IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 6, June 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906217 |

# Breast Cancer Classification Using Convolutional Neural Networks (CNN)

**Nutakki Prasanthi, Kondeti Krishna Priya, Kanikaram Tulasi, Kalava Ramya Sree**

Under Graduate Students, Department of Electronics and Communications Engineering, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, AP, India

**ABSTRACT:** Breast cancer is a disease in which cells in the breast grow out of control rapidly. It occurs when a malignant(cancerous) tumor originates in the breast cells. It is the most commonly occurring cancer in women and the second most common cancer overall. Around 2 million cases were observed in 2018. The early diagnosis of breast cancer can improve the prognosis and chance of survival significance, as it can promote timely clinical treatment to patients affected. Further accurate classification from the data of benign tumors can prevent patients from undergoing unnecessary treatments. Thus, the correct diagnosis of breast cancer and the classification of patients into malignant or benign groups is the subject of all research done and observed. Because of its unique advantages in critical features detection from complex breast cancer dataset, machine learning(ML) is widely recognized as the methodology of choice in breast cancer pattern classification.

## I. INTRODUCTION

Breast cancer (BC) is the most common cancer in women, affecting about 10 percent of all the women at some stages of their life. Due to the varying nature of breast symptoms, patients are often subjected to a barrage of tests, including but not limited to mammography, ultrasound and biopsy, to check their likelihood of being diagnosed with breast cancer. Biopsy is the most indicative among these procedures, which involves extraction of sample cells or tissues for examination. The sample of cells is obtained from a breast fine needle aspiration(FNA) procedure and then sent to a pathology laboratory to examine under a microscope. Numerical features, such as radius, texture, perimeter and area, can be measured from microscope images. Data, later on, obtained from FNA are analyzed in combination with various imaging data to predict the probability of the patient having a malignant breast cancer tumor.

## II. CONTRIBUTIONS

In our project, we design a classification program using CNN. We concluded our result as benign and malignant. To perform this we took a data set from kaggle i.e., image patches which does not require any pre processing. Using CNN, features are extracted and based on that features classification algorithm is carried out. We considered patch wise images. A data augmentation method is used in this work to enlarge the training set and to raise the efficiency.

## III. MATERIALS AND METHODS

### *DATASET*

We use IDC(Invasive Ductal Carcinoma) regular dataset from Kaggle. This dataset holds patches of size 50x50 which are scanned at 40x. File names in the dataset look likes this:

8863_idx5_x451_y1451_class0

Here, 8863_idx5 is the patient ID, 451 and 1451 are the x and y coordinates of the crop respectively, and 0 is the class label(0 denotes absence of IDC, 1 denotes IDC).

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906217 |

The dataset is downloaded and placed in the specified location i.e., in the original directory in the datasets directory. We have a directory for each patient ID. And in each such directory, we have the 0 and 1 directories for images with benign and malignant content.

***CONVOLUTIONAL NEURAL NETWORKS (CNNS)***

CNNs are applied to explore patterns in an image. CNN's are the neural networks that are specialized in visual pattern recognition. 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, as shown in figure below. 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. Convolutional layers bring out the features of images with precise positions. With convolutional layers, down sampling can 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. Data augmentation is an effective and widely used tool to avoid the over fitting problem by creating additional data.

***CNNs for Patch Wise Classification***

In the proposed architecture we have two classes, which are benign class and malignant class. In our work we used 50*50 image patches. Each neuron is connected to all the neurons in fully connected layer sharing some weights. By this convolution map is able to identify the same patterns at all the image positions and to reduce the total number of parameters obtained. The network is designed in hierarchical manner. Initially the image patches are divided into training and testing sets. The training is given as input to convolution neural network algorithm.

The convolutional layers and max-pooling layers are fully connected and which are then connected to the Softmax classifier. The Softmax classifier contains the number of output classes equal to the number of the outputs. In fully connected layer, the values are obtained by taking parameters as weights, bias and ReLU is performed to avoid negative and zero values. Next sigmoid function is performed and finally it gives output 0 or 1.

ReLU works as $f(x) = \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases}$

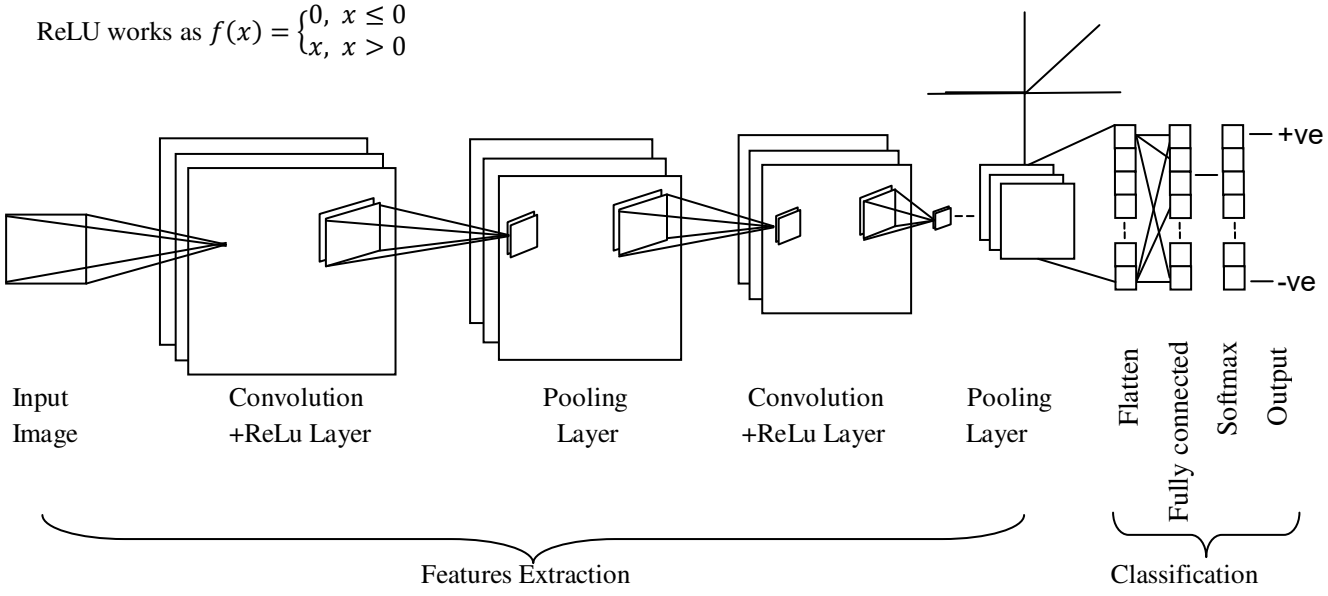


**Figure 1:** Typical CNN architecture for automatic detection of IDC breast cancer.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906217 |

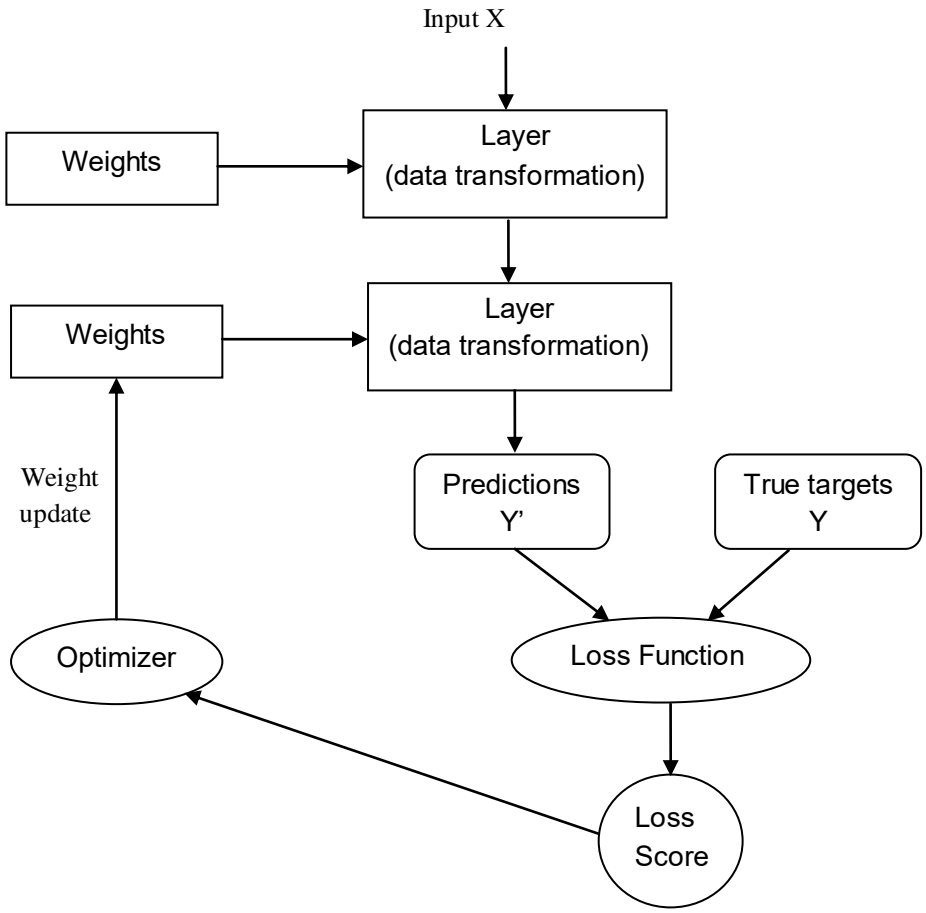


**Figure 2:** Detailed Process of Neural Networks

## IV. RESULT

The performance of a method is evaluated in terms of sensitivity and accuracy. This evaluation is performed patch wise. A softmax classifier is used to classify as benign and malignant.

Based on True Positive(TP), True Negative(TN), False Positive(FP) and False Negative(FN) values some parameters are obtained i.e., precision, recall and F1 score.

***Precision*** is the ratio of correctly predicted positive observations to the total predicted positive observations Or it is a proportion of true positive values.

$$Precision = \frac{TP}{TP + FP}$$

***Recall*** is the ratio of correctly predicted positive observations to all the observations in actual class.

Or it is a measure of true positive values in all true positive and false negative values.

$$Recall = \frac{TP}{TP + FN}$$

***F1 score*** is the weighted average of precision and recall.

$$F1\ Score = \frac{2 * (Precision * Recall)}{(Precision + Recall)}$$

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906217 |

| | Precision | Recall | F1-Score |
|---|---|---|---|
| IDC(-)0 | 0.99 | 0.78 | 0.87 |
| IDC(+)1 | 048 | 0.95 | 0.64 |

***Confusion Matrix***

Confusion Matrix is a very important metric when analyzing misclassification. Each row of the matrix represents the instances in a predicted class while each column represents the instances in an actual class. The diagonals represent the classes that have been correctly classified.

It is generally represented as

$$Confusion\ Matrix = \begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix}$$

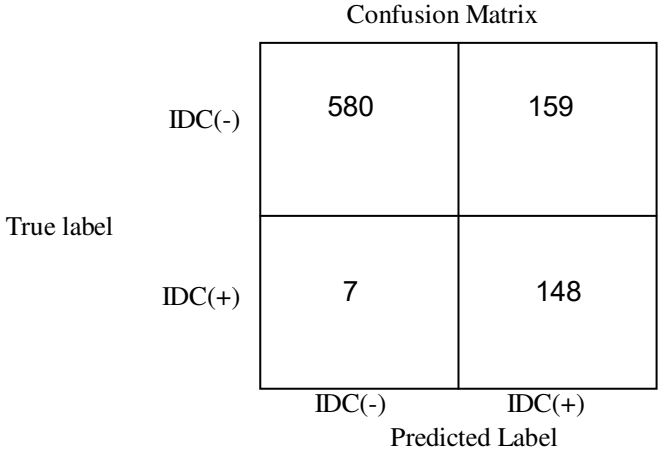


**Figure 3:** Confusion Matrix for the data.

***Accuracy*** is specified as the percentage of correct predictions for a model.

Mathematically, accuracy can be calculated as the follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

***Sensitivity*** is a measure of the proportion of actual positive cases that got predicted as positive (or true positive). Sensitivity is also termed as Recall.

Mathematically, sensitivity can be calculated as the follows:

$$Sensitivity = \frac{TP}{TP + FN}$$

***Specificity*** is defined as the proportion of actual negatives, which got predicted as the negative (or true negative).

Mathematically, specificity can be calculated as the follows:

$$Specificity = \frac{TN}{TN + FP}$$

The proposed classifier distinguished all the benign and malignant samples respectively. This model produces an overall accuracy of 81.4 3%, specificity of 95.4 % and sensitivity of 78.4 %.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906217 |

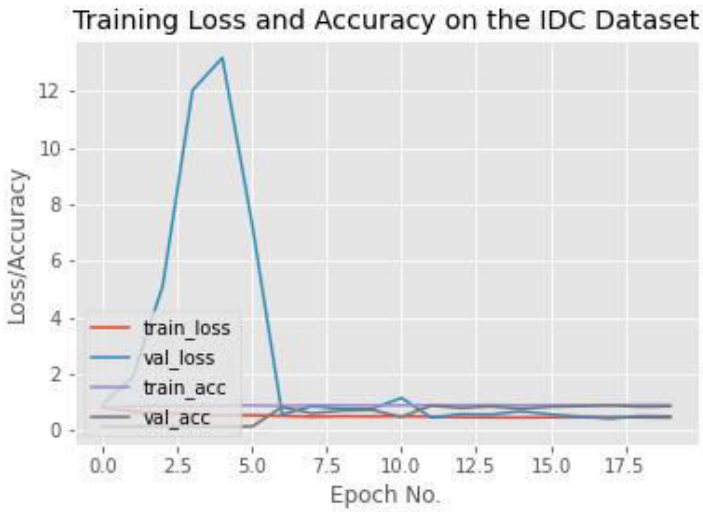


**Figure 4:** The graphical representation of Loss/Accuracy with respect to number of epochs.

We observe that loss decreases as number of epochs increases.

## V. CONCLUSION

In the project we build a breast cancer classifier on the IDC dataset using keras. Convolution Neural Network is used here to train the images. For each epoch, loss and accuracy are determined. We conclude the predicted values have a good agreement with the effective values of the model.

## REFERENCES

1. M. Masud, A. E. Eldin Rashed, and M. S. Hossain, "Convolutional neural network based models for diagnosis of breast cancer," Neural Computing and Applications, vol. 5, 2020.
2. J. L. Wang, A. K. Ibrahim, H. Zhuang, A. Muhamed Ali, A. Y. Li, A.Wu, "A study on automatic detection of IDC breast cancer with convolutional neural networks," in proceedings of the 2018 international conference on Computational Science And Computational Intelligence (CSCI), pp. 703-708, Las Vegas, NV, USA, December 2018.
3. A. Jemal, R.S., E. Ward, Y. Hao, J. Xu, T. Murray, M.J. Thun, "Cancer statistics", CA Cancer J Clin, Mar-Apr;58(2):71-96. 2008, doi: 10.3322/CA.2007.0010.
4. Jafari-Marandi, R., Davarzani, S., Gharibdousti, M.S., and Smith, B.K., An optimum ANN based breast cancer diagnosis: Bridging gaps between ANN learning and decision-making goals. Appl. Soft Comput. 72:108–120, 2018.
5. Übeyli, E.D. Implementing automated diagnostic systems for breast cancer detection. Expert Systems With Applications, vol. 33, pp. 1054–1062, 2007.
6. Guo, H., and Nandi, A.K.: Breast cancer diagnosis using genetic programming generated feature. 2005 IEEE Workshop on Machine Learning for Signal Processing, Mystic, CT. , pp. 215–220, 2005.
7. Dheeba, J., Singh, N.A., & Selvi, S.T,. Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach. Journal of Biomedical Informatics, vol. 49, pp. 45–52, 2014.
8. A. Cruz-Roa, A. Basavanhally, F. Gonzalez et al., "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks," Medical Imaging 2014: Digital Pathology, vol. 9041, no. 216, p. 904103, 2014.
9. G. Murtaza, L. Shuib, A. W. Abdul Wahab et al., "Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges," Artificial Intelligence Review, vol. 53, no. 3, pp. 1655–1720, 2020.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details