

Analysis and Comparison for Innovative Prediction Technique of Breast Cancer Tumor by Linear Discriminant Analysis Algorithm over Support Vector Machine Algorithm with Improved Accuracy

Srinivasulureddy Ch, Neelam Sanjeev Kumar

Department of Biomedical Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode: 602105.

ABSTRACT

Aim: The objective of this study is to use machine learning algorithms to detect the presence of breast cancer tumors and compare accuracy, sensitivity, and precision between the Linear Discriminant Analysis (LDA) and Support vector machine algorithm (SVM). **Materials and Methods:** The research uses two sets of data from the Wisconsin Breast Cancer dataset, which is obtained from the UCI Machine Learning Repository. Linear Discriminant Analysis (N=20) and Support vector machine (N=20) with sample size in accordance to total sample size calculated using clincalc.com by keeping alpha error-threshold at 0.05, confidence interval at 95%, enrollment ratio as 0:1, and power at 80%. The accuracy, sensitivity, and precision are calculated using MATLAB software. **Results:** Comparison of accuracy (%), sensitivity (%), and precision (%) are done using SPSS software using independent sample t-test. Linear Discriminant Analysis algorithm results in an accuracy of 88.25% ($p < 0.001$), the sensitivity of 94.68% ($p < 0.001$), and precision of 84.35% ($p < 0.001$). Support vector machine algorithm results in an accuracy of 97.50%, sensitivity of 95.83%, and precision of 100%. **Conclusion:** Support vector machine algorithm performed significantly better than Linear Discriminant Analysis algorithm with improved accuracy of 97.50% for breast cancer prediction.

Keywords

Innovative Breast Cancer Prediction, Machine Learning Algorithms, Linear Discriminant Analysis, Support Vector Machine, Accuracy, Sensitivity.

Imprint

Srinivasulureddy Ch, Neelam Sanjeev Kumar. Analysis and Comparison for Innovative Prediction Technique of Breast

Cancer Tumor by Linear Discriminant Analysis Algorithm over Support Vector Machine Algorithm with Improved Accuracy. *Cardiometry*; Special issue No. 25; December 2022; p. 885-890; DOI: 10.18137/cardiometry.2022.25.885890; Available from: <http://www.cardiometry.net/issues/no25-december-2022/linear-discriminant-analysis-algorithm>

INTRODUCTION

Breast cancer is the most common cancer among women. It is the second leading cause of death in women, after lung cancer (Ferlay et al. 2015). Many machine learning algorithms are used to predict breast cancer with improved diagnostic accuracy (Karabatak and Cevdet Ince 2009). The major purpose of this Innovative breast cancer prediction technique is to compare and analyze the accuracy, sensitivity, and precision of the LDA and SVM algorithm. Supervised machine learning algorithms are the most prominent of the present cancer detection methodologies. Breast cancer can be discovered early and treated more effectively, potentially saving thousands of lives. In a range of medical applications, machine learning algorithms are being utilized to detect and categorize cancer (Mao et al. 2019) (Elmadhun and Finlayson 2009).

Approximately 286 ScienceDirect and 10 IEEE Xplore were found to be related to this work, which was completed in recent years and reported the developed algorithms and models. (Nematzadeh, Ibrahim, and Selamat 2015) performed a comparison study on a decision tree, NB, NN, and SVM with three different kernel functions over the Wisconsin breast cancer dataset, results showed NN achieved the highest accuracy of 98.09%. The relevance vector machine (RVM) was compared to distinct machine learning methods for diagnosing breast cancer by (Gayathri and Sumathi 2016) to decrease characteristics of the linear discriminant analysis used to classify data. The simulation findings yielded sensitivity (98%), specificity (94%), and precision (96%). Performance comparison is done by (Asri et al. 2016) over different classifiers DT, SVM, KNN, and NB using WEKA data mining tools. The approach results showed that SVM has performed better with an accuracy of 97.13%, and a precision of 98% with a low error rate. Research by (Bazazeh and Shubair 2016) examined three of the most prominent machine learning (ML) algorithms for breast cancer detection and diagnosis, namely SVM, RF, and Bayesian

