OPTIMAL BIOPSY DECISION MAKING IN BREAST CANCER USING REINFORCEMENT LEARNING

By

Ruchi Parmar
501034872

Literature Review &
Exploratory Data Analysis

Master of Science
in the Program of
Data Science and Analytics

Toronto, Ontario, Canada, 2024

Table of Contents

List of Figures

List of Tables

1. Introduction

This document covers the Introduction, Literature Review and Exploratory Data Analysis for the first deliverable of my Major Research Project (MRP). It begins with a brief background on the topic and datasets, defines the problem, and states the research question. This is followed by a literature review and a detailed exploratory analysis of the dataset.

A. Background

Breast Cancer remains significant global health challenge, with increasing incidence rates. It is estimated that 2.3 million new cases of breast cancer were diagnosed in 2022 and it caused 670,00 total deaths globally[34]. Annual Mammography is the standard practice followed by most of the advanced countries to detect Breast Cancer in early stages. In the United States approximately 39 million mammograms are performed annually as of year 2021[12]. Post mammogram more than 1 million women undergo breast biopsies annually. These biopsies are crucial for diagnosing breast abnormalities and determining whether a lump or mass is benign or cancerous. Approximately 20% of breast biopsies performed in the U.S. each year turn out to be cancerous, while the remaining 80% are false positives (benign)[22]. This cause unnecessary treatments, patient anxiety, and expenditures. This project focuses on exploring Reinforcement Learning(RL) techniques to optimize breast cancer biopsy decision-making. This approach aims to optimize decision-making, reduce unnecessary procedures, and mitigate the economic impact of false positives, ultimately improving patient outcomes. I have utilized data of State-Transition probabilities for breast cancer risks available on GitHub Repository[9] to build RL framework.

B. Research Objectives

Through this research project, I aim to achieve several objectives in two major parts. First, I will implement traditional Markov Decision Process(MDP) based Reinforced Learning methods for biopsy decision-making and obtain results with available data in line with existing research. Next, I will apply advanced reinforcement learning techniques, such as Q-Learning and Deep Q-Learning, to further optimize the biopsy decision-making process. My goals are to assess the impact of these methods on reducing false positives (unnecessary procedures) while maintaining true positives (actual cancer cases) and to compare the effectiveness of these techniques against traditional MDP-based (Markov Decision Process) methods. Additionally, I want to formulate

reward functions to guide RL agents in making decisions, considering factors such as the disutility of biopsy, to achieve improved performance.

2. Literature Review

The primary focus of this project is to understand the approaches used by radiologists in making decisions based on mammogram screenings and to replicate traditional reinforcement learning (RL) methods for training an RL agent to optimally make biopsy decisions while balancing false positives (FP) and true positives (TP). Next, utilize advance RL methods to further optimize biopsy decision making process. Although several studies have explored traditional RL methods in this domain, applying Q-Learning and Deep Q-Learning for optimizing biopsy decision-making is a novel approach.

In this section, I provide an overview of the articles referenced for this project. I began by reviewing articles related to traditional methods for detecting breast cancer. Screening mammography is the standard practice for early cancer detection in asymptomatic women. Organizations such as the American Cancer Society (ACS) advocate mammography as the primary method for identifying breast cancer. The ACS advises that women undergo clinical breast examinations every three years between the ages of 20 and 40 and annually after age 40. For women at average risk, screening mammography is recommended annually starting at age 40 and should continue as long as the woman is eligible for basic breast cancer treatment (Smith et al. (2010) [27]). This practice has been shown to significantly reduce breast cancer mortality (Nelson et al. (2009) [23]).

After screening mammography, radiologists evaluate the likelihood of cancer based on the findings and decide on the appropriate management, balancing the need for early cancer detection against the risk of false positives. Depending on the patient's cancer risk, radiologists typically choose one of three options: (1) routine follow-up mammography in a year, (2) short-term follow-up mammography in six months, or (3) immediate diagnostic actions such as biopsy (Ayvaci and Burnside (2012) [3]).

Mammogram interpretation can go wrong in two ways. First, a mammogram can be interpreted as normal when cancer is present, leading to delayed diagnosis and increased risk to the patient's life. While some false negatives are due to cancers that are not visible on the mammogram, the majority are cancers that could have been detected in hindsight (Baines

and Dayan (1999) [4], Bird, Wallace, and Yankaskas (1992) [8]). Second, a mammogram can be incorrectly labeled as positive when there is no cancer, resulting in over-treatment and unnecessary anxiety for patients. The false-positive rate of mammography is nearly 80% in United States ([22]). These error rates highlight the complexity of interpreting mammograms.

The accuracy of mammogram reading differs with the radiologist's skills and training (Barlow et al. (2004) [6], Beam, Layde, and Sullivan (1996) [7]). Radiologists, with fellowship training in mammography and extensive focus on this domain, generally perform better than general radiologists, who interpret mammograms as part of a broader practice (Sickles, Wolverton, and Dee (2002) [26]). Additionally, there are performance differences between countries also. Research found that while cancer detection rates are similar in the U.S. and the U.K., U.S. radiologists labeled many more mammograms as uncertain or suspicious, leading to at least twice as many follow-up tests, such as biopsies (Smith-Bindman et al. (2003) [28]).

