iJRASET
International Journal For Research in
Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 8 **Issue:** V **Month of publication:** May 2020

**DOI:** http://doi.org/10.22214/ijraset.2020.5043

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

# Breast Cancer Detection and Classification using Artificial Neural Network

Saiprasad Balraj[1], Aakash Chaurasiya[2], Dipan Kanabar[3], Prathamesh Gosavi[4], Sunil Patil[5]
[1, 2, 3, 4, 5]*Department of Electronics and Telecommunication, K.J. Somaiya Institue of Engineering and I.T., Mumbai, India*

***Abstract: Breast cancer is one of the leading causes of women fatalities. Breast cancer is a worldwide disease with symptoms and features being similar in almost all regions. In India, in the past 25-30 years there has been a steep increase in cases of breast cancer being reported in a comparatively younger age group (40+). Indian women have a higher rate of fatality due to breast cancer compared to the global average. There is severe lack of awareness among rural as well as urban women resulting in the delay in diagnosis of cancer makes it difficult to fight. There are many instances when specialists disagree on opinions related to diagnosis hence, forcing patients to look for another opinion from a different specialist hereby wasting crucial time and monetary resources of a patient. Hence, with the aim to support a doctor's diagnosis with an accurate, precise and quick diagnosis this Computer Aided Diagnosis (CADx) system has been developed. This model takes H and E stained images of breast tissues as input and classifies them into four classes namely Benign, In-Situ, Invasive and Normal. This program uses image processing techniques and Deep learning which is a class of Machine learning and AI. A Convolutional Neural Network has been used as the primary brain of this model since it detects distinctive features from the data without any human supervision or predefined set of features. This model is implemented using python and some useful deep learning libraries like TensorFlow and Keras. This model yields a 95.6 percent sensitivity for cancer. This model can help doctors and lab technicians to provide quick diagnosis and reduce conflict in opinions of specialists. A patient can get diagnosed in a more efficient and cost effective manner.***

## I. INTRODUCTION

Breast cancer is the most common cancer in women world-wide. Breast cancer is a disease where cells in the breast tissue get deformed, divide uncontrollably, typically resulting in a lump or mass. Breast cancer has no specific symptoms.

The most common symptom is the presence of painless lumps in the breast. Sometimes the breast cancer affects the lymph nodes present in the underarm region even causing swollen nodes before the main tumour is large enough to be felt or detected. Some of the less common symptoms are heaviness of breasts, changes in skin around the breast and changes in the nipple including abnormal discharges (especially blood). Any such change should be reported to a doctor as soon as possible. Breast cancer is detected by screening for early detection or by the presence of lumps in the breasts by self-examination or by examination by doctors. Many masses are detected in mammograms and most of the lumps turn. For confirmation of the presence of cancerous cells usually a needle biopsy is done. Many factors affect the type of biopsy being used.

## II. TYPES OF BREAST CANCER

There are many types of breast cancer and they have dif-ferent types of descriptions. The type of cancer is determined by checking the type of cells and tissues that have undergone abnormality. There are many different ways to classify cancer. Majority of breast cancers are Adeno- carcinomas. Adeno-carcinomas are formed in many other cancers and form in glands and ducts that secrete fluid. The most broadly used classification for cancer is based on whether the cancer has spread to surrounding tissues or not. Based on this criteria Breast cancer is classified into two broad categories namely In-Situ and Invasive.

