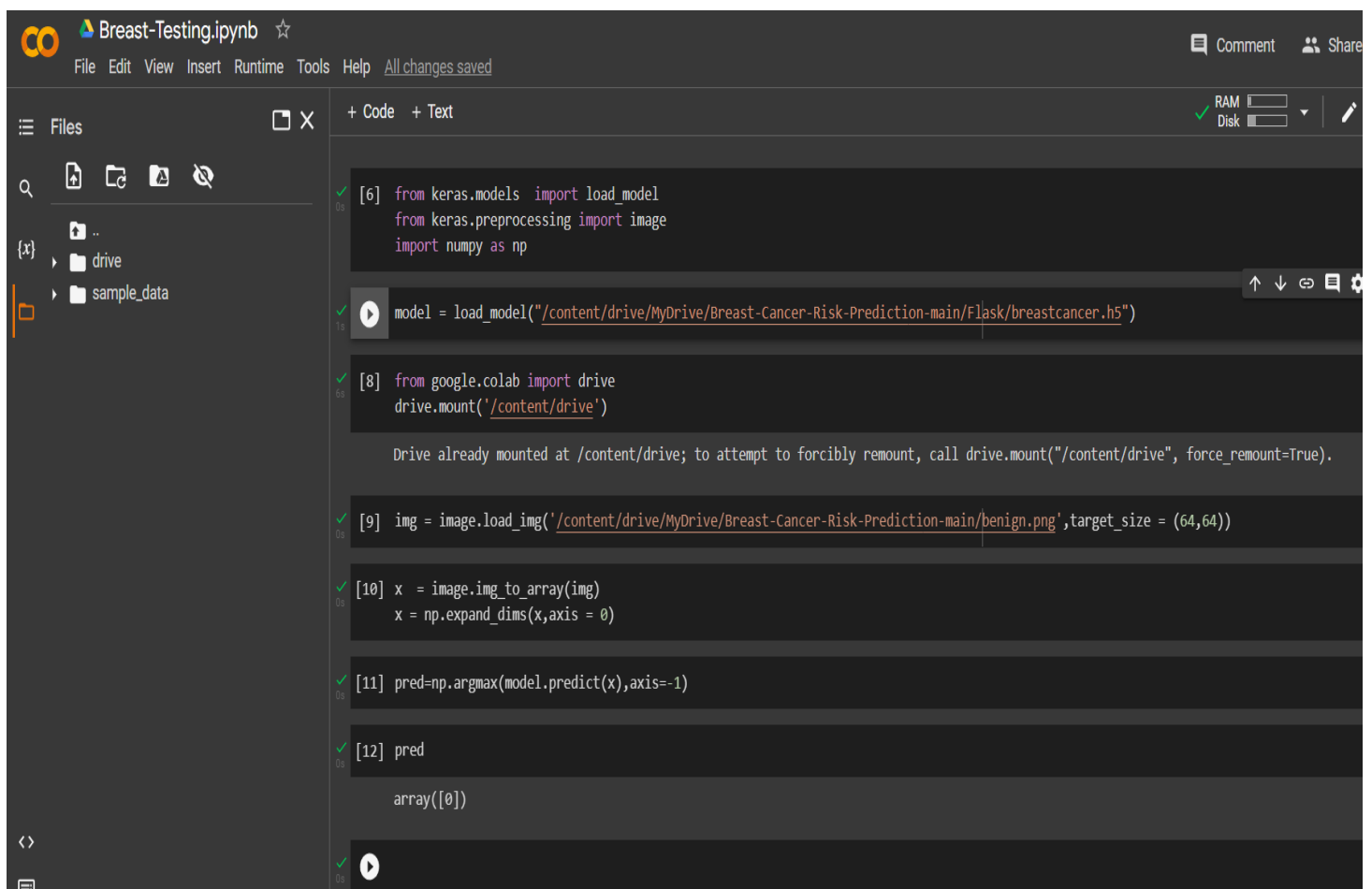
[13] model.compile(loss = 'binary_crossentropy',optimizer = "adam",metrics = ["accuracy"])

[14] model.fit_generator(x_train, steps_per_epoch = 5,epochs = 10,validation_data = x_test,validation_steps = 40)