To ensure consistency in mammography practices, the American College of Radiology (ACR) developed the Breast Imaging Reporting and Data System (BI-RADS) lexicon (BI-RADS (1998) [25], Liberman and Menell (2002) [20]). This system includes descriptors that are key indicators for benign or malignant diagnoses, assisting radiologists and referring physicians in making informed breast cancer decisions and managing patients. Each mammogram is described using standardized descriptors and categorized into one of six final assessment categories (BI-RADS codes) based on the mammogram's observations as seen in Table 1. For instance, BI-RADS 0 requires additional imaging; BI-RADS 1 and 2 suggest no further action and continued routine screening for no or benign findings; BI-RADS 3 indicates a probably benign finding needing short-term follow-up (typically six months to two years); and BI-RADS 4 and 5 suggest suspicious or high-risk findings, recommending an immediate biopsy (Chhatwal, Alagoz, and Burnside (2010) [14]).

Despite this standardization, the current use of the BI-RADS lexicon has unresolved issues. Notably, the BI-RADS categories do not account for age in biopsy decision-making, despite older age groups having specific characteristics that need particular consideration. For example, breast cancers in older women are often less aggressive, implying that a higher action threshold might be suitable (Fowble et al. (1994) [17]). Moreover, false-positive mammograms that lead to unnecessary invasive procedures can be more challenging or risky for older individuals with other health conditions. These unique features of breast cancer in older women suggest that biopsy probability thresholds might differ from those in younger women (Chhatwal, Alagoz, and Burnside (2010) [14]).

Table 1: BI-RADS assessment codes

BI-RADS	Explanation	Recommended Action
0	Need additional Imaging evaluation	additional Imaging evaluation
1	Negative finding	routine yearly screening
2	Benign Finding	routine yearly screening
3	Probably Benign (less than 2% risk of cancer)	Short-term follow-up (6 months)
4	Suspicious (risk of cancer is between 2% and 95%)	Biopsy
5	Highly suggestive of malignancy (95% risk of cancer)	Biopsy

Computer-aided diagnostic (CADx) models show promise to enhance early breast cancer diagnosis as per different researches. For example, Few models like Bayesian network to predict malignancy risk using mammographic features and demographic data (developed by Burnside et al. (2006) [11], Burnside et al. (2009) [10]), artificial neural networks (used by [5] and Ayer et al. (2010) [2]), decision fusion(used by Jesneck et al. (2006) [19]) and logistic regression (used by Chhatwal et al. (2009) [15]) performed better and gave more accurate results. This models generally provides probability of cancer instead of BI-RADS categories, which can be more helpful to make accurate decisions. However, this methods also doesn't provide any threshold for biopsy decision with respect to patient related factors such as age, cost etc.(Chhatwal, Alagoz, and Burnside (2010) [14]).

These methods doesn't help to determine an optimal threshold for recommending biopsies to reduce high number of false-positives. Furthermore, false-positive mammograms cause unnecessary anxiety, pain, and potential complications for patients. Additionally, biopsies can introduce changes (such as distortion) in future mammograms, complicating subsequent diagnoses ([14]). There are several studies which utilised traditional Markov Decision Based Reinforcement Learning methods for various use cases such as finding optimal biopsy decision making, to find optimal screening intervals, to find screening polices under various constraints such as cost or to reduce the over diagnosis.

Reinforcement learning (RL) (Sutton and Barto (2018) [31], Szepesvári (2010) [32], Powell (2011) [24], Hengst et al. (2020) [18], Dong, Ding, and Zhang (2020) [16]), is a type of adaptive mechanism and a sub-field of machine learning where agents learn dynamically through heuristic methods by interacting with their environment. Unlike other computational approaches, RL allows agents to learn without needing specific supervision or a complete model of the environment.

Typically, an agent's actions in the environment result in both immediate and delayed rewards. This interaction prompts the agent to search for the optimal next action within the state space. The agent operates through a trial-and-error process to maximize the total accumulated rewards over time (Yang et al. (2023) [35]).

In one of the studies, author developed Optimal Biopsy Decision Model(OBDM) that integrates both mammographic and Demographic features using finite-horizon discrete time Markov Decision Process(MDP). With nearly 65,892 clinical data, model parameters such as state transition probabilities were estimated. This model recommends fewer biopsies compared to radiologists, reducing false positives without significantly increasing false negatives. Age-dependent biopsy thresholds indicate higher thresholds for older women, aligning with less aggressive cancer progression and comorbidities in this group (Chhatwal, Alagoz, and Burnside (2010) [14]).

In another paper, author utilised large-scale finite-horizon Markov Decision Process (MDP) model incorporating 4.6 million states to reduce number of over diagnosis in breast cancers which leads to unnecessary treatments and healthcare costs. This model aims to distinguish between invasive and indolent breast cancers. It employed divide-and-search algorithm for optimal decision threshold identification. This resulted into potential 20% reduction in over diagnosis through tailored diagnostic guidelines. This study also found that age-specific diagnostic policies could result in substantial quality assured life years(QALY) gains and cost savings and identified that current practices could lead to aggressive diagnostic actions, especially in older women (Tunç, Alagoz, and Burnside (2022) [33]).

