DOI: https://doi.org/10.52756/ijerr.2024.v42.005 *Int. J. Exp. Res. Rev.*, Vol. 42: 50-59 (2024)

Original Article | *Peer Reviewed* | *Open Access*

International Journal of Experimental Research and Review (IJERR)

ISSN: 2455-4855 (Online)
www.iaph.in
ISSN 2455-4855

Correlation of Prognostic Factors of Invasive Lobular Carcinoma and Invasive Ductal Carcinoma

Apurva Vashist*, Anil Kumar Sagar and Anjali Goyal

Department of Computer Science and Engineering, Sharda University, Greater Noida, U.P., India

E-mail/Orcid Id:

AV, apurva.vashist@gmail.com, https://orcid.org/0009-0004-3941-0709; *AS,* anil.sagar@sharda.ac.in, https://orcid.org/0000-0002-5991-2835; *AG,* anjali1@sharda.ac.in, https://orcid.org/0000-0003-3003-9275

Article History:

Received: 10th Oct., 2023
Accepted: 29th Jul., 2024
Published: 30th Aug., 2024

Keywords:

Breast cancer prediction, biomarkers, gaussian method, gene, invasive lobular carcinoma, invasive ductal carcinoma, multi-omics dataset, QLattice

How to cite this Article:

Apurva Vashist, Anil Kumar Sagar and Anjali Goyal (2024). Correlation of Prognostic Factors of Invasive Lobular Carcinoma and Invasive Ductal Carcinoma. *International Journal of Experimental Research and Review*, *42*, 50-59.

DOI:
https://doi.org/10.52756/ijerr.2024.v42.005

Abstract: One of the biggest risks facing women in the twenty-first century is breast cancer. Invasion lobular carcinoma and invasion ductal carcinoma are the two main categories into which it is divided. Omics data is used to identify predictive biomarker signatures for clinical applications to detect breast cancer. Predictive performance has significantly improved because of recent advancements in machine learning techniques. Here, we are using an approach built on symbolic regression called the QLattice on a variety of clinical omics data sets. Through the identification of potential regulatory interactions between biomolecules, this method creates efficient, high-performing models that can forecast and explain the results of a specific omics experiment. The models have the potential to make it easier to find new biomarker signatures due to their clarity and obvious functional shape, which make them simple and easy to comprehend. A comprehensive experimental investigation was conducted to assess the machine learning model's efficacy in terms of the Area under the Curve (AUC) for breast cancer. The outcomes, which were contrasted with other approaches, demonstrate the suggested framework's efficacy and capacity to beat the alternative algorithms in terms of AUC, which is 0.66. Here, we profiled breast tumors in detail, including ductal carcinoma, mixed carcinoma, and invasive lobular carcinoma, by using the Gaussian method and TNXB gene.

Introduction

Breast cancer (BC) is one type of cancer that begins in the breast cells (Rami et al., 2023; Yadav et al., 2024). It is one of the most frequent cancers that strike women, although being much less common in men. Examining breast cancer's forms, risk factors, symptoms, signs and treatment choices is necessary to comprehend the disease. Breast cancer is a complicated illness with many subtypes and contributing variables. Patients benefit most from early identification and a comprehensive approach to treatment. Research advancements continue to enhance knowledge, diagnosis, and treatment, improving the prognosis and quality of life for patients with breast cancer.

With more than 280000 diagnoses and 40000 predicted deaths from invasive breast cancer in the US in 2021, it is the most prevalent cancer among women. For women 20 to 59 years old, it is still the top cause of death and mortality decreases have regularly plateaued across all age categories.

The amount of diagnoses for non-invasive ductal carcinoma in situ (DCIS) has increased because of improvements in mammography screening. A second breast cancer (SBC) can occur in up to 40% of women after DCIS, 28% of which are invasive breast cancers (Siegel, 2021; Sagar, 2020). Despite the largely positive outcomes of DCIS. Choosing the best therapeutic and clinical follow-up methods for DCIS is still a hot topic of debate. It requires thought to prevent overtreating women with low-risk diseases and undertreating those at high risk of developing an invasive SBC (Tseng, 2019). It's essential to determine which women are most prone to develop a second invasive SBC to individualize care and therapy as much as feasible for each patient, as shown in figure 1.

***Corresponding Author:** apurva.vashist@gmail.com

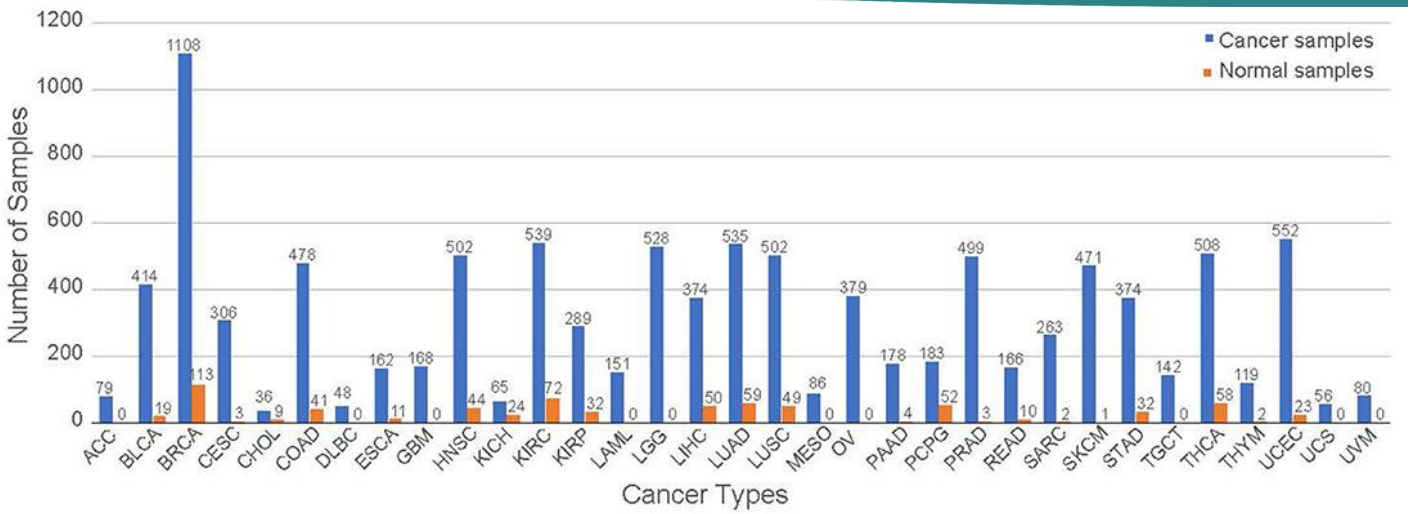


Figure 1 Types of cancer (Ramirer, 2020).