Networks (BN). The Wisconsin breast cancer dataset was used as a training set to evaluate and compare the performance of the three machine learning classifiers in terms of key metrics like accuracy, recall, precision and area under the ROC curves. SVM had the best performance in terms of accuracy, specificity, and precision, according to the findings. RF, on the other hand, has the best chance of correctly identifying tumors. Our team has extensive knowledge and research experience that has translate into high quality publications (Chellapa et al. 2020; Lavanya et al. 2021; Raj Ret al. 2020; Shilpa-Jain et al. 2021; S et al. 2021; Ramadoss et al. 2022; Wuet al. 2020; Kalidoss et al. 2021; Kaja et al. 2020; Antink et al. 2020; Paulet al. 2020; Malaikolundhan et al. 2020)

The key difficulty that has inspired me to perform this research on enhancing the accuracy of machine learning and the prediction of breast cancer in early stages is inefficient early identification of breast cancer and human error in the detection of breast cancer by conventional approaches. The fundamental issue with current research is that the breast cancer prediction systems are inaccurate. The authors compared machine learning algorithms since they were expected in machine learning algorithms and deep learning technology. The aim of this Innovative breast cancer prediction technique is to compare and study breast cancer detection strategies using machine learning algorithms such as LDA and SVM with improved accuracy.

MATERIALS AND METHODS

The research was conducted at the Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, in the University simulation lab. The data used in this Innovative breast cancer technique is the Wisconsin breast cancer data set obtained from UCI Machine Learning Repository. This data set is used to classify the benign and malignant cells in a data set by using columnar properties to describe the cells. Support comes in the form of visualization and analysis. The measurements for the digital images of fine-needle aspirate of a breast mass were contributed by research from the University of Wisconsin. The information is separated into two categories. The sample size was calculated based on the findings of prior studies (Li 2018) (Bharat, Pooja, and Anishka Reddy 2018) by clincalc.com by keeping alpha error-threshold value 0.05, confidence interval at 95%, enrollment ratio as 0:1, and power at 80%. For

the data acquired from the Wisconsin breast cancer data collection, sample preparation is done in this research work of Innovative breast cancer prediction.

The total sample size is 40, and group 1 is the LDA with N value 20, and group 2 is the SVM with N value 20. A sample dataset for both the LDA and the SVM is exported to a Microsoft Excel document and then imported into Matlab as an input. For training the source dataset, the Matlab 2021 software must be installed on the PC. The imported data is trained using classification learner tools in MATLAB. k-fold cross-validation is performed on the data to improve the machine learning model performance TP, TN, FP, and FN values are recorded in a confusion matrix. The confusion matrix is used to calculate the accuracy (%) sensitivity (%) and precision (%) values.

Statistical analysis

IBM SPSS V26.0 was utilized for statistical analysis in this study. This software compares the accuracy, sensitivity, and precision of LDA with the SVM algorithm. Because the variables are independent of one another, an independent sample T-test was used to compare the two groups performance. Radius mean, texture mean, perimeter mean, area mean, smoothness mean, concavity mean, concave points mean, symmetry mean are the characteristics used in the innovative breast cancer prediction (Maglogiannis, Zafiroopoulos, and Anagnostopoulos 2009). There are no dependent variables in this study.

RESULTS

The results of this Innovative breast cancer prediction comparing the accuracy, sensitivity, and precision of the LDA and SVM algorithms for breast cancer prediction on the Wisconsin breast cancer dataset show that the SVM outperforms the LDA algorithm with improved accuracy. Table 1a and Table 1b show the accuracy, sensitivity, and precision values of LDA and SVM respectively. SVM has an accuracy (97.50%), sensitivity (95.83%), and precision (100%), whereas LDA results have an accuracy (88.25%), sensitivity (94.68%), and precision (84.35%). Table 2 demonstrates that the SVM shows less error and no standard deviation with improved accuracy when compared to the LDA. Table 3 shows the results of the Independent sample T-test which shows the difference in accuracy ($P < 0.001$), sensitivity ($P < 0.001$), and precision ($P < 0.001$) are statistically significant.

Table 1a

Breast cancer prediction using LDA Algorithm.

