International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-4, November 2019

Classification of Breast Cancer using Deep Learning Architecture

B.Krishnakumar, K.Kousalya, R.S.Mohana, K.Dinesh, S.Santhiya

Abstract— Human beings are affected by several diseases nowadays. All those diseases are healable with minimal amount of treatment when they are identified at its early stage. Several patients were not serious enough in diagnosing the disease initially, which makes the disease incurable for the patient lifelong. Hence in recent days number of death rates are getting increased. Cancer is the most dangerous diseases. Among several types of cancers women are mostly affected by breast cancer. In most of the developing and under developing countries breast cancer is the most prominent reason for women mortality. It is also curable when it is identified at its starting stage. During the later stages the cancer cells will be disseminated all over body hence it is difficult to remove it completely. Hence it has to be identified at its initial stage in order to give best treatment for the patient at right time. In this paper , Convolution Neural Network (CNN) a deep learning model is proposed for the investigation of breast cancer images for finding whether the person is affected by cancer or not. In the proposed work , features from images are extracted using convolution layers and then it is passed to the fully connected layer where it classifying the images as either malignant or benign. Experiments using standard benchmark datasets for the proposed CNN Model and standard Visual Geometry Group Network (VGGNet) has been conducted to measure its performances. From the results ,it is clear that CNN outperformed with the accuracy of 86.32% when compared to VGGNet which provides only 50% accuracy for the identification of breast cancer.

Keywords- Breast Cancer, Classification, CNN , VGGNet,

I. INTRODUCTION

In machine learning a model is created for categorizing the given input as a output class. In order to create a model, several machine learning algorithms were used. Training data set is used by the machine learning algorithms to train and deploy a model. Once the model is created it is used to find the output class label for the given input data. Accuracy of the model is calculated by using both predicted output and actual output for the given input data. When the prediction accuracy has reached some threshold value then the training process will come to an end, otherwise the model will be trained again and again using the machine learning algorithm until accuracy reaches the threshold value.

But machine learning techniques will not yield fruitful results when working with high dimensional data where the number of input features were large.

Then the machine learning algorithms are not capable of extracting the features automatically. These two drawbacks are the biggest challenges for machine learning algorithms.

To overcome the disadvantage of machine learning, a Deep learning techniques is developed which is capable of handling large amount of data and it can also extract features automatically from the input.

Deep learning models can generate the features on which the outcome will depend on. To build deep learning model that mimic the working of brain. It is achieved through Neural Networks. The Motivation behind Neural Networks is biological Neuron. All the input features were given to the first layer which is called as input layer. Several computation process will be carried out in input layer and the hidden layer. All the layers in between the input layer and the output layer are called hidden layers. The output is provided by the output layer.

II. LITERATURE SURVEY

Amir Fallahi and Shahram Jafari (2011) employed classifier based on Bayesian network to establish an automated methodology for identifying breast cancer. Gorgel, Sertbas (2015) proposed to detect abnormal cancer cell based on Support vector machine and spherical wavelet transform . Jankovic (2015) used two matrix based on gray-scale occurance and gray level spatial dependence for the identification of unusual breasts.IssacNiwas (2010) classified the image as malignant or benign. Lukasz (2008) used circular hough transform to identify circular shaped nuclei. Microscopic images were taken for the analysis of malignant cell and for find their grading value. In this paper support vector machine is employed for the purpose of classification. MarekKowal (2011) extracted features using clustering based on Gaussian mixture along with a use of adaptive threshold. Classification techniques such as KNN, Decision tree classifiers are used. Uncommon breast is identified by Nakamura (2017) using optimization techniques such as particle swarm and biogrography optimization.

Manuscript published on November 30, 2019.
* Correspondence Author
B.Krishnakumar*, Department of CSE, Kongu Engineering College,Erode, Tamilnadu, India.(Email: krishnakumar@kongu.ac.in)
Dr.K.Kousalya, Department of CSE, Kongu Engineering College,Erode, Tamilnadu, India. (Email: kouse@kongu.ac.in)
Dr.R.S.Mohana, Department of CSE, Kongu Engineering College,Erode, Tamilnadu, India. Email: mohanapragash@kongu.ac.in
Dr.K.Dinesh, Department of CSE, Kongu Engineering College, Erode, Tamilnadu, India. (Email: dinesh.cse@kongu.ac.in)
S.Santhiya, PG Scholar, Department of CSE, Kongu Engineering College, Erode, Tamilnadu, India. (Email:santhiya123cse@gmail.com)

Pawel (2013) used tiny cytology based images for the identification of breast cancer. Shanthi& Murali Bhaskaran (2013, 2014) developed a method to detect all types of abnormalities in breast cancer.

III. PROBLEM STATEMENT

When breast cells starts to develop in an uncontrolled manner in breast then it is causes breast cancer. It is a type of cancer which occurs mostly for women and rarely for men. Breast consist of three important components such as lobules, ducts, and connective tissue. Lobules are the type of gland which is responsible for the production of milk. The milk is transferred to the nipple with the help of ducts. Connective tissues are the one which connects every parts of the breast. Brest cancer cells may develop at lobules or duct or any other part of the breast. It will be formed as a tumor and it can be absorbed as lump.

There are two types of tumors such as benign and malignant. Benign are the type which are not expressed as dangerous or cancerous since they were developed slowly. Benign tumor will not affect nearby tissues or any other different part of the breast or body. The other type of tumor is malignant , which will spread quickly and it is very dangerous to health. Finding the category of the tumor is very important for proper treatment..

IV. PROPOSED WORK

The proposed framework is based on comparison of convolutional neural network and VGGNet architectures used for analysing and to find benign and malignant cells from the breast cancer affected images. Features are extracted from convolutional layer, pooling layer. The features are given as input to a fully connected layer for the purpose of classification.

The computer reads an image through three channels. They are red, green, blue. Each channel has own pixel value. The size of the image will be rows* cols*3. Real-life images are not that small. Consider the image size of 200*200*3 input pixels. It is then passed to a fully connected layer. The number of filter values needed for the first layer will be 1,20,000. It requires more neurons. It leads to overfitting data. Network would become huge and nearly impossible to train. So we cannot use a fully connected network for image classification.

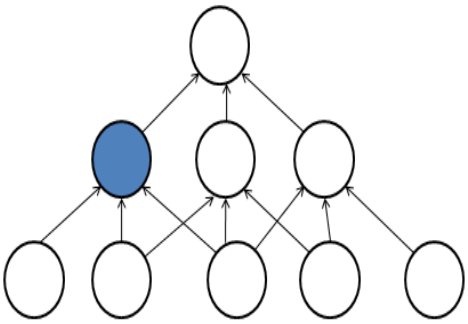

Fig.1. Basic Structure of Neurons and Layer

For example: In Fig.1, shaded neuron in first hidden layer is linked to only a set of neurons from its previous layer.. Unlike regular neural networks, the shaded neuron is connected to all the five neurons before it. Each and every layer is made up with set of neurons, but the neurons are not fully connected with its pervious layers. Hence CNN takes only less number of links compared with Artificial Neural Networks

A. Convolutional Neural network

Convolutional Neural network involves the following layers. Convolution layer, ReLu Activation layer, Pooling layer, Flattening and Fully connection layer. Convolutional, pooling layers are used for learning the features which is shown in Fig.2. Flattening and a Fully connection layer is used for classification.

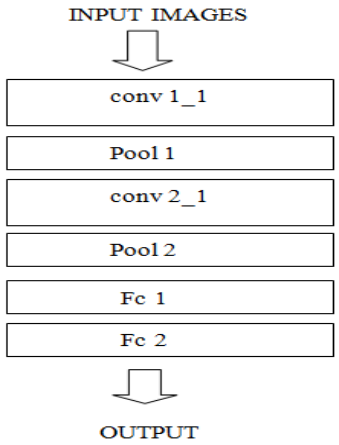

Fig.2. Basic Model of CNN

- *Convolutional Layer*

The convolutional layer consists of an input image and filter. Filter is also called a spatial extent. The spatial extent may be 3*3 or 5*5 or 7*7. The input image is convolving by filter with padding, stride, and number of the filter. While convolving corner pixel, corner pixel will get cover one time. Corner pixel information is loss. To overcome this problem padding is introduced. There are two types of padding. Same padding and Valid padding. In the same padding, we add zeros around the input image. Corner pixel gets cover more than once. So that input and output image has the same dimension. Valid padding where zeros are not added around the input image.

Stride defines the interval at which filter is applied. The number of pixels by which the filter is slide over the input image. The filter matrix will be drifted one pixel when the value of stride is equal to one. If it is two then two pixels will be moved. After convolving it passed to ReLU function. ReLU function is to remove negative values by changing the negative value to zero to produce the output image. The output image is also called a feature map.

- *Pooling Layer*

The result of the Conv layer's output is redundant because of much information. Convolutional layers will produce similar values for neighboring pixels in outputs. Convolutional layer output is passed as input to the next layer which is pooling. The main objective of using pooling layer is to diminish the size the input.

The process of pooling is executed by taking minimum, maximum or average of the pixel values. Max Pooling takes the maximum value from the input matrix. Average pooling generates the output by calculating average value of input. Whereas sum pooling provides the sum of the input pixel as output.The pooling layer is also performed with padding and stride

- *Flattening*

Flattening is done before transmitting the input features to the fully connected layer from the convolution layers. Flattening transforms a two-dimensional matrix of pooling layer output into a vector. The vector will be given as input to a fully connected neural network.

- *Fully Connected Layer*

The fully connected layer consist of neurons which are linked to all other neurons present in its previous layer. In the output layer softmax activation function has been used for the purpose of classification. Feature which are expressed by the previous layers are put together in this fully connected layer and by utilizing it the classification process is done. This layer is used for the prediction of classes

B.VGGNet

VGGNet was developed by Simonyan and Zisserman from the university of oxford. VGGNet was the developed before ILSVRC-2014. ILSVRC is ImageNet LargeScale Visual Recognition Challenge. ImageNet consist of 15million labeled high-resolution images. It belongs to 22,000 categories. The goal of the ImageNet challenge is to train the model that can correctly classify the input image into 1000 object categories. 1000 image categories represent object classes. VGGNet achieved an error rate of 7.3. The input VGGNet Architecture is of fixed 224*224. VGGNet as shown in Fig. 3, uses the Convolutional layer, ReLU layer, Pooling layer, Flattening, Full connection layer as same as a convolutional layer.

- *Convolutional Layer:*

VGGNet consists of 16 convolutional layers. The convolutional layer consists of an input image and filter. Filter is also called a spatial extent. The spatial extent is of fixed 3*3 for all 16 layers. The input image is convolving by filter with padding, stride, and number of the filter. The convolutional layer uses the same padding and stride of one.

- *Pooling Layers*

The pooling operation is done to reduce the dimension of the input image.

- *Flattening Layer*

Flattening transforms a two-dimensional matrix of pooling layer output into a vector. The output from the previous layer is flattened with 25089 feature maps each of size 1*1.

- *Fully Connected Layer*

VGGNet consists of two fully connected layers with 4096 units. It is followed by the softmax layer for predicting two classes.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. Dataset

The dataset [9] consists of an invasive breast cancer image. The cancer cells initialize at the milk ducts and it spreads to all other parts of the breast which has normal tissues. Invasive ductal carcinoma is also a type of breast cancer cell that spread through blood to all other parts of the body it is also called as infiltration ductal carcinoma.

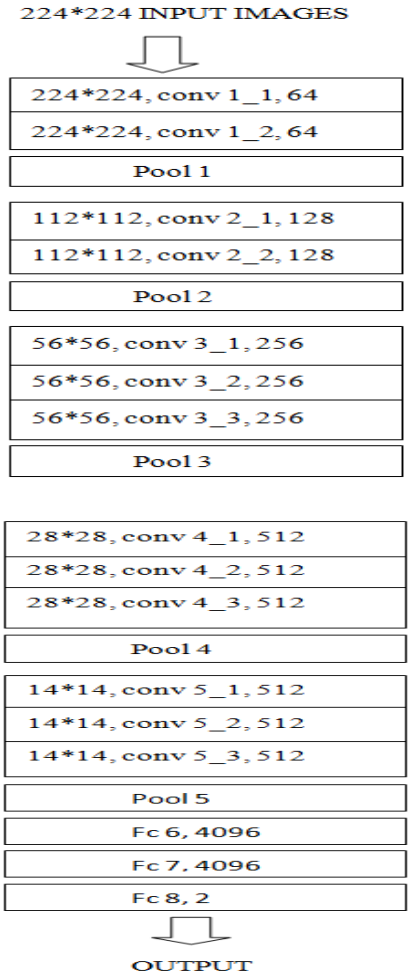

Fig.3. Basic Architecture of VGGNet

The dataset consist of 162 breast cancer image specimens which is mounted as slide images taken using 40x scanner. From that, 277,526 patches were extracted. Each patch of size is 50 x 50. There are 198,739 negative patches and 78,787 positive patches. The image name represents the category of the image. For example 10_idx4_x13_y11_class0.png is the image name then it is given by using the following format, u1_x1X1_y1Y1_classC1.png, Where the patient id is given as U1, X1 and Y1 represents the x and y coordinate of the patches and C represents the class as positive or negative

B. Results and Analysis

The proposed framework based on comparison between convolutional neural network and VGGNet is done with breast cancer dataset. Table-I shows the results of convolutional neural network and VGGNet. Convolutional neural network shows better accuracy than VGGNet.

Table-I Accuracy

Architecture	Accuracy%
Convolutional Neural Network	86.32
VGGNet	50.00

The proposed convolutional neural network model with 2 convolution layer outperformed with the accuracy of 86% when compared with the CNN Model used in [13] with 3 convolution layers which shows accuracy of about 80%. The comparison is done based on the same histopathological dataset for both proposed and existing CNN model. Table-II shows confusion matrix for the histopathological image dataset and precision and recall values for two different classes is shown in Table-III.

VI. CONCLUSION

In this framework, features are extracted from breast cytology images that was trained, tested and validated. Finally the performance of proposed framework is the comparison of Convolutional Neural Network and VGGNet gives the better accuracy for Convolutional Neural Network than VGGNet. In the future we shall use a pretrained VGGNet to improve the accuracy of VGGNet for classifying the benign and malignant cells.

Table -II Confusion matrix

		Actual	
		benign	**malignant**
Predicted	**benign**	9	8
	malignant	0	41

Table-III Precision and Recall

	Precision	Recall
benign	0.53	1
malignant	1	0.81

REFERENCE

1. Amir Fallahi, and S. Jafari,. "An expert system for detection of breast cancer using data preprocessing and bayesian network." international Journal of Advanced Science and Technology, 34, 2011,pp. 65-70.
2. P. Gorgel, A. Sertbas, O.N. Ucan, "Computer-aided classification of breastmasses in mammogram images based on spherical wavelet transform andsupport vector machines", Expert Syst. 32 (1),2015,pp.155–164.
3. M. Milosevic, D. Jankovic, A. Peulic, "Comparative analysis of breast cancerdetection in mammograms and thermograms", Biomed. Eng.-Biomed. Tech. 60(1),2015,pp. 49–56.
4. Niwas, I., Palanisamy, P. and Sujathan, K. "Wavelet based feature extraction method for breast cancer cytology images". In Industrial Electronics and Applications (ISIEA), IEEE Symposium on. Penang, IEEE, 2010,pp. 686-690.
5. Lukasz Jelen , Thomas Fevens, Adam Krzyz˙ Ak, "Classification Of Breast Cancer Malignancy Using Cytological Images Of Fine Needle Aspiration Biopsies" Int. J. Appl. Math. Comput. Sci, Vol.18, 2008, pp.75–83.
6. Kowal, M., P. Filipczuk, A. Obuchowicz, and J. Korbicz, "Computer aided diagnosis of breast cancer using Gaussian mixture cytological image segmentation." Journal of Medical Informatics & Technologies, 17,2011, pp. 257-262
7. K.Nakamura, "Abnormal Breast detection via combination of particle swarm optimization and biogeography-Based optimization", Adv. Comput. Sci. Res. 70, 2017,pp.646–650.
8. Paweł Filipczuk., T. Fevens, A. Krzyzak, and R. Monczak, "Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies." Medical Imaging, IEEE Transactions 32 (12), 2013,pp. 2169-217
9. https://www.kaggle.com/paultimoth mooney/breast-histopatholog images
10. Fabio Alexandre Spanhol, "Breast cancer histopathological image classification using convolutional neural networks,2016.
11. Shanthi & Murali Bhaskaran, "A Novel Approach for Detecting and Classifying Breast Cancer in Mammogram Images", International Journal of Intelligent Information Technologies, 9(1), 21-39, 2013
12. .Shanthi & Murali Bhaskaran, "A novel approach for classification of abnormalities in digitized mammograms", Sadhana -Academy Proceedings in Engineering Science journal Vol. 39, Part 5, 2014, pp. 1141–1150

AUTHORS PROFILE

Mr. B.Krishnakumar is pursuing Ph.D., in Machine Learning under Anna University. Currently he is a Assistant Professor in the department of Computer Science & Engineering in Kongu Engineering College, Tamilnadu, India. He has completed 12 years of teaching service. He has published 5 articles in International &National Journals. He has published 9 articles in International & National Conference.He has authored 1 book chapter with reputed publishers. He has organized 2 funded seminars and workshops.

Dr.K. Kousalya, received the B.E. and M.E. degrees in Computer Science and Engineering from Bharathiar University, Coimabatore, India, in 1993 and 2001, respectively. She completed her research work and awarded Ph.D degree by Anna University in the year 2010. She is working as an Professor in the department of Computer Science and Engineering, Perundurai, Tamilnadu. She has published more than 50 research articles in National/International Journal. She has guided more than 10 phd scholars. Her areas of interest are Deep Learning, Machine Learning, Cloud Computing, Compiler Design and Theory of Computation.

Dr. R.S.Mohana, received the Bachelor of Engineering in Computer Science in the year 2002 from National Institute of Technology, Trichy. She completed her Master of Engineering in Computer Science in the year 2009 from Anna University. She completed her research work on "Machine Learning approaches for scheduling and admission control in Cloud Computing" and awarded Ph.D degree by Anna University in the year 2016. She has presented and published many papers in national, international conferences and journals. Currently she is working as Assistant Professor (Senior Grade) in the Department of Computer Science and Engineering at Kongu Engineering College, Erode. Her area of research interest is Cloud Computing and big data analytic.

Dinesh Komarasamy, is currently working as an Assistant Professor in the Department of Computer Science and Engineering, Kongu Engineering College, Anna University, India. He has awarded as Ph.D degree under the faculty of Information and Communication Engineering in the year 2019 from Anna Univerisity. He has obtained his M.E graduate in the specification of Computer Science and Engineering and B.E graduate in the specification of Computer Science and Engineering during the year 2012 and 2010 respectively. Also, His research area include Cloud Computing and Optimization Techniques. He has presented 6 research publications in conferences and published 9 papers in the reputed journals.He has also published a book chapter in the area of cloud computing.

Ms.S.Santhiya received the Bachelor of Engineering in Computer Science in the year 2018 from Builders Engineering College, Kangayam. Currently she is doing Master of Engineering in the Department of Computer Science and Engineering at Kongu Engineering College, Erode. Her area of research interest is Deep Learning and Machine Learning.