Title
A Multi-part Optimization Framework for POMDPs in Lung Cancer Screening

Permalink
https://escholarship.org/uc/item/6t43g1n2

Author
Han, Simon X

Publication Date
2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

A Multi-part Optimization Framework for POMDPs in Lung Cancer Screening

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Biomedical Engineering

by

Simon Han

2023

ABSTRACT OF THE DISSERTATION

A Multi-part Optimization Framework for POMDPs in Lung Cancer Screening

by

Simon Han

Doctor of Philosophy in Biomedical Engineering

University of California, Los Angeles, 2023

Professor Alex Anh-Tuan Bui, Co-Chair

Professor Denise R. Aberle, Co-Chair

Currently, low-dose computed tomography (LDCT) is the only recommended screening test for patients who are at high risk of lung cancer. However, the cost-benefit analysis of LDCT must be weighed against the high number of false positives, radiation exposure, unnecessary procedures, and the associated patient distress as a result of the aforementioned possibilities. Sequential decision making models such as the partially observable Markov decision process (POMDP) have seen success in making recommendations in clinical applications such as lung cancer screening. Enabled by the availability of longitudinal datasets that track patient health over time, these models make predictions toward long-term health outcomes. A key challenge in lung cancer screening is the balancing between true positives and false positives, that is, maximizing true positives while minimizing false positives. This dissertation attempts to address this challenge by leveraging a variety of techniques toward optimizing decision making over time. First, the modularized POMDP (MODPOMDP) framework is developed to account for temporal variations within a POMDP model. Each time point is optimized separately to ensure "earlier" detection by maximizing positive predictions over

the entire screening duration. Second, a two-part model framework (MODPOMDP2) is developed to differentiate true and false positive predictions from each other. This method combines classic machine learning (ML) techniques and MODPOMDP to maintain true positives while decreasing false positives. Third, the validity of these approaches is demonstrated in an external testing dataset.

The dissertation of Simon Han is approved.

Corey Wells Arnold

William Hsu

Denise R. Aberle, Committee Co-Chair

Alex Anh-Tuan Bui, Committee Co-Chair

University of California, Los Angeles

2023

To my family and friends.

Life is a finite horizon POMDP, choose a sensible rewards function.

— PANAYIOTIS PETOUSIS

Try not to become a person of success but rather try to become a person of value.

— ALBERT EINSTEIN

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGMENTS

Rina (Ruiwen) Ding, Jennifer Polson, Wenyuan Li, Yiwen Meng, Leihao Wei, Anna (Juan) Wu, Kyle Singleton, Jean Garcia-Gathright, and Mary Zide. I am also deeply appreciative of the many other friends I have met at UCLA: Fei Han, Awella (Ai) Ye, Chunan Liang, Harsha Kittur, Janay Kong, Mayuri Kerr, James Che, Mrinal Rath, Mayank Jog, and many more.

Lastly, I could not have completed this journey without the support of my family. My mother, Amy Guo, who raised me and whom I know will always be there for me. James Freeman, for being such a role model in my life. Hui-Hsuan Huang, my partner, best friend, and wife. My son, Andrew James Han, whose infectious laughter fill me with joy. I feel incredibly lucky to have you all in my life.

<div align="right">

Simon X. Han

Cerritos, CA

10 March 2023

</div>

2005–2009   B.S. Bioengineering, UCSD, La Jolla.

2009–2011   Database Manager, eMolecules, San Diego.

2011–2013   M.S. Biomedical Engineering, UCLA, Los Angeles.

2016–2022   Graduate Student Researcher, Medical & Imaging Informatics,
             Department of Radiology, UCLA, Los Angeles.

PUBLICATIONS

Abtin F, Suh RD, Nasehi L, **Han SX**, Hsu W, Quirk M, Genshaft S, Gutierrez AJ, Cameron RB. Percutaneous cryoablation for the treatment of recurrent thymoma: preliminary safety and efficacy. Journal of Vascular and Interventional Radiology. 2015 May 1;26(5):709-14.

Hsu W, **Han SX**, Arnold CW, Bui AA, Enzmann DR. A data-driven approach for quality assessment of radiologic interpretations. Journal of the American Medical Informatics Association. 2016 Apr 1;23(e1):e152-6.

Petousis P, **Han SX**, Aberle D, Bui AA. Prediction of lung cancer incidence on the low-dose computed tomography arm of the National Lung Screening Trial: A dynamic Bayesian network. Artificial Intelligence in Medicine. 2016 Sep 1;72:42-55.

Shen S, **Han SX**, Petousis P, Weiss RE, Meng F, Bui AA, Hsu W. A Bayesian model for

estimating multi-state disease progression. Computers in Biology and Medicine. 2017 Feb 1;81:111-20.

Abtin F, Quirk MT, Suh RD, Hsu W, **Han SX**, Kim GH, Genshaft S, Sandberg JK, Olevsky O, Cameron RB. Percutaneous cryoablation for the treatment of recurrent malignant pleural mesothelioma: safety, early-term efficacy, and predictors of local recurrence. Journal of Vascular and Interventional Radiology. 2017 Feb 1;28(2):213-21.

Bakir M, Jackson NJ, **Han SX**, Bui A, Chang E, Liem DA, Ardehali A, Ardehali R, Baas AS, Press MC, Cruz D. Clinical phenomapping and outcomes after heart transplantation. The Journal of Heart and Lung Transplantation. 2018 Aug 1;37(8):956-66.

Petousis P, **Han SX**, Hsu W, Bui AA. Generating reward functions using IRL towards individualized cancer screening. In Artificial Intelligence in Health: First International Workshop, AIH 2018, Stockholm, Sweden, July 13-14, 2018, Revised Selected Papers 1 2019 (pp. 213-227). Springer International Publishing.

Shen S, **Han SX**, Aberle DR, Bui AA, Hsu W. An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification. Expert Systems with Applications. 2019 Aug 15;128:84-95.

Lin Y, Wei L, **Han SX**, Aberle DR, Hsu W. EDICNet: An end-to-end detection and interpretable malignancy classification network for pulmonary nodules in computed tomography. In Medical Imaging 2020: Computer-Aided Diagnosis 2020 Mar 16 (Vol. 11314, pp. 344-355). SPIE.

CHAPTER 1

Introduction

1.1 Motivation

As more biomedical data become available, especially longitudinal datasets, researchers have gained an unprecedented ability to track patient health over time. Prognostic models can incorporate these data towards long-term health outcome predictions. In particular, sequential decision making under a partially observable context is uniquely suited to support such medical decision making problems as the true health state of a patient is often unknown, requiring series of external tests that elucidate the belief about possible conditions. Moreover, such clinical tests and observations have an inherent amount of sensitivity and detection error (i.e., false positive/negative rates). Partially observable Markov decision processes (POMDPs) aim to determine an optimal set of actions over time, referred to as a *policy*, maximizing a desired outcome.

In the healthcare setting, for instance, POMDPs are used to create policies to guide lung cancer screening [1, 2], balancing detection rates, quality of life, and cost. While lung cancer incidence and mortality have been decreasing for decades, it generally has a poor prognosis and is the number one death-causing cancer. There will be an estimated 238,340 new lung cancer cases representing 12.2% of all new cancers and 127,070 lung cancer deaths representing 20.8% of all cancer-related mortality in the United States in 2023 [3]. Since 2013, the U.S. Preventive Services Task Force (USPSTF) recommends annual low-dose computed tomography (LDCT) screening for high-risk patients [4], with an updated high-risk definition

in 2021 [5]. A meta-analysis found that the false positive rate of LDCT at baseline screening ranged between 7.9% to 49.3% among 27 publications [6]. Excess screening can lead to radiation-induced cancer, unnecessary tests such as invasive procedures, overdiagnosis, incidental findings, and increases in patient distress. At the same time, missing a true positive cancer case can incur a large quality of life cost on the patient. Thus a POMDP-based lung cancer screening model must optimize towards detecting true positives earlier while reducing false positives over time.

Predicated on a stationarity assumption, classic POMDPs assume that the transition, observation, and reward submodels are time-invariant. *However, in healthcare this assumption is often untrue.* In breast cancer screening for example, the incidence rate is highly dependent on age, thus affecting transition probabilities over time. Likewise, the accuracy of mammograms depend on the age of the woman screened and other time-varying features. The rewards model, if framed as quality-adjusted life years (QALY), not only depends on the age of the patient, but on patient preferences (e.g., the disutility of the imaging exam may outweigh potential benefits for older women). Currently, researchers applying POMDP models to healthcare problems use a limited simulation-based approach (e.g., enumerating all sample paths) or do not explicitly address these changes over time. A systematic strategy is needed to explicitly represent changing transition, observation, and rewards in POMDPs over the expected modeling period, adapting over time. This problem also extends to the domain of lung cancer screening.

1.2 Contributions

This dissertation addresses the optimization of POMDP predictions by fulfilling the following two aims:

- **Aim 1**: *Improving POMDP predictions by optimizing each modularized POMDP submodel.* **Aim 1** investigates the use of POMDPs under a novel framework, namely the

decomposition of POMDP problems into time-dependent constituent components. This is the **modularized POMDP** (**modPOMDP**) framework. Formally, a MODPOMDP model is composed of I POMDP submodels joined in a sequential fashion, where each $POMDP_i$ has a defined applicable time range $\sigma_i \in \Sigma$. Each submodel can take on different transition, observation, and reward functions that better characterize an evolving disease trajectory and its progressive assessment and treatment. In a domain-specific manner, the values for I and Σ depend on the decision-making problem, as well as the availability of data. In effect, Σ reflects the time duration over which the problem or disease is considered (e.g., a 5-year perspective vs. lifetime), while I captures the temporal granularity based on the frequency of available observations (e.g., biannual vs annual screening). Optimization of each $POMDP_i$ is in itself a balancing act, but this aim will primarily bias towards recall for capturing positive cancer cases.

- **Aim 2**: *Optimizing POMDP positive predictions with classifiers in a two-part model approach.* **Aim 2** further explores optimization in POMDPs through use of a two-part model approach. While two-part models are not commonly seen with electronic health records (EHR) data, they are more common in other artificial intelligence (AI) fields. One application is in highly skewed data in which one model reduces the search space and the second model optimizes on the remaining samples. In this work, the two-part model consists of MODPOMDP from **Aim 1** and classifiers common in the machine learning (ML) field. More specifically, strategies are developed for optimizing classifiers and experiments are designed for selecting classifiers. This two-part MODPOMDP is collectively called MODPOMDP2.