Sample	Accuracy (%)	Sensitivity (%)	Precision (%)
1	0.875	0.95	0.826087
2	0.875	0.95	0.826087
3	0.9	0.952381	0.869565
4	0.85	0.869565	0.869565
5	0.85	0.947368	0.782609
6	0.875	0.95	0.826087
7	0.9	0.952381	0.869565
8	0.875	0.95	0.869565
9	0.9	0.952381	0.826087
10	0.875	0.95	0.913043
11	0.875	0.954545	0.826087
12	0.975	0.95	0.826087
13	0.85	0.95	0.913043
14	0.875	0.954545	0.913043
15	0.9	0.954545	0.826087
16	0.875	0.95	0.869565
17	0.975	0.952381	0.826087
18	0.875	0.95	0.913043
19	0.875	0.95	0.826087
20	0.9	0.952381	0.869565

Table 1b

Breast cancer prediction using SVM Algorithm.

Sample	Accuracy (%)	Sensitivity (%)	Precision (%)
1	0.975	0.958333	1
2	0.975	0.958333	1
3	0.975	0.958333	1
4	0.975	0.958333	1
5	0.975	0.958333	1
6	0.975	0.958333	1
7	0.975	0.958333	1
8	0.975	0.958333	1
9	0.975	0.958333	1
10	0.975	0.958333	1
11	0.975	0.958333	1
12	0.975	0.958333	1
13	0.975	0.958333	1
14	0.975	0.958333	1
15	0.975	0.958333	1
16	0.975	0.958333	1
17	0.975	0.958333	1
18	0.975	0.958333	1
19	0.975	0.958333	1
20	0.975	0.958333	1

Table 2

Comparison of mean accuracy, mean sensitivity and mean precision of LDA and SVM.

Parameters	Group	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	Linear Discriminant Analysis	20	88.25	.02161	.00483
	Support Vector Machine	20	97.50	.00000	.0000
Sensitivity	Linear Discriminant Analysis	20	94.68	.01827	.00409
	Support Vector Machine	20	95.83	.00000	.00000
Precision	Linear Discriminant Analysis	20	84.35	.03569	.00798
	Support Vector Machine	20	100	.00000	.00000

Figure 1 illustrates the bar chart comparing the LDA and SVM algorithms accuracy, specificity, and precision. With the accuracy ranging from (88.25-97.50%), sensitivity ranging from (94.68-95.83%) and precision ranging from (84.35-100%) both procedures appear to give the same variable results.

Figures 2a and 2b represent the confusion matrix of the LDA and SVM respectively, where TP, FP, TN, and FN values are utilized to calculate accuracy, sensitivity, and precision.

DISCUSSION

In this study of Innovative breast cancer prediction techniques, the SVM algorithm have the higher accuracy (97.50%), sensitivity (95.83%), and precision (100%) when compared to LDA accuracy (88.25%), sensitivity (94.68%) and precision (84.35%) as shown in Table 2. In terms of accuracy, sensitivity and precision there appears to be a statistically significant difference.

SVM provides high accuracy and consistent diagnosis in predicting breast cancer using the Wisconsin breast cancer dataset (Oyewola et al. 2017) performed research on five machine learning algorithms in which LDA has less accuracy of 93.5% when compared to SVM accuracy of 95.8% in predicting breast cancer. Many researchers developed algorithms and models using machine learning to predict breast cancer disease (Pawar, Sharma, and Sapate 2021). Some of the works include integrating the machine learning techniques with feature selection and feature extraction to identify better models (Omondigbe, Veeramani, and Sidhu 2019) on the Wisconsin diagnostic breast

Table 3

Independent sample T-test in predicting the accuracy, sensitivity, and precision of breast cancer prediction using LDA and SVM. There appears to be a statistically significant difference ($p < 0.001$) in both methods.

