BREAST CANCER RISK PREDICTION

Using Convolutional Neural Networks

Developed by: Arpita Prakash Hegde, Neha Pawar KR, Nidhi S Jain, Nikita B N, Yashaswini Bandloor. V

Smart Bridge-Remote Summer Internship Program

1. INTRODUCTION

Deep learning applications are used in many industries like Automated driving, Aerospace and Defense, Medical Research, Industrial Automation, Electronics, etc. Cancer researchers use deep learning to automatically detect cancer cells. Teams at UCLA built an advanced microscope that yields a high-dimensional data set used to train a deep learning application to accurately identify cancer cells. Breast cancer is one of the main causes of cancer death 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 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. There are two most important goals for prediction and description. Prediction involves using some variables in data set to predict unknown values of other variables and Description concentrates on finding patterns describing the data that can be interpreted by human.

There are huge number of phases in the prediction in Deep Learning. The first phase is Data collection. In this phase, we create two folders namely testset and trainset and collect large number of images of cancer positive patients and cancer negative patients. The second phase is Data Preprocessing, it contains a lot of sub-phases for the processing of data, which includes importing libraries that is ImageDataGenerator, configuration of the ImageDataGenerator class, and applying the ImageDataGenerator functionality to trainset and testset. The third phase is Model Building, it includes importing the Model building libraries, initializing the model, loading the Preprocessed data, adding CNN layers, adding Dense layers, configuring the learning process, training and testing the model, optimizing the model, and saving the model. Thus, the last phase is Application Building, that is building a web application which includes an HTML file and a python code which is integrated to the built model.

1.1 OVERVIEW

A risk factor is anything that increases chances of getting a disease, such as cancer. But having a risk factor, or even many, does not confirm to get the disease. While we cannot change some breast cancer risk factors like family history and aging, etc. While there are some factors that can be controlled. Thus diagnosing the cancer at the earliest stage is very important. Thus Deep Learning techniques like Convolutional Neural Networks are used widely for the prediction of the breast cancer risk. This prediction is done using Python Flask model, in which a web application is built to analyze the cancer risk.

1.2 PURPOSE

Breast cancer is the main cause of death among women. Early prediction of breast cancer will help with the survival of breast cancer patients. Machine learning and Deep Learning have been widely used in the diagnosis of breast cancer and on the early detection of breast cancer. The main aim of this research is to review the role of deep learning techniques in breast cancer detection and diagnosis.

2. LITERATURE SURVEY

A literature review showed that there have been several studies on the survival prediction problem using statistical approaches and Convolutional Neural Networks. However, we could only find a few studies related to medical diagnosis and recurrence using data mining approaches, artificial neural networks, convolutional neural networks to develop prediction models for breast cancer survival by analyzing a large dataset. They studied 951 breast cancer patients and used tumor size, axillary nodal status, histological type, mitotic count, nuclear pleomorphism, tubule formation, tumor necrosis, and age as input variables. Pendharker patterns in breast cancer, in this study, they showed that data mining could be a valuable tool in identifying similarities (patterns) in breast cancer cases, which can be used for diagnosis, prognosis, and treatment purposes. These studies are some examples of researches that apply data mining and deep learning to medical fields for prediction of diseases.

2.1 PROPOSED SYSTEM

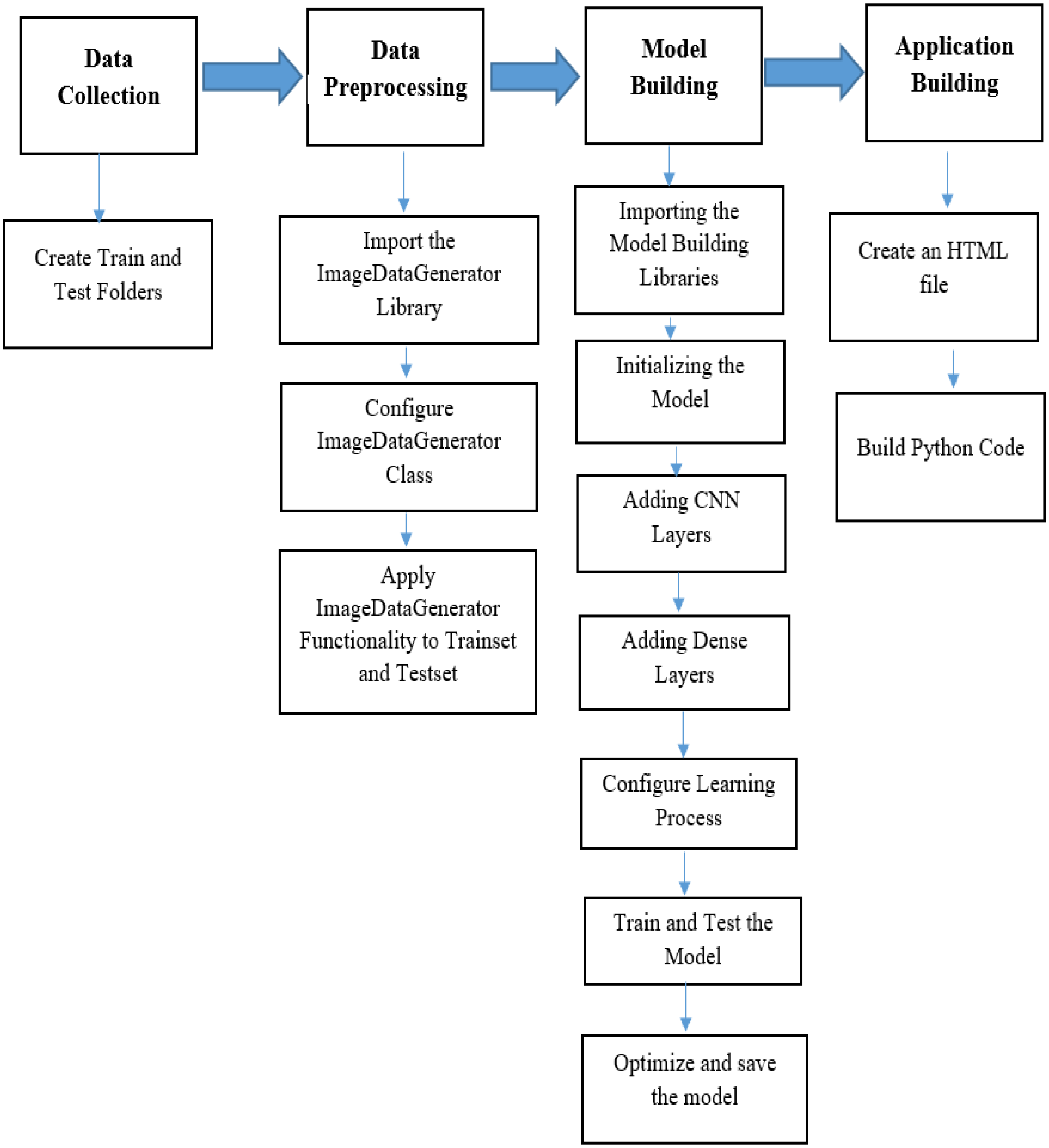
Deep Learning (Convolutional Neural Networks):

In this proposed system we are able to identify the rate of breast cancer. By prediction of Breast cancer in healthcare we are able to take suitable measures accordingly to control the cancer tissue. Due to the large size of each image in the training dataset, we propose a technique which consists of two consecutive convolutional neural networks. By using deep neural networks we are predicting if the person has cancer or not. Users have feasibility to upload scanned images on a web page to know about the status.

3. THEORETICAL ANALYSIS

It is important to detect breast cancer as early as possible. Manual detection of a cancer cell is a tiresome task and involves human error, and hence computer-aided mechanisms are applied to obtain better results as compared with manual pathological detection systems. In **deep learning**, this is generally done by extracting features through a convolutional neural network (**CNN**) and then classifying using a fully connected network.We have trained a convolutional neural network and obtained a prediction accuracy of up to 83.19%. CNN is a modified variety of deep neural net which depends upon the correlation of neighbouring pixels. It uses randomly defined patches for input at the start, and modifies them in the training process. Once training is done, the network uses these modified patches to predict and validate the result in the testing and validation process. Convolutional neural networks have achieved success in the image classification problem, as the defined nature of CNN matches the data point distribution in the image. As a result, many image processing tasks adapt CNN for automatic feature extraction.

3.1 Block Diagram



3.2 Software Designing

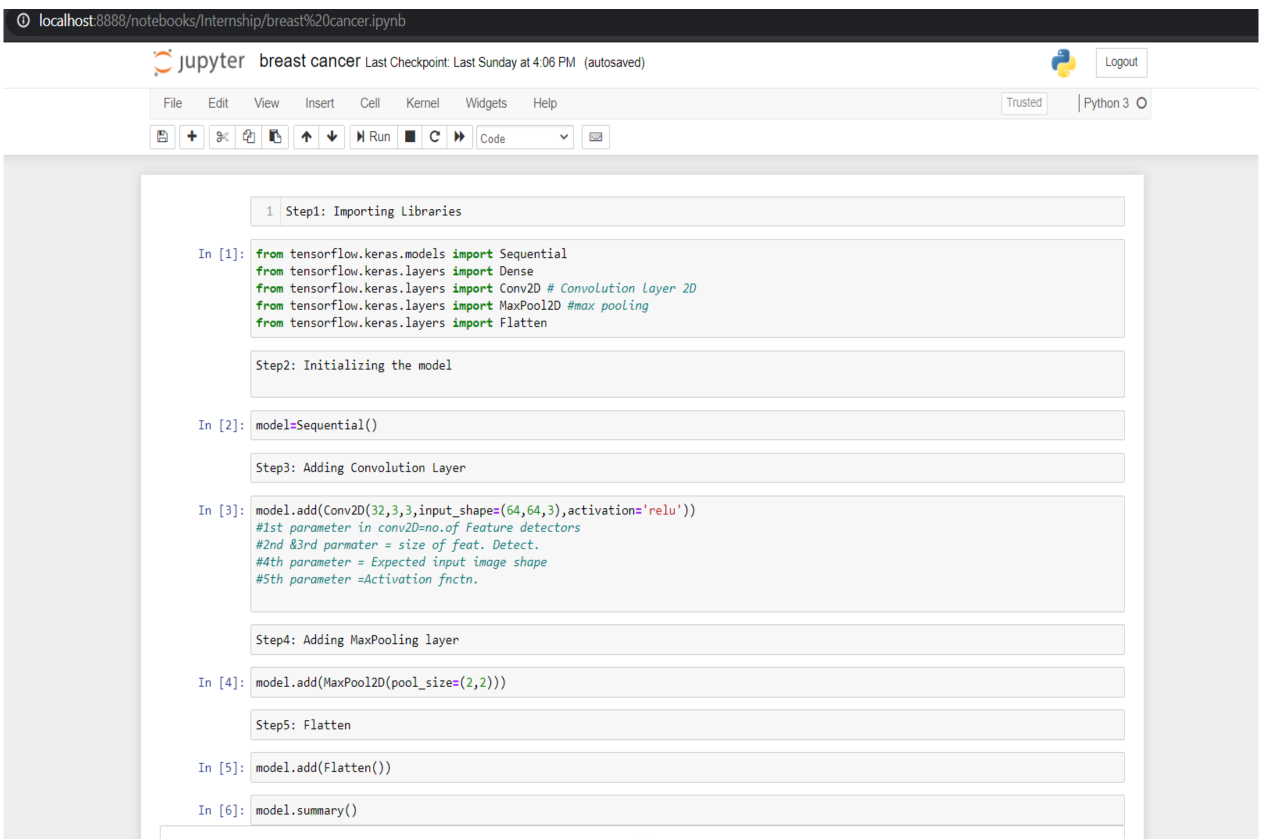
- Jupyter Notebook Environment
- Spyder
- Deep Learning Algorithms
- Python(Sequential,Dense,Conv2D,MaxPool2D,Flatten)
- HTML

- Flask

We developed this breast cancer prediction by using the Python language, which is a high level programming language along with Deep Learning Algorithm such as CNN. For coding we used the Jupyter Notebook of Anaconda distributions and Spyder, an integrated scientific programming in python language. Flask is used as a user interface for the prediction. Hypertext Markup Language (***HTML***) is the standard markup language for documents designed to be displayed in a web browser.

4. EXPERIMENTAL INVESTIGATION

In our project, we have used the Breast Cancer Dataset. This dataset contains two folders: test_set and training_set. In test_set folder, we have two categories called Positive and Negative, where, Positive has the images having breast cancer and Negative has the images which doesn't have breast cancer. Similarly in the training_set folder. Having 202 images belonging to 2 classes and 59 images belonging to 2 classes.



localhost:8888/notebooks/Internship/breast%20cancer.ipynb

jupyter breast cancer Last Checkpoint: Last Sunday at 4:06 PM (autosaved)

File Edit View Insert Cell Kernel Widgets Help

```
Step6: ANN layers

In [7]: model.add(Dense(units=128,activation='relu',kernel_initializer='random_uniform'))

In [8]: model.add(Dense(units=1,activation='sigmoid',kernel_initializer='random_uniform'))

In [9]: model.summary()

In [10]: model.compile(optimizer='adam',loss='binary_crossentropy',metrics=['accuracy'])

In [11]: # loading the dataset

In [12]: from keras.preprocessing.image import ImageDataGenerator

Using TensorFlow backend.

In [13]: train_datagen=ImageDataGenerator(rescale=1./255,shear_range=0.2,horizontal_flip=True,zoom_range=0.2)
         test_datagen=ImageDataGenerator(rescale=1./255)

In [14]: x_train=train_datagen.flow_from_directory(r'C:\Users\User\Internship\Breast cancer dataset\dataset\training_set',target_size=(64
         x_test=test_datagen.flow_from_directory(r'C:\Users\User\Internship\Breast cancer dataset\dataset\test_set',target_size=(64,64),b

Found 202 images belonging to 2 classes.
Found 59 images belonging to 2 classes.

In [15]: print(x_train.class_indices)

{'Negative': 0, 'Positive': 1}

In [16]: model.fit_generator(x_train,steps_per_epoch=202,validation_data=x_test,epochs=40,validation_steps=59)

In [17]: model.save('testmodel3.h5')
```

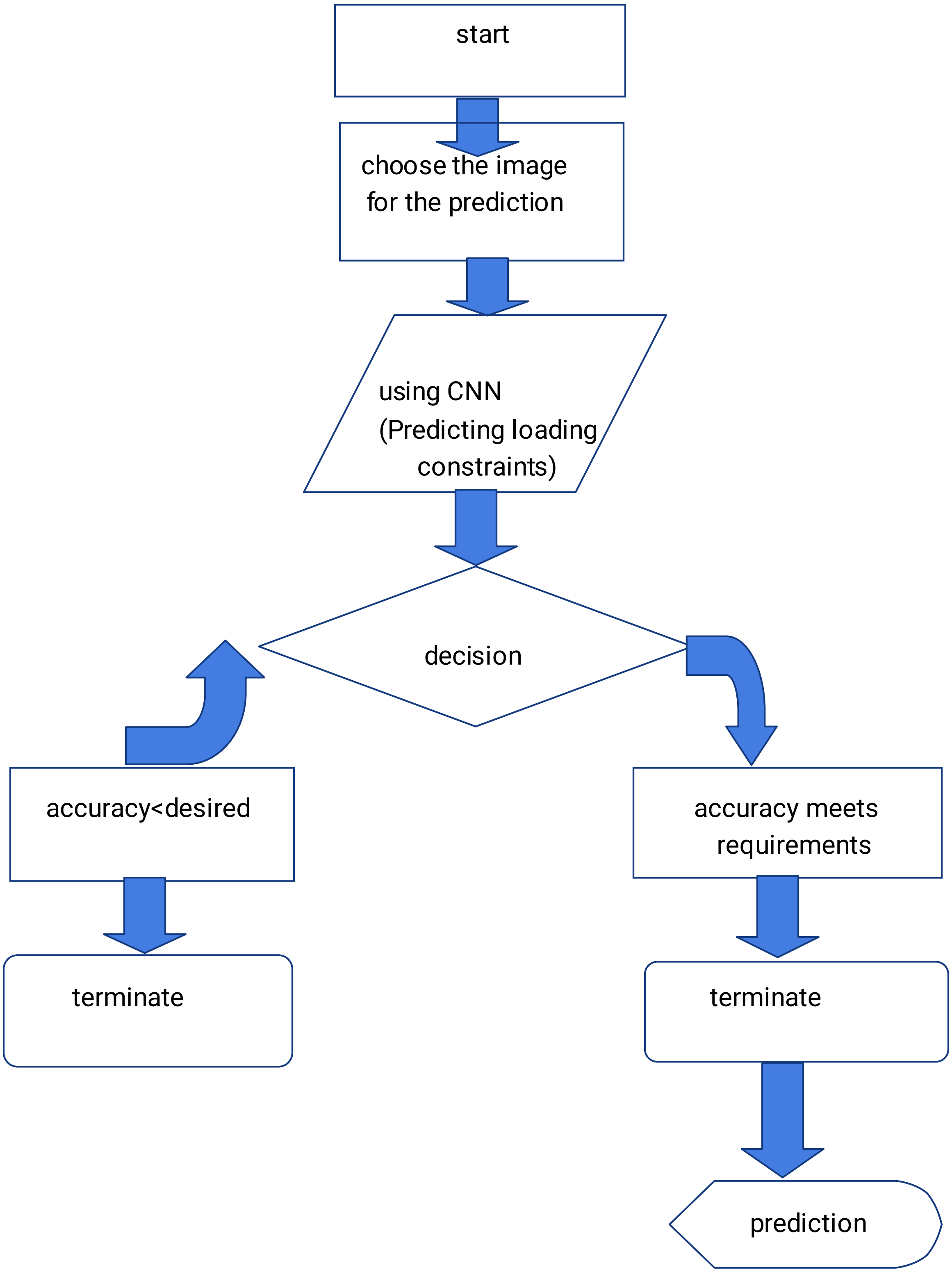
localhost:8888/notebooks/Internship/breast%20cancer.ipynb

jupyter breast cancer Last Checkpoint: Last Sunday at 4:06 PM (autosaved)

File Edit View Insert Cell Kernel Widgets Help

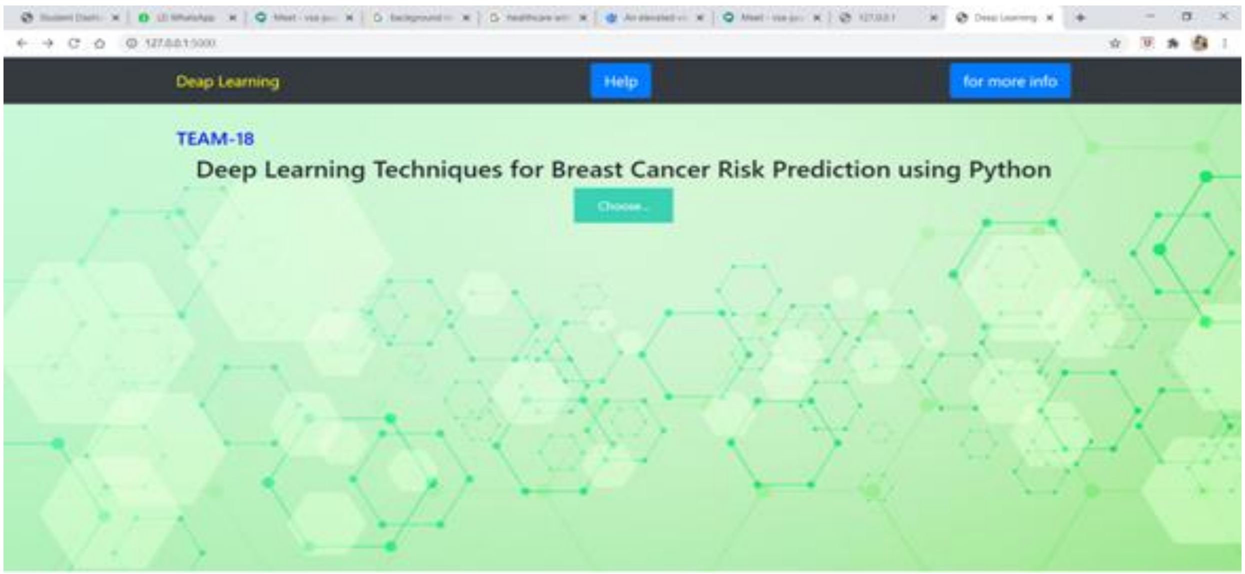
```
x_test=test_datagen.flow_from_directory(r'C:\Users\User\Internship\Breast cancer dataset\dataset\test_set',target_size=(64,64),b

Found 202 images belonging to 2 classes.
Found 59 images belonging to 2 classes.

In [15]: print(x_train.class_indices)

{'Negative': 0, 'Positive': 1}

In [16]: model.fit_generator(x_train,steps_per_epoch=202,validation_data=x_test,epochs=40,validation_steps=59)

Epoch 35/40
202/202 [==============================] - 39s 192ms/step - loss: 0.0123 - accuracy: 0.9950 - val_loss: 1.4462 - val_accurac
y: 0.7969
Epoch 36/40
202/202 [==============================] - 43s 215ms/step - loss: 0.0036 - accuracy: 0.9990 - val_loss: 1.3645 - val_accurac
y: 0.8296
Epoch 37/40
202/202 [==============================] - 42s 208ms/step - loss: 7.5068e-04 - accuracy: 1.0000 - val_loss: 1.5452 - val_accu
racy: 0.8302
Epoch 38/40
202/202 [==============================] - 40s 199ms/step - loss: 0.0062 - accuracy: 0.9985 - val_loss: 1.3106 - val_accurac
y: 0.8474
Epoch 39/40
202/202 [==============================] - 39s 191ms/step - loss: 0.0498 - accuracy: 0.9873 - val_loss: 1.5582 - val_accurac
y: 0.7986
Epoch 40/40
202/202 [==============================] - 40s 196ms/step - loss: 0.0023 - accuracy: 1.0000 - val_loss: 1.3609 - val_accurac
y: 0.8319

Out[16]: <tensorflow.python.keras.callbacks.History at 0x1b7345c93c8>

In [17]: model.save('testmodel3.h5')
```

5. FLOWCHART :

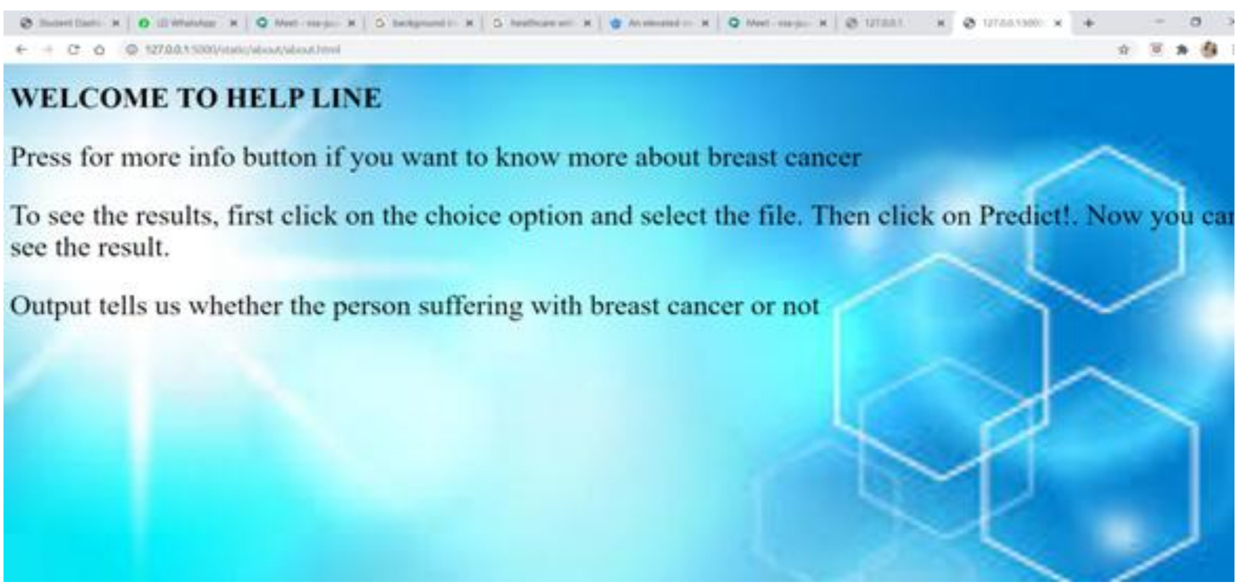


6. RESULT :

In this paper, the CNN algorithm is used to predict its performance.The results shows 83.19% accuracy after running 40 epochs.



<u>Fig:1-Home Page</u>



<u>Fig:2-Help option</u>

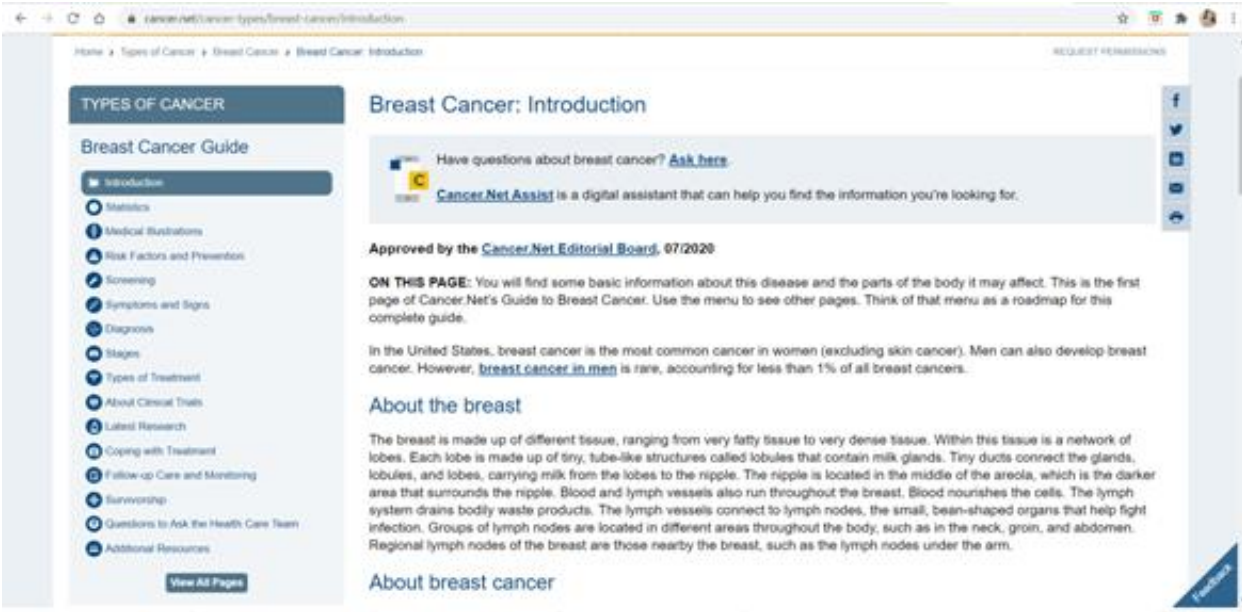


Fig:3-when you clicked on for more info button it will redirect to https://www.cancer.net/cancer-types/breast-cancer/introduction this website.

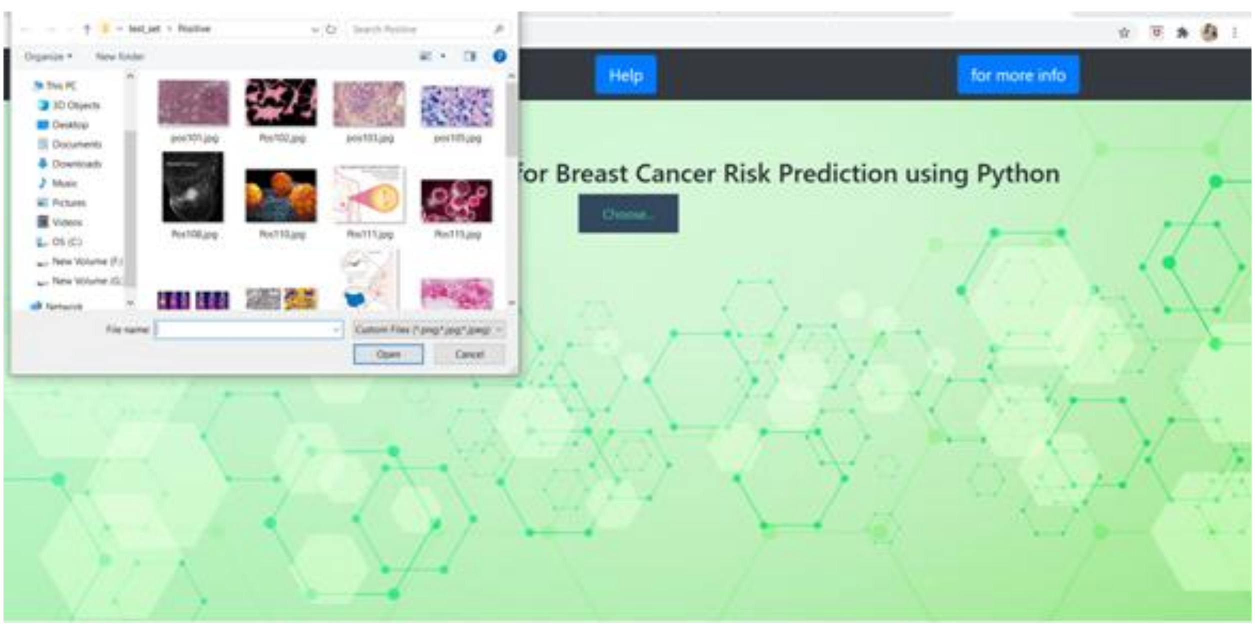


Fig:4-when you pressed choose, it will ask you to choose a file from the localhost.

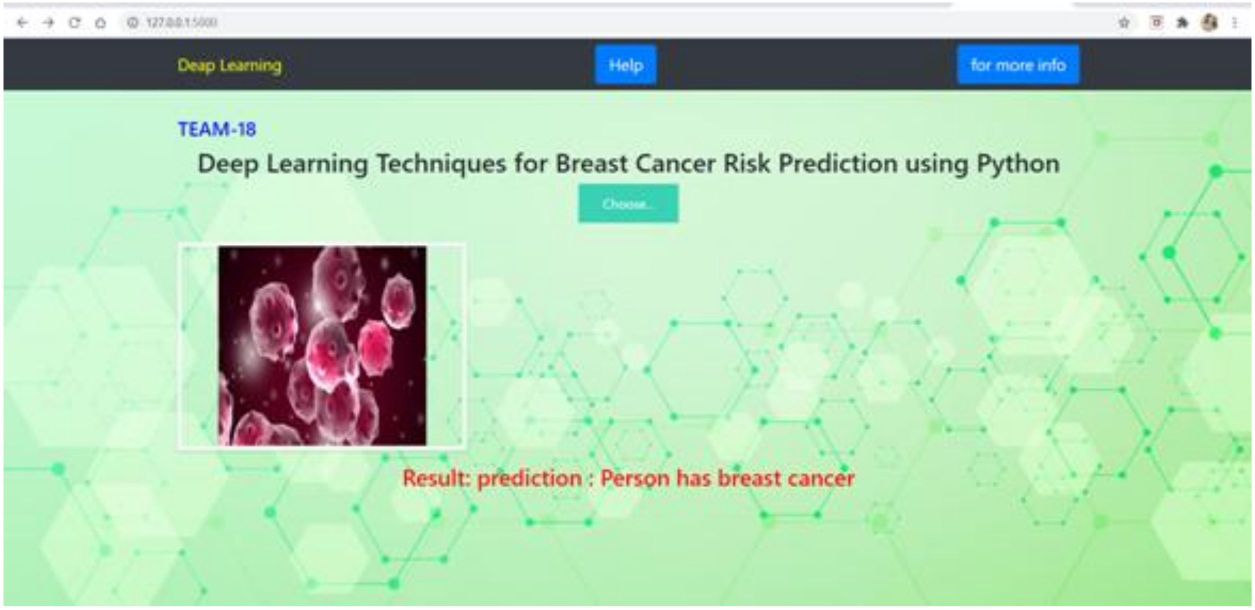


Fig:5-click on **predict** to see the output.(positive prediction)

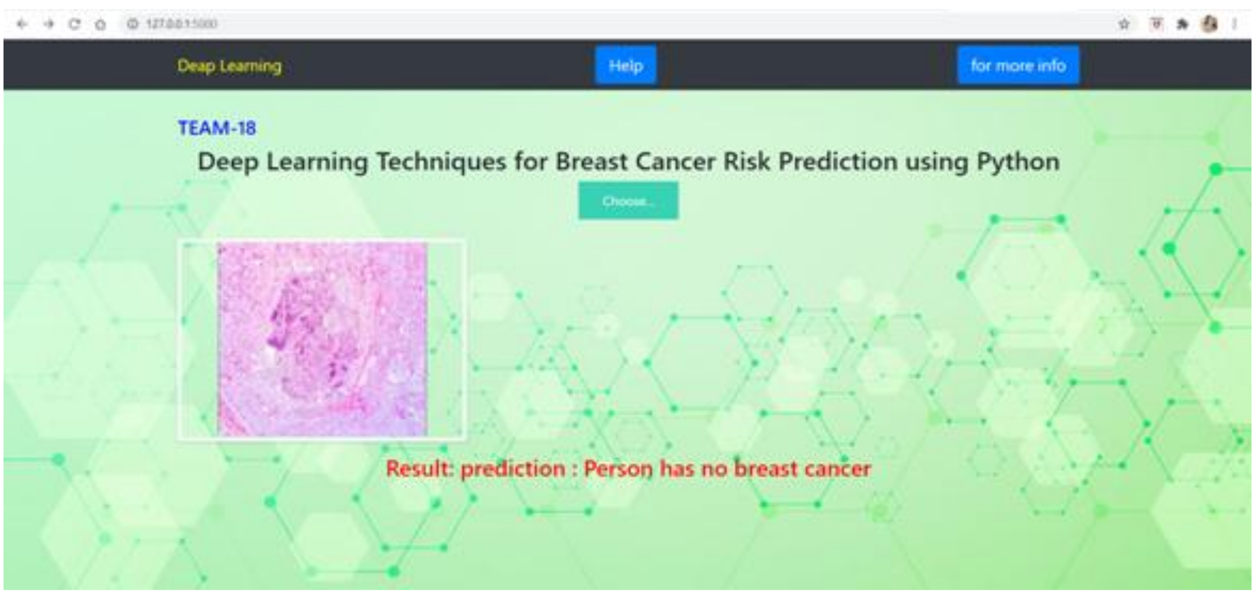


Fig:6- negative prediction output

7. ADVANTAGES AND DISADVANTAGES

ADVANTAGES:

- **Reduces the risk of dying from breast cancer:**

Of 1,000 women who have a mammogram every 2 years for 20 years, 7 deaths are prevented.

- **Reduces the risk of having to undergo chemotherapy:**

Screening often allows for the detection of cancers at an early stage of development. Treatment is then possible without chemotherapy.

- **Allows women to know the health of their breasts:**

The vast majority of women (nearly 98 %) will not have breast cancer if their mammograms and additional examinations do not reveal cancers.

DISADVANTAGES:

- **Periods of waiting and anxiety when additional examinations are required:**

Almost half the women who participate in the screening for 20 years (453 in 1,000) have at least one additional examination. This represents 156 more women than in the 1,000 who do not participate in the screening.

- **Possible overdiagnosis:**

Of 77 breast cancer diagnoses, 10 would be cases of overdiagnosis.

Overdiagnosis is the discovery of a cancer that would never have been detected without screening. It can happen that a woman receives a diagnosis for cancer that would never have had an effect on her health or consequences on her life – like a cancer that develops very slowly or a benign cancer. This could happen to participants in the screening program because a mammogram detects breast cancer in the early stages of development.

However, given that it is still impossible to differentiate between harmless cancers and deadly ones, all cancers are treated. As such, women in the screening program could:

- Receive treatment that would not be necessary
- Suffer the side effects of these treatments
- Have to live with the experience of having been diagnosed with cancer
- Have frequent medical appointments to ensure that the cancer does not return

8. APPLICATIONS

• As our logistic regression, linear discriminant analysis, and neural network models with the broader set of inputs effectively predicted five-year breast cancer risk, these models could be used to inform and guide screening and preventative measures.
• Our models could easily be incorporated into phone application or website breast cancer risk prediction tools. Using our models as such would be convenient and cost-effective as our personal health data inputs are easy and inexpensive to obtain from electronic health records or an office visit. While the purpose of such a tool would not be to replace physician advice or mammograms, risk predictions derived from our models could contribute to both early breast cancer detection and breast cancer prevention.
• Women who receive high estimated risks could be motivated to seek out a doctor or take other preventative actions. Models could be used to guide immediate personal decisions such as screening.
• Doctors could also use these models to inform decisions on whether and when to recommend long-term preventative actions such as chemoprevention and hormone replacement therapy.

9. CONCLUSION

Breast cancer if found at an early stage will help save lives of thousands of women or even men. These projects help the real world patients and doctors to gather as much information as they can. The research on nine papers has helped us gather the data for the project proposed by us. By using machine learning algorithms we will be able to classify and predict the cancer into being or malignant.
Machine learning algorithms can be used for medical oriented research, it advances the system, reduces human errors and lowers manual mistakes.

10. FUTURE SCOPE

The model for supervised learing can be iterated so that they can achived the higher accuracy.Parameters of these classifeication techniques can be tweaked to optimized. Overall due to less number of samples ,it was difficult to build the models with higher accuracy. Hence as more data comes in, these models can be reiterated. Also ,GUI can be made more comprehensive and more flexiblity can be provided to these model for the end user to tweak the parameters visualize the results

11.BIBLIOGRAPHY

- American Cancer Society .Cancer facts and figures 2017.Avaliable online at: www.cancer.org/research/cancer-facts-satistics/all-cancer-facts-figures/cancer-facts-figures-2017.html.Last accessed july 20,2017

- Susan G.Komen Northeast Ohio.Executive summary.Susan. G.Komen Community Profile-Report2015.Avaliable,online.at:komenno-hio.org/wp-content/uploads/2014/12/Komen-North-east-Ohio-2015-Community-profile-Report.pdf.Laste accessed july 18,2017

- National Cancer Institute.SEER state cancer statistic review 1975-2014. Avaliable online at:seer.cancer.gov/cov/1975-2014.Last accessed july 18,2017

- National-Institute.SEER,state-cancer-profiles.Avaliableonline,at:statecancerprofiles.cancer.gov/.Last accessed july18,2017

- Institute of Medicine.Delivering High_Quality Cancer Care :Charting a New Course for a System in Crisis.Washington, DC : National Academies Press; 2013.Avaliable online at:www.nap.edu/cata-log/18359/delivering-high-quality-cancer-care-charting-new-course-for Last accessed December 26,2017.

- Breast Cancer Initiative 2.5.Planning:improving access to breast cancer .Available online at:fredhuct.org/en/labs/phs/projects/breast-can-cer-initiative_2.5.html%20# content .Last accessed july 25,2017

- Kolcaba K.Comfort Theory and Practice: A Vision for Hoslistic Health Care and Research. New york , NY: Springer; 2002

- Nardi EA, Wolfson JA ,Rosen ST, et al Value, access ,and cost of cancer care delivery at academic cancer centers .J Netw Compr Canc Netw.2016;14(7):837-847.

11. APPENDIX

INDEX.HTML :

```
{% extends "base.html" %} {% block content %}
<style type="text/css">
body {
   width=100%;
   height: 100%;
  background-
image:url("https://healthtechmagazine.net/sites/healthtechmagazine.net/files/styles/c
dw_hero/public/articles/%5Bcdw_tech_site%3Afield_site_shortname%5D/202003/2020
0326_HT_Web_HIMSS_Nutanix_Cloud.jpg?itok=-34HbJSi");
   background-position: centre;
   background-repeat: no-repeat;
   background-size: cover;
}
</style>
<h4 style="color:blue">TEAM-18</h4>
<h2 style= "text-align:center" >Deep Learning Techniques for Breast Cancer Risk
Prediction using Python</h2>

<style>
   #align{
         text-align:center;
        }
   #but{
         text-align:center;
      }
</style>

<div id="align">
   <form id="upload-file" method="post" enctype="multipart/form-data">
      <label for="imageUpload" class="upload-label">
         Choose...
       </label>
      <input type="file" name="file" id="imageUpload" accept=".png, .jpg, .jpeg">
   </form>
```

```
    <div class="image-section" style="display:none;">
        <div class="img-preview">
            <div id="imagePreview">
            </div>
        </div>
        <div id="but">
            <button type="button" class="btn btn-primary btn-lg " id="btn-predict" style="text-
align:center">Predict!</button>
        </div>
    </div>

    <div class="loader" style="display:none;"></div>

    <h3 style="color:red" "text-align:center"  id="result">
        <span> </span>
    </h3>

</div>

{% endblock %}
```

APP.PY :

```
from _future_ import division, print_function
import sys
import os
import glob
import numpy as np
from keras.preprocessing import image
from keras.applications.imagenet_utils import preprocess_input, decode_predictions
from keras.models import load_model
from keras import backend
from tensorflow.keras import backend
from keras.initializers import glorot_uniform

import tensorflow as tf
global graph
```

```
tf.compat.v1.disable_eager_execution()
graph=tf.compat.v1.get_default_graph()

#global graph
#graph = tf.get_default_graph()


from skimage.transform import resize

# Flask utils
from flask import Flask, redirect, url_for, request, render_template
from werkzeug.utils import secure_filename
from gevent.pywsgi import WSGIServer

# Define a flask app
app = Flask(_name_)

# Model saved with Keras model.save()
#MODEL_PATH = 'models\mymodel3.h5'

# Load your trained model
#model = load_model(os.path.join(MODEL_PATH))
#model = load_model(os.path.join("mymodel.h5"))

#model=tf.keras.models.load_model(r"C:\Users\User\Internship
2020\cnnflask\models\mymodel.h5")
model_path="models/mymodel.h5"
model=tf.keras.models.load_model(model_path)
# Load your trained model
#model = load_model(mymodel3)
    # Necessary
# print('Model loaded. Start serving...')

# You can also use pretrained model from Keras
# Check https://keras.io/applications/
#from keras.applications.resnet50 import ResNet50
#model = ResNet50(weights='imagenet')
```

```
#model.save('')
print('Model loaded. Check http://127.0.0.1:5000/')




@app.route('/', methods=['GET'])
def index():
    # Main page
    return render_template('index.html')

@app.route("/about")
def about():
  return render_template("about.html")
  return render_template('index.html')



@app.route('/predict', methods=['GET', 'POST'])
def upload():
    if request.method == 'POST':
        # Get the file from post request
        f = request.files['file']

        # Save the file to ./uploads
        basepath = os.path.dirname(_file_)
        file_path = os.path.join(
            basepath, 'uploads', secure_filename(f.filename))
        f.save(file_path)
        img = image.load_img(file_path, target_size=(64, 64))
        x = image.img_to_array(img)
        x = np.expand_dims(x, axis=0)
        with graph.as_default():
            preds = model.predict_classes(x)
        index = ['Negative, does not have breast cancer','Positive, has a breast cancer']
        text = "prediction : "+index[preds[0][0]]
        if preds[0][0]==0:
```

```
        prediction="Person has no breast cancer"
    else:
        prediction="Person has breast cancer"
    text = "prediction : "+prediction

    return text

            # ImageNet Decode

if _name_ == '_main_':
    app.run(debug=False,threaded = False)
```