1.3 Organization

The remaining chapters of this dissertation are organized as follows:

- Chapter 2 describes the background of sequential decision making, such as POMDPs

and their applications in healthcare. Methods of building and solving POMDPs are also described, including the datasets used in the evaluation of this work.

- Chapter 3 describes the design and implementation of MODPOMDP as stated in **Aim 1**.
- Chapter 4 describes the design and implementation of MODPOMDP2 as stated in **Aim 2**.

Finally, Chapter 5 summarizes the contributions and findings from this dissertation and future directions that can be extended from this research.

CHAPTER 2

Background

2.1 Sequential Decision Making

Sequential decision making (SDM) is concerned with how a decision maker (agent) interacts in a modeled world (environment) and makes a series of decisions that optimizes an outcome [7]. Each decision may have immediate and long-term consequences that depends heavily on the future objective. The agent must take into account these future possibilities and at the same time balance the current and later states to achieve the desired goal. As such, SDM has a significant role in a range of clinical decision-making tasks (e.g., selection of a set of tests to perform to quickly diagnose a patient; series of actions to optimally treat a disease, etc.). SDM encompasses a variety of different approaches, briefly covered below.

2.1.1 Multi-armed Bandits

One type of SDM is the multi-armed bandit (MAB) problem, or simply bandit problem. In its classical formulation, it is defined as a set of actions (also called bandits, arms, machines, etc.), each with an associated independent reward distribution [8–14]. At each iteration, the agent picks one action to take and receives some reward from the corresponding distribution while other actions remain inactive and receive no reward. The strategy for picking the action is referred to as the *policy*. Through successive actions, the agent explores the rewards distributions and determines the one that returns the highest reward while updating the parameters of the policy. This process, however, introduces the exploration-exploitation

trade-off problem: the agent must balance between gathering information by taking actions with lower reward distributions and picking the action currently known to give the highest reward. An extension of the bandit problem is the contextual bandit problem where the action taken depends on a set of covariates [15–17]. The bandit family of approaches has been used in many different fields, such as personalized advertisement and news placement [18, 19]. In healthcare, it is traditionally used in patient allocation of clinical trials [8, 20]. More recently, it was discussed in the context of clinical decision support in breast cancer screening [21] and for delivering personalized interventions in mobile health (mHealth) [22].

2.1.2 Markov Decision Process

Another approach to SDM is the Markov decision process (MDP) [7]. MDPs can be described by a four element tuple that consist of: 1) states that describe the world; 2) actions that describe possible interactions with the environment; 3) transitions that are a probabilistic description of moving from one state to another having taken an action; and 4) immediate rewards of having taken an action in a particular state. In breast cancer screening for instance, states could be stages of cancer; actions could be to have a mammogram or wait; transitions could be the natural progression of disease; and rewards could be the expected quality-adjusted life years (QALY). In an MDP, transitions satisfy the Markov assumption such that future states only depend on the current state for a given action. Solving an MDP involves producing a mapping from states to actions [23], that is, the action that is expected to give the highest reward in a given state. The solution to the MDP (i.e., policy) is a set of actions derived from this mapping and the policy that gives the highest reward is the optimal policy. The resulting policy depends on the *horizon*, which is the number of decisions the agent is expected to make. When there is a fixed number of horizons, the MDP is referred to as finite-horizon problem. Conversely, an infinite horizon MDP refers to a situation when there is no expected termination. The ability to maximize long-term rewards is one reason for the popularity of MDPs, as the agent can forgo short-term rewards

to achieve a greater goal. In medicine, MDPs have been used in a variety of problems such as when to recommend liver transplantation [24], breast cancer biopsy [25, 26], breast cancer screening [27], and intervention for a joint breast and ovarian cancer problem [28].

2.1.3 Partially Observable Markov Decision Process

One requisite assumption of MDPs is that the agent knows the environment completely — there is no uncertainty regarding the agent's observations. In real-world applications and in particular medicine, the true health state of a patient is often unknown, and clinical observations have an inherent degree of uncertainty (e.g., a test's false positive rate). An extension of MDPs, called partially observable Markov decision processes (POMDPs), permits uncertainty over the state space by adding two more elements: 1) the set of possible observations available to the agent; and 2) the probability of making an observation while in a state [29, 30]. Extending the prior breast cancer screening example, observations could be results of mammogram exams (e.g., BI–RADS [31] scores) and observation probabilities could be the probability that a BI–RADS score is some cancer stage. Because a POMDP agent cannot directly access states to produce a mapping from states to actions, the POMDP agent instead maintains a probability distribution over the states, called a *belief*. Thus, the goal of a POMDP policy is to derive a mapping of beliefs to actions and does not depend on the state itself [7]. POMDPs too, have been used in a variety of medical problems such as when to recommend breast cancer biopsy [25], breast cancer screening [32], colorectal cancer screening [33, 34], prostate cancer biopsy [35], prostate cancer screening [36], antibiotic administration for sepsis [37], surgery for Parkinson's disease [38], amputation for diabetic foot disease [39], and intervention for ischemic heart disease [40].

Formally, a POMDP is defined as a tuple $\langle S, A, T, R, \Omega, O \rangle$ consisting of:

- States $s \in S$, which represent the possible states for a given phenomenon/process to be in (e.g., a disease state);

- Actions $a \in A$, which are acts that the agent can take within the environment;

- Transition function $T(s, a, s') = \Pr(s'|s, a)$, the probability of ending in state s' for taking action a in state s;

- Reward function $R(s, a)$, the immediate reward for taking action a in state s;

- Observations $o \in \Omega$, which define the observations the agent can make about the environment; and

- Observation function $O(s', a, o) = \Pr(o|s', a)$, the probability of observing observation o for taking action a and ending in state s'.

The state, action, and observation sets are typically finite. The agent is expected to reason about the system by selecting actions that maximize expected rewards over time. Given that the underlying state space is unknown to the agent, the agent reasons using a *belief*. A belief state b is a probability distribution over S, where $b(s)$ is the probability of being in state s. The belief state space is represented by an n-dimensional triangle, the $(|S| - 1)$-dimensional simplex Δ. All belief states rest within the simplex and have the property that $\sum_{s \in S} b(s) = 1$, where $0 \le b(s) \le 1$ for all $s \in S$. Given an existing belief state b, an action a, and an observation o, a state-estimation function $\mathrm{SE}(b, a, o)$ is used to update the belief. Each $b'(s')$ can be updated via:

$$b'(s') = \frac{O\left(s', a, o\right) \sum_{s \in S} T\left(s, a, s'\right) b\left(s\right)}{\Pr\left(o|a, b\right)} \tag{2.1}$$

where $\Pr(o|a, b)$ is a normalizing factor that ensures b' sums to 1. That belief update requires only the current belief state b demonstrates that POMDPs are Markovian processes. Prior to solving, a POMDP is often recast as a MDP over the belief space, or *belief MDP*. The belief MDP is defined as $\langle \Delta, A, \tau, \rho \rangle$ consisting of:

- Belief states $b \in \Delta$, which are the set of possible belief distributions over the POMDP states;

8
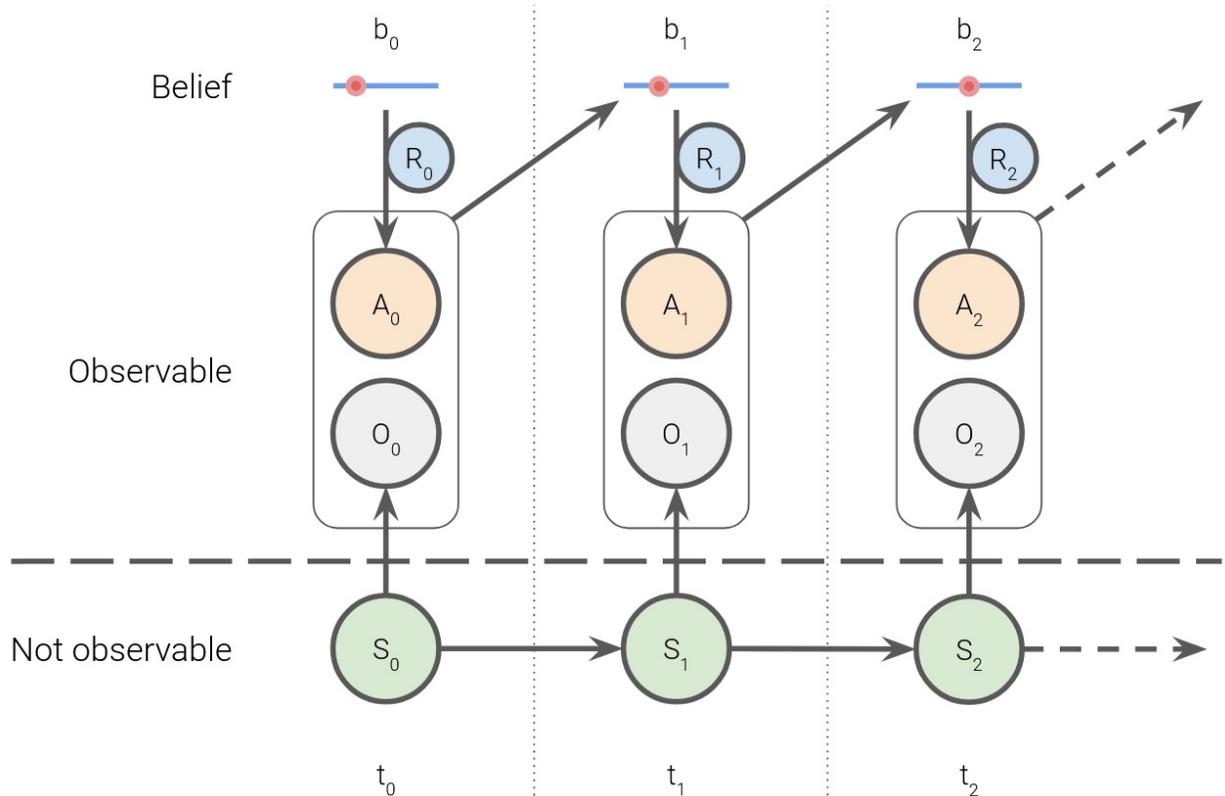
Figure 2.1: A simplified representation of a POMDP over 3 time points illustrating the unobservable nature of the underlying state S_t, which emits an observation O_t. An optimal action A_t is selected based on belief b_t while generating a reward R_t. b_{t+1} is then updated based on O_t and A_t. Dashed arrows indicate transitions to a future time point (not shown).