Parameter	Equal Variances	Levene's Test for Equality of Variances		T-test for Equality of Means					
		F	Sig	t	df	Significance (one-Sided p)	Mean Difference	Std. Error Difference	95% Confidence interval (Upper)
Accuracy	Assumed	38.67	<.001	-19.14	38	<.001	-.09250	.00483	-.08272
	Not assumed			-19.14	19.00	<.001	-.09250	.00483	-.08238
Sensitivity	Assumed	4.396	.043	-2.831	38	.004	-.01157	.00409	-.00330
	Not assumed			-2.831	19.00	.005	-.01157	.00409	-.00302
Precision	Assumed	49.46	<.001	-19.61	38	<.001	-.15652	.00798	-.14037
	Not assumed			-19.61	19.00	<.001	-.15652	.00798	-.13982

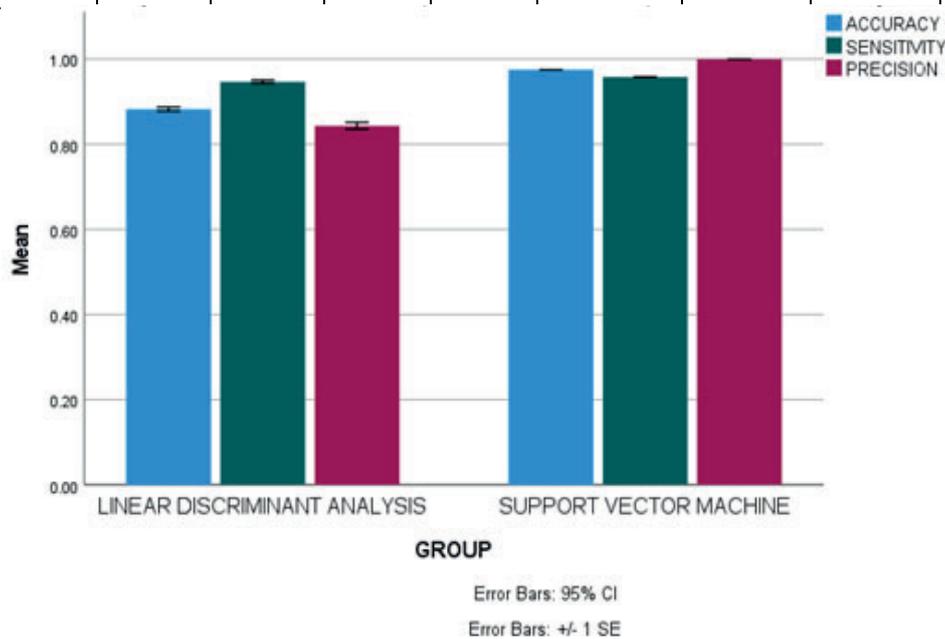


Fig. 1. Bar chart representing the comparison of mean accuracy, mean sensitivity, and mean precision of breast cancer prediction with LDA and SVM algorithms. Both techniques appear to produce the same variable results with accuracy ranging from (88.25-97.50%), sensitivity ranging from (94.68-95.83%) and precision ranging from (84.35-100%). X-axis: LDA vs SVM, Y-axis: Mean accuracy, mean sensitivity, and mean precision of detection ± 1 SD.

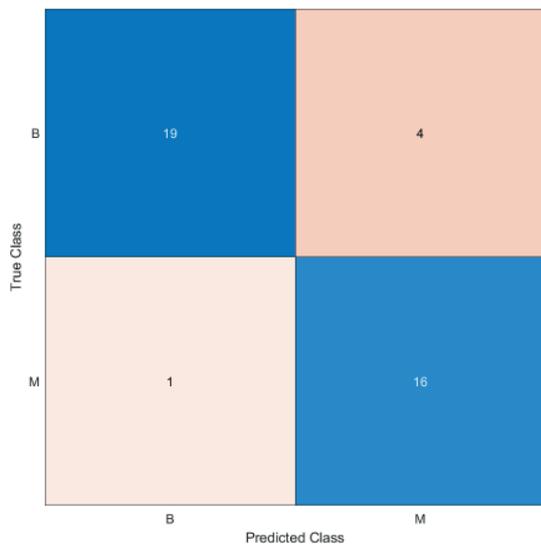


Fig. 2a. Confusion matrix of LDA for $k = 10$. TP accounts for 19, FP accounts for 4, FN accounts for 1 and TN accounts for 16. The total accuracy was found to be 88.25%.

