Transactions on Networks and Communications - Vol. 9, No. 4
Publication Date: August, 25, 2021
DOI: 10.14738/tnc.94.10663.
Idowu, I. R., David, A. O., Elelu, S., & Adefisayo, T. O. (2021). Prediction of Breast Cancer Images Classification Using Bidirectional Long Short Term Memory and Two-Dimensional Convolutional Neural networ. Transactions on Networks and Communicaitons, 9(4). 29-38.

Prediction of Breast Cancer Images Classification Using Bidirectional Long Short Term Memory and Two-Dimensional Convolutional Neural network

Ifedotun Roseline Idowu
Department of Computer Science,
Federal College of Animal Health and Production Technology
Moor Plantation, Ibadan, Nigeria

Adeniji Oluwashola David
Department of Computer Science, University of Ibadan
Ibadan, Nigeria

Shehu-Alimi Elelu
Department of Chemistry, Delaware State University
Dover Delaware, USA

Togun Oluwapelumi Adefisayo
Department of Computer Science,
Federal College of Animal Health and Production Technology
Moor Plantation, Ibadan, Nigeria

ABSTRACT

Breast cancer is most prevalent among women around the world and Nigeria is no exception in this menace. The increased in survival rate is due to the dramatic advancement in the screening methods, early diagnosis, and discovery in cancer treatments. There is an improvement in different strategies of breast cancer classification. A model for training deep neural networks for classification of breast cancer in histopathological images was developed in this study. However, this images are affected by data unbalance with the support of active learning. The output of the neural network on unlabeled samples was used to calculate weighted information entropy. It is utilized as uncertainty score for automatic selecting both samples with high and low confidence. A threshold that decays over iteration number is used to decide which high confidence samples should be concatenated with manually labeled samples and then used infine-tuning of convolutional neural network. The neural network was optionally trained using weighted cross-entropy loss to better cope with bias towards the majority class. The developed model was compared with the existing model. The accuracy level of 98.3% was achieved for the developed model while the existing model 93.97%. The accuracy gain of 4.33%. was achieved as performance in the prediction of breast cancer.

Keywords: breast cancer, image classification, prediction, biLSTM, conv2D

INTRODUCTION

In Africa, cervical cancer was the commonest cause of cancer- related deaths among women for several decades but breast cancer is the leading cause of cancer related deaths among Nigerian women. This is not due to a reduction in cervical cancer but an increase in the incidence of breast cancer. Breast cancer is a malignant cancerous) growth that begins in the tissues of the breast. Cancer is a disease in which abnormal cells grow in an uncontrolled way (Siegel, Miller & Jemal, 2016).It also affects one in eight women during their lives. It is the commonest site specific malignancy affecting women that leads to cancer mortality in women. Breast cancer is commonly seen in four stages that represents its progression (Lee et al, 2010).Deep learning with convolutional neural networks (CNNs) has proved to be a powerful technique in general object recognition, learning high-level image features directly from images, and yielding improved image classification performance in (Li et al. 2012).

RELATED WORK

Classification is one of the most important and essential task in machine learning and data mining. Research has been conducted to apply data mining and machine learning on different medical datasets to classify Breast Cancer. Many of them show good classification accuracy. (Chaurasia & Pal, 2014). compared the performance criterion of supervised learning classifiers; such as Naïve Bayes, SVM-RBF kernel, RBF neural networks, Decision trees (J48) and simple CART; to find the best classifier in breast cancer datasets. The experimental result shows that SVM-RBF kernel is more accurate than other classifiers; it scores accuracy of 96.84% in Wisconsin Breast Cancer (original) datasets.

(Djebbari, Liu, Phan & Famili, 2008), consider the effect of ensemble of machine learning techniques to predict the survival time in breast cancer. Their technique shows better accuracy on their breast cancer data set comparing to previous results.

(Hiba,Hajar,Hassa &Thomas,2016), compare the performance of C4.5, Naïve Bayes, Support Vector Machine (SVM) and K- Nearest Neighbor (K-NN) to find the best classifier in WBC. SVM proves to be the most accurate classifier with accuracy of 96.99%.

(Christopher, Yetunde & Dada, 2018), achieve accuracy of 69.23% using decision tree classifier (CART) in breast cancer datasets. The accuracy of data mining algorithms SVM, IBK, BF Tree was compared by (Andhra, 2011).The performance of SMO shows a higher value compared with other classifiers.

(Thorsten, 1999) reaches accuracy of 95.06% with neuron fuzzy techniques when using Wisconsin Breast Cancer (original) datasets. In this study, a hybrid method is proposed to enhance the classification accuracy of Wisconsin Breast Cancer (original) datasets (95.96) with 10 fold cross validation.

(Liu , Cheng & Lu,2009) experimented on breast cancer data using C5 algorithm with bagging; by generating additional data for training from the original set using combinations with repetitions to produce multisets of the same size as you're the original data; to predict breast cancer survivability.

(Delen , Walker , Kadam,2005), took 202,932 breast cancer patients records , which then pre-classified into two groups of "survived" (93,273) and "not survived" (109,659). The results of predicting the survivability were in the range of 93% accuracy.

(Spanhol, Oliveira, Petitjean,Heutte,2016), investigated a dataset of 7909 breast cancer histopathology images acquired on 82 patients, which is now publicly available from http://web.inf.ufpr.br/vri/breast-cancer-database. The dataset includes both benign and malignant images. The results obtained with state-of-the-art image classification systems had an accuracy ranges from 80% to 85%. (Vikas & Thimmaraju, 2019), applied support vector machine (SVM) with diverse dataset on breast cancer diagnosis and classification. Their experimental procedure is based on the following Technique. (1) The given dataset is isolated into 85% preparing and 15% testing sets dependent on the 10-overlay cross approval system (2) they visualize the data using density plots to get a sense of the information conveyance. (3) The testing set is encouraged into the built classifiers preceding examination of their classification accuracy, precision, and F-measure rates. (4) The classifier preparing times are likewise contrasted with investigate the computational complexities of preparing distinctive classifiers. A graph is plotted to compare performance of SVM, KNN and Gaussian Naive Bayes. They claimed that SVM reached an accuracy of 98% which I feel is credited to model overfitting. (Koziarski, Kwolek& Cyganek, 2019), conducted an investigation on Breast Cancer Classification on histopathological images affected by data imbalance using active learning and convolutional neural network. They reported an output of the neural network on unlabeled samples and used it to calculate weighted information entropy. The output is utilized as uncertainty score for automatic selecting both samples with high and low confidence, on the initial training pool consisting of only labeled data, the algorithm based on entropy achieved 86% accuracy, whereas algorithm based on weighted entropy achieved 83.9% accuracy. In seventh iteration the entropy-based algorithm learned on 3035 samples from PL pool and 3146 training samples from PH pool, and achieved on such training data 93.3% classification accuracy, whereas weighted entropy-based algorithm learned on 3035 samples from PL pool and 3335 samples from PH pool and achieved 94.62% classification accuracy.

MATERIALS AND METHODS

Description of the Existing model

The Existing model create the dataset, fluid samples taken from patients with solid breast masses and an easy-to-use graphical computer program called 'Xcyt' which is capable of performing the analysis of cytological features based on a digital scan. The program uses a curve-fitting algorithm, to compute ten features from each one of the cells in the sample, then it calculates the mean value, extreme value and standard error of each feature for the image, returning a 30 real-valuated vector. (UCI Repository Information)

Description of the Developed model

The Developed model combines a bidirectional long short term memory encoder and a two dimensional convolutional neural network. The bidirectional long short term memory will actively learn the dataset and the two dimensional convolutional neural network will classify the dataset into the required class of either malignant or benign. The procedure for the developed model is shown below.

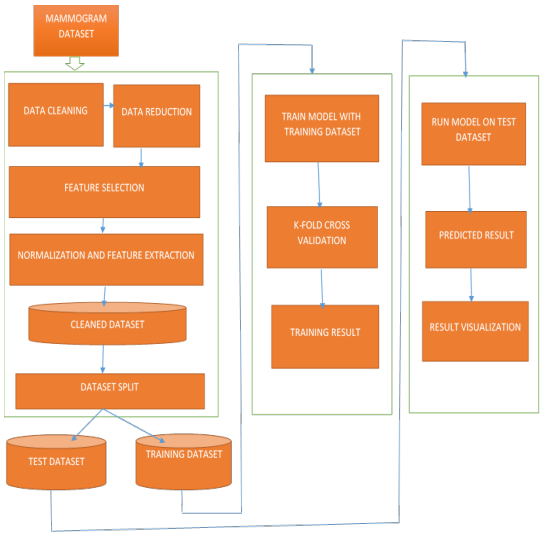


Figure 1: Procedure of the Developed model

Attribute Information required are:

1. ID number
2. Diagnosis (M = malignant, B = benign)

Ten real-valued features are computed for each cell nucleus:

1. radius (mean of distances from center to points on the perimeter)
2. texture (standard deviation of gray-scale values)
3. perimeter
4. area
5. smoothness (local variation in radius lengths)
6. compactness ($perimeter^2$ / area — 1.0)
7. concavity (severity of concave portions of the contour)
8. concave points (number of concave portions of the contour)
9. symmetry
10. fractal dimension ("coastline approximation" — 1)

The mean, standard error and "worst" or largest (mean of the three largest values) of these features were computed for each image, resulting in 30 features. For instance, field 3 is Mean Radius, field 13 is Radius SE, field 23 is Worst Radius.

Data mining algorithms for the model

CNN is one of most commonly used connectionism model for classification. Connectionism models focus on learning from environmental stimuli and storing this information in a form of connections between neurons Figure 2. The weights in a neural network were adjusted according to the training data by some learning algorithm. That is, the greater the difference in the training data, the more difficult for the learning algorithm to adapt the training data, and the worse classification results it will produce.

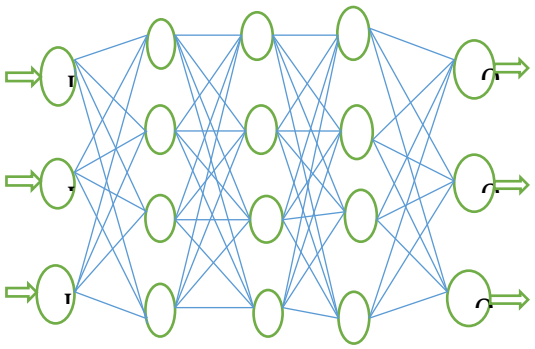


Figure 2: Convolutional Neural Networks

In the convolution layer, multiple filters with different window size move on the output from the Bi-LSTM to perform one-dimensional convolution. As the filter moves on, many sequences, which capture the precipitation data features in the filtered n-gram, are generated. Many feature sequences are combined into a feature map. In the pooling layer, a max-overtime pooling operation is applied to capture the most useful local features from feature maps. Activation functions are added to incorporate element-wise non-linearity. The outputs of multiple filters are concatenated in the merge layer. After another dropout process, a fully connected *softmax* layer output the probability distribution over labels from multiple classes.
To create a BiLSTM network for sequence-to-label classification, the study perform the following:

1. First create a layer array containing a sequence input layer, a BiLSTM layer, a fully connected layer, a softmax layer, and a classification output layer.
2. Specify the size of the sequence input layer to be the number of features of the input data.
3. Specify the size of the fully connected layer to be the number of classes. We do not need to specify the sequence length. And,
4. For the BiLSTM layer, we specify the number of hidden units and the output mode 'last'.

To train the one dimensional neural network to classify sequence data, we use the output from the BiLSTM network. The BiLSTM network enables us to input sequence data into the two dimensional neural network, and make predictions based on the individual time steps of the sequence data. Next, we load the training data from the output of the BiLSTM. Let's assume an array variable called XTrain. XTrain is a cell array containing say N sequences of dimension 32. Y is a categorical vector of labels, which correspond to the 32 features. The entries in XTrain are matrices with 32 rows (one row for each feature) and varying number of columns (one column for each time step). This will allow the two dimensional convolutional neural network to analyze and classifier the data to make a prediction. The flow of the model is given below

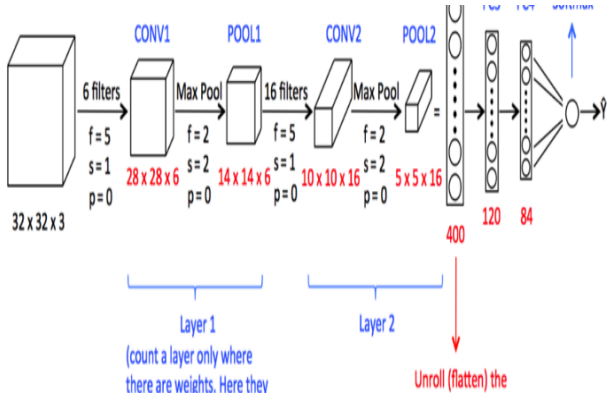


Figure 3: The flow on the network

The states, can treat the boundary points differently, which is helpful in many cases. Also the states as the hidden representations of moving objects, a BiLSTM and 2dCNN with a larger transitional kernel should be able to capture faster motions while one with a smaller kernel can capture slower motions. To ensure that the states have the same number of rows and same number of columns as the inputs, padding is needed before applying the convolution operation. Here, padding of the hidden states on the boundary points can be viewed as using the state of the outside world for calculation. The input comes first followed by initialization of all the states of the BiLSTM to zero which corresponds to "total ignorance" of the future. Similarly, if zero padding is performed on the hidden states, the actual setting of the state outside world to zero will assume no prior knowledge.

Experiment for the Model

The mammogram images from the dataset were labelled as depicted. It also shows the predicted label by the model.

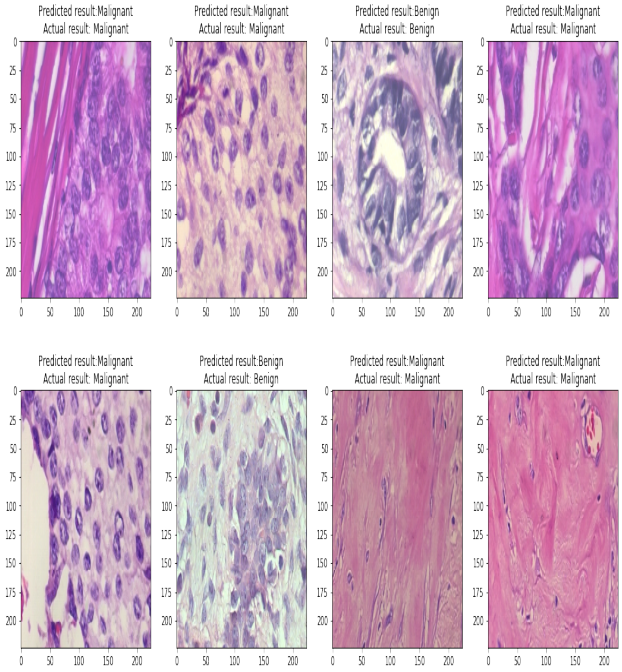


Figure 4: Mammogram image of the breast cancer dataset

The Table 1 provide the Subclass samples for the two classification of dataset with relation to patient.

Table 1: Classification of Dataset of Subclass samples.

Classes	Subclasses	Number of Patients	Magnification factors				Total
			40×	100×	200×	400×	
Benign	A	4	114	113	111	106	444
	F	10	253	260	264	237	1014
	TA	3	109	121	108	115	453
	PT	7	149	150	140	130	569
Malignant	DC	38	864	903	896	788	3451
	LC	5	156	170	163	137	626
	MC	9	205	222	196	169	792
	PC	6	145	142	135	138	560
Total		82	1995	2081	2013	1820	7909

This dataset contains 7909 samples each falling within two main classes: benign or malignant. The benign subset contains 2440 samples and the malignant subset contains 5429 samples. The samples are collected from 82 patients with different magnification factors including 40×, 100×, 200×, 400×. Some of the example images with a 400× magnification factor are shown in table 1. Each class has four subclass, the four types of benign cancer are Adenosis (A). Fibroadenoma (F), Tubular Adenoma (TA), and Phyllodes Tumor (PT). The four subclasses of malignant cancer are Ductal Carcinoma (DC), Lobular Carcinoma (LC), Mucinous Carcinoma (MC), and Papillary Carcinoma (PC).

RESULT AND DISCUSSION

In this experiment, 70% of the samples were used for training and 30% of the samples for testing of the dataset in table 1. To generalize the classification task to perform successfully when testing new patients, it was ensured that the patients selected for training were not used during testing. With the experimental design by Bogdan et al, the model reported the average accuracy after successfully completing five trials. The loss function of the training dataset in the experiments using both image-wise and patch-wise evaluation was conducted. The labels of the patches are considered to have same class label as the original images. On the other hand, using patch-wise approach: first, 128×128 pixel center patches were cropped from an input sample. Second, the augmentation techniques were applied on the center patches and 20 augmented samples per patch were generated. Third, we evaluated the model with 200 randomly selected patches with a size of 128×128 pixels from a single image. The *Loss function of the trained Dataset is below.*

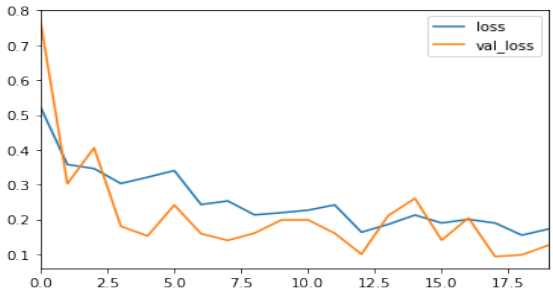


Figure 5: Loss function of the Dataset

The down sampled and center patches are shown for two different input samples. In the first experiment, the dataset was trained with the developed model using the stochastic gradient descent (SGD) optimization function. The study set the momentum to 0.9 and decay is calculated based on the initial learning rate and number epochs of the respective trial. The experiment for the three trials where 500 epochs are used in each trail. After 500 epochs, the learning rate is decreased by the factor of 10.

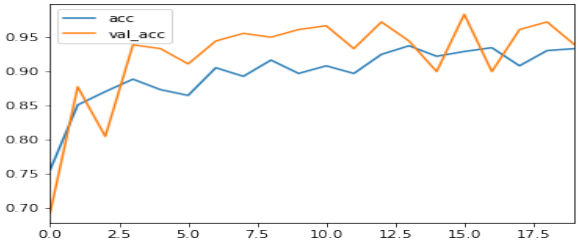


Figure 6: Accuracy Prediction of Test Dataset

To achieve an accurate prediction, the augmented of the dataset with patches to balance gives a better prediction. The Figure7 shows the confusion matrix of the proposed model. 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.

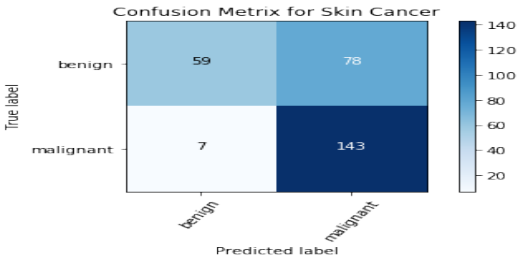


Figure 7: Prediction of Confusion Matrix

The 45 degree line shown in figure 8 is the random line, where the Area Under the Curve or AUC is 0.5. To further the curve from this line, the higher the AUC and better the model. The highest a model can get is an AUC of 1, where the curve forms a right angled triangle. The ROC curve can also help debug a model. For example, if the bottom left corner of the curve is closer to the random line, it implies that the model is misclassifying at Y=0. Whereas, if it is random on the top right, it implies the errors are occurring at Y=1.

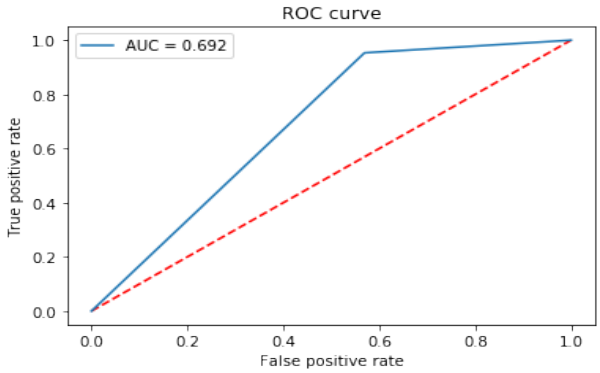


Figure 8: Area under the curve (AUC)

The result of the Precision, F-Score and Recall is given in table 1 below: Table 2 result of Precision, F - Score and Recall for the developed solution.

Table 2: Analysis of Result for the Model

Accuracy	Precision	Recall	F1 score	ROC-AUC
98.3%	0.65	0.95	0.77	0.692

In the table above, the study pick the best results for each of the execution of the model. This includes First run, second run Third run, forth run and Final run.

The evaluation of the developed model with the work of Bogdan et al method is tabulated in table 3. The result of the comparison is presented as follow:

Table 3: Performance Evaluation of both Models

MODEL	*DEVELOPED MODEL*	*BOGDAN ET AL*
ACCURACY	***98.3%***	***93.97%***

The developed model perform better with an accuracy gain of 4.33%. i.e accuracy level of 98.3% while Bogdan et al reported an accuracy level of 93.97%.The performance gain was mainly due to the fact that a different method of learning the dataset before actually performing the prediction was adopted.

CONCLUSION

In this research work, an active learning method on breast cancer dataset was developed using a bidirectional long short term memory (BiLSTM) and a two dimensional convolutional neural network model (Conv2D). The performance of the developed method was evaluated via image level, patient level, image based, and patch based analysis. The developed model perform better with an accuracy gain of 4.33%.This research work has improved the prediction of breast cancer. It will be interesting to see whether it will be able to differentiate between overlapping and single cancer cells. In the future, a study of this problem is recommended such as cell overlap and uneven color distribution in the pathological images of breast cancer obtained from different staining methods.

References

[1] Andhra Pradesh, (2011) "Analysis of Feature Selection with Classification: Breast Cancer Datasets," Indian Journal Computer Science& Engineering, vol. 2, no. 5, pp. 756–763.

[2] Chaurasia, Vikas & Pal, Saurabh. (2014). Data mining techniques: To predict and resolve breast cancer survivability. 3. 10-22. https://www.researchgate.net/publication/290483856_Data_mining_techniques_To_predict_and_resolve_breast_cancer_survivability

[3] Christopher, Ngene & Yetunde, Daramola & Dada, Emmanuel. (2018). Performance Comparison of Machine Learning Techniques for Breast Cancer Detection. 10.20286/nova-jeas-060105.

[4] Delen D, Walker G, Kadam A. (2005), Predicting breast cancer survivability: a comparison of three data mining methods. Artificial Intelligence in Medicine. 2005 Jun; 34(2):113-27.doi;10.1016/j.artmed.2004.07.002.PMID:15894176

[5] Djebbari, Amira & Liu, Ziying & Phan, Sieu & Famili, Fazel. (2008). An ensemble machine learning approach to predict survival in breast cancer. International journal of computational biology and drug design. 1. 275-94. 10.1504/IJCBDD.2008.021422.

[6] Hiba Asria ,Hajar Mousannifb ,Hassan Al Moatassime c ,Thomas Noeld.: Using machine learning algorithms for breast cancer risk prediction and diagnosis. Procedia Computer Science 83 (2016) 1064 – 1069

[7] Komura, D., Ishikawa, S.(2018): Machine learning methods for histopathological image analysis. Computational and Structural Biotechnology Journal 16 (2018) 34 – 42 8.

[8] Koziarski, Michał & Kwolek, Bogdan & Cyganek, Boguslaw. (2019). Convolutional Neural Network-Based Classification of Histopathological Images Affected by Data Imbalance: 17th International Conference, ICAISC 2018, Zakopane, Poland, June 3-7, 2018, Proceedings, Part II. 10.1007/978-3-030-12177-8_1.

[9] Lee CH, Dershaw DD, Kopans D, Evans P, Monsees B, Monticciolo D, Brenner RJ, Bassett L, Berg W, Feig S, Hendrick E, Mendelson E, D'Orsi C, Sickles E, Burhenne LW.(2009), Breast cancer screening with imaging: recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. J Am Coll Radiol. 2010 Jan;7(1):18-27. doi: 10.1016/j.jacr.2009.09.022. PMID: 20129267.

[10] Li H, Giger ML, Lan L, Bancroft Brown J, MacMahon A, Mussman M, Olopade OI, Sennett C. (2012) Computerized analysis of mammographic parenchymal patterns on a large clinical dataset of full-field digital mammograms: robustness study with two high-risk datasets. J Digit Imaging. 2012 Oct;25(5):591-8. doi: 10.1007/s10278-012-9452-z. PMID: 22246204; PMCID: PMC3447101.

[11] Liu, Ya-qin, Wang,Cheng, and Zhang, Lu, (2009),"Decision tree based predictive models for breast cancer survivability on imbalanced data," 3rd International Conference on Bioinformatics and Biomedical Engineering book ISBN :978-1-4244-2901-1book e-ISBN :978-1-4244-2902-8DOI10.1109/ICBBE.2009.5162571pp. 1–4,

[12] Peter Adebayo Idowu, Kehinde Oladipo Williams, Jeremiah Ademola Balogun & Adeniran Ishola Oluwaranti (2015). Breast cancer risk prediction using data mining classification techniques on transactions on networks and communication volume 3, Issue 2 issn 2054 -7420 DOI: 10.14738/tnc.32.662PublicationDate:10thMarch2015URL: http://dx.doi.org/10.14738/tnc.32.662

[13] Siegel R. L., Miller K. D., Jemal A., "Cancer statistics, 2016," CA Cancer J. Clin. 66(1), 7–30 (2016).10.3322/caac.21332 [PubMed] [CrossRef] [Google Scholar]

[14] Spanhol, F., Oliveira, L., Petitjean, C., Heutte, L. (2016): A dataset for breast cancer histopathological image classification. IEEE Transactions on Biomedical Engineering, VOL. 63, NO. 7, 1455–1462 9.

[15] Thorsten Joachims (1999). Transductive Inference for Text Classification Using Support Vector Machines. Icml. 1999; 99:200-209. doi:10.4218/etrij.10.0109.0425.

[16] Vikas S,Thimmaraju S N, (2019). "Breast Cancer Diagnosis and Classification Using Support Vector machines with Diverse Datasets". International Journal of Computer Sciences and Engineering, 7(4), 442 -446