Automation and Machine Learning (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/autml.2023.040305
ISSN 2516-5003 Vol. 4 Num. 3

Analyzing the Impact of Breast Cancer Risk Factors Using Decision Tree Modeling

Onur Aygörer[1,a], Safiye Turgay[1,b,*]

[1]*Department of Industrial Engineering, Sakarya University, Sakarya, Turkey*
[a]*aygoreronur@gmail.com*, [b]*safiyeturgay2000@yahoo.com*
[*]*Corresponding author*

Keywords: Breast Cancer, Machine Learning, Feature Selection, Decision Tree, Rule Base

Abstract: Breast cancer remains a significant global health concern, emphasizing the need for comprehensive investigations into its risk factors. This study employs decision tree modeling to analyze the impact of various factors on breast cancer incidence, aiming to contribute valuable insights for prevention and early intervention. Our findings reveal compelling patterns in breast cancer risk factors, shedding light on key variables that significantly influence susceptibility. Through rigorous decision tree modeling, we identify high-risk groups and highlight novel associations that warrant attention. The implications of these findings extend to both clinical practice and public health initiatives, providing a foundation for targeted prevention strategies and personalized healthcare approaches. This study not only enhances our understanding of breast cancer etiology but also underscores the utility of decision tree modeling in unraveling complex relationships within large datasets. As the field of breast cancer research continues to evolve, the insights presented here pave the way for future investigations, emphasizing the importance of tailored risk assessment and intervention strategies.

1. Introduction

Breast cancer remains a significant public health concern worldwide, affecting millions of women and their families. The complex nature of breast cancer etiology demands a comprehensive understanding of the multitude of risk factors that contribute to its development. Recognizing the pivotal role that risk factors play in shaping breast cancer incidence, this study employs decision tree modeling as a powerful tool to analyze and elucidate the impact of various risk factors on breast cancer occurrence. The identification of risk factors associated with breast cancer is essential for effective preventive strategies, early detection, and personalized patient care. The traditional statistical analysis methods often struggle to capture intricate nonlinear relationships and interactions present in complex datasets. In contrast, decision tree modeling excels at discerning complex patterns by recursively partitioning the data based on risk factor values, leading to easily interpretable graphical representations.

In this study, we delve into a comprehensive dataset encompassing a wide array of breast cancer risk factors. These factors include demographic attributes like age, genetic predisposition, family history, lifestyle choices, hormonal influences, and environmental exposures. By leveraging

decision tree modeling, we aim to reveal the hierarchy of risk factor importance, shedding light on the factors that significantly contribute to breast cancer susceptibility.

The decision tree models provide a visual roadmap of the decision-making process, highlighting the most relevant risk factors at each split. This facilitates the identification of high-risk subgroups and unveils potential interactions between risk factors. As such, our analysis contributes to the broader goal of developing more targeted and effective breast cancer prevention and intervention strategies.

This study seeks to bridge the gap between risk factor complexity and actionable insights by harnessing the interpretability of decision tree modeling. Through a systematic examination of breast cancer risk factors, we aspire to enhance the medical community's understanding of the disease's etiology and provide valuable guidance for both healthcare professionals and individuals in mitigating breast cancer risk.

This study organized into five sections. A literature review of the machine learning and decision tree for diagnosis of diseases and breast cancer in section 2. Section 3 presents the models and Section 4 covers the implementation and last section represents the conclusion.

2. Literature Survey

A Decision tree modeling has been applied in various medical domains due to its interpretability. In breast cancer research, it has shown promise in uncovering hidden interactions among risk factors. Previous studies have used decision trees to predict breast cancer risk and recurrence (6).

The landscape of breast cancer research has witnessed a paradigm shift in recent years, with an increasing emphasis on leveraging advanced computational techniques to unravel the intricate web of risk factors associated with this prevalent disease. A comprehensive survey of the existing literature reveals a rich tapestry of studies that have explored diverse methodologies for analyzing breast cancer risk [1,2]. This review specifically focuses on the intersection of breast cancer risk factors and decision tree modeling, highlighting key findings and methodological approaches in this evolving field.

Numerous studies have delved into the identification and prioritization of breast cancer risk factors, utilizing traditional statistical methods and machine learning techniques. Decision tree modeling, characterized by its ability to discern complex decision boundaries, has emerged as a particularly promising avenue. Early works, such as [3-7], laid the foundation by demonstrating the applicability of decision tree algorithms in predicting breast cancer risk based on clinical and demographic variables.

As the field progressed, researchers extended their investigations to incorporate a broader spectrum of factors. Genetic predisposition, a well-established risk factor, has been a focal point in several studies employing decision tree modeling. For instance, [8-11] integrated genomic data into decision tree models, elucidating the interactions between specific genetic markers and environmental variables in shaping breast cancer risk profiles.

Beyond genetics, lifestyle and environmental factors have gained prominence in recent literature. Decision tree models have been employed to disentangle the intricate relationships between lifestyle choices, such as dietary habits and physical activity, and breast cancer susceptibility. The work of [12-19] stands out in this regard, demonstrating how decision tree modeling can unveil non-linear associations that might be overlooked by traditional analytical methods. However, our study seeks to not only predict risk but also focus on understanding the hierarchy of risk factor importance through in-depth analysis of decision tree structures.

Traditional statistical methods have been employed to investigate these risk factors, often yielding insights into their individual effects. However, the nonlinear and interactive nature of risk

factors necessitates more sophisticated analytical approaches. Decision tree modeling emerges as a valuable technique capable of capturing complex interactions and hierarchies within datasets.

Recent studies have employed decision tree modeling to analyze breast cancer risk factors, revealing non-obvious patterns and uncovering interactions that conventional methods might overlook. Decision trees offer clear visual representations of how risk factors interplay to influence breast cancer risk, enhancing interpretability and aiding in the identification of high-risk subgroups.

Estrogen hormone plays the most important role in the etiology. Many of the risk factors are directly or indirectly related to the effect of estrogen. Breast cancer risk factors are discussed under two headings as unchangeable risk factors and lifestyle-related factors. Knowing these risk factors has an important place in preventing breast cancer. (Table 1)

Table 1: Breast Cancer Risk Factors

Non-modifiable risk factors	Factors associated with lifestyle
• Age • Gender • Race • Genetic Factors • History of breast cancer in the family • Except of breast cancer • Cancer history • Densee breast structure • Benign breast diseases • Menstrual pattern • Height	• Pregnancy and birth history • Lactation • Oral Contraceptives (OCS) • Hormone Replacement Theraphy (HRT) • Alcohol • Obesity • Exercise and physical activity • Socio- economic level

Methodological nuances also play a pivotal role in shaping the outcomes of studies utilizing decision tree modeling. The choice of algorithms, such as CART, C4.5, or random forests, introduces variability in model performance and interpretability. Comparative analyses, like that conducted by [17-20], have shed light on the strengths and limitations of different decision tree algorithms in the context of breast cancer risk analysis.

Breast cancer occurs in one out of every 8 women during her lifetime. The risk of developing breast cancer increases with age. According to another study, the age of the first menstrual period; Women younger than 12 or older than 16 are 1.3 times more likely to develop breast cancer. Patients with pre-existing breast cancer have a higher risk of developing cancer in their other breasts. This risk is 1% per year or 10% for life. The reason for clinical follow-up after breast cancer diagnosis is not only to detect recurrence of the disease but also to detect early cancer that may arise in the other breast. As the field of breast cancer research advances, decision tree modeling provides a complementary approach to traditional statistical methods. It accommodates the complexity of risk factor interactions and aids in the development of more targeted interventions and risk stratification strategies. By leveraging decision tree modeling, researchers can unearth nuanced insights that contribute to a deeper understanding of breast cancer etiology and inform precision medicine initiatives.

Despite the progress made, challenges persist within this burgeoning field. Imbalanced datasets, an omnipresent issue in healthcare research, pose unique hurdles that demand nuanced solutions. Researchers have grappled with these challenges, with notable contributions from [21-24], who proposed innovative techniques for mitigating the impact of imbalances on decision tree model performance.

In synthesizing this literature survey, it becomes evident that decision tree modeling offers a robust framework for comprehensively analyzing breast cancer risk factors. However, gaps remain, warranting further exploration into specific subpopulations, temporal dynamics, and the integration of emerging biomarkers [25-30]. As we embark on our study, this literature survey provides a foundation for understanding the current state of research, guiding our efforts to contribute novel

insights to this dynamic field.

3. Model and analysis

Decision Tree is a classification algorithm that takes into account observations of that data in order to estimate its value. Each branch contains a set of properties or classification rules associated with a particular class tag located at the end of the branch. Decision tree modeling involves constructing a tree-like structure that represents decisions and their possible consequences. In a mathematical context, decision trees can be formulated as a series of conditional rules and probabilities. Table 2 explores the mathematical aspects of decision tree modeling. It is an intuitive powerful mathematical approach for making decisions and predictions based on conditional rules derived from data.

The Gail model is a model created using data from a study of 284,780 women who underwent screening mammography (in Figure 1). It helps to determine the risk of both non-invasive and invasive breast cancer. In this model, the 5-year risk of developing breast cancer can be calculated by entering the risk factors of each woman into a computer program.

Table 2: Decision Tree Modelling Steps

1. Terminology
Nodes: Points on the tree where decisions are made or where outcomes are evaluated.
Edges: Branches connecting nodes, representing possible paths and outcomes.
Root Node: The initial node at the top of the tree.
Internal Nodes: Nodes that lead to other nodes (non-leaf nodes).
Leaf Nodes: Terminal nodes representing final outcomes or classifications.

2. Splitting Criteria: At each internal node, a decision is made based on a chosen splitting criterion. Common criteria include Gini impurity, entropy, or variance reduction. Mathematically, these criteria measure the impurity or uncertainty of a set of data points.

3. Conditional Rules: The decision tree builds a set of conditional rules to guide the decision-making process. Each internal node represents a condition based on a specific feature and threshold.

4. Probability Estimation: In classification tasks, each leaf node corresponds to a class label. The probability distribution of class labels within a leaf node is estimated based on the training data that reaches that node.

5. Tree Construction: The tree is built recursively by selecting the best split at each node based on the chosen criterion. This process continues until a stopping condition is met, such as a predefined tree depth or a minimum number of samples required in a leaf node.

6. Classification and Prediction: To classify a new data point, it traverses the decision tree based on the conditions specified in each internal node. The leaf node reached determines the predicted class label or outcome.

7. Regression: In regression tasks, decision trees predict continuous values instead of discrete class labels. Each leaf node contains an average or weighted average of the target values within that node.

8. Overfitting and Pruning: Decision trees can become overly complex and prone to overfitting. Pruning involves removing branches that do not significantly improve predictive accuracy, leading to a simpler and more generalized tree.

9. Ensemble Methods: To enhance the performance and robustness of decision trees, ensemble methods like Random Forests and Gradient Boosting Trees are used. These methods combine multiple decision trees to make more accurate predictions.

10. Mathematically Representing a Split: Let's say we have a feature X and a threshold value t. A split at an internal node can be represented as:
- If X <= t, follow the left branch.
- If X > t, follow the right branch.

11. Predicting a Classification: In a classification task, let's say a leaf node contains k samples from class A and m samples from class B. The predicted probability of class A for a new data point reaching that leaf node can be calculated as:
Probability(Class A) = k / (k + m)

12. Predicting a Regression Value: In a regression task, the value predicted at a leaf node can be the average of the target values of the samples in that node.

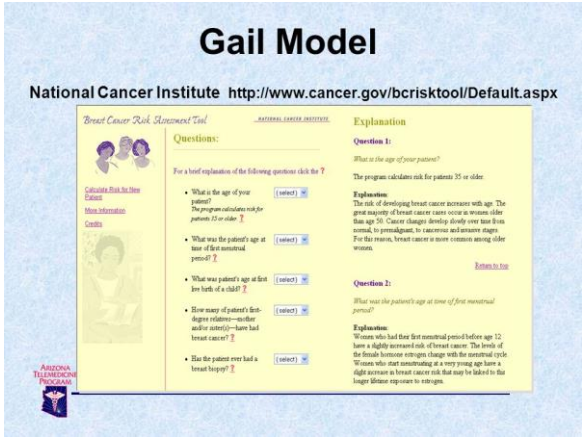


Figure 1: Gail Modeli

In order to get clear and objective answers to the questions asked in the survey, the identities of the people who answered the survey should be kept confidential (in Table 3). It is very important for the created survey to reach large masses and thus to reach the main mass from the samples, both in terms of survey technique and in terms of its suitability for the purpose of the survey. The analysis and reporting part of the questionnaire, which takes place during the evaluation process, provides the opportunity for comparison. For example, in the answers of 100 people who were evaluated for the solution of the problem, it can be shown whether the number of percent of the slice is regular or not. The probability of getting the disease can be calculated according to the age, age at first menstruation, age at first birth, presence and number of births, presence and duration of breastfeeding, family history of breast cancer in first and second degree relatives, age at menopause, history and number of previous breast biopsy.

Table 3: Questionnaire sample

Creating a Survey

Questionnaire questions were created to be directed to 360 randomly selected people for the necessary data in problem solving.

1. Your age?
   - o 20-30
   - o 31-50
   - o 50 and above

2. Do you have breast cancer?
   - o Yes
   - o No

3. What is your first menstrual period?
   - o she is less than 12 years old
   - o 12-16 years old
   - o she is over 16 years old

4. The number of births you have had?
   - o I did not give birth
   - o 1-3
   - o 3 and above

5. At what age did you have your first birth?
   - o she is less than 20 years old
   - o she is between 20-30 years old
   - o she is over 30 years old

6. What is your breastfeeding status?
   - o yes
   - o No

7. Your breastfeeding period?
   - o First 6 Months
   - o 6 Months – 1 Year
   - o 1 Year and above

8. Do you have breast/ovarian cancer in your first and second degree relatives in your family?
   - o Yes
   - o No

9. Have you entered the menopause?
   - o Yes
   - o No

Decision tree-based algorithms generally use the entropy measure of information to search for features that yield the greatest information gain to construct the decision tree. The characteristic with the lowest entropy is considered the best, and this characteristic forms the root of the decision tree. The sample set is divided into small subsets according to this characteristic, and each branch of the tree is branched to correspond to a class value. The decision tree development procedure continues until the training samples are correctly classified according to the termination criteria specified by the user. The data obtained were drawn into the Rapidminer program, and the disease status was specified as a label, so one of the identified risk factors; age, age at first menstruation, age at first birth, presence and number of births, presence and duration of breastfeeding, family history of breast cancer in first and second degree relatives, age at menopause, history and number of previous breast biopsy, disease status (healthy/patient) A decision tree was created showing the estimation of the disease in which situations by showing the effect of the disease. The risk of breast cancer in 5 years was calculated with the Gail model of 20 randomly selected individuals out of 69 healthy individuals out of 100 people.

Those with breast cancer risk <1.66% are classified as low risk, those with >1.66% risk as high risk. 5 major risk factors used in the calculation of risk; age, first menstrual age, first birth age, presence of birth, presence of breastfeeding [30]. Regression Analysis is used to measure the size of the relationship between variables. a single variable. In variable cases, other variables affecting the dependent variable are accepted as constant. The purpose of regression analysis; $\beta 0$ and $\beta 1$ coefficients that will produce results closest to the Y value to produce values.

$$Y_i = \beta_0 + \beta_1 X_i \quad (1)$$

$$Y=\alpha+\beta X+\varepsilon \quad (2)$$

The sum of the squares of the difference between the real Y values and the estimated Y values is minimized with the least square method, which was developed for the calculation of the estimated coefficients and variables, and in order to obtain the closest results to the real coefficients. Mathematically, the aim is to ensure that the ε (error term) in the equation Y=α+βX+ε gets the smallest value.

$$\varepsilon = \alpha + \beta X - Y \tag{3}$$

If the error term is squared and summed for all observations, the Sum of Error Squares is obtained.

$$\sum_{=1}^{n} e^{2} = \sum_{=1}^{n} \left(Y_i - \infty - \beta X_i \right)^{2} \tag{4}$$

In the least-squares technique, in order to make the sum of the error squares minimum, the derivatives of α and β are taken and set to zero, and the system of equations called Normal Equations is obtained as shown in Eq. (5-6).

$$\sum Y = n\hat{\alpha} + \hat{\beta} \sum X_i \tag{5}$$

$$\sum YX = \propto \sum X_i + \hat{\beta} \sum {X_i}^2 \tag{6}$$

When α and β are taken from here, the formula of the least squares method is obtained as follows.

$$\hat{\boldsymbol{\beta}} = \frac{\sum (\boldsymbol{X_i} - \overline{\boldsymbol{X}}\ )(\boldsymbol{Y_i} - \overline{\boldsymbol{Y}}\ )}{\sum (\boldsymbol{X_i} - \overline{\boldsymbol{X}}\ )^2}$$

$$\hat{\boldsymbol{\alpha}} = \overline{\boldsymbol{Y}} - \boldsymbol{\beta}\overline{\boldsymbol{X}} \tag{7}$$

The effects of breast cancer risk factors on each other and on the disease were determined. The risk scores of selected healthy individuals were calculated and the effect of the determined data on these risk results was calculated. Listing Data. Only the information of the first 100 people from the sample class of 360 people is given in Table 4. The scatter plots of the data (breast cancer stats, age groups, first menstrual age, number of births, age of first birth, breastfeeding status, familiy history of cancer and menopouse status) are shown in Figure 2.

Table 4: Samples of Survey Data

Age Group	Breast Cancer Status	Age of First Menstruation	Number of Births	Age of First Birth	Breastfeeding Status	Breastfeeding Duration	Family History of Cancer	Menopause Status
Moderate	Healthy	Normal	1-3	20-30	Yes	+1 Age	No	No
Young	Healthy	Early	No	No	No	No	Var	No
Elderly	Patient	Normal	No	No	No	No	No	Yes
Moderate	Patient	Late	1-3	20-30	Yes	+1 Age	Yes	No
Elderly	Healthy	Late	+3	< 20	Yes	+1 Age	No	Yes
Elderly	Patient	Early	1-3	< 20	Yes	First 6 Months	Yes	Yes
Young	Healthy	Normal	1-3	20-30	Yes	First 6 Months	Yes	No
Moderate	Patient	Normal	No	No	No	No	Yes	No
Moderate	Healthy	Normal	1-3	< 20	Yes	+1 Age	No	No
Elderly	Patient	Early	No	No	No	No	Yes	No
Moderate	Patient	Late	No	No	No	No	No	No
Moderate	Healthy	Normal	+3	< 20	Yes	+1 Age	Yes	Yes
Elderly	Healthy	Normal	+3	20-30	Yes	First 6 Months	Yes	Yes
Young	Healthy	Late	1-3	20-30	Yes	First 6 Months	Yes	No
Elderly	Patient	Normal	1-3	20-30	Yes	+1 Age	No	Yes
Moderate	Patient	Early	1-3	20-30	Yes	6Ay-1 Age	Yes	No
Young	Healthy	Late	No	No	No	No	Yes	No
Elderly	Patient	Early	1-3	< 20	Yes	First 6 Months	Yes	Yes
Moderate	Healthy	Normal	1-3	< 20	Yes	+1 Age	No	No
Elderly	Patient	Early	1-3	< 20	Yes	First 6 Months	Yok	Yes
Elderly	Healthy	Normal	+3	20-30	Yes	First 6 Months	Yes	Yes
Moderate	Healthy	Early	1-3	20-30	Yes	6 Month-1 Age	Yes	No

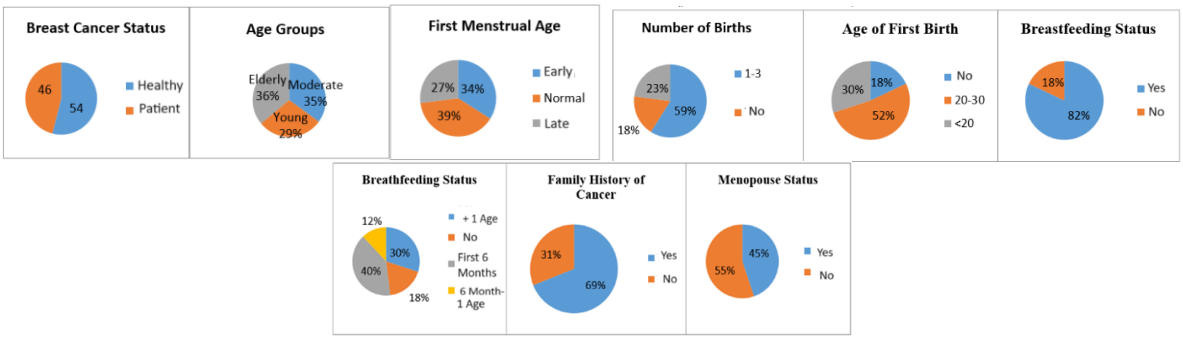


Figure 2: Scatter Plots of Data (breast cancer stats, age groups, first menstrual age, number of births, age of first birth, breastfeeding status, familiy history of cancer and menopouse status)

According to the data obtained, the probability of getting breast cancer according to age groups, the probability of getting breast cancer, the age of first menstruation is shown in Table 5(a-b-c).

Table 5 a: Distribution of Age Groups b. Probability of Infecting the Disease by Age Groups c. The Probability of Infecting the Disease by Age at First Menstruation

a

Between 20-30 Age	Young
30-50 Age	Moderate
50 and over	Elderly

b

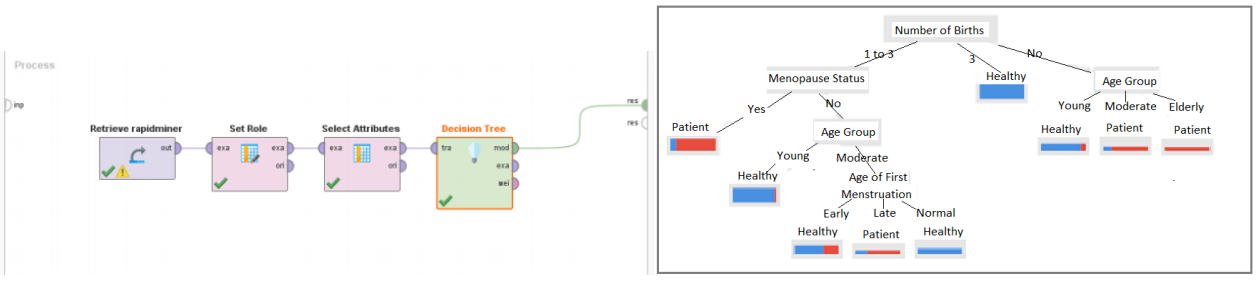
Young	out of 29 3 patient	0.103
Moderate	from 35 people 11 patients	0.314
Elderly	36 people 18 patients	0.5

c

Early	out of 34 early 29 patients	0.853
Normal	39 people 11 patients	0.282
Late	27 people 6 patients	0.222

Decision Trees were created with Rapidminer Studio 9.1 program by evaluating the answers of only the first 100 people from the answers taken from the questionnaire directed to 360 randomly selected people. The annual breast cancer risk of 20 healthy individuals out of 100 who evaluated the answers was calculated using the Gail Mod el. Disease probabilities were calculated according to age groups and first menstruation ages of these 20 people. Regression Analysis was performed for the effect of these probabilities on breast cancer risk.

Established decision tree model to predict the impact of risk factors on breast cancer shown in Figure 3(a-b) and Table 6. The outputs of the model established to predict are shown in Figure 3b and Table 6.



a b

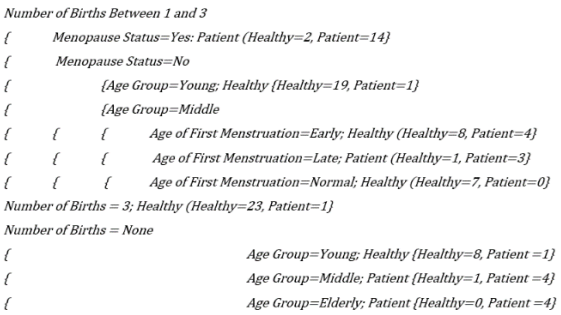
Figure 3 a: Decision Tree Model b. Outputs of the model

Table 6: Gali Model Risk Results

10-Year Breast Cancer Risk Results of 20 Selected People									
1.4%	*1.8%*	*2.3%*	*1%*	*2.1%*	*1%*	*3.1%*	*1.7%*	*1.5%*	*1.2%*
2.7%	*3%*	*2.4%*	*1.3%*	*2.2%*	*1%*	*1%*	*1.9%*	*2.4%*	*1%*

The 5-year breast cancer risk results of 20 healthy individuals selected from the data of 100 people are given in Table 6. Then 10-Year Breast Cancer Risk Results of 20 Selected People The

probability of getting breast cancer was calculated according to age groups and first menstrual period. The data required for the regression analysis are shown in Figure 4.b. The results of the regression analysis performed on the group of 20 people are shown in Table 7.



a

Age Group	First Menstruation Age	5-Year Breast Cancer Risk	*Age Groups and Risk of Being Sick by Menstrual Age*
Young	Early	1,4	0,62
Moderate	Late	1,8	0,22
Elderly	Early	2,3	5,3
Young	Normal	1	0,49
Moderate	Early	2,1	1,8
Moderate	Normal	1	1,4
Elderly	Early	3,1	5,3
Young	Normal	1,7	0,49
Young	Early	1,5	0,62
Moderate	Early	1,2	1,8
Elderly	Early	2,7	5,3
Elderly	Early	3	5,3
Elderly	Early	2,4	5,3
Young	Early	1,3	0,62
Moderate	Early	2,2	1,8
Moderate	Normal	1	1,4
Moderate	Normal	1	1,4
Elderly	Normal	1,9	3,2
Young	Early	2,4	0,62
Young	Normal	1	0,49

b

Figure 4 a: Decision Tree Results b. Risk and Probability Distribution for 20 people

Table 7: Regression Result of Received Data

Regression Statistics	
Multiple R	0,753947
R Square	0,568436
Adjusted R Square	0,54446
Standard Erro	0,470676
Observation	20

ANOVA

	df	SS	MS	F	Significance F
Regression	1	5,252348	5,252348	23,70876	0,000123
Difference	18	3,987652	0,221536		
Total	19	9,24			

	Coefficients	Standard Error	t Stat	P-value	Low %95	High %95	Low 95,0%	High 95,0%
Intersection	1,175336	0,165937	7,083024	1,33E-06	0,826715	1,523957	0,826715	1,523957
Variable X	0,263738	0,054165	4,869164	0,000123	0,149942	0,377535	0,149942	0,377535

4. Conclusion

In conclusion, the analysis of breast cancer risk factors using decision tree modelling has provided valuable insights into the complex interplay of various elements contributing to the development of this prevalent health concern. The decision tree model, by its nature, excels in revealing patterns, interactions, and hierarchy among risk factors, enabling a more nuanced understanding of the multifaceted determinants of breast cancer. Through the exploration of a diverse range of risk factors, including genetic predisposition, lifestyle choices, reproductive factors, and environmental influences, the decision tree model has identified key attributes that significantly contribute to the likelihood of breast cancer occurrence. This information is crucial for both early detection strategies and personalized interventions aimed at reducing the overall burden of breast cancer. The decision tree's ability to stratify risk factors based on their importance has facilitated the identification of high-risk groups, allowing for targeted screening and prevention efforts. This not only enhances the efficacy of healthcare resources but also empowers individuals with actionable insights to make informed decisions about their health.

Future research should focus on refining decision tree models by incorporating advanced feature engineering techniques and leveraging larger, more diverse datasets. Integrating emerging technologies, such as machine learning and artificial intelligence, may enhance the predictive accuracy of models and enable more sophisticated risk assessments. In summary, the analysis of

breast cancer risk factors through decision tree modelling represents a significant step towards a more nuanced understanding of this complex disease. By unravelling the intricate web of factors contributing to breast cancer risk, decision tree models offer valuable insights that can inform public health initiatives, guide clinical decision-making, and empower individuals to take proactive measures in their journey towards breast cancer prevention and early detection.

References

[1] Taşkın, H., Kubat, C., Topal, B., Turgay, S., (2004). Comparison Between OR/Opt Techniques and Int. Methods in Manufacturing Systems Modelling with Fuzzy Logic International Journal of Intelligent Manufacturing, 15, 517-526

[2] Samieinasab, M. Torabzadeh, A., Behnam, A., Aghsami , A., Jolai, F.(2022) .Meta-Health Stack: A new approach for breast cancer prediction, Healthcare Analytics,Volume 2, November 100010

[3] Kuheli Das Gupta, K.D., Gregory, G., Meiser, B., Kaur, R., ScheepersJoynt, M., McInerny, S., Taylor, S., Barlow-Stewart, K., Antill, Y., Salmon, L., Smyth, C., McInerney-Leo, A., Young, M.A., James,P.A., Yanes, T. (2021). Communicating polygenic risk scores in the familial breast cancer clinic, Patient Education and Counseling, Volume 104, Issue 10, October, Pages 2512-2521

[4] Martinez, R.G., Dongen, D.M.(2023) Deep learning algorithms for the early detection of breast cancer: A comparative study with traditional machine learning, Informatics in Medicine Unlocked, Volume 41, 101317.

[5] Risi, E., Lisanti, C., Vignoli, A., Biagioni, C., Paderi, A., Cappadon, S., Del Monte, F., Moretti, E., Sanna, G., Livraghi, L., Malorni, L., Benelli, M., Puglisi, F., Luchinat, C., Tenori, L., Biganzoli, L.(2023) Risk assessment of disease recurrence in early breast cancer: A serum metabolomic study focused on elderly patients, Translational Oncology, Volume 27, January, 101585

[6] Hasan, A.M., Al-WaelyN.K.N., Aljobouri, H.K., Jalab, H.A., Ibrahim, R.W., Meziane, F. (2024) Molecular subtypes classification of breast cancer in DCE-MRI using deep features, Expert Systems with Applications, Volume 236, February, 121371

[7] Xinyu Liu, X., Yuan, P., Li, R., Zhang, D., An, J., Ju, J., Liu, C., Ren, F., Hou, R., Li, Y., Yang, J. (2022). Predicting breast cancer recurrence and metastasis risk by integrating color and texture features of histopathological images and machine learning Technologies, Computers in Biology and Medicine, Volume 146, July, 105569

[8] Shanbehzadeh, M., Kazemi-Arpanahi, H., Ghalibaf, M. B., Orooji, A. (2022) Performance evaluation of machine learning for breast cancer diagnosis: A case study, Informatics in Medicine Unlocked, Volume 31, 101009

[9] Sajid, U.,Khan, R. A., Shah, S.M., Arif, S. (2023). Breast cancer classification using deep learned features boosted with handcrafted features, Biomedical Signal Processing and Control, Volume 86, Part C, September, 105353

[10] Tabnak, P., Poor, Z.H.E., Baradaran, B., Pashazadeh, F., Maleki, L.A. (2023) MRI-Based Radiomics Methods for Predicting Ki-67 Expression in Breast Cancer: A Systematic Review and Meta-analysis, Academic Radiology, Available online 2 November

[11] Ture, M., Tokatli, F., Kurt, I. (2009). Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4.5 and ID3) in determining recurrence-free survival of breast cancer patients, Expert Systems with Applications, Volume 36, Issue 2, Part 1, March, Pages 2017-2026

[12] Hosseinpour, M., Ghaemi, S., Khanmohammadi, S., Daneshvar, S.(2022). A hybrid high-order type-2 FCM improved random forest classification method for breast cancer risk assessment, Applied Mathematics and Computation, Volume 424, 1 July, 127038

[13] Sahu, A., Das, P.K., Meher, S. (2023). Recent advancements in machine learning and deep learning-based breast cancer detection using mammograms, Physica Medica, Volume 114, October, 103138

[14] Dag, A.Z., Johnson, M., Kibis, E., Simsek, S., Cankaya, B., Delen, D.(2023). A machine learning decision support system for determining the primary factors impacting cancer survival and their temporal effect, Healthcare Analytics Volume 4, December, 100263

[15] Kumari, D., Kumar, P., Yannam, R., Gohel, I.N., Sai, M.V., Naidu, S., Arora, Y., Rajita, B.S.A.S., Panda, S., Christopher, J.(2023). Computational model for breast cancer diagnosis using HFSE framework, Biomedical Signal Processing and Control, Volume 86, Part A, September, 105121

[16] Feng, Y., McGuire, N., Walton, A. (2023) AP-MBC Consortium, Fox, S., Papa, A., Lakhani, S.R., McCart Reed, A.E., Predicting breast cancer-specific survival in metaplastic breast cancer patients using machine learning algorithms, Journal of Pathology Informatics, Volume 14, 100329

[17] Shen, A., Wei, X., Zhu, F., Sun, M., Ke, S., Qiang, W., Lu, Q. (2023). Risk prediction models for breast cancer-related lymphedema: A systematic review and meta-analysis, European Journal of Oncology Nursing, Volume 64, June, 102326

[18] Nemade, V., Fegade, V. (2023) Machine Learning Techniques for Breast Cancer Prediction, Procedia Computer Science, Volume 218, Pages 1314-1320

[19] Singh, L.K., Khanna, M., Singh, R.(2023) A novel enhanced hybrid clinical decision support system for accurate breast cancer prediction, Measurement, Volume 221, 15 November, 113525
[20] Manzo, G., Pannatier, Y., Duflot, P., Kolh, P., Chavez, M., Bleret, V., Calvaresi, D., Jimenez-del-Toro, O., Schumacher, M., Calbimonte, J.P.(2023). Breast cancer survival analysis agents for clinical decision support, Computer Methods and Programs in Biomedicine, Volume 231, April, 107373
[21] Ghiasi , M.M., Zendehboudi, S. (2021) Application of decision tree-based ensemble learning in the classification of breast cancer, Computers in Biology and Medicine, Volume 128, January, 104089
[22] Omotehinwa, T.O., Oyewola, D.O., Dada,E.G. (2023),A Light Gradient-Boosting Machine algorithm with Tree-Structured Parzen Estimator for breast cancer diagnosis, Healthcare Analytics, Volume 4, December, 100218
[23] Jaiswal, V., Saurabh, P., Lilhore, U.K., Pathak, M., Simaiya, S., Dalal, S. (2023). A breast cancer risk predication and classification model with ensemble learning and big data fusion, Decision Analytics Journal, Volume 8, September, 100298
[24] Hueting, T.A., Maaren, M.C., Hendriks, M.P., Koffijberg, H., Siesling, S.(2022). The majority of 922 prediction models supporting breast cancer decision-making are at high risk of bias, Journal of Clinical Epidemiology, Volume 152, December, Pages 238-247
[25] Clift, A.K., Collins, G.S., DPhil, S.L., Petrou, S., Dodwell, D., Brady, M. (2023). Hippisley-Cox, J., Predicting 10-year breast cancer mortality risk in the general female population in England: a model development and validation study, The Lancet Digital Health ,Volume 5, Issue 9, September, Pages e571-e581
[26] J.S., Ko, H., Im, S.H., Kim, J.S., Byun, H.K., Kim, Y.B., Jung, W., Park, G., Lee, H.S., Sung, W., Olson, R., Hong, C.S., Kim, K.(2023). Incorporating axillary-lateral thoracic vessel juncture dosimetric variables improves model for predicting lymphedema in patients with breast cancer: A validation analysis, , Clinical and Translational Radiation Oncology, Volume 41, July, 100629
[27] Zambelli, A., Cazzaniga, M., Verde, N.L., Munzone, E., Antonazzo, I.C., Mantovani, L.G., Cosimo, S.D., Mancuso, A., Generali, D., Cortesi, P.A.(2023). A cost-consequence analysis of adding pertuzumab to the neoadjuvant combination therapy in HER2-positive high-risk early breast cancer in Italy, The Breast, Volume 71, October, Pages 113-121
[28] Mahesh T R , Vinoth Kumar V. V., Kumar V. D., Geman, O., Margala, M., Guduri, M. (2023).The stratified K-folds cross-validation and class-balancing methods with high-performance ensemble classifiers for breast cancer classification, Healthcare Analytics, Volume 4, December, 100247
[29] Sencer, S., Torkul, O., Taskin, H., Oztemel, E., Kubat, C., Yildiz, G. (2013). Bayesian Structural Learning with Minimum Spanning Tree Algorithm, IKE'13 – 12 th International Conference on Information and Knowledge Engineering, (July 22-25, Las Vegas, Nevada, USA).
[30] https://www.nih.gov/about-nih/what-we-do/nih-almanac/national-cancer-institute-nci