- Actions $a \in A$, which are acts that are the same as the POMDP;

- Transition function $\tau(b, a, b') = \Pr(b'|a, b) = \sum_{o \in \Omega} \Pr(b'|b, a, o) \Pr(o|a, b)$, where:

$$\Pr(b'|b, a, o) = \left\{ \begin{array}{ll} 1, & \text{if } \mathrm{SE}\,(b, a, o = b') \\ 0, & \text{otherwise} \end{array} \right\}$$

- Reward function $\rho(b, a) = \sum_{s \in S} b(s) R(s, a)$, for taking action a in belief state b.

Because the belief value is continuous, the belief MDP is effectively a continuous space MDP. Thus, the dynamic programming approach used in solving MDPs applies. The goal of

dynamic programming is to update the value function V_t using a given value function V_{t-1} where t is the number of steps left in the decision process. When applied to POMDPs, the value function can be defined in terms of the belief state:

$$V_t(b) = \max_{a \in A} \left[\rho(b, a) + \gamma \sum_{o \in \Omega} \Pr(o|a, b) V_{t-1}(b') \right] \quad (2.2)$$

where $0 \leq \gamma \leq 1$ is a discount factor to reduce future reward contributions.

Solving a POMDP. Sondik [41] demonstrated that the value function is piece-wise linear and convex (PWLC) and can be represented finitely. Thus, the value function is a set of linear equations, one for each of the discretized regions in the belief simplex Δ. Many exact algorithms [41–43] take advantage of the PWLC property to derive the value function, with value iteration [44,45] being the more popular method. The value function can be represented as:

$$V_t(b) = \max_{\alpha \in \Gamma_t} b \cdot \alpha \quad (2.3)$$

where Γ_t is a set of $|S|$-dimensional alpha vectors α representing the value function at t. Each α is associated with an action a and composed of coefficients of its linear equation. To determine the best action to take for a belief state b at t, calculate the dot product between b and Γ_t and select the α with the highest dot product. The associated action is then the optimal action. Γ_t is then the result and also the representation of the optimal policy π^* at t, since an optimal action is given for any belief $b \in \Delta$. The optimal policy π^* is a deterministic list of actions that when followed, produces the highest reward over a given time frame.

For finite-horizon problems, it is typical to use a non-stationary policy approach. The number of horizons H in value iteration is equal to the number of decisions the agent is expected to make. The rationale is that if $h = 1$, the agent only needs to make one decision and hence value iteration only needs to be solved for one iteration; additional iterations are wasted and $\pi_{h=1}$ is then taken as the optimal policy for $h = 1$. For a finite-horizon problem, the optimal policy is then $\pi^* = \{\pi_H, \pi_{H-1}, \cdots, \pi_1\}$. For infinite-horizon problems, because

the agent continuously makes decisions, the stop condition is different. A discount factor γ is typically used to reduce future reward contributions and value iteration continues until convergence, that is, $|V_h - V_{h-1}| < \epsilon$, where ϵ is a difference metric. Assuming it takes H horizons to reach convergence, the optimal policy in the infinite-horizon case is $\pi^* = \pi_H$.

In value iteration, the number of alpha vectors can grow exponentially and many researchers have improved the efficiency of value iteration through enumeration or pruning methods such as linear support by Cheng [46]; witness algorithm by Littman [47] and Littman et al. [48]; and incremental pruning by Cassandra et al. [49]. More recent approximate POMDP algorithms seek to avoid the exponential growth in the number of alpha vectors and reduce the computational costs of exact methods. Many of these newer approaches [50–55] use a point-based method where alpha vectors are only generated for selected belief points. Point-based methods mainly differ in the way the belief points are selected and the order in which operations are then performed on those belief points. A review of point-based methods can be found in Shani et al. [56].

The QMDP Algorithm. In this work, the POMDPs are solved using an implementation of an approximate algorithm called QMDP [57,58] and the pseudocode is shown in Algorithm 1. In QMDP, the algorithm solves POMDPs by solving for the Q values of the underlying MDP. That is, the observation model is set aside and the POMDP is solved as an MDP by computing $Q(s, a)$ values using transitions and rewards only. $Q(s, a)$ values can then be used to estimate Q values for a belief state b using $Q_a(b) = \sum_s b(s) \text{QMDP}(s, a)$. An optimal action a_{opt} can be selected for a belief b as shown by Algorithm 2. By maximizing over all actions, the action with the highest expected reward is returned at each step, maximizing rewards over time. Thus, QMDP approximates the value function of a POMDP by making use of the computed value function of the underlying MDP. The advantage of QMDP is it has the speed and efficiency of solving MDPs while also suitable for large POMDP problems.

Algorithm 1: QMDP Algorithm; adapted from [2].

Input : S, A, R, ϵ
Output: Q matrix

Computing the Q matrix;
$V(s) = \text{MDP\_VI}(S, A, R, \epsilon)$ # Value Iteration algorithm
for $s_i \in S$ **do**
    for $a \in A$ **do**
        $Q(s_i, a) = R(s, a) + \sum_{s_j \in S} T(s_j, a, s_i)V(s_i)$
Return: Q

Algorithm 2: Action Selection Algorithm; adapted from [2].

Input : Q, b
Output: a_{opt} optimal action

Given belief b;
$a_{opt} = \arg\max_a \sum_{s_i \in S} b(s_i)Q(s_i, a)$
Return: a_{opt}

2.2 Building a POMDP Model

Building a POMDP involves defining and estimating the values of its various components. Recent research in lung and breast cancer screening have used novel ways for this task [1,2,59]. Namely, dynamic Bayesian networks are developed to provide the transition and observation probabilities, and inverse reinforcement learning is applied to learn the reward function from clinicians.

2.2.1 Dynamic Bayesian Networks

Dynamic Bayesian networks (DBNs) [60] are extensions of Bayesian networks (BN) for modeling dynamic processes. That is, variables are modeled over time and can represent the changes in a system over the modeling interval. The probability distribution over time is

represented below, where $X_i^{(t)}$ is a random variable at time t:

$$P\left(X_1, ..., X_n\right) = \prod_{n=1}^{n} P\left(X_i | X_1, ..., X_{i-1}\right) \tag{2.4}$$

A DBN can be represented in two parts: 1) a prior model for the initial distribution and 2) a transition model for the changes to the process across time points:

$$P\left(X^{(t+1)} | X^{(t)}\right) = \prod_{X^{(t+1)} \in \mathbf{X}^{(t+1)}} P\left(X^{(t+1)} | Par\left(X^{(t+1)}\right)\right) \tag{2.5}$$

where $Par\left(X^{(t+1)}\right)$ is the parent. DBNs are used to estimate conditional probabilities which are often in the form of conditional probability tables (CPTs). DBNs in this work are solved using the Bayes Net Toolbox (BNT) [61].

DBNs and BNs have been used in many medical domains such as breast cancer [62,63], lung cancer [64], cervical cancer [65], colon cancer [66], oral cancer [67], among many others [68–72]. A review of changes and advancements of DBNs in recent years can found in [73].

2.2.2 Adaptive Step-size Maximum Entropy Inverse Reinforcement Learning

The reward function is one of the most important components of the POMDP and actively drives the agent's decision making. In SDM models in healthcare, rewards are typically in terms of cost [33, 34, 39, 74–76], mortality risk [77], QALY [78], or even a combination of cost and QALY [79]. However, the cost of healthcare does not necessarily reflect patient outcomes. Moreover, rewards such as mortality and QALY can be difficult to acquire and or define. A recent method of using adaptive step-size maximum entropy (MaxEnt) inverse reinforcement learning (IRL) algorithm to compute the reward function was proposed by Petousis et al. and demonstrated in lung and breast cancer screening [1]. The goal of MaxEnt IRL [80] is to learn the rewards of the POMDP through available data, that is the expert optimal policy π^*. The pseudocode is shown in Algorithm 3.

The reward function r is defined in Equation 2.6. τ, or a trajectory, is a series of state-action pairs where states s is the series of states visited and actions a are the actions that the

---
Algorithm 3: MaxEnt Inverse Reinforcement Learning; adapted from [59].

    Input : MDP and trajectories D
    Output: State-rewards

    Randomly initialize θ;
    repeat
        Solve for the optimal stochastic policy using $r(\tau) = \theta^T f_\tau$ with value iteration;
        Use a dynamic algorithm to compute $p(s|\tau;\theta) = D_s$, the state visitation
         frequencies;
        Compute the gradient $\nabla_\theta L$;
        Update θ;
    until *convergence*;
---

agent took to arrive at those states. f_s represent feature vectors and θ represent the features' respective weights. f_τ is the sum of feature counts across states visited in a trajectory.

$$r\left(\tau;\theta\right) = \theta^T f_\tau = \sum_{s \in \tau} \theta^T f_s \tag{2.6}$$

Algorithm 3 first solves for the optimal policy then solve for the state visitation frequencies. The gradient $\nabla_\theta L$ can be computed with:

$$\nabla_\theta L = \widetilde{f} - \sum_{s_i} D_{s_i} f_{s_i} \tag{2.7}$$

where $\widetilde{f}$ is the average feature counts across all trajectories. Finally, θ is updated with:

$$\theta_{i+1} = \theta_i + \eta \nabla_\theta L \tag{2.8}$$

where η is a learning rate. Adaptive step-size is accomplished by first defining $\eta = \frac{\alpha}{(t+A)^\alpha}$ and t as:

$$t_{i+1} = \max\left(t_i + f\left(\langle -\nabla_\theta L_i, \nabla_\theta L_{i-1}\rangle\right), 0\right) \tag{2.9}$$

$f(\cdot)$ is a sigmoidal function where its constant values, as well as α and A are defined in [81].

A multiplicative rewards model. The output from the IRL algorithm is state-rewards and not the $R(s,a)$ function required in a POMDP. To address this issue, MaxEnt IRL can

be used to solve two MDPs, a state MDP model for $R(s)$ and an action MDP model for $R(a)$. The implementation details of the two MDPs are described in [1]. Both state-rewards are normalized to $[-1, 1]$ and the final rewards function is then computed using:

$$R(s, a) = R(s) \cdot R(a) \tag{2.10}$$

If a POMDP has only two actions, then solving an action MDP with MaxEnt IRL for $R(a)$ is unnecessary. As a result of normalization, $R(a)$ would become $[-1, 1]$ regardless of the output from MaxEnt IRL. This is likewise the case for $R(s)$ if the POMDP only has two states.