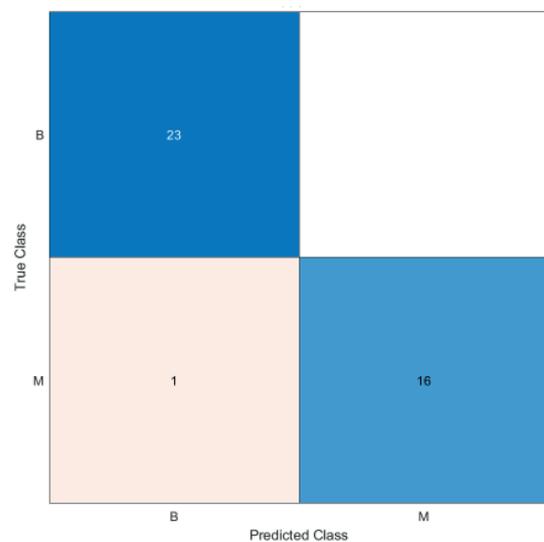


Fig. 2b. Confusion matrix of SVM for $k = 10$. TP accounts for 23, FP accounts for 0, FN accounts for 1 and TN accounts for 16. The total accuracy was found to be 97.25%.

cancer dataset, proposed hybrid model was developed by lowering the high dimensionality features of the dataset using LDA and SVM, yielding an accuracy of 98.82%, a sensitivity of 98.41% and specificity 99.07%. Another model proposed by (Rajaguru and Prabhakar 2017) obtained an average classification accuracy of 83.45% when a Bayesian Linear discriminant classifier is used. Experimental results showed that the LDA has achieved maximum sensitivity of 96% when compared to SVM sensitivity of 85% (Fusco et al. 2016).

The accuracy of the machine learning methods LDA and SVM is limited in this study due to the short sample size. Further increase in the sample may yield better accuracy of the algorithms in predicting breast cancer. The factors affecting the study are unhealthy lifestyle, family hereditary, and gene mutations. Combining the large dataset of real-time applications with other machine learning and deep learning methods could lead to enhanced performance in the future. The most up-to-date deep learning approaches can be integrated to produce a model capable of more accurately detecting and diagnosing breast cancer. Furthermore, Innovative breast cancer prediction techniques can be used to identify various stages of breast cancer in the near future.

CONCLUSION

When compared to LDA the SVM algorithm that runs in MATLAB proved to offer high results with improved accuracy of 97.50% in predicting breast cancer tumors. Furthermore, unlike other methods, the algorithm's performance improved as the amount of data rose. This model is highly efficient and has a lot of potential in terms of improving the diagnostic efficiency of breast cancer, thus it can be used in hospitals and testing facilities.

DECLARATION

Conflicts of Interest

No conflict of interest in the manuscript

Authors Contributions

Data collection, data analysis, and manuscript writing were all done by author CHSR. NSK was involved in manuscript computerization, data validation, and critical evaluation.

Acknowledgments

We would like to thank Saveetha School of Engineering, Saveetha Institute of Medical and Technical

Sciences (Formerly known as Saveetha University) management for providing me with the necessary infrastructure to complete the project.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

1. Venus Electronics Tamilnadu.
2. Saveetha University.
3. Saveetha Institute of Medical And Technical Sciences.
4. Saveetha School of Engineering.

REFERENCES

1. Asri, Hiba, Hajar Mousannif, Hassan Al Moatassime, and Thomas Noel. 2016. "Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis." *Procedia Computer Science*. <https://doi.org/10.1016/j.procs.2016.04.224>.
2. Bazazeh, Dana, and Raed Shubair. 2016. "Comparative Study of Machine Learning Algorithms for Breast Cancer Detection and Diagnosis." 2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA). <https://doi.org/10.1109/icedsa.2016.7818560>.
3. Bharat, Anusha, N. Pooja, and R. Anishka Reddy. 2018. "Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis." 2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C). <https://doi.org/10.1109/cimca.2018.8739696>.
4. Elmadhun, N. Y., and C. A. Finlayson. 2009. "The Value of Mammography vs. Physical Examination in Detecting a Recurrent Breast Cancer in Women with a Previous Diagnosis of Breast Cancer." *Poster Session Abstracts*. <https://doi.org/10.1158/0008-5472.sabcs-5014>.
5. Ferlay, Jacques, Isabelle Soerjomataram, Rajesh Dikshit, Sultan Eser, Colin Mathers, Marise Rebelo, Donald Maxwell Parkin, David Forman, and Freddie Bray. 2015. "Cancer Incidence and Mortality Worldwide: Sources, Methods and Major Patterns in GLOBOCAN 2012." *International Journal of Cancer. Journal International Du Cancer* 136 (5): E359–86.
6. Fusco, Roberta, Mario Sansone, Salvatore Filice, Guglielmo Carone, Daniela Maria Amato, Carlo Sansone, and Antonella Petrillo. 2016. "Pattern Recognition Approaches for Breast Cancer DCE-MRI Clas-

- sification: A Systematic Review.” *Journal of Medical and Biological Engineering*. <https://doi.org/10.1007/s40846-016-0163-7>.
7. Gayathri, B. M., and C. P. Sumathi. 2016. “Comparative Study of Relevance Vector Machine with Various Machine Learning Techniques Used for Detecting Breast Cancer.” 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICIC). <https://doi.org/10.1109/icic.2016.7919576>.
 8. Karabatak, Murat, and M. Cevdet Ince. 2009. “An Expert System for Detection of Breast Cancer Based on Association Rules and Neural Network.” *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2008.02.064>.
 9. Li, Yixuan. 2018. “Performance Evaluation of Machine Learning Methods for Breast Cancer Prediction.” *Applied and Computational Mathematics*. <https://doi.org/10.11648/j.acm.20180704.15>.
 10. Maglogiannis, Ilias, Elias Zafiroopoulos, and Ioannis Anagnostopoulos. 2009. “An Intelligent System for Automated Breast Cancer Diagnosis and Prognosis Using SVM Based Classifiers.” *Applied Intelligence*. <https://doi.org/10.1007/s10489-007-0073-z>.
 11. Mao, Ning, Ping Yin, Qinglin Wang, Meijie Liu, Jianjun Dong, Xuexi Zhang, Haizhu Xie, and Nan Hong. 2019. “Added Value of Radiomics on Mammography for Breast Cancer Diagnosis: A Feasibility Study.” *Journal of the American College of Radiology: JACR* 16 (4 Pt A): 485–91.
 12. Nematzadeh, Zahra, Roliana Ibrahim, and Ali Selamat. 2015. “Comparative Studies on Breast Cancer Classifications with K-Fold Cross Validations Using Machine Learning Techniques.” 2015 10th Asian Control Conference (ASCC). <https://doi.org/10.1109/ascc.2015.7244654>.
 13. Omondiagbe, David A., Shanmugam Veeramani, and Amandeep S. Sidhu. 2019. “Machine Learning Classification Techniques for Breast Cancer Diagnosis.” *IOP Conference Series: Materials Science and Engineering*. <https://doi.org/10.1088/1757-899x/495/1/012033>.
 14. Oyewola, David, Danladi Hakimi, Kayode Adeboye, and Musa Danjuma Shehu. 2017. “Using Five Machine Learning for Breast Cancer Biopsy Predictions Based on Mammographic Diagnosis.” *International Journal of Engineering Technologies IJET*. <https://doi.org/10.19072/ijet.280563>.
 15. Pawar, Shivaji D., Kamal Kr Sharma, and Suhas G. Sapate. 2021. “Advances in Machine Learning and Deep Learning Approaches for Mammographic Breast Density Measurement for Breast Cancer Risk Prediction: An Overview.” *Design of Intelligent Applications Using Machine Learning and Deep Learning Techniques*. <https://doi.org/10.1201/9781003133681-8>.
 16. Rajaguru, Harikumar, and Sunil Kumar Prabhakar. 2017. “Bayesian Linear Discriminant Analysis for Breast Cancer Classification.” 2017 2nd International Conference on Communication and Electronics Systems (ICCES). <https://doi.org/10.1109/cesys.2017.8321279>.