### *A. In-Situ*

In-Situ simple means in place. In an In-Situ carcinoma the abnormal cells are present at the location of origin and have not yet spread to neighbouring breast tissues, lobules and ducts. It is also sometimes referred to as stage 0 of breast cancer. This type of carcinoma usually occurs at the ducts and hence most of the cases are termed as Ductal Carcinoma In-Situ (DCIS). DCIS is a non-invasive cancer. The DCIS is termed as non-invasive because it hasn't affected the neighbouring normal tissues and is present only in the milk ducts. DCIS is not necessarily fatal but it puts a patient at high risk of developing an invasive cancer. A person with a history of DCIS has a higher chance of getting breast cancer than a person who has never had breast cancer. There are many cases of recurrences reported at under 30 percent. Women who have undergone Breast Conserving Surgery to treat DCIS which does not use radiation therapy have 25-30 percent chances of breast cancer recurrence at any period in their lives. Women who have taken radiation therapy post their surgery have a lower risk of recurrence at under 15 percent.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

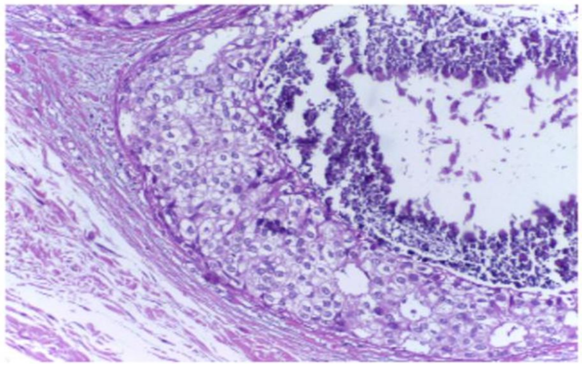


Fig.2.1

*B. Ivasive*

Most of the breast cancers which have spread to the surrounding tissues are termed as invasive carcinoma. This type of cancer occurs when abnormal (cancerous) cells from the milk ducts or lobules invade the neighbouring tissues. The most common types Invasive breast cancers are Invasive Ductal Carcinoma and Invasive Lobular Carcinoma. Invasive Ductal Carcinoma is more common than the latter. Survival chances for a person after she has Invasive breast cancer solely depends on individual diagnosis and treatment. The treatment usually involves a combination surgery, radiotherapy and chemotherapy. The order in which the therapies are given to patients depends on the stage of the cancer and the different characteristics of the tumour.

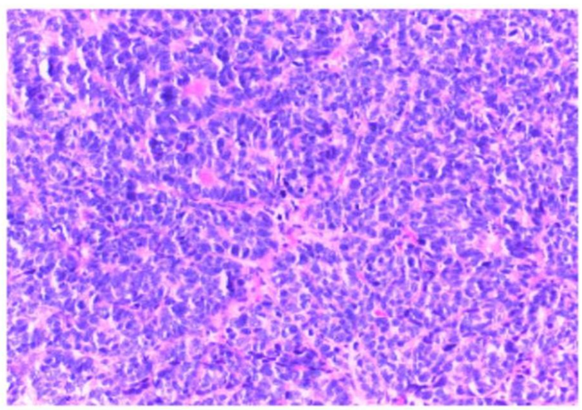


Fig.2.2

## III. ABBREVIATIONS

CADe - Computer Aided Detectors CADx - Computer Aided Diagnosis DCIS - Ductal Carcinoma In-Situ IDC - Invasive Ductal Carcinoma ML - Machine Learning
CNN - Convolutional Neural Network RELU - Rectified Linear Unit