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: UserWarning: `Model.fit_generator` is deprecated and will be removed in a future
***Entry point for launching an IPython kernel.
Epoch 1/10
4/5 [=====>.....] - ETA: 9s - loss: 2.3443 - accuracy: 0.2621 WARNING:tensorflow:Your input ran out of data; interrupting tra
WARNING:tensorflow:Your input ran out of data; interrupting training. Make sure that your dataset or generator can generate at least `steps_per_e
5/5 [=====] - 39s 9s/step - loss: 2.3443 - accuracy: 0.2621 - val_loss: 1.4728 - val_accuracy: 0.2727
<keras.callbacks.History at 0x7f4f0d211dd0>

[15] model.save("breastcancer.h5")
```

Testing Model:



```
[6] from keras.models import load_model
from keras.preprocessing import image
import numpy as np

model = load_model("/content/drive/MyDrive/Breast-Cancer-Risk-Prediction-main/Flask/breastcancer.h5")

[8] from google.colab import drive
drive.mount('/content/drive')

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

[9] img = image.load_img('/content/drive/MyDrive/Breast-Cancer-Risk-Prediction-main/benign.png',target_size = (64,64))

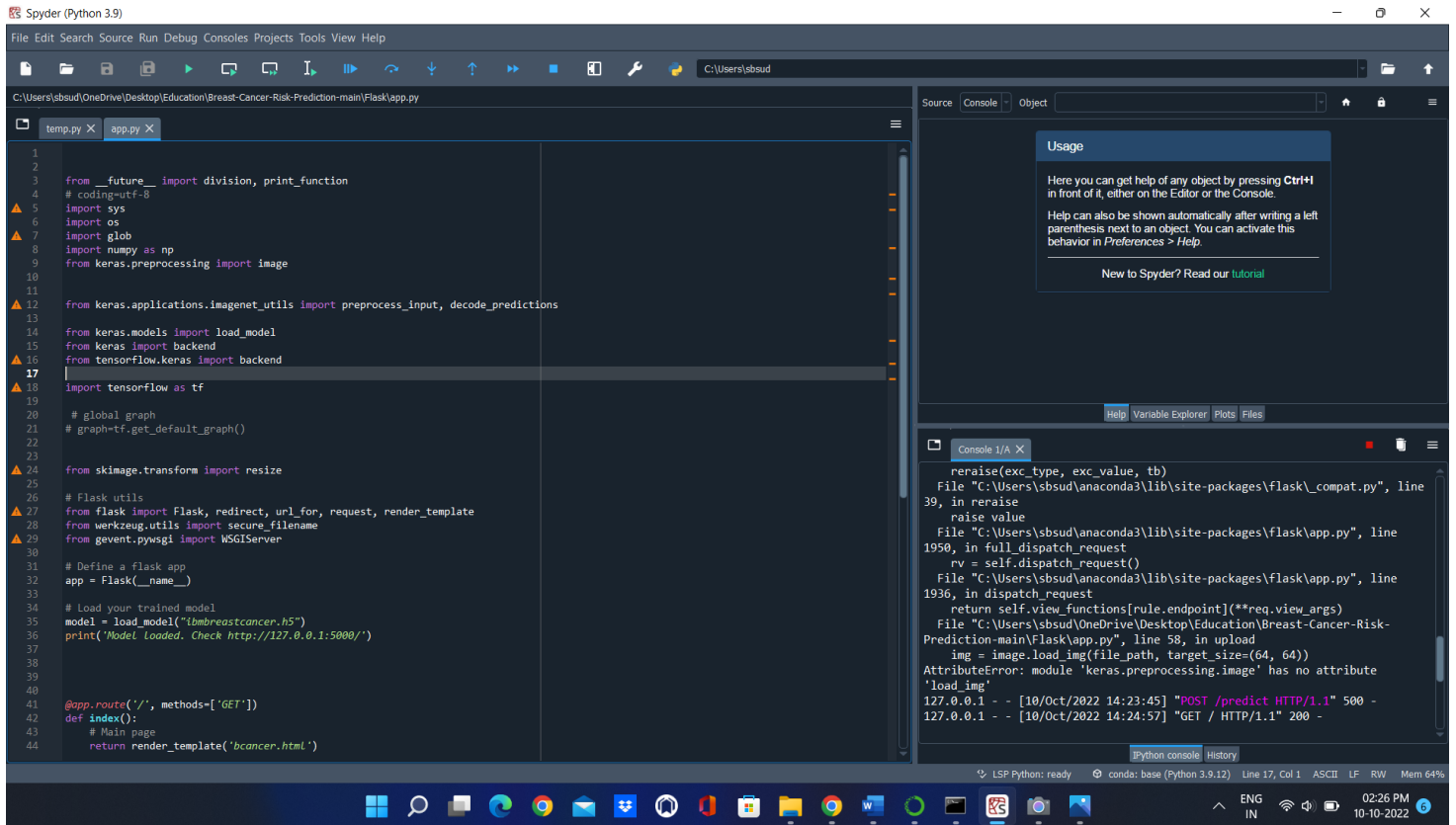
[10] x = image.img_to_array(img)
x = np.expand_dims(x,axis = 0)

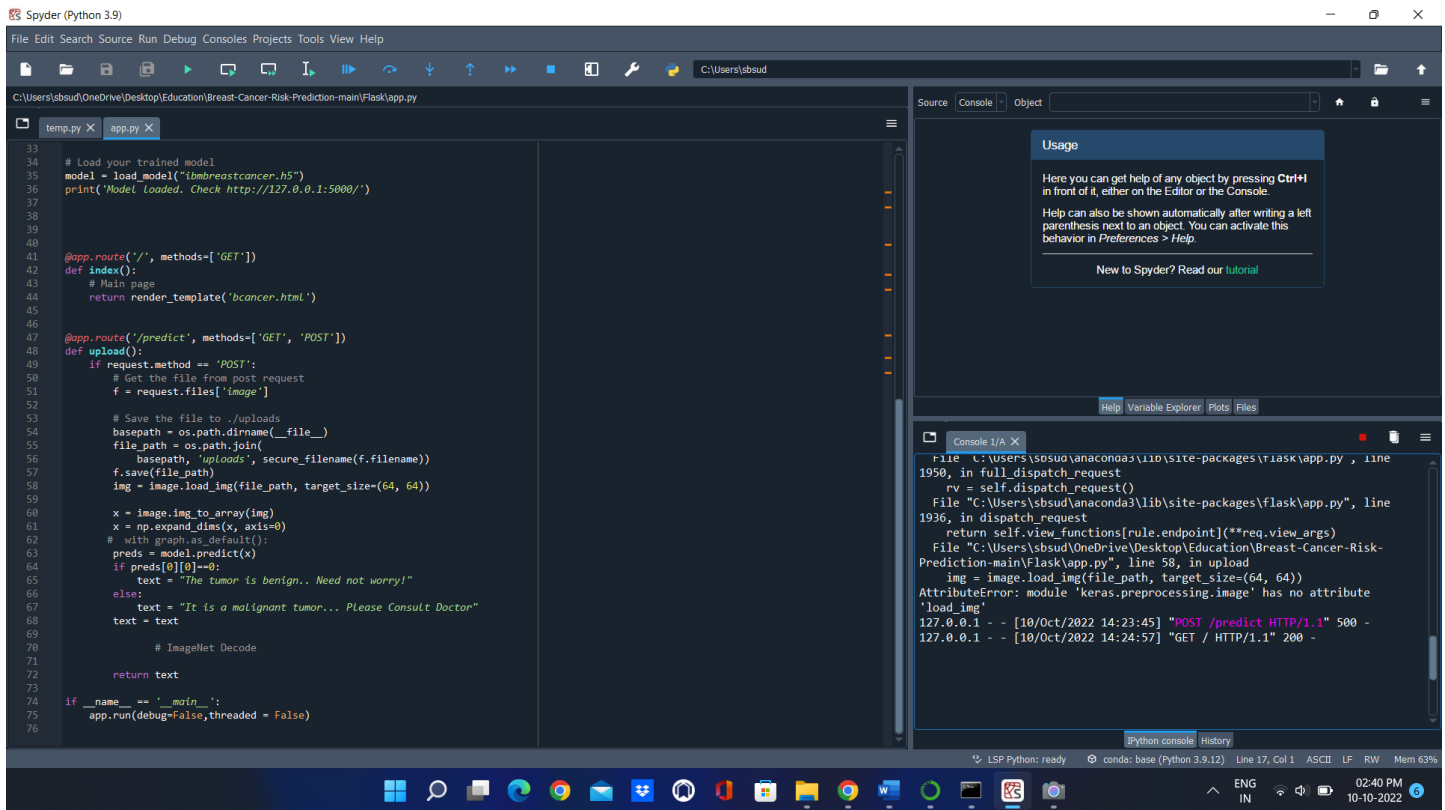
[11] pred=np.argmax(model.predict(x),axis=-1)

[12] pred

array([0])
```

Interface: App





Output:

