Advances in Parallel Computing Algorithms, Tools and Paradigms
D.J. Hemanth et al. (Eds.)

doi:10.3233/APC220043

Detection of Tumor for Breast Cancer Patient Using Regression Tree Compared with Linear SVM Algorithm

R. Saimonika[a] and P.Sriramya[b,1]
[a]*Research Scholar, Dept. of CSE, Saveetha School of Engineering,*
[b]*Professor, Dept. of AI&DS, Saveetha School of Engineering,*
[a.b]*SIMATS, Chennai, Tamil Nadu, India*

Abstract: The paper aims to detect the tumor in breast cancer images using a Regression Tree and compare it with a Support Vector Machine to evaluate accuracy values.The regression tree algorithm is applied to 10 sample images from a dataset of more than 250 images. For the same input images taken accuracy, values are evaluated. The dataset taken consists of breast cancer images in this research study a Novel Regression Tree and Linear Support Vector Machine is evaluated with a total of 20 sample sizes. Based on the statistical analysis the significance value for calculating accuracy was found to be $p<0.05$(significant two-tailedtests). Detection of breast cancer was performed by using a Regression Tree and Support Vector Machines which achieved the accuracy of 95% and 86% respectively. By the above two results, we conclude that the performance of the Novel Regression tree is significantly better than the Linear Support Vector Machine in terms of accuracy.

Keywords. Linear Support Vector Regression, Novel Regression tree, Machine learning, Breast cancer diagnosis, Detection of the tumor.

1. Introduction

To detect breast cancer with an algorithm using a regression tree and comparing to evaluate accuracy using a support vector machine [1]. The new diagnostic system at the Wisconsin hospitals was trained using patient samples [1,2]. The gender population breast cancer is the most prevalent cancer cause but genetic and environmental factors are most likely to blame and is used as a mathematical tool for predicting survival and the factors that inflame have been increasingly used to predict data in recent decades [3]. The first significant cancer is the presence of masses calculation of um sediments in the breast tissue is the second pointer to calculation in the mammogram description; these appear as tiny bright spots [4]. The development of additional cells results in the formation of a lump of tissue, the early cancer detection is more than important mammography isa screening procedure for breast cancer [5].

A previously identified and tested algorithm was used to classify possible subjects by evaluating Medicare claims for newly diagnosed breast cancer cases patients with

[1]P.Sriramya, Dept. of AI&DS, Saveetha School of Engineering, SIMATS, Chennai, India. E-mail: sriramyap@saveetha.com

breast cancer were recruited to take part in a survey evaluating breast cancer outcomes [6]. The existing technique used to develop the classification system had a high accuracy rate, projected a better survival ability of women diagnosed with breast cancer, and could be used as a guide when developing a medical decision [7]. To minimize the number of patients that need to be tested for Oncotype DX [8]. A regression tree is used to evaluate the innovative prediction to determine whether the variable combinations are correlated with greater or smaller predicted values incross-validationof arecursive manner [9]. The proposed system was obtained using discriminant analysis, artificial neural network and multivariate adaptive regression splines, and a visible alternative in the diagnosis of breast cancer [10]. A descriptive study was carried out on women who had been diagnosed with breast cancer tumor type, the histologic and nuclear grading axillary nodal status lymph node involvement [11]. The drawback of the existing system is to identify the tumor where the cell is present and the tumor will be detected with a low-efficiency image. They are often relatively inaccurate diagnoses. The main aim of the study is to improve the accuracy of detection of tumors in breast cancer images by using a Regression tree compared with Linear Support Vector Machine regression.

2. Materials and Methods

The research was carried out at SIMATS' Saveetha School of Engineering, Department of Computer Science and Engineering. A sample of more than 250 photos was used in the procedure. Two groups were tested to see how accurate they were in detecting images. To improve accuracy, each group was subjected to a total of ten iterations. The Support vector machine was used in Group 1 and the Regression tree technique was used in Group 2. The G power calculation is used to calculate the required samples for this assay. [12]. The minimum power of the analysis is fixed as 0.8 and the maximum accepted error is fixed as 0.5. Implemented breast cancer detection using a Regression tree with a sample size of 20 is taken for testing and divided into 2 groups tested.

Table 1. Different types of intensity values (Malignant, Benign, or Normal) for the 5 sample images taken

Image No	Malignant	Benign	Normal
Image 1	0	1	0
Image 2	0	0	1
Image 3	1	0	0
Image 4	1	0	0
Image 5	0	1	0

In the **Table 1** The dataset was taken from the standard medical database kaggle. The parameters include tumor size, histologic type, histologic grade, axillary lymph node status, and tumor growth function. The independent variables are texture_mean, area_mean, smoothness_mean, concave_mean, concavity_mean, compactness_mean, perimeter_mean, fractal_dimension_mean and the dependent variables are texture_se, area_se, smoothness_se, concave_se, concavity_se, compactness_se, perimeter_se, fractal_dimension_se. The training dataset is 70% and the testing dataset is 30%. The Figure. 1 shows the process flow of the tumor detection process.

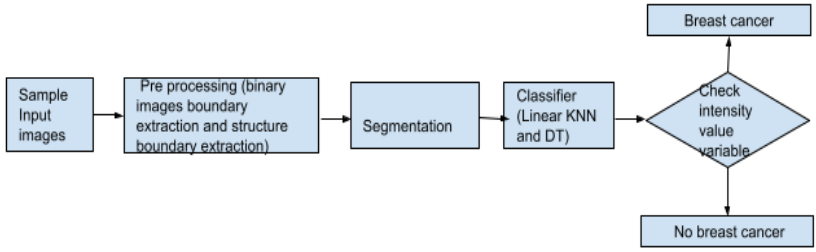

Figure 1. Overall process flow of breast cancer tumor detection

2.1. Support Vector Machine

An SVM classifier (SVM) is a binary linear classification with an explicitly defined decision boundary to reduce generalization error. It's a sophisticated and adaptable Machine Learning model that can handle linear and nonlinear classification, regression, and even outlier detection. SVM is highly suited to the categorization of complicated datasets that are small or medium in size.

The intensity of the tumor region is extracted using binary images boundary extraction and structure boundary extraction the noise will be removed. Cancer refers to any malignant tumor that arises from the cell in form breast. To eliminate the undesirable piece of the image and section the part with higher thickness it is reducing a large amount of data. The tumor detection method follows the process and filter of using breast cancer.

2.2. Regression Tree Algorithm

The patient diagnoses are made by cross-referencing new attribute values in the regression tree and maintaining the path until the tree node is reached, which indicates whether the tumor is benign or malignant. Breast cancer patients will initially only experience one or two symptoms and signs. The presence of these signs and symptoms does not necessarily imply that the individual has breast cancer. It is a method of reducing large amounts of data. If the machine learning algorithm is too large to perform then it is transformed into a set of features the tumor will be detected and can identify the type of tumor. The mammography image is converted into new pixels and calculation is applied to contribute to the image goal.

The mammography image is converted into new pixels and calculation is applied to contribute to the image goal. The purpose of software specifications is to identify the resources that must be installed on a device for an application to run. Before the software can be installed, certain prerequisites must be installed. Below are the functional requirements at a bare minimum: Matlab version 2019 for Windows.

In addition to **Table 2**, experimental research, the work was statistically analyzed using IBM's Statistical Package for Social Sciences (SPSS). The average, standard deviation, and standard error mean were calculated. To compare the two groups, an independent sample T-test analysis was used. The "IBM SPSS Independent T-test Analysis" was used to perform statistical analysis for two independent variables utilizing a support vector machine and a regression tree.

Table 2. Experimental Results of the accuracy achieved for Regression tree and Linear Support Vector machine regression

Iterations	Accuracy (%)	
	Regression Tree	Linear SVM
1	95.1328	86.5328
2	95.1342	86.5342
3	95.1370	86.5370
4	95.1313	86.5393
5	95.1393	86.5370
6	95.1338	86.5297
7	95.1345	86.5310
8	95.1374	86.5460
9	95.1406	86.5322
10	95.1378	86.5374

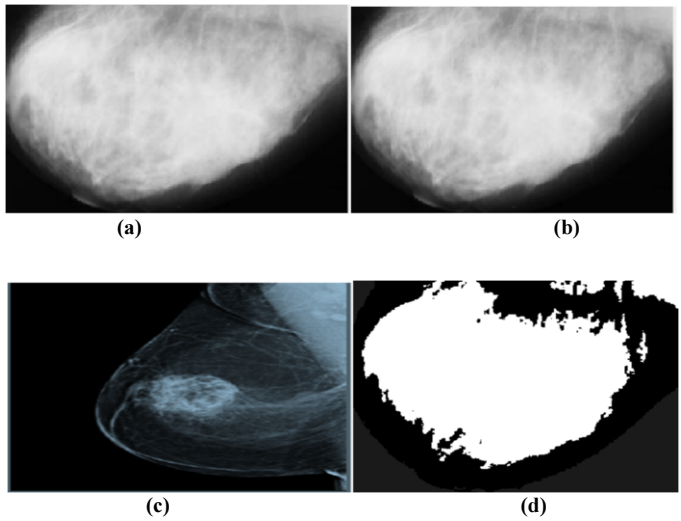

Figure 2. Simulation results of Regression tree-based breast cancer detection system (a) Input image (b)preprocessing to remove the noise (c) Segmented image (d) Tumor detected image

3. Results

In the Figure 2 gives the method of noise removal that is used to transform an image into a binary image during preprocessing. The intensity tumor may be expressed. The tumor region is extracted using binary images boundary extraction and structure boundary extraction The tumor noise will be removed in this way of data. If the algorithm is too large to perform then it is transformed into a set of features. The tumor will be detected and have to identify the type of tumor whether it is malignant or benign or normal breast cancer. Figure 2(a) shows the input database is a preprocessing segment. Figure 2(b) Remove the noise with an adaptive median filter from the input image. Figure 2(c) The threshold value is converted into new pixels and forms the

tumor with line formation. Figure 2(d) identifies cancer with that image whether it is malignant or benign or normal.

The breast cancer detection accuracy of the Regression tree was much superior to Linear Support vector machine regression, as shown in Table 1. The regression tree's detection accuracy was shown to be significantly superior to the support vector machine 0.05 in all iterations. From the above values, it was clear that the regression tree machine learning algorithm performed significantly better than support vector machine regression. In **Table 3**, tells about the Group Statistics: The mean accuracy, standard deviation, and standard error mean have been calculated for the T-Test sample. The sample size of each algorithm is 10. The mean accuracy of the Voting Regression is significantly high when compared with the Linear SVM. The mean of Regression Tree is 95% more compared with Linear Support Vector Machine 86% and std. Error Mean for Regression Tree is 0.1388 and Linear Support Vector Machine is 0.1336.

Table 3. Group Statistics for Regression Tree is significantly high when compared with the Linear SVM.

Algorithm	N	Mean	Std.Deviation	Std.Error Mean
Regression Tree	10	95.1369	.0042	.00132
Linear Support Vector Machine Regression	10	86.53549	.0049	.00155

The confidence interval of the difference as lower and upper values range as shown in **Table 4**. The table 3 shows an independent sample T-test comparing significance level with p0.05 and fixed confidence interval (regression tree performs considerably better than Linear SVM with $p<0.05$).

Table 4. Statistical analysis of mean and standard deviation for Decision tree and K-neighbourhood regression independent sample test got a significance of 0.000 which is less than(0.05).

		F	Sig	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
r -square	Equal variances	.251	.622	4198 511	18	.000	8.6014 60675	0.00204 86929	8.5971 56531	8.6057 64819
	Equal variances Not assumed			4198 511	17.560	.000	8.6014 60675	0.00204 86929	8.5971 48795	8.6057 72554

In the **Figure 3,** shows a bar chart comparing the mean accuracy of breast cancer diagnosis computed using a Regression Tree vs a Linear Support Vector Machine Regression. The regression tree technique and the SVM approach are compared. The regression tree's mean accuracy is higher than SVM, while its standard deviation is slightly lower. With its standard deviation, support vector machines appear to offer the most varied outcomes; the difference between Regression Tree and Linear Support Vector Machine Regression is statistically significant. According to the graph, the Regression Tree has higher accuracy than the Linear Support Vector Machine regression.

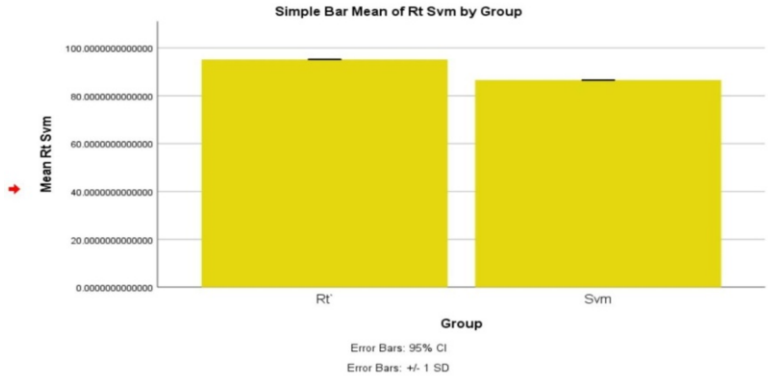

Figure 3. X-Axis: Regression tree vs SVM Algorithm and Y-Axis: Mean accuracy of detection ± 1 SD.

4. Discussion

Based on the significance value obtained in statistical analysis, the Novel Regression tree technique has significantly greater accuracy than the Support vector machines in this study. The Mean Std.deviation and Std.Error Mean values for the Novel regression tree and support vector machine regression techniques are shown in Table 4. Figure. 3 shows the mean error of accuracy is to be lesser in a Novel Regression Tree (95%) than in Linear Support Vector Machine regression (86%).

In this research breast cancer was based on the images using KNN compared with a genetic algorithm for the classification withan accuracy of 79% [13]. This paper presents information about breast cancer based on the radial basis and gaussian mixture model for the classification of the accuracy of 89% [14]. This article is about breast cancer cells using Naive Bayes and SVM and ensemble algorithms. It has increased the accuracy by 93% [15]. In this article the breast cancer diagnosis using SVM and KNN classifier with the classification of the accuracy of 98% [16]. Although the study's results are superior in both experimental and statistical analysis, it does have certain drawbacks. Breast cancer survivors have restrictions in terms of their functional abilities. Disease-related inability to carry out typical everyday activities. Treatments are more likely to die as a result of poor overall health. The noise from the tumor will be removed by breast cancer. The tumor will be detected at a particular point. Identify the type of tumor in this benign cancer. The Regression Tree has a high accuracy value and it has the image quality to identify the tumor.

5. Conclusion

Because of their incapacity to conduct routine everyday tasks as a result of their condition or treatment, they are more likely to die as a result of their overall bad health. Image processing is used for enhancing accuracy in early innovative prediction of breast cancer using a Novel Regression Tree compared with Linear Support Vector Machine Regression. The significant value p less than 0.05 is achieved based on the statistical analysis performed and the Independent sample test run outcomes. As a

result, the Regression Tree approach performs better than the Linear Support Vector Machine Algorithm in terms of accuracy.

References

[1] Filipczuk P, Stevens T, Krzyzak A, Monczak R. Computer-Aided Breast Cancer Diagnosis Based on the Analysis of Cytological Images of Fine Needle Biopsies. IEEE Trans Med Imaging. 2013 Dec;32(12):2169–78.

[2] Mangasarian OL, Nick Street W, Wolberg WH. Breast Cancer Diagnosis and Prognosis Via Linear Programming [Internet]. Vol. 43, Operations Research. 1995. p. 570–7. Available from: http://dx.doi.org/10.1287/opre.43.4.570

[3] Faradmal J, Soltanian AR, Roshanaei G, Khodabakhshi R, Kasaeian A. Comparison of the performance of log-logistic regression and artificial neural networks for predicting breast cancer relapse. Asian Pac J Cancer Prev. 2014;15(14):5883–8.

[4] Hoque MO, Feng Q, Toure P, Dem A, Critchlow CW, Hawes SE, et al. Detection of aberrant methylation of four genes in plasma DNA for the detection of breast cancer. J Clin Oncol. 2006 Sep 10;24(26):4262–9.

[5] Su X, Meneses K, McNees P, Johnson WO. Interaction trees: exploring the differential effects of an intervention program for breast cancer survivors [Internet]. Vol. 60, Journal of the Royal Statistical Society: Series C (Applied Statistics). 2011. p. 457–74. Available from: http://dx.doi.org/10.1111/j.1467-9876.2010.00754.x

[6] Kong AL, Pezzin LE, Nattinger AB. Identifying patterns of breast cancer care provided at high-volume hospitals: a classification and regression tree analysis. Breast Cancer Res Treat. 2015 Oct;153(3):689–98.

[7] Chao C-M, Yu Y-W, Cheng B-W, Kuo Y-L. Construction thofe model for the breast cancer survival analysis usesa support vector machine, logistic regression, and decision tree. J Med Syst. 2014 Oct;38(10):106.

[8] Ingoldsby H, Webber M, Wall D, Scarrott C, Newell J, Callagy G. Prediction of Oncotype DX and TAILORx risk categories using histopathological and immunohistochemical markers by classification and regression tree (CART) analysis. Breast. 2013 Oct;22(5):879–86.

[9] Fonarow GC, Adams KF Jr, Abraham WT, Yancy CW, Boscardin WJ, ADHERE Scientific Advisory Committee, Study Group, and Investigators. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. JAMA. 2005 Feb 2;293(5):572–80.

[10] Chou S-M, Lee T-S, Shao YE, Chen I-F. Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines [Internet]. Vol. 27, Expert Systems with Applications. 2004. p. 133–42. Available from: http://dx.doi.org/10.1016/j.eswa.2003.12.013

[11] Klauber-DeMore N. Breast Cancer in Young Women. IOS Press; 2006. 121 p.

[12] Anjia M, Zhizhong G. The influence of model mismatch to power system calculation, part II: on the stability calculation [Internet]. 2005 International Power Engineering Conference. 2005. Available from: http://dx.doi.org/10.1109/ipec.2005.207076

[13] Pawlovsky AP, Matsuhashi H. The use of a novel genetic algorithm in component selection for a CNN method for breast cancer prognosis [Internet]. 2017 Global Medical Engineering Physics Exchanges/Pan American Health Care Exchanges (GMEPE/PAHCE). 2017. Available from: http://dx.doi.org/10.1109/gmepe-pahce.2017.7972084

[14] Rajaguru H, Prabhakar SK. A Comprehensive Analysis on Breast Cancer Classification with Radial Basis Function and Gaussian Mixture Model [Internet]. IFMBE Proceedings. 2017. p. 21–7. Available from: http://dx.doi.org/10.1007/978-981-10-4220-1_5

[15] Hazra A, Kumar S, Gupta A. Study and Analysis of Breast Cancer Cell Detection using Naïve Bayes, SVM and Ensemble Algorithms [Internet]. Vol. 145, International Journal of Computer Applications. 2016. p. 39–45. Available from: http://dx.doi.org/10.5120/ijca2016910595

[16] Rong L, Yuan S. Diagnosis of Breast Tumor Using SVM-KNN Classifier [Internet]. 2010 Second WRI Global Congress on Intelligent Systems. 2010. Available from: http://dx.doi.org/10.1109/gcis.2010.278