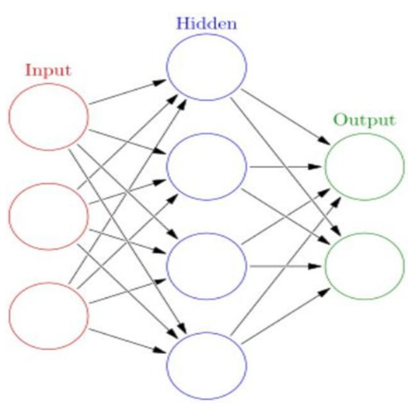



Fig.A

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

## IV. IMPLEMENTATION

### A. Block Diagram

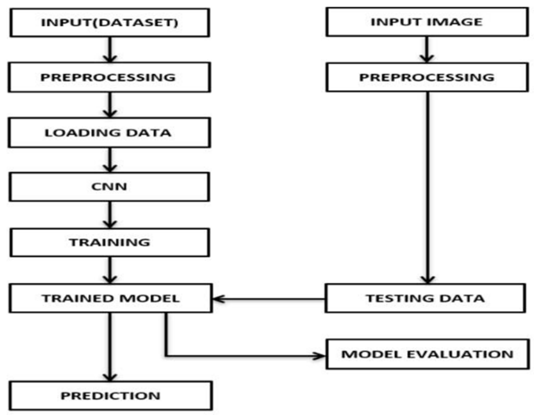



Fig. 4.1

### B. Software Implementation

For this project, Python is used as the primary programming language. The input images being used here are H and E stained images. Hematoxylin and Eosin stain (H and E) is one of the principal stains in histology. It is also considered as the golden standard for many diagnoses. Data set used is “Breast Histopathological images from Kaggle”. The original dataset consisted of 162 whole mount slide images of Breast Cancer specimens scanned at 40x. From that, 277,524 patches of size 50 x 50 were extracted (198,738 IDC (Invasive ductal carcinoma) negative and 78,786 IDC positive). https://www.kaggle.com/paultimothymooney/breast-histopathology-images/data

### C. Preprocessing

The data set contains raw images. The images are named according to their actual classifications (Benign, In-situ, Invasive and Normal). It is necessary to normalize data so that the neural network yields accurate outputs. The images are cropped to a suitable size uniformly throughout the entire database and rotated if needed. A function ‘getasSoftmax’ is defined to retrieve the name of the image. This information will be used as labels. The Images are then stored as numpy arrays of size 256, called as ‘x’ data, their respective labels are stored as ‘y’ data. The data is then split into training data and testing data using the function defined as ‘load-dataset’. This function returns training data (x and y) and test data (x and y). Four classes are defined according to label names and likewise prediction indices are defined for prediction namely Benign (0), In-situ(1), Invasive (2) and Normal(3).

### D. Convolutional Neural Network Mod-elling

The CNN contains the following layers:

1) *Input Layer:* This is the starting layer of the CNN, it takes image data in a numpy array and forwards it to the next layer.
2) *Activation Layer:* It maps the resulting values in between 0 to 1 or between -1 to 1 etc. (Depending on the function used). There are two types of activation functions, linear and non-linear activations. The activation function is also known as Transfer Function. The RELU (Rectified Linear Unit) function is used in the network as activation func-tion in all instances. This function changes all negative values to zero.

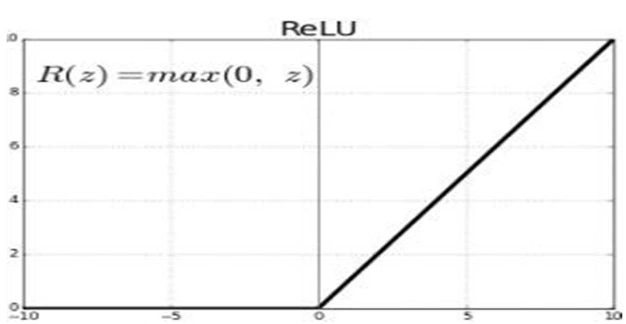



Fig. 4.2

3) *Convolution Layer:* In this layer filters are applied to the images or feature maps. This is one of the most important layers in a CNN. Depending on the shape of the data the convolution takes place. This layer has the maximum user defined parameters like number of kernels and kernel size.
4) *Max-Pooling Layers:* This layer is used in downsizing the input data thereby reducing its dimensionality. This helps classification to be done based on the features pinned in the remaining sub-region. It reduces the com-putational cost by reducing the number of parameters to learn and provides basic translation invariance to the internal representation.