Other study aimed to assess the value and effectiveness of dynamic mammography screening policies over static ones. Author Created a partially observed Markov chain model to emulate breast cancer progression and investigate dynamic policies that adjust screening intervals based on age-related risk factors. By analyzing wide range of two-phase screening policies for their impact on lifetime mortality risk, author concluded that dynamic, two-phase screening policies could be more beneficial than the current recommendations (Maillart et al. (2008) [21]). A. Ayer and Stout used a Partially Observable Markov Decision Process (POMDP) to optimize screening schedules. They aimed to develop a personalized screening model that considers a broader range of individual risk factors, including personal screening history improving quality-adjusted life years(QALYs) and reducing unnecessary screenings. This study concluded that personalised screening strategies could potentially lead to fewer mammograms and false positives while increasing QALYs, especially in high-risk groups (Ayer and Stout (2012) [1]).

In another study, author focused on optimizing diagnostic decisions after mammography, maximizing Quality-Adjusted Life Years (QALYs) within budget constraints. They created a finite-horizon constrained Markov Decision Process (MDP) model. Results showed that Optimal decision-making could lead to approximately 22% savings in costs without compromising QALYs. Short-term follow-ups identified as primary targets for cost reduction when under budget constraints. The study's findings highlight the importance of integrating cost considerations with health outcomes to improve healthcare decision-making (Ayvaci and Burnside (2012) [3]). Cevik et al. focuses on optimal allocation of limited mammography resources for effective breast cancer screening. Early detection of breast cancer through mammography is crucial but, limited resources (including cost and availability of trained personnel and diagnostic machines) hinder widespread adoption of mammography screening. The study proposes a constrained partially observable Markov decision process (CPOMDP) model. It concludes that efficient allocation of resources based on different risk levels significantly improves QALY gains, especially for patients at higher breast cancer risk. As mammography screening capacity decreases, patients in the 40–49 age group receive the least priority for screening (Cevik et al. (2018) [13]).

Most of these studies used traditional MDP based approaches. To the best of my knowledge, no study has explored the use of advanced reinforcement learning methods, such as Q-Learning or Deep Q-Networks(DQN) specifically for the used case of Optimal Biopsy Decision Making. As many other studies in healthcare domain have used these methods, they have observed better performances. Traditional RL methods follows model/algorithm based approach, whereas Q-learning and DQN follows learning from environment by following action-reward strategy. One of the study on brain tumor localization study addresses limitations in AI for radiology, including large annotated datasets, non-generalizability, and inadequate explainability and explores reinforcement learning (RL) as a solution for these challenges. They trained Deep Q Network and incorporated image exploration with rewards and punishments to localize lesions.This study observed that RL predictions consistently improved during training and achieved 85% accuracy in predicting lesion locations, outperforming supervised deep learning (7% accuracy) (Stember and Shalu (2020) [30]).

This research project focus on using these techniques to find threshold for biopsy recommendation in Breast cancer for Optimal Biopsy Decision making and compare results with transitional methods.

3. Exploratory Data Analysis

This section aims to provide a comprehensive understanding of the dataset used for the research project. It contains the information on Data Source and Files, description of data's basic features. It includes more detailed analysis, such as relation between variables, identifying trends. By performing EDA, I aim to gain a deeper understanding of the data, which will inform the subsequent steps in the research and help address the research questions effectively.

A. Data Source and Files

Data for this project is acquired from Github open Repository - Breast Cancer transition Probabilities[9]. The datasets are open-sourced and compliant with the MRP requirements. It contains two csv files -

- "all_women_data.csv" - This file contains data of 10000 women's probability of cancer at early stage and advanced stage at each time stamp(age 40-100). Patients start screening at age 40 and goes under screening each year as recommendation.
- "state_transition_probabilities.csv" - This file contains state transition probabilities at each age for RL model for Regular Mammogram and Biopsy actions. It considers cancer risk score as a state of RL model(i.e. MDP) and calculate transition probabilities of states at different age.

Data files are available on GitHub. Link to the GitHub repository is included in References.

B. Data Analysis

In this section we will analyze we will analyze each feature of both data files, exploring their individual properties and interrelationships. Additionally, we will examine the rewards distribution for each action (Annual Mammogram & Biopsy) of the RL agent used in the project.

- Patient ID - ID of patient(10000 women in total). This column has no nulls.
- Age - Age of patient at time of screening. Age ranges between 40 to 100. As per the guidelines women usually starts regular screening at the age of 40.

- Stages - There are total 5 features according to 5 different stages of cancer and death - Stage 0 - Healthy, Stage 1 - Early stage cancer, Stage 2 - Advanced stage cancer, Stage 3 - Death from cancer, Stage 4 - Death from other causes. These features contains values of the probability of patient being in each stage. i.e. if probability if patient being healthy at given age.

  These probabilities ranges between 0 to 1. In the following graphs(Figure 1 & Figure 2) we can see the distribution of probabilities of patient being in early or advanced stage cancer.

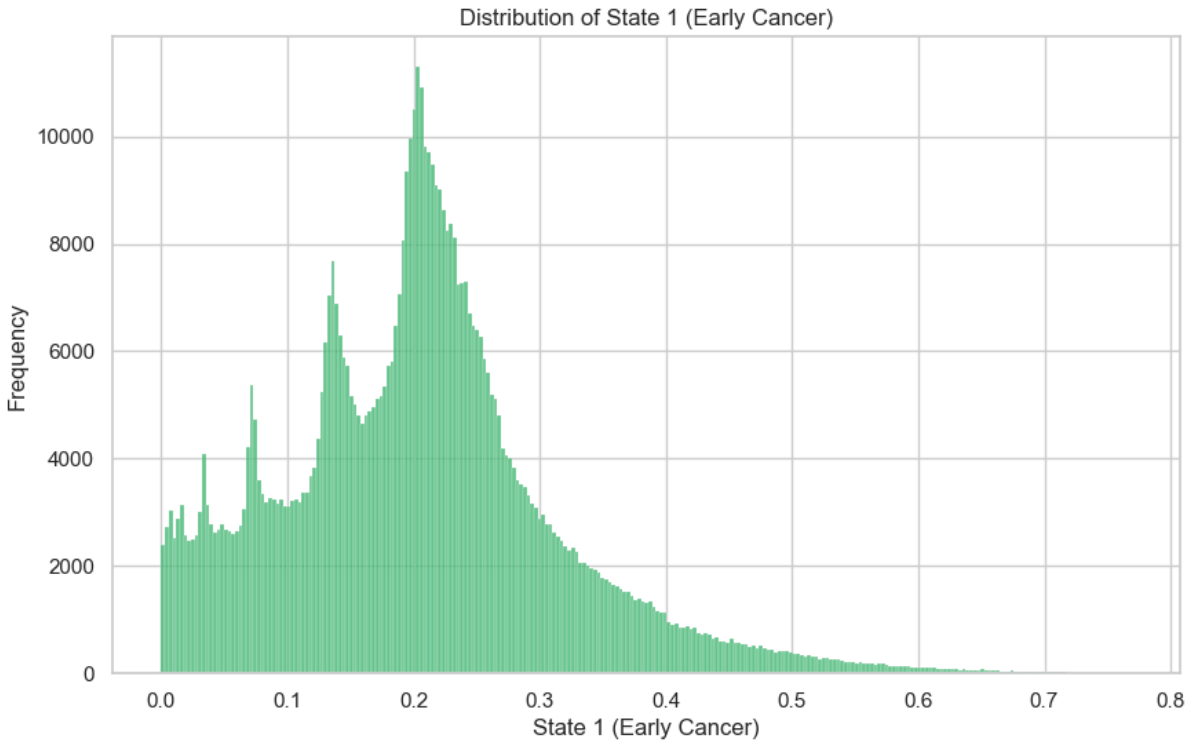

Figure 1: Distribution of Feature - State 0(Early Cancer)

- Risk Score - Risk score is the current probability of cancer for patient. It is calculated using [(1 - probability of being healthy) * 100]. The risk score is calculated by converting the probability of breast cancer into a number ranging from 0 to 100. For example, if the probability of patient x in state 0 - Healthy at age t is 0.7500, then probability of breast cancer is 0.2500, then the risk score for patient x at age t is 25.

  This Risk scores will be used in the RL algorithms as state space S, where S = 0, 1, 2,...., 100. Where S = 0 meaning patient has 0 probability of cancer. At any time t patient will be in state s, $s \in \mathbf{S}$.

- state transition probabilities - Data features From State, To state and Probability represents the state transition probabilities at each age from 40 to 100.

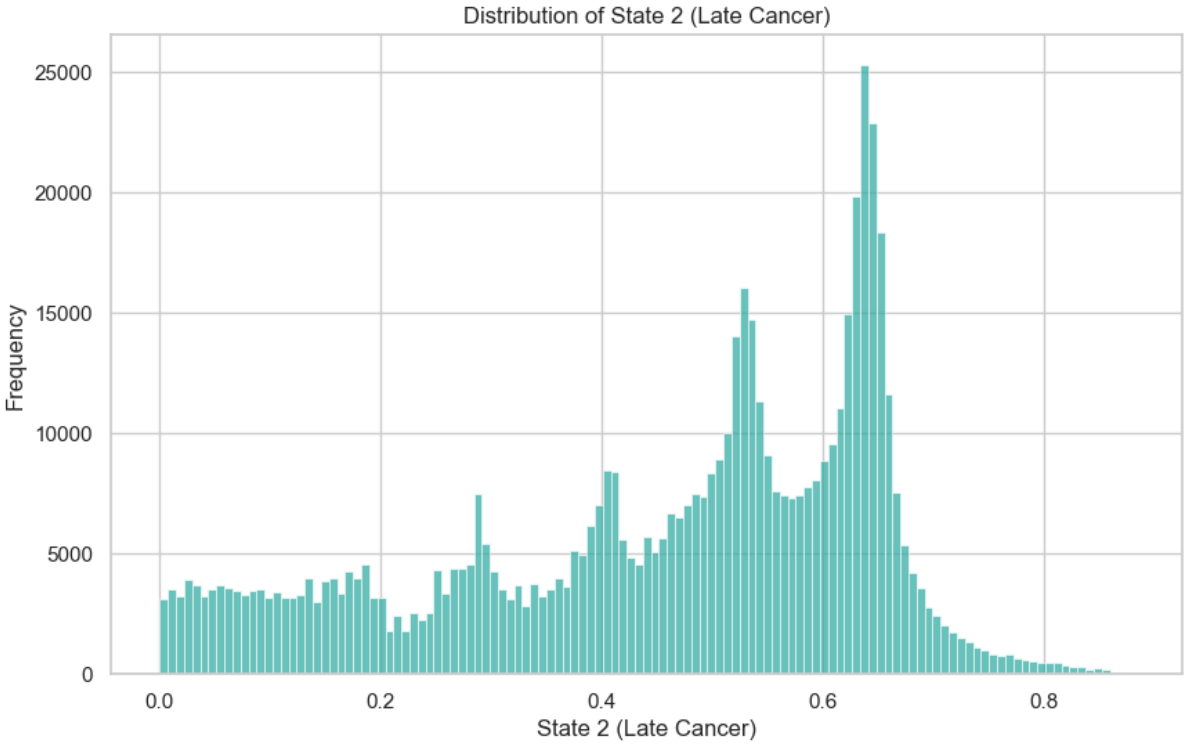

Figure 2: Distribution of Feature - State 1(Advanced Cancer)

For example, if patient at timestamp(age) t = 40 is in stage s1 = 5 (5% probability of breast cancer), then at next timestamp t + 1 = 41, patient being in any state s2, $s2 \in \mathbf{S}$ is probability P. In Figure 3, example of transition probabilities are given. This probabilities are taken from study by Maillart et al. (2008) [21]. This model has only 5 stages and they have probabilities for age groups such as 40 - 44, 45-49, etc.

$$\alpha \in [45, 49] \qquad \alpha \in [50, 54]$$

$$P_\alpha = \begin{pmatrix} 0.99784 & 0.00097505 & 0 & 0 & 0.0011804 \\ 0 & 0.79049 & 0.20833 & 0 & 0.0011804 \\ 0 & 0 & 0.87870 & 0.12011 & 0.0011804 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad P_\alpha = \begin{pmatrix} 0.99703 & 0.0012814 & 0 & 0 & 0.0016911 \\ 0 & 0.86317 & 0.13514 & 0 & 0.0016911 \\ 0 & 0 & 0.85732 & 0.14099 & 0.0016911 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Figure 3: Example of sample transition probabilities

Same way, our data has transition probabilities at each age 40, 41,..., 100 from stat s1 to s2 where $s1, s2 \in \mathbf{S}$. These probabilities are available in "state_transition_probabilities.csv" file due to large scale.

- Rewards - In Reinforcement Learning, an RL agent receives rewards for its actions, aiming to maximize the cumulative rewards to gain the greatest overall benefit.

For this project, I have developed an RL agent with two possible actions: Annual Mammogram and Biopsy. At any age between 40 and 100, the agent receives rewards based on the probability of cancer (either early or advanced stage) and the action taken. These rewards are derived with reference from multiple studies (mentioned in the Literature Review) and general research on breast cancer and biopsies. The rewards are expressed in terms of the total expected years of life following each action. If the action is an Annual Mammogram, the patient receives a reward ranging from 1 year to a few months, depending on their age and cancer probabilities, until the next screening. If the action is a Biopsy, the patient receives a one-time lump sum reward in terms of the total expected life years after the biopsy.

As shown in Figure 4 and 5, for the action Annual Mammogram, at a very young age and with a low probability of cancer, the reward is given as 1 year and decreases with age. Similarly, the rewards for Biopsy also decrease with age. The rewards also diminish as cancer probabilities (early stage or advanced stage) increase. Also, Early and Advance stages of cancer varies with different age groups. i.e., for very young women, small probability of cancer like 2% is advanced whereas with increase in age it can be higher[29].

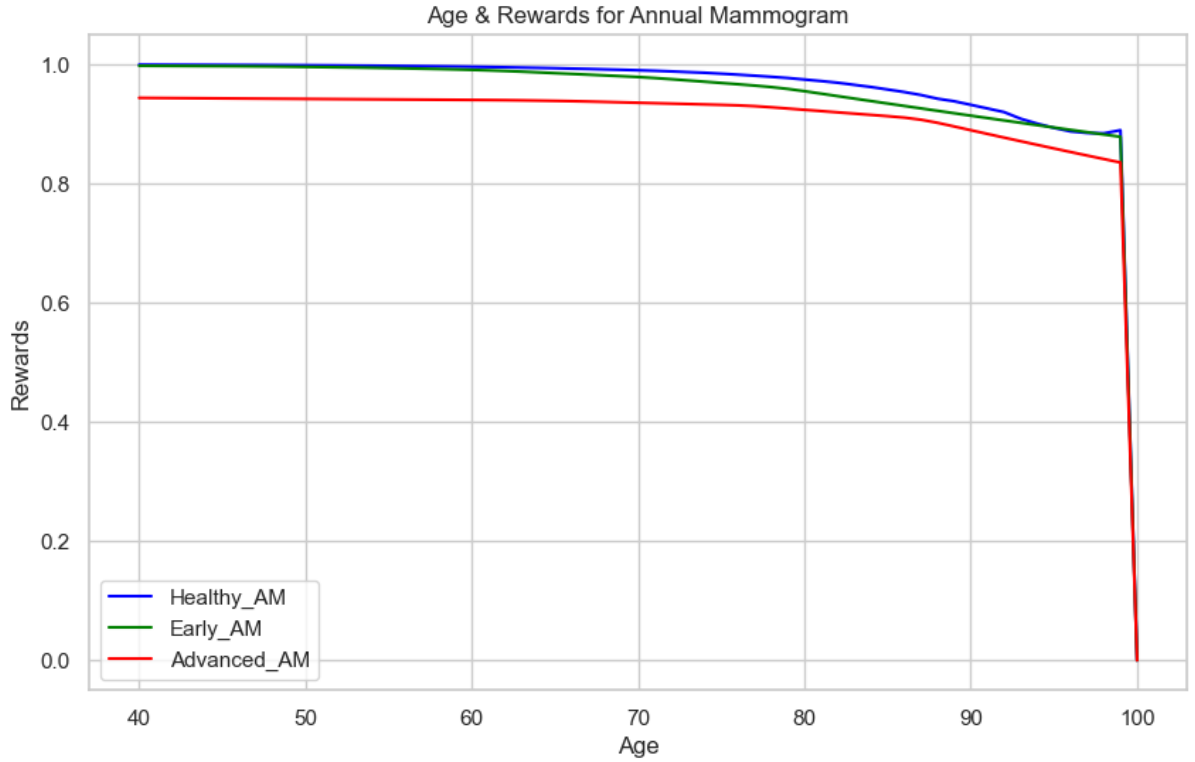

Figure 4: Age - Rewards for Healthy (s = 0), Early & Advance stages with action - Annual Mammogram

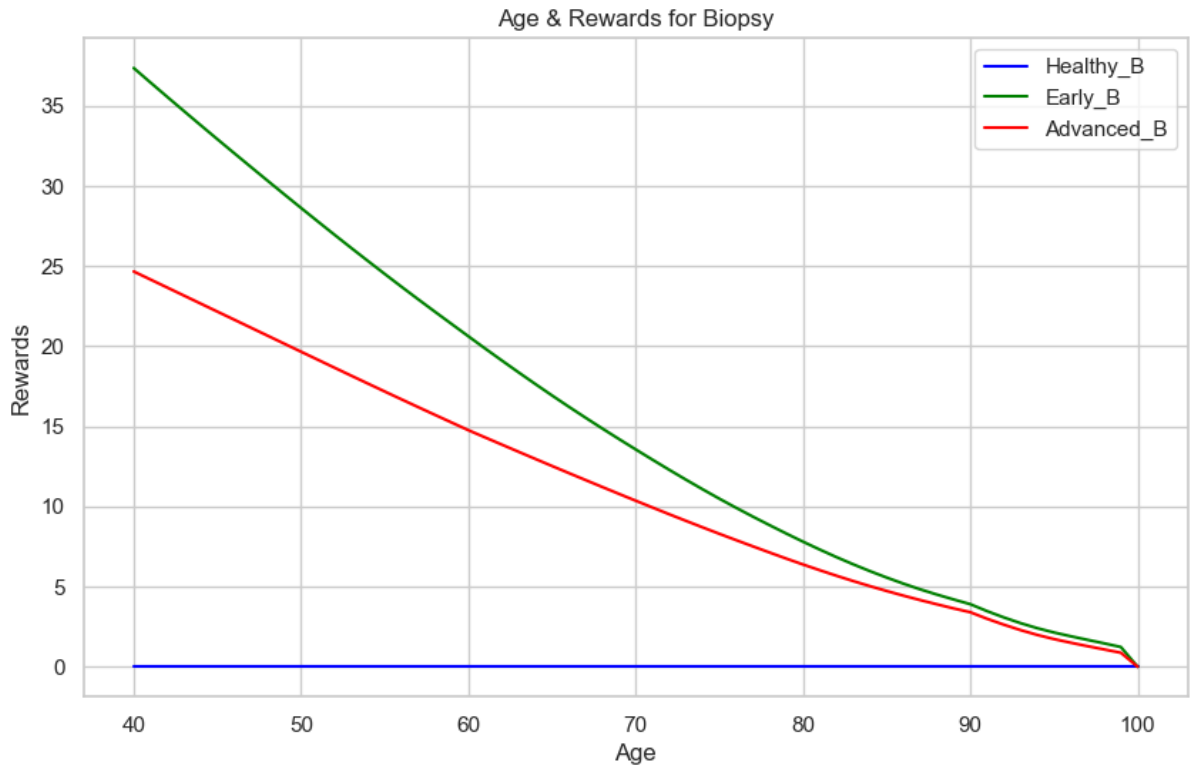

Figure 5: Age - Rewards for Healthy (s = 0), Early & Advance stages with action - Biopsy

C. Correlation between age and Probability of being Healthy

From the data, it is observed that probability of being Healthy decreases with increase in age. Same way, probability of cancer increases with increase in age. In Figure 6, it can be seen that If we take average probability of being healthy and cancer for each age, with age Chances of Being Healthy decreases and Chances of cancer increases. This figure has mean probability for patients from dataset at each age.

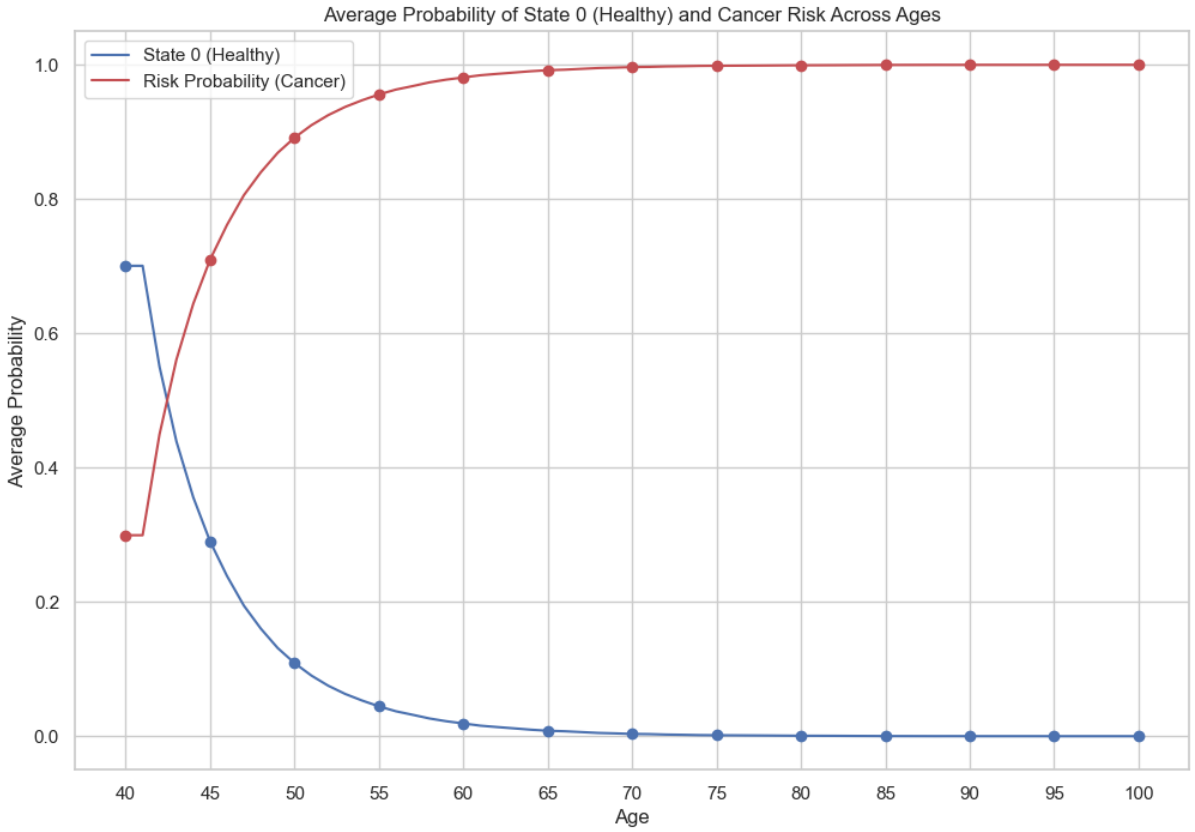

Figure 6: Average Probability of State 0 (Healthy) and Cancer Risk Across Ages

References

[1] Alagoz Ayer and Stout. "Personalized Mammography Screening Operations Research". In: *Operations Research Articles in Advance INFORMS* (2012), pp. 1–16.

[2] T. Ayer et al. "Breast cancer risk estimation with artificial neural networks revisited: Discrimination and calibration". In: *Cancer* 116.14 (2010), pp. 3310–3321.

[3] Alagoz Ayvaci and Burnside. "Optimizing Breast Cancer Diagnosis Under Resource Constraints Manufacturing & Service Operations Management". In: *INFORMS* 14.4 (2012), pp. 600–617.

[4] C. Baines and J. R. Dayan. "A tangled web: Factors likely to affect the efficacy of screening mammography". In: *J. National Cancer Inst.* 91.10 (1999), pp. 833–838.

[5] J. A. Baker et al. "Breast cancer: Prediction with artificial neural network based on BI-RADS standardized lexicon". In: *Radiology* 196.3 (1995), pp. 817–822.

[6] W. E. Barlow et al. "Accuracy of screening mammography interpretation by characteristics of radiologists". In: *J. National Cancer Inst.* 96.24 (2004), pp. 1840–1850.

[7] C. A. Beam, P. M. Layde, and D. C. Sullivan. "Variability in the interpretation of screening mammograms by U.S. radiologists. Findings from a national sample". In: *Arch. Intern. Med.* 156.2 (1996), pp. 209–213.

[8] R. E. Bird, T. W. Wallace, and B. C. Yankaskas. "Analysis of cancers missed at screening mammography". In: *Radiology* 184.3 (1992), pp. 613–617.

[9] *Breast Cancer transition Probabilities*. URL: `https://github.com/atlas3107/Breast_Cancer_transition_probabilities`.

[10] E. S. Burnside et al. "A probabilistic computer model developed from clinical data in the national mammography database format to classify mammography findings". In: *Radiology* 251.3 (2009), pp. 663–672.

[11] E. S. Burnside et al. "Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results: Initial experience". In: *Radiology* 240.3 (2006), pp. 666–673.

[12] Radiology Business. *Breast imaging*. URL: `https://radiologybusiness.com/topics/medical-imaging/womens-imaging/breast-imaging/3d-mammography-approaching-50-percent-breast`.

[13] M. Cevik et al. "Analysis of Mammography Screening Policies under Resource Constraints". In: *Production and Operations Management* 0.0 (2018), pp. 1–24.

[14] Chhatwal, Alagoz, and Burnside. "Optimal Breast Biopsy Decision-Making". In: *Operations Research INFORMS* 58.6 (2010), pp. 1577–1591.

[15] J. Chhatwal et al. "A logistic regression model based on the national mammography database format to aid breast cancer diagnosis". In: *Amer. J. Roentgenol* 192.4 (2009), pp. 1117–1127.

[16] H. Dong, Z. Ding, and S. Zhang. "Deep Reinforcement Learning: Fundamentals, Research and Applications editors (1st ed.)" In: *Springer, Singapore* (2020).

[17] B. L. Fowble et al. "The influence of young age on outcome in early stage breast cancer". In: *Internat. J. Radiat. Oncol. Biol. Phys.* 30.1 (1994), pp. 23–33.

[18] F.D. Hengst et al. "Reinforcement learning for personalization: a systematic literature review". In: *Data Sci.* 3.2 (2020), pp. 107–147.

[19] J. L. Jesneck et al. "Optimized approach to decision fusion of heterogeneous data for breast cancer diagnosis". In: *Medical Phys.* 33.8 (2006), pp. 2945–2954.

[20] L. Liberman and J. H. Menell. "Breast imaging reporting and data system(BI-RADS)". In: *Radiol. Clin. North Amer* 40.3 (2002), pp. 409–430.

[21] Lisa M. Maillart et al. "Assessing Dynamic Breast Cancer Screening Policies". In: *Operations Research INFORMS* 56.6 (2008), pp. 1411–1427.

[22] MedicineNet. *Breast Biopsies*. URL: `https://www.medicinenet.com/percentage_of_breast_needle_biopsies_are_cancer/article.htm`.

[23] H. D. Nelson et al. "Screening for Breast Cancer: An Update for the U.S. Preventive Services Task Force". In: *Ann. Internal Medicine* 151.10 (2009), pp. 727–737.

[24] W.B. Powell. "Introduction to approximate dynamic programming W.B. Powell (Ed.), Approximate Dynamic Programming (2nd ed.)" In: *John Wiley & Sons, New York* (2011), pp. 111–165.

[25] BI-RADS. "Breast Imaging Reporting and Data System (BI-RADS) 3rd ed." In: *American College of Radiology, Reston, VA* (1998).

[26] E. A. Sickles, D. E. Wolverton, and K. E. Dee. "Performance parameters for screening and diagnostic mammography: Specialist and general radiologists". In: *Radiology* 224.3 (2002), pp. 861–869.

[27] R. A. Smith et al. "Cancer Screening in the United States, 2010: A Review of Current American Cancer Society Guidelines and Issues in Cancer Screening". In: *CA: A Cancer J. for Clinicians* 60.2 (2010), pp. 99–119.

[28] R. Smith-Bindman et al. "Comparison of screening mammography in the United States and the United Kingdom". In: *JAMA* 290.16 (2003), pp. 2129–2137.

[29] American Cancer Society. *Cancer Staging*. URL: `https://www.cancer.org/cancer/diagnosis-staging/staging.html`.

[30] Joseph Stember and Hrithwik Shalu. "Deep reinforcement learning to detect brain lesions on MRI: a proof-of-concept application of reinforcement learning to medical images". In: *Production and Operations Management* (2020).

[31] R.S. Sutton and A.G. Barto. "Reinforcement Learning: An Introduction (2nd ed.), MIT Press, Cambridge, Mass". In: *Adaptive Computation and Machine Learning* (2018).

[32] C. Szepesvári. "Algorithms for reinforcement learning". In: *Synth. Lect. Artif. Intell. Mach. Learn.* 4.1 (2010), p. 103.

[33] S. Tunç, O. Alagoz, and E. S. Burnside. "A new perspective on breast cancer diagnostic guidelines to reduce overdiagnosis". In: *Production and Operations Management* 31 (2022), pp. 2361–2378.

[34] WHO. *Breast Cancer*. URL: `https://www.who.int/news-room/fact-sheets/detail/breast-cancer`.

[35] Chan-Yun Yang et al. "Reinforcement learning strategies in cancer chemotherapy treatments: A review". In: *Computer Methods and Programs in Biomedicine* 229 (2023).