Arab American University

Faculty of Graduate Studies

The Role of Mammography and Innovative Health Informatics Utilization on Breast Cancer Detection in Palestine

By

Abdurrahman Fayez Abdurrahman Juma

Supervisor

Dr. Yousef Mimi

Co- Supervisor

Prof. Mohammed Awad

This thesis was submitted in partial fulfillment of the requirements for the Master's degree in Health Informatics

July/2024

Thesis Approval

The Role of Mammography and Innovative Health Informatics Utilization on Breast Cancer Detection in Palestine

By

Abdurrahman Fayez Abdurrahman Juma

This thesis was defended successfully on 04/07/2024 and approved by:

Committee members	Signature
1. Dr. Yousef Mimi: Supervisor	Y.Mimi
2. Prof. Mohammed Awad: Co-Supervisor	M.Awad
3. Dr. Ahmad Ewais: Internal Examiner	A.Ewais
4. Dr. Hala Al Labadi: External Examiner	H.Labadi

Declaration

I, Abdurrahman Fayez Abdurrahman Juma, student number 201812616, solemnly declare that the master's thesis titled "The Role of Mammography and Innovative Health Informatics Utilization on Breast Cancer Detection in Palestine", submitted to the Arab American University's College of Graduate Studies for the master’s degree in health informatics, is my own original work.

This thesis has been completed during my period of study at the university and has not been submitted for any other degree or qualification. In it, I have clearly stated any portions previously submitted, duly attributed all consulted sources, acknowledged all quotations from others' works, and clearly distinguished between my contributions and those made in collaboration.

I affirm that this work has not been published prior to its submission and understand that any breach of these declarations could lead to the withdrawal of my degree. This thesis is supervised by Dr. Yousef Mimi and co-advised by Prof. Mohammad Awad.

The Name of The Student: Abdurrahman Fayez Abdurrahman Juma

ID: 201812616

Signature: *Abdurrahman Juma*

Date: 2024-10-09

Acknowledgment

I would like to extend my heartfelt gratitude to my supervisors, Dr. Yousef Mimi, and Prof. Mohammed Awad, for their invaluable guidance, insightful critiques, and unwavering support throughout the course of this research. Their expertise and mentorship have been paramount in shaping not only this thesis but also my growth as a researcher.

We sincerely thank Mr. Amjad Atallah, IT Administrator at the Ministry of Health, for facilitating access to the critical mammogram data, and Dr. Huda Haddad, Head of Breast Cancer at the Ministry of Health, for her assistance in providing essential breast cancer-related attributes. Their respective contributions have been pivotal to the success of our study.

I am also profoundly grateful to Mr. Mohammad Bani Odeh, Director at the World Health Organization, for his significant contributions in advancing multiple research topics. His advice in identifying the scope and navigating the current literature in Palestine has been incredibly beneficial. His perspective and support have been truly enlightening and have helped to ensure that this research is as comprehensive and impactful as possible.

To all, I offer my sincere appreciation for their contributions to this research.

Abstract

Background: Breast cancer is the most diagnosed cancer among women worldwide, presenting significant public health challenges. The importance of early detection in improving survival rates is well-recognized, facilitated by advancements in diagnostic technologies such as mammography. This thesis explores the efficacy of early detection strategies in the Palestinian context, in collaboration with the Ministry of Health, using data from two significant datasets: "Mammogram" and "Breast Cancer".

Methods: A predominantly quantitative approach, complemented by qualitative insights, is used, with secondary data analyzed through a retrospective cohort design and observational epidemiological model. The research involves all the women in the West Bank who underwent mammography screenings at the Ministry of Health, using a non-probability census sampling method to examine 30,000 records over four years.

Results: This study finds that the governorate of residence, first pregnancy age and heart diseases are strong predictor of BIRAD scores. It shows an 83.23% detection completeness for high BIRADS cases, indicating effective alignment between mammogram and cancer registries. Regional disparities are evident, with Jenin having the highest case concentration. While most high BIRAD cases are migrated to the breast cancer registry within one month, some experience delays of up to several years.

Conclusion and Recommendation: This thesis highlights the vital role of health informatics in early detection and enhancing breast cancer outcomes. The recommendations focus on upgrading data systems, better training for healthcare providers, and advocating for stronger policies to enhance breast cancer management and patient care, urging collective action from all stakeholders in Palestine's healthcare system.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ABVS: Automated Breast Volumetric Scanning

AI: Artificial Intelligence

API : Application Programming Interface

BCSC: Breast Cancer Surveillance Consortium

BIRADs: Breast Imaging-Reporting and Data System

BMI: Body Mass Index

BPN: Back-Propagation Neural Network

BSE: Breast Self-Examination

CAD: Computer-Aided Diagnostic

CBE: Clinical Breast Examination

DB: Data Base

EHRs : Electronic Health Records

FDA: Food and Drug Administration

GA: Genetic Algorithm

GLCM: Gray-Level Co-occurrence Matrix

HBM: Health Belief Model

HER: Electronic Health Record

HHUS: Hand-Held Bilateral Ultrasound

HOG: Histogram of Oriented Gradients

HRT: Hormonal Replacement Therapy

KNCSP: Korea National Cancer Screening Program

MOH : Ministry of Health

NCCN : National Comprehensive Cancer Network

OPT: Occupied Palestinian Territory

PHP: Hypertext Preprocessor (Programming Language)

PNIPH: Palestinian National Institute of Public Health

Chapter 1: Introduction

1.1 Background

Mammography stands as a crucial screening method for early breast cancer detection, effectively revealing potential abnormalities. By identifying suspicious breast images, mammography aids in reducing unnecessary biopsies and additional tests, enhances the detection of multiple breast tumors, and increases the accuracy in determining the characteristics of breast abnormalities (Bancej et al., 2003). Effective mammography programs are essential for diagnosing breast cancer at an early stage, thereby reducing mortality rates. National efforts focus on raising awareness to encourage women's participation, leading to the early identification of at-risk cases and their timely transition to treatment. Breast cancer is the most prevalent cancer among Palestinian women, accounting for 32% of all cancer diagnoses in the West Bank and 18% in the Gaza Strip. However, over 60% of breast cancer cases in Palestine are diagnosed at a late stage due to lack of awareness and limited access to screening services (Johns Hopkins Medicine, n.d.; GlobalGiving, n.d.; United Nations Population Fund, 2020; Anera, 2010).

Health informatics plays a critical role in improving breast cancer detection and treatment outcomes. By integrating vast amounts of health data, including mammography results, patient histories, and genetic information, health informatics enables more personalized and accurate screening protocols. This integration facilitates the seamless sharing of information among healthcare providers, improving coordination and continuity of care. Advanced data analytics and machine learning algorithms can enhance the interpretation of mammograms, reducing false positives and negatives, thereby improving the overall accuracy of breast cancer detection (Obermeyer & Emanuel, 2016; Shadbolt et al., 2021; Uslu & Stausberg, 2021).

This study aims to explore the role of mammography in breast cancer detection by analyzing data from the Ministry of Health's mammography screening electronic registry (MOH mammogram registry). It will examine the extent to which mammography risk stages correlate with actual breast cancer cases as recorded in the Real Breast Cancer Registry.

1.2 Problem Statement

In Palestine, the integration between mammogram screenings and the Real Breast Cancer Registry (RBCR) is currently insufficiently coordinated. Patients undergo mammography screenings and receive results later, which may prompt a subsequent breast cancer screening based on those results (United Nations Population Fund, 2018). The lack of real-time synchronization between these two registries hinders the efficient alignment, completeness, and accuracy of case migration.

1.3 Significance of the Study

This study explores mammography's role in breast cancer detection. It aims to establish the reliability of mammograms as an early screening tool for breast cancer, applicable to both symptomatic and asymptomatic women. By determining the effectiveness of mammography screening at various stages, the research will contribute to enhancing early breast cancer detection. This will enable the recommendation of suitable follow-up procedures for cases identified as considerable risk, utilizing a robust informatics framework.

Furthermore, the availability of a comprehensive mammogram dataset allows for an in-depth analysis of attendee demographics, including factors such as educational background, age, and medical history. This analysis is crucial for understanding the influence of these factors on breast cancer detection and screening efficacy, thereby providing valuable insights into personalized and community-level healthcare strategies.

1.4 Study Objectives

The primary objective of this thesis is to leverage health informatics to automate and enhance the integration of mammogram and cancer records, improving data quality and linkage for early breast cancer detection.

The specific objectives are:

Evaluate the Integration of High-Risk Mammogram Cases: Assess how effectively high-risk mammogram cases are integrated into the breast cancer registry and determine the alignment between these registries in capturing cases associated with high-risk mammogram results.

Analyze Timeliness of Data Transfer: Investigate the timeliness of transferring high BIRAD scores from mammogram records to the breast cancer registry and identify factors influencing this transfer process.

Examine Relationships Between Mammogram Attributes and Breast Cancer Evidence: Explore the relationship between specific mammogram attributes and breast cancer evidence, and analyze the geographical distribution of these cases across West Bank governorates.

1.5 Research Questions

1- What is the relationship between specific mammogram attributes and breast cancer evidence, and how does the geographical distribution of breast cancer cases vary across West Bank governorates?

2- How effectively are high-risk mammogram cases integrated into the breast cancer registry, and to what extent do these registries align in capturing breast cancer cases associated with high-risk mammogram results?

3- What is the timeliness of transferring high BIRAD scores from mammogram records to the breast cancer registry, and what factors contribute to variations in the duration of this process?

1.6 Overview

Chapter one introduces the study's background, problem statement, significance, objectives, and research questions.

Chapter two reviews relevant literature, identifies gaps, discusses current research implications, and concludes with key findings.

Chapter three outlines the conceptual framework, defining key concepts, study variables, and operational definitions.

Chapter four describes the materials and methods, including study design, settings, population, sampling, data collection, integration, preparation, feature engineering, ethical considerations, and performance metrics.

Chapter five presents the study results, analyzing mammogram attribute weights, correlations between findings, registry integration, timeliness of data migration, and regional case distribution.

Chapter six discusses the findings, emphasizing mammogram features, data linkage completeness, timeliness, regional case distribution, mammogram performance, and concludes with recommendations, potential obstacles, points of strength, and future research directions.

Chapter 2: Literature Review

2.1 Introduction

Breast cancer remains a significant global health challenge, ranking as the most common cancer in women and the second most prevalent cancer worldwide, with over two million new cases in 2018 (Bancej et al., 2003). The cornerstone of breast cancer early detection is mammography, a screening tool that has been instrumental in identifying cancerous changes before clinical symptoms arise, thereby contributing to a reduction in breast cancer mortality (WHO, The Palestinian National Institute of Public Health, 2014). Mammography screenings utilize X-ray imaging to detect abnormal tumours in the breast. The procedure involves compressing the breast between two plates to obtain clear images, which are then classified using the Breast Imaging-Reporting and Data System (BI-RADS) (National Cancer Institute, 2018).

This system, developed by the American College of Radiology, provides a standardized reporting schema for breast imaging, categorizing results into seven assessment categories. Women with BI-RADS scores above 3 typically undergo further diagnostic procedures, including biopsies, to confirm the presence of breast cancer. The results from these screenings and subsequent tests are vital for devising appropriate treatment plans and are recorded in databases like the Real Breast Cancer Registry (RBCR), which collects comprehensive patient data.

This literature review delves into the global and Palestinian-specific landscape of mammography screening and health informatics in breast cancer prediction. Studies highlighted in the review encompass various aspects, such as the efficacy of mammography in early-stage cancer detection, comparison with other imaging techniques, and the impact of health informatics tools like AI in improving diagnosis accuracy and patient management. The review also addresses the challenges and limitations of mammography, including issues like overdiagnosis, false positives, and the effectiveness of screening programs in different geographical and economic settings. Notably, the review identifies a significant gap in the integration between mammogram registries and cancer registries, particularly in the context of Palestine. This lack of integration impacts the timely and accurate migration of high-risk cases from mammogram detection to breast cancer treatment, underscoring the need for innovative informatics solutions to bridge this gap.

The review concludes with the recognition that while mammography is the most effective method for early breast cancer detection, broader societal, political, and economic factors influence the accessibility and effectiveness of screening programs. The current study aims to investigate the extent to which the mammogram registry aligns with the cancer registry in Palestine and explore how informatics can enhance the screening process and patient outreach, improving early breast cancer detection and treatment. This introduction encapsulates the breadth of the literature review, setting a solid foundation for understanding the significance of the study in the context of existing research and identifying areas for further investigation.

2.2 Related Work

This literature review explores the effectiveness of mammography and the application of health informatics in breast cancer prediction, highlighting both the benefits and limitations of various studies:

2.2.1 International Studies

A study by the U.S. Food and Drug Administration on how well mammograms work found that they are very good at finding tumors up to two years before they can be felt by touch. This means mammograms can detect cancer early when it is easier to treat. The study showed that mammograms can detect 85-90% of cancer cases. However, it did not talk about the issue of overdiagnosis, which happens when non-threatening cancers are found and treated unnecessarily, causing potential harm (American Cancer Society, 2019). In a large-scale study at the University of North Carolina, researchers looked at cancer detection and recall rates in over 117,000 women. They found many false positives, meaning many women were told they might have cancer when they actually did not. This was especially common among women aged 35-39. These false positives can cause a lot of anxiety and lead to unnecessary tests and treatments, showing that more accurate and reliable screening methods are needed (Yankaskas et al., 2010).

A comprehensive study in the U.S. from 1976 to 2008 showed that mammography significantly increased the early detection of breast cancer, which is beneficial because early detection can lead to better treatment outcomes. However, the study also found an increase in overdiagnosis, meaning some cancers that were detected would not have caused problems if left untreated. This raises questions about the overall benefits of mammography in reducing breast cancer deaths (Bleyer & Welch, 2012). A review conducted in Canada gathered health utility values for breast cancer, which measure the quality and length of life-related to different treatments. The review found a lot of variability in these values, especially with systemic treatments like chemotherapy. This makes it difficult to establish consistent standards for evaluating the effectiveness and impact of different treatments (Kaur et al., 2022).

A study in São Paulo, Brazil, focused on identifying key indicators to improve breast cancer reporting in developing countries. The study emphasized important factors that can enhance the accuracy and efficiency of breast cancer data collection and reporting. However, it did not consider how these findings would apply in different economic and social settings, which can vary widely (Vieira et al., 2017). The nationwide screening program in the Netherlands, which started in the 1990s, achieved high participation and was effective at detecting cancer early. The study supported the program's success, showing that many women attended screenings and cancers were found at an earlier, more treatable stage. However, it did not address issues of overdiagnosis or how well the program worked for different demographic groups, which could affect its overall effectiveness (Van Luijt et al., 2013). A study from Northwestern Germany found that mammography screening reduced the number of advanced breast cancer cases in women aged 55 and older. This means fewer women in this age group were diagnosed with late-stage breast cancer, which is harder to treat. However, the study did not see the same results in younger women, suggesting that screening strategies might need to be tailored to different age groups to be more effective (Simbrich et al., 2016).

A study of the Korean National Cancer Screening Program showed that women who did not get screened had more cases of late-stage breast cancer. This highlights the importance of early detection through mammography, as catching cancer early can lead to better outcomes. However, the study also pointed out concerns about overdiagnosis, where cancers that might not have caused problems are treated unnecessarily (Choi et al., 2018). Research by the Indonesian Medical Association compared mammography with other methods for detecting breast cancer. The study highlighted the benefits of mammography in finding cancer early but also noted challenges like overdiagnosis and false positives. These false positives can lead to unnecessary worry and medical procedures. The study also identified significant barriers in low-income countries, such as limited infrastructure and a lack of awareness about breast cancer screening (Soekersi et al., 2022).

An Indian study used artificial intelligence (AI) and computer vision to distinguish between benign (non-cancerous) and malignant (cancerous) tumors in breast cancer detection. The AI was very accurate, showing that it can be a powerful tool for improving diagnosis. However, the study also highlighted ongoing challenges in accurately identifying different types of tumors, which is crucial for appropriate treatment (Khan et al., 2023). Research on Open AI's

ChatGPT examined how this AI tool can help doctors make decisions, especially in cancer screening and medical documentation. The study highlighted the benefits of using AI to assist with these tasks, such as improving efficiency and accuracy. However, it also pointed out potential risks, like over-reliance on technology and possible errors in AI-generated recommendations (Liu et al., 2023).

An observational study compared AI algorithms with the Breast Cancer Surveillance Consortium (BCSC) risk model to see which was better at predicting breast cancer risk. The study found that combining AI with the BCSC model improved predictive accuracy. This suggests that AI could play a bigger role in assessing breast cancer risk in the future, but it should be used alongside existing clinical models to ensure balanced and accurate predictions (Arasu et al., 2023). Research on short-term follow-ups for BI-RADS 3 mammogram findings suggested that this approach could be a good alternative to immediate biopsies. BI-RADS 3 findings are likely to be benign (not cancerous), so short-term follow-ups can help monitor these cases without the need for immediate invasive procedures, reducing patient anxiety and medical costs (Berg et al., 2020).

A study introduced a Computer-Aided Diagnosis (CAD) system for categorizing mammographic masses, which achieved high accuracy in classification. This represents a significant advancement in using technology to help radiologists assess the likelihood of malignancy and improve diagnostic accuracy, ultimately aiding in better patient management (Boumaraf et al., 2020). A study looked at how using electronic health records (EHR) in decision-making for mammography screening can improve patient satisfaction. It found that clear, data-driven communication between doctors and patients greatly improved satisfaction levels. This research emphasizes the importance of effective communication in healthcare, particularly when discussing screening results and treatment options (Liu et al., 2022).

A comparative study examined two breast imaging techniques: automated breast volumetric scanning (ABVS) and hand-held bilateral whole breast ultrasound (HHUS). The findings suggested that while ABVS can lead to more tests and potential overdiagnosis, it is important to balance these technological advances with patient comfort and diagnostic accuracy. This means ensuring that new technologies not only improve detection rates but also consider the patient's experience and the risk of unnecessary treatments (Tutar et al., 2020).

2.2.2 Regional Studies

An Iranian study explored the potential of AI in breast cancer detection by applying machine learning techniques to mammography and demographic data. The research demonstrated that random forests, a type of AI model, were effective in predicting breast cancer. However, it also emphasized the necessity for human oversight in AI-based diagnostics to ensure accuracy and reliability (Rabiei et al., 2022). A study conducted in Oman assessed the performance of mammography by looking at its positive predictive value (PPV), particularly focusing on BI-RADS 5 categorizations, which are highly suspicious for cancer. The findings revealed a strong correlation between BI-RADS 5 and actual cancer diagnoses, supporting the reliability of the BI-RADS system in predicting breast cancer. This research highlights the effectiveness of using BI-RADS 5 as a crucial indicator in mammography for accurate cancer prediction (Taif et al., 2014).

A study focused on breast cancer screening among Middle Eastern women used the health belief model to assess awareness levels and identify barriers to screening. The research highlighted significant variations in awareness and emphasized the need for comprehensive educational programs that are tailored to overcome cultural and psychological obstacles. These findings underscore the importance of addressing specific cultural contexts to improve participation in breast cancer screening and health outcomes (Bahri et al., 2022).

2.2.3 Palestinian Studies

A study conducted by A Najah University investigated mammography participation among Palestinian women and found that screening rates were low despite high levels of awareness about breast cancer. The study revealed that psychological barriers, such as fear of the results and anxiety about the screening process, and cultural barriers, such as societal norms and stigmas associated with breast cancer, significantly hindered women from participating in mammography screening. These findings highlight the need for targeted interventions, such as culturally sensitive education programs and psychological support services, to address these barriers and encourage more women to undergo screening (Hamshari et al., 2021).

A review of mammographic screening in the occupied Palestinian territory highlighted a critical gap in providing balanced information about the risks and benefits of screening. The study found that many women were not fully informed about what mammography entails, leading to misconceptions about the procedure and its importance. This lack of information contributed to a reluctance to participate in screening programs. The study emphasized the need for developing culturally sensitive educational campaigns that clearly communicate the benefits and potential risks of mammography. Additionally, it stressed the importance of clear and accurate communication from healthcare providers to ensure women can make informed decisions about their health (AlWaheidi et al., 2020).

A study by the United Nations Population Fund (UNFPA) mapped breast cancer care in Palestine, highlighting the diverse range of healthcare services available and the government's initiatives for early detection. The research identified several systemic challenges, such as limited medical resources, inadequate healthcare infrastructure, and a lack of trained medical personnel, which impede the effective delivery of breast cancer care. The study underscored the importance of developing integrated healthcare services that can address these challenges by improving resource allocation, enhancing healthcare infrastructure, and providing better training for medical staff. Such improvements are crucial for increasing the effectiveness of early detection and treatment programs for breast cancer patients (Jubran et al., 2018).

Research by the Palestinian National Institute of Public Health (PNIPH) and the World Health Organization (WHO) examined mammography screening practices in Palestine. The study recommended several improvements, particularly in the review of medical records and communication among healthcare professionals, to ensure accurate cancer staging. Accurate staging is essential for determining the most appropriate treatment plan for patients. The research highlighted the need for comprehensive data management systems that can track patient information accurately and facilitate effective communication among healthcare providers. These strategies are critical for enhancing the accuracy and efficiency of cancer diagnosis and treatment in Palestine (WHO, PNIPH, 2014).

A cross-sectional study conducted in Nablus, West Bank, examined breast cancer screening practices and found low engagement among women. The study identified significant barriers to participation, including psychological factors such as fear of the screening process and potential diagnosis, financial constraints that make screening unaffordable for many women, and cultural factors such as societal stigmas and misinformation about breast cancer. These findings highlight the complex interplay of psychological, financial, and cultural elements that influence women's decisions to participate in breast cancer screening programs. The study underscores the need for targeted interventions, such as financial assistance programs, community education initiatives, and culturally sensitive counseling services, to address these challenges and encourage more women to undergo screening (Al-Tell et al., 2019).

Each of these studies contributes valuable insights into the effectiveness, challenges, and future directions of mammography and health informatics in breast cancer detection, providing a nuanced understanding of the field.

2.3 Literature Gap

The limited knowledge of Palestinian women about the importance of mammography influences their attitude towards attending mammography examinations; this will be enhanced in the study by utilizing innovative informatic follow-up techniques. Furthermore, the mammogram registry and cancer registry are loosely separated registries. They have no integrative flow (i.e., detecting high BIRADs cases in mammogram registry are not migrated or communicated to cancer registry). In this study comparing the presence of a high BI-RAD case in the mammogram registry with its presence in the breast cancer registry in a timely flow is the study's major objective, i.e., check the completeness of a case in both registries (mammogram and cancer) and finally find the quantitative contribution of mammography screening on early breast cancer detection.

Chapter 3: Conceptual and Operational Frameworks

3.1 Introduction

This research aims to critically evaluate breast cancer detection methods in Palestine, focusing primarily on mammography as a screening tool. It investigates the relationship between mammogram attributes and BI-RADS scores, alongside variables like patient demographics, reproductive history, lifestyle factors, and chronic diseases. The study also examines geographical trends in breast cancer incidence and scrutinizes the integration of mammogram and breast cancer registry data. Furthermore, it assesses technological and human factors affecting mammogram accuracy. Ultimately, the study seeks to enhance breast cancer screening and healthcare policies in Palestine, improving detection rates and patient outcomes.

3.2 Conceptual Definition

This research constitutes an in-depth analysis of breast cancer detection methodologies within the Palestinian healthcare context, with a specific focus on the efficacy and application of mammography as a primary screening tool. The study is anchored in the evaluation of mammogram attributes and their correlation with Breast Imaging Reporting and Data System (BI-RADS) scores, which serve as indicators for the likelihood of malignant breast lesions. Key variables under examination include patient age, body mass index (BMI), age at menarche and menopause, age during first pregnancy, contraceptive use, hormonal replacement therapy, smoking habits, and the presence of chronic conditions such as diabetes or hypertension.

A significant aspect of the research involves geographical analysis, assessing breast cancer distribution and incidence rates across various governorates in the West Bank. This analysis provides insights into regional disparities, potential environmental influences, and the accessibility of healthcare services. The study also scrutinizes the integration and data completeness between mammogram screenings and the breast cancer registry, emphasizing the importance of timely and accurate data transfer for high-risk cases. Technological and human factors in mammogram performance are critically assessed, with an emphasis on the quality of mammography equipment and the expertise of radiologists in interpreting mammogram images. This encompasses an evaluation of the accuracy and reliability of mammogram results and their implications for breast cancer detection and subsequent care. By synthesizing these elements, the study aims to comprehensively understand the current state of breast cancer screening in Palestine. It seeks to identify critical areas for improvement in screening practices, data management, and healthcare policy breast cancer detection rates, patient outcomes, and overall public health strategy.

Following is a highlight of some conceptual definitions used in the study:

Breast Cancer Registry: “repository for socio-demographic, environmental, clinical history, family history, and biospecimen data collected at the National Comprehensive Cancer Network (NCCN) centres for participants with a personal history of breast cancer and/or characteristics of hereditary breast cancer.” (Fred & Pamela Buffett Cancer Centre at the University of Nebraska)

Mammogram: “A mammogram is an X-ray image of your breasts used to screen for breast cancer. Mammograms play a key role in early breast cancer detection and help decrease breast cancer deaths.” (Mayo Clinic)

BI–RADS: “stands for Breast Imaging Reporting and Data System and was established by the American College of Radiology.BI-RADS is a system that was developed by radiologists to report mammogram results using a common language. The radiologist assigns a single digit BI-RADS score (ranging from 0 to 5) when the report of your mammogram is created”. (DIS Diagnostic Imaging Services)

3.3 Study Variables

- Mammogram Features in Defining High BI-RADS: This includes various attributes such as age, menarche, medical history, breast density, BMI, and their predictive power for high BI-RADS values indicative of malignant breast lesions.
- Geographical Variability: The study considers the governorate or regional factor and its influence on BI-RAD scores, affected by environmental factors, healthcare access, and regional health practices.
- Body Mass Index (BMI): The relationship between BMI and BI-RADS scores is analyzed, although it is indicated that the correlation is very weak.
- Menarche Age: The study investigates the impact of the age at which menstruation begins on BI-RAD scores in breast imaging.
- First Pregnancy Age: This variable considers how the age during the first pregnancy might influence BI-RAD scores.
- Contraceptive Use: The study examines the relationship between contraceptive use and BI-RAD scores.
- Menopause Age: The study looks at how menopause age correlates with BI-RAD scores in breast imaging.
- Hormonal Replacement Therapy (HRT): The impact of HRT on breast tissue density and the potential influence on the risk of breast diseases and BI-RAD scores.
- Smoking Status: This variable examines the association between smoking and BI-RAD scores.
- Presence of Chronic Diseases: The study assesses the impact of chronic diseases like hypertension or diabetes on BI-RAD scores.

- Heart Diseases, Hypertension, and Diabetes: These are considered as individual factors influencing BI-RAD scores.
- Family History of Breast Cancer: The influence of family history on BI-RAD scores is studied.
- Breast Density: This variable looks at the direct impact of breast tissue density on the visibility of lesions in mammography and its relation to BI-RAD scores.
- Completeness between Mammogram and Breast Cancer Registries: The study evaluates the percentage of high-risk BI-RAD cases that are recorded in the breast cancer registry.
- Timeliness between High BI-RAD Mammogram Detection and Breast Cancer Evidence: This includes the analysis of the time difference between high BI-RAD values and breast cancer registry attendance.
- Distribution of Breast Cancer Cases among West Bank Governorates: The study investigates the geographical distribution of breast cancer cases and their implications.
- Mammogram Performance in Breast Cancer Early Detection across Governorates: This variable analyzes mammograms' effectiveness, uncertainty, and failure rates in predicting breast cancer across different regions.

3.4 Operational Definition

- Mammogram Attributes**:** This includes variables such as age at menarche, medical history, breast density, and Body Mass Index (BMI). These attributes are considered to assess their weight or significance in predicting breast cancer risk.
- BI-RADS Values: The Breast Imaging Reporting and Data System (BI-RADS) values are crucial variables. They categorize the results of mammography screening and indicate the likelihood of malignant breast lesions.

- Correlation Between Mammogram Attributes and BI-RADS Values: This involves examining the relationship between the mammogram attributes and the BI-RADS values.
- Timeliness and Completeness of High BI-RADS Values and Breast Cancer Incidence: This variable focuses on the promptness and completeness in recording high BI-RADS values and their correlation with actual breast cancer cases.
- Statistical Models and Analytical Methods: These are utilized to quantify relationships between mammogram attributes, BI-RADS values, and breast cancer incidence. This includes logistic regression, correlation analysis, and various statistical tests like Chi-square and ANOVA.
- Chronic Diseases, Diabetes, Heart Diseases, Hypertension, Contraceptive Use, Breast Density, First Pregnancy Age, Menopause Age, Body Mass, Personal History of Ovarian Cancer, Family History of Breast Cancer, Smoking Habits, and Menarche Age: These are considered as individual variables to assess their correlation with BI-RAD scores.
- Integration between Mammogram and Breast Cancer Registries: This involves measuring the completeness and integrity of data between these two registries, specifically for cases with high-risk BI-RAD scores.
- Geographical Variance: Examining the distribution of BI-RAD categories across different directorates or regions.
- Age at Diagnosis: This variable refers to the age range of patients at the time of breast cancer diagnosis.
- Breast Density Categories in Breast Cancer Patients: Analysing how different breast density categories are represented among breast cancer patients.

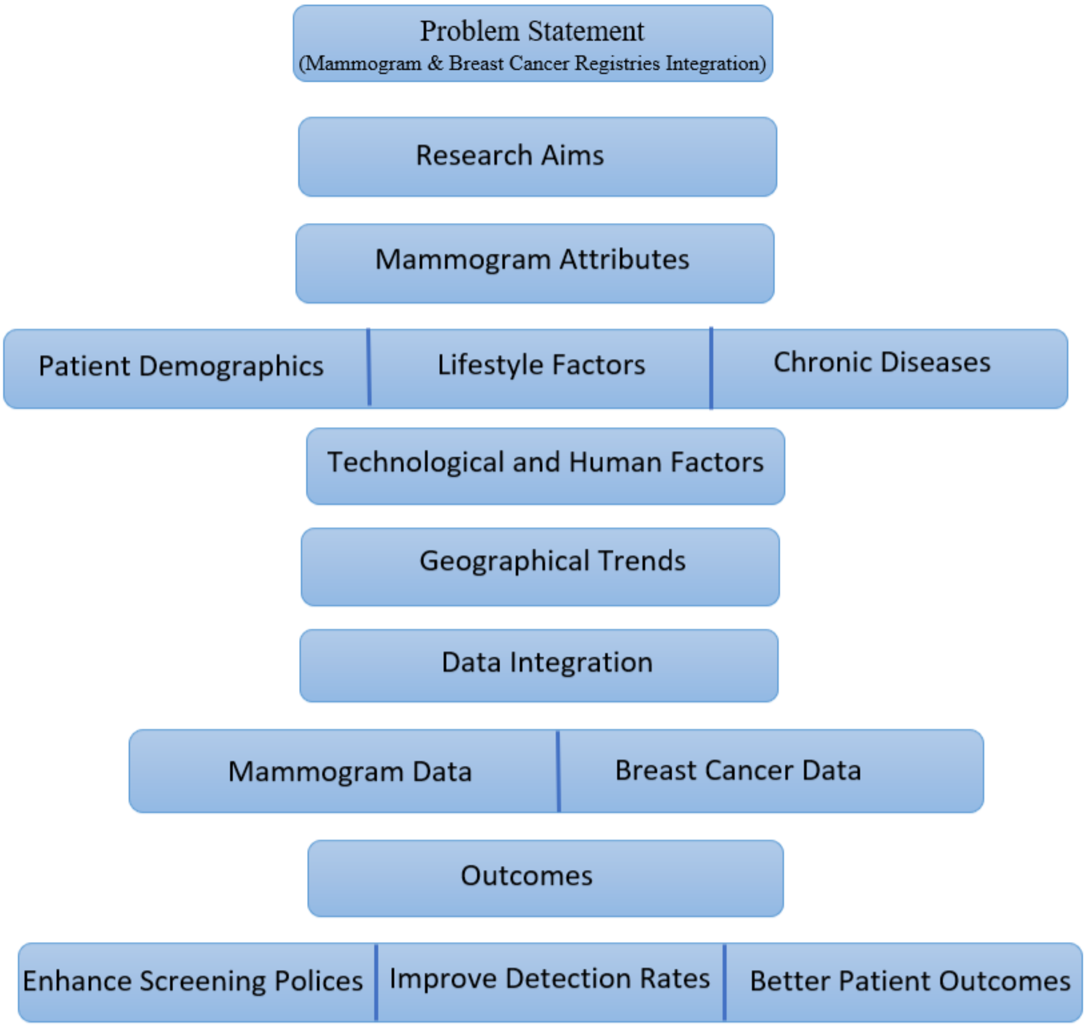

Figure 3.1: Conceptual Framework for Breast Cancer Detection Methods in Palestine

The conceptual framework for breast cancer detection methods in Palestine is visualized in the diagram (Figure 3.1). The framework highlights the relationships between various factors such as mammogram attributes, geographical variability, data completeness, and timeliness, as well as technological and human factors affecting mammogram accuracy and breast cancer detection. This diagram helps in understanding the interconnectedness of different variables and their impact on breast cancer screening and detection outcomes.

Chapter 4: Methodology

4.1 Introduction

This study primarily adopts a quantitative approach, supplemented by qualitative methods, to analyze mammography screening and cancer detection. It uses secondary data analysis from registry records in an observational epidemiological model within a retrospective cohort study design, focusing on all women in the West Bank who underwent mammography screenings at the Ministry of Health (MOH) using a non-probability census sampling method.

The quantitative analysis applies algorithms to uncover insights. Pearson correlation evaluates relationships between 14 mammogram attributes and BIRADS scores, while matching algorithms check data accuracy by comparing high BIRADS scores with cancer registry entries. Time-series analysis examines delays in transferring BIRADS data to the cancer registry, and geographical analysis maps breast cancer rates across West Bank regions. These methods enhance detection and data integration.

In addition, data collection, merging, variable selection, data cleaning, and normalization ensure data quality. Qualitative methods assess completeness of registries through informatic-based follow-ups.

Finally, the study examines the automated migration of high BI-RADS scores from mammogram records to the cancer registry, ensuring smooth data integration and efficient cancer detection.

4.2 Study Design

This research adopts an observational study design, specifically a retrospective cohort study, to explore the role of mammography and health informatics in breast cancer detection in Palestine.

Observational Study:

The study analyzes pre-existing data without experimental intervention, focusing on naturally occurring patterns and outcomes related to breast cancer detection.

Secondary Data Analysis using Retrospective Cohort Approach:

This study employs a secondary data analysis method, utilizing a retrospective cohort approach to examine historical data from mammography and cancer registries. The cohort includes women in Palestine who have undergone mammography screenings. The research aims to explore the relationships between mammography utilization, the implementation of health informatics, and outcomes in breast cancer detection.

Data Sources:

Mammography registry records

Cancer detection databases

Population:

All women in the west bank in Palestine who have undergone mammography and are recorded in mammogram and breast cancer registries.

Variables:

Independent Variables: Mammography utilization and the integration of health informatics

Dependent Variables: Breast cancer detection rates, cancer stage at diagnosis, and patient outcomes

Analysis:

Secondary data analysis methods are applied to explore the relationship between mammography use, health informatics, and breast cancer detection. Comparisons are made to assess detection rates and outcomes before and after the introduction of health informatics innovations.

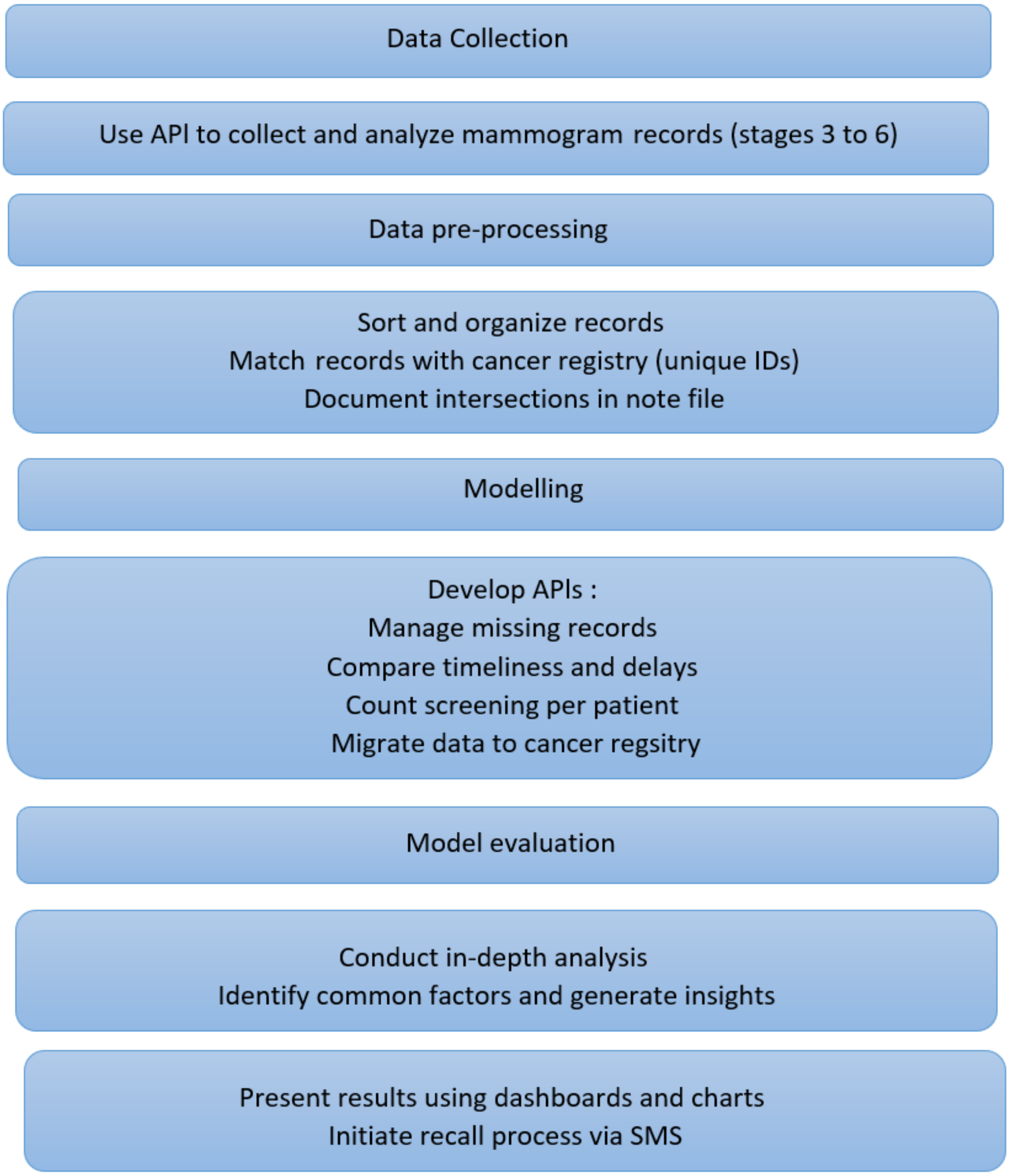

Figure 4.1: Study Design Diagram

Here is the diagram representing the study design. Each step is illustrated to show the process flow from one to nine.

4.3 Study Settings

The research focuses on women residing in the West Bank who sought mammography screenings at the Ministry of Health (MOH) care centres. The study area encompasses the geographical regions of the West Bank, where MOH facilities provide healthcare services, particularly mammography screenings. The target population consists of women who attended these MOH care centres for mammography screenings. The primary data source for the research is the mammogram registry maintained by the Ministry of Health. This registry contains detailed records of screening results for each woman who underwent mammography. The information logged in the registry includes the outcomes of the screenings, such as the presence or absence of abnormalities, cancer stages, and other relevant diagnostic details. The primary participants of this study were women who underwent mammography screenings at the MOH care centres. The study focused on women diagnosed with cancer stages 3 to 6, thereby excluding normal and benign cases (stages 0, 1, 2). Figure 3.1: Venn diagram illustrating the intersection between the Mammogram Registry with high BI-RAD and the Breast Cancer Registry. The overlapping area between the two circles represents patients who appear in both registries, highlighting the intersection.

4.4 Study Focus

This research aims to enhance data quality and registry integration in Palestine to improve breast cancer screening and treatment. Key objectives include ensuring accurate and comprehensive data, integrating cancer and mammogram registries for a holistic view, and maintaining timely data updates. These efforts are crucial for early detection, informed decision-making, and advancing breast cancer research and treatment strategies in the region.

4.5 Population and Sampling

4.5.1 Population

The population for this study includes all women in the West Bank region of Palestine who underwent mammography screenings at the Ministry of Health (MOH) care centers. This population is represented by the cases recorded in both the old mammogram registry (2016-2019) and the new DHIS2 system (end of 2019-2024), along with cases in the breast cancer registry.

4.5.2 Sampling Technique

The study uses census sampling, meaning no specific sampling technique was applied; all available cases from the registries during the specified time frames were included.

4.5.3 Inclusion Criteria

Time Frame: Women whose cases were recorded in the mammogram registry from 2016 to 2024.

Diagnosis: Women diagnosed with breast cancer at stages 3 to 6, as indicated in the intersecting cases between the mammogram and breast cancer registries.

4.5.4 Exclusion Criteria

Time Frame: Cases outside the 2016-2024 period for the mammogram registry.

Diagnosis: Women with normal or benign mammogram results (cancer stages 0, 1, 2).

Cases that do not have an intersection between the mammogram and breast cancer registries, suggesting that they either were not diagnosed with breast cancer, or their cancer stages fall outside of 3 to 6.

4.5.5 Sample Size

- Mammogram Registry (Legacy PHP-based Website 2016-2019): 17,351 cases.
- Mammogram Registry (New DHIS2 System End of 2019-2024): 22,335 cases.
- Breast Cancer Registry (Intersecting with Mammogram Records): 1,683 cases.

The mammogram registries contain a total of 39,686 cases: 17,351 from the legacy PHP-based system and 22,335 from the newly implemented DHIS2 system. Of these, 1,683 cases are identified as having cancer. This analysis involves intersecting the breast cancer registry with the mammogram registry, focusing on cases that meet the inclusion criteria of having a cancer diagnosis and mammograms with BIRAD stages ranging from 3 to 6.

4.6 Data Collection

In this study, two primary data sources are drawn upon: the mammogram registry and the breast cancer registry. The mammogram registry provided us with a comprehensive dataset pertaining to individuals who underwent mammography screening, encompassing various aspects such as patient demographics, examination results, and follow-up information. On the other hand, the breast cancer registry supplied critical information about patients diagnosed with breast cancer, including cancer type, stage, treatment, and outcomes. These registries, both administered by the Ministry of Health in Palestine, played a pivotal role in ensuring standardized data collection and management practices across healthcare facilities. This oversight by the Ministry not only bolstered the reliability of our data but also promoted consistent, high-quality care for patients undergoing mammography screening and those facing a breast cancer diagnosis throughout the region.

4.6.1 Data Sources

This study is managing with two registries administered by the Ministry of Health in Palestine: The Mammogram Registry and the Breast Cancer Registry.

4.6.1.1 Mammogram Registry

The Mammography-e-Registry in Palestine tracks and monitors women undergoing mammography in the West Bank, aiming to improve screening programs and guide breast cancer interventions. Key features include a centralized electronic system storing patient data, unique identifiers for each woman, automated appointment reminders, standardized BI-RAD classification, and performance reporting through dashboards. Benefits include early cancer detection, improved patient care, informed decision-making, and optimized resource allocation. Developed by the PNIPH and Ministry of Health, it is operational in 14 West Bank facilities with plans to expand coverage to the Gaza Strip and private facilities. Future goals include integrating with other registries for comprehensive data and improving data analysis for research and policy. (Palestinian National Institute of Public Health, n.d.)

4.6.1.2 Breast Cancer Registry

The Breast Cancer Registry aims to improve breast cancer management in Palestine amid challenging conditions. Developed by the PNIPH and the Ministry of Health, it collects data from primary healthcare facilities in the West Bank. The dataset, spanning 14 years, includes demographics, medical history, screening details, outcomes, and administrative data. Strengths include aiding early detection, supporting quality control, and providing data for policy and resource allocation. Limitations are geographical coverage restricted to the West Bank and potential data gaps. Future directions involve expanding to the Gaza Strip and private facilities, integrating with other registries, enhancing data utilization for research and policymaking, and improving data quality and completeness. (Palestinian National Institute of Public Health, n.d.)

4.6.2 Registry Administration

The Ministry of Health plays a pivotal role in overseeing and managing critical healthcare registries in the Palestinian healthcare system. These registries serve as essential tools for data collection, analysis, and management across healthcare facilities, contributing to the enhancement of public health and the overall quality of healthcare services in Palestine. This comprehensive overview explores the Ministry of Health's role in administering these registries, with a focus on ensuring standardized data collection and management practices.

4.6.3 Data Access Approval

This research project was facilitated by a collaborative partnership established with the Arab American University and the Ministry of Health in Palestine. This collaboration was initiated following extensive and successful communication between the university and the Ministry of Health, highlighting the importance of cooperation between academic institutions and governmental bodies in advancing healthcare research. Through this partnership, official authorization was obtained to access and utilize the comprehensive datasets from both the mammogram registry and the breast cancer registry. This authorization was a pivotal milestone in our study, as it granted us unrestricted access to a treasure trove of healthcare information. (See Appendix D).

4.7 Data Integration and Preparation

4.7.1 Data Merging

To gain insights, data merged from the mammogram and breast cancer registries using unique IDs, ensuring integrity and privacy. This unified dataset revealed patterns and correlations between mammogram findings and cancer diagnoses, allowing us to evaluate mammography effectiveness, risk factors, and patient outcomes. It provided a comprehensive view of the patient journey and formed the basis for evidence-based recommendations to improve breast healthcare in Palestine.

4.7.2 Variable Selection

Initial Variable Count: The combined dataset initially contained more than one hundred variables from both registries. This extensive list included a wide range of data points, from basic demographic information to specific medical history details.

Focused Selection: A focused approach was adopted to refine the dataset to align with the research objectives. Out of the original one hundred variables, twenty-six variables were selectively retained for detailed analysis. This decision was based on the relevance of each variable to the study's goals and the potential insights they could provide. The selected variables cover crucial aspects such as patient demographics (e.g., age, district), medical history (e.g., diabetes, hypertension), and specific details regarding the mammogram (e.g., density, BIRAD).

Criteria for Selection: Variables were selected based on their relevance to key factors identified in the current body of literature, prioritizing those most critical to the study's objectives. To ensure the robustness of the analysis, variables with significant levels of missing data or those of questionable quality were excluded. Additionally, variables that demonstrated limited variability or low statistical power in preliminary analyses were omitted to avoid skewed results. Consideration was also given to the feasibility of data collection and the potential for bias, with efforts made to include only those variables that could be reliably measured and reported. A sample table was created to provide an overview of the selected variables, showcasing ten rows of the twenty-six variables chosen for detailed analysis. Table 4.1 below presents a sample of these variables from the mammogram dataset.

Table 4.1: Mammogram Registry Sample of Variables

ID	1	2	3
District ID	Bethlehem	Al-Ram	Al-Ram
Birth Date	11/12/1969 0:00	1/26/1959 0:00	6/28/1969 0:00
Appointment Date	8/24/2016 0:00	11/23/2016 0:00	11/23/2016 0:00
Mass (KG)	71	88	81
BMI	26.7229	35.2508	34.1542
Referral Status	Yes	Yes	No
Source of Referral	Non-Governmental		
Previous Mammogram	No Mammo	> 2 Yr	> 2 Yr
Menarche Age	> 12	> 12	> 12
First Pregnancy Age	< 35 Yr	Not Pregnant	< 35 Yr
Contraceptive Use	Never	Never	Never
Menopause	Not Reach	40-55 Yr	Not Reach
Hormonal Replacement Therapy	Never	Never	Never
Smoking	Neither	Neither	Both
Chronic Diseases	None	Hypertension, Diabetes	Hypertension
Hypertension	No	Yes	Yes
Diabetes	No	Yes	No
Heart Diseases	No	No	No
Personal History of Breast Cancer	Neither	Neither	Neither
Family History of Breast Cancer	Mother	None	None
Personal History of Ovarian Cancer	No	No	No
Family History of Ovarian Cancer	No	No	No
Mammogram Reason	Screening	Diagnostic	Diagnostic
Density	Heterogeneous Dense	Scattered fibro glandular Density	Scattered fibro glandular Density
BIRAD	1	3	1

This table should provide a clear example of how the variables are structured and what kind of data they contain, which can be crucial for your research analysis.

4.7.3 Data Cleaning

This section describes the data cleaning methods used to improve data quality. The process involved identifying and correcting errors and inconsistencies using Microsoft Excel, which offers versatile functionalities for sorting, filtering, and modifying data.

- **Missing Values:** Managing missing values is crucial in data cleaning to prevent biased estimates and reduced statistical power. The approach involved removing records with missing values or imputing them using methods like nearest neighbour value substitution and mean calculation. Nearest neighbour value substitution replaces missing values with similar ones, while mean calculation replaces numerical missing values with the average, maintaining data distribution.
- **Inconsistent Formats:** Standardizing data formats is crucial for accurate analysis, especially with large datasets from multiple sources. Text data was reformatted by removing extraneous HTML tags, standardizing date formats, and standardizing BI-RAD classifications to ensure consistency in categorization and interpretation.
- **Outliers**: Outliers, which significantly differ from other data points, can skew, and mislead machine learning models. To manage outliers, the Winsorization method was used, which involves clipping the extreme values and replacing them with nearby non-outliers. This technique reduces the impact of outliers, ensuring more accurate and reliable results.
- **Translation**: In datasets containing multilingual data, it is essential to translate all text into a single language for uniformity. This ensures data interpretation and analysis consistency, particularly in machine learning models where linguistic consistency is key.

In this dataset, variables containing a mix of Arabic and English were uniformly translated into English. This step is crucial for maintaining data consistency and ensuring that all analysts and stakeholders can understand and use the data effectively, regardless of their language proficiency.

Table 4.2: Mammogram Sample before Cleaning and Translation

Personal history of Ovarian cancer	Family history of Ovarian cancer	Clinical examination result	Mammogram reason	Menarche	BIRAD	Recommendations
لا (No)	لا (No)	b	تشخيصي (Diagnostic)	1	1	Continue Routine Screening
لا (No)	لا (No)	a	تقصي (Screening)	1	3	To do breast and axillary U/S
لا (No)	نعم (Yes)	b	تقصي (Screening)	3	1	For routine follow-up.
لا (No)	لا (No)	a	تشخيصي (Diagnostic)	1	0	To do breast US.
لا (No)	لا (No)	b	تشخيصي (Diagnostic)	1	0	To do breast and axillary U/S
لا (No)	لا (No)	a	تشخيصي (Diagnostic)	1	1	For routine follow-up
نعم (Yes)	نعم (Yes)	c	تقصي (Screening)	2	2	For routine follow-up

This table contains raw mammogram data, including personal and family medical history, clinical results, mammogram reasons, and recommendations, prior to data cleaning and translation.

4.7.4 Data Normalization

Data normalization is a critical step in the research methodology that involves transforming and pre-processing raw data to make it suitable for analysis. This step aims to ensure that the data is in a consistent format, reduces the impact of outliers, and helps in making meaningful comparisons across different variables. In the context of the research, data normalization encompasses two main processes: scaling and categorization.

Scaling: Scaling ensures continuous variables have equal weighting in the analysis, preventing variables with larger ranges from dominating. In this research, continuous variables like age and appointment dates were scaled using normalization techniques such as Min-Max scaling or Z-score standardization. Min-Max scaling transforms data to a common range, while Z-score standardization centres the data around the mean with a standard deviation of one.

Categorization: Categorization converts variables into discrete groups, simplifying qualitative or nominal data for easier interpretation and comparison. In the research, variables like district ID, BI-RAD classifications, chronic diseases, smoking habits, and mammogram reasons were categorized. For example, district IDs were grouped into urban, suburban, and rural categories, BI-RAD classifications into benign, suspicious, and malignant groups, and smoking habits into non-smoker, occasional smoker, and regular smoker categories. This process simplifies analysis, facilitates the use of statistical tests, and helps draw meaningful conclusions. Overall, categorization and scaling ensure data is ready for analysis, making it easier to explore relationships and patterns, and enhancing research findings.

Table 4.3: Mammogram Sample of un-Categorized Features

Previous mammogram	Menarche	First pregnancy age	Contraceptive use	Menopause	Smoking	Chronic diseases	Personal History of BC
e	b	A	d	d	d	a	d
d	b	C	d	b	d	b	d
d	a	A	d	d	c	d	d
e	b	A	d	d	d	a, b	d
e	b	A	d	d	d	d	d
d	b	A	d	d	d	d	d
c	a	A	d	d	d	d	d
e	b	A	d	b	d	d	d
d	b	A	d	d	d	d	d
e	b	A	d	b	d	a	d
d	b	A	d	d	d	d	d
d	b	A	d	d	d	d	d
d	b	A	d	d	d	d	d
d	b	A	d	b	d	d	d
d	a	A	d	d	d	d	d
d	c	A	d	d	d	d	d

This table contains mammogram data with various features such as previous mammogram results, menarche age, first pregnancy age, contraceptive use, menopause, hormonal replacement therapy, smoking, chronic diseases, and personal history of breast cancer.