Types of breast cancer

Based on where it starts, breast cancer can be roughly divided into two categories:

Ductal carcinoma: The ducts that deliver milk to the nipple are the site of initiation for ductal carcinoma. Whereas invasive ductal carcinoma (IDC) has expanded outside of the duct walls, ductal carcinoma in situ (DCIS) is non-invasive.

Lobular Carcinoma develops in the glands that produce milk (lobules). While invasive lobular carcinoma (ILC) has the potential to spread to other areas of the breast and beyond, lobular carcinoma in situ (LCIS) is a sign of an elevated risk of breast cancer (Vashist et al., 2023; Sagar et al., 2021).

The majority of ILC genomic investigations to date have concentrated on mRNA expression and DNA copy-number analysis, offering little insight into the underlying biology of this disease. Four hundred sixty-six breast tumors from six distinct expertise platforms were analyzed for the inaugural TCGA breast cancer study published in Cancer Genome Atlas in 2012. There were only 36 samples from ILC, and there were no lobular-specific characteristics other than CDH1 mutations and decreased mRNA and protein expression (Rezaeijo et al., 2023). We examined 817 breast tumors from the TCGA, including 127 ILC, which is almost twice as many as we typically do. This study found numerous genetic changes that distinguish between ILC and IDC, proving at the molecular level that ILC is a unique breast cancer subtype and offering fresh information on the biology of ILC tumors and treatment options (Singh et al., 2024; Rezaeijo et al., 2023).

Tenascin-X is a protein that can be produced using instructions from the TNXB gene. The connective tissues, which support the body's muscles, joints, organs, and skin, are organized and partly maintained by this protein (Kanehisa, 2016; Tonmoy, 2021; Sandhu, 2022). A family of proteins known as collagens supports and supports connective tissues across the body. Additionally, tenascin-X controls the stability and structure of elastic fibers, giving connective tissues flexibility and stretchiness (elasticity). Tenascin-X is a protein that regulates the extracellular matrix (ECM), cellular adhesion, and tissue structure. Its potential correlation between genetic variants or expression levels and the risk or advancement of breast cancer is what makes it useful in predicting the disease. A vital part of the extracellular matrix (ECM) is the glycoprotein family of tenascins, which includes tenascin-X. It contributes to tissue healing, structural integrity, and cell signaling. Researchers could find possible associations with cancer formation, aggressiveness, or responsiveness to treatment by looking at Tenascin-X levels or gene variations in breast cancer patients. Current study focuses on Tenascin-X's potential as a biomarker for cancer risk, prognosis, and therapy responsiveness to determine its function in predicting breast cancer.

The QLattice: A new machine learning model

QLattice is a supervised machine-learning tool for symbolic regression. The QLattice graph is neither a neural network nor a model based on decision trees. The QLattice is like a decision tree in that it explains ability and interpretability by dissecting the black box neural network.

QLattice graph:

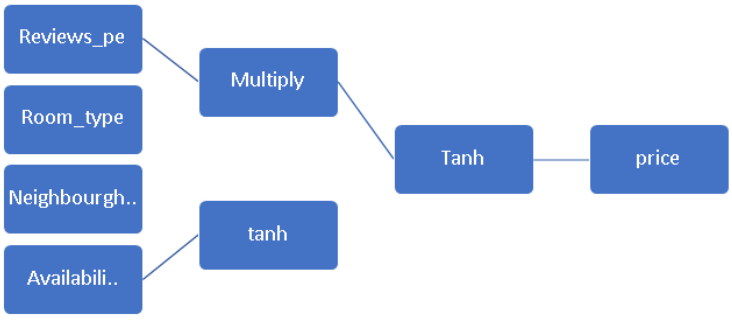


Figure 2. QLattice graph.

Thousands of possible models are found by QLattice, which then looks for the one graph that has the ideal characteristics and interaction combinations to provide the precisely adjusted model for our issue. When combined, the multiply, linear, sine, tanh and Gaussian data transformations almost completely cover all naturally occurring dependencies shown in figure 2.

DOI: https://doi.org/10.52756/ijerr.2024.v42.005

These are the data transformations available in the QLattice.

Organization

The first section introduces breast cancer and briefly explains its various types. The second section is a literature review, while the third section describes the multi-omics dataset. In the fourth section, we outline the methodology, where we use Gaussian methods to calculate accuracy, comparing results from single and multiple iterations. Additionally, we discuss the impact of the TNXB gene mutation in breast cancer, comparing results with and without this mutation. The fifth section focuses on the results and their analysis.

Literature Review

Taghizadeh, 2022, 762 BC patients and 138 solid tissue normal participants were used to investigate relevant BC characteristics. Three categories of machine learning algorithms were used:

1. Feature selection techniques are used, and the most valuable feature is chosen by comparing them.
2. A feature extraction approach, Principal Component Analysis (PCA).
3. We used 13 classification algorithms along with automated ML hyper-parameter adjustment.

Singh et al.2024, examined the relationships between proteins, copy number variations, mutations, and RNA expression in their 2024 study on breast cancer prediction using multi-omics datasets. A heatmap that displayed the correlation patterns throughout the multi-omics dataset was used to visualize the relationships between these various data types.

