International Conference on Advances in Materials, Computing and Communication Technologies
In Association withInternational Journal of Scientific Research in Science and Technology
Volume 9 | Issue 1 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)

Analysis of Breast Cancer in CNN Using Modified Logistic Regression Algorithm

Ajitha T, Reya Mory M*
*Department of Electronics and Communication Engineering, St. Xaviers Catholic College of Engineering Kanyakumari, India

ABSTRACT

Primary identification and prediction of type of the cancer ought to develop a compulsion in cancer study, in order to assist and supervise the patients. The significance of classifying cancer patients into high or low risk clusters needs commanded many investigation teams, from the biomedical and the bioinformatics area, to learn and analyze the application of machine learning (ML) approaches. Logistic Regression method and Multi-classifiers has been proposed to predict the breast cancer. To produce deep predictions in a new environment on the breast cancer data. It explores the different data mining approaches using Classification which can be applied on Breast Cancer data to build deep predictions. Besides this, this study predicts the best Model yielding high performance by evaluating dataset on various classifiers

Keywords: bio informatics, machine learning, logistic regression, clusters, classifiers

I. INTRODUCTION

Breast cancer remains to be the outmost identified cancer in the whole universe and is the prime source of cancer demise amid women. Earlier detection of breast cancer can save many lives in a best effective manner. Without reaching for surgical biopsy, prompt diagnosis entails accurate and steadfast diagnosis process that permits medical practitioner to differentiate benign breast tumors from malignant tumors. Every single minute, anywhere all over the globe context of breast cancer is identified amongst females and every one minute, [7]everywhere all over the globe and somebody expires from breast cancer. Breast tumors can be identified and further classified into three different categories known as benign breast cancers, in situ cancers, and invasive cancers. Benign breast tumors fall into the Chief category of tumors often detected by undergoing mammography. They cannot extent to external organs as by nature they are non-cancerous. By means of mammography in rare cases, this one is hard to discriminate specific bulk of benign from malignant lesions. Insitu or noninvasive cancer, is fully confined in the ducts.

Breast cancer is a serious issue worldwide and the major problem is to find the cancer at the early stages and there are some existing methods to find the cancer at benign stages, detection of spiculated lesions in digital mammograms [10].It enables computation of the future burden of cancer according to projected population changes[4]. CADe has been shown to help radiologists find more cancers both in observer

studies and in clinical evaluations [8]. (SVM) have greater accurate diagnosis ability [1] SVM-based method combined with feature selection have been conducted on different training-test partitions of the Wisconsin breast cancer dataset (WBCD) method. The digitalization of the mammographic images is made with suitable contrast and spatial resolution for processing purposes. The broad recuperation system allows the user to search for different images, exams, or patient characteristics [5]. Mammographic Imaging describes a Practical Guide [2], retains information on analog mammography, builds upon ongoing developments for breast imaging. Screening MRI is recommended for women with an approximately 20–25% or greater lifetime risk of breast cancer[6]. Breast parenchymal density is considered a strong indicator of breast cancer risk and therefore useful for preventive tasks [3]. Treatments for breast cancer cause some type of side effects. Surgery can cause pain and lymphedema [9].

II. METHODOLOGY

Supervised learning is performed for classification tasks. The input is fed to the median filter which is used to reduce the noise from the image. Then it is given to the region of interest and logistic regression to extract the feature. The clustering is used to divide groups into similar characteristics and assign them into cluster. The CNN classifier is used to predict the tumour.

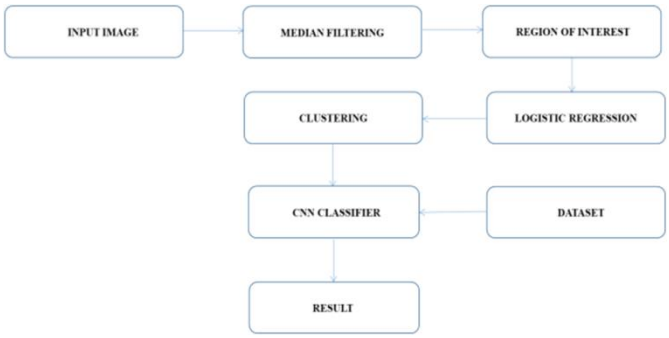

Fig 1 BLOCK DIAGRAM

A. Logistic Regression

Logistic regression is a predictive analysis. It measures the relationship between dependent variable(our predicted output)and one or more independent variables (our features) by estimating probabilities using logistic regression method. Logistic Regression is a Machine Learning algorithm which is used for the classification problems. It is a predictive analysis algorithm and is used to predict whether the tumor is malignant or benign.

Logistic regression is mainly used for predicting the continuous variables. So, we are using the hypothesis function,

$h\theta\,(x) = 1/1+ e^{-x}$

Where,$h\theta(x)$-prediction function θ-model parameter x-input value

$$if h\Theta(x) \geq 0.5, y = 1(malignant)$$
$$h\Theta(x) \leq 0.5, y = 0(benign)$$

Two main steps are used,

- Calculate the logistic function
- Find the model parameters

Here the data should be linearly seperable then transfer the data points using logistic function or sigmoid function

$$h\theta(x) = \frac{1}{1 + e^{-x}}$$

This function transforms each input value in the range between 0 to 1.The smallest negative numbers resulted in values is close to zero. The largest positive numbers resulted in values is close to one.

To calculate the probablity of the class,

$$p(class = 0/1) = 1/1+ e^{-(bo+b1x1+b2x2)}$$

Find the model parameters using the training set of inputs. Then we calculate the prediction values of the model parameters for each training set.

B. CNN CLASSIFIER

A Convolution neural network (CNN) is a neural network that has one or more convolution layers and is used mainly for image processing, classification, segmentation. The Convolution neural Network (CNN) is a deep learning neural network. It is used for image recognition and classification.

1) Convolution Layer:

A convolution layer is a class of deep neural networks, most commonly applied to analyzing visual imagery. This means that learned features of one section of the image can also be applied to other parts, and similar features are used in all image sections. After the features are acquired, the features of Convolutional layer are used to categorize images

2) Pooling Layer

Pooling layers are used to reduce the dimensions of the feature maps.It reduces the number of parameters and the amount of computation performed in the network The pooling layer shows the features present in region of the feature map generated by a convolution layer. It makes the model more robust to variations in the position of the features in the input image.

3) Fully connected Layer

The output of the last convolution layer and connect every node of the current layer with the other nodes of the next layer. Neurons in a fully connected layer have full connections to all activations in the previous layer.

III. RESULT AND DISCUSSION

The medical images like MRI or binary image are given as input image. The original data set consist of 162whole mount slide image of breast cancer (BCa) specimen scanned at 40 xs.

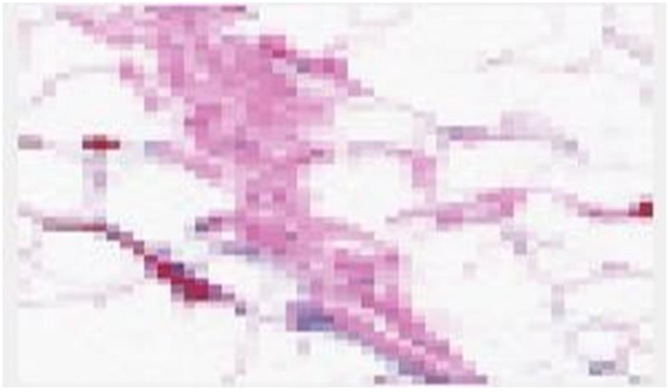

Fig 2 ORIGINAL IMAGE

The original image is converted in to gray scale image. This process removes the entire color

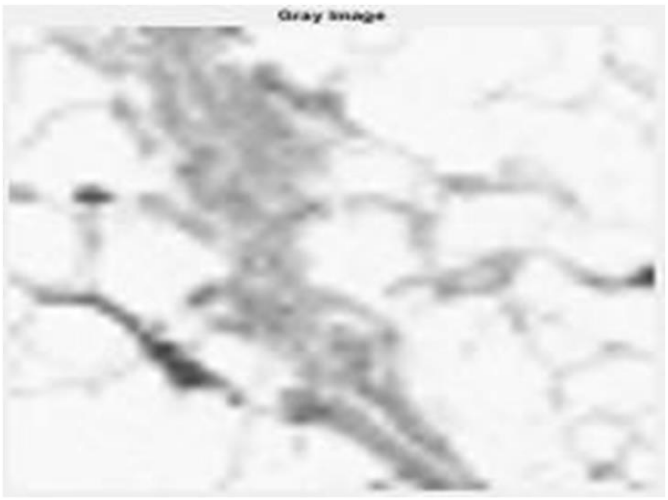

Fig 3 GRAY IMAGE

Image leaving only the luminance of each Pixel
The grey level or grey value indicates the brightness of a pixel. A greyscale or image is one in which the value of each pixel is a single sample representing only an amount of light

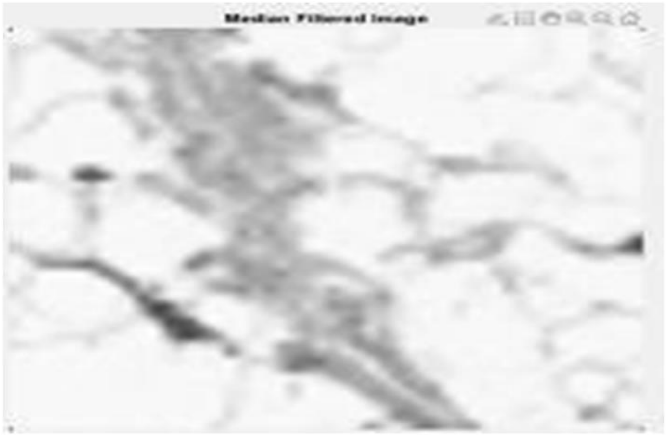

Fig 4 MEDIAN FILTERED IMAGE

Median filter is used to remove noise from an image or signal.ROI is used to find the particular area of an image.

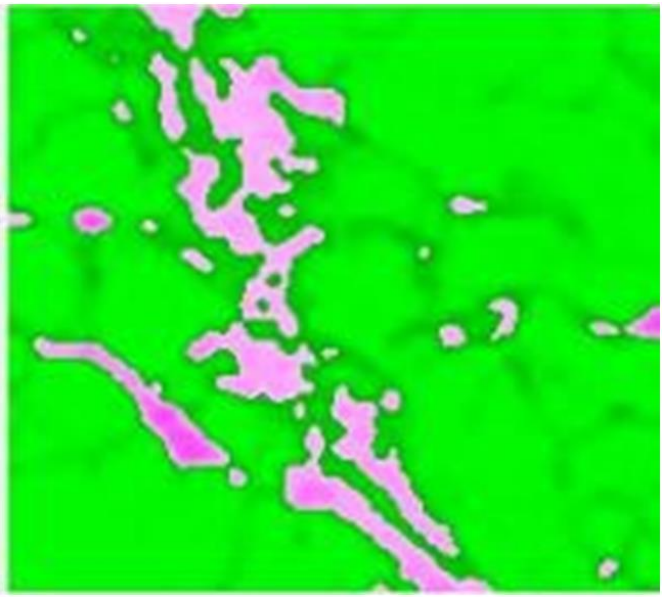

Fig 5 PREDICTED TUMOR IMAGE

CNN is used to classify the image and predict whether the tumor is malignant or benign.

METHOD	FEATURE	CLASSIFIER	NO OF IMAGES	ACCURACY %
Genetic algorithm	Region based	Decision tree	196	91%
Naïve Bayes	Shape based	SVM	226	93%
Linear Regression	Intensity based	SVM	322	95%
Logistic Regression	Size and Volume based	CNN	569	97%

TABLE 1 COMPARSION WITH OTHER METHODS

IV.CONCLUSION

Data Mining could be used to predict the kind of tumor the patient is suffering from in the predict the class of cancer to which a patient may classified, we need to extract the hidden knowledge pertaining to various attributes that could be used to boost the efficiency in general by utilizing the best resources available. In our paper, CNN and Simple Logistic regression classifier is used. Results conclude that Simple Logistic regression method obtains the Best Model to predict breast cancer by means of different data mining techniques.Results indicate that Simple Logistic regression obtained best performance in general compared to the other classifiers in terms of classification accuracy, RMSE, specificity and sensitivity, F-measure, ROC curve area, time taken to build the model and Kappa Statistics.

V. REFERENCES

[1]. Akay.M.F.,(2009) "Support vector machines combined with feature selection for breast cancer diagnosis," Expert Syst. With Applicat., vol. 36, no. 2, pt. 2, pp. 3240–3247, Mar.

[2]. Andolina.V., Lillé.S., Willison.K.,(2009) Mammographic Imaging: A Practical Guide, Williams & Wilkins, Lippincott, vol 13 ,pp.466-467.

[3]. Bandyopadhyay.S.K., Banerjee.S., Maitra.I.K.,(2010) Digital imaging in pathology towards detection and analysis of human breast cancer. International conference on computational intelligence, communication systems and networks (CICSYN), Liverpool, United Kingdom, pp. 295–300.

[4]. Ferlay J, Shin H, Bray F, Forman D, Mathers C, Parkin D. GLOBOCAN (2008), Cancer incidence and mortality worldwide, pp 109-110.

[5]. Fonseca, P., Mendoza, J., Wainer, J., Ferrer, J., Pinto, J., Guerrero, J., & Castaneda, B. (2015). Automatic breast density classification using a

convolutional neural network architecture search procedure. In SPIE Medical Imaging.vol 34,pp. 233-350.

[6]. Garcia.M., Jemal.A., E. Ward.M. et al.,(2007) Global Cancer Facts & Figures 2007, American Cancer Society, Atlanta, Ga, USA ,pp 311-400.

[7]. Herranz.M., Ruibal.A.,(2012) Optical imaging in breast cancer diagnosis: the next evolution. Hindawi Publ. Corp. J. Oncol.vol 25, pp 300-450.

[8]. Jackson A, Buckley DL, Parker GJM.(2005) Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Oncology, Berl in germany,pp.440-500.

[9]. kim.B,(2017) Breast Cancer Organization, "Breast cancer treatment causes severe side effects in many woman", Ardmore: Breast Cancer Organization, February.vol 45,pp 514-545.

[10]. Kisilev, P., Sason, E., Barkan, E., Hashoul, S., (2016). Medical image description usingmulti-task-loss CNN. In: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis. Springer, pp. 121–129.