2.3 Data

The two datasets used in this work are the National Lung Screening Trial (NLST) [82] and the Integrated Diagnostics (IDx)[1] lung cancer screening database at UCLA (hereafter referred to simply as UCLA data). Both datasets are collected over time and provide longitudinal patient data relevant for an SDM model. The NLST dataset was used to develop the MODPOMDP as well as the MODPOMDP2 model. The UCLA data was used towards external validation.

2.3.1 NLST

The NLST was a randomized controlled trial (RCT) conducted in 33 institutions across the United States with the aim of comparing the effectiveness of screening with low-dose computed tomography (LDCT) and chest x-ray (CXR) at reducing lung cancer-specific mortality. $53,454$ current or former smokers considered to be high risk for developing lung cancer were recruited into the study between August 2002 and April 2004. Eligibility at baseline were between 55–74 years old, ≥ 30 pack-years of smoking history, < 15 years of quit time for

---

[1]https://diaag.medsch.ucla.edu/index.html

former smokers, no history of lung cancer, and among others. Participants were randomized into either the LDCT or CXR arm and had up to 3 screenings with ~1 year between screenings. Screenings were conducted between 2002 and 2007, with follow-ups through December 2009. Findings suspicious for lung cancer (e.g., nodules $\geq$ 4 mm) were referred for further follow-up and were counted as positive for lung cancer if confirmed through a diagnostic procedure (e.g., biopsy). Participants confirmed for lung cancer were given medical follow-up and did not receive further screenings in the trial.

2.3.2 UCLA

The IDx Research Program is an initiative of the Department of Radiological Sciences, David Geffen School of Medicine at UCLA. This group provided lung cancer screening data for patients at UCLA. Unlike the NLST, patients at UCLA are not part of a RCT and not a selected group of high-risk individuals. The last data pull was provided in September 2021 and contained a total of 4,966 patients with screenings spanning decades. The exact years are unknown because the provided dataset anonymized patient data including shifting all dates. There were 11,471 reports associated with a lung nodule and 20,499 unique nodules. 3,185 patients answered a lung cancer survey of demographics and medical history questions. The population was very diverse in age, from 18–99 years old, with 2,580 patients between 55–74 years old.

CHAPTER 3

modPOMDP: The Modularized POMDP

3.1 Introduction

In the context of Markov models, stationarity means that the transition probabilities do not change over time [83]. In multi-armed bandits (MABs), the literature on stationarity is usually in terms of the environment, while in Markov decision processe (MDPs) and partially observable MDPS (POMDPs), the literature on stationarity is discussed in terms of the policy [7] or environment. In MDPs and POMDPs, a stationary policy does not change with different horizons (i.e., it is often a feature of infinite-horizon problems). In a finite-horizon problem, while the optimal policy could be stationary, it is not guaranteed and the mapping from state or belief to action could change as the MDP or POMDP is being solved [84]. The non-stationary policy then, selects actions from the policy at each intermediate time step. The research in this chapter is primarily concerned with stationarity of the environment and will be referred to as stationarity unless otherwise specified.

In many real-world applications, stationarity is often not true. As such, recently researchers have started investigating reward distributions in MABs that change over time and how to balance the exploration-exploitation trade-off in a changing environment [85–87]. In dynamic Bayesian networks (DBNs), others have started exploring non-stationary DBNs (i.e., Bayesian networks whose structure and parameters evolve over time) [88–92]. In MDPs and POMDPs, stationarity (e.g., time-invariant transitions) are often assumed. Indeed, the formal definition of a POMDP by Monahan [29] states that a POMDP's core process is

to be a finite state Markov chain with stationary transition probabilities. The problematic nature of this assumption, especially in a biomedical context, has been pointed out by several [24, 32].

Approaches for overcoming stationarity. In Abdollahian et al. [28], the authors explored an MDP model for breast and ovarian cancer intervention strategies. Though the authors did not set out to address stationarity in particular, by explicitly modeling age-dependency in the state space (for a total of 8,492 states) they effectively model changing disease and intervention dynamics over time. Maillart et al. [78] developed a partially observable Markov chain to study mammography screening policies where the authors enumerated all sample paths that changed over time. Ayer et al. [32] solved a finite-horizon POMDP to address decision support in mammography screening. The authors took the non-stationary policy approach and used age-specific transitions and rewards at different time steps during the solving process. Despite recent interest, the stationarity problem in a POMDP setting remains underexplored. While the non-stationary policy approach is the usual and recommended solution method [84], it does not guarantee optimal expected reward over the decision horizon.

In this work, stationarity was addressed in a lung cancer screening POMDP by modularizing the POMDP through the MODPOMDP framework. Specifically, each POMDP module is optimized individually for its applicable timeframe by tuning the reward function separately. Optimized rewards from each time point are then serialized into the combined non-stationary reward function to account for changes over time. The framework is implemented on a cohort of $5,089$ NLST participants and externally validated on a patient population at UCLA. In the NLST dataset, it is found that 1) results from non-stationary rewards outperformed results from each of the stationary rewards; and 2) results from non-stationary rewards are comparable to those of experts in true positives with a small increase in false positives. In the UCLA dataset, results from non-stationary rewards also outper-

formed results from each of the stationary rewards.

3.2 Method

3.2.1 Data Preprocessing

Table 3.8 and Table 3.9 contain the list of variables used in building the POMDP model as well as their distributions.

Tammemägi. For variables that were used in the Tammemägi $PLCO_{M2012}$ model [93], the preprocessing steps taken were the same as in $PLCO_{M2012}$. Patients with missing values required for calculating the $PLCO_{M2012}$ 6-year probability of lung cancer were removed from the datasets. Patients in the Tammemägi model can only belong to one combined race and ethnicity category, and as a result, multiracial patients were removed. Hispanics of any race were counted only as Hispanic and only non-Hispanics of a race were counted in their respective races. This method of grouping race and ethnicity appears in other papers as well, such as Tuttle et al. [94]. For family history and personal history of cancer, a patient was marked as "yes" if the patient was "yes" in any of its subcategories. Table 3.1 summarizes subcategories in family history and personal history of cancer and shows the difference between NLST and UCLA subcategories. While the personal history of cancer categories are broadly similar, each dataset collects cancer types not in the other set. Patients with prior history of lung cancer including diagnosis and procedures (e.g., surgery, chemotherapy, etc.) were also removed. Body mass index (BMI), smoking quit time, pack-years, and lastly the risk scores were calculated for remaining participants and patients. NLST criteria (e.g., age, smoking history, etc.) were reimposed on the NLST dataset as there are a few participants who fell outside of the inclusion criteria but were in the NLST database. NLST criteria were also applied to the UCLA dataset to get a comparable population for external validation.

Nodule-related features. Nodules were binned based on size. Bin 0: < 4 mm, Bins 1–8: 1 mm bins from 4–11 mm, Bin 9: $(11, 27]$ mm, and Bin 10: > 27 mm. Bin 1's size of 4 mm was informed by NLST's definition of suspicious findings (≥ 4 mm). Some nodules have inconsistencies in the recorded sizes where the larger diameter is less than the smaller diameter, and these were removed. Screenings with abnormalities but missing nodule sizes were also removed. Screenings without abnormalities were assumed to be in bin 0.

NLST screening-related. Participant data from the LDCT arm were used in building the model and must have an abnormality in at least one screening. Participants who did not develop cancer must have completed all three rounds of screenings. Because NLST does not keep track of individual nodules across screenings, an assumption was made that a nodule in the same lung lobe across different screening time points is the same nodule. Participants with only one identified finding in the same lobe across time were kept. For patients who were diagnosed with cancer, a location match between the diagnostic procedure identified primary tumor location and the screening tracked nodule was done. Patients whose lobes did not match were removed. Physician intervention was defined as having a diagnostic follow-up after the screening. To augment screening data, half-year data were interpolated, and nodule size was taken to be the mean between annual screenings. The total time range is thus $\Sigma = [1, 3]$ years and a horizon of 5 (5 time points). Figure 3.1 shows all possible screening trajectories in NLST without interpolated time points.

UCLA screening-related. As UCLA data is real-world observational data from a clinical lung cancer screening program, the screenings are not neatly separated into annual screenings as NLST was. As such, the patient's first available survey, which was necessary for calculating the Tammemägi 6-year lung cancer risk score, was used as the anchor time point. Some patients may receive a screening exam with a suspicious finding and return within a couple of weeks to months for a diagnostic imaging follow-up, at which point the patient fills out the survey. It therefore seemed reasonable to utilize screening and physician intervention data
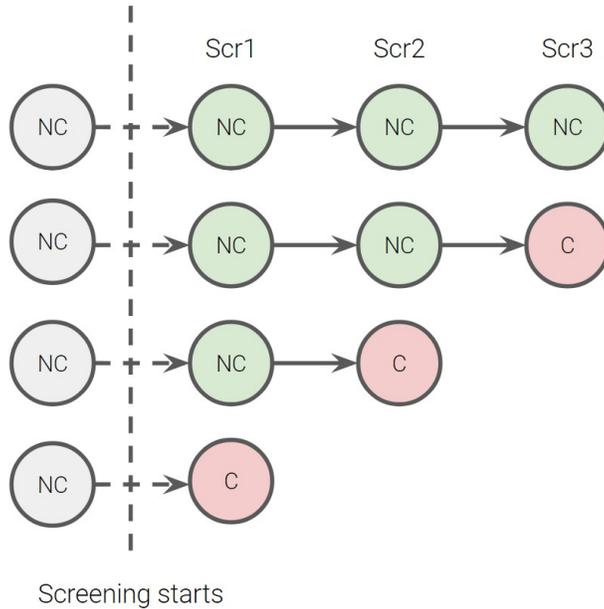
Figure 3.1: Possible screening trajectories of NLST participants. Participants start screening with no cancer and are removed from the trial if diagnosed with cancer. Non-cancer participants must have all three screenings. **NC:** no cancer; **C:** cancer.