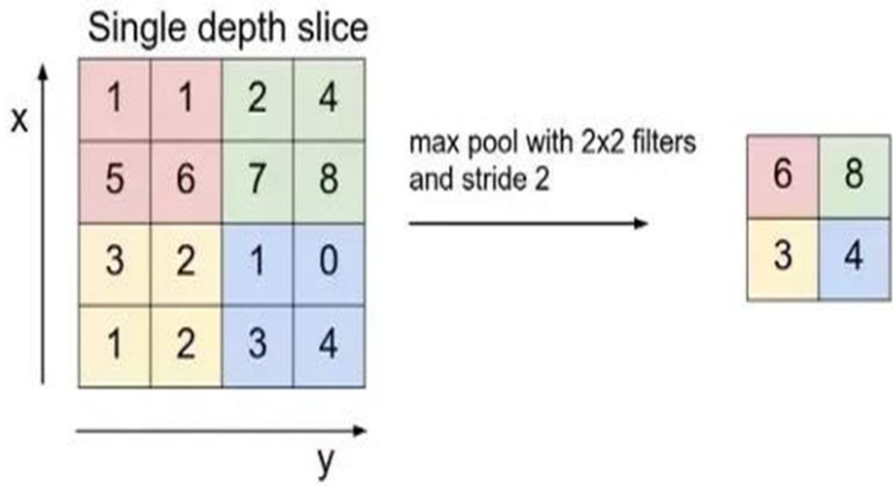



Fig. 4.3

5) *Dense Layers:* It is the regular deeply connected neu-ral network layer. The output of this layer is out-put=activation(dot (input, kernel)). The input here is the actual input to this layer and the kernel represents the corresponding weights.

### E. Training (Learning Algorithm)

The training process that is learning values of weights and biases is an iterative process. Since this is a supervised learning algorithm, there is forward as well as backward propagation of information in the process.
In the forward propagation, information from the input layer is passed on to subsequent layers after certain computations and the output layer assigns accordingly to match the desired label.
Once the data is passed through all the layers all the neurons complete their calculations and with each pass of information the values are stored in the neurons.
A loss function is used to measure the magnitude of error in the prediction of this network. The weight of neurons will be adjusted. Ideally a zero divergence is desired in the output predictions or label matches. The divergence is also termed as cost. In this algorithm a ‘categorical cross entropy’ loss function is used.
The loss calculated is propagated backwards to the pre-vious layers/hidden layers. This process continues until all neurons receive the loss signal and compute their contribution in the loss.
A gradient descent is used to change the weights which increments them by minute values. This process calcu-lates the derivative of the loss function which gives an idea about the direction of descent needed.
This process is done in batches, which in this case the batch size is set to 16. This means at instance during the training process the model will be processing only 16 images. The process is completed in successive iterations over the entire data set. The number of iterations is called as epochs and in this algorithm it is set to 10. These parameters can be tweaked according to computational capacity of the system and to achieve changes in accu-racy.
The last step is evaluation of the model which is done based on the results from the testing data.
A softmax function is used to convert the values from the output layer into probabilities. All the outputs for a specific prediction once passed through the softmax function will add up to 1.
An argmax function can also be used to fetch the ar-gument with maximum value from the output function, which will be the prediction index.
The model now is fit for detecting cancer from HE stained images and can classify them into the defined classes.

## V. RESULT

We have implemented this program for detection of breast cancer using HE stained images and it classifies the tissue present in the slide as Benign, In-Situ, Invasive and Normal. A 95.6 percent sensitivity for cancer is achieved using our proposed method. This method uses one of the most popular and accurate deep learning algorithms in Convolutional Neural Networks. This program reduces time taken by specialists to finalize their diagnosis by supporting their decision with a precise diagnosis.

1) *Applications:* The program can detect breast cancer in HE stained images in a very short time. This project can provide help to specialists to avoid disagreement in the diagnosis between them by supporting the closest diagnosis. This program is a cost effective, efficient and a time saving application for patients who seek for a "2nd opinion" from other doctors. This method can reduce patient's expenditure by reducing preventive therapy prescribed by doctors.
2) *Limitations:* We have implemented this program using python in an Anaconda Spyder environment, this program requires python being installed in the host computer. This is an efficient program which gives quick diagnosis. In future, even though it yields a 95.6 percent sensitivity for cancer the accuracy for classification can be further increased and this application can be designed to be more doctor and lab technician friendly.

## VI. CONCLUSION

The program, Breast Cancer Detection and classifica-tion using Artificial Neural Network was designed and modelled for Doctors and Lab Technicians using Deep Learning. The program classifies HE stained images of breast tissue collected by doing a biopsy as Benign, In-Situ, Invasive and Normal. It was tested efficiently and gives an effective 95.6 percent sensitivity to cancer. As it is essential to keep the diagnosis swift and cost effective this model will help the specialists in doing so. The efficient and fast diagnosis provided by this system will provide support to the decisions made by doctors over any diagnosis. Since it is a Computer Aided Diagnosis System it requires no maintenance and is very cost effective. Because of these factors it makes this model fit for being included side by side to the existing systems for diagnosis in hospitals and diagnosis laboratories with approval from specialists.

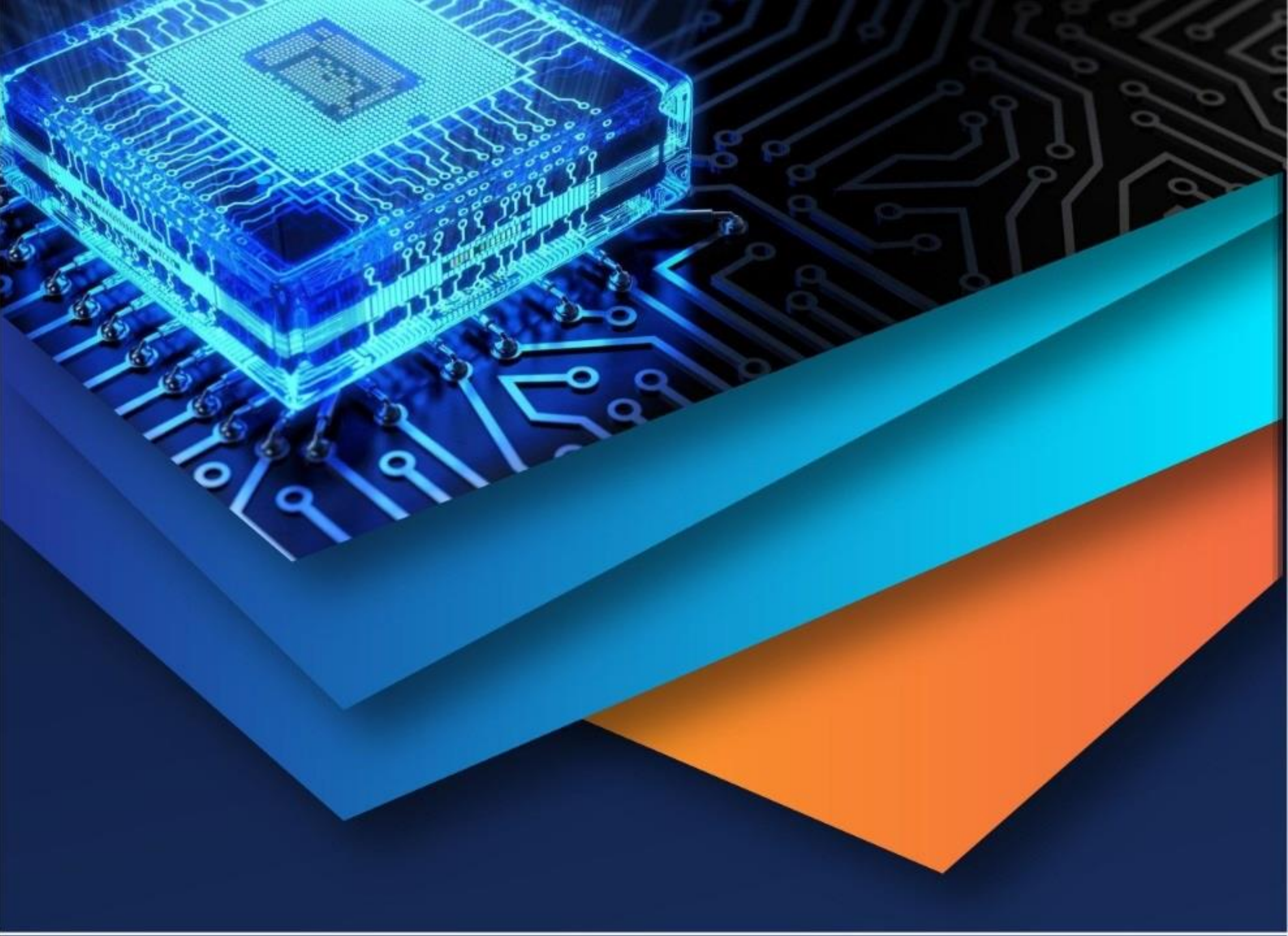
## REFERENCES

[1] HARSHITA SRIVASTAVA ON APRIL 24, 2018 IN ARTIFICIAL INTELLIGENCE Convolutional Neural Networks Explained.
[2] Y. A. Hamad, K. Simonov and M. B. Naeem, "Breast Cancer Detection and Classification Using Artificial Neural Networks," 2018 1st Annual International Con-ference on Information and Sciences (AiCIS), Fallujah, Iraq, 2018, pp. 51-57.
[3] Wikipedia contributors, 'Convolutional neural network', Wikipedia, The Free Encyclopedia, ¡https://en.wikipedia.org/w/index.php?title=Convolutional neural networkoldid=950807179¿
[4] Anirudh Rao, May 29,2019, Convolutional Neural Net-work Tutorial (CNN) – Developing An Image Classifier

In Python Using TensorFlow Convolutional Neural Net-

work (CNN) Tutorial In Python Using TensorFlow.

5) Breast Cancer Detection Using Convolutional Neural Networks, 2003
6) S. Guan and M. Loew, ”Breast Cancer Detection Using Transfer Learning in Convolutional Neural Networks,” 2017 IEEE Applied Imagery Pattern Recognition Work-shop (AIPR), Washington, DC, 2017, pp. 1-8.
7) Samuel Aji Sena, Panca Mudjirahardjo and Sholeh Hadi Pramono Breast Cancer Detection using Residual Convolutional Neural Network and Weight Loss Jurnal Infotel. 2019;11(2):43-47 DOI 10.20895
8) Introduction to TensorFlow
9) Applied Deep Learning - Part 4: Convolutional Neural Networks.
10) American Cancer Society. Breast Cancer Facts Fig-ures 2019-2020. Atlanta: American Cancer Society, Inc. 2019. Breast Cancer Facts Figures 2019-2020.
11) Ductal Carcinoma In Situ: Symptoms, Diagnosis, and Treatment.
12) Invasive Breast Cancer — Susan G. Komen®
13) Breast cancer by World Cancer Research Fund.
14) Types of Breast Cancer: Common, Rare and More Varieties by Cancer Treatment Centres of America.M. G. Kanojia and S. Abraham, ”Breast cancer detection using RBF neural network,” 2016 2nd International Con-ference on Contemporary Computing and Informatics (IC3I), Noida, 2016, pp. 363-368.
15) Simple Introduction to Convolutional Neural Networks by Matthew Stewert Feb 27,2019.
16) I. Singh, K. Sanwal and S. Praveen, ”Breast cancer detection using two-fold genetic evolution of neural network ensembles,” 2016 International Conference on Data Science and Engineering (ICDSE), Cochin, 2016, pp. 1-6.




10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429


# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY