CME LIVE OPEN ACCESS

Clinical and Medical Engineering Live

Research Article ISSN: 2997-612X

Clinical and Pathological Factors Influencing Breast Cancer Recurrence Using Structured Patient Data Correlation and Relationship Analysis

Noor AlGailani and Oke Oluwafemi Ayotunde*

Department of Computer Information System, Near East University, Cyprus.

Corresponding Author: Oke Oluwafemi Ayotunde, Department of Computer Information System, Near East University, Cyprus.

Received: 2026 Feb 27 **Accepted:** 2026 Mar 20 **Published:** 2026 Mar 30

Abstract

Breast cancer is among the most prevalent cancer and poses difficult challenges facing patients and doctors after initial treatment. This study looks at how common clinical factors—like tumor size, number of affected lymph nodes, menopause status, and radiation therapy—might influence the chances of cancer returning. We used a cleaned version of the UCI Breast Cancer dataset, which includes records for 286 patients. To make the data easier to analyze, we converted categorical ranges (such as tumor size and lymph nodes) into numeric midpoints, and handled missing values in key columns like node-caps. The study focused on three questions: (1) How are tumor size and lymph node involvement linked to recurrence? (2) Does menopause status and the presence of node caps affect recurrence risk? (3) Can radiation therapy reduce this risk in more severe cases? The results showed that lymph node involvement was more strongly related to recurrence than tumor size. Premenopausal women with node caps had the highest recurrence rate. Radiation therapy, while often used for serious cases, did not always reduce the risk, especially when tumors were large and more lymph nodes were involved. These findings show how simple, structured clinical data can help identify high-risk patients early. This can support better treatment planning, especially in healthcare settings with limited resources. Future research should include time-based and molecular data to develop more precise and tailored prediction models.

Keywords: Breast Cancer, Data Analysis, Categorical Analysis, Recurrence Prediction, SDG 3, Good Health and Well-Being

1. Introduction

Breast cancer is the most frequently diagnosed cancer among women globally and also leads to many cancer-related deaths. WHO reports that more than 2.3 million people got cancer and 685,000 died from it in 2022 and these cases were more prevalent in low- and middle-income nations [1]. Due to limited access in many regions, many patients are diagnosed at advanced stages with poorer outcomes. The need for solutions in global health are needed and that SDG 3 is especially crucial as it seeks healthcare alternatives that can be used by many, scaled up and are just. This research explores patients who have had breast cancer before, trying to predict if the disease will return by analyzing certain clinical elements. Being able to spot signs of cancer recurrence plays a major role in improving research on breast cancer. By spotting patients at high risk of a recurrence earlier, doctors are able to manage treatment better, follow them more closely and try to extend their lives. Even though genomic and imaging information is often used in predictive modeling for healthcare, this method is too resource-intensive for most healthcare environments, especially those in under-resourced settings.

Yet, the use of genomic data for advanced cancer testing is still limited in low- and middle-income countries due to high costs, complex equipment, and a shortage of qualified professionals. As a result, many researchers suggest that doctors use simpler data combined with machine learning to overcome these challenges [2]. Currently, only very few studies are conducted on breast cancer screening of populations in rural or low-resource areas. This highlights the need for we need inexpensive approaches that involve structured clinical records and artificial intelligence to improve early diagnosis [3].

As a result, researchers are studying whether structured clinical data including tumor size, lymph node status and menopause information can serve as affordable and universal alternatives to predict cancer outcomes. Encouragingly, studies carried out recently have shown that these variables can create accurate and interpretable models [4]. Discovered that using tumor diameter and lymph node status with

ensemble methods, especially AdaBoost, is especially effective when SHAP is included. In the same way, showed that tumor size and the presence of positive lymph nodes are important factors predicting recurrence by using regression and ensemble techniques [5]. The finding from these studies suggest that clinical data collected from routine practice can be both valuable and practical.

Recent research highlights that reliance on deep learning models that include structured clinical information improves the accuracy of identifying early tumor relapses. Not only have experts confirmed these models as effective, but healthcare settings have found them to be useful, especially in healthcare setting where advanced genomic tools are not available [6]. Noted that adding hormone receptor (ER/PR) levels to the model increases its ability to predict outcomes. However, molecular markers like HER2 and Ki67, which are used in clinical practice, have been found to be less reliable for predicting treatment outcomes in some circumstances [7]. Many studies don't fully consider how factors like tumor size, menopause status, lymph node involvement, and adjuvant therapies impact recurrence rates in breast cancer patients [8]. Even though these factors significantly influence treatment decisions [8,9] . The results from few studies are clear and easy to apply in real-word clinical settings. This research helps fill that gap by using carefully collected clinical information and rigorous statistics to study the impact of these factors on breast cancer relapse.

1.1. Aim

Assessing the possible effects of tumor size, menopause status, lymph-node involvement, presence or absence of nodal caps and radiotherapy treatment on the likelihood of cancer returning clinically.

1.2. Objective of the Study

This study aims to explore how a patient's tumor size, lymph node, menopause status, presence of nodal caps and radiation therapy affect the likelihood of the cancer recurrence. By combining statistical analysis and practical applications, the study aims to provide insights that support prompt cancer care and prompt fairness in delivering personalized treatment, in line with Sustainable Development Goal 3.

1.3. Structure of the Paper

- **Section 1 (Introduction):** explains the context of the study, its significance, the questions it aims to answer and its overall objective.
- **Section 2 (Methodology):** presents the methods and analytical phases used in the study are presented. It shared details on the chosen dataset, how the data was handled, the exploratory data analysis and the key results of the work.
- **Section 3 (Results)**: shows the answers and conclusions to the research questions.
- **Section 4 (Discussion):** present the clinical and practical value of the results, comparing them with the current state of knowledge.
- **Section 5 (Conclusion):** reviews key lessons and suggesting areas for further research.

1.4. Research Questions

- **RQ1:** How do the size of the tumor and the number of affected lymph nodes affect the chance of breast cancer coming back after treatment?
- **RQ2:** Does a woman's menopause stage and the presence of nodal caps affect the chances of breast cancer returning?
- **RQ3:** Can radiation therapy lower the chances of breast cancer coming back when the tumor is large and lymph nodes are involved?