4.8 Feature Engineering and Dimensionality Reduction

Feature engineering and dimensionality reduction are crucial steps in the research methodology aimed at enhancing the quality of data and improving the efficiency of data analysis. These steps involve the creation of new variables (feature engineering) and the elimination of less important variables (dimensionality reduction).

4.8.1 Feature Engineering

Feature engineering, essential for capturing underlying patterns and relationships in data, involved creating new variables from existing data in the research. This included calculating age by subtracting the birth date from the appointment date and computing BMI from mass and height data to explore their impact on health outcomes. Additionally, variables were created to identify prevalent chronic diseases like hypertension, diabetes, and heart disease. These steps enriched the dataset, uncovered hidden patterns, and improved its overall quality for analysis.

4.8.2 Dimensionality Reduction

Dimensionality reduction simplifies the dataset by removing variables that do not significantly contribute to the research or introduce noise. This process involves eliminating high-deviation variables to reduce complexity and improve analysis efficiency. Techniques like variance analysis and feature selection algorithms help retain the most informative variables. Together with feature engineering, which creates new relevant variables, dimensionality reduction streamlines the dataset, enhancing the quality of analysis and increasing the likelihood of meaningful findings.

4.9 API Utilization and Algorithm Selection

This study employed PostgreSQL for data management and integration, leveraging its advanced capabilities to handle and query large datasets effectively. Approximately 30,000 mammogram records and 1,500 breast cancer records were stored in PostgreSQL tables. An outer join operation was executed using the patient index as a key, facilitating the creation of a comprehensive dataset that combined mammogram and cancer registry data.

API Utilization:

PostgreSQL API: The PostgreSQL API was instrumental in managing and querying the large datasets. It provided robust support for SQL queries, which were used for data merging, retrieval, and manipulation. PostgreSQL's advanced indexing and optimization features ensured efficient handling of complex joins and queries, which was crucial for integrating and analyzing the datasets.

Algorithm Selection and Application:

Correlation Analysis: Pearson correlation coefficients were calculated to assess the relationships between 14 selected mammogram attributes and BIRADS scores. This algorithm was chosen for its effectiveness in measuring the strength and direction of linear relationships between continuous variables.

Matching Algorithms: To evaluate the alignment between high BIRADS scores and breast cancer registry cases, matching algorithms were employed. These algorithms compared high-risk mammogram cases with corresponding cancer registry records to verify the accuracy and completeness of the data integration.

Time-Series Analysis: The timeliness of transferring high BIRADS scores from the mammogram registry to the cancer registry was analyzed using time-series algorithms. These algorithms helped quantify the duration of the transfer process and identify factors influencing delays, providing insights into the efficiency of data handling procedures.

Geographical Analysis: To analyze the geographical distribution of breast cancer prevalence across West Bank governorates, geographical analysis algorithms were used. These algorithms calculated regional percentages and mapped the data to highlight variations in breast cancer incidence across different areas.

These algorithms and APIs were selected based on their ability to provide precise, reliable, and actionable insights from the data, supporting the study's objective to enhance breast cancer detection and registry practices in Palestine.

4.10 Ethical Considerations

This research requires the management of sensitive patient data, with identities rigorously masked and encrypted to ensure privacy. Data from two distinct registries is carefully merged to maintain accuracy and uphold data quality. The study measures the performance of technicians, equipment, and medical centers across both governmental and non-governmental organizations, with strict confidentiality applied to all findings. Continuous communication with stakeholders is prioritized to maintain transparency and address any concerns. These ethical considerations are essential for protecting patient and organizational privacy, ensuring data integrity, and maintaining the credibility of the research outcomes.

Table 4.4: Data Sample for Encrypting Some Features in Mammogram

Event	Program stage	Tracked entity instance	ID
C8PM7ZpG7xd	FBqE4jwHLgu	e7aPoNOAtEZ	ouyF8vGXRev
eDhW1r8DHOP	FBqE4jwHLgu	e7aPoNOAtEZ	fGUiTn8acpf
zopKdQJTcl2	FBqE4jwHLgu	D4Fgg8wdB89	sZtS97dqwpu
EWwtd4u7Usr	FBqE4jwHLgu	FLt2zGVagn5	Br8Wddd7Nqj

This table contains a sample of encrypted data for various features in a mammogram dataset, including event IDs, program stages, tracked entity instances, and unique identifiers.

Chapter 5: Results

5.1 Overview

This chapter presents the results of the research, organized around three key questions. First, it explores the relationship between specific mammogram attributes and breast cancer evidence, along with the geographical distribution of cases across West Bank governorates. Second, it assesses the effectiveness of integrating high-risk mammogram cases into the breast cancer registry and how well these registries capture such cases. Finally, it examines the timeliness of transferring high BIRAD scores to the registry, identifying factors contributing to variations in this process.

5.2 Calculation of Mammogram Attribute Weights in BI-RAD

5.2.1 BI-RADS vs BMI

Descriptive Statistics

Table 5.1: BI-RADS BMI Descriptive Data

BI-RADS	Mean BMI	Standard Deviation	Count
0	30.2	5.7	4620
1	31.33	6.24	8698
2	32.07	7.62	10684
3	32.17	7.4	4526
4	29.65	7.9	650
5	29.54	5.5	309
6	28.49	6.97	135

Individuals with higher BMI tend to have benign BI-RADS scores (1-3), while those with suspicious or malignant findings (4-6) have lower BMI. Regular screenings and comprehensive health management are essential.

Correlation Analysis

Correlation Coefficient (BMI with BI-RADS): 0.046489

Interpretation

The correlation coefficient between BI-RADS scores and BMI is approximately 0.046. The result 0.046, which indicates a very weak positive linear relationship. This suggests that while there is a slight tendency for BMI to increase with higher BI-RADS scores, the relationship is not strong and should be interpreted with caution. In summary, the correlation analysis shows that BMI and BI-RADS scores are not strongly correlated, implying that other factors might play a more significant role in determining BI-RADS scores.

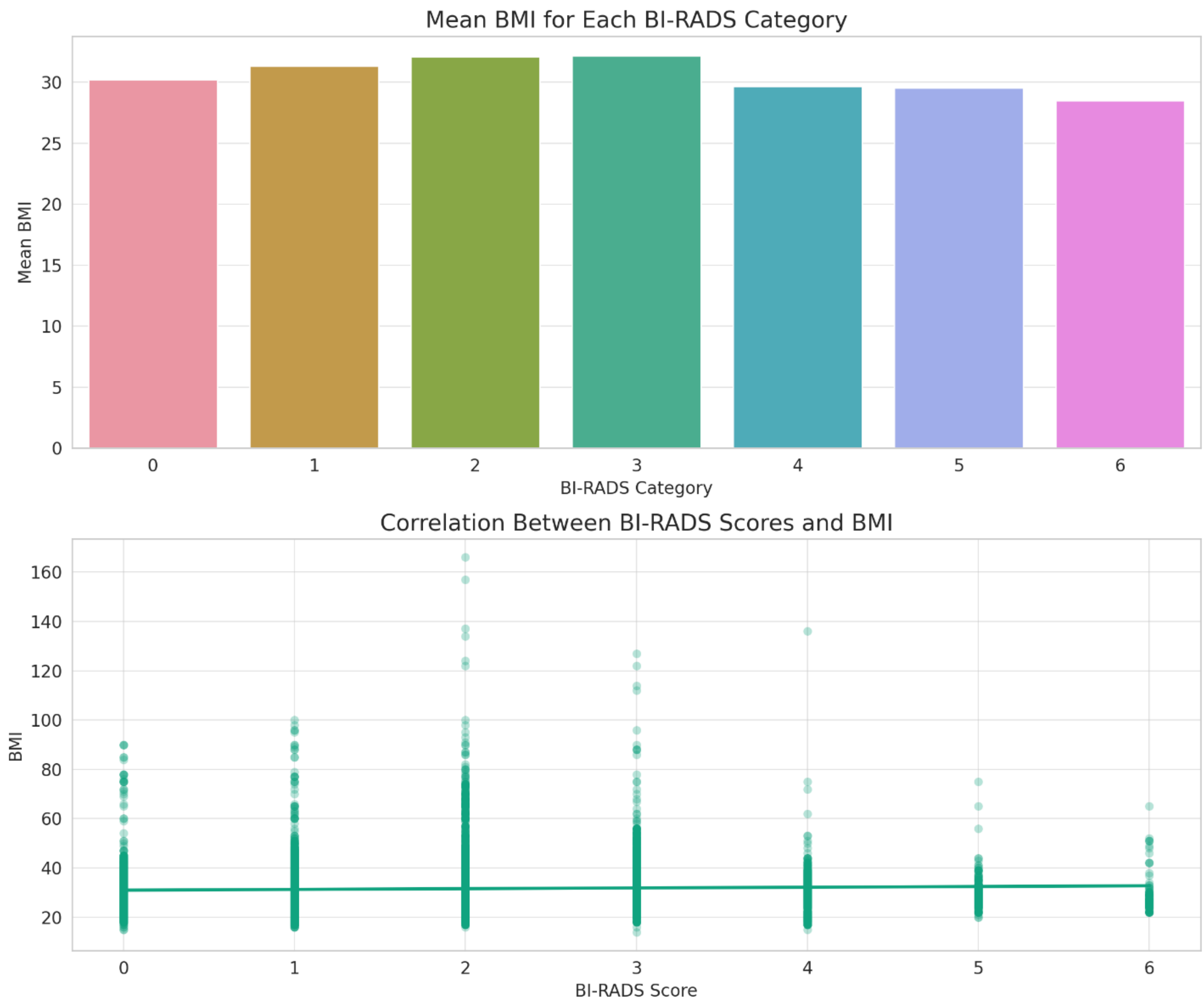

Figure 5.1: BMI Mean per BI-RADS Category

Figure 5.1 shows two graphs: The Mean BMI for Each BI-RADS Category: This bar graph shows the average BMI corresponding to each BI-RADS category. It illustrates the variations in mean BMI across different BI-RADS scores. Correlation Between BI-RADS Scores and

BMI: The scatter plot with a linear fit demonstrates the relationship between BI-RADS scores and BMI. The plot visually represents the weak positive correlation indicated by the correlation coefficient. The points are distributed widely around the regression line, underscoring the weak nature of this correlation.

5.2.2 BI-RADS vs Chronic Diseases

Table 5.2: BI-RADS and Chronic Diseases Descriptive Statistics

Count	Mean	Standard Deviation	Minimum	25th percentile	Median (50th percentile)	75th percentile	Maximum
29622	1.64	1.09	0	1	2	2	6

In 29,622 entries, chronic diseases yield a mean BI-RAD score of 1.64, mostly benign, with some variability (SD 1.09). Regular screenings are essential.

Table 5.3: BI-RADS with and without Chronic Diseases

BI-RAD Score	No Chronic Diseases	Yes Chronic Diseases
0	3614	1006
1	5728	2970
2	5521	5163
3	2364	2162
4	439	211
5	256	53
6	126	9

Chronic diseases result in benign BI-RAD scores with fewer negative findings (score 0). Regular screenings are essential

Table 5.4: Chi-square test results for BI-RADS with Chronic Diseases

Chi-square statistic	P-value	Degrees of Freedom
1331.16	1.95E-284	6

Given the extremely low p-value (far below any standard significance level, e.g., 0.05), null hypothesis was rejected. This indicates there is a statistically significant association between BI-RADS scores and the presence of chronic diseases.

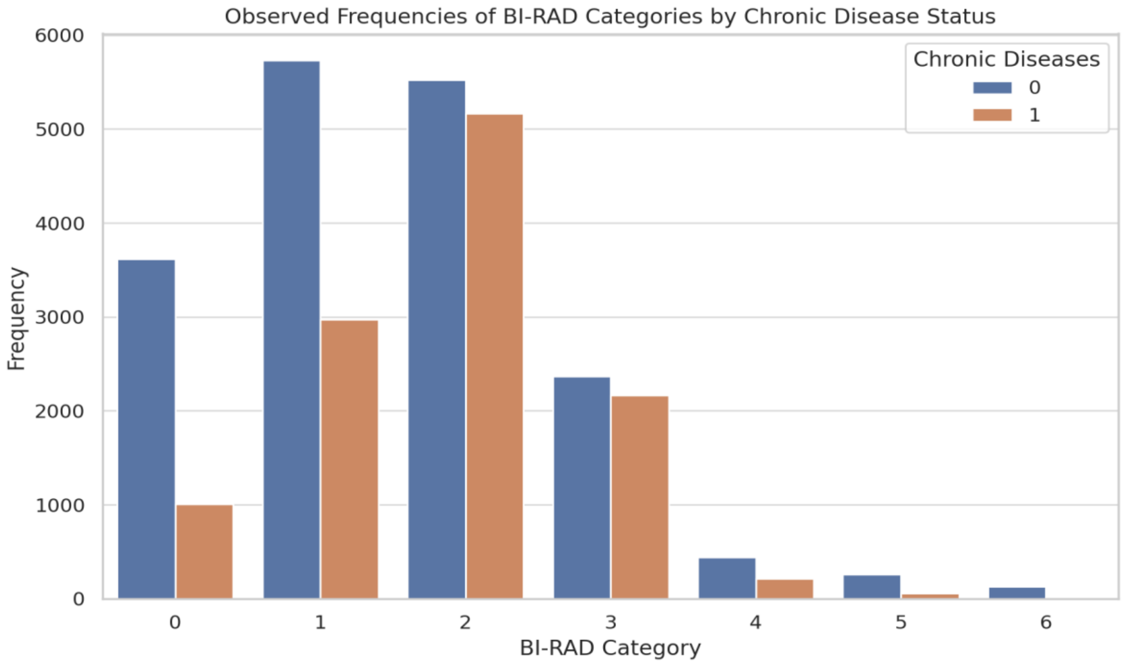

Figure 5.2: BI-RAD for Chronic Diseased Patients

The bar chart shows the frequencies of BI-RAD categories, comparing individuals with and without chronic diseases. It highlights significant differences in BI-RAD category distribution between these groups.

5.2.3 BI-RADS vs Diabetes

Table 5.5: Descriptive statistics for BI-RADs with Diabetes

Mean BI-RAD	Standard Deviation BI-RAD	Range BI-RAD	Total Diabetes Entries	No Diabetes Entries
1.65	1	0 to 6	16924	13590

In 16,924 entries, diabetes patients have a mean BI-RAD score of 1.65. Diabetes does not significantly raise average BI-RAD scores, but regular screenings are essential due to score variability.

Correlation Analysis: The correlation coefficient between 'BI-RAD' and 'Diabetes' is approximately 0.100. This suggests a weak positive correlation. That is, there is a slight tendency that higher BI-RAD scores are associated with the presence of diabetes, but this relationship is not strong.

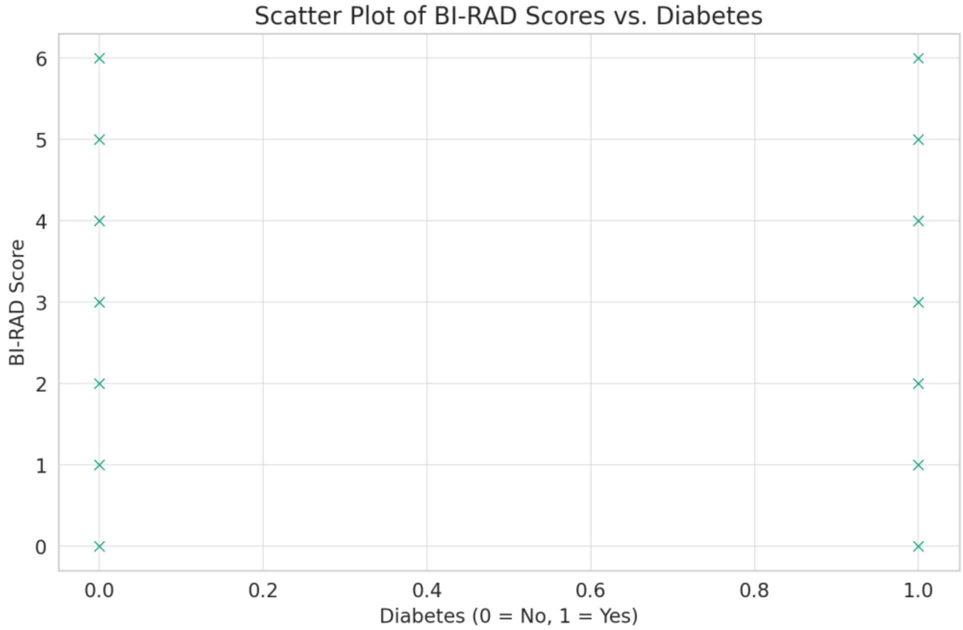

Figure 5.3: BI-RADS vs Diabetes

This scatter plot shows the relationship between BI-RAD scores and diabetes presence (0 for 'No', 1 for 'Yes'). The x-axis represents diabetes presence, and the y-axis represents BI-RAD scores. The plot illustrates a weak positive correlation, visually highlighting the distribution of data points.

5.2.4 BI-RADS vs Heart Diseases

Table 5.6: BI-RADS vs Heart Diseases Descriptive Statistics

Total Entries	Yes Heart Diseases	No Heart Diseases	Percentage No Heart Diseases	Missing Values
16,924	640	16,284	96.20%	None

In 16,924 entries, 96.2% have no heart diseases, and 3.8% do. Detailed analysis is needed to understand the impact on BI-RAD scores.

Correlation Analysis: The correlation coefficient between 'BI-RAD' and 'Heart Diseases' is approximately 0.038, indicating a very weak positive correlation. This suggests a slight, negligible association between higher BI-RAD scores and heart diseases. Further analysis is needed for a comprehensive understanding.

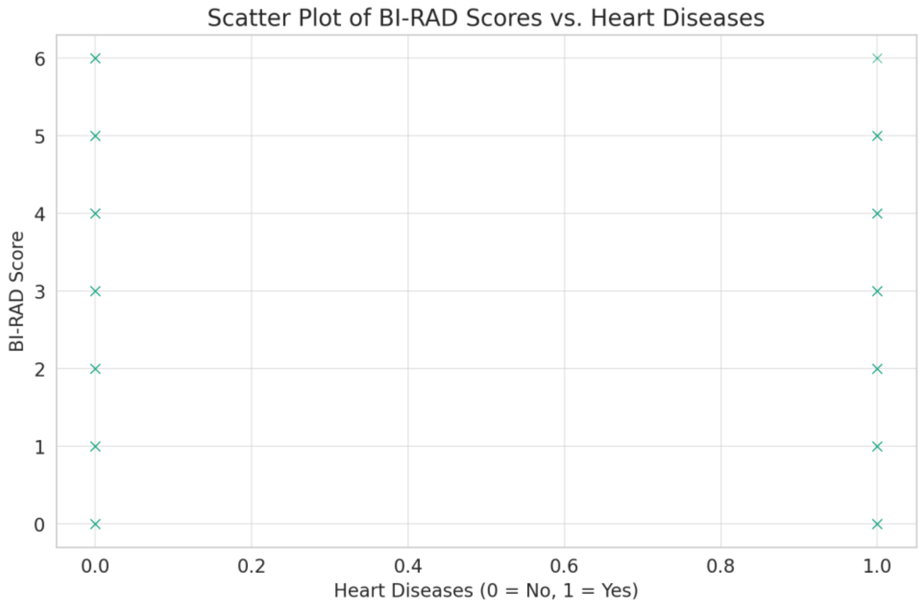

Figure 5.4: BI-RADS vs Heart Diseases

This scatter plot shows the relationship between BI-RAD scores and heart diseases (0 for 'No', 1 for 'Yes'). The distribution of points indicates a very weak positive correlation, with most points clustered at '0' for heart diseases, reflecting their low prevalence in the dataset. This visual aligns with the minimal association found in the correlation analysis.

5.2.5 BI-RADS vs Hypertension

Descriptive Statistics:

Table 5.7: BI-RADS vs Hypertension Descriptive Statistics

Total Entries	BI-RAD Value Range	Hypertension Representation	Mean BI-RAD	Standard Deviation BI-RAD	Hypertension Prevalence	Missing Values
16,924	0 to 6	1 for 'Yes', 0 for 'No'	1.65	1	0.17%	None

In 16,924 entries, hypertension (0.17%) shows a mean BI-RAD score of 1.65, indicating mostly benign findings. Hypertension has little impact on average BI-RAD scores. Regular screening is essential.

Correlation Analysis: The correlation coefficient between 'BI-RAD' and 'Hypertension' is approximately -0.018, indicating a very weak negative correlation. This suggests minimal association between BIRAD scores and hypertension. Given the low prevalence of hypertension in the dataset, further analysis with more sophisticated methods may be needed for a detailed understanding.

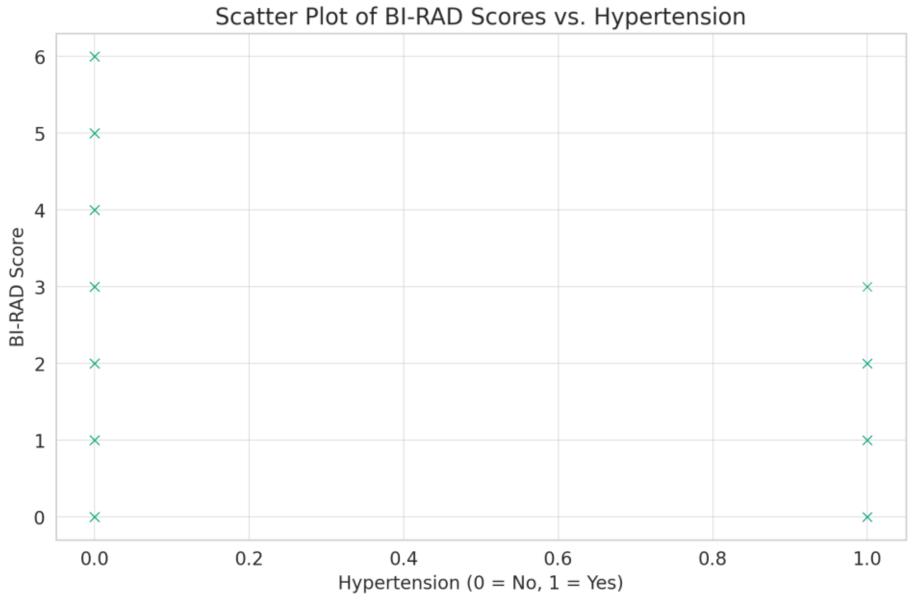

Figure 5.5: BI-RADS vs Hypertension

This scatter plot shows the relationship between BI-RAD scores and hypertension (0 for 'No', 1 for 'Yes'). It reveals a very weak negative correlation, with most points clustered at '0' for hypertension, reflecting its low prevalence. This supports the minimal association found in the correlation analysis.

5.2.6 BI-RADS vs Contraceptive Use

Table 5.8: BI-RADS vs Contraceptive Use Descriptive Statistics

Total Entries	BI-RAD Value Range	Contraceptive Use Representation	Mean BI-RAD	Standard Deviation BI-RAD	Mean Contraceptive Use	Missing Values
16,924	0 to 6	0 for 'Never', 1 for '< 10 Year', 2 for '=> 10 Year', 3 for 'Now'	1.65	1	0.24	None

Contraceptive use shows a mean BI-RAD score of 1.65 with moderate variability (standard deviation 1.0) across 16,924 entries. Most findings are benign, and no missing values ensure reliable analysis. Tailored screening based on contraceptive use is recommended.

Correlation Analysis: The correlation coefficient between 'BI-RAD' and 'Contraceptive Use' is approximately 0.083, indicating a weak positive correlation. This suggests a slight tendency for higher BI-RAD scores to be associated with longer or current contraceptive use. Further analysis may be needed due to the data's nature and categorical representation.

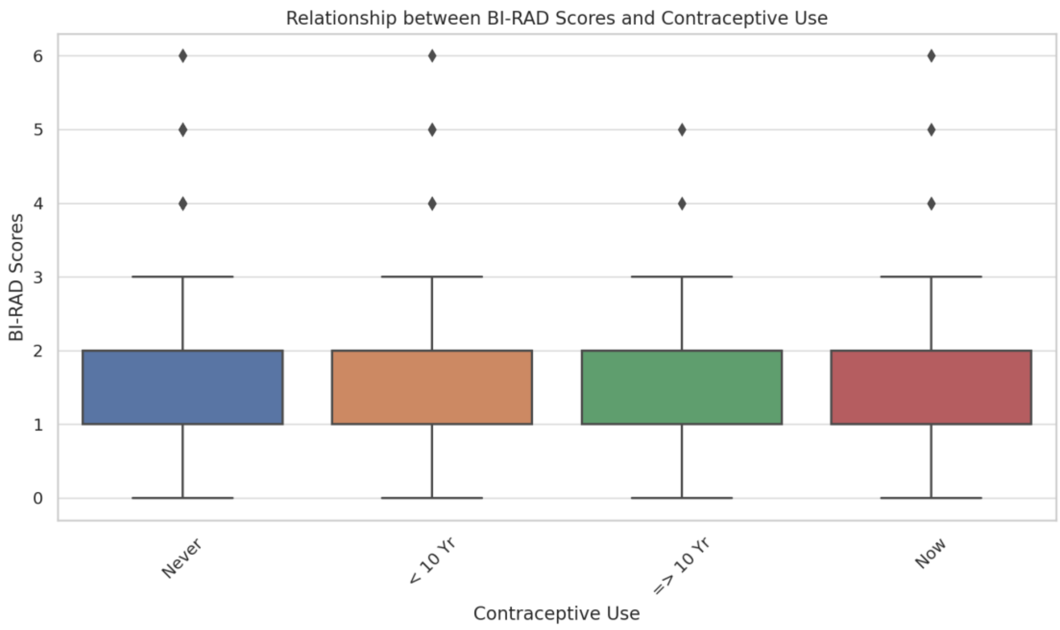

Figure 5.6: Contraceptive Use per BI-RAD Category

This box plot shows the relationship between BI-RAD scores and contraceptive use categories. It displays the quartiles and distribution of BI-RAD scores for each contraceptive use category, highlighting differences between groups and any outliers.

5.2.7 BI-RADS vs Breast Density

The "Density" column has four categories: Scattered fibro glandular density, Heterogeneous dense, almost entire fatty tissue, and extremely dense.

.

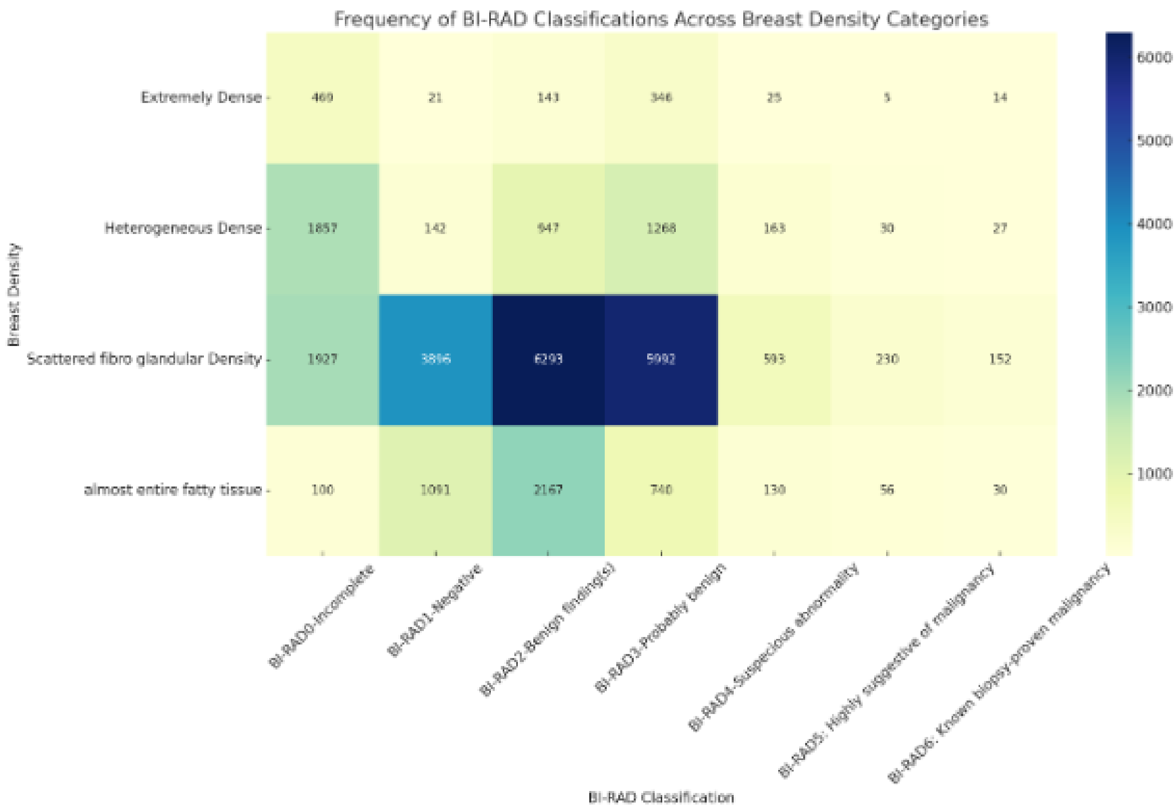

Figure 5.7: Distribution of BI-RAD across Breast Density Categories

The heatmap and frequency table reveal BI-RAD classification distributions across breast density categories. 'BI-RAD2-Benign finding(s)' and 'BI-RAD3-Probably benign' are common in 'Scattered fibro glandular Density'. 'Almost entire fatty tissue' shows more 'BI-RAD1-Negative' and 'BI-RAD2-Benign finding(s)'. 'Extremely Dense' has a balanced but lower BI-RAD distribution. A Chi-square test can confirm the statistical significance of these patterns.

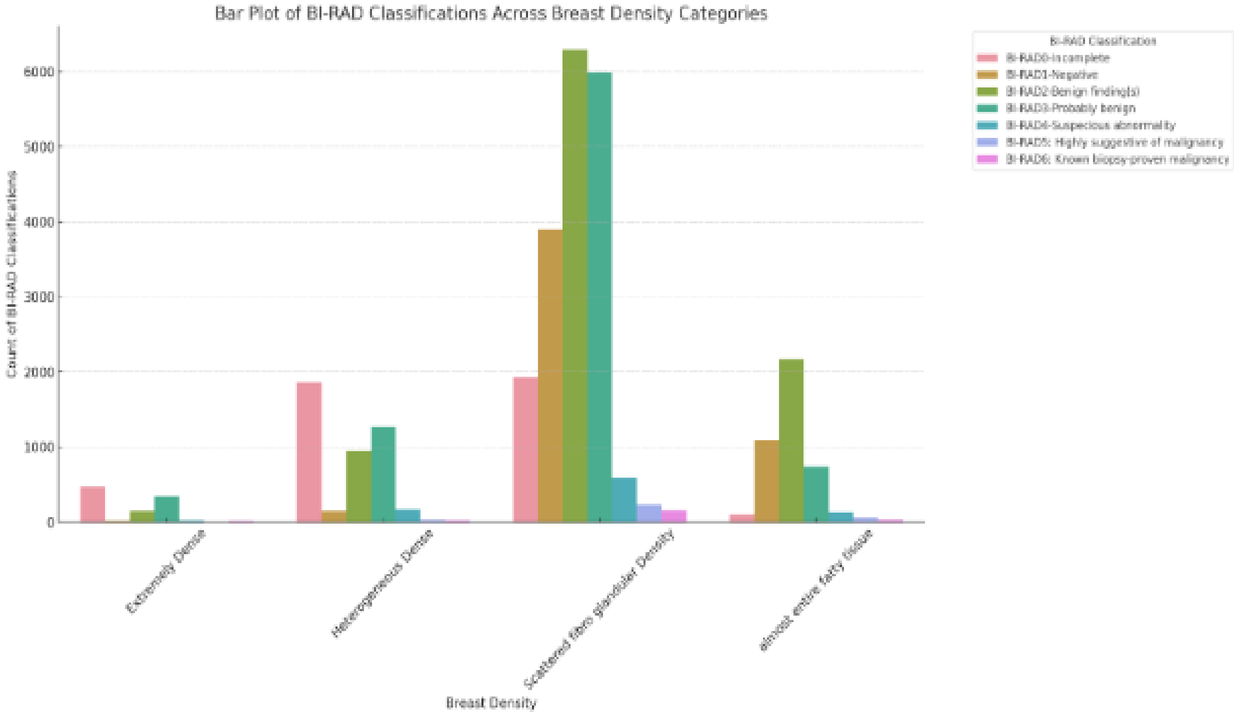

Figure 5.8: Bar Plot of BI-RAD Classifications across Breast Density Categories

The bar plot displays the count of each BI-RAD classification across breast density categories, using colors to differentiate BI-RAD scores. The bar heights show how BI-RAD frequencies vary by density, illustrating the relationship between them. Next steps: create a table for Chi-square test results and frequency analysis, and draw a chart showing the correlation between BI-RAD scores and breast density.

Chi-Square Test and Frequency Analysis Table

Table 5.9 below represents the Chi-square test results and the frequency analysis of BI-RAD classifications across different breast density categories:

Table 5.9: Total Counts of each Density per BI-RAD Category

Breast Density Category	BI-RAD0-Incomplete	BI-RAD1-Negative	BI-RAD2-Benign	BI-RAD3-Probably benign	BI-RAD4-Suspicious	BI-RAD5-Highly suggestive	BI-RAD6-Known malignancy	Total
Extremely Dense	469	21	143	346	25	5	14	1023
Heterogeneous Dense	1857	142	947	1268	163	30	27	4434
Scattered fibro glandular Density	1927	3896	6293	5992	593	230	152	19083
Almost entire fatty tissue	100	1091	2167	740	130	56	30	4314
Total	4353	5150	9550	8346	911	321	223	28854

The 'Total' row shows the overall counts and the results of the Chi-square test. The Chi-square statistic is 5375.47 with a p-value of 0.0 and 18 degrees of freedom, indicating a significant association between breast density and BI-RAD classifications.

5.2.8 BI-RADS vs First Pregnancy Age

Table 5.10: BI-RADS vs First Pregnancy Age Descriptive Analysis

First Pregnancy Age	Count	Mean BI-RAD Score	Standard Deviation	Min BI-RAD Score	Max BI-RAD Score	Median (50%) BI-RAD Score
< 35 Year	14,642	1.63	0.99	0	6	2
>= 35 Year	376	1.77	1.01	0	6	2
Not Pregnant	1,906	1.77	1.09	0	6	2

Women with first pregnancies under 35 have lower BI-RAD scores (mean 1.63). Those with first pregnancies at 35 or older and those never pregnant have higher scores (mean 1.77), indicating increased risk. Tailored screening is advised.

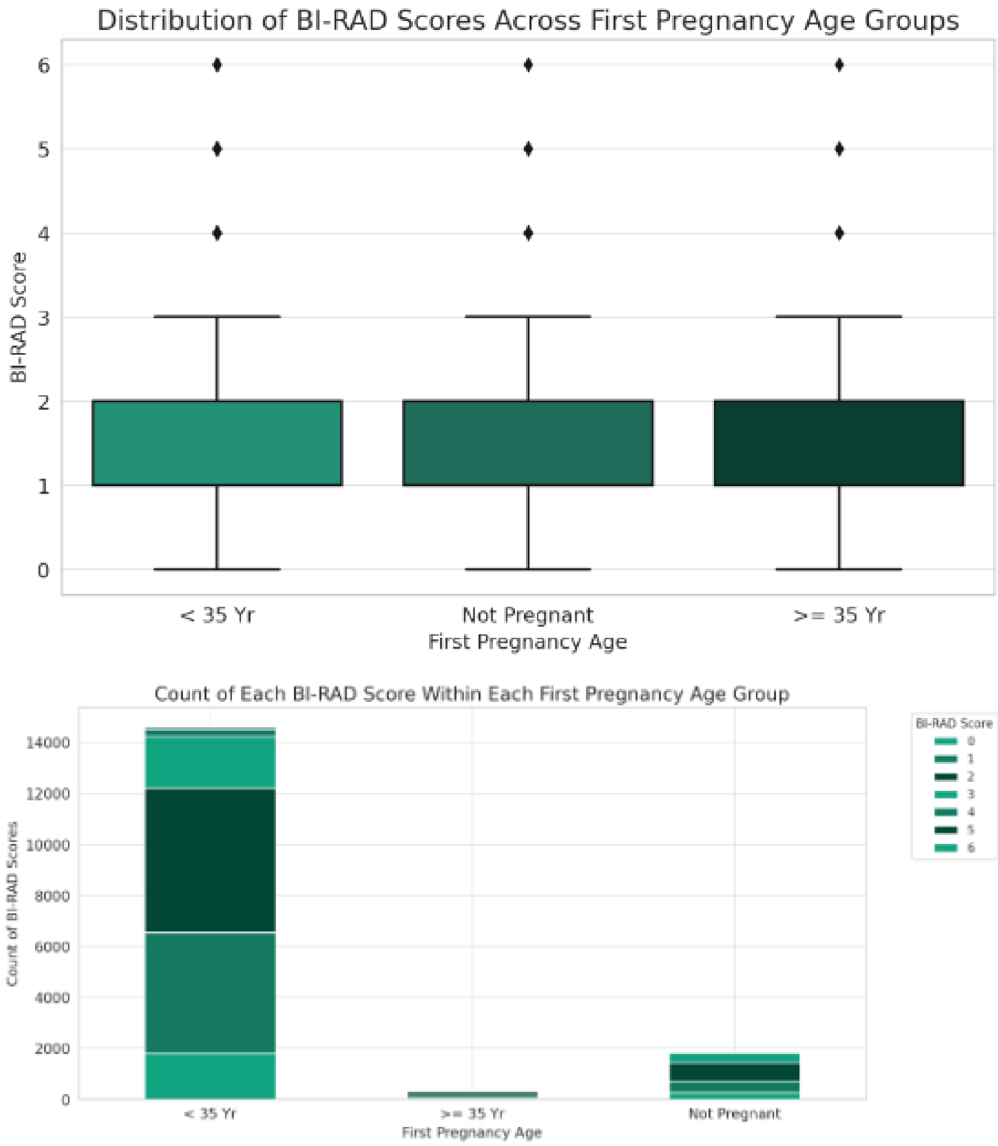

Figure 5.9: Distribution Of BI-RADS Scores across First Pregnancy Age Groups

The visualizations reveal the distribution of BI-RAD scores across first pregnancy age groups:

Box Plot: All groups have a median BI-RAD score of 2. The "Not Pregnant" and ">= 35 Yr" groups show more variability than the "< 35 Yr" group.

Bar Plot: Most BI-RAD scores in each group are 1, 2, and 3. The "< 35 Yr" group has the most pronounced distribution due to its larger size.

A Chi-squared test indicates a significant association between first pregnancy age and BI-RAD scores (Chi-squared: 107.52, p-value: 1.85×10−17, df: 12), suggesting that first pregnancy age influences BI-RAD scores, though causation is not implied.

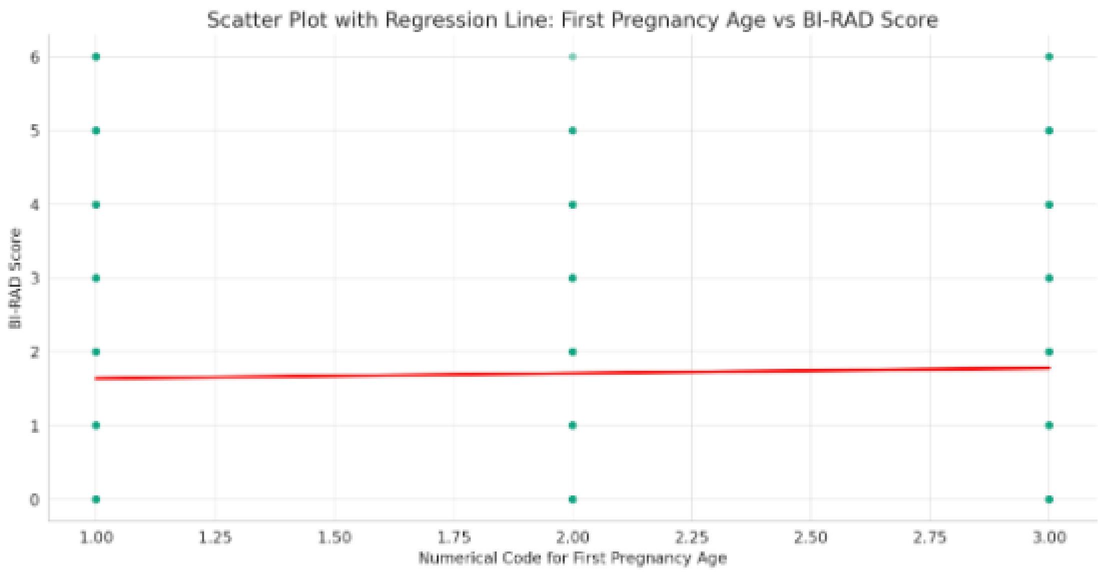

Figure 5.10: Scatter Plot with Regression Line: First Pregnancy Age vs BI-RAD Score

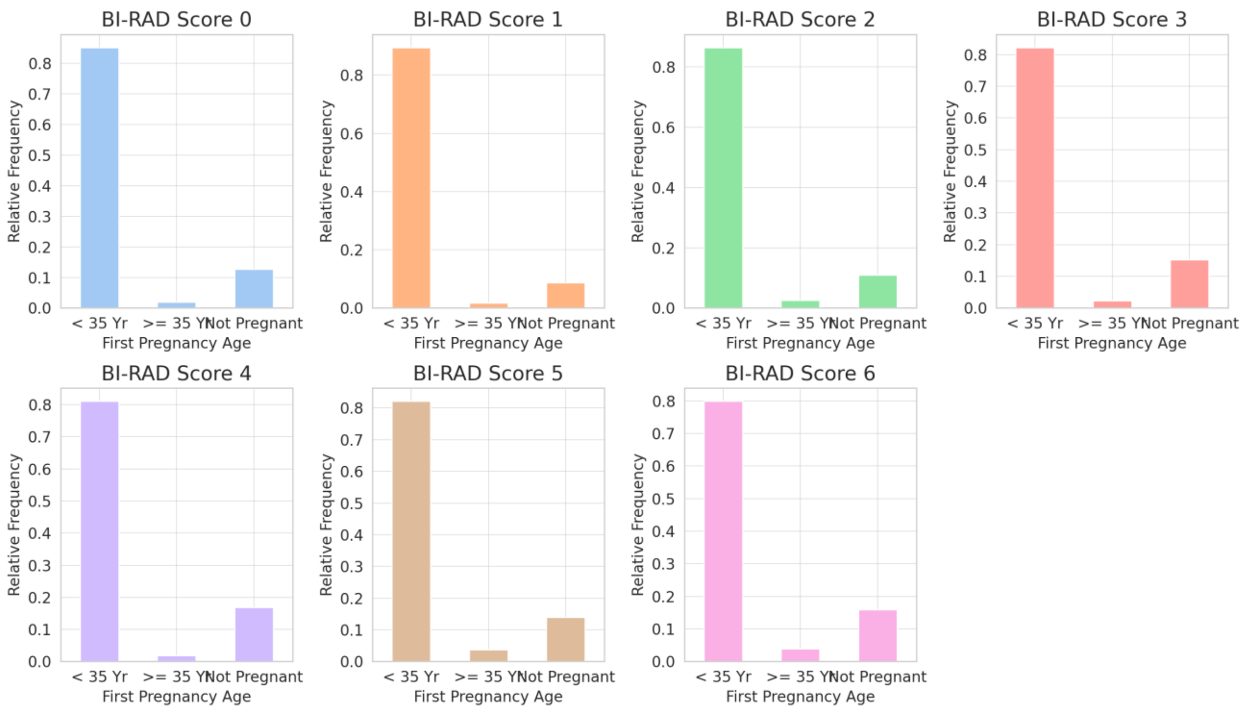

Figure 5.11: Bar Chart of Relative Frequency of BI-RAD Score vs First pregnancy Age

The scatter plot with a regression line shows a weak positive trend between first pregnancy age and BI-RAD score, indicating a slight increase in BI-RAD scores with older age at first pregnancy. The bar charts illustrate the relative frequency of each BI-RAD score (0-6) across different first pregnancy age groups ("< 35 Yr", ">= 35 Yr", "Not Pregnant"), highlighting how the occurrence of each score varies by age group.

5.2.9 BI-RADS vs Hormonal Replacement Therapy (HRT)

Statistical Tests: Depending on the distribution of the data, tests like Chi-square for categorical data or ANOVA if the BI-RAD scores behave numerically to understand if there are significant differences between groups.

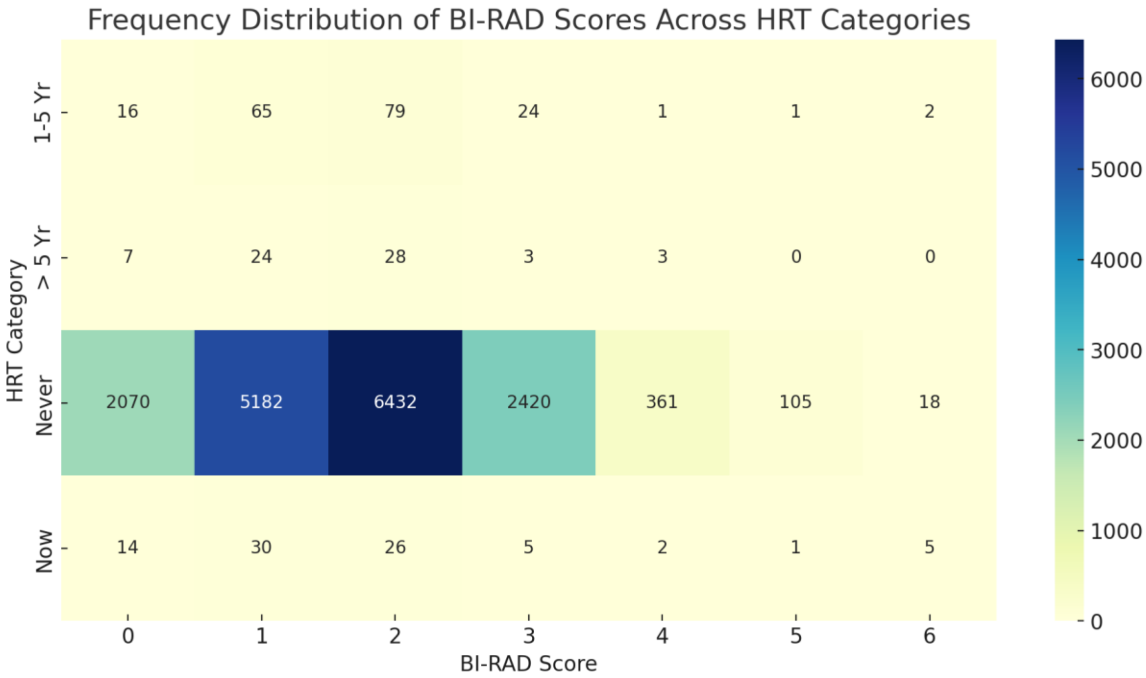

Figure 5.12: Frequency Distribution of BI-RAD Scores across HRT Categories

The analysis reveals interesting insights into the relationship between Hormonal Replacement Therapy (HRT) and BI-RAD scores:

Table 5.11: BI-RAD Scores across HRT Categories Descriptive Statistics

Statistic	Never	1-5 Year	5+ Year	Now
Most Observed BI-RAD Scores	0, 1, 2	Even distribution (0-3)	Even distribution (0-3)	Similar to '1-5 Year'
Mean BI-RAD Score	1.65	1.68	1.55	1.69

HRT use shows even BI-RAD scores. Current users have a higher mean (1.69) than never users (1.65), while long-term users have the lowest mean (1.55). Regular screening is essential.

Statistical Tests

The Chi-Square test shows a significant association between HRT status and BI-RAD score categories (Chi-square: 228.27, p-value: 2.07e-38), indicating that BI-RAD score distribution varies by HRT category. However, the ANOVA test reveals no significant difference in mean BI-RAD scores across HRT categories (ANOVA: 0.30, p-value: 0.83), suggesting that average BI-RAD scores are similar regardless of HRT status. The large number of cases in the "Never" category might influence these results.

5.2.10 BI-RADS vs Smoking

The dataset has "BI-RAD" scores (0-6) and "Smoking" with four categories: Neither, Both (Shisha and Cigarettes), Shisha, and Cigarettes. To analyse the relationship between smoking habits and BI-RAD scores, planned to: Analyse the frequency distribution of BI-RAD scores within each smoking category. Calculate the mean BI-RAD score for each smoking category. Perform statistical tests (Chi-square for categorical data, ANOVA for numerical data) to check for significant differences between groups.

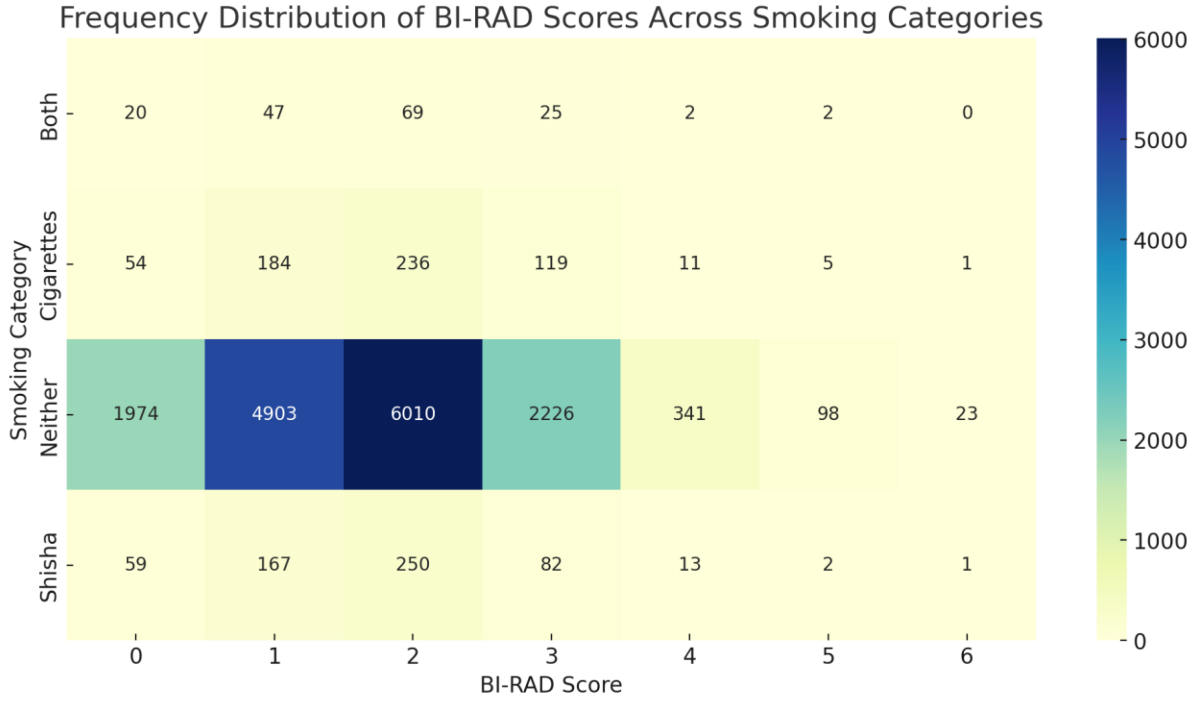

Figure 5.13: Frequency Distribution of BI-RAD Scores across Smoking Categories

The heatmap shows the distribution of BI-RAD scores within each smoking category. The statistical analysis of the relationship between smoking habits and BI-RAD scores reveals several insights:

Table 5.12: BI-RAD Scores across Smoking Categories Descriptive Statistics

Statistic	Neither	Cigarettes	Shisha	Both
Most Observed BI-RAD Scores	Highest number of cases across all scores	Higher concentration in lower scores (0-2)	Higher concentration in lower scores (0-2)	Fewer cases, similar distribution
Mean BI-RAD Score	1.64	1.78	1.71	1.68

The "Neither" category dominates all BI-RAD scores. Cigarette smoking slightly increases with higher BI-RAD scores, suggesting a potential correlation, while "Both" and "Shisha" categories remain low.

Statistical Tests

The Chi-Square test reveals a significant but weak association between smoking status and BI-RAD scores (Chi-square: 29.24, p-value: 0.046). The ANOVA test shows significant differences in mean BI-RAD scores across smoking categories (ANOVA: 4.59, p-value: 0.003), with cigarette smokers having slightly higher scores. The "Neither" category's large size may influence these results, so the effect size should be interpreted cautiously.

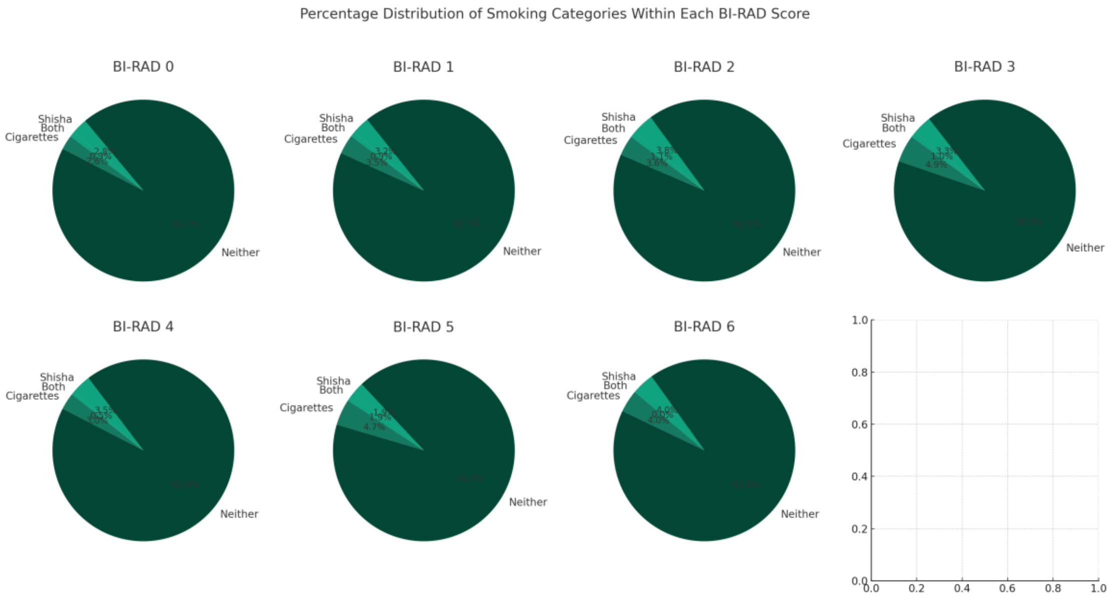

Figure 5.14: Percentage Distribution of Smoking Categories with BI-RAD Scores

Pie Charts: Each pie chart represents the percentage distribution of smoking categories for a specific BI-RAD score:

BI-RAD 0 to 6: The charts demonstrate the distribution of smoking habits across different BI-RAD scores. In all categories, the "Neither" group is the most significant segment, indicating that most individuals across BI-RAD scores do not engage in smoking. The other categories (Both, Cigarettes, and Shisha) show a smaller proportion. These visualizations provide a clear perspective on the prevalence of different smoking habits within each BI-RAD category, highlighting the dominance of the "Neither" group across all BI-RAD scores.

Table 5.13: Percentage Distribution of Smoking Categories with BI-RAD Scores

BI-RAD Score	Both	Cigarettes	Neither	Shisha
0	0.95%	2.56%	93.69%	2.80%
1	0.89%	3.47%	92.49%	3.15%
2	1.05%	3.59%	91.55%	3.81%
3	1.02%	4.85%	90.78%	3.34%
4	0.54%	3.00%	92.92%	3.54%
5	1.87%	4.67%	91.59%	1.87%
6	0.00%	4.00%	92.00%	4.00%

The "Neither" category dominates all BI-RAD scores. Cigarette smoking slightly increases with higher BI-RAD scores, suggesting a potential correlation, while the "Both" and "Shisha" categories remain low.

5.2.11 BI-RADS vs Menarche Age

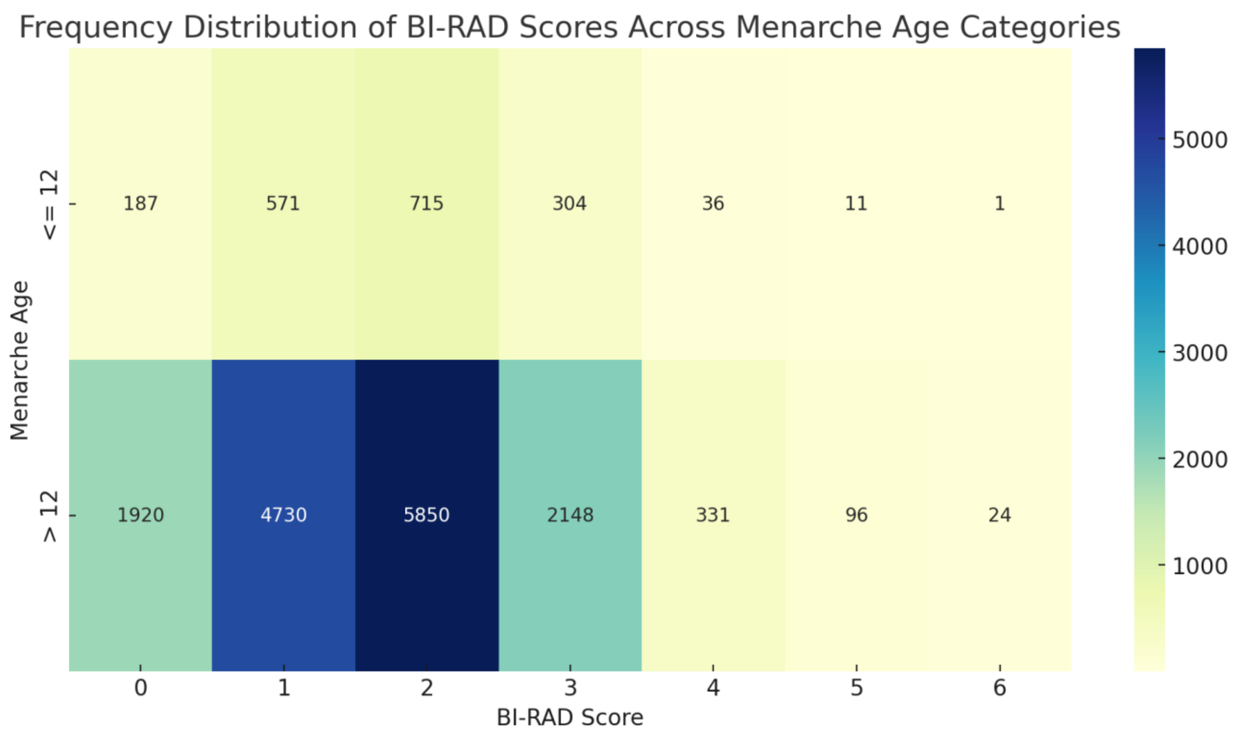

Figure 5.15: Frequency Distribution of BI-RAD Scores across Menarche Age Categories

The statistical analysis of the relationship between menarche age and BI-RAD scores reveals the following insights:

Table 5.14: BI-RAD Scores across Menarche Age Categories Descriptive Statistics

Menarche Age Category	Most Observed BI-RAD Scores	Mean BI-RAD Score
<= 12 years	Even distribution (0-3)	1.71
> 12 years	Even distribution with higher concentration	1.64

Menarche age ≤ 12 years shows an even BI-RAD score distribution (mean 1.71). Menarche age > 12 years has a higher score concentration (mean 1.64). Menarche age modestly impacts BI-RAD scores.

Statistical Tests: The Chi-Square test shows a significant association between menarche age and BI-RAD scores (Chi-square: 16.29, p-value: 0.012). The ANOVA test indicates significant differences in mean BI-RAD scores between menarche age groups (ANOVA: 6.77, p-value: 0.009), with later menarche associated with slightly lower scores. The differences are statistically significant but subtle.

5.2.12 BI-RADS vs Menopause Age

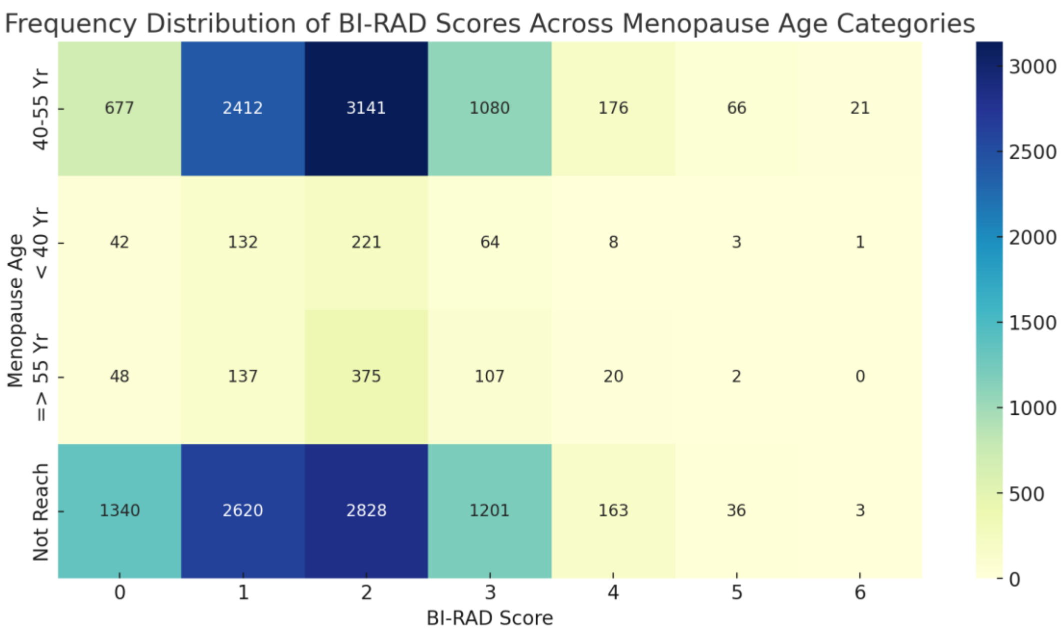

Figure 5.16: Frequency Distribution of BI-RAD Scores across Menopause Age Categories

The statistical analysis of the relationship between menopause age and BI-RAD scores provides the following insights:

Table 5.15: BI-RAD Scores across Menopause Age Categories Descriptive Statistics

Statistic	40-55 Year	< 40 Year	=> 55 Year	Not Reach
Most Observed BI-RAD Scores	Higher concentration across all scores	Higher frequency in lower scores (0-2)	Higher frequency in lower scores (0-2)	Mostly concentrated in lower scores
Mean BI-RAD Score	1.73	1.74	1.88	1.55

BI-RAD scores are higher for 40-55 (1.73) and ≥ 55 years (1.88) compared to < 40 (1.74) and Not Reached (1.55). Older and perimenopausal women need tailored screening.

Statistical Tests: The Chi-Square test shows a significant association between menopause age and BI-RAD scores (Chi-square: 368.06, p-value: 4.06e-67), indicating menopause age significantly affects the distribution and mean of BI-RAD scores, affecting breast imaging categorizations.

5.2.13 BI-RADS vs Family History

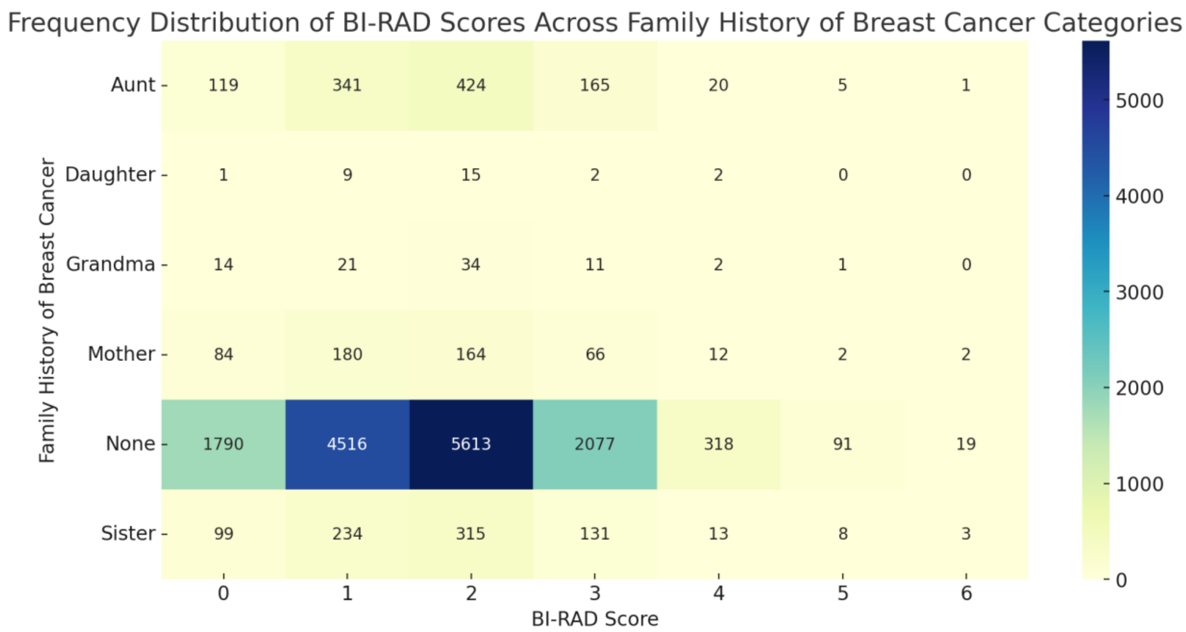

Figure 5.17: Frequency Distribution of BI-RAD Scores across Family History of Breast Cancer

The statistical analysis of the relationship between family history of breast cancer and BI-RAD scores provides the following insights:

Table 5.16: BI-RAD Scores across Family History of Breast Cancer Descriptive Statistics

Statistic	Aunt	Daughter	Grandma	Mother	None	Sister
Most Observed BI-RAD Scores	Varying distributions with fewer cases	Varying distributions with fewer cases	Varying distributions with fewer cases	Varying distributions with fewer cases	Highest number of cases across all scores	Varying distributions with fewer cases
Mean BI-RAD Score	1.67	1.83	1.63	1.52	1.65	1.7

Family history shows varied BI-RAD distributions but similar mean scores (1.52-1.83), indicating mostly benign findings. The "None" category has the highest case count, suggesting family history does not significantly impact average scores. Regular screening is crucial for everyone.

Statistical Tests: The Chi-Square test shows a trend towards an association between family history and BI-RAD scores (Chi-square: 41.67, p-value: 0.076), but it is not statistically significant. However, the ANOVA test indicates significant differences in mean BI-RAD scores across family history categories (ANOVA: 2.40, p-value: 0.035), suggesting family history modestly impacts BI-RAD scores. Overall, family history influences BI-RAD scores.

5.2.14 BI-RADS vs Governorate

The bar chart shows the distribution of BI-RAD categories across different directorates, with each bar representing a directorate and coloured segments indicating the counts of various BI-RAD categories. This provides a comparative view of BI-RAD distributions among directorates. For more detailed analysis, pie charts can be created for individual directorates to show the proportion of each BI-RAD category. If you have specific directorates of interest, pie charts can be generated for those.

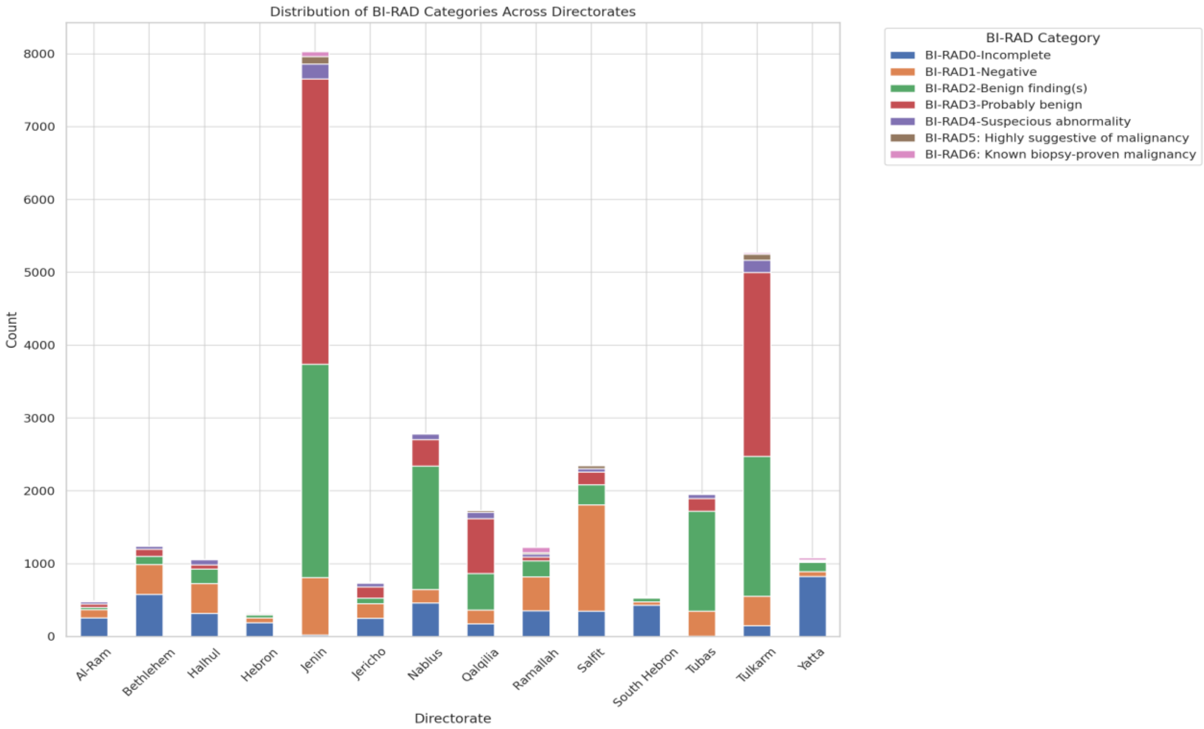

Figure 5.18: Distribution of BI-RAD Categories across Directorates

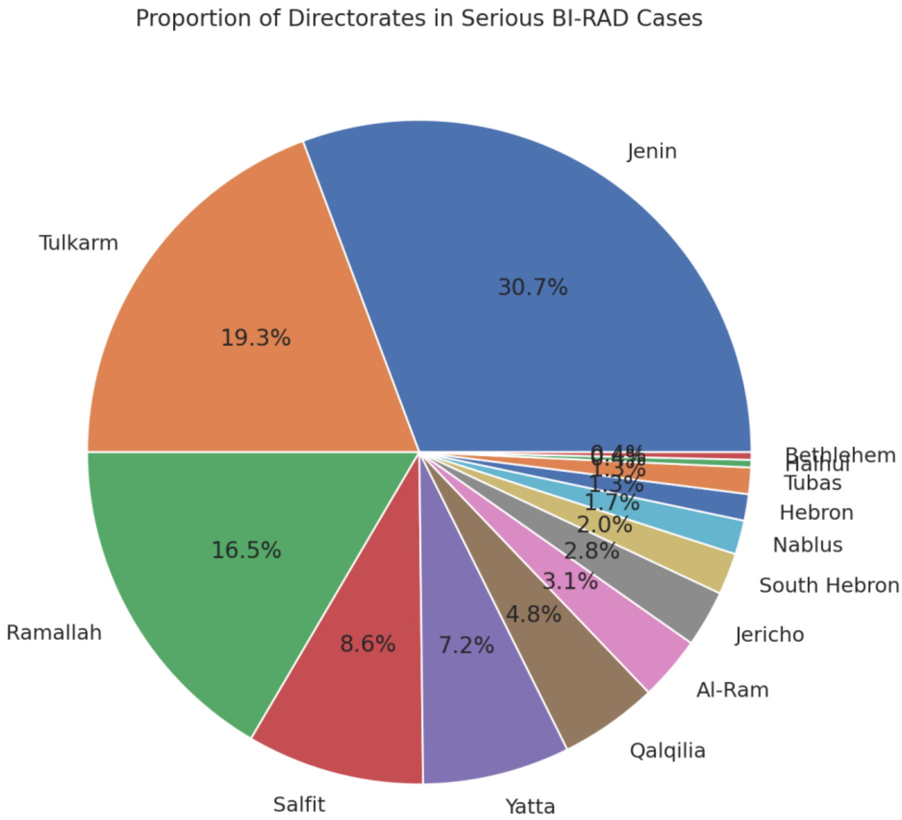

Figure 5.19: Proportion of Directorates in Serious BI-RAD Cases

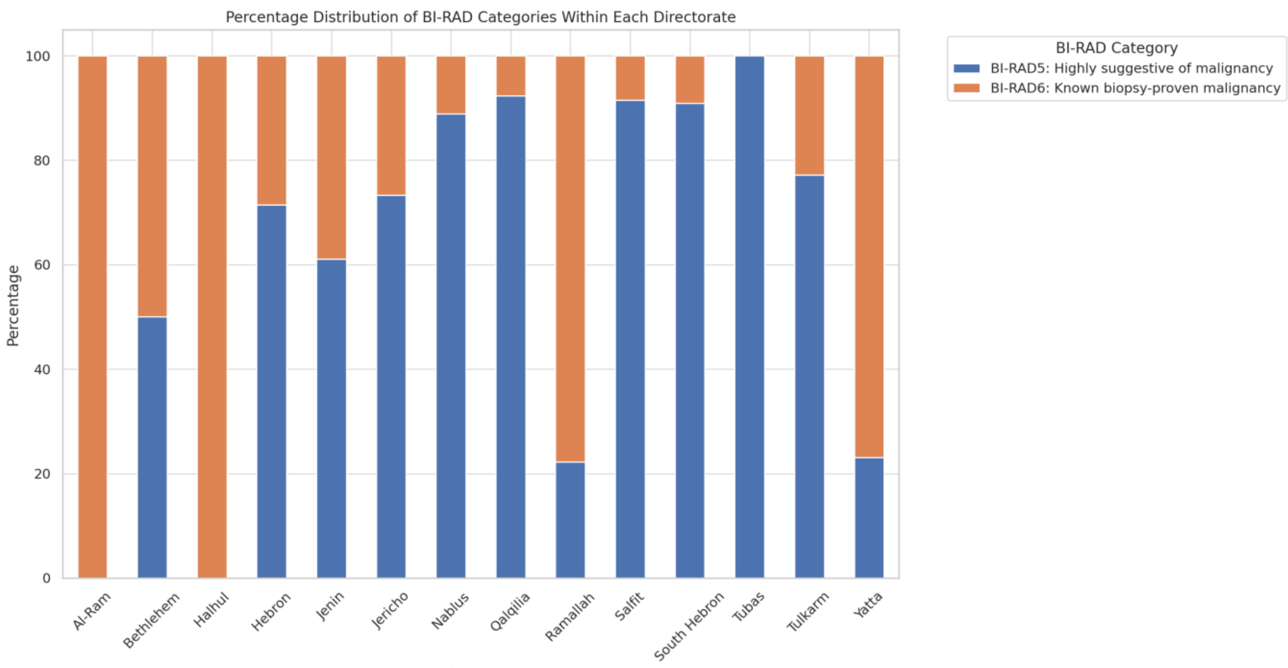

Figure 5.20: Percentage Distribution of BI-RAD Categories within each Directorate

The visualizations provide insights into the distribution of serious BI-RAD cases across different directorates and within each directorate:

Pie Chart (Proportion of Directorates in Serious BI-RAD Cases): This chart shows the relative proportion of serious BI-RAD cases across different directorates. Directorates with larger pie segments, such as Jenin and Tulkarm, have a higher number of serious BI-RAD cases compared to others. This indicates a varying prevalence of serious BI-RAD cases across different regions.

Bar Chart (Percentage Distribution of BI-RAD Categories Within Each Directorate): The chart shows the percentage distribution of "BI-RAD5: Highly suggestive of malignancy" and "BI-RAD6: Known biopsy-proven malignancy" across directorates. Some directorates, like Hebron, have more "BI-RAD5" cases, while others, like Al-Ram, have more "BI-RAD6" cases. This variation may reflect different diagnostic patterns or prevalence of certain conditions in these regions.

Interpretation: These visualizations are crucial for understanding the geographic distribution of serious BI-RAD cases, which can inform healthcare resource allocation and targeted intervention strategies. The differences in the percentage distributions of BI-RAD5 and BI-RAD6 categories suggest varying stages of cancer detection or differing levels of healthcare access and quality across the directorates. It is important for healthcare providers and policymakers to consider these variations to improve breast cancer screening, diagnosis, and treatment strategies in these areas.

5.3 Correlation between Breast Cancer Findings and Specific Mammogram Characteristics

This section explores the correlations between breast cancer findings and key mammogram characteristics, including BI-RAD scores, geographic variations, breast density, and patient age. It highlights the importance of mammography in detecting breast cancer and examines how these factors influence diagnosis and understanding of the disease, providing insights for tailored screening and diagnostic approaches.

5.3.1 Distribution of BIRAD Scores across Breast Cancer Registry

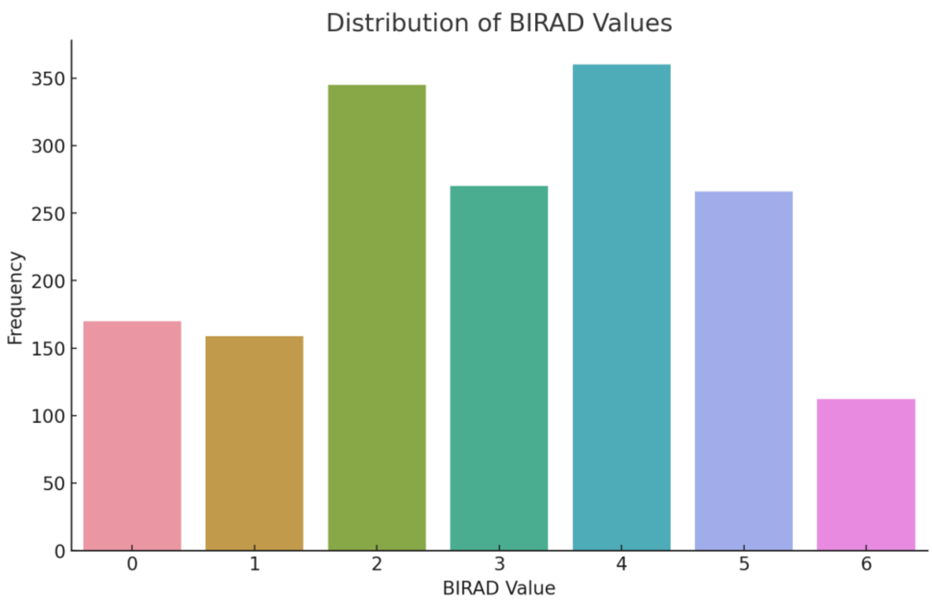

Figure 5.21: Distribution of BIRAD Scores across Breast Cancer Registry

Based on the analysis of the dataset, here are some key findings:

Table 5.17: BIRAD Scores across Breast Cancer Registry Descriptive Statistics

Statistic	**Count**	**Mean**	**Standard Deviation**	**Min**	**Max**
Value	1682	3.03	1.71	0	6

Table 5.18: BIRAD Scores across Breast Cancer Registry Counts

BIRAD 0	**BIRAD 1**	**BIRAD 2**	**BIRAD 3**	**BIRAD 4**	**BIRAD 5**	**BIRAD 6**
170	159	345	270	360	266	112

The mean BIRAD score is 3.03 among 1682 observations, with notable concentrations in BIRAD 4 and 5 (37%), indicating many cases require urgent follow-up and intervention.

Visualization: The bar plot shows that BIRAD values 2 (benign findings) and 4 (suspicious abnormalities) are the most common, followed by BIRAD 3 (probably benign) and 5 (highly suggestive of malignancy). This distribution indicates a diverse range of diagnoses among the patients.

Interpretation: The distribution of BIRAD values highlights breast cancer screening outcomes in the dataset. Higher scores (4, 5, 6), indicating a greater likelihood of malignancy, make up a significant portion, emphasizing mammography's role in cancer detection. BIRAD 0 and lower scores (1, 2, 3), suggesting benign conditions, are also well represented, showcasing the variety of breast imaging outcomes. This analysis provides an overview of the relationship between BIRAD values and breast cancer diagnoses in the patient cohort.

5.3.2 Diagnosis Age Range per Breast Cancer Patient

Diagnosis Age represents the ages at which patients were diagnosed with breast cancer. To analyse this, need to generate descriptive statistics and create visualizations to understand the distribution of diagnosis ages. This will help in exploring the relationship between age and breast cancer diagnosis.

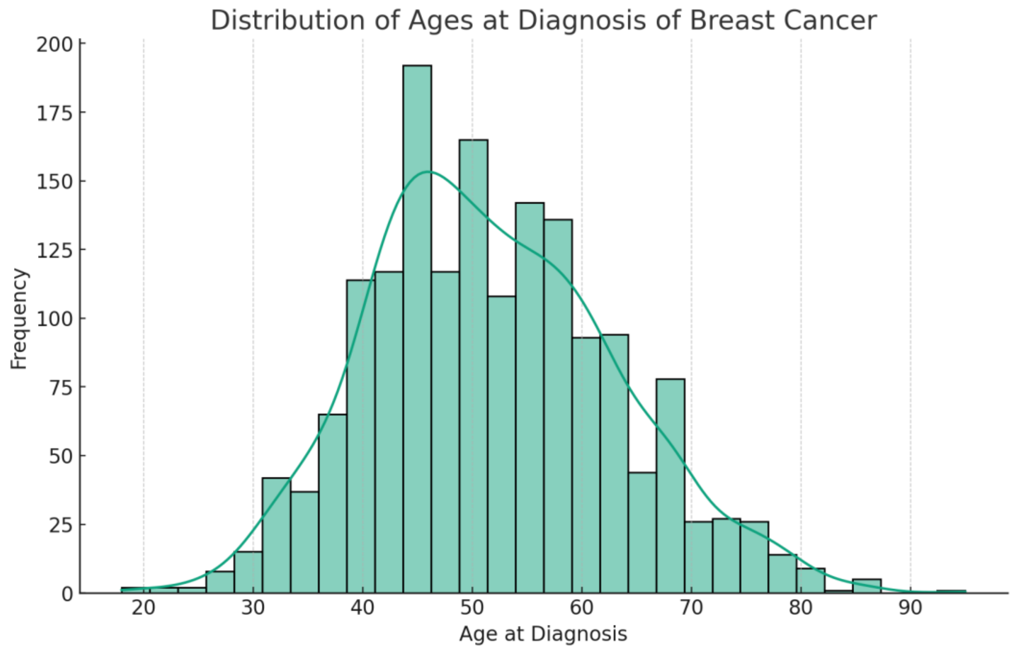

Figure 5.22: Distribution of Ages at Diagnosis of Breast Cancer

Table 5.19: Ages at Diagnosis of Breast Cancer Descriptive Statistics

Statistic	Count	Mean	Standard Deviation	Min	Max	25th Percentile	Median (50th Percentile)	75th Percentile
Value	1682	51.77	11.29	18	95	44	51	59

The average age at diagnosis for 1682 patients is 51.77 years, ranging from 18 to 95. Most diagnoses occur between 44 and 59 years, indicating the need for targeted screening across diverse ages.

Visualization: The histogram of breast cancer diagnosis ages shows a normal distribution slightly skewed towards older ages, peaking around the early fifties, consistent with the mean diagnosis age.

Interpretation: The age distribution shows breast cancer is most common in middle-aged to older adults, particularly those in their 40s to 60s. Cases emphasize the need for awareness and screening across all age groups. The median age of 51 years marks the peak diagnosis frequency, highlighting age as a significant factor in breast cancer occurrence and detection.

5.3.3 Breast Cancer and Breast Density Combination

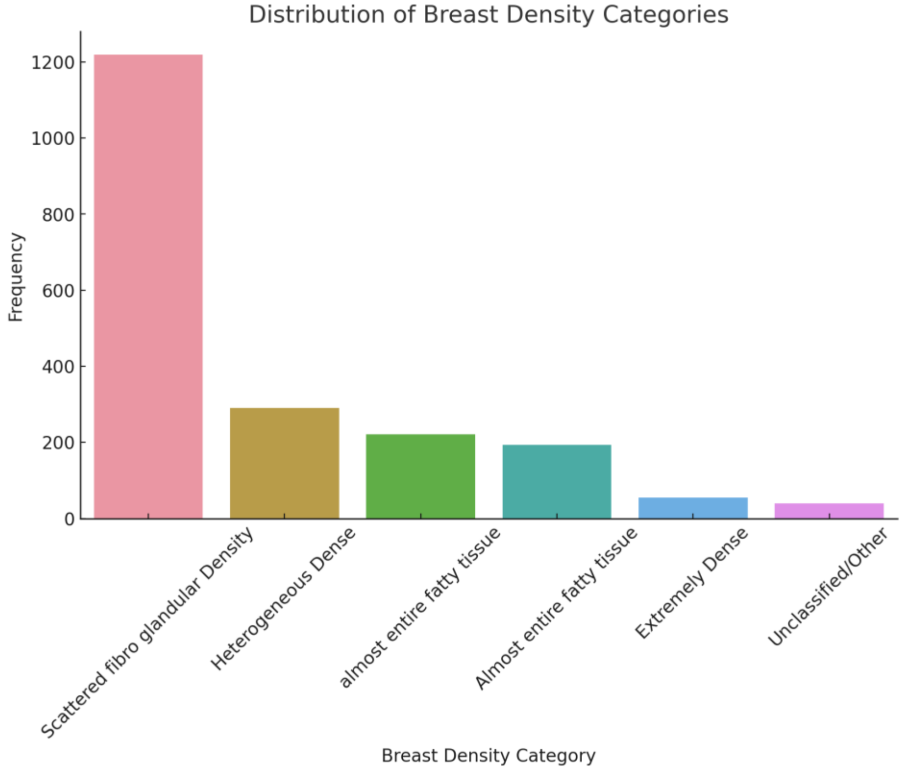

Figure 5.23: Distribution of Breast Density Categories across Breast Cancer Patients

Table 5.20: Breast Density Categories across Breast Cancer Patients Descriptive Statistics

Statistic	**Scattered Fibro Glandular Density**	**Heterogeneous Dense**	**Almost Entire Fatty Tissue**	**Extremely Dense**	**Unclassified/Other**
Frequency	1220	291	415	55	39

The bar plot shows that "Scattered Fibro Glandular Density" is the most prevalent breast density category, followed by "Heterogeneous Dense" and "Almost Entire Fatty Tissue." The "Extremely Dense" category is less common. This distribution suggests higher detection rates in patients with "Scattered Fibro Glandular Density" due to clearer mammogram results, while the lower frequency of "Extremely Dense" tissue reflects challenges in cancer detection. Understanding these distributions is crucial for improving breast cancer screening and detection strategies, especially in denser breast tissues.

5.4 Integration between Mammogram and Breast Cancer Registries

Table 5.21: BIRADS 5 and 6 and Breast Cancer Detection Descriptive Statistics

Statistic	Total Cases with BIRAD 5 or 6	Cases Detected as Breast Cancer	Completeness (%)
Value	465	387	83.23

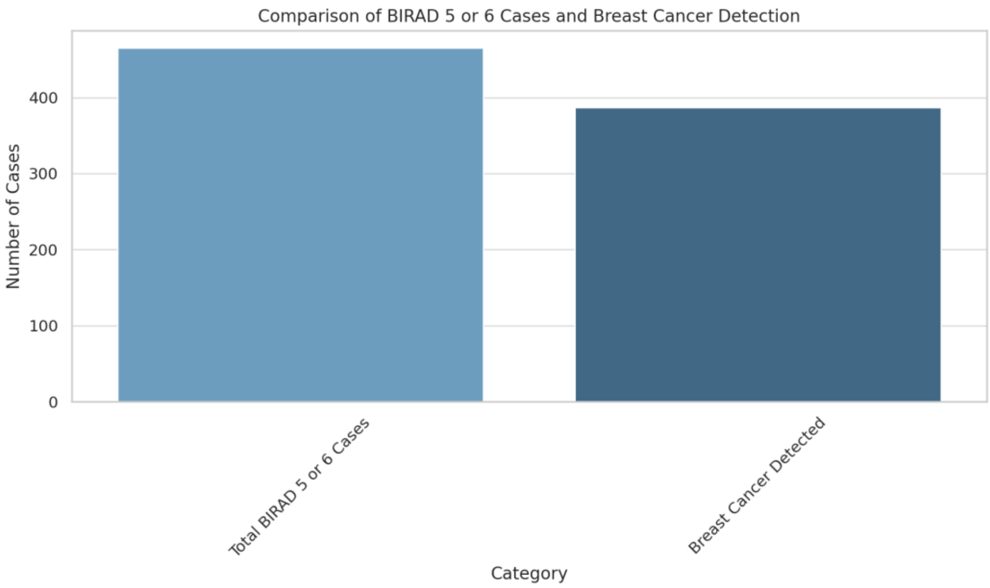

Figure 5.24: Comparison of BIRADS 5 and 6 and Breast Cancer Detection.

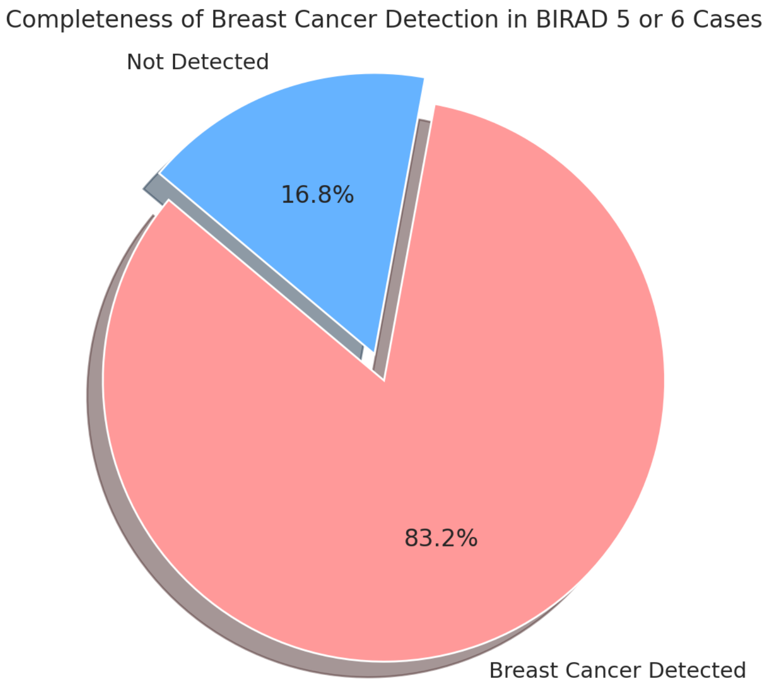

Figure 5.25: Completeness of Breast Cancer Detection in Mammogram BIRADS 5 and 6

The bar chart compares the total cases with a BIRAD score of 5 or 6 to the number detected as breast cancer, illustrating the proportion of high-risk cases confirmed as breast cancer. Next, a pie chart will visually represent the completeness percentage, showing the proportion of confirmed cases versus those not confirmed. The pie chart shows the completeness of breast cancer detection for cases with a BIRAD score of 5 or 6, illustrating the proportion confirmed as breast cancer versus those not detected. These visuals and the completeness percentage highlight the alignment between the mammogram and breast cancer registries, revealing strong detection coverage and identifying gaps where some high-risk cases may be missed.

5.5 Timeliness of Migrating Mammogram High BIRAD Scores into Breast Cancer Registry

To analyze the timeliness and reachability between the breast cancer and mammogram registries, the plan includes Calculating summary statistics (mean, median, range) of the time differences. Creating visualizations to understand the distribution of these time differences.

The summary statistics for the time difference in months between mammogram enrollment and breast cancer registry incidence are as follows:

Table 5.22: Mammogram Entry and Breast Cancer Reachability Timeliness Statistics

Statistic	**Count**	**Mean (months)**	**Standard Deviation (months)**	**Minimum (months)**	**25th Percentile (months)**	**Median (50th Percentile) (months)**	**75th Percentile (months)**	**Maximum (months)**
Value	694	5.28	12.25	0	0	1	2	65

This data suggests that while most cases are recorded in both registries within a brief time frame, there are some outliers with significantly longer delays. Next, a histogram is created to visualize the distribution of these time differences, which will help in understanding the spread and concentration of the data points.

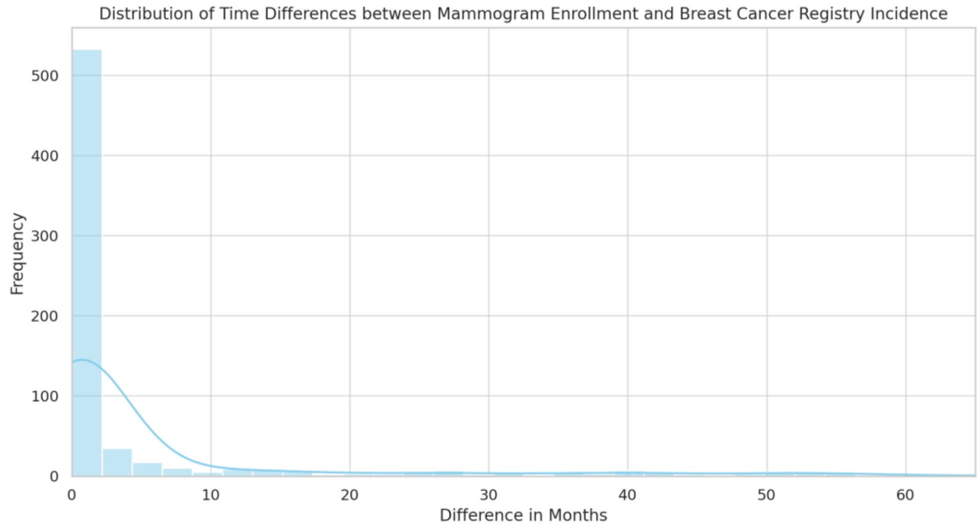

Figure 5.26: Timeliness between Mammogram Registration and Breast Cancer Reachability

The histogram illustrates the distribution of time differences (in months) between mammogram enrollment and breast cancer registry incidence. This visualization helps to understand the frequency and spread of these time differences. To further enhance the analysis, a box plot was created, which is useful for visualizing the range, interquartile range, median, and potential outliers in the dataset. This will provide additional insights into the timeliness and reachability between the two registries.

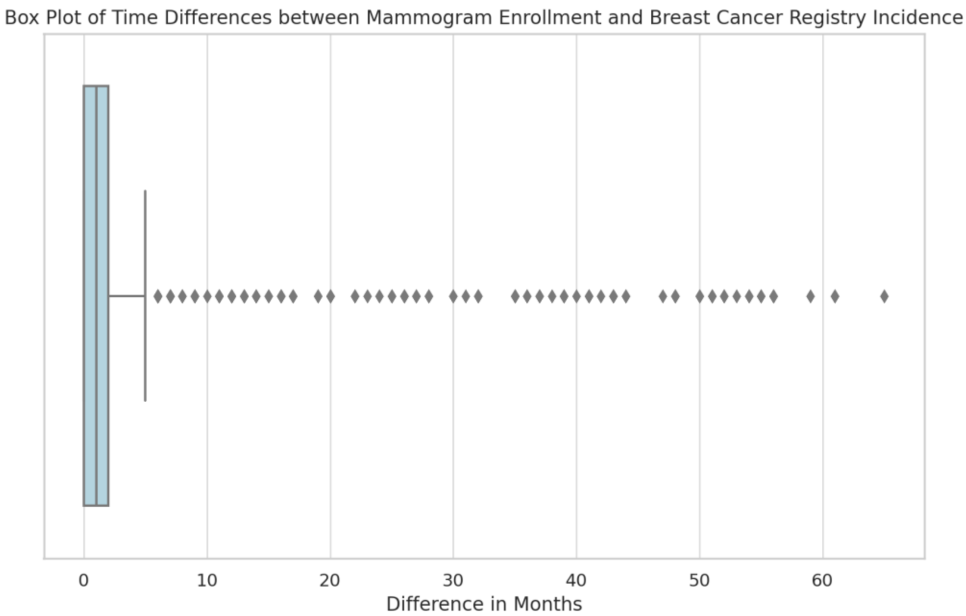

Figure 5.27: Box Plot of Time Diff between Mammogram & Breast Cancer Registry

The box plot shows the time differences between mammogram enrollment and breast cancer registry incidence, highlighting the median, quartiles, and outliers. Combined with summary statistics, it provides a comprehensive analysis of timeliness and reachability between the registries. While most cases are promptly recorded, significant variations and delays exist, indicating areas for improvement in registration and data management processes.

5.6 Distribution of Breast Cancer Cases among West Bank Governorates

To analyze breast cancer distribution among governorates, follow these steps: calculate the count and percentage of cases for each governorate, create a table with these counts and percentages, and visualize the data using a pie chart. For example, Jenin has the highest number of cases (451), accounting for 31.60% of the total. The next top governorates are Tubas (9.74%), Nablus (8.06%), and Tulkarm (7.92%). A pie chart will visually represent these proportions.

The pie chart visually represents the distribution of breast cancer cases among different governorates. This chart highlights the proportion of cases in each governorate relative to the total, providing a clear and immediate understanding of the geographical spread of breast cancer cases. This comprehensive analysis, with the provided table and pie chart, offers valuable insights into the regional distribution of breast cancer cases, which can be crucial for public health planning and resource allocation.

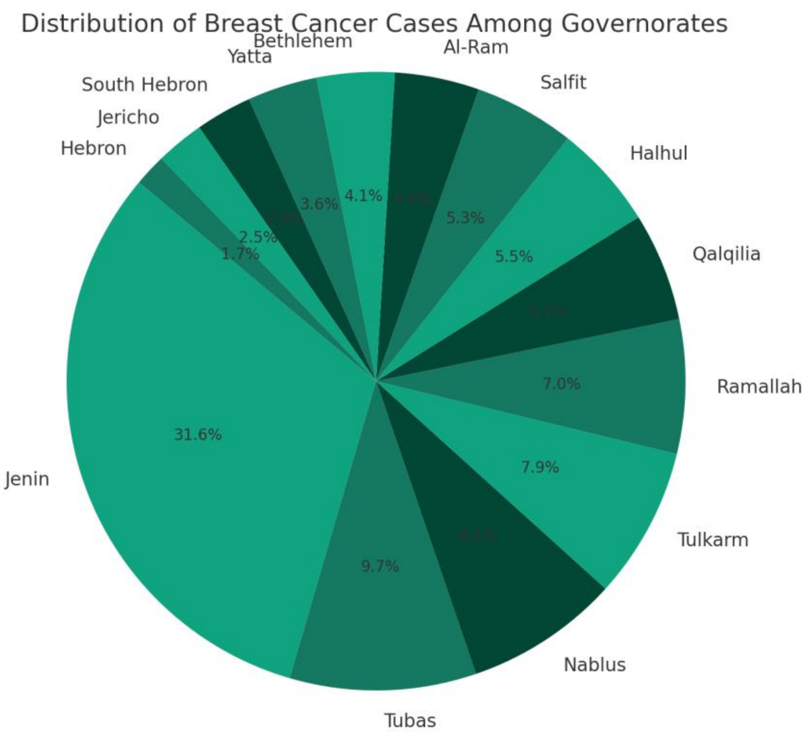

Figure 5.28: Distribution of Breast Cancer Cases among Governorates

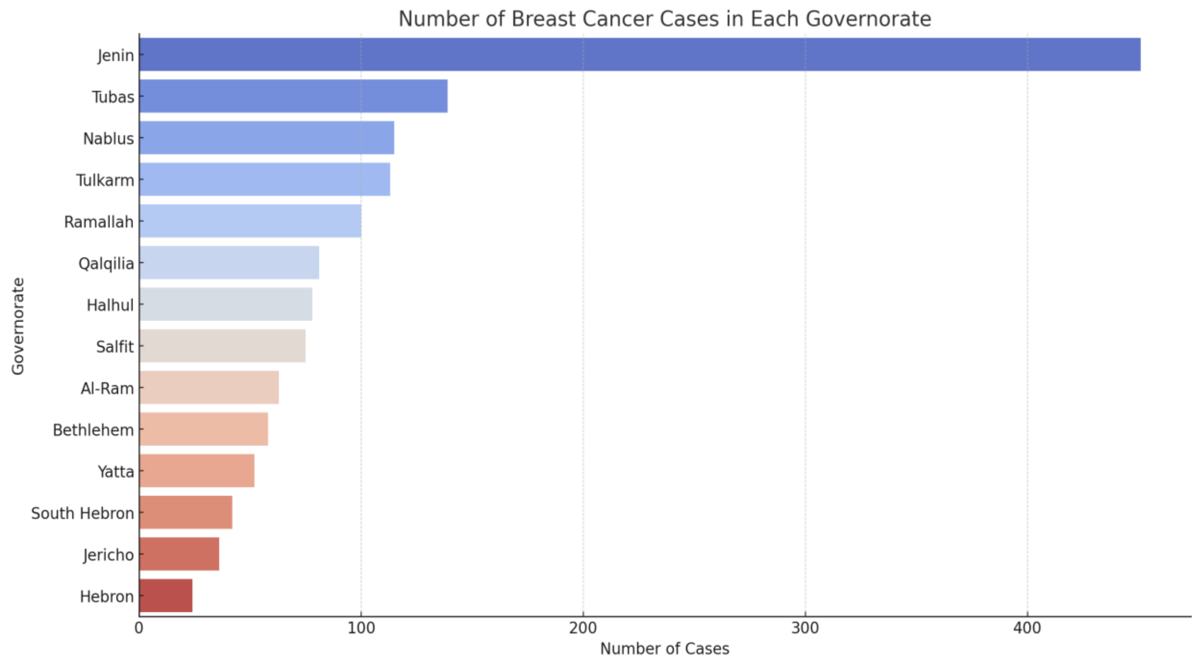

Figure 29: Number of Breast Cancer Cases per Governorate

Bar Chart: This chart visually represents the number of breast cancer cases in each governorate, providing a clear comparison of case counts across regions. You can download and view the bar chart using this link: Download Bar Chart.

Table 5.23: Breast Cancer Counts and Percentages per Governorate

Governorate	Count
Jenin	451
Tubas	139
Nablus	115
Tulkarm	113
Ramallah	100
Qalqilia	81
Halhul	78
Salfit	75
Al-Ram	63
Bethlehem	58
Yatta	52
South Hebron	42
Jericho	36
Hebron	24

Table 5.23: A table displaying the count and percentage of breast cancer cases for the fourteen governorates. This table offers a detailed view of the data in a structured format. These additional visualizations and the table complement the pie chart and provide a comprehensive understanding of the geographical distribution of breast cancer cases.

Performance of Mammogram in Breast Cancer Detection per Governorate: Here is the bar chart representing the performance of mammograms in different governorates. The green bars indicate the percentage of cases where mammograms detected cancer ("Yes"), the red bars represent the percentage of cases where mammograms did not detect cancer ("No"), and the grey bars show the percentage of uncertain cases ("May Be"). Each set of bars corresponds to a specific governorate, providing a clear visualization of the data.

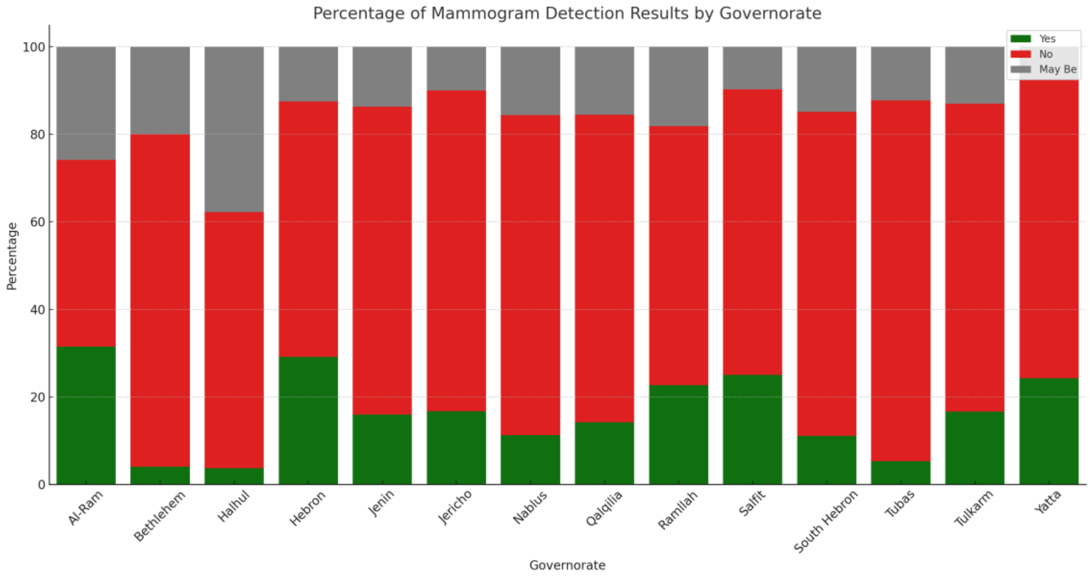

Figure 5.30: Performance of Mammogram per Governorate

Table 5.24: Mammogram Detection Results per Governorate

Governorate	Detected	Missed	Probable
Jenin	118 (16%)	521 (70%)	101 (14%)
Tubas	10 (6%)	155 (82%)	23 (12%)
Nablus	31 (11%)	201 (73%)	43 (16%)
Tulkarm	37 (17%)	157 (70%)	29 (13%)
Ramallah	60 (23%)	157 (59%)	48 (19%)
Qalqilia	22 (14%)	109 (70%)	24 (15%)
Halhul	3 (4%)	48 (58%)	31 (38%)
Salfit	36 (25%)	94 (65%)	14 (10%)
Al-Ram	28 (32%)	38 (42%)	23 (26%)
Bethlehem	5 (4%)	95 (76%)	25 (20%)
Yatta	16 (24%)	45 (68%)	5 (8%)
South Hebron	9 (11%)	60 (74%)	12 (15%)
Jericho	10 (17%)	44 (73%)	6 (10%)
Hebron	7 (29%)	14 (58%)	3 (13%)

Table 5.24 provides a detailed breakdown for each governorate, allowing for a comprehensive analysis of the mammogram detection results.

Chapter 6: Discussion

6.1 Introduction

This chapter discusses how advanced mammography technology aids breast cancer detection and treatment in Palestine. It covers identifying key mammogram features, linking mammogram data with cancer records, quickly updating high-risk cases, and mapping breast cancer cases in the West Bank. These analyses highlight the importance of modern technology in improving breast cancer outcomes in Palestine.

6.2 Importance of Mammogram Features in Defining High BIRADS

This research focuses on determining the significance of various mammogram attributes, including age, menarche, medical history, breast density, and BMI, in predicting breast cancer risk. Employing machine learning techniques, specifically logistic regression and correlation analysis, the study calculates the weight of each mammogram attribute. The objective is to understand the predictive power of these attributes for high BI-RADS values, indicative of a higher likelihood of malignant breast lesions.

Table 6.1 ranks the factors based on their correlation with BIRAD scores, from the highest positive correlation to the most negative. Remember, these correlations do not imply causation but indicate the strength of association between each factor and the BIRAD scores.

Governorate: This column indicates the geographical region or governorate. The relation between the governorate and the BIRAD score could be influenced by environmental factors, healthcare access, and regional health practices. The relative proportions of serious BI-RAD cases among different directorates. Directorates like Jenin, Tubas and Tulkarm have larger segments, indicating a higher prevalence of serious BI-RAD cases in these areas. This highlights the uneven distribution of serious BI-RAD cases across various regions.

Table 6.1: Importance of Mammogram Features in Defining High BIRADS

Predictor	Correlation with BIRAD
Governorate	0.20177
Diabetes	0.10021
First Pregnancy Age	0.04635
Heart Diseases	0.03884
Previous Mammogram	0.01768
Family History of Breast Cancer	-0.00009
Hormonal Replacement Therapy (HRT)	-0.00119
Smoking	-0.01059
BMI	-0.01214
Hypertension	-0.01262
Density	-0.01704
Menarche Age	-0.02079
Menopause	-0.08281
Contraceptive Use	-0.01043
Chronic Diseases	-0.01101

- **BMI (Body Mass Index):** BMI is a measure of body fat based on height and weight. Higher BMI may be associated with increased risk factors for breast diseases, affecting the BIRAD score. The analysis of the relationship between BI-RADS scores and Body Mass Index (BMI) reveals that while there is variation in average BMI across different BI-RADS categories, with the highest averages in categories 2 and 3 and the lowest in 5 and 6, the overall correlation between BMI and BI-RADS scores is very weak (correlation coefficient approximately 0.046). This indicates a slight tendency for BMI to increase with higher BI-RADS scores, but the relationship is not strong and varies significantly among individuals. Consequently, this suggests that factors other than BMI might be more significant in determining BI-RADS scores.
- **Menarche Age:** The age at which menstruation begins. Earlier menarche age has been linked in some studies to a higher risk of breast cancer, influencing the BIRAD score. The Chi-square and ANOVA tests indicate a significant but modest relationship between the age of menarche and BI-RAD scores in breast imaging. The age at which menstruation begins is associated with variations in BI-RAD scores, with a slightly lower average score

observed in individuals who experienced menarche after 12 years of age. Although statistically significant, the impact of menarche age on BI-RAD scores is small, suggesting a subtle influence on the categorization of breast imaging findings.

- **First Pregnancy Age:** Age during the first pregnancy. Earlier or later ages of the first pregnancy might have different implications for breast health, impacting the BIRAD score. The analysis shows a statistically significant association between the age at first pregnancy and BI-RAD scores, with a p-value much lower than 0.05, indicating that BI-RAD scores are related to the age of first pregnancy. First pregnancy ages of more than 35 years are associated with higher BIRAD scores.
- **Contraceptive Use**: Contraceptive use could have hormonal implications that might affect breast tissue and thus influence the BIRAD score. The analysis reveals a weak positive correlation (coefficient of 0.083) between BI-RAD scores and contraceptive use, implying a slight increase in BI-RAD scores with more extended contraceptive use. However, the interpretation of this relationship requires further detailed analysis due to the categorical nature of the contraceptive use data.
- **Menopause Age**: Indicates the menopausal status of the individual. Post-menopausal women have different breast cancer risk profiles compared to pre-menopausal women, which might affect their BIRAD scores. The Chi-square and ANOVA tests indicate a strong association between menopause age and BI-RAD scores in breast imaging. Women experiencing menopause at or after 55 years tend to have higher average BI-RAD scores. This significant difference suggests that the age at menopause notably influences BI-RAD scores, impacting how breast imaging findings are categorized.
- **Hormonal Replacement Therapy (HRT):** The use of HRT can impact breast tissue density and potentially affect the risk of breast diseases, thereby influencing the BIRAD score. The ANOVA test results suggest similar average BI-RAD scores across these

groups, implying consistent average risk levels despite different HRT statuses. The high number of cases in the "Never" HRT category could influence these statistical interpretations.

- **Smoking:** Smoking status is a known risk factor for many diseases and could indirectly affect breast health, thus influencing the BIRAD score. The analysis indicates a moderate association between smoking status and BI-RAD scores, with cigarette smokers showing slightly higher average scores. However, the large number of non-smoking cases could affect these results. Therefore, while there is some link between smoking and BI-RAD scores, its significance should be interpreted with caution due to the uneven sample distribution.
- **Chronic Diseases:** The presence of chronic diseases like hypertension or diabetes could indicate an overall health status that might indirectly affect breast health and the BIRAD score. The extremely low p-value in the analysis leads to the rejection of the null hypothesis, indicating a statistically significant association between BI-RADS scores and the presence of chronic diseases. The distribution of BI-RADS scores differs notably between individuals with chronic diseases and those without inversely with higher BIRADS.
- **Hypertension**: High blood pressure might be associated with certain health conditions that could influence breast health. The analysis indicates a very weak negative correlation (coefficient -0.018) between BI-RAD scores and hypertension, suggesting minimal association. However, due to the low prevalence of hypertension in the dataset, this finding should be interpreted cautiously, and further analysis with a more balanced dataset or advanced methods may be needed for a clearer understanding.
- **Diabetes**: Diabetes might influence hormonal balance and overall health, potentially impacting breast health and the BIRAD score. The analysis shows a weak positive

correlation (coefficient of 0.100) between BI-RAD scores and diabetes, indicating a slight tendency for higher BI-RAD scores to be associated with the presence of diabetes, although this relationship is not strong.

- **Heart Diseases**: The presence of heart diseases could indicate an overall different health profile, which might have implications for breast health.

  The correlation between BI-RAD scores and heart diseases is very weak (coefficient 0.038), indicating a minimal association. Further, due to the low prevalence of heart diseases in the dataset, more detailed analysis may be necessary.

- **Family History of Breast Cancer**: A family history of breast cancer is a known risk factor and can significantly influence the BIRAD score. Chi-square and ANOVA tests suggest a statistically significant but impact of family history of breast cancer on BI-RAD scores, with variations across different family history groups. Family history may be one of several factors influencing BI-RAD scoring.

- **Density**: Breast density is a direct indicator of breast tissue composition. Different densities can affect the visibility of lesions in mammography and are directly related to the BIRAD score. The very low p-value (0.0) suggests a statistically significant association between BI-RAD classifications and breast density categories, indicating a strong relationship between these variables. Observed frequencies in BI-RAD classifications vary significantly from expected frequencies under the assumption of independence.

6.3 Linking Mammogram Data with Breast Cancer Records (Completeness)

The completeness percentage of about 83.23% indicates that a significant majority of high-risk BIRAD cases (scores 5 or 6) were confirmed as breast cancer cases, as per the breast cancer registry. However, there remains a portion (around 16.77%) of cases that were not matched, suggesting room for improvement in registry integration or data capturing processes. To better illustrate these findings, visual representations such as pie charts are used, showing the distribution of BIRAD 5 or 6 cases and the number of these cases identified as breast cancer.

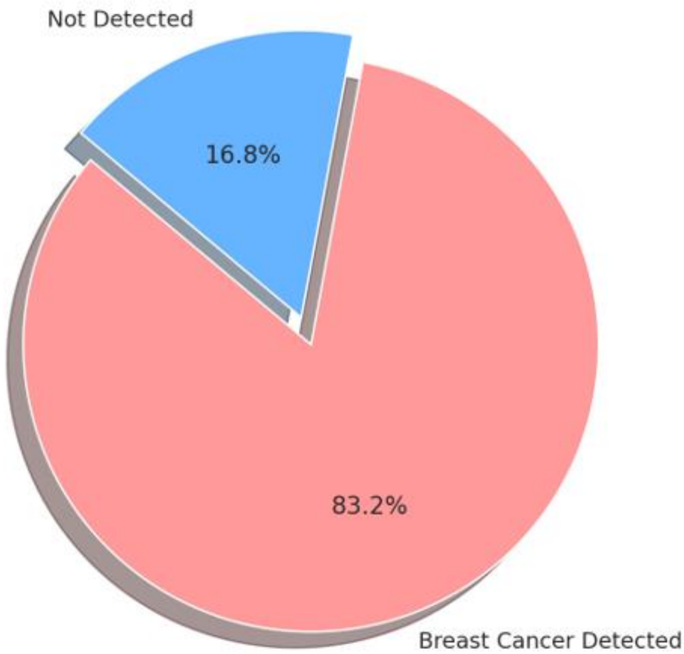

Figure 6.1: Mammogram High BIRAD Scores Completeness in Breast Cancer

The pie chart visually depicts the completeness of breast cancer detection in cases with a BIRAD score of 5 or 6, showing the proportion of cases confirmed as breast cancer against those not detected. Together with the completeness percentage, these visualizations offer a detailed view of how well the mammogram registry aligns with the breast cancer registry, highlighting effective detection in high-risk cases while also pointing out a gap where some cases may not be sufficiently captured or recorded in the breast cancer registry.

6.4 Timeliness between High BIRAD Mammogram and Breast Cancer Evidence

Based on the analysis of the time difference between high BIRADS value measurement and breast cancer registry attendance, here are five key points incorporating statistical insights:

- Early Detection and Treatment: Approximately 42.42% of cases were attended to within one month of a high BIRADS score being recorded. This suggests a prompt response in these cases, which is crucial for early treatment and better outcomes.
- Healthcare System Efficiency: The data shows that 34.49% of cases had a time difference of 1-3 months. While this is a reasonable timeframe, it may indicate some delays in the referral or diagnostic process. The longest delays (>3 months) were seen in about 23.09% of cases, highlighting areas where the healthcare system could potentially be improved.
- Patient Follow-up and Compliance: The median time difference is one month, suggesting that half of the patients are brought into the breast cancer registry quickly after their mammogram. However, a range in response times (minimum 0 months, maximum 65 months) indicates variability in patient follow-up and compliance.
- Statistical Analysis and Public Health Policy: With a mean time, difference of approximately 5.28 months and a standard deviation of 12.25 months, there is significant variation in the time taken to attend to high-risk cases. This variability could inform public health policies to target areas or groups with longer wait times.
- Quality of Care: The fact that 75% of cases are attended to within two months is positive, but the presence of extreme cases (up to 65 months) might raise concerns about the quality of care and delayed diagnosis or treatment in certain instances. This variation suggests that while many patients receive timely care, there are still significant gaps that need to be addressed. These insights indicate that while a substantial portion of high-risk cases are attended to promptly, there is still a considerable percentage of cases experiencing delays, highlighting areas for potential improvement in healthcare delivery and patient management.

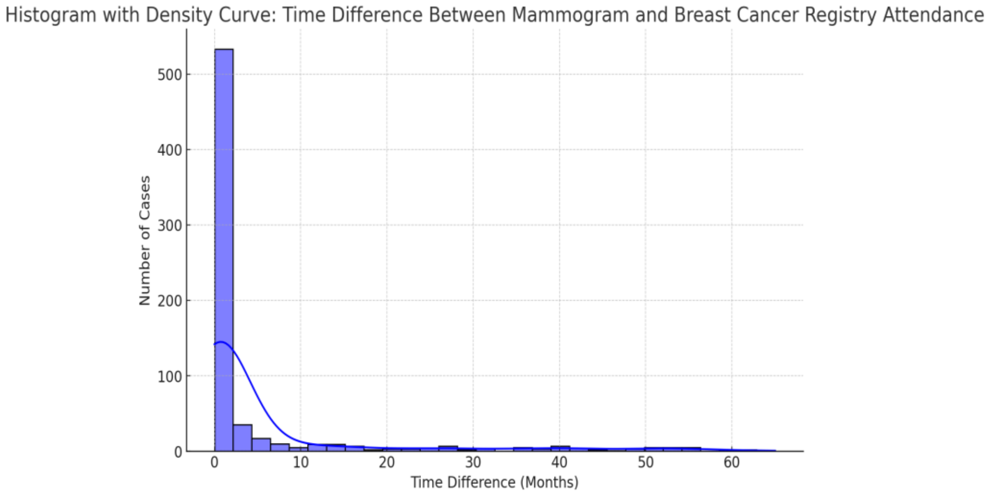

Figure 6.2: Time Difference between Mammogram and Breast Cancer Registry Attendance

Here is a histogram with a density curve overlay for the time difference between mammogram high BIRADS value measurement and breast cancer registry attendance. This chart combines the histogram's discrete frequency representation with a smooth density estimate, providing a comprehensive view of the data distribution.

6.5 Distribution of Breast Cancer Cases among West Bank Governorates

Based on the dataset of breast cancer cases across different governorates, here is a comprehensive discussion incorporating the statistics and notes for each point:

1. Distribution Across Governorates: Jenin reports the highest number of cases (451), significantly more than other cities. This could indicate a higher prevalence or better detection/reporting mechanisms in Jenin. Tubas, Nablus, and Tulkarm follow with 139, 115, and 113 cases respectively, suggesting these areas also have a notable number of cases. The dataset includes fourteen different cities, with Hebron reporting the fewest cases (24).
2. Statistical Overview: On average, each governorate reports about 102 cases (mean), but this number is influenced by the high count in Jenin. The standard deviation is approximately 106, indicating a wide variation in the number of cases across different

cities. The median number of cases across these governorates is 76.5, which is lower than the mean, reflecting the skewness towards cities with fewer cases.

3. Potential Implications: The variation in case numbers might reflect differences in population size, healthcare access, awareness levels, and screening practices across these regions. Governorates with lower case numbers might not necessarily have lower prevalence but could have issues related to underreporting or lack of adequate screening facilities. The high number of cases in certain areas like Jenin could prompt targeted public health interventions, increased resource allocation, and further epidemiological investigations to understand the underlying causes.
4. Further Research and Policy Actions: Understanding the reasons behind the geographical disparities is crucial. Factors such as socioeconomic status, environmental influences, genetic predispositions, and lifestyle factors should be investigated. Policies focusing on improving breast cancer awareness, screening programs, and healthcare infrastructure, especially in lower-reporting areas, might be beneficial. Collaboration between healthcare providers, local authorities, and communities is essential to enhance reporting accuracy and breast cancer management across all governorates.

Here is the pie chart illustrating the distribution of breast cancer cases across different governorates. Each segment of the pie represents the proportion of cases in a specific governorate, providing a visual overview of how breast cancer cases are distributed geographically.

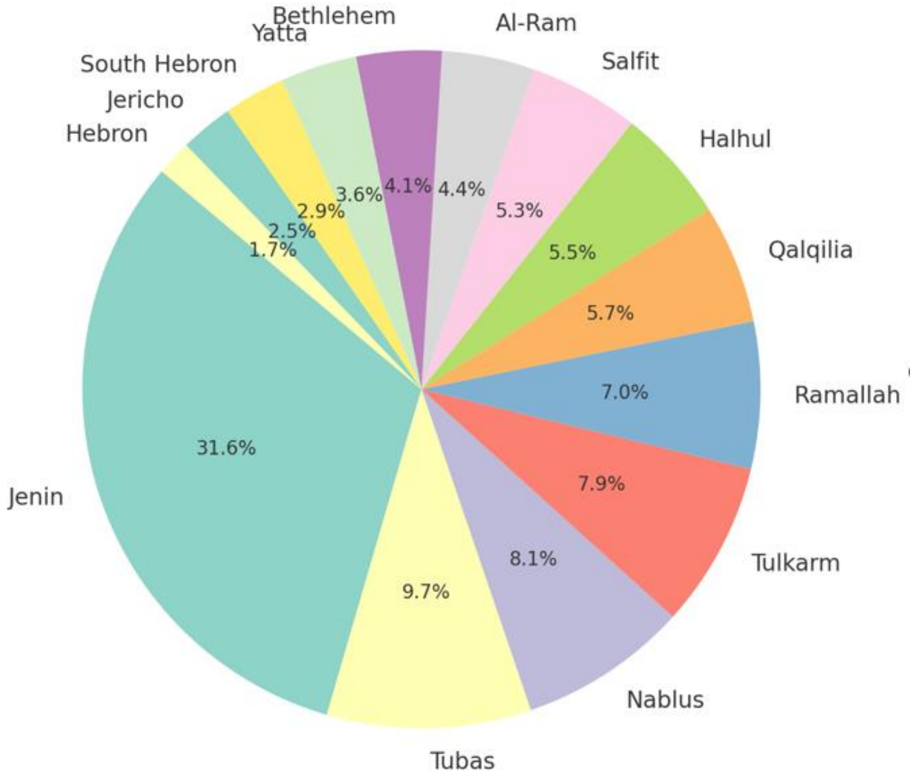

Figure 6.3: Distribution of Breast Cancer Cases across Governorates

6.6 Mammogram Performance in Breast Cancer Early Detection across Governorates

The overall results from the data indicate varied performance in mammogram screenings for breast cancer detection across different Governorates. The effectiveness of mammograms in accurately predicting breast cancer ('Yes'), the uncertainty ('May Be'), and the failure to detect cancer ('No') differ significantly from one region to another. Here is a comprehensive analysis:

1- Effectiveness of Mammograms (Yes): The success rate of mammograms in predicting breast cancer ('Yes') varies considerably. Governorates like Al-Ram (25.84%) and Halhul (37.80%) demonstrate higher success rates. Meanwhile, regions like Salfit (9.72%) and Yatta (7.58%) show lower effectiveness. This variation could be attributed to differences in equipment quality, radiologists' expertise, or population demographics.

2- Uncertainty in Predictions (May Be): A notable aspect is the high rate of 'May Be' outcomes across all regions, indicating a significant level of uncertainty in mammogram results. Tubas (82.45%) and South Hebron (74.07%) have particularly high rates of uncertainty. This could suggest a need for improved mammogram technology or better training for technicians and radiologists in interpreting results.

3- Failure to Detect Breast Cancer (No): The failure rate, where mammograms did not predict existing breast cancer, also shows variation. Governorates like Salfit (25.00%) and Ramallah (22.64%) exhibit higher rates of missed detections. In contrast, Bethlehem (4.00%) and Halhul (3.66%) have low failure rates. This discrepancy might reflect differences in the stages at which cancer is detectable, the quality of mammogram equipment, or the prevalence of types of breast cancer that are harder to detect.

4- Overall Implications: The data reveals the critical need for advancements in mammographic screening techniques and more stringent protocols or training for radiologists. The high percentage of 'May Be' results across the board suggests that many women might be undergoing unnecessary stress and additional testing due to inconclusive initial screenings.

There is also an evident need for targeted improvements in specific regions, especially those with high 'No' percentages, as these are missed opportunities for early intervention, which is crucial in cancer treatment.

Table 6.2: Breast Cancer Detection Performance by Governorate

Governorate	Succeeded	Not Sure	Failed
Jenin	101 13.65%	521 70.41%	118 15.95%
Nablus	43 15.64%	201 73.09%	31 11.27%
Ramallah	48 18.11%	157 59.25%	60 22.64%
Tulkarm	29 13.0%	157 70.4%	37 16.59%
Tubas	23 12.23%	155 82.45%	10 5.32%
Qalqilia	24 15.48%	109 70.32%	22 14.19%
Salfit	14 9.72%	94 65.28%	36 25.0%

Bethlehem	25 20.0%	95 76.0%	5 4.0%
Al-Ram	23 25.84%	38 42.7%	28 31.46%
Halhul	31 37.8%	48 58.54%	3 3.66%
South Hebron	12 14.81%	60 74.07%	9 11.11%
Yatta	5 7.58%	45 68.18%	16 24.24%
Jericho	6 10.0%	44 73.33%	10 16.67%
Hebron	3 12.5%	14 58.33%	7 29.17%

Each cell contains the count of cases followed by the percentage for the respective outcome ("Yes", "May Be", "No") in each Governorate. The percentages are color-coded as per your request: Green for "Yes", Orange for "May Be", and Red for "No".

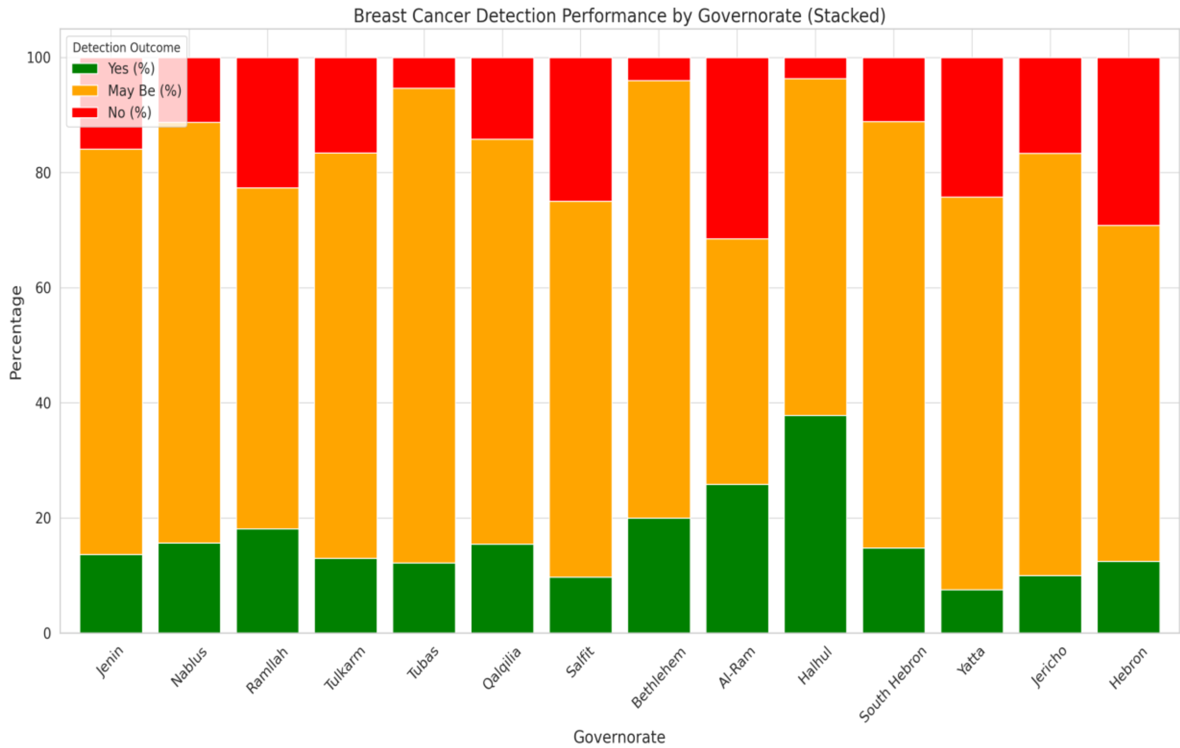

Figure 6.4: Breast Cancer Detection Performance by Governorate

The bar graph has been modified to have thicker (wider) bars, which also enlarges the labels for the Governorates. This adjustment enhances the overall readability and visual impact of the graph, providing a clearer and more detailed view of the breast cancer detection performance across each Governorate.

Chapter 7: Conclusion and Recommendation

7.1 Conclusion

This study delves into the critical aspects of breast cancer detection and screening, particularly focusing on the Palestinian context. It offers insightful observations and raises pertinent questions about the current state of breast cancer diagnosis, highlighting the need for enhanced awareness, better technology, and improved healthcare practices. Below the key findings were simplified of this research, making them more accessible and straightforward.

- High Risk Indicated by BIRADS 5 and 6: Scores of 5 and 6 in BIRADS are strong signs that suggest a high likelihood of breast cancer.
- Limited Screening Among Palestinian Women: There's a worryingly low rate of breast cancer screening among Palestinian women, only 30K screenings per four years for every one million adult women.
- Reasons Behind Low Screening Rates: Many women avoid breast screening due to fear of finding out they have the disease and a lack of knowledge about its impact on themselves and their families.
- Factors Linked to Breast Cancer: A high risk of breast cancer is associated with several factors, including family history, dense breast tissue, early start of menstruation, late menopause, and having a first child after 35 in addition to the geographical area and so it is seen that most of the cases found in northern governorates like Jenin, Tubas, and Tulkarm.
- Other Secondary Factors: Using Hormonal Replacement Therapy for over five years, smoking cigarettes, and having chronic diseases like heart problems, diabetes, or high blood pressure are moderately linked to breast cancer.

- Completeness between Mammogram and Breast Cancer Registries: About 83.2% of mammogram high BIRADS cases are moved and recorded in the breast cancer registry, indicating that not all cases are being tracked effectively.
- Variance of BIRADS in Breast Cancer Registry: With 1682 breast cancer cases there exists an average BIRAD score of around three, the registry shows a range of breast conditions, from benign to very malignant, suggesting that mammogram results and BIRADS scores might sometimes be inaccurate.
- Timeliness of High Mammogram BIRADS into Breast Cancer Care: Half of the patients get into the breast cancer registry quickly after getting high mammogram scores, 75% of the cases attend within two months, but others face delays, with about 23% waiting more than three months, raising concerns about the quality of care and delayed treatment.
- Regional Differences in Cancer Rates: The incidence of breast cancer is notably higher in the northern West Bank, especially in Jenin and Tubas, pointing to possible environmental causes.
- Variation in Detection Rates: The high detection of cancer in Halhul and Al-Ram suggests they might have better-trained staff or superior equipment.
- Issues with Mammogram Accuracy: There's a significant number of unclear mammogram results and some cases that go undetected, possibly due to technology limitations or interpretation errors.
- Importance of Technology and Expertise: The accuracy of breast cancer detection depends on the quality of mammogram machines and the skill of the radiologists reading the images.
- Implications for Public Health Policies: These findings emphasize the need for improved healthcare infrastructure, training for healthcare workers, and more focus on areas with higher cancer rates.

- The Importance of Early Detection: Catching breast cancer early through effective screenings is key to better treatment outcomes, highlighting the need for improvements in this area.

This study sheds light on the urgent need for advancements in breast cancer screening and diagnosis. By addressing these challenges, detection rates and patient outcomes can be improved, calling for joint efforts from healthcare policymakers, practitioners, and researchers in combating breast cancer.

7.2 Recommendation

These recommendations aim to address the gaps identified in the study and contribute to the improvement of breast cancer detection and treatment in Palestine. Improve Mammogram and Breast Cancer Records Integration: The study noted a significant majority of high-risk BI-RAD cases were confirmed as breast cancer cases, yet there was a notable portion of cases that were not matched. This suggests a need for research focused on improving the integration of mammogram data with breast cancer registries. Investigating the barriers to effective data integration and proposing solutions could be beneficial.

