BREAST CANCER RISK PREDICTION

using deep learning techniques-convolution neural networks

Developed by:Pulyala Sony,Puskuri Jyothi,Aedavelli Sai Roshini,Yenreddy Bikshapathi

Smart Bridge - Remote Summer Internship Program

1.INTRODUCTION

Most common type of medical hazard found in the middle aged women is,breast cancer.Mortality rate of women due to breast cancer can be reduced if can be detected at a relatively early stage. With the help of latest, efficient and advanced screening methods, the majority of such cancers are diagnosed when the disease is still at a localized stage [1]. The utility of machine learning techniques in healthcare analysis is growing progressively . Certainly analysis of patient's clinical data and physician's judgment are the most considerable features in diagnosis. Most of the possible medical flaws can be avoided by the using classification systems, and also offer healthcare data to be analyze in lesser time and in more exhaustive manner. Accurate and timely prediction of breast cancer allows physicians and healthcare providers to make most favorable decision about the patient treatment. Of all the cancers, breast cancer constitutes of 11.6% in new cancer cases and come up with 24.2% of cancers among women. In case of any sign or symptom, usually people visit doctor immediately, who may refer to an oncologist, if required. The oncologist can diagnose breast cancer by: Undertaking thorough medical history, Physical examination of both the breasts and also check for swelling or hardening of any lymph nodes in the armpit.

A. Imaging tests:

Mammogram, Magnetic resonance imaging (MRI) of breast, Ultrasound of breastX-ray of the breast, Tissue biopsy: Removal of the tissue of the breast for examination by a pathologist. Sentinel node biopsy: Once breast cancer is confirmed, patients regularly undergo sentinel node biopsy. This helps to detect cancerous cells in lymph nodes to confirm metastasis of breast cancer into lymphatic system. If required, oncologist may also order additional tests or procedures. In the conventional way of diagnosing breast cancer some tests and procedures are carried out. These tests include Breast exam Mammogram Breast ultrasound Biopsy. As an alternative we can also use Machine Learning techniques for the classification of benign and malignant tumors. The prior diagnosis of Breast Cancer can enhance the prediction and survival

rate notably [1], so that patients can be informed to take clinical treatment at the right time. Classification of benign tumors can help the patients avoid undertaking needless treatments. Thus the research is to be carried for the proper diagnosis of Breast Cancer and categorization of patients into malignant and benign groups. Machine Learning, with its advancements in detection of critical featuresfrom the complex datasets is largely acknowledged as the method in the prediction of breast cancer. Application of data mining techniques in the medical field can help in prediction of outcomes, minimizing the cost of medicines, aid people's health, upgrade the healthcare value and to rescue lives of people. This process of classifying benign and malignant tumors can be best done by the application of Classification techniques of machine learning and deep learning. Lot of research is being conducted in this area by the application of various machine learning,deep learning and data mining techniques for many different datasets on Breast Cancer. Most of them show that classification techniques give a good accuracy in prediction of the type of tumor.

1.1 Overview

Breast cancer became the major source of mortality between women.The accessibility of healthcare datasets and data analysis promote the researchers to apply study in extracting unknown pattern from healthcare datasets. The intention of this study is to design a prediction system that can predict the incidence of the breast cancer at early stage by analyzing through the set of training and testing images that has been selected from clinical datasets.Wisconsin breast cancer dataset (WBCD) have been used to conduct the proposed experiment.Data Mining is one of the most motivating and vital area of research with the aim of extracting information from tremendous amount of accumulated data sets.Here a new model for classifying type of breast cancer by using deep learning concepts and as alternative machine learning concepts can also be used for classifying whether the cancer is malignant or benign. The model has been built using data from imaging test and MRI ,ultrasound x-ray of breast to predict the type of breast cancer. By using the neural networks a Flask model has been implemented and tested.The potential of the proposed method is obtained using Imagedata generator class which was obtained by preprocessing the training and testing images to predict the image by a series of random translations. The outcome confirms that the maximum classification accuracy (85%) and configures the learning process by optimizing and saving the model ,is achieved for this study.

1.2 Purpose

Our aim from the project is to make use of numpy,keras,tensorflow ,Imagedatagenerator libraries from python to extract the libraries for deep learning for the prediction of breast cancer. Secondly, to learn how to hyper tune the parameters

using image search cross validation for the convolution neural networks algorithm.And in the end, to predict whether the type of cancer is benign or malignant ensemble techniques of combining the predictions from deep learning algorithms and withdrawing the conclusions.

2. LITERATURE SURVEY

Deep Learning is the process of analyzing data from different perspectives and extracting useful knowledge from it. It is the core of knowledge discovery process..
Different data learning techniques include

- Convolutional **Neural Networks**
- Recurrent **Neural Networks** (RNNs)
- Generative Adversarial Networks
- Self-Organizing Maps
- Boltzmann Machines
- Deep **Reinforcement Learning**
- Autoencoders.

Convolution neural network is the most applied deep learning technique,which detects predictions based on image classification and uses object detection code to predict image through video analysis.Identifying type of cancer and identifying the object based on images are particularly well suited to deep learning techniques.In CNN,training set is used to build the model as the Imagedatagenerator class which can classify the data items into its images into appropriate classes. A test set is used to validate the model.

2.1 Existing Problem

Lots of breast cancer research has been reported in the literature of medical data analysis, and most of them turn up with good classification accuracies. Polat et al , proposed LS-SVM classifier algorithm for the diagnosis of breast cancer and achieved the classification accuracy of 98.53% using 10-fold cross validation. Akay, proposed a new method for the breast cancer diagnosis using support vector classification algorithm on the most predictive features and obtain the classification accuracies of 99.02% without cross-validation . Yeh et al. present a innovative technique for breast cancer detection, by using statistical methods in combination with swarm optimization and reported the accuracy of 98.71%. Marcano-Cede~no et al. proposed a new method AMMLP for the classification of breast cancer datasets by using an Artificial Neural Network over the biological metaplasticity property and acquire classification accuracy of 99.26 . Kaya and Uyar in their work presented a hybrid approach for the detecting

hepatitis disease by means of rough set and extreme machine learning algorithm. The selected hepatitis disease dataset was from UCI repository. 20 reducts containing three to seven attributes were produced by using rough set theory. The reducts are selected and then the records with missing value are removed from each reduct. Classification is done on selected reducts by using back propagation neural network and obtained the accuracy of 98.6%. All of the above listed works are just a small representative of the existing huge number of research in utilizing machine learning and data mining techniques to a range of healthcare domains for forecasting and pattern recognition purposes.

2.2 Proposed Solution

Deep Learning (Convolutional Neural Networks):

In vision of the problem statement described in the introduction section, a CNN model is proposed with boosted accuracy to predict the breast cancer patient. The framework is composed of the following important phases:

- Dataset Collection(creating training and testing folders)
- Data Preprocessing.
- Model Building.
- Achieving trained model with highest accuracy
- Using trained model for prediction
- Application Building

Classifications of the data sets are done on the basis of specific properties posses by the sample variable that the capable to classify them, and each sample variable is assigned a malignant or benign class. Classification is principally done by making predictions based on known sample data that has been learned from training data. Designed algorithm is first trained on the known data labels and further uses this learning to predict the class labels for the new unknown set of data sample. The classification objective set for this study is to achieve enhanced accuracy by using Imagedatagenerator classifiers and determine which one suits the most for breast cancer classification technique. We train the classifier with known sample data in a training dataset and check its performance by examining the test .

And also we have created an UI using the Flask for the breast cancer prediction, this UI will allow the users to predict the type of cancer very easily and the User interface is user friendly not at least one complication in using the interface, and it can be used just by entering some necessary details or images into the UI in real time it'll give the predicted value like if the cancer is malignant or benign and how often does he takes the precautions to control the cancer in the body.

Basically this model will give the predicted image when a person with breast cancer will confirm whether the type of cancer is malignant or benign, by just examining

some necessary symptoms of cancer in real time, and those symptoms of person will be examined by the physician and provide some precautions to the person to get rid of cancer sooner.

3. THEORITICAL ANALYSIS

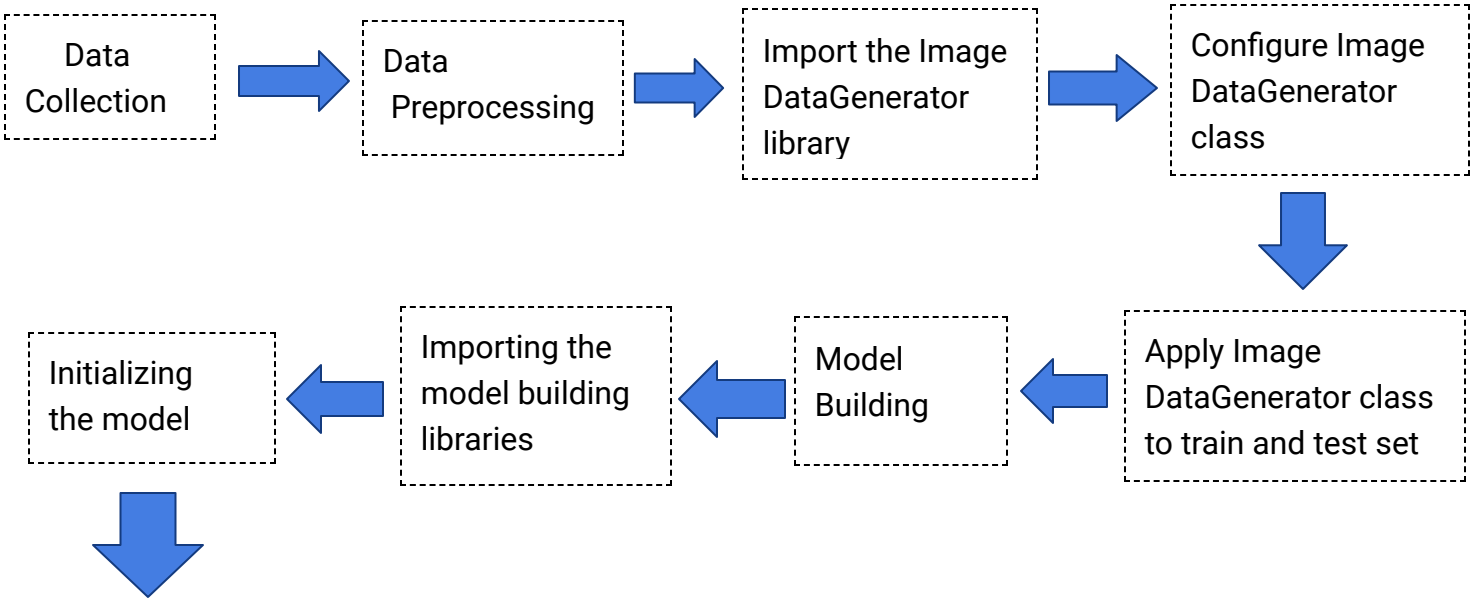
While selecting the algorithm that gives an accurate prediction we gone through lot of algorithms which gives the results abruptly accurate and from them we selected only one algorithm for the prediction problem that is convolution neural networks, it assumes that the presence of a particular feature detector in a imagedatagenerator class detects the image and transforms the image in the batch by a series of random translations and these translations are based on arguments and replaces original batch with new data.and also performs data augmentation. Thats how the prediction work great with the neural networks.

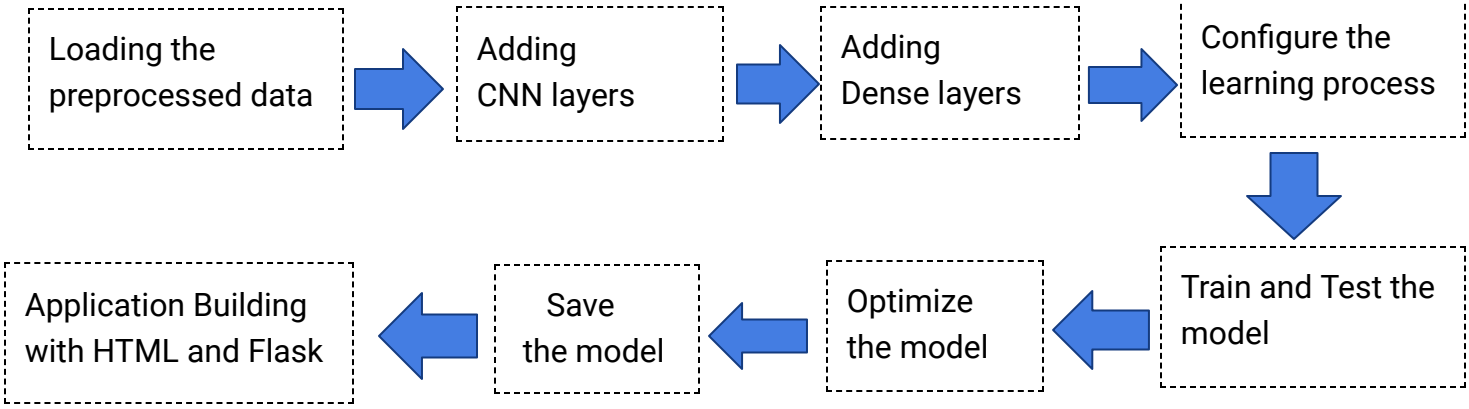
The peculiarity of this problem is collecting the data of cancer affected persons details real time and working with the prediction at the same time, so we developed an user interface for the people who'll be accesssing for the type of breast cancer prediction. Accuracy is defined as the ratio of the number of samples correctly classified by the classifier to the total number of samples for a given test data set. The formula is as follows

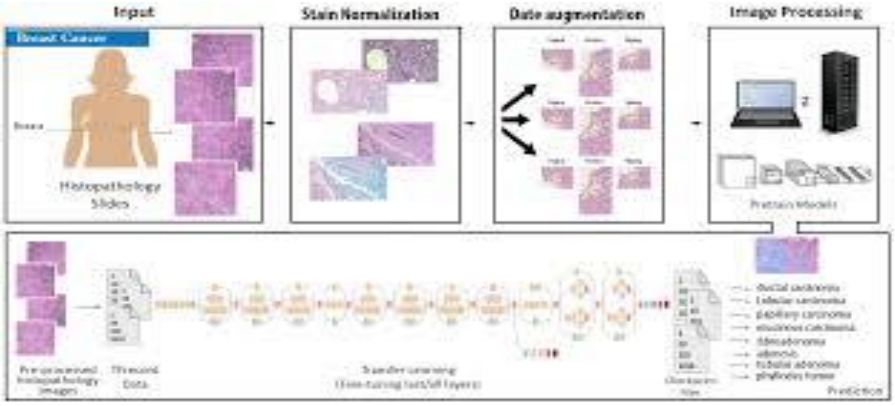
Accuracy=TP+TN/TP+TN+FT+FN

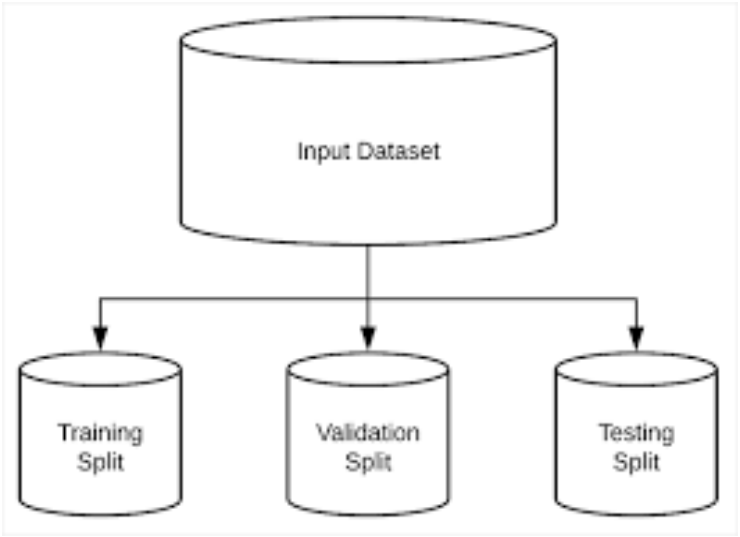
At first we got like lot of worst accuracies because we tried lot of trained images by giving more number of epochs for the best accurate algorithm , finally after all of that we tried the best suitable algorithm which gives the prediction accurately is convolution neural networks. And developed it to use as a real time prediction probelm for the breast cancer risk prediction.

3.1 Block Diagram




Loading the preprocessed data
Adding CNN layers
Adding Dense layers
Configure the learning process
Train and Test the model
Optimize the model
Save the model
Application Building with HTML and Flask


Input
Stain Normalization
Data augmentation
Image Processing


Input Dataset
Training Split
Validation Split
Testing Split

3.2 Software Designing

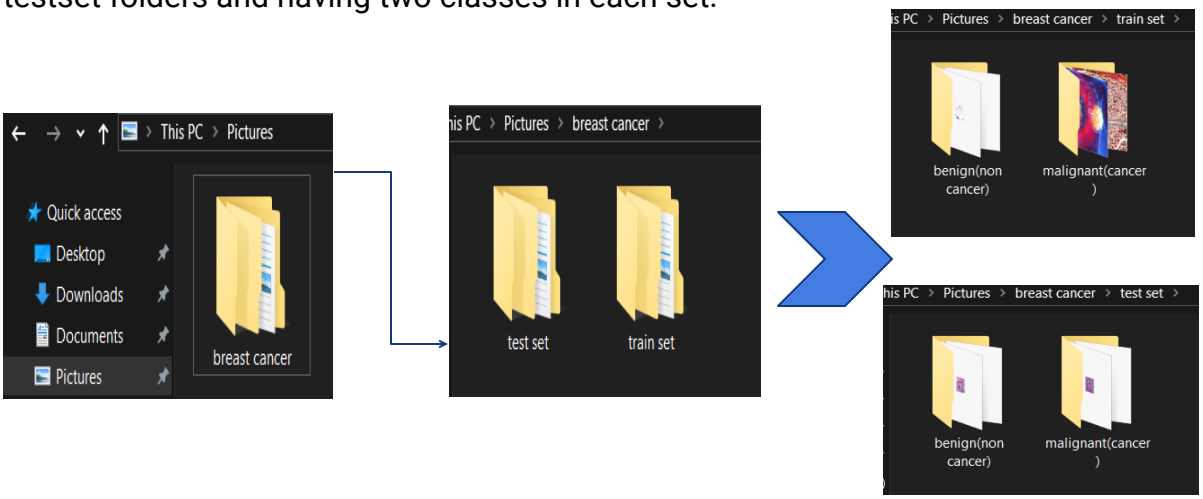
- Jupyter Notebook Environment
- Spyder Ide
- Deep Learning Algorithms
- Python (pandas, numpy,keras,tensorflow)
- HTML
- Flask

We developed this breast cancer risk prediction by using the Python language which is a interpreted and high level programming language and using the Deep Learning algorithms. for coding we used the Jupyter Notebook environment of the Anaconda distributions and the Spyder, it is an integrated scientific programming in the python language.

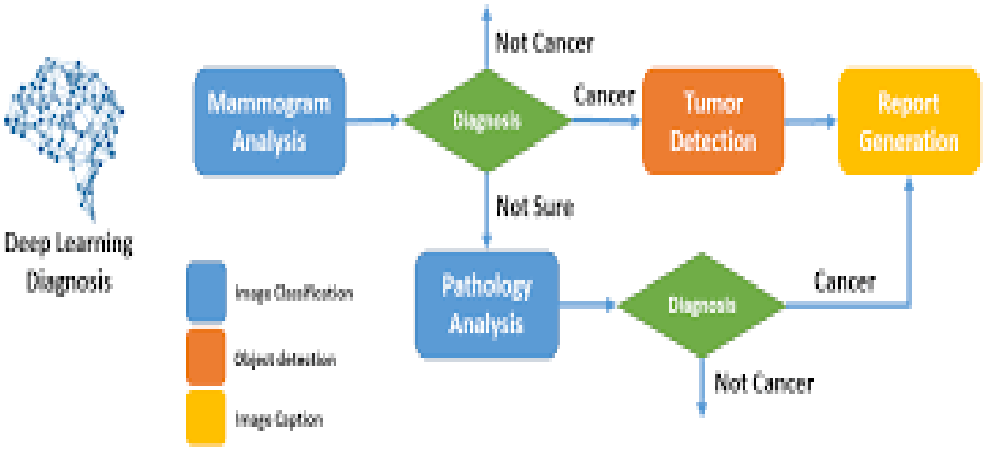
For creating an user interface for the prediction we used the Flask. It is a micro web framework written in Python and uses WSGI for web development. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions, and a scripting language to create a webpage is HTML by creating the templates to use in the functions of the Flask and HTML.

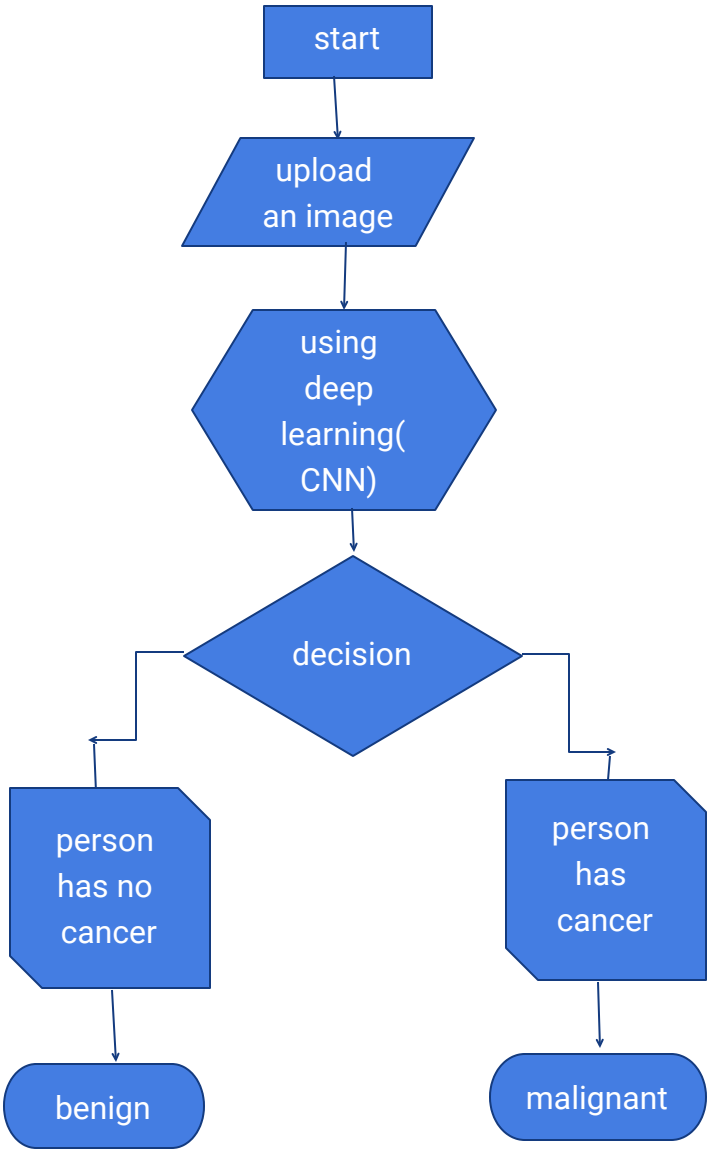
4. EXPERIMENTAL INVESTIGATION

In this experiment,we have used dataset by creating training and testing folders by giving 314 training images into 2 classes i.e malignant(class 1) and benign (class 2) and 136 testing images 2 classes i.e.malignant(class1) and benign(class 0). 70% of images are given for training and 30% of images are given for testing.The screenshots below shows an idea about the data collection by creating trainset and testset folders and having two classes in each set.



5.FLOWCHART

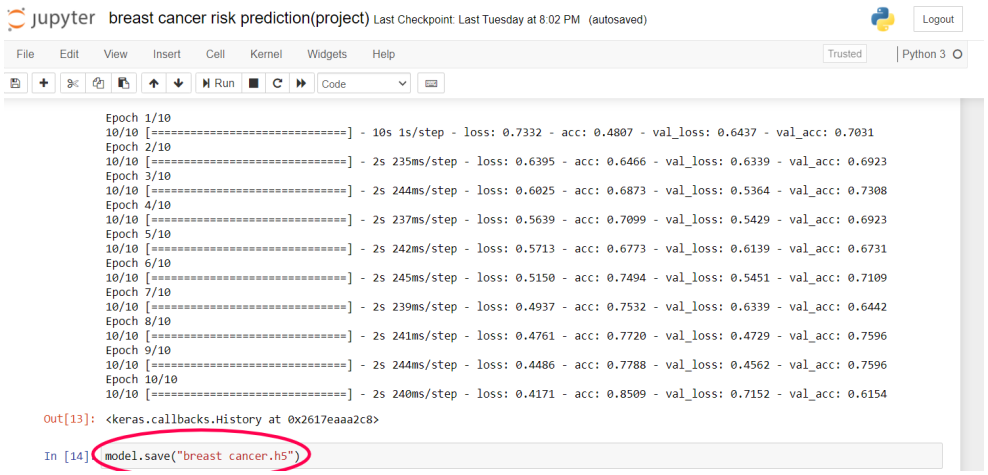

Deep Learning
Diagnosis
Mammogram Analysis
Diagnosis
Not Cancer
Cancer
Tumor Detection
Report Generation
Not Sure
Pathology Analysis
Diagnosis
Cancer
Not Cancer
Image Classification
Object detection
Image Caption

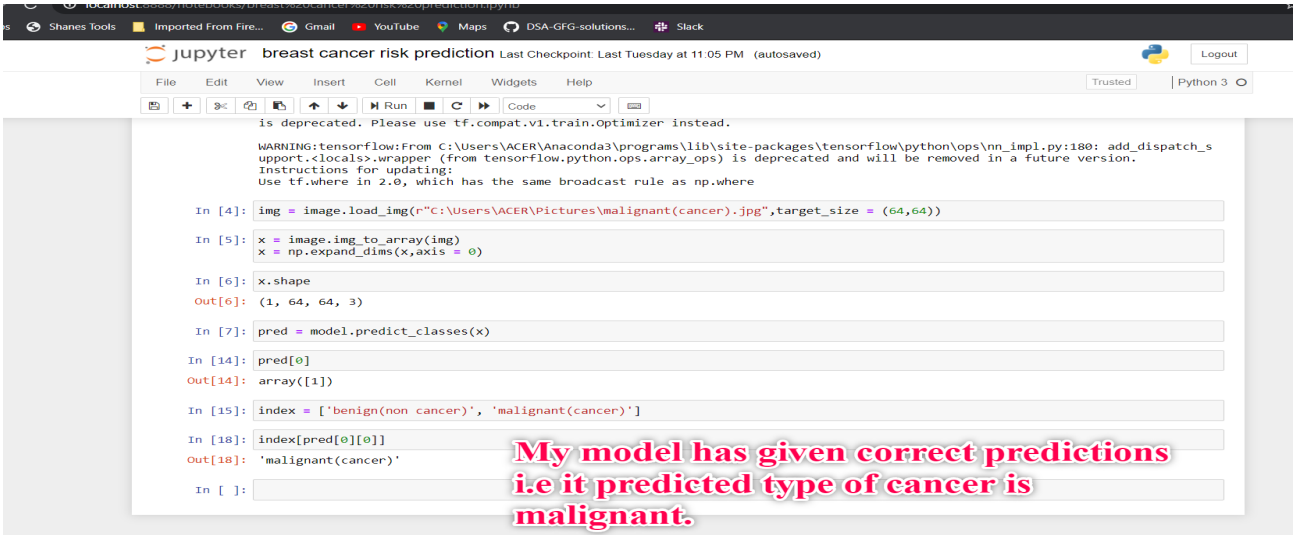

start
upload an image
using deep learning(CNN)
decision
person has no cancer
person has cancer
benign
malignant

6.RESULT

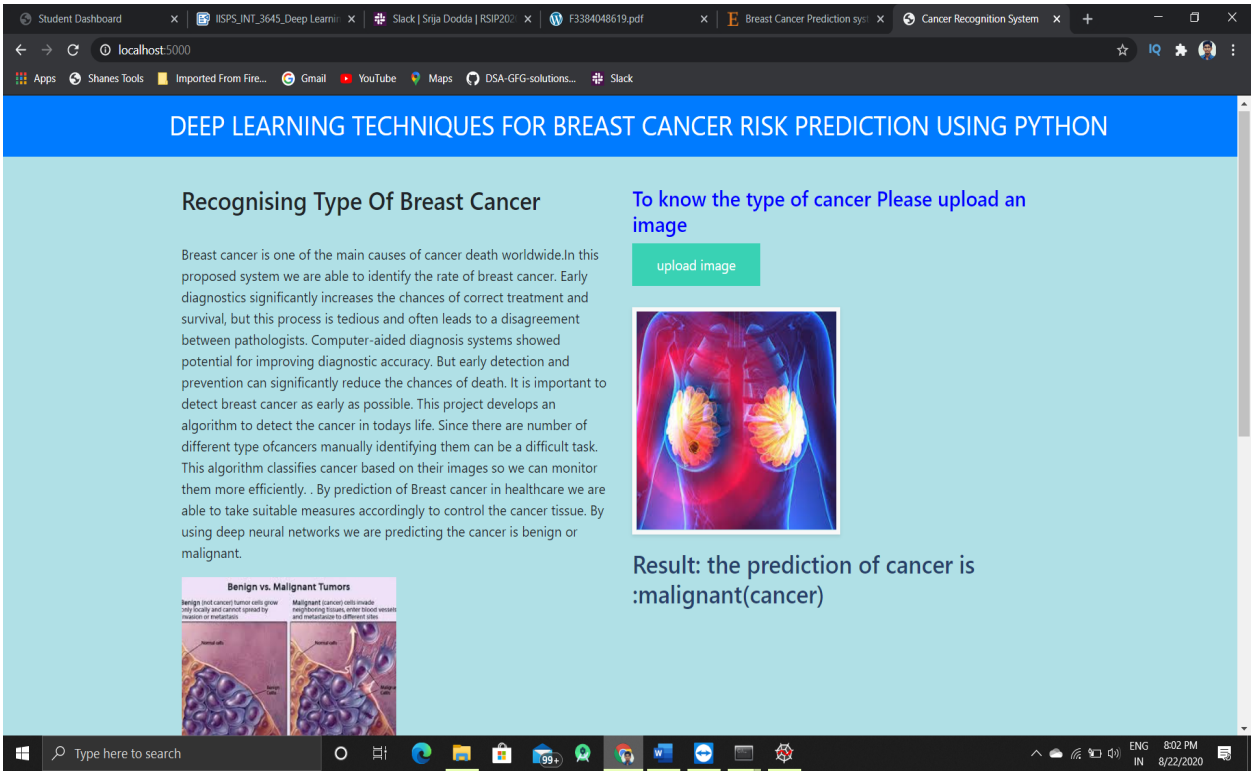
In this paper, the convolution neural networks is used to predict the outcome based on image and video analysis by giving an object detection code.there are many algorithms in machine learning techniques which are used to predict the breast cancer...but the technique I have used in this project is deep learning technique.the neural networks using ImageDataGenerator class and feature detectors it predicts an image whether it is a malignant cancer or benign cancer.By having good amount of training images for training and testing images for testing my model has predicted good accuracy with 85% and validation accuracy as 61%...

firstly I have saved my model and predicted the type cancer whether it is malignant or benign.secondly after getting correct predictions I have built an web application using flask and html.The screenshots below shows the model building and application building in the UI.





The below screenshots indicates the web page development in the UI by using flask and html.....



7. ADVANTAGES AND DISADVANTAGES

Advantages:

- Reduces the risk of dying from breast cancer.
- Reduces the risk of having to undergo chemotherapy.
- Allows women to know the health of their breasts.
- Easy and simple user interface for the people to recognise the type of cancer easily by detecting an image.
- Convolution neural networks gives the accurate result of predictions upto 85% which is we used for prediction.
- It is composed using the HTML and Python for the web usage in real time.
- It can work in real time and predict as soon as the necessary images for prediction are given to the model.

Disadvantages:

- CNN do not encode the position and orientation of object.
- Lack of ability to be spatially invariant to the input data.
- Periods of waiting and anxiety when additional examinations are required.
- Possible overdiagnosis.

8.APPLICATIONS

- In an effort to differentiate between benign and malignant tumours, early work concentrated on the MRI scan of human breast tumour extracts.
- It is one of the most widely used areas of data mining in the health care industry.
- 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.
- 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 the cancer is benign or malignant.
- Users have feasibility to upload scanned images on a web page to know about the status.

➢ **PRECAUTIONS:**

9. CONCLUSION

Our work mainly focused in the advancement of predictive image analysis and video analysis to achieve good accuracy in predicting valid disease outcomes using Deep learning methods. The analysis of the results signify that the integration of data augmentation images along with different ,feature detectors and Imagedatagenerator class to transform the orihginal data intlo new data using random translations. Further research in this field should be carried out for the better performance of the classification techniques so that it can predict on more images.

A decision support system for predicting breast cancer helps and assist physician in making optimum, accurate and timely decision, and reduce the overall cost of treatment. The proposed system greatly reduces the cost of treatment and improves the quality of life by predicting breast cancer at early stage of development.

10.FUTURE SCOPE

The future work will focus on exploring more of the dataset values and yielding more interesting outcomes. This study can help in making more effective and reliable disease prediction and diagnostic system which will contribute towards developing better healthcare system by reducing overall cost, time and mortality rate. In further study, we will try to conduct experiments on larger data sets or try to tune the model so as to achieve the state -of-art performance of the model and a great UI support system making it complete web application model.

11. BIBILIOGRAPHY

- Jemal A, Murray T, Ward E, Samuels A, Tiwari RC, Ghafoor A, Feuer EJ, Thun MJ. Cancer statistics, 2005. CA: a cancer journal for clinicians. 2005 Jan 1;55(1):10-30.
- Polat K, Güneş S. Breast cancer diagnosis using least square support vector machine. Digital Signal Processing. 2007 Jul 1;17(4):694- 701.
- Akay MF. Support vector machines combined with feature selection for breast cancer diagnosis. Expert systems with applications. 2009 Mar 1;36(2):3240-7.
- Yeh WC, Chang WW, Chung YY. A new hybrid approach for mining breast cancer pattern using discrete particle swarm optimization and statistical method. Expert Systems with Applications. 2009 May 1;36(4):8204-11.

- Marcano-Cedeño A, Quintanilla-Domínguez J, Andina D. WBCD breast cancer database classification applying artificial metaplasticity neural network. Expert Systems with Applications. 2011 Aug 1;38(8):9573-9.
- Kaya Y, Uyar M. A hybrid decision support system based on rough set and extreme learning machine for diagnosis of hepatitis disease. Applied Soft Computing. 2013 Aug 1;13(8):3429-38.
- Nahato KB, Harichandran KN, Arputharaj K. Knowledge mining from clinical datasets using rough sets and backpropagation neural network. Computational and mathematical methods in medicine. 2015;2015.
- Liu L, Deng M. An evolutionary artificial neural network approach for breast cancer diagnosis. InKnowledge Discovery and Data Mining, 2010. WKDD'10. Third International Conference on 2010 Jan 9 (pp. 593-596). IEEE.
- Chen HL, Yang B, Liu J, Liu DY. A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis. Expert Systems with Applications. 2011 Jul 1;38(7):9014-22.

APPENDIX

A.Source Code

HTML:

```
<html lang="en">
<head>

  <meta charset="UTF-8">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <meta http-equiv="X-UA-Compatible" content="ie=edge">
  <title>Cancer Recognition System</title>
        <link  href="https://cdn.bootcss.com/bootstrap/4.0.0/css/bootstrap.min.css"
rel="stylesheet">
  <script src="https://cdn.bootcss.com/popper.js/1.12.9/umd/popper.min.js"></script>
  <script src="https://cdn.bootcss.com/jquery/3.3.1/jquery.min.js"></script>
  <script src="https://cdn.bootcss.com/bootstrap/4.0.0/js/bootstrap.min.js"></script>
  <link href="{{ url_for('static', filename='css/main.css') }}" rel="stylesheet">
      <style>
      .bg-dark {
            background-color: #42678c!important;
      }
      #result {
```

```
            color: #0a1c4ed1;
        }
        </style>
</head>

<body style="background-color:powderblue">
  <nav class="navbar navbar-dark bg-primary">
    <div class="container">
            <a  class="navbar-brand"  href="#"  style="font-size:30px">DEEP  LEARNING TECHNIQUES FOR BREAST CANCER RISK PREDICTION USING PYTHON</font></a>
    </div>

  </nav>
  <div class="container">

    <div id="content" style="margin-top:2em">
            <div class="container">
             <div class="row">
                  <div class="col-sm-6 bd" >
                   <h3>Recognising Type Of Breast Cancer </h3>
                   <br>
                    <p>Breast cancer is one of the main causes of cancer death worldwide.In this proposed system we are able to identify the rate of breast cancer. 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. This project develops an algorithm to detect the cancer in todays life. Since there are number of different type ofcancers manually identifying them can be a difficult task. This algorithm classifies cancer based on their images so we can monitor them more efficiently. . By prediction of Breast cancer in healthcare we are able to take suitable measures accordingly to control the cancer tissue. By using deep neural networks we are predicting the cancer is benign or malignant.</p>
                        <img src="https://encrypted-tbn0.gstatic.com/images?q=tbn%3AANd9GcToXWreLreIT1URBLGSkV_MrD2_GEF_xmFnFQ&usqp=CAU.jpg"  style="height:180px"class="img-rounded" alt="Gesture">
                        <h1 style="margin:20px;">Types of Breast Cancers :</h1>
```

```
                    <h2 style="margin:20px;color:red;">1. Malignant<br>2. Benign</h2>
                    <h3 style="margin:20px;font-size:20px;color:green;"> Malignant :</h3>
<p style="margin-left:20px;font-size:20px;">Malignant refers to cancer cells that can invade and kill nearby tissue and spread to other parts of your body.A malignant tumor is a tumor that may invade its surrounding tissue or spread around the body. Malignant tumors are cancerous growths. They are often resistant to treatment.</p>
<h5 style="margin:20px;font-size:20px;color:green;"> Benign :</h5>
<p style="margin-left:20px;font-size:20px;">A benign tumor is a tumor that does not invade its surrounding tissue or spread around the body. In most cases, the outlook with benign tumors is very good. But benign tumors can be serious if they press on vital structures such as blood vessels or nerves.</p>

                </div>
                <div class="col-sm-6">
                    <div>
                        <h4><font color="blue">To know the type of cancer Please upload an image</font></h4>
                    <form action = "http://localhost:5000/predict" id="upload-file" method="post" enctype="multipart/form-data">
                        <label for="imageUpload" class="upload-label" font color="yellow">
                            upload image
                        </label>
                        <input type="file" name="image" id="imageUpload" accept=".png, .jpg, .jpeg">
                    </form>


                    <div class="image-section" style="display:none;">
                        <div class="img-preview">
                            <div id="imagePreview">
                            </div>
                        </div>
                        <div>
                            <button type="button" class="btn btn-info btn-lg " id="btn-predict">submit</button>
                        </div>
```

```
                        </div>

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

                        <h3>
                                <span id="result"> </span>
                        </h3>

                </div>
                        </div>

                  </div>
                </div>
                </div>
   </div>
</body>

<footer>
   <script src="{{ url_for('static', filename='js/main.js') }}" type="text/javascript"></script>

</footer>

</html>
```

APP.PY:

```
from flask import Flask, render_template,request
import os
from keras.preprocessing import image
from werkzeug.utils import secure_filename
from keras.models import load_model
import tensorflow as tf
global graph
graph = tf.get_default_graph()
import numpy as np
model = load_model("breast cancer.h5")
app = Flask(__name__)
@app.route('/')
def index():
```

```
    return render_template("base.html",methods = ['GET'])
@app.route('/predict',methods = ['GET','POST'])
def pred():
    if request.method == "POST":
        f = request.files["image"]
        print("hie")
        """take the path of  the current
        running prog,and concatenate to the
        folder where you wouldlike to save the file"""
        basepath = os.path.dirname(__file__)
        print(basepath)
        file_path = os.path.join(basepath,"uploads",secure_filename(f.filename))
        print(file_path)
        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():
            p = model.predict_classes(x)
            print(p)
        index = ["benign(non cancer)","malignant(cancer)"]
        text = "the prediction of cancer is :" +index[p[0][0]]
        return text
if __name__ == "__main__":
    app.run(debug = True)
```