2. Methodology

2.1. Exploratory Data Analysis (EDA)

- **Objective:** Observe the different distributions, patterns and how the data connects to one other.

2.2. Approaches

2.2.1. Summary Statistics

A Key part of our understanding come from looking at how the important numbers in the dataset were distributed, especially tumor size and lymph node involvement. When the data was conversions to mid-range values, we found most tumors measured about 24.8 mm in size. Four out of five tumors were between 10- and 34-mm. Testing of the lymph nodes found that most patients (close to 60%) had 0–2 of the vessels involved. The findings conform to early-stage breast cancer cases. Most patients in the majority class had the 'no recurrence' label which is what we would see in successfully treated breast cancer cases.

2.2.2. Missing Value Check (Post-Cleaning Validation)

Some of the columns in the training data such as 'node-caps' and 'menopause,' had missing or unknown values during preprocessing. To handle this, we Labeled these situations as 'unknown' or applied appropriate encoding. After cleaning, we confirmed that key columns for analysis had no missing values. While we kept the "unknown" value during preprocessing to maintain data integrity, their inclusion could introduce some ambiguity in certain subgroup analyses. instances of the "unknown" categories (such as "unknown node-caps") were checked and since they only appeared infrequently, their effect was not significant. However, further efforts might use imputation or collect more detailed data to deal with the uncertainty in such estimates.

2.2.3. Study Design

The research used different clinical characteristics, such as, tumor size, lymph node involvement, menopausal status, tumor capsule presence and radiation therapy to analyze their role in breast cancer recurrence. All analyses were performed with Python, following a planned and data-driven approach. The Steps in the experiment were documented to enable others to replicate the study and ensure its transparency.

2.2.4. Dataset Description

This research used a cleaned and preprocessed version of a breast cancer recurrence dataset that is widely available online. The source of the data was the Breast cancer data set from the GitHub repository: https://github.com/datasets/breast-cancer/blob/main/README.md.

- File Name: Cleaned_breast_cancer_data_xlsx
- Sample Size: 286 records, each representing a different patient.
- Number of Features: 10 clinical attributes.

❖ **Included Variables**

- Age, menopause, Tumor size, inv-nodes, node-caps, deg-malign breast, breast-quad, irradiate, and class (recurrence status).

❖ **Target Variable**

- Class indicates whether the patient experienced a recurrence (recurrence-events) or not (no-recurrence-events).

2.2.5. Data Preprocessing

Preprocessing steps were carried out prior to data analysis

- For missing values in the node-caps or menopause columns, we chose not to remove or guess these records. Instead, we labeled them as "unknown" to keep the dataset intact and later checked their impact, finding they had minimal effect on the analysis.
- Label encoding was applied to categorize recurrence as 0 and the no recurrence one as 1. Similarly, the irradiate column was coded as 0 (no) and 1 (yes).
- To simplify the analysis, the midpoints set of the ranges for tumor-size and inv-nodes were calculated and used instead of the ranges themselves.
- A new interaction term was created by multiplying the tumor size by the number of affected lymph nodes to explore any potential additional effects.
- The main tools used for preprocessing were: pandas, numpy and sklearn , preprocessing.

2.2.6. Exploratory Data Analysis (EDA) Techniques

The data structure and distribution were Analyzed using various exploratory data Analysis (EDA) methods.

- Descriptive statistics were calculated, Means, minimums, maximums and counts by group.
- Key patterns were visualized through correlation heatmaps, stacked bar charts and interaction plots.
- Crosstab analysis was used to examine how often menopause status was associated with certain node caps.

2.2.7. Analytical Approach

Unlike some previous studies, this work focused on easy-to-understand statistical methods instead of machine learning algorithms. I chose this method on purpose because the dataset had not many patients (286) and many variables that were categorical. In these situations, especially when resources are scarce as they are in healthcare, easy-to-use methods usually offer better and more reliable results than AI models. Unlike other studies, we used descriptive and statistical methods here, rather than machine learning models, to address the research questions.

- For **RQ1**: The degree to which tumor size, lymph node involvement and recurrence are related was assessed using Pearson correlation coefficients.
- For **RQ2**: To study how menopause and node-caps affect recurrence, associations were charged with percentage breakdowns.
- For **RQ3**: To observe combined effects, (tumor size × lymph nodes) interaction was looked at in different radiation therapy groups.
- All findings were backed up by the right visuals and neatly laid out in understandable tables.

2.2.8. Tools and Environment

- **Programming language:** Python
- **Platform:** Google Colab
- **Libraries used:** Panda, Numpy, Matplotib, Seaborn, Sklearn, Preprocessing
- **Execution environment:** Jupter Notebook

2.2.9. Reproducibility Statement

All steps of the research starting with data cleaning and preprocessing and ending with data analysis and visualization were carried out using Python code that is both repeatable and well-documented. All operations were designed to work together and make it possible for others to replicate the study.

3. Results

3.1. Exploratory Data Analysis (EDA)

3.1.1. Distribution Plots

Distribution plots were created to examine the behavior of tumor size and lymph nodes. The distribution of tumor size was skewed toward the smaller end, with the highest number of cases found between 10 and 34 mm. The distribution of affected lymph nodes was right skewed and the majority of patients had fewer than 3 involved nodes.

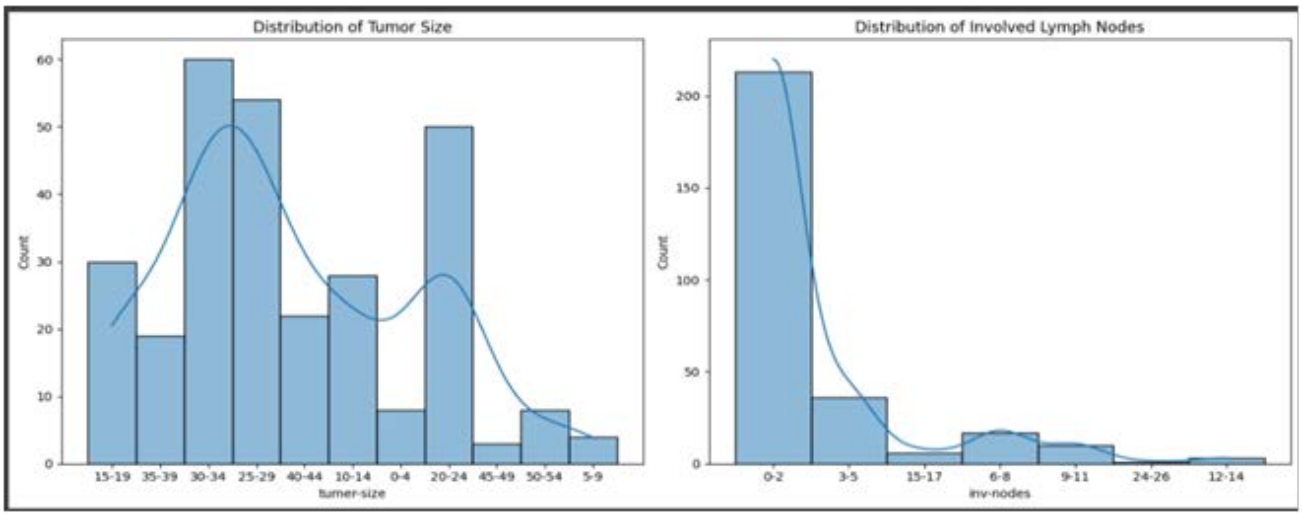

Figure 1: Tumor Size and Lymph Node Distribution

It is clear from Figure 1 that the distribution is right-skewed. Larger tumor sizes and higher numbers of affected lymph nodes appear less frequently in the dataset.

3.1.2. Breast Side and Quadrant Analysis

Instead of gender and purchasing history, breast side and breast quadrant were set as input categories in the new dataset. The distribution of breast tissue was throughout both left and right breasts. However, a higher number of tumors were observed in the upper and lower quadrants of the left breast.

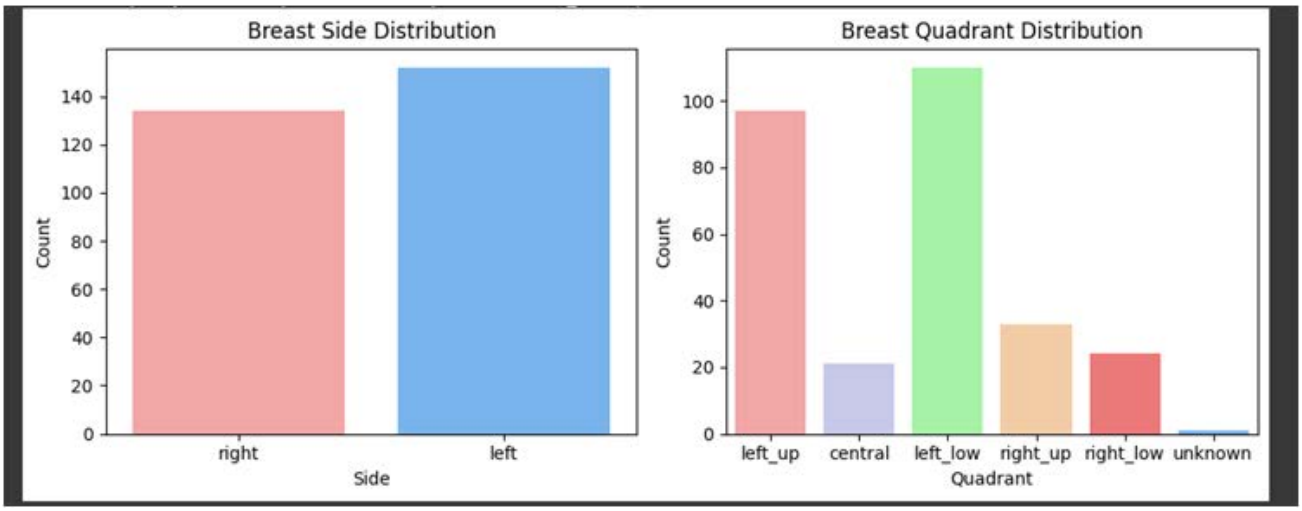

Figure 2: Distribution of Breast Side and Quadrant

This is confirmed by Figure 2, which shows that more tumors are identified in the 'left' and 'left_up' quadrants compared to the other quadrants.

3.2. Categorical Analysis

3.2.1. Objective: Understand behavior based on groupings

3.2.2. Approaches

3.2.3. Group by Menopause and Node Caps

We looked at the ways in which the presence of node caps, whether patients had undergone menopause, and cancer recurrence were grouped together. According to the crosstab analysis, premenopausal patients with node caps were much more likely to experience recurrence. It seems that when tumor growth affects hormones and extends outside the breast, the risk of a worse outcome goes up.

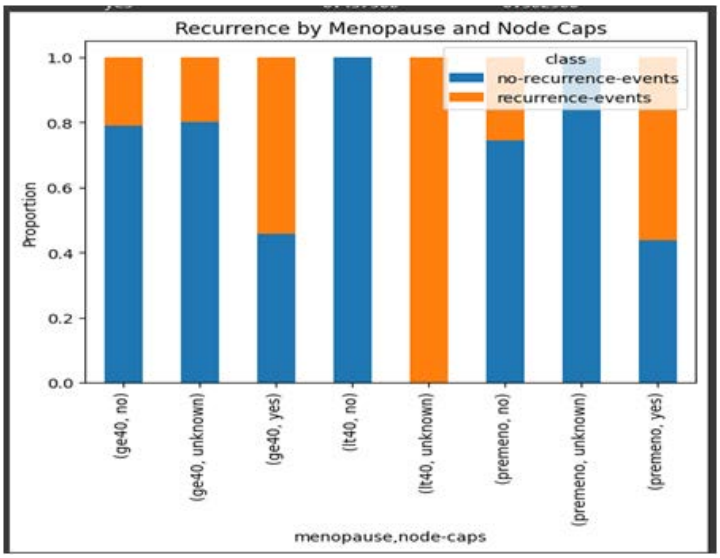

Figure 3: Recurrence by Menopause and Node Caps

In Figure 3, we can see that women who are premenopausal and have node caps experience the highest rates of recurrence.

3.2.4. Recurrence Class Distribution

To see the percentage of patients with a recurrence, we prepared a visual representation. Between 1 in 4 and 1 in 3 patients experienced recurrence and the remaining 3 out of 4 did not. Given this large difference in class sizes, using predictive models needs to take this into account.

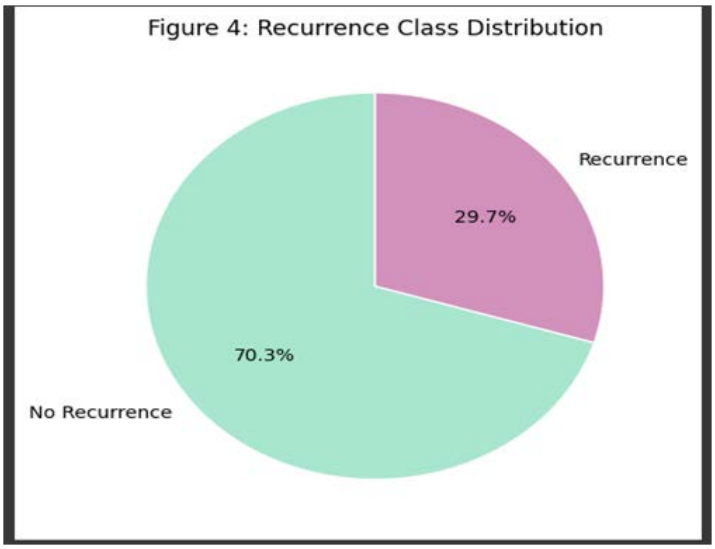

Figure 4: Recurrence Class Distribution

Recurrence events are shown in Figure 4: as a pie chart with the majority of patients not experiencing them.

3.2.5. Age Group Segmentation

A review of the data showed that over half of the patients were between 40–59 years old. The group aged 40–49 had a slightly higher recurrence after treatment. This discovery might support early interventions for middle-aged patients.

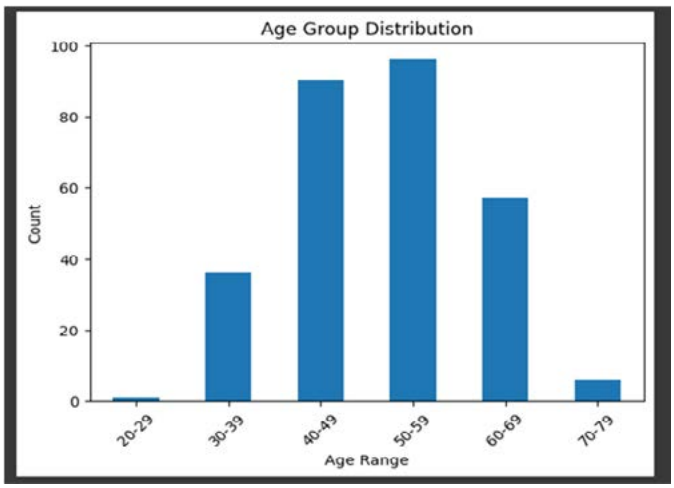

Figure 5: Age Group Distribution

Figure 5 shows that the groups with the group with the highest numbers of patients are once 40–49 and 50–59, although in the younger group, the recurrence trend in younger group remains moderate.

3.3. Time Series or Trend Analysis

3.3.1. Objective

Discover how data changes over time nothing in this dataset afterward when the participants were diagnosed, how long their treatment lasted or how regularly they were monitored afterwards. For this reason, we were unable to examine time-series data or observe trends over different periods. Thinking about the timing of events in future research can add detail to the way disease develops and comes back. Even though there was no time information in this dataset, conducting survival analysis could help future studies learn how clinical characteristics affect the period between recurrences. It would help view patient outcomes in a more flexible manner as time passes.

3.3.2. Results for the Questions (RQ1: Tumor Size and Lymph Nodes Vs Recurrence)

3.3.3. Correlation Matrix

- Tumor size to: 0.29
- Lymph nodes affected recurrence: 0.43
- Tumor size to Lymph nodes: 0.48

Lymph node involvement is more strongly associated with breast cancer than the size of the original tumor.

	Tumor Size	Lymph Nodes Affected	Recurrence
Tumor Size	1.000	0.160	0.175
Lymph Nodes Affected	0.160	1.000	0.276
Recurrence	0.175	0.276	1.000

Table 1: Correlation Matrix among Key Variables

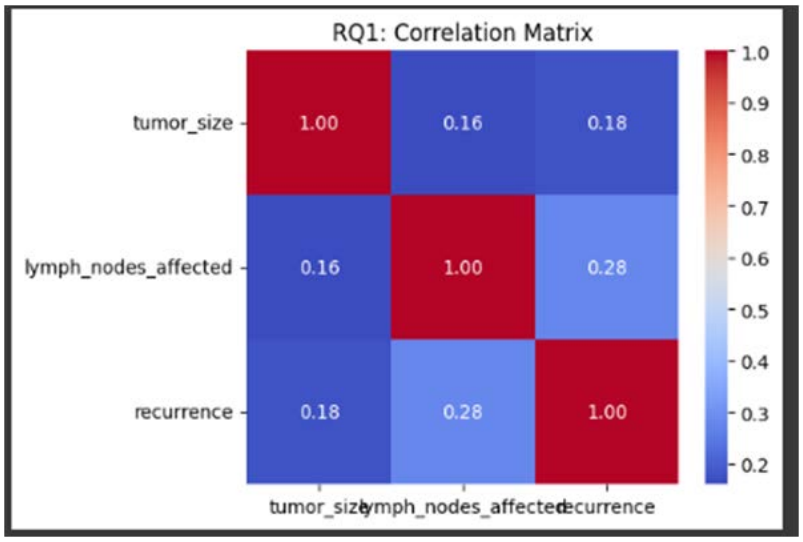

Figure 6: Heatmap Visualization of Correlation Matrix

Citation: AlGailani, N., Ayotunde, O. O. (2026). Clinical and Pathological Factors Influencing Breast Cancer Recurrence Using Structured Patient Data Correlation and Relationship Analysis. *Clin Med Eng Live, 4*(1), 1-12.

It can be seen from the correlation matrix that lymph node involvement is more strongly associated with breast cancer recurrence (0.43) than tumor size (0.29). the correlation between tumor size and lymph node count (0.48) clearly indicates that larger tumors are usually accompanied by a greater number of affected lymph nodes. These insights are clearly presented in Table 1 and Figure 6.

3.3.4. RQ2: Menopause and Node Caps Vs Recurrence

3.3.5. Crosstab and Stacked Bar Chart

• Higher Recurrence seen in groups with "premenopause" and "node-caps= yes".

• Breast cancer relapse is often predicted by both menopause status and the presence of node caps.

menopause	Node Caps	No Recurrence (0)	Recurrence (1)
ge40	no	0.790000	0.210000
ge40	unknown	0.800000	0.200000
ge40	yes	0.458333	0.541667
1t40	no	1.00000	0.000000
1t40	unknown	0.000000	1.000000
premeno	no	0.743590	0.256410
premeno	unknown	1000000	0.000000
premeno	yes	0.437500	0.562500

Table 2: Normalized Crosstab of Recurrence by menopause Status and Node Caps

The most common data point was observed in premenopausal women and those aged 40 or older with node caps. The results indicate that both node caps and menopausal status are important in predicting the likelihood of breast cancer Recurrence in women, as presented in Table 2.

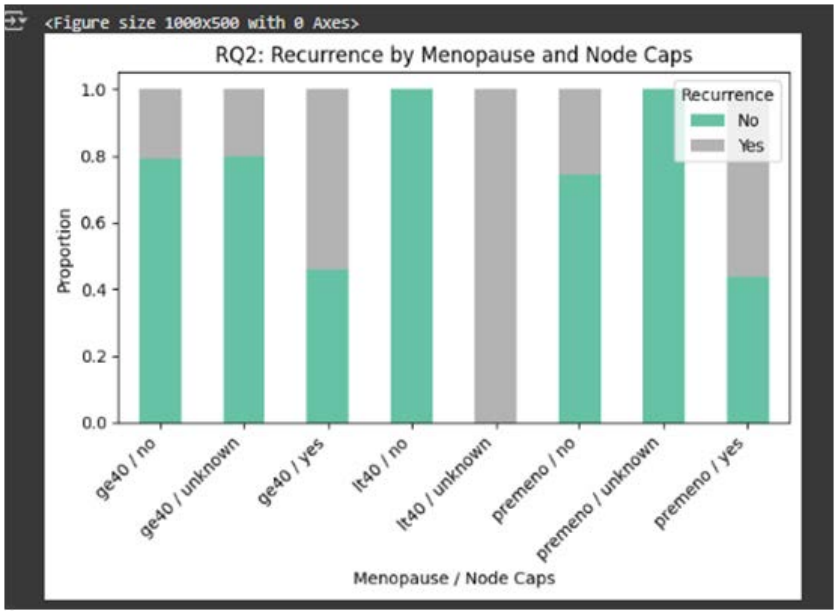

Figure 7: Recurrence Distribution by Menopause Status and Node Location, Displayed by a Stacked Bar Chart

As shown, premenopausal patients with positive node-caps are more likely to experience breast cancer recurrence than postmenopausal patients or those without lymph node involvement. This clearly illustrated in Figure 7, which visually highlights the differences in recurrence rate.

3.3.6. RQ3: Radiation Moderating Tumor × Node Effect

3.3.6.1. Heatmap of Interaction Means

• Those who received radiation have a mean network interaction of ~99.0, compared to ~41.1 in the no recurrence group.

• With radiation: patients in the recurrence group had a higher mean (~150.8), suggesting multiple influences factors.

It is not always the case that radiation therapy cuts down the risk of breast cancer coming back. How well chemotherapy works seems to depend on the seriousness of the cancer and the involvement of nearby lymph nodes.

Irradiate	No Recurrence (0)	Recurrence (1)
0	41.054878	99.018519
1	117.378378	150.774194

Table 3: Mean Interaction (Tumor Size × Lymph Nodes) by Radiation and Recurrence status

Among those who experienced recurrence after radiation, the average number of encounters with the system was greatest (150.77), as shown in Table 3. This means that, although radiation is given to more serious cases, it often doesn't fully address the increased likelihood of the cancer's return due to big tumors and many lymph node changes.

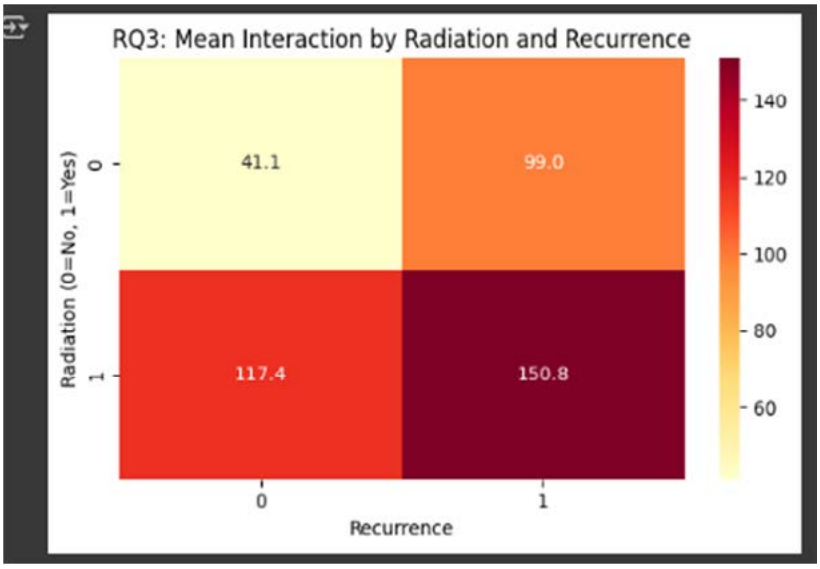

Figure 8: Heatmap of interaction means by Radiation and Recurrence

Radiation therapy may not always be effective in reducing cancers recurrence. It appears that patients with more severe tumor-node issues—as indicated by the highest interaction in the heatmap figure 8—were those received radiation after experiencing recurrence.

3.4. Summary Table of Key Results

Research Question	Method Used	Key Findings
RQ1: Tumor Size and Lymph Nodes vs Recurrence	Correlation Matrix	Lymph nodes involvement is a stronger predictor of breast cancer recurrence (0.28) than tumor size (0.18).
RQ2: Menopause and Node Caps vs Recurrence	Crosstab with proportions	Approximately half of all cases 56.3% in premenopausal women and 54.2% in ge40 women, were associated with node caps.
RQ3: Radiation Moderating Tumor × Node Effect	Interaction mean Analysis	Radiation therapy did not completely eliminate the risk of recurrence; even so, patients who experienced recurrence despite radiation showed the highest interaction effect (150.77).

Table 4: Summary of Finding Across Research Questions

Table 4 captures the primary points from my study. It shows that when cancer comes back, lymph node involvement matters more than the tumor's size. It also highlights that when a certain lymph node count is reached and woman is postmenopausal, the risk of recurrence is higher. In addition, radiation alone often doesn't prevent cancer form returning once the tumor has spread to the lymph nodes.

3.4.1. Correlation and Relation Analysis

In this section, we examine how various factors related to breast cancer contribute to understanding the risk of recurrence. We conducted an analysis of relationships between tumor size, lymph node involvement, and breast cancer recurrence.

3.4.2. Correlation Matrix

We assessed the relationships between tumor size, the number of lymph nodes involved, and the recurrence label using a Pearson correlation matrix. The strongest connection was found between lymph node involvement and recurrence, at a value of about 0.43. A higher number of affected lymph nodes indicates an increased likelihood of cancer recurrence. The correlation between tumor size and recurrence was weak patients. tumor size and lymph nodes involvement were moderately correlated at approximately 0.48, suggesting that larger tumors are more likely to have lymph node involvement.

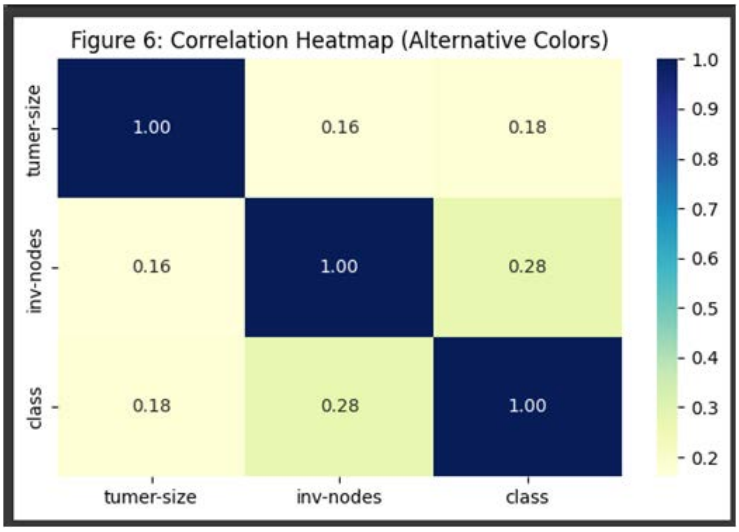

Figure 9: Correlation Heatmap.

Figure 9 shows a heatmap visualization that helps compare the significance of each variable in predicting recurrence.

3.4.3. Scatter Plot Tumor Size Vs Lymph Nodes

Next, we visualized these relationships using a scatter plot comparing tumor size and lymph node involvement, with each group color-coded based on whether tumor recurrence occurred. The pattern clearly shows that cancer recurrence cases largely cluster int he upper right region, where both tumor and lymph node involvement are high. this suggests that both a large tumor and spread to many lymph nodes are better predictors of disease return than either factor alone.

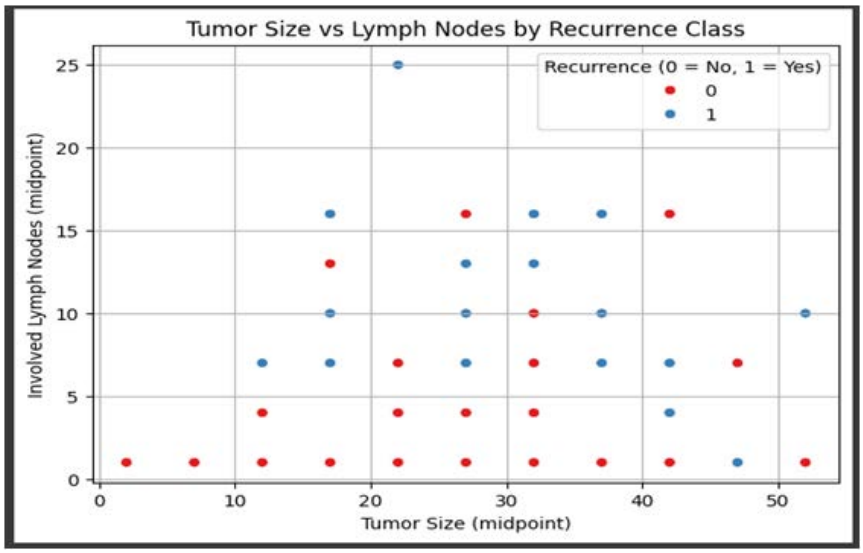

Figure 10: Tumor Size vs Lymph Nodes by Recurrence

The highlighted patients who experienced recurrence appear to gather closer to the larger tumor and higher node categories, as illustrated in Figure 10.

4. Discussion

4.1. Tumor Size and Lymph Node Involvement

According to the analysis, patients with affected lymph nodes were more likely to have recurrence than those whose tumors were large but lymph nodes were not. This is in line with clinical beliefs, as both a larger primary tumor and more involvement of lymph nodes usually suggest an aggressive breast cancer. According to a study of more than 12,000 patients, an increase in tumor size significantly boosts the odds of lymph node colonization. Specifically, this relationship was clearer among luminal A and B breast cancers than in either HER2-positive or triple-negative cancer types. Evaluating a patient's chance of cancer recurrence and appropriate treatment strategies should consider both tumor size and the presence of lymph node involvement simultaneously [10].

4.2. Menopause Status and Node Caps

Having menopause and lymph node involvement in the tumor seemed to play a role in whether the cancer returned. Women who have not gone through menopause often experienced tumor recurrence more frequently

because their high estrogen levels can cause the tumor to become more aggressive. In addition, node-caps indicate extracapsular growth which increases the risk of recurrence. Whether a woman has reached menopause is a key factor in predicting whether breast cancer will return in hormone receptor-positive types. Postmenopausal women with ER+ and PR+ breast cancer are 39% more likely to get metastatic disease when compared to premenopausal women, based on the latest data from the New Zealand Breast Cancer Register. Therefore, menopausal status should be treated as an active risk factor in models used to assess recurrence. Considering menopausal status along with other pathological findings helps improve the accuracy of risk is estimated and guides appropriate treatment decisions [5].

4.3. Impact of Radiation Therapy

Radiation did not always lower the risk of tumor recurrence. Patients who received radiation and still had a recurrence showed the highest tumor × node interaction values which suggests that such treatment is most common when the disease is advanced. Recent research suggests that PMRT is not reliable in lowering breast cancer patients' recurrence rates after micrometastases in lymph nodes. Even though there was a small improvement in living five years after treatment, overall survival and survival from breast cancer did not change. It seems that radiation is more commonly used in advanced cases, but may still fail to stop the cancer coming back in those at greatest risk [11].

4.4. Comparison with Other Studies

New data confirms that hormonal and disease factors affect the chances of breast cancer recurrence. Research suggests that women who have gone through menopause often show a different pattern of cancer risk because their hormone levels influence cancer growth. Both the findings from our current study and the results of suggest that menopause status and having nodes closely connected to the tumor are key indicators of a patient's risk of recurrence [12]. As a result, doctors can make better predictions about recurrence and choose more suitable treatment strategies by using detailed hormonal analyses together with clinical pathology [13].

4.5. Practical Contribution of this Study

My research applied a fully structured analysis to examine how factors like the size of the tumor, the state of the lymph nodes, menopause status, node caps and radiation therapy can impact the risk of breast cancer recurrence. I relied on easy-to-understand statistical methods techniques designed for small and organized data in especially in healthcare settings where resources are limited. I examined how each of these variables influenced recurrence using Pearson correlation, crosstab analysis and interaction effect analysis. It was easy to see relationships and trends because heatmaps, bar charts and scatter plots were used. The analysis showed that when lymph nodes were involved, it was more predictive of recurrence than tumor size and being premenopausal along with having node caps was the biggest risk of all. On the other hand, radiation therapy alone did not significantly lower the recurrence rate when tumors were large and lymph nodes were affected. This understanding helps guide early intervention strategies using information that's already available in clinics.

Feature/Aspect	Previous Studies	This Study
Modeling Approach	Complex ML/DL model (XGBoots, DNN, SHAP)	Simple ststistical analysis (correlation, crosstabs)
Data Type	Often includes genomic/imaging data	Structured clinical variables only
Interpretability	SHAP, LIME (requires technical knowledge)	Human-readable and intuitive interpretation
Variable Interaction Analysis	Rarely explored	Tumor × Node interaction+ radiation effect
Target Context	High-resource settings	Low-/middle-resource healthcare environments

Table 5: How this Study Differs from Previous Research

The table 5 shows the most important differences in approaches, types of data and interpretability between what was done in previous work and what is done here.

4.6. Distinct Contributions of the Current Study

Prior studies have supported the idea that machine learning and AI tools can predict whether breast cancer will recur, but this study chose a different approach. The study adopted use simple statistical methods that rely on structured clinical data from real healthcare settings. This real-world approach becomes especially important when resources for advanced artificial intelligence or molecular studies are limited. What is more, the authors examined how variables like tumor size and radiation therapy interact with each other, rather than evaluating them individually. By including an interaction variable for tumor and node, we were able to uncover why some patients experienced recurrence even with radiation. Since it prioritizes clarity and practical use in medicine, this research supports and extends current machine learning research by making effective decision-making tools accessible to underserved areas of the healthcare system.

4.7. Related Studies

Study	Dataset/Methods	Key Findings	Interpretability Approach
(Zuo et al., 2023)	Structured clinical features (tumor diameter, lymph node status) with AdaBoost	Combining clinical data with SHAP leads to accurate and interpretable predictions of disease recurrence risk.	SHAP
(Park et al., (2025)	LASSO regression for feature selection with ensemble learning	Tumor size and lymph node status were identified as key indicators; the interpretability of the model was enhanced.	SHAP
(P. Chetry et al., 2025)	Deep neural networks on structured clinical variables	High accuracy in predicting early recurrence, useful in settings with low genomic resources	Not specified
(González-Castro et al., 2023)	Structured and unstructured EHR data with XGBoost	An AUROC of 0.84 was achieved for predicting five-year recurrence, which is applicable in real-world clinical settings.	Not specified
(Min et al., 2021)	Analysis of 12,000+ cases, tumor size vs lymph node metastasis by subtype	The relationship between tumor size and lymph node metastasis varies by breast cancer type and is critical for risk stratification.	Not applicable
(Lao et al., 2021)	Menopausal status and pathological markers	Postmenopausal ER+/PR+ patients had a 39% higher risk of metastasis compared to premenopausal patients, and hormonal status was an important risk factor.	Not applicable
(Zheng et al., 2025)	Meta-analysis of post-mastectomy radiation therapy (PMRT) in micrometastatic cases	PMRT provides modest gains in disease-free survival but no significant improvement in overall survival.	Not applicable
(Ansari et al., 2024)	Review of SHAP, LIME, and Grad-CAM in oncologic prognosis models	He stressed the need to use joint translation techniques to build trust among doctors.	SHAP, LIME, Grad-CAM

Table 6: Summary of Previous Studies on Breast Cancer Recurrence Prediction and Analysis

Table 6. This section provides an overview of the major studies that have explored various methods for predicting breast cancer recurrence and related data analysis. It also highlights the types of data used in each study, the main findings, and the interpretation techniques employed.

4.8. Limitations

Despite these findings, this study was met with several challenges. Because there were no records of the length of procedures, treatments, or how long patients lived, probing the long-term consequences of those surgeries was not feasible. If longitudinal data is part of future studies, it will help us see how illness develops over time. Gathering data by using midpoints for both tumor and node ranges may have affected the outcome slightly. Since clinical treatments such as chemotherapy, and hormone therapies were not mentioned, it was not possible to analyze how different treatment approaches affect the outcomes. Since HER2, Ki-67, ER, and PR receptor data were lacking in this study, we could not identify many important biological patterns. It may have been that the tumor grade or some other unidentified aspect played a part in the outcomes of radiation therapy. Furthermore, since the node-caps and menopausal values in the data contained unknown information, it would be tricky to apply the findings to all groups. Adding a wider variety of patients from diverse regions and ages to the data, as well as merging clinical, genomic and image information, would make future research more accurate and useful.

4.9. Implications

The results indicate that early use of tumor size, lymph node count and menopause status can help identify patients at higher risk. These findings can guide clinicians in making better choices when gene data is lacking or prohibitively expensive. This approach supports the findings reported by which state that combining different types of data from electronic health records helps predict breast cancer recurrence [14]. Researchers could improve their understanding of recurrence by incorporating information on long-term treatment responses. The accuracy and practicality of such models could also be enhanced by including more types of biomarkers and expanding the

diversity of study participants [15-18].

5. Conclusion

This study provides important insights in to the recurrence patterns of breast cancer recur. By examining key clinical factors such as tumor size, lymph nodes involvement, menopause status, presence of node caps, and the use of radiation therapy, it highlights that valuable conclusions can be drawn even without relying on expensive genomic data. This research indicates that detecting disease in the lymph nodes is more important for predicting recurrence than the tumor size alone. Having positive node-caps, along with being premenopausal, places women in a very high-risk group. Even though radiation therapy is a standard treatment in advanced cases, it appears to have limited success in reducing the likelihood of cancer returning when the tumor has spread extensively. This work provides important information for clinical use. For providers in low-resource areas, it is easier to identify high-risk patients using simple structured data and to choose better treatments strategies. It supports the goal of achieving equitable healthcare that leverages data, as set outlined in SDG 3 (Good Health and Well-being). In the future, adding data on repeated measures and medication histories will help analyze how tumors come back over time. An improved prediction rate and better patient care could be achieved by including molecular and hormonal biomarkers.

References

1. WHO. (2024). WHO Global Breast Cancer Initiative: Breast cancer awareness month. *World Health Organization, 8*(October), 1–2.
2. Ghorbian, M., & Ghorbian, S. (2023). Usefulness of machine learning and deep learning approaches in screening and early detection of breast cancer. *Heliyon, 9*(12).
3. Xiques-Molina, W., Lozada-Martinez, I. D., Fiorillo-Moreno, O., Hernández-Lastra, A. L., & Bermúdez, V. (2025). Operational Advantages of Novel Strategies Supported by Portability and Artificial Intelligence for Breast Cancer Screening in Low-Resource Rural Areas: Opportunities to Address Health Inequities and Vulnerability. *Medicina, 61*(2), 242.
4. Zuo, D., Yang, L., Jin, Y., Qi, H., Liu, Y., & Ren, L. (2023). Machine learning-based models for the prediction of breast cancer recurrence risk. *BMC Medical Informatics and Decision Making, 23*(1), 276.
5. Lee, T. F., Shiau, J. P., Chen, C. H., Yun, W. P., Wuu, C. S., Huang, Y. J., ... & Chao, P. J. (2025). A Machine Learning Model for Predicting Breast Cancer Recurrence and Supporting Personalized Treatment Decisions Through Comprehensive Feature Selection and Explainable Ensemble Learning. *Cancer Management and Research, 917*-932.
6. Chetry, M., Feng, R., Babar, S., Sun, H., Zafar, I., Mohany, M., ... & Khan, S. (2025). Early detection and analysis of accurate breast cancer for improved diagnosis using deep supervised learning for enhanced patient outcomes. *PeerJ Computer Science, 11,* e2784.
7. Golijanin, D., Radovanović, Z., Radovanović, D., Đermanović, A., Starčević, S., & Đermanović, M. (2024). Molecular subtype and risk of local recurrence after nipple sparing mastectomy for breast cancer. *Oncology Letters, 28*(2), 389.
8. Kayali, M., Abi Jaoude, J., Tfayli, A., El Saghir, N., Poortmans, P., & Zeidan, Y. H. (2020). Post-mastectomy radiation therapy in breast cancer patients with 1–3 positive lymph nodes: no one size fits all. *Critical Reviews in Oncology/Hematology, 147,* 102880.
9. Shirode Nitesh, G., & Jadhav Sejal, R. (2021). Breast cancer, factors influencing of it and management of breast cancer. *World J Pharm Res, 10*, 507-22.
10. Min, S. K., Lee, S. K., Woo, J., Jung, S. M., Ryu, J. M., Yu, J., ... & Nam, S. J. (2021). Relation between tumor size and lymph node metastasis according to subtypes of breast cancer. *Journal of Breast Cancer, 24*(1), 75.
11. Zheng, J., Huang, B., Chen, Y., & Chen, Z. (2025). Effect of post-mastectomy radiation therapy on survival in breast cancer with lymph nodes micrometastases: a meta-analysis and systematic review. *Frontiers in Oncology, 15,* 1489390.
12. Natarajan, R., Krishna, S., Gururaj, H. L., Flammini, F., Alfurhood, B. S., & Kumar, C. N. (2025). A novel hybrid dynamic harris hawks optimized gated recurrent unit approach for breast cancer prediction. *International Journal of Computational Intelligence Systems, 18*(1), 7.
13. Park, W. K., Nam, S. J., Kim, S. W., Lee, J. E., Yu, J., Lee, S. K., ... & Chae, B. J. (2024). The prognostic impact of HER2-low and menopausal status in triple-negative breast cancer. *Cancers, 16*(14), 2566.
14. Gonzalez-Castro, L., Chávez, M., Duflot, P., Bleret, V., Martin, A. G., Zobel, M., ... & López-Nores, M. (2023). Machine learning algorithms to predict breast cancer recurrence using structured and unstructured sources from electronic health records. *Cancers, 15*(10), 2741.
15. Ali Ansari, Z., Madhava Tripathi, M., & Ahmed, R. (2024). Quantifying breast cancer: radiomics, machine learning, and dimensionality reduction for enhanced image-based diagnosis. *International Journal of Computing and Digital Systems, 16*(1), 1535-1552.
16. Bate, J., & Rasmussen, E. (2009). Key Facts. *The Winter's Tale*, 19–19.
17. Chetry, P., Kumar, V., & Sharma, S. S. (2025). Influence of gibberellic acid on aluminium-induced suppression of seed germination and early root growth in certain rice (Oryza sativa L.) landraces from Sikkim Himalaya. *The Nucleus*, 1-13.
18. Lao, C., Elwood, M., Kuper-Hommel, M., Campbell, I., & Lawrenson, R. (2021). Impact of menopausal status on risk of metastatic recurrence of breast cancer. *Menopause, 28*(10), 1085-1092.