100 days before and up to 6 years after the date of the survey. While most categorical values could be mapped 1-to-1 between UCLA and NLST, consistency could not be. Table 3.2 shows the final mapping between UCLA consistency and NLST predominant attenuation[1]. Another major difference is that UCLA data tracks individual nodules over time. And as such, each sample in the UCLA dataset is a nodule. Nodules that do not have any screenings within the defined timeframe or had interventions prior to any screenings were removed. Unlike NLST, which uses "positive" and "negative" for screening results, UCLA screening results are given using Lung-RADS [95,96]. Following the provided definitions, Lung-RADS Categories 1 and 2 are negative screenings and Categories 3 and above are positive screenings. As UCLA data is used for external validation, negative nodules can have any number of screenings up

---

[1]Finalized mapping after email correspondence with an expert thoracic radiologist at UCLA (Prosper). "Solid" can be both "soft tissue" and "fat", but as there are only three cases of "fat", "solid" is mapped to "soft tissue".

to five and always start at the first screening. The exception is when the negative nodule belongs to a patient with a positive nodule. In such a case, the negative nodule must have a screening on the same date as the positive nodule and the negative nodule screenings may shift so that the end dates of both nodules align. Figure 3.2 shows examples of nodule trajectory alignment. Ground truth is defined as having a biopsy confirmed cancer or having resection/ablation procedure done on the nodule. Physician intervention is defined in two ways. The first is when a procedure such as a core biopsy, biopsy, ablation, or resection was done. The second is when the patient undergoes diagnostic imaging or non-screening LDCT. However, if the patient's first screen was a diagnostic imaging or non-screening LDCT, then it was likely an incidental finding and *not* considered physician intervention until a procedure was performed.

The final counts per screen are shown in Table 3.3.
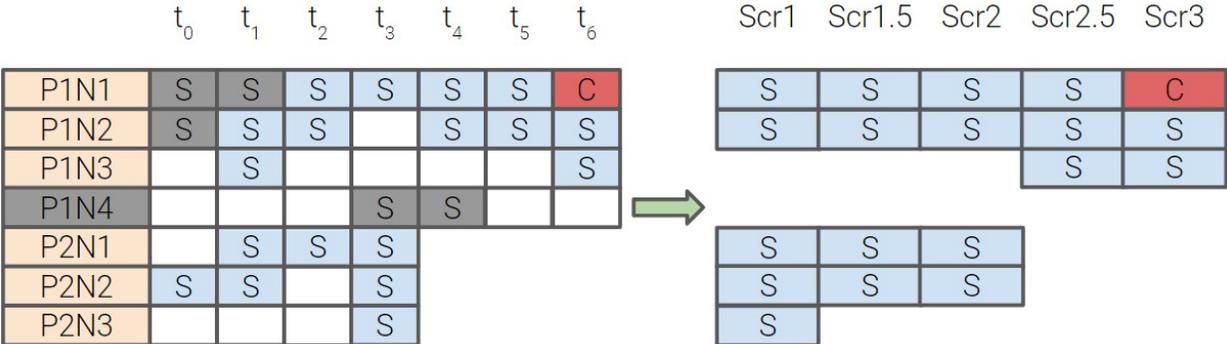
Figure 3.2: Examples of nodule trajectory alignment in UCLA data before and after reformatting. Patient 1 (P1) has 4 nodules (N1-4), one of which is a procedure confirmed cancer (C). Screenings (S) are shifted to align with C, but P1N4 is removed because it does not have a screening on the same date as C. P2 only has negative nodules and all start at Scr1.

Table 3.1: Comparison between NLST and UCLA in two categorical variables. Differences are shaded. Some differences are due to combining subcategories, for example NLST's Larynx, Nasal, Oral, and Pharynx were merged into UCLA's Head & Neck. Notably, UCLA collects more cancer types, but did not collect Stomach and Transitional Cell cancers that were collected in NLST.

	NLST				UCLA			
Family History of Lung Cancer	Father	Mother	Brother	Sister	Father	Mother	Sibling	Child
	Child							
Personal History of non-Lung Cancer	Bladder	Breast	Cervical	Colorectal	Bladder	Breast	Cervix	Colon
	Esophageal	Kidney	Pancreatic	Thyroid	Esophagus	Kidney	Pancreas	Thyroid
	Larynx	Nasal	Oral	Pharynx	Head & Neck	Gall Bladder	Liver	Lymphoma
	Stomach	Transitional Cell			Mesothelioma	Ovary	Prostate	Skin
					Uterus			

Table 3.2: Mapping between UCLA consistency and NLST predominant attenuation. **GGN:** ground-glass nodule; **PSN:** part-solid nodule.

UCLA	NLST
Solid	Soft tissue
Pure GGN	Ground-glass
PSN	Mixed
All others	Other

3.2.2 Defining POMDP Components

States (s) and actions (a). To model the lung cancer screening process, three states [97] and two actions are defined. The `No cancer (NC)` state is for participants who do not have suspicious findings (e.g., nodule size < 4 mm). The `Uncertain (U)` state is for participants who do have suspicious findings (e.g., nodule size $\geq$ 4 mm), but not diagnosed with cancer. The `Lung cancer (LC)` state is for participants who have a confirmed cancer (e.g., through biopsy). Participants who enter `LC` are removed from the screening process to emulate trial proceedings in NLST. At each time point, the agent either recommends a LDCT if belief for cancer is low or an intervention to confirm cancer if belief for cancer is high enough.

Table 3.3: NLST and UCLA counts of cancer and non-cancer cases per time. In NLST, each participant was preprocessed to have only one nodule and the count is both the number of participants and the number of nodules. For UCLA, each patient can have multiple nodules and therefore nodule counts are provided. For UCLA's cancer cases, each nodule belonged to only 1 patient.

	NLST Participants					UCLA Nodules				
	Cancer (%)		Non-Cancer (%)		Total	Cancer (%)		Non-Cancer (%)		Total
Scr1	128	(2.5)	4961	(97.5)	5089	33	(0.9)	3569	(99.1)	3602
Scr1.5	0		4961		4961	8	(0.7)	1104	(99.3)	1112
Scr2	70	(1.4)	4891	(98.6)	4961	1	(0.2)	402	(99.8)	403
Scr2.5	0		4891		4891	0	(0.0)	150	(100.0)	150
Scr3	88	(1.8)	4803	(98.2)	4891	4	(5.9)	64	(94.1)	68
Sum	278					46				

Observations (z). The observation space mainly consists of model-derived values representing the probability of observing cancer. Information (nodule size, attenuation, and margins) from the three screenings and two interpolated time points are used as input to a DBN to infer the probability of cancer at each time point. These probabilities are then discretized into a value between 0 and 1 representing their closest percent probability. The DBN was trained using five bootstrap iterations of 5-fold cross validation using NLST. Stratified sampling was used to ensure each sample was used as training four times and once as testing per bootstrap. The mean probabilities of the testing set over the five bootstraps were used as the final probabilities for discretization. To get observations for UCLA, the NLST testing fold was replaced with UCLA data, with means and discretization operations occurring thereafter. Together with a cancer after intervention observation and a no cancer regardless of intervention observation, the observation space consists of 102 possibilities.

Transitions and observation probabilities. The transition and observation probabilities were calculated using a DBN. Figure 3.3(c) shows the transition model. Each node
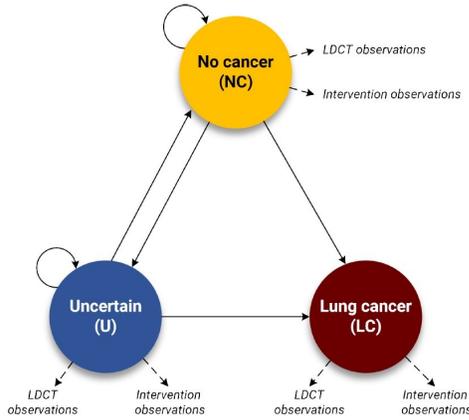
(a) Lung cancer screening POMDP.
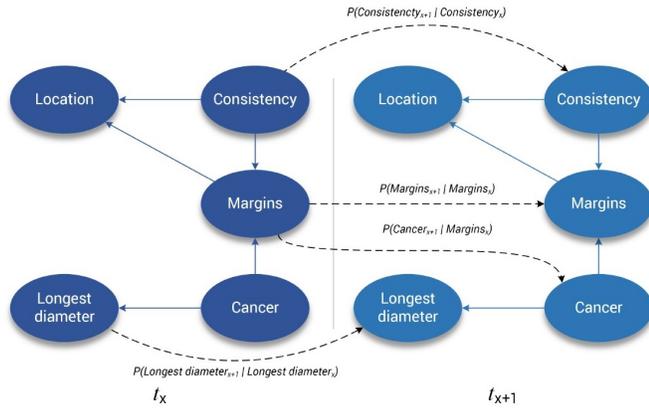
(b) DBN-based observation model.
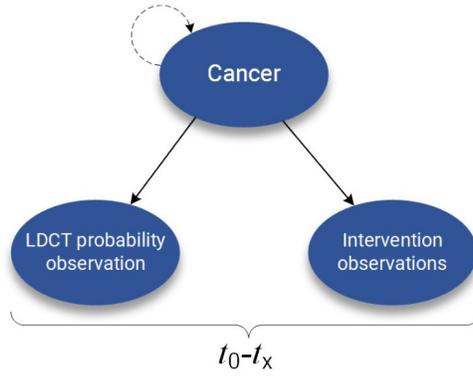
(c) Transition model.

Figure 3.3: Various components of lung cancer screening POMDP; adapted from [2].

represents a conditional probability table (CPT). The `Cancer` node has three categories for the three defined states. `LDCT` has 100 categories for the discretized observation probabilities. `Intervention` has two categories, a cancer after intervention and a no cancer regardless of intervention. In instances where the conditional probabilities contain 0, the 0 was replaced with 0.0001.

Rewards. Calculation of the reward function $R(s, a)$ via adaptive step-size maximum entropy (MaxEnt) inverse reinforcement learning (IRL) was previously described in Chapter

2.2.2. While normally both the state MDP and action MDP are necessary, it is apparent that as there are only two defined actions in the POMDP, $R(a)$ would default to $[-1, 1]$. As such, MaxEnt IRL was only used for the state MDP model prior to employing the multiplicative model. When these rewards were used to select actions, the recall values shown in Table 3.4 suggest additional optimization of the rewards are necessary. Reward tuning is accomplished by progressively increasing the relative rewards for transitioning into the LC state. The rewards that produced the highest true positives were then selected for rewards that produced the lowest false positives. A stop condition is set such that the false positive rate does not exceed 90%.

Table 3.4: Results of using MaxEnt IRL $R(s, a)$ to select actions on the NLST dataset. Model performance (mean counts $\pm$ 95% CI) is calculated over 100 instances of bootstrap on the testing set.

NLST		TN	FP	FN	TP	Pre.	Rec.
	Scr1	857.46 ± 2.01	136.54 ± 2.01	$3.40 \pm .30$	$22.60 \pm .30$	$0.1425 \pm .0024$	$0.8692 \pm .0117$
	Scr2	706.96 ± 4.39	273.04 ± 4.39	$1.99 \pm .26$	$12.01 \pm .26$	$0.0423 \pm .0009$	$0.8579 \pm .0188$
$R(s, a)$	Scr3	629.78 ± 4.95	332.22 ± 4.95	$3.04 \pm .38$	$14.96 \pm .38$	$0.0431 \pm .0009$	$0.8311 \pm .0208$
	Sum	2194.20	741.80	8.43	49.57		

Initial beliefs. Initial beliefs b_0 are the belief distribution over POMDP states at t_0. As the Tammemägi risk scores are calculated for participants at baseline, they are a natural candidate for initial beliefs. The initial belief for state LC is defined as $b_{0,\text{LC}} = 2 \times \text{PLCO}_{\text{M2012}}$. $b_{0,\text{U}}$ is assumed to be 0 while $b_{0,\text{NC}} = 1 - b_{0,\text{LC}}$.

3.2.3 The modPOMDP Framework

Figure 3.4 shows the experimental setup of modPOMDP. The initial data are split such that 80% were used in training the model and tuning the rewards. The remaining 20% were used for testing. This was done over 100 instances of bootstrap. Under the modPOMDP

framework, the reward tuning step is done per screen. This implies $I = 3$ POMDP submodels where the relevant year ranges are $\sigma_1 = [1, 2)$, $\sigma_2 = [2, 3)$, and $\sigma_3 = [3, 3]$. When externally validating on the UCLA data, the 20% testing set is replaced with the UCLA data. Thus the rewards from $POMDP_1$ would be applied to screens 1 and 1.5, $POMDP_2$ to screens 2 and 2.5, and $POMDP_3$ to screen 3.
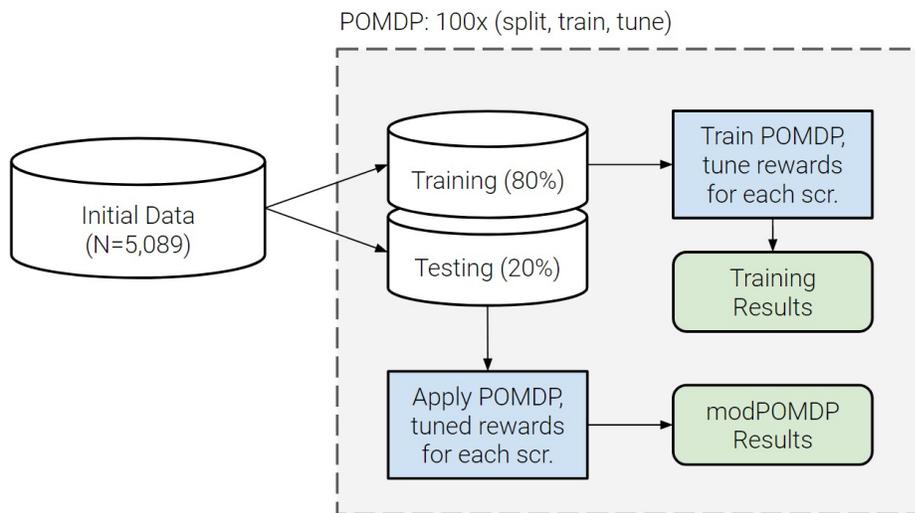
Figure 3.4: Experimental setup of MODPOMDP. A POMDP model is trained per bootstrap iteration and rewards are tuned for each screen. For external validation, the `Testing (20%)` box is replaced with UCLA data.

3.2.4 The Brock Model

The Brock model [98] was implemented on the NLST dataset to compare to MODPOMDP results. The Brock model consists of four models: 1a) parsimonious without spiculation, 1b) full without spiculation, 2a) parsimonious with spiculation, and 2b) full with spiculation. The full model with spiculation (Brock2b) was implemented in this work. The Brock risk scores were calculated using data (including nodule features) from the time of screening (i.e., screen 2 Brock2b scores were calculated using data from screen 2). As age is variable in the model and NLST screenings were roughly one year apart, age was incremented by 1 and 2 for

screens 2 and 3, respectively. Similar to the MODPOMDP setup, there were 100 bootstrap instances where 80% of the data were used for threshold tuning for recall and 20% were used for testing. The variables and distributions for each screening are shown in Table 3.10.

3.3   Results

NLST.   Table 3.5 summarizes MODPOMDP (mp) results for NLST. Screen 1 Optimized (Scr1-O) results are derived from applying Scr1-O rewards to all three screens. Likewise, Screen 2 Optimized (Scr2-O) and Screen 3 Optimized (Scr3-O) results are derived by applying their respective optimized rewards to all three screens. Scr1-O and Scr2-O performed similarly in terms of TPs (Scr1-O: 57.40 v. Scr2-O: 57.76) across three screens. However, Scr2-O had an overall higher FP count (Scr1-O: 2206.63 v. Scr2-O: 2489.32). While Scr3-O had overall low TPs and FPs, it performed well for the screen that it was optimized for (TP: 17.73 and FP: 411.88). When these screen-optimized results are collected into MOD-POMDP, the results show similar TPs to physicians (p) as well as to Scr1-O and Scr2-O. On the otherhand, MODPOMDP had lower FPs compared to Scr1-O and Scr2-O, and only slightly higher FPs relative to the physicians (mp: 1975.95 vs. p: 1926.45).

UCLA.   Table 3.6 summarizes MODPOMDP results for UCLA. In terms of TPs, both MODPOMDP and physicians did well, achieving 45 out of 46 TPs over 5 time points. For FPs, physician performance immediately stands out, where screen 1 FP rate is only 0.4%. This is far lower than the baseline LDCT FP rates (7.9% to 49.3% among 27 publications) reported by Jonas et al. [6]. By contrast, FP rates of physicians for screens 1.5 to 3 range from 22.3% to 48.4%. For Scr1-O, the FP rates of screens 1 and 1.5 are 39.6% and 47.6% respectively, whereas NLST Scr1-O is 62.1%. In other words, screen 1 optimized rewards generalized to UCLA better. For Scr2-O, the FP rates of screens 2 and 2.5 are 96.6% and 97.4% respectively, comparable to NLST Scr2-O's 96.6%. For Scr3-O, the FP rate of screen

Table 3.5: Results for MODPOMDP on the NLST dataset. Model performance (mean counts $\pm$ 95% CI) is calculated over 100 instances of bootstrap on the testing set. **Highlight:** results collected for MODPOMDP.

NLST		TN	FP	FN	TP	Pre.	Rec.
	Scr1	480.43 ± 2.89	513.57 ± 2.89	$0.00 \pm .00$	$26.00 \pm .00$	$0.0482 \pm .0003$	$1.0000 \pm .0000$
	Scr2	322.14 ± 2.83	657.86 ± 2.83	$0.74 \pm .14$	$13.26 \pm .14$	$0.0198 \pm .0002$	$0.9472 \pm .0102$
Physician	Scr3	206.98 ± 2.39	755.02 ± 2.39	$0.00 \pm .00$	$18.00 \pm .00$	$0.0233 \pm .0001$	$1.0000 \pm .0000$
	Sum	1009.55	1926.45	0.74	57.26		
	Scr1	376.87 ± 14.23	617.13 ± 14.23	$0.26 \pm .12$	$25.74 \pm .12$	$0.0406 \pm .0010$	$0.9900 \pm .0047$
Screen 1	Scr2	209.58 ± 24.38	770.42 ± 24.38	$0.34 \pm .11$	$13.66 \pm .11$	$0.0178 \pm .0006$	$0.9757 \pm .0081$
Optimized	Scr3	142.92 ± 18.62	819.08 ± 18.62	$0.00 \pm .00$	$18.00 \pm .00$	$0.0218 \pm .0005$	$1.0000 \pm .0000$
	Sum	729.37	2206.63	0.60	57.40		
	Scr1	386.53 ± 23.26	607.47 ± 23.26	$0.10 \pm .06$	$25.90 \pm .06$	$0.0430 \pm .0024$	$0.9962 \pm .0023$
Screen 2	Scr2	33.06 ± 26.06	946.94 ± 26.06	$0.14 \pm .12$	$13.86 \pm .12$	$0.0148 \pm .0006$	$0.9900 \pm .0083$
Optimized	Scr3	27.09 ± 21.42	934.91 ± 21.42	$0.00 \pm .00$	$18.00 \pm .00$	$0.0193 \pm .0008$	$1.0000 \pm .0000$
	Sum	446.68	2489.32	0.24	57.76		
	Scr1	817.31 ± 4.86	176.69 ± 4.86	$2.15 \pm .27$	$23.85 \pm .27$	$0.1206 \pm .0029$	$0.9173 \pm .0104$
Screen 3	Scr2	664.79 ± 3.70	315.21 ± 3.70	$1.19 \pm .21$	$12.81 \pm .21$	$0.0392 \pm .0007$	$0.9150 \pm .0148$
Optimized	Scr3	550.12 ± 3.51	411.88 ± 3.51	$0.27 \pm .09$	$17.73 \pm .09$	$0.0413 \pm .0002$	$0.9850 \pm .0049$
	Sum	2032.22	903.78	3.61	54.39		
	Scr1	376.87 ± 14.23	617.13 ± 14.23	$0.26 \pm .12$	$25.74 \pm .12$	$0.0406 \pm .0010$	$0.9900 \pm .0047$
	Scr2	33.06 ± 26.06	946.94 ± 26.06	$0.14 \pm .12$	$13.86 \pm .12$	$0.0148 \pm .0006$	$0.9900 \pm .0083$
modPOMDP	Scr3	550.12 ± 3.51	411.88 ± 3.51	$0.27 \pm .09$	$17.73 \pm .09$	$0.0413 \pm .0002$	$0.9850 \pm .0049$
	Sum	960.05	1975.95	0.67	57.33		

3 is 52.1% vs. NLST Scr3-O 42.8%. As Scr2-O and Scr3-O performed within expectations, it appears that UCLA Scr1-O's better than expected results lead to the comparatively worse performance of MODPOMDP vs. Scr1-O (mp: 2505.5 vs. Scr1-O: 2348.97) in terms of FPs.