- **Study Timeliness of Care Post-Mammography**: Investigating the reasons for delays in treatment after high BI-RAD scores are recorded and proposing solutions to reduce these delays would be a valuable area of research. This could involve studying healthcare system efficiency, patient follow-up, and compliance, as well as the quality of care provided.
- **Assess Effectiveness of Mammogram Screenings Across Governorates**: The study indicates varied performance in mammogram screenings across different regions. Research to understand the causes of these variations, including the quality of equipment, expertise of radiologists, and demographic differences, could inform targeted improvements.

- **Investigate Technological and Human Factors in Mammography Effectiveness**: The study points to the influence of technological and human factors on the effectiveness of mammograms. Future research could focus on the impact of mammogram machine quality, the resolution of images, and the proficiency of radiologists in interpreting these images. This could lead to recommendations for equipment upgrades and enhanced training programs for healthcare professionals.
- **Enhance Breast Cancer Screening and Public Awareness Programs**: The low engagement in breast screening in Palestine, driven by fear and lack of awareness, is a critical issue. Research to develop and assess the effectiveness of targeted awareness programs and screening campaigns, especially in regions with lower engagement, could be instrumental in increasing participation in breast cancer screening.
- **Examine Geographic Variance in Breast Cancer Cases**: The study found a higher prevalence of breast cancer in certain areas, particularly in the northern governorates of the West Bank. Future research could explore the underlying reasons for these geographical disparities, considering factors such as environmental and pollution influences, socioeconomic status, agriculture crops and lifestyle factors.
- **Further Investigate Mammogram Features and Breast Cancer Risk**: The study highlighted the importance of various mammogram attributes like menopause age, menarche age, family history, breast density, first pregnancy age, and compound attributes in predicting breast cancer risk. Future research could delve deeper into these attributes, using more advanced machine learning techniques or larger datasets, to enhance predictive accuracy and understand how these factors influence breast cancer risk.
- **Introduce Hybrid Model of Deep Learning and Machine Learning in Breast Cancer Detection:** Future research should focus on refining and evaluating hybrid deep learning

and machine learning models for breast cancer detection, as these combined approaches have demonstrated significant performance improvements over single-model methods.

- **Enhancing Data Quality in Mammogram and Breast Cancer Databases:** Future tools maintenance should focus on improving mammogram and breast cancer data quality by implementing advanced validation rules in their respective application programming interfaces. This would ensure more accurate and reliable data for effective breast cancer detection.
- **Automating Integration between Mammogram and Breast Cancer Registries:** Research should aim to develop automated systems linking mammogram databases and breast cancer registries. This automation would facilitate the timely transfer of high-risk BI-RAD mammogram records into breast cancer registries for efficient monitoring and follow-up by healthcare professionals.
- **Leveraging Health Informatics for Enhanced Case Management:** Improve the utilization of health informatics tools for better case follow-up and alerting mechanisms in breast cancer management. This includes the development of sophisticated dashboards that prioritize and highlight high-risk cases, ensuring timely and effective medical interventions.
- **Implementing Automated Messaging in Breast Cancer Follow-Up and Awareness:** Explore the implementation of automated messaging systems in mammogram and breast cancer care. This would include sending alerts for high-risk BI-RAD scores or when specific criteria are met, annual awareness messages, and the provision of online consultation services through platforms like WhatsApp. Such initiatives aim to enhance patient follow-up, increase awareness, and maximize the use of health informatics in patient engagement and care management.

7.3 Points of Strength

The study is notable for its multifaceted approach to breast cancer research, combining extensive data analysis, advanced statistical methods, and a focus on impactful findings relevant to Palestinian women's health. Its key strengths can be summarized under the following headings:

- **Extensive Data Analysis:** The study utilizes a large dataset of over 30,000 records, providing a robust foundation for its findings.
- **Advanced Data Analytical Techniques**: Employing advanced technologies, particularly AI tools, enhances the depth and accuracy of the analysis. Cutting-edge machine learning methods, such as logistic regression, are used for in-depth analysis, and visual tools like graphs are employed to present complex data effectively.
- **Data Integration across multiple datasets:** The integration of multiple datasets and linking their results offers a more comprehensive understanding of breast cancer risk factors.

  Relevant and Impactful Findings: The study uncovers multiple findings that directly relate to the health and daily life of Palestinian women, addressing the most critical cancer among this demographic.

  Potential for Real-world Impact: If its findings are implemented, the research has the potential to create tangible effects, significantly influencing breast cancer detection and management strategies.

  Using Broad Health Prediction Factors: Extensive examination of multiple health factors like age, first menstrual period, family history, menopause age, breast density in mammograms, etc. Detailed study of how BI-RADS scores relate to various health conditions.

- **Proposing Regional Insights:** A thorough study of breast cancer cases in different areas offers crucial information for public health planning.

Overall, the study's methodological rigor, use of sophisticated statistical techniques, and focus on a critical health issue make it valuable to public health and breast cancer research.

7.4 Potential Obstacles

The following potential obstacles and limitations highlight areas for further research and consideration in applying the study's findings to broader contexts and healthcare settings:

- **Data Quality and Completeness**: The study reveals a completeness percentage of around 83.23% in matching high-risk BI-RAD cases to breast cancer cases, indicating a gap in data integration processes. This suggests potential limitations in data quality and completeness that could affect the study's findings.
- **Geographical Variation in Breast Cancer Cases**: The study notes significant geographical disparities in breast cancer cases across the West Bank, particularly in northern governorates. This variation could indicate underlying limitations in the study, such as potential biases due to environmental factors, healthcare access, or regional health practices that were not fully accounted for.

  Timeliness in Patient Care: The study highlights delays in the treatment of high-risk cases, with a significant portion of cases experiencing delays of more than three months. This suggests potential limitations in the healthcare system's efficiency and patient follow-up, which could impact the study's implications for public health policy.
- **Technological and Human Factors**: The study acknowledges the influence of both technological and human factors on the effectiveness of mammograms. Limitations in mammogram technology and the interpretation of images could lead to indeterminate or undetected cases, affecting the accuracy of breast cancer detection.

- **Research-to-Policy Translation in Palestine's Healthcare System**: there may be limitations in how these findings can be translated into effective actions, considering the diverse healthcare landscape of Palestine.

Limitations in Generalizability Due to Regional Screening Variations: The effectiveness of mammogram screenings, which seems to vary across regions due to differences in radiologist expertise or demographic factors, could challenge the universal applicability of the study's findings.

References

Akosa, J. (2017). Predictive accuracy: A misleading performance measure for highly imbalanced data. In Proceedings of the SAS Global Forum 2017. Retrieved from https://support.sas.com/resources/papers/proceedings17/0921-2017.pdf

Al-Tell, M., Natour, N., & Younes, D. (2019). Barriers to breast cancer screening among Palestinian women in Nablus region, Northern West Bank. Palestinian Medical and Pharmaceutical Journal, 5(2), 75-82. https://doi.org/10.59049/2790-0231.1061

AlWaheidi, S., McPherson, K., Chalmers, I., Sullivan, R., & Davies, E. A. (2020). Mammographic screening in the occupied Palestinian territory: A critical analysis of its promotion, claimed benefits, and safety in Palestinian health research. JCO Global Oncology, 6, 1772-1790. https://doi.org/10.1200/JGO.19.00383

American Cancer Society. (2019). Breast cancer facts & figures 2019-2020. American Cancer Society. Retrieved from https://www.cancer.org/research/cancer-facts-statistics.html

Anera. (2010). Early breast cancer detection in the West Bank. Anera. Retrieved May 24, 2024, from https://www.anera.org/stories/early-breast-cancer-detection-in-the-west-bank/

Arasu, V. A., Habel, L. A., Achacoso, N. S., Buist, D. S. M., Cord, J. B., Esserman, L. J., Hylton, N. M., Glymour, M. M., Kornak, J., Kushi, L. H., Lewis, D. A., Liu, V. X., Lydon, C. M., Miglioretti, D. L., Navarro, D. A., Pu, A., Shen, L., Sieh, W., Yoon, H. C., & Lee, C. (2023). Comparison of mammography AI algorithms with a clinical risk model for 5-year breast cancer risk prediction: An observational study. Radiology, 307(5), e222733. https://doi.org/10.1148/radiol.222733

Bahri, N., Mardani, F., Sharifi, N., & Dashti, S. (2022). Predicting factors for breast cancer screening in Middle Eastern women based on health belief model: A systematic review. Journal of the Egyptian National Cancer Institute, 34(1), 50. https://doi.org/10.1186/s43046-022-00150-3

Bancej, C., Decker, K., Chiarelli, A., Harrison, M., Turner, D., & Brisson, J. (2003). Contribution of clinical breast examination to mammography screening in the early detection of breast cancer. Journal of Medical Screening, 10(1), 16-21. https://doi.org/10.1177/096914130301000104

Bassett, L. W., & Butler, D. L. (1991). Mammography and early breast cancer detection. American Family Physician, 43(4), 547-557. Available from https://www.aafp.org/afp/1991/0401/p547.html

Berg, W. A., Berg, J. M., Sickles, E. A., Burnside, E. S., Zuley, M. L., Rosenberg, R. D., & Lee, C. S. (2020). Cancer yield and patterns of follow-up for BI-RADS category 3 after screening mammography recall in the national mammography database. Radiology, 296(1), 32-41. https://doi.org/10.1148/radiol.2020192641

Bleyer, A., & Welch, H. G. (2012). Effect of three decades of screening mammography on breast-cancer incidence. New England Journal of Medicine, 367, 1998-2005. https://doi.org/10.1056/NEJMoa1206809

Bleyer, A., & Welch, H. G. (2012). Effect of three decades of screening mammography on breast-cancer incidence. Obstetrical & Gynecological Survey, 67(10), 634-635. https://doi.org/10.1097/OGX.0b013e31826c8bb7

Boumaraf, S., Liu, X., Ferkous, C., & Ma, X. (2020). A new computer-aided diagnosis system with modified genetic feature selection for BI-RADS classification of breast masses in mammograms. Biomedical Research International, 2020, 7695207. https://doi.org/10.1155/2020/7695207

Busch, S. H., Hoag, J. R., Aminawung, J. A., Xu, X., Richman, I. B., Soulos, P. R., Kyanko, K. A., & Gross, C. P. (2019). Association of state dense breast notification laws with supplemental testing and cancer detection after screening mammography. American Journal of Public Health, 109(5), 762-767. https://doi.org/10.2105/AJPH.2019.304967

Choi, K. S., Yoon, M., Song, S. H., Suh, M., Park, B., Jung, K. W., & Jun, J. K. (2018). Effect of mammography screening on stage at breast cancer diagnosis: Results from the Korea

National Cancer Screening Program. Scientific Reports, 8(1), 8882. https://doi.org/10.1038/s41598-018-27152-3

Coleman, C. (2017). Early detection and screening for breast cancer. Seminars in Oncology Nursing, 33(2), 141-155. https://doi.org/10.1016/j.soncn.2017.02.009

Febles, G. (2012). False negatives in mammography. In Proceedings of the International Radiology Congress (pp. 45-50). Barcelona, Spain. International Radiology Society.

GlobalGiving. (n.d.). Breast cancer patient navigation program in Palestine. GlobalGiving. Retrieved May 24, 2024, from https://www.globalgiving.org/projects/breast-cancer-patient-navigation-palestine/

Hamshari, S., Nazzal, Z., Altell, M., Nanaa, I., Jbara, R., & Sabri, R. (2021). Mammogram uptake and barriers among Palestinian women attending primary health care in North Palestine. European Journal of General Practice, 27(1), 264-270. https://doi.org/10.1080/13814788.2021.1985996

JetBrains. (2021). DataGrip (Version 2021.1.3) [Computer software]. JetBrains. https://www.jetbrains.com/datagrip/

Johns Hopkins Medicine. (n.d.). Breast Imaging Reporting & Data System (BI-RADS). Johns Hopkins University. Retrieved from https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/breast-imaging-reporting-and-data-system-birads

Jubran, J., Shaar, A., Jarrar, K., & Hammad, S. (2018). Pathway to survival - The story of breast cancer in Palestine. United Nations Population Fund (UNFPA). Retrieved from https://palestine.unfpa.org/en/publications/pathway-survival-story-breast-cancer-palestine

Kaur, M. N., Yan, J., Klassen, A. F., David, J. P., Pieris, D., Sharma, M., Bordeleau, L., & Xie, F. (2022). A systematic literature review of health utility values in breast cancer. Medical Decision Making, 42(5), 704-719. https://doi.org/10.1177/0272989X211065471

Khan, A. I., Abushark, Y. B., Alsolami, F., Almalawi, A., Alam, M. M., Kshirsagar, P., & Khan, R. A. (2023). Prediction of breast cancer based on computer vision and artificial intelligence techniques. Measurement, 218, 113230. https://doi.org/10.1016/j.measurement.2023.113230

La Vecchia, C., Negri, E., Bruzzi, P., Franceschi, S., Bucchi, L., & Parazzini, F. (1993). The impact of mammography on breast cancer detection. Annals of Oncology, 4(1), 41-44. https://doi.org/10.1093/oxfordjournals.annonc.a058567

Liu, J., Wang, C., & Liu, S. (2023). Utility of ChatGPT in clinical practice. Journal of Medical Internet Research, 25, e48568. https://doi.org/10.2196/48568

Liu, Y., Kornfield, R., Yang, E. F., Burnside, E., Keevil, J., & Shah, D. V. (2022). Patient-provider communication while using a clinical decision support tool: Explaining satisfaction with shared decision making for mammography screening. BMC Medical Informatics and Decision Making, 22(1), 323. https://doi.org/10.1186/s12911-022-02058-3

Løberg, M., Lousdal, M. L., Bretthauer, M., & Kalager, M. (2015). Benefits and harms of mammography screening. Breast Cancer Research, 17(63). https://doi.org/10.1186/s13058-015-0525-z

Mammography. (2018). NIH: National Cancer Institute. Retrieved from https://www.cancer.gov/types/breast/mammograms-fact-sheet

Palestinian National Institute of Public Health. (n.d.). Mammography-e-Registry. World Health Organization, Regional Office for the Eastern Mediterranean. Retrieved from https://www.emro.who.int/opt/priority-areas/palestinian-national-institute-of-public-health.html

Rabiei, R., Ayyoubzadeh, S. M., Sohrabei, S., Esmaeili, M., & Atashi, A. (2022). Prediction of breast cancer using machine learning approaches. Journal of Biomedical Physics and Engineering, 12(3), 297-308. https://doi.org/10.31661/jbpe.v0i0.2109-1403

Simbrich, A., Wellmann, I., Heidrich, J., Heidinger, O., & Hense, H. W. (2016). Trends in advanced breast cancer incidence rates after implementation of a mammography screening program in a German population. Cancer Epidemiology, 44, 44-51. https://doi.org/10.1016/j.canep.2016.07.006

Soekersi, H., Azhar, Y., & Akbari, K. S. (2022). Role of mammography in breast cancer screening: A literature review. Journal of the Indonesian Medical Association, 72(3), 144-150. https://doi.org/10.47830/jinma-vol.72.3-2022-627

Taif, S., Tufail, F., & Alnuaimi, A. S. (2016). Mammography performance in Oman: Review of factors influencing cancer yield and positive predictive value. Asia-Pacific Journal of Clinical Oncology, 12(2), e250-e258. https://doi.org/10.1111/ajco.12181

Tutar, B., Esen Icten, G., Guldogan, N., Kara, H., Arıkan, A. E., Tutar, O., & Uras, C. (2020). Comparison of automated versus hand-held breast US in supplemental screening in asymptomatic women with dense breasts: Is there a difference regarding woman preference, lesion detection, and lesion characterization? Archives of Gynecology and Obstetrics, 301(5), 1257-1265. https://doi.org/10.1007/s00404-020-05501-w

United Nations Population Fund (UNFPA). (2020). For one breast cancer survivor in Gaza Strip, a journey of hardship and hope. Retrieved May 24, 2024, from https://www.unfpa.org/news/one-breast-cancer-survivor-gaza-strip-journey-hardship-and-hope

United Nations Population Fund. (2018). Pathway to Survival - the Story of Breast Cancer in Palestine. UNFPA.https://palestine.unfpa.org/en/publications/pathway-survival-story-breast-cancer-palestine

Uslu, A., & Stausberg, J. (2021). Value of the electronic medical record for hospital care: Update from the literature. Journal of Medical Internet Research, 23(12), e26323. https://doi.org/10.2196/26323

Van Luijt, P. A., Fracheboud, J., Heijnsdijk, E. A., den Heeten, G. J., de Koning, H. J., & National Evaluation Team for Breast Cancer Screening in the Netherlands Study Group

(NETB). (2013). Nation-wide data on screening performance during the transition to digital mammography: Observations in 6 million screens. European Journal of Cancer, 49(16), 3517-3525. https://doi.org/10.1016/j.ejca.2013.06.019

Vieira, R. A. C., Biller, G., Uemura, G., Ruiz, C. A., & Curado, M. P. (2017). Breast cancer screening in developing countries. Clinics (Sao Paulo), 72(4), 244-253. https://doi.org/10.6061/clinics/2017(04)09

Wang, S., Minku, L. L., & Yao, X. (2016). Dealing with multiple classes in online class imbalance learning. In Proceedings of the 25th International Joint Conference on Artificial Intelligence (pp. 2118-2124). AAAI Press. https://doi.org/10.24963/ijcai.2016/295

World Health Organization (WHO), & Palestinian National Institute of Public Health. (2014). Performance of mammography screening in the national breast cancer screening program: A retrospective cohort study. WHO Regional Office for the Eastern Mediterranean. Retrieved from https://www.emro.who.int/opt/priority-areas/palestinian-national-institute-of-public-health.html

Yankaskas, B. C., Haneuse, S., Kapp, J. M., Kerlikowske, K., Geller, B., & Buist, D. S. (2010). Performance of the first mammography examination in women younger than 40 years. JNCI: Journal of the National Cancer Institute, 102(10), 692-701. https://doi.org/10.1093/jnci/djq131

Appendices

Appendix A: Mammogram Registry DHIS2

Person ID	1	2	3	4	5
Organization Code	GP301790	GP301790	GP301790	GP050610	GP100646
Organization Name	(Al.Biereh) أمومة البيرة	(Al.Biereh) أمومة البيرة	(Al.Biereh) أمومة البيرة	(Tubas Central) طوباس المركزية	(Center of Mammography \Tulkarm) مركز تصوير الثدي الترميمي
Date of Enrollment	00:00.0	00:00.0	00:00.0	00:00.0	00:00.0
BI-RAD	BI-RAD1-Negative	BI-RAD0-Incomplete	BI-RAD0-Incomplete	BI-RAD2-Benign finding(s)	BI-RAD3-Probably benign
BMI	34	34	34	49	29
Age	48	46	45	51	50
Menarche Age	13	13	12	12	14
Menopause Age	NaN	NaN	NaN	48	NaN
Marital Status	Married	Married	Married	Married	Married
Took Contraceptives	0	0	0	0	0
Got Pregnant	1	1	1	1	1
Has Ovarian Cancer	0	0	0	0	0
Family Breast Cancer History	1	1	1	0	0
Has Breast Cancer	0	0	0	0	0
Mass (Kg)	87	87	87	110	65
Density	Scattered fibro glandular Density	Heterogeneous Dense	Heterogeneous Dense	Scattered fibro glandular Density	Scattered fibro glandular Density
Recommendations	For routine follow up	left breast us show cystic lesion...	for breast us	repeat mammography after 2 years	ULTRASOUND: mass in deep MLO for diff. LN from tumor...
Mammogram Reason	Screening	Diagnosis	Screening	screening	screening
Has Chronic Diseases	No	No	No	Yes	Yes

Note: The displayed data is a small portion of the entire dataset for illustration purposes and may have been truncated for brevity.

Appendix B: Mammogram Registry Legacy PHP-based Website

Attribute	Record1	Record2	Record3	Record4
Person ID	1	2	3	4
District ID	Bethlehem	Bethlehem	Bethlehem	Bethlehem
Birth Date	12/20/1960	6/7/1962	6/7/1962	2/22/1970
Appointment Date	12/27/2016	12/27/2016	10/17/2018	12/27/2016
BI-RAD	0	0	3	2
Mass (KG)	79	66	75	75
Height	158	145	145	155
BMI	32	31	36	31
Referral Status	Yes	No	No	Yes
Source of Referral	Governmental	No Referral	No Referral	Governmental
Previous Mammogram	2 Yr	2 Yr	> 2 Yr	2 Yr
Menarche Age	> 12	> 12	> 12	> 12
First Pregnancy Age	< 35 Yr	Not Pregnant	< 35 Yr	< 35 Yr
Contraceptive Use	Never	Never	Never	Never
Menopause	40-55 Yr	40-55 Yr	40-55 Yr	Not Reach
Hormonal Replacement Therapy (HRT)	Never	Never	Never	Never
Smoking	Neither	Neither	Neither	Neither
Chronic Diseases	Hypertension	None	Diabetes	Hypertension
Hypertension	Yes	No	No	Yes
Diabetes	No	No	Yes	No
Heart Diseases	No	No	No	No
Personal History of Breast Cancer	Neither	Neither	Neither	Neither
Family History of Breast Cancer	None	Sister	Sister	None
Personal History of Ovarian Cancer	No	No	No	No
Mammogram Reason	Diagnostic	Screening	Screening	Screening
Density	Scattered fibro glandular Density	Heterogeneous Dense	Scattered fibro glandular Density	Scattered fibro glandular Density
Recommendations	Ultrasound is recommended.	Ultrasound is recommended.	Ultrasound is recommended.	Continue routine screening.

Note: The displayed data is a small portion of the entire dataset for illustration purposes and may have been truncated for brevity.

Appendix C: Breast Cancer Registry

PersonalID	Year	AddCode	AgeDiag	BirthDate	ICD10	IncidentDate
1	2021	311145	37	19840824	C508	20211209
2	2021	812485	47	19740611	C502	20210815
3	2014	812205	42	19711122	C509	20140928
4	2014	210520	45	19681214	C509	20141210
5	2023	231165	56	19661010	C508	20230921
6	2019	411415	48	19701023	C504	20190102
7	2022	812205	47	19750701	C505	20221110
8	2019	812230	48	19700808	C509	20190620
9	2021	511700	54	19670910	C509	20211212
10	2017	210220	59	19580126	C509	20170815

Note: The displayed data is a small portion of the entire dataset for illustration purposes

Appendix D: University Letter to Ministry of Health for Data Authorization

Arab American University
Faculty of Graduate Studies

الجامعـــة العربيـــة الأمريكيـــة
كلية الدراسات العليا

2020-05-12

حضرة الدكتورة أمل أبو عوض المحترمة
مدير عام التعليم الصحي في وزارة الصحة الفلسطينية

<u>تسهيل مهمة بحثية</u>

تحية طيبة وبعد،

تهديكم الجامعة العربية الأمريكية أطيب التحيات، وبالإشارة الى الموضوع أعلاه تشهد كلية الدراسات العليا في الجامعة ان الطالب عبدالرحمن فايز عبدالرحمن جمعة والذي يحمل الرقم الجامعي 201812616 هو طالب ماجستير في الجامعة العربية الأمريكية تخصص معلوماتية صحية، ويعمل على دراسة علمية لمساق الرسالة بعنوان "دور تصوير الثدي الشعاعي واستخدام المعلوماتية الصحية في الكشف المبكر عن سرطان الثدي في فلسطين" بإشراف الدكتور يوسف الميمي والدكتور محمد عوض، نأمل من حضرتكم الايعاز لمن يلزم لمساعدته للحصول على بيانات غير تعريفية من سجل السرطان وسجل تصوير الثدي الشعاعي اللازمة للدراسة، علماً ان المعلومات ستستخدم لغاية البحث فقط وسيتم التعامل معها بغاية السرية.

وتفضلوا بقبول فائق الاحترام

عميد كلية الدراسات العليا

د. عبد الرحمن ابو لبدة

كلية الدراسات العليا
FACULTY OF GRADUATE STUDIES

Jenin Tel: +970-4-2418888 Ext:1471,1472 Fax: +970-4-2510810 P.O. Box:240
Ramallah Tel: +970- 2- 2941999 Fax: +970-2-2941979 Abu Qash - Near Alrehan
E-mail: FGS@aaup.edu ; PGS@aaup.edu Website: www.aaup.edu

الملخص

يُعد سرطان الثدي الأكثر انتشارًا بين النساء على مستوى العالم، ويشكل تحديات صحية عامة جسيمة. يُعتبر الكشف المبكر عن سرطان الثدي عاملاً محوريًا في تحسين معدلات البقاء على قيد الحياة، خاصة في ظل التطورات الكبيرة في التقنيات التشخيصية مثل التصوير الشعاعي للثدي (الماموغرام). تسعى هذه الأطروحة إلى دراسة فعالية استراتيجيات الكشف المبكر عن سرطان الثدي في السياق الفلسطيني بالتعاون مع وزارة الصحة الفلسطينية، مستندةً إلى قاعدتي بيانات أساسيتين: "التصوير الشعاعي للثدي" و"سجل سرطان الثدي".

اعتمدت الدراسة على منهجية كمية مدعومة بتحليلات نوعية، حيث تم تحليل البيانات الثانوية باستخدام تصميم بأثر رجعي. شملت الدراسة جميع النساء في الضفة الغربية اللاتي خضعن لفحوصات التصوير الشعاعي للثدي لدى وزارة الصحة، باستخدام أسلوب المعاينة غير الاحتمالي لتحليل 30,000 سجل مريض طبي على مدار أربع سنوات.

أظهرت النتائج وجود ارتباط وثيق بين مكان الإقامة، وعمر الحمل الأول، وأمراض القلب وبين احتمالية الحصول على درجات عالية في نظام تصنيف كثافة الثدي. كما أظهرت الدراسة نسبة تكامل تصل إلى 83.23% في ربط الحالات ذات التصنيف المرتفع في نظام تصنيف كثافة الثدي بسجل سرطان الثدي، مما يشير إلى فعالية الربط بين سجل التصوير الشعاعي وسجل السرطان. كما تم رصد تباينات إقليمية واضحة، حيث سجلت محافظة جنين أعلى معدلات الإصابة. وبالرغم من أن معظم الحالات عالية التصنيف في سجل فحص الثدي الشعاعي تُضاف إلى سجل سرطان الثدي خلال شهر واحد، إلا أن بعض الحالات شهدت تأخيرات تصل إلى عدة سنوات.

في الختام، تؤكد هذه الأطروحة على الدور الحيوي الذي تلعبه المعلوماتية الصحية في تحسين الكشف المبكر عن سرطان الثدي ورفع كفاءة نتائجه العلاجية. تشمل التوصيات تحديث أنظمة البيانات، تعزيز تدريب الكوادر الصحية، وتطوير سياسات صحية أكثر صرامة لتحسين إدارة حالات سرطان الثدي والرعاية المقدمة للمرضى، مع التأكيد على ضرورة تكاتف الجهود بين جميع الأطراف المعنية في النظام الصحي الفلسطيني لتحسين جودة الرعاية الصحية.