Rezaeijo (2023) assesses how well six machine learning models predict brain metastases in lung cancer by utilizing EGFR analysis and PET/CT radiomics. In 2020, the Cancer Hospital Affiliated with Shandong First Medical University diagnosed 204 patients with lung adenocarcinoma. The researchers retrospectively analyzed these patients. Before starting any medication, these individuals had EGFR gene testing and PET/CT imaging.

According to several recent research, the performance of classifiers can be improved by removing noise and unimportant data during data preparation using a feature selection strategy, such as the GA (Nouira, 2020). The comparatively high accuracy of some machine learning regression approaches was also highlighted as a result.

A diverse array of feature selection models has been employed for cancer classification and predicting clinical outcomes, primarily leveraging mRNA gene expression data. Hybrid bioinspired algorithms have emerged as a valuable approach for identifying a subset of pertinent genes relevant to cancer prediction. For instance, Coleto-Alcudia and Vegas-Rodrigues,2020 have introduced a hybridization of teaching models and the artificial bee colony (ABC) algorithm. In this approach, the initial step involves reducing the dimensionality of the feature space through a ranking method, followed by the ABC algorithm selecting the most significant gene subset.

Masoudi-Sobhanzadeh et al. (2021) The authors provide a technique to deal with the difficulty of feature selection in biological data analysis by fusing evolutionary algorithms and algorithms from globally recognized competitions. In many bioinformatics and biomedical applications, feature selection is a crucial stage since it aids in the identification of pertinent genes or features that may be utilized for tasks like illness classification or clinical outcome prediction.

The combination of two forms of molecular data, RNA-Seq and Reverse Phase Protein Array (RPPA), is explored by Isik and Ercan (2017) for the prediction of cancer patients' survival times. The creation of a prediction model using data from RNA-Seq, which provides gene expression data, and RPPA, which provides protein expression data, appears to be the main goal of this study. The accuracy of survival time projections for cancer patients may be improved by integrating these two forms of molecular data since it enables a more thorough knowledge of the molecular mechanisms causing the illness.

Lænkholm et al. (2020) the Prosigna-PAM50 assay's prognostic value in postmenopausal women with estrogen receptor-positive (ER+) and HER2-negative (HER2-) invasive lobular or ductal breast cancer is the subject of this study, which is most likely a population-based analysis. A genetic test called the Prosigna-PAM50 assay assists patients with breast cancer in determining their risk of recurrence. The author gave useful information for determining risk and preparing a treatment strategy for postmenopausal patients with ER+ and HER2-positive breast cancer.

The Dataset

705 breast tumor samples (611 patients survived, 94 patients died)

Four Data Types (n features):

- Copy Number Variations (860)
- Somatic Mutations (249)
- Gene Expression (604)
- Protein Expression (223)

Total: 1936 features

mu: Somatic mutation (yes, no) [somatic mutation – An alteration in DNA that occurs after conception. Somatic mutations can occur in any of the cells of the body except germs cells (sperm and egg) and, therefore, are not passed to children] (Zenbout et al., 2022; Ghosh, 2009; Biswas, 2020).

cn: Copy number variation as calculated by gistic (-2,-1,0,1,2)

rs: RNA (Ribonucleic acid) sequencing i.e., gene expression

pp: phosphor-protein levels

Methodology

Collect the dataset, and after the data preprocessing choose the model, here we have chosen Gaussian model. Utilizing the training dataset, train both models. Set the

DOI: https://doi.org/10.52756/ijerr.2024.v42.005

number of iterations for the first model to 1 (a single iteration), and for the second model, set it to 200.

Analyze both models' performance using the testing/validation dataset. Evaluate by metrics include accuracy, precision, recall and ROC-AUC. To prepare a ROC curve, confusion matrix and partial plots for the data. Keep track of the performance metrics for both models over the course of one iteration and 200 iterations and compare them. Then, find the associations between gene expression levels and survival outcomes in individuals with and without TNXB mutations.

A machine learning technique called a confusion matrix (CM) is used to evaluate a model's performance. The CM aids in the computation of numerous important metrics that assess a model's efficacy. Among these metrics are:

- Accuracy: The ratio of correct predictions (true positives and negatives) to the total number of predictions.
- Precision: The percentage of actual positive predictions among all the model's positive predictions. It's a metric for positive prediction accuracy.
- Recall: The percentage of real positives that the model properly detected is known as recall (sensitivity). It shows how well the model can extract pertinent information.
- Specificity: The percentage of real negatives that the model accurately detected. It shows how well the model can prevent false alerts.
- F1 Score: The harmonic mean of recall and precision is the F1 score. When striking a balance between recall and precision is crucial, this statistic can be helpful.
- The Receiver Operating Characteristics (ROC) curve illustrates the true positive rate (recall) in relation to the false positive rate. It is a useful tool for assessing the diagnostic performance of the model at different thresholds.
- Area Under the ROC Curve (AUC): The area under the ROC curve is expressed as a numerical value. It shows how well a model can distinguish between classes and ranges from 0 to 1.

These metrics thoroughly understand a model's performance and are frequently employed in machine learning research to assess and contrast other models or methodologies.

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \quad \text{(i)}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad \text{(ii)}$$

$$Recall\ or\ Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad \text{(iii)}$$

$$Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive} \quad \text{(iv)}$$

$$F1\ Score = \frac{2\ X\ (Precision\ x\ Recall)}{Precision + Recall} \quad \text{(v)}$$

Result and discussion

A model train for single iteration

Inputs
rs_APOB
rs_HBB

Training Metrics		Test
Accuracy	0.89	0.871
AUC	0.633	0.641
Precision	0.739	0.571
Recall	0.274	0.125

Figure 3. Constrain the model to have 3 edges (e.g., 2 features and one interaction).

In terms of accuracy, our model appears to perform well, but there is some potential for improvement in terms of AUC and recall, particularly if correctly recognizing positive cases is essential for our application, as shown in figure 3.

It's important to consider the context of our problem and the potential consequences of false positives and false negatives. We may adjust the model's threshold depending on the application to optimize precision or recall.

Further analysis, such as a confusion matrix, can provide more insights into the model's performance, including the distribution of true positives, true negatives, false positives, and false negatives. By using the ROC curve, we can calculate the AUC. At the training time, the AUC is 0.66 but at the testing time AUC is 0.64. This is less than the training time shown in figure 4.

Partial Plots for single iteration

In machine learning, partial plots, also known as partial dependence plots (PDPs), are a visualization approach used to comprehend the relationship between a particular feature and the anticipated outcome of a model while maintaining the constant values of other features shown in figure 5. They are very beneficial for deciphering complicated models, such as ensemble approaches.

DOI: https://doi.org/10.52756/ijerr.2024.v42.005

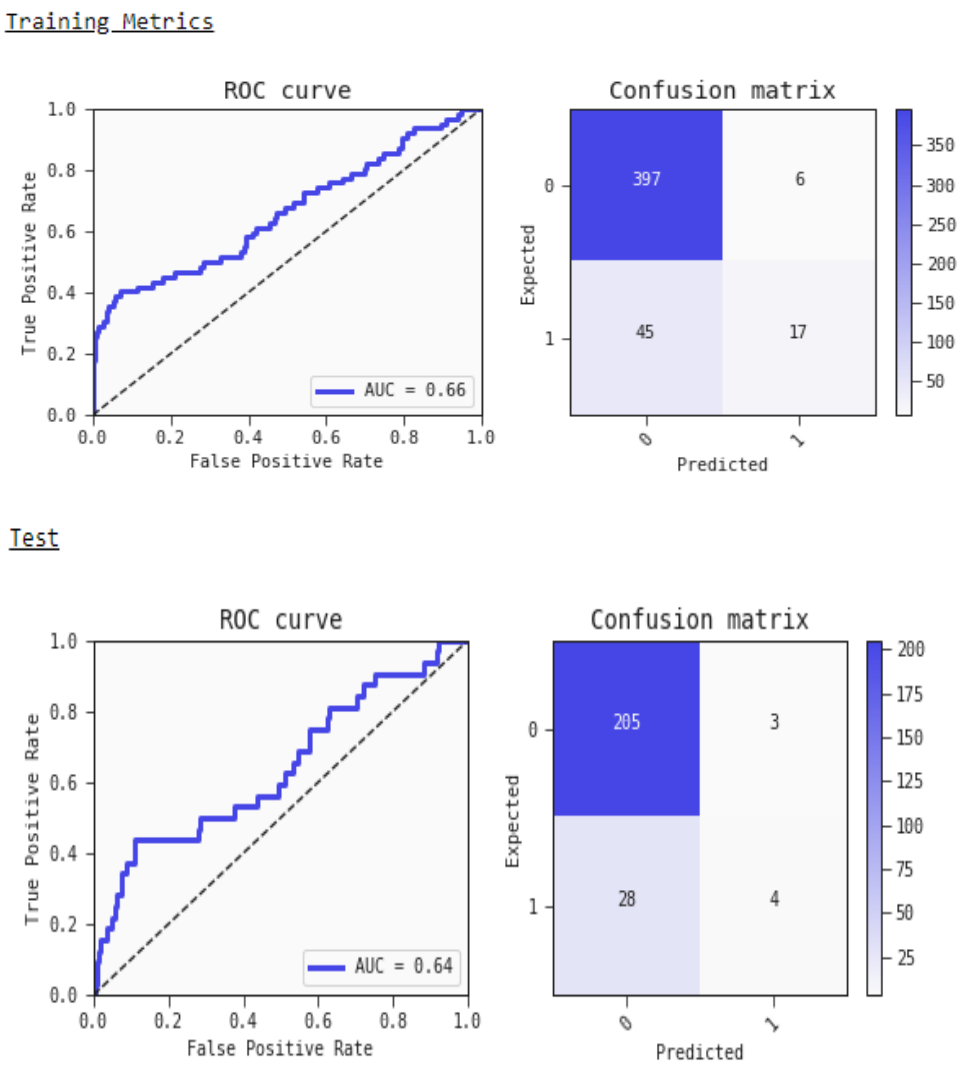


Figure 4. ROC curve for single iteration.

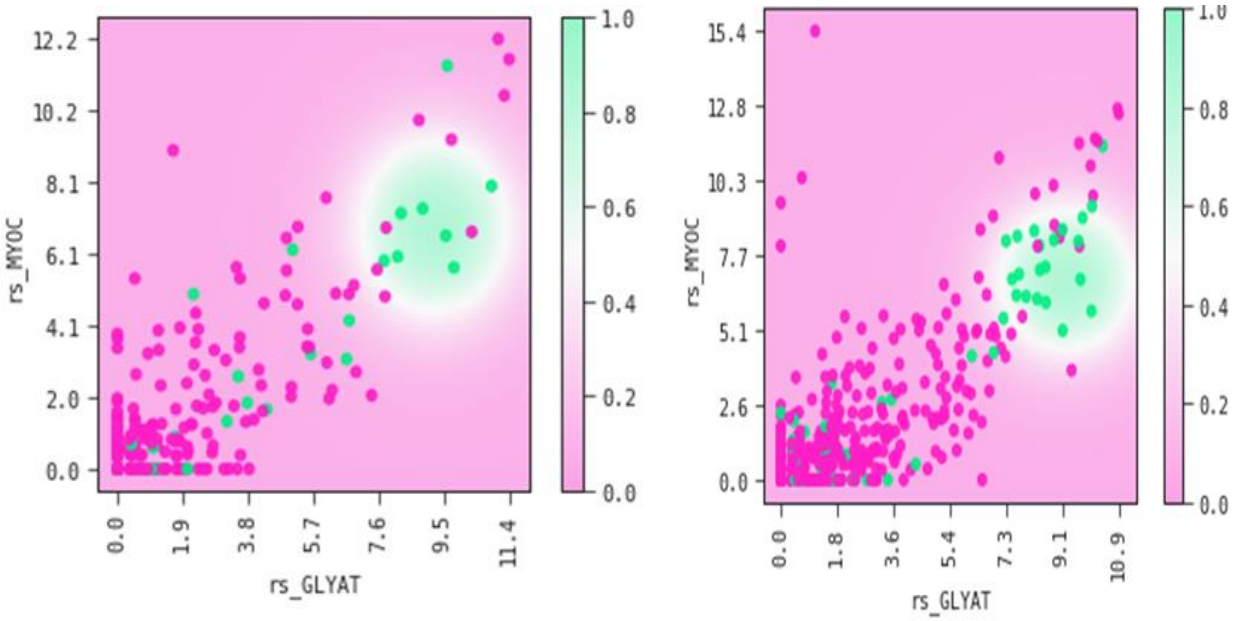


Figure 5. Partial Plots for one iteration.

DOI: https://doi.org/10.52756/ijerr.2024.v42.005

The model generalizes well on unseen data. It is a sign that a machine learning model has learned the underlying patterns in the training data and can make precise predictions or classifications on fresh, previously unexplored cases when a model generalizes successfully on unexplored data. Because machine learning aims to create models that can perform well in real-world scenarios where the data is not restricted to the training set, generalization is a fundamental goal in this field.

Looking at the ROC curve

In the prediction of breast cancer, situations where there is class imbalance or when we wish to examine the trade-off between sensitivity and specificity, the Receiver Operating Characteristic (ROC) curve is a graphical tool used to assess the performance of binary classification algorithms.

Sensitivity (True Positive Rate): On the y-axis of the ROC curve is a representation of the model's sensitivity (True Positive Rate). Sensitivity quantifies the share of true positive predictions (positive cases that were successfully detected) among all real positive cases. Sensitivity increases as the ROC curve is moved upward.

Specificity (True Negative Rate): On the x-axis, the ROC curve also shows data on specificity (True Negative Rate). Out of all negative situations, specificity measures the percentage of true negative predictions (negative cases correctly detected). Specificity rises as you move right along the ROC curve shown in figure 6, 7 and 8 and table 1 shows the accuracy of different algorithms by using a single iteration.

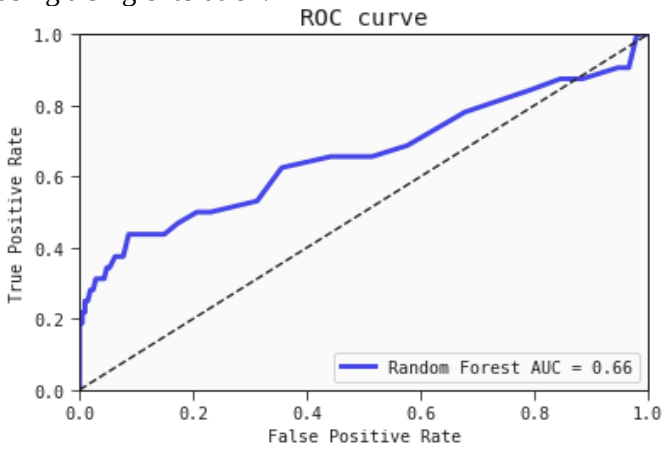


Figure 6. ROC curve by random forest.

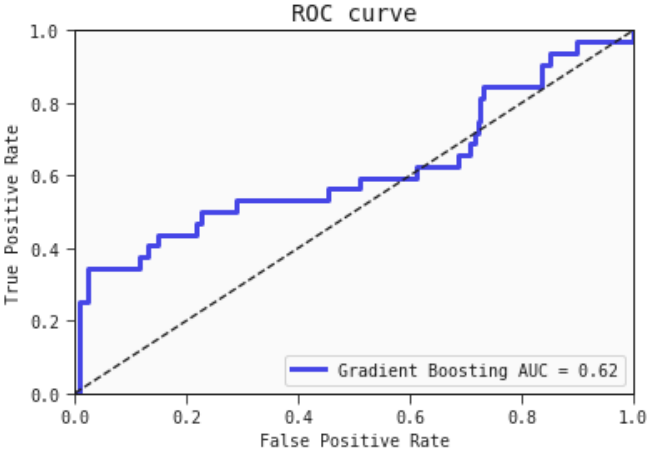


Figure 7. ROC curve by gradient boosting.

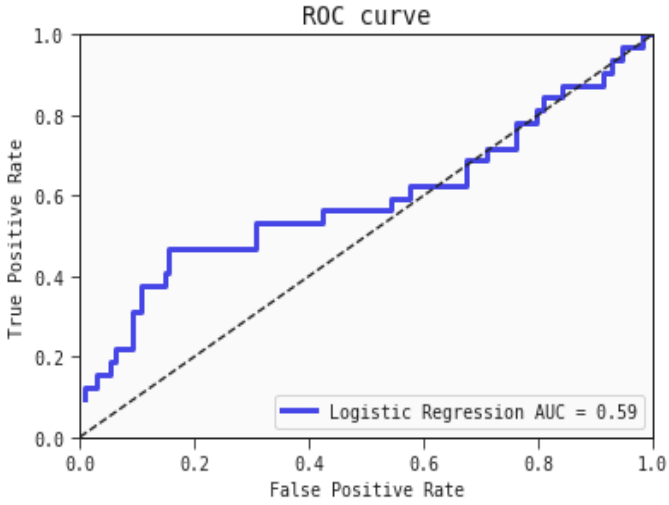
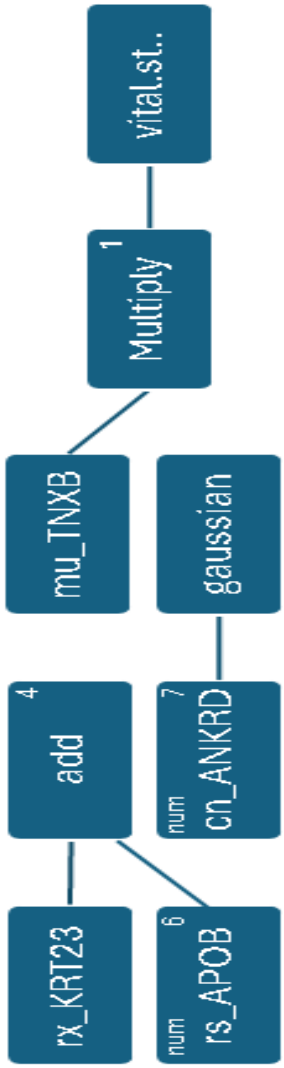


Figure 8. ROC curve by logistic Regression.

Table 1. Showing the accuracy of different algorithms by using a single iteration.

Algorithms	AUC
Random Forest	0.66
Gradient boosting	0.62
Logistic Regression	0.59

A model train for 200 iterations



DOI: https://doi.org/10.52756/ijerr.2024.v42.005

Inputs
cn_ANKRD308
mu_TNXB
rs_APOB
rs_KRT23

Training Metrics		Test
Accuracy	0.895	0.875
AUC	0.648	0.66
Precision	0.81	0.667
Recall	0.274	0.125

Figure 9. A Model trained for 200 iterations.

Table 2. Showing the accuracy by single iteration and 200 iteration.

		Accuracy	AUC	Precision	Recall
Single iteration	Training	0.89	0.663	0.739	0.274
	Testing	0.871	0.641	0.571	0.125
Multiple (200) iteration	Training	0.895	0.648	0.81	0.274
	Testing	0.875	0.66	0.667	0.125

After the 200 iteration AUC at the training time 0.65 and at the testing time, AUC is 0.66. This result is better than the single iteration shown in table 2.

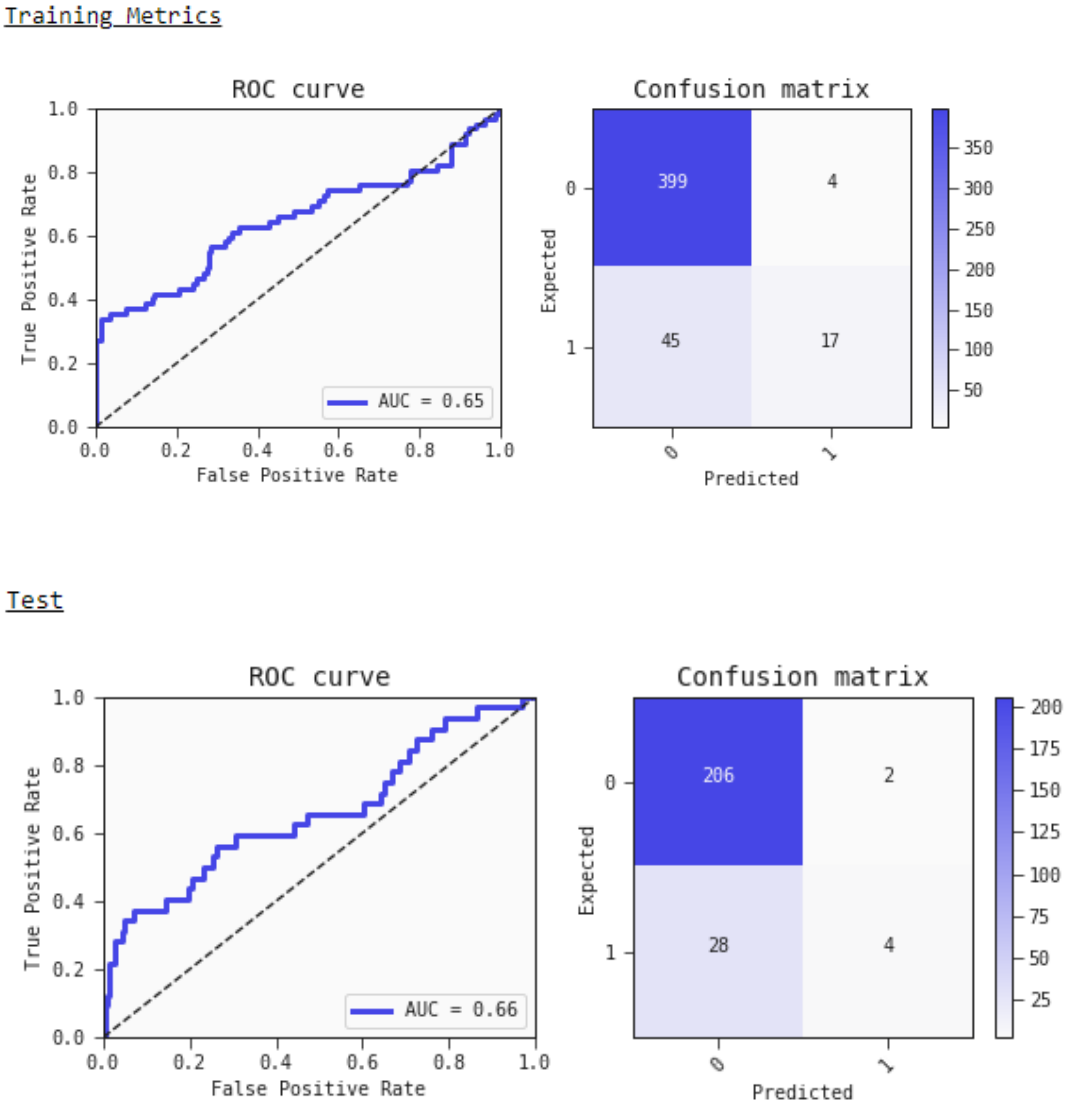


Figure 10. ROC curve for 200 iterations.

Looking at people without TNXB mutations

In people without TNXB mutations, high APOB and KRT23 gene expression are associated with death, as shown in figure 11.

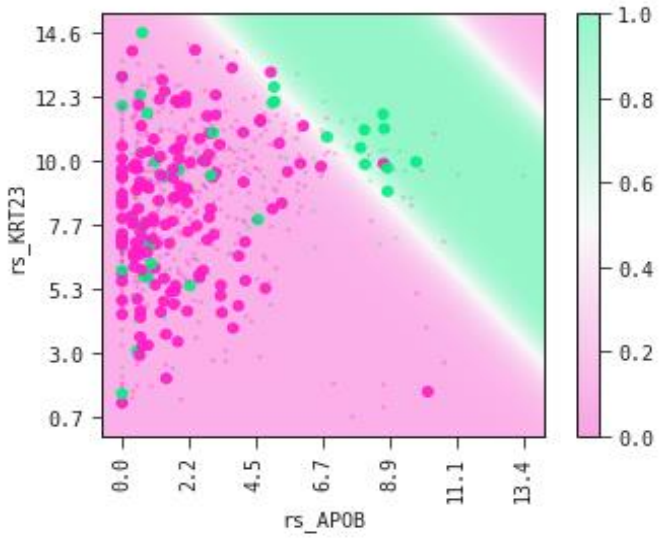


Figure 11. non-TNXB mutation carriers.

cn_ANKRD30B

mu_TNXB

In individuals without TNXB mutations, we observed that both high gene expression levels of APOB and KRT23 are associated with an increased risk of death. This suggests that the combination of elevated expression of both APOB and KRT23 might be a predictive factor for adverse health outcomes or mortality in this group.

It's essential to consider the biological context of these genes. APOB is involved in lipid metabolism and has been linked to cardiovascular health, while KRT23 is a keratin protein that can be associated with various cellular processes. High expression levels of these genes in individuals without TNXB mutations may indicate underlying health issues or specific disease pathways.

Looking at TNXB mutation carriers

In people with a TNXB mutation, only a high APOB is required for a case to be fatal, as shown in figure 12.

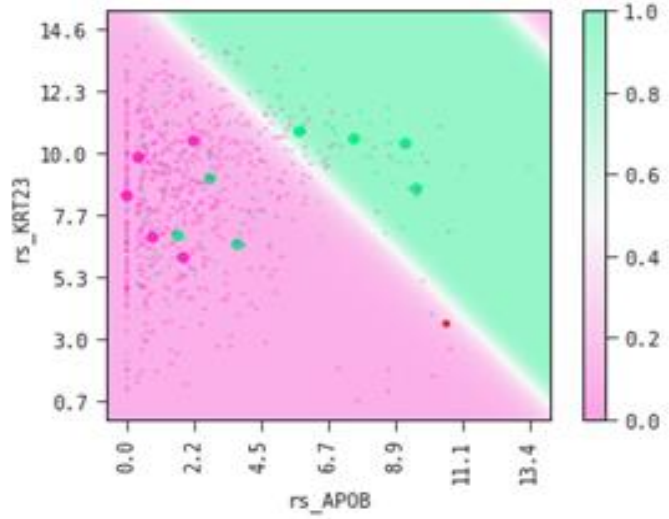


Figure 12. The predictions for TNXB-mutation carriers.

In individuals with TNXB mutations, we found that only high APOB gene expression is required for a case to be fatal. This suggests that in this genetic context, APOB expression levels might be more critical in determining survival outcomes compared to KRT23.

The observation that high APOB expression alone is associated with a fatal outcome in TNXB mutation carriers could be indicative of a unique genetic interaction or pathway specific to this subgroup. It might also point to a potential genetic vulnerability or susceptibility to certain health conditions that are influenced by APOB expression.

Discussion

Biological Mechanisms: Investigating the roles of APOB and KRT23 in relevant biological pathways and disease processes could provide insights into why their expression levels are linked to mortality in these groups.

Clinical Implications: These findings may have clinical implications. For individuals without TNXB mutations, monitoring APOB and KRT23 expression levels could help identify those at higher risk for adverse health outcomes. In contrast, for TNXB mutation carriers, focusing on APOB expression may be particularly important in assessing their health risks and designing potential interventions.

Genetic Interactions: Consider exploring potential interactions between TNXB mutations and the expression of APOB and KRT23. Genetic interactions can provide valuable insights into how specific genes or mutations modulate each other's effects. Tenascin-X is utilized in diagnostic procedures or as a component of a risk assessment instrument if it demonstrates itself to be a dependable prediction marker for the early diagnosis or tracking of breast cancer. It might also be a target for novel treatments intended to sabotage ECM pathways that support malignancy.

Validation and Further Research: It's crucial to validate these findings with larger and independent datasets to ensure their reliability. Additionally, further research can investigate the causality and underlying molecular mechanisms driving these associations.

Clinical Decision-Making: Depending on the strength and consistency of these associations, they could inform clinical decision-making, risk assessment, and personalized medicine approaches for individuals with and without TNXB mutations.

Conclusion

In conclusion, using the Gaussian method to calculate accuracy, the accuracy after a single iteration is 89% during training and 87.1% during testing with the omics dataset. When we used multiple iterations (200), the accuracy increased to 89.5% during training and 87.5% during testing. And our findings highlight intriguing associations between gene expression levels and survival outcomes in individuals with and without TNXB mutations. A new study is being done on the function of tenascin-X in breast cancer prediction. It includes investigating the effects of this protein on the onset and

DOI: https://doi.org/10.52756/ijerr.2024.v42.005

spread of breast cancer as well as its interactions with other elements of the extracellular matrix.

The QLattice identified a **genetic switch**, i.e., a mutation in a gene (TNXB) that seems to drive cancer severity. In Figure.10, we show the decision boundary for non-TNXB mutation carriers: Here, individuals with high APOB and KRT23 gene-expression seem to be at risk of dying. In Figure 11, we show the predictions for TNXB-mutation carriers. Here, high levels of APOB are predicted to be detrimental, no matter the levels of KRT23.

Conflict of Interest Statement

This is to certify that there is no conflict of interest in this paper.

References

Biswas, N., & Chakrabarti, S. (2020). Artificial intelligence (AI)-based systems biology approaches in multi-omics data analysis of cancer. *Frontiers Oncol., 10*, 588221. https://doi.org/10.3389/fonc.2020.588221.

Coleto, A., & Vega-Rodríguez M.A. (2020). Artificial bee colony algorithm based on dominance (ABCD) for a hybrid gene selection method. *Knowl. Based Syst., 205*, 106323.

Ghosh, D., & Poisson, L.M. (2009). Omics" data and levels of evidence for biomarker discovery. Genomics, 93, 13–16.

Isik, Z., & Ercan, M.E. (2017). Integration of RNA-Seq and RPPA data for survival time prediction in cancer patients. *Comput. Biol. Med., 89*, 397–404.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research, 44*(D1), D457–D462. https://doi.org/10.1093/nar/gkv1070

Khanfari, M. S., Cheki, M., Mohammadi, S. M., Moniri, S., Heydarheydari, S., & Rezaeijo, S.M. (2023). Exploring the efficacy of multi-flavored feature extraction with radiomics and deep features for prostate cancer grading on mp MRI. *BMC Med. Imaging, 23*(1), 195.

Lænkhol, A.V., Jensen, M.B., & Eriksen, J.O., (2020). Population-based Study of Prosigna-PAM50 and Outcome Among Postmenopausal Women with Estrogen Receptor-positive and HER2-negative Operable Invasive Lobular or Ductal Breast Cancer. *Clinical Breast Cancer, 20*(4), e423-e432. https://doi.org/10.1016/j.clbc.2020.01.013.

Masoudi-Sobhanzadeh, Y., Motieghader, H., Omidi, Y., & Masoudi-Nejad, A. (2021). A machine learning method based on the genetic and world competitive contests algorithms for selecting genes or features in biological applications. *Sci. Rep., 11*(1), 1–19.

Nouira, K., Maalej, Z., Rejab, F. B., Ouerfelly, L., & Ferchichi, A. (2020). Analysis of breast cancer data: a comparative study on different feature selection techniques. IEEE, *In 2020 International MultiConference on Organization of Knowledge and Advanced Technologies* (OCTA), pp. 1-11.

Rami, N., Kulkarni, B., Chibber, S., Jhala, D., Parmar, N., & Trivedi, K. (2023). In vitro antioxidant and anticancer potential of Annona squamosa L. Extracts against breast cancer. *Int. J. Exp. Res. Rev., 30*, 264-275. https://doi.org/10.52756/ijerr.2023.v30.024

Rezaeijo, S.M., Chegeni, N., Baghaei, N.F., Makris, D., Bakas, S. (2023). Within-Modality Synthesis and Novel Radiomic Evaluation of Brain MRI Scans. *Cancers* (Basel), *15*(14), 3565. https://doi.org/10.3390/cancers15143565.

Sagar, A. K., Banda, L., Sahana, S., Singh, K., & Singh, B. K. (2021). Optimizing quality of service for sensor-enabled Internet of healthcare systems. *Neuroscience Informatics, 1*(3), 100010.

Sagar, A. K., Singh, S., & Kumar, A. (2020). Energy-aware WBAN for health monitoring using critical data routing (CDR). *Wireless Personal Communications, 112*(2020), 273-302.

Sandhu, J.K., Kaur, A., & Kaushal, C. (2022). Analysis of Breast Cancer in Early Stage by Using Machine Learning Algorithms: A Review. *2022 IEEE International Conference on Current Development in Engineering and Technology* (CCET), Bhopal, India. pp. 1-7. https://doi.org/10.1109/CCET56606.2022.10080757.

Siegel, R.L., Miller, K.D., Fuchs, H.E., & Jemal, A. (2021). Cancer statistics. *CA Cancer J. Clin., 71*(1), 7-33. https://doi.org/10.3322/caac.21654

Singh, J., Goyal, D., & Vashist, A. (2024). Feature selection using correlation analysis for accurate breast cancer diagnosis in application of Synthetic High Dimensional Data in IGI Global, 2024.

Taghizadeh, E., Heydarheydari, S., Saberi, A., Jafarpoor, N.S., & Rezaeijo, S.M. (2022). Breast cancer prediction with transcriptome profiling using feature selection and machine learning methods. *BMC Bioinformatics, 23*(6), 1-9. https://doi.org/10.1186/s12859-022-04965-8

Tonmoy, S., Hiya, H.K., Hasan, M.Z., Zubair-Hassan, K.M., & Zahan, N. (2021). Breast Cancer Prediction with Gaussian Process Using Anthropometric Parameters. *12th International Conference on Computing Communication and*

Networking Technologies (ICCCNT), *2021*, 1-6. https://doi.org/10.1109/ICCCNT51525.2021.9579704

Tseng, (2019). Predicting breast cancer metastasis by using serum biomarkers and clinicopathological data with machine learning technologies. *International Journal of Medical Informatics*, *128*, 79–86. https://doi.org/10.1016/j.ijmedinf.2019.05.003

Vashist, A., Sagar, A., & Goyal, A. (2023). Breast Cancer Detection by Using Decision Tree. In book: *Advanced Communication and Intelligent Systems*, pp. 71-79. https://doi.org/10.1007/978-3-031-45121-8_7

Yadav, P., Bhargava, C., Gupta, D., Kumari, J., Acharya, A., & Dubey, M. (2024). Breast Cancer Disease Prediction Using Random Forest Regression and Gradient Boosting Regression. *International Journal of Experimental Research and Review*, *38*, 132-146. https://doi.org/10.52756/ijerr.2024.v38.012

Zenbout, I., Bouramoul, A., Meshoul, S., & Amrane, M. (2022). Efficient Bioinspired feature selection and machine learning-based frame work using omics data and biological knowledge data bases in cancer clinical endpoint prediction. *IEEE Access*, *11*, 2022.

How to cite this Article:

Apurva Vashist, Anil Kumar Sagar and Anjali Goyal (2024). Correlation of Prognostic Factors of Invasive Lobular Carcinoma and Invasive Ductal Carcinoma. *International Journal of Experimental Research and Review*, *42*, 50-59.

DOI : https://doi.org/10.52756/ijerr.2024.v42.005