Brock2b. Table 3.7 shows Brock2b results for NLST. When compared to MODPOMDP, Brock2b (B2b) had lower TPs (B2b: 55.52 vs. mp: 57.33) and lower FPs (B2b: 1151.87 vs. mp: 1975.95) across three screenings. For screen 1 only, both Scr1-O and B2b performed well in TPs (both: 25.74), while B2b had much lower FPs (B2b: 361.87 vs. Scr1-O: 617.13).

Table 3.6: Results for MODPOMDP on the UCLA dataset. Model performance (mean counts $\pm$ 95% CI) is calculated over 100 instances of bootstrap on the testing set. **Highlight:** results collected for MODPOMDP.

UCLA		TN	FP	FN	TP	Pre.	Rec.
	Scr1	$3555.00 \pm .00$	$14.00 \pm .00$	$0.00 \pm .00$	$33.00 \pm .00$	$0.7021 \pm .0000$	$1.0000 \pm .0000$
	Scr1.5	$858.00 \pm .00$	$246.00 \pm .00$	$0.00 \pm .00$	$8.00 \pm .00$	$0.0315 \pm .0000$	$1.0000 \pm .0000$
	Scr2	$284.00 \pm .00$	$118.00 \pm .00$	$0.00 \pm .00$	$1.00 \pm .00$	$0.0084 \pm .0000$	$1.0000 \pm .0000$
Physician	Scr2.5	$91.00 \pm .00$	$59.00 \pm .00$	$0.00 \pm .00$	$0.00 \pm .00$	$0.0000 \pm .0000$	n/a
	Scr3	$33.00 \pm .00$	$31.00 \pm .00$	$1.00 \pm .00$	$3.00 \pm .00$	$0.0882 \pm .0000$	$0.7500 \pm .0000$
	Sum	4821.00	468.00	1.00	45.00		
	Scr1	2157.50 ± 53.95	1411.50 ± 53.95	$0.00 \pm .00$	$33.00 \pm .00$	$0.0238 \pm .0011$	$1.0000 \pm .0000$
	Scr1.5	578.35 ± 15.09	525.65 ± 15.09	$1.00 \pm .00$	$7.00 \pm .00$	$0.0135 \pm .0004$	$0.8750 \pm .0000$
Screen 1	Scr2	146.09 ± 14.88	255.91 ± 14.88	$0.00 \pm .00$	$1.00 \pm .00$	$0.0042 \pm .0002$	$1.0000 \pm .0000$
Optimized	Scr2.5	41.95 ± 4.33	108.05 ± 4.33	$0.00 \pm .00$	$0.00 \pm .00$	$0.0000 \pm .0000$	n/a
	Scr3	16.14 ± 1.63	47.86 ± 1.63	$0.00 \pm .00$	$4.00 \pm .00$	$0.0787 \pm .0020$	$1.0000 \pm .0000$
	Sum	2940.03	2348.97	1.00	45.00		
	Scr1	2155.50 ± 65.05	1413.50 ± 65.05	$0.00 \pm .00$	$33.00 \pm .00$	$0.0244 \pm .0014$	$1.0000 \pm .0000$
	Scr1.5	219.22 ± 63.88	884.78 ± 63.88	$0.38 \pm .12$	$7.62 \pm .12$	$0.0101 \pm .0009$	$0.9525 \pm .0149$
Screen 2	Scr2	13.51 ± 10.73	388.49 ± 10.73	$0.00 \pm .00$	$1.00 \pm .00$	$0.0027 \pm .0002$	$1.0000 \pm .0000$
Optimized	Scr2.5	3.95 ± 3.15	146.05 ± 3.15	$0.00 \pm .00$	$0.00 \pm .00$	$0.0000 \pm .0000$	n/a
	Scr3	1.34 ± 1.06	62.66 ± 1.06	$0.00 \pm .00$	$4.00 \pm .00$	$0.0605 \pm .0020$	$1.0000 \pm .0000$
	Sum	2393.52	2895.48	0.38	45.62		
	Scr1	3063.64 ± 24.44	505.36 ± 24.44	$2.61 \pm .89$	$30.39 \pm .89$	$0.0585 \pm .0016$	$0.9209 \pm .0268$
	Scr1.5	864.20 ± 9.13	239.80 ± 9.13	$2.00 \pm .12$	$6.00 \pm .00$	$0.0255 \pm .0012$	$0.7500 \pm .0000$
Screen 3	Scr2	280.85 ± 3.57	121.15 ± 3.57	$0.00 \pm .00$	$1.00 \pm .00$	$0.0084 \pm .0003$	$1.0000 \pm .0000$
Optimized	Scr2.5	86.20 ± 1.73	63.80 ± 1.73	$0.00 \pm .00$	$0.00 \pm .00$	$0.0000 \pm .0000$	n/a
	Scr3	30.64 ± 0.89	33.36 ± 0.89	$0.00 \pm .00$	$4.00 \pm .00$	$0.1089 \pm .0030$	$1.0000 \pm .0000$
	Sum	4325.53	963.47	4.61	41.39		
	Scr1	2157.50 ± 53.95	1411.50 ± 53.95	$0.00 \pm .00$	$33.00 \pm .00$	$0.0238 \pm .0011$	$1.0000 \pm .0000$
	Scr1.5	578.35 ± 15.09	525.65 ± 15.09	$1.00 \pm .00$	$7.00 \pm .00$	$0.0135 \pm .0004$	$0.8750 \pm .0000$
	Scr2	13.51 ± 10.73	388.49 ± 10.73	$0.00 \pm .00$	$1.00 \pm .00$	$0.0027 \pm .0002$	$1.0000 \pm .0000$
modPOMDP	Scr2.5	3.95 ± 3.15	146.05 ± 3.15	$0.00 \pm .00$	$0.00 \pm .00$	$0.0000 \pm .0000$	n/a
	Scr3	30.64 ± 0.89	33.36 ± 0.89	$0.00 \pm .00$	$4.00 \pm .00$	$0.1089 \pm .0030$	$1.0000 \pm .0000$
	Sum	2783.95	2505.05	1.00	45.00		

Table 3.7: Results for Brock2b on the NLST dataset. Model performance (mean counts $\pm$ 95% CI) is calculated over 100 instances of bootstrap on the testing set. **B2b:** Brock2b.

		TN	FP	FN	TP	Pre.	Rec.	ROC-AUC	PR-AUC
	Scr1	632.13 ± 3.23	361.87 ± 3.23	$0.26 \pm .10$	$25.74 \pm .10$	$0.0665 \pm .0005$	$0.9900 \pm .0039$	$0.8130 \pm .0016$	$0.5284 \pm .0018$
	Scr2	461.82 ± 20.42	518.18 ± 20.42	$1.36 \pm .22$	$12.64 \pm .22$	$0.0249 \pm .0012$	$0.9029 \pm .0155$	$0.6870 \pm .0089$	$0.4646 \pm .0075$
B2b	Scr3	690.18 ± 4.22	271.82 ± 4.22	$0.86 \pm .16$	$17.14 \pm .16$	$0.0596 \pm .0009$	$0.9522 \pm .0089$	$0.8348 \pm .0040$	$0.5063 \pm .0044$
	Sum	1784.13	1151.87	2.48	55.52				

3.4 Discussion

This chapter described the design and implementation of a modularized POMDP framework for lung cancer screening. The key realization comes from the observation that due to the Markovian nature of belief states in POMDPs, each time point can be evaluated separately. The temporal dynamics of the system is captured through reward function optimization and represents a simple and effective way of addressing the stationarity assumption. In this work, while the conditional probabilities of POMDP subcomponents are jointly learned across time points, the reward functions are optimized individually for each time point.

Given the nascent nature of lung cancer screening and the setup of the NLST, a modularized POMDP framework was tested given the observation that most participants/patients when first starting a cancer screening program have a higher potential for (an initial) positive finding vs. those who have been continuously monitored over time – suggesting that the observation, transition, and reward functions are non-stationary and thus should be modeled differently over time.

In NLST, the results show that MODPOMDP performs better than each of the screen-optimized results – that is, maintaining comparable true positives while having lower false positives. When compared to experts, MODPOMDP performs at the same level for true positives with only a small increase in false positives (49.5 across three screenings). The Brock2b model on the hand other, was not able to capture true positives to the same extent as MOD-POMDP overall. However, Brock2b performed very well in the first screen, capturing the

same number of true positives while having much fewer false positives than MODPOMDP. This performance did not extend to screens 2 and 3, indicating that the temporal dynamics and dependencies learned in the POMDP cannot be replicated by a model that predicts each time point completely independently. These results demonstrate the applicability of the MODPOMDP framework.

When applied to the UCLA external validation set however, MODPOMDP received higher false positives than screen 1 optimized rewards alone, but maintained the same number of true positives. As previously pointed out in Section 3.3, UCLA's optimized screens 1 and 1.5 results are the outliers in having a much lower false positive rate than NLST's screen 1 while the false positive rates are comparable for other time points. One possibility involves the fact that NLST is a RCT with multiple eligibility criteria, one of which is that the participant cannot have a CT within 18 months of recruitment. Coupled with the high-risk for lung cancer smoking criteria, NLST's screen 1 would naturally be expected to have higher than normal positive cases. UCLA data on the other hand is electronic health record (EHR) data and many patients had CTs within 18 months of the baseline survey. Many abnormalities would have already been found and one would not expect the front-loaded distribution of cancers that appears in NLST. Indeed, at baseline, 40.1% of the NLST participants had no or < 4 mm abnormalities whereas 79% of UCLA nodules had no or < 4 mm abnormalities. To an agent trained and tuned on the NLST data, many of the UCLA nodules would appear as negative and no intervention recommended. While screen 1 optimized rewards performed well for earlier screens, it was not able to capture the temporal dynamics of screen 3. That is, screen 3 optimized rewards showed better results than screen 1 optimized rewards showed in screen 3 (NLST Scr1-O FP: 819.08 vs. Scr3-O FP: 411.88; UCLA Scr1-O FP: 47.86 vs. Scr3-O FP: 33.36).

One assumption made about the UCLA data is the characterization of physician intervention. Intervention is defined as having a procedure (e.g., biopsy), diagnostic imaging (DXI), or non-screening LDCT (NSL). If the patient's first screen is a DXI or NSL however,

then it and subsequent DXI/NSL were assumed to be incidental findings and also treated as non-intervention until a procedure is done. As most non-cancer patients and nodules only have one or two screens, these trajectories are clustered at screens 1 and 1.5. At screen 1, DXI and NSL are 344 out of 3602 screenings. At screen 2, DXI and NSL are 320 out of 1112 screenings. How physician intervention is defined clearly has an effect on UCLA's false positive rates at screens 1 and 1.5.

Another observation is that regardless of the dataset, screen 2 appears to be a difficult time point. The precision at screen 2 is the lowest for the physicians, MODPOMDP, and Brock2b. In NLST, screen 2 is the only time point where the physicians have false negatives. When optimizing rewards for screen 2, true positives were not improved much despite the very aggressive rewards, but false positives did increase substantially relative to applying screen 1 optimized rewards on screen 2. In Brock2b, true positives were the lowest among the three groups, despite tuning the decision threshold for recall. This suggests that the Brock2b model cannot differentiate very well between positive and negative nodules in some circumstances, a finding previously reported in Winter et al. [99]. The authors noted the differentiation difficulty even when Brock2b achieved high area under the curve (AUC). It is likely then, that there are cancer patients whose screening results look very similar to non-cancer patients. It is also noted that in the NLST, applying screen 1 optimized rewards on screen 2 (instead of screen 2 optimized rewards on screen 2) would decrease overall true positives by 0.2 while decreasing false positives by 176.52. This largely maintains the same true positives as physicians but lowers false positives significantly (mp: 1799.43 vs. p: 1926.45). Markedly, if applied to the UCLA cohort, this same trend is observed. That is, true positives did not change, but false positives decreased by 170.58 to 2334.47. This observation suggests that the MODPOMDP setup of $I = 3$ submodels where $\sigma_1 = [1, 2)$, $\sigma_2 = [2, 3)$, and $\sigma_3 = [3, 3]$ may not be optimal, and instead can be $I = 2$ where $\sigma_1 = [1, 3)$ and $\sigma_2 = [3, 3]$. It appears that the dynamics of screen 1 can apply to screen 2 well and how MODPOMDP setup is defined can influence outcome.

One limitation of this study is in the preprocessing of UCLA EHR data. When building the NLST POMDP model, the time interval was set to half-year by interpolating values between two screenings. In the UCLA data, the time interval between screenings were used as they occur, meaning the time between two screenings is unlikely to be six months. The average number of days for a nodule between screens 1 and 1.5 is 402.6 ± 256.9 (mean $\pm$ std), 367.6 ± 203.9 between screens 1.5 and 2, 316.4 ± 185.0 between screens 2 and 2.5, and 274.8 ± 242.6 between screens 2.5 and 3. The high standard deviation shows substantial differences in individual nodule trajectories. For example, some patients may be referred for 3-month follow-ups while others return to screening after multiple years. Modeling trajectories to allow a more fitting time interval between screenings may improve POMDP performance.

Table 3.8: Baseline characteristics of NLST participants and UCLA patients. UCLA Scr2.5 has no cancer nodules and therefore omitted. For numeric variables, the values are mean $\pm$ standard deviation. For categorical or binary variables, the values are count (%). **C:** cancer; **NC:** non-cancer; **LC:** lung cancer; **NH/PI:** Native Hawaiian or other Pacific Islander; **AI/AN**: American Indian or Alaskan Native.

Characteristics	NLST (Participant count)				UCLA (Patient count)				
	Scr1 C (128)	Scr2 C (70)	Scr3 C (88)	NC (4803)	Scr1 C (33)	Scr1.5 C (8)	Scr2 C (1)	Scr3 C (4)	NC (726)
PLCO$_{M2012}$ risk score	$0.049 \pm .039$	$0.051 \pm .037$	$0.044 \pm .034$	$0.033 \pm .028$	$0.06 \pm .055$	$0.063 \pm .041$	$0.07 \pm .0$	$0.109 \pm .086$	$0.041 \pm .037$
≥ 0.0151	114 (89.1)	63 (90)	77 (87.5)	3640 (75.8)	29 (87.9)	7 (87.5)	1 (100)	4 (100)	590 (81.3)
Age	63.9 ± 5.3	62.5 ± 4.5	62.8 ± 5.2	61.5 ± 5	66.8 ± 4.9	66.7 ± 5.5	$68 \pm .0$	67.3 ± 7	64.9 ± 5.4
$[55 - 60)$	35 (27.3)	19 (27.1)	31 (35.2)	2011 (41.9)	4 (12.1)	1 (12.5)	0 (0)	1 (25)	160 (22)
$[60 - 65)$	34 (26.6)	29 (41.4)	23 (26.1)	1490 (31)	6 (18.2)	1 (12.5)	0 (0)	0 (0)	199 (27.4)
$[65 - 70)$	38 (29.7)	17 (24.3)	22 (25)	868 (18.1)	13 (39.4)	4 (50)	1 (100)	1 (25)	212 (29.2)
$[70 - 75)$	21 (16.4)	5 (7.1)	12 (13.6)	434 (9)	10 (30.3)	2 (25)	0 (0)	2 (50)	155 (21.3)
BMI	26.6 ± 4.1	26.8 ± 6.6	26.3 ± 4.1	27.7 ± 4.9	26.9 ± 5.1	26.2 ± 4.5	$37.8 \pm .0$	23.6 ± 3.3	27.1 ± 5.4
< 18.5	1 (0.8)	2 (2.9)	2 (2.3)	43 (0.9)	1 (3)	0 (0)	0 (0)	0 (0)	20 (2.8)
$[18.5\text{-}25)$	46 (35.9)	25 (35.7)	27 (30.7)	1441 (30)	13 (39.4)	5 (62.5)	0 (0)	3 (75)	244 (33.6)
$[25\text{-}30)$	59 (46.1)	33 (47.1)	45 (51.1)	2001 (41.7)	10 (30.3)	2 (25)	0 (0)	1 (25)	286 (39.4)
≥ 30	22 (17.2)	10 (14.3)	14 (15.9)	1318 (27.4)	9 (27.3)	1 (12.5)	1 (100)	0 (0)	176 (24.2)
Gender									
Male	73 (57)	43 (61.4)	52 (59.1)	2806 (58.4)	24 (72.7)	3 (37.5)	0 (0)	1 (25)	444 (61.2)
Female	55 (43)	27 (38.6)	36 (40.9)	1997 (41.6)	9 (27.3)	5 (62.5)	1 (100)	3 (75)	282 (38.8)
Education									
$<$ High school	16 (12.5)	6 (8.6)	8 (9.1)	277 (5.8)	0 (0)	0 (0)	0 (0)	0 (0)	23 (3.2)
High school	31 (24.2)	17 (24.3)	19 (21.6)	1192 (24.8)	9 (27.3)	1 (12.5)	1 (100)	0 (0)	116 (16)
Post high school	10 (7.8)	10 (14.3)	14 (15.9)	675 (14.1)	1 (3)	0 (0)	0 (0)	1 (25)	51 (7)
Some college	30 (23.4)	20 (28.6)	23 (26.1)	1143 (23.8)	7 (21.2)	3 (37.5)	0 (0)	1 (25)	186 (25.6)
College	21 (16.4)	7 (10)	14 (15.9)	825 (17.2)	12 (36.4)	3 (37.5)	0 (0)	0 (0)	207 (28.5)
Graduate	20 (15.6)	10 (14.3)	10 (11.4)	691 (14.4)	4 (12.1)	1 (12.5)	0 (0)	2 (50)	143 (19.7)
Race/ethnicity									
White	115 (89.8)	62 (88.6)	81 (92)	4405 (91.7)	27 (81.8)	7 (87.5)	1 (100)	3 (75)	616 (84.8)
Black	4 (3.1)	7 (10)	3 (3.4)	198 (4.1)	0 (0)	1 (12.5)	0 (0)	1 (12.5)	32 (4.4)
Hispanic	5 (3.9)	0 (0)	1 (1.1)	81 (1.7)	2 (6.1)	0 (0)	0 (0)	0 (0)	32 (4.4)
Asian	4 (3.1)	1 (1.4)	2 (2.3)	97 (2)	4 (12.1)	0 (0)	0 (0)	0 (0)	40 (5.5)
NH/PI	0 (0)	0 (0)	0 (0)	11 (0.2)	0 (0)	0 (0)	0 (0)	0 (0)	6 (0.8)
AI/AN	0 (0)	0 (0)	1 (1.1)	11 (0.2)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Hist. of COPD (% Yes)	13 (10.2)	8 (11.4)	7 (8)	255 (5.3)	12 (36.4)	3 (37.5)	1 (100)	2 (50)	191 (26.3)
Prior non-LC (% Yes)	10 (7.8)	4 (5.7)	3 (3.4)	206 (4.3)	12 (36.4)	4 (50)	0 (0)	3 (75)	134 (18.5)
Fam. hist. of LC (% Yes)	34 (26.6)	23 (32.9)	22 (25)	1067 (22.2)	10 (30.3)	2 (25)	1 (100)	1 (25)	155 (21.3)
Cigarettes per day	30.1 ± 13.5	30.6 ± 10.7	29.8 ± 13	28.7 ± 11.3	23 ± 6.7	20.1 ± 4.9	$20 \pm .0$	25 ± 8.7	22.4 ± 8.2
Years smoked	43.2 ± 7.6	43.2 ± 6.3	43 ± 7	40 ± 7.2	43.1 ± 6.1	45.5 ± 7.3	$47 \pm .0$	47.5 ± 12.5	41.6 ± 8.1
Current smokers	64 (50)	44 (62.9)	47 (53.4)	2397 (49.9)	13 (39.4)	3 (37.5)	0 (0)	1 (25)	288 (39.7)
Smoking quit time	6.6 ± 4.3	6.2 ± 4.4	7.2 ± 4.8	7.3 ± 4.6	5 ± 4.3	5 ± 1.9	$3 \pm .0$	7.4 ± 4.7	6.4 ± 4.5

Table 3.9: Baseline characteristics of NLST and UCLA nodules. Baseline nodule sizes were computed from nodules with diameter > 0 mm. UCLA Scr2.5 has no cancer nodules and therefore omitted. For numeric variables, the values are mean $\pm$ standard deviation. For categorical or binary variables, the values are count (%). **C:** cancer; **NC:** non-cancer.

	NLST (Participant count)				**UCLA (Nodule count)**				
Characteristics	Scr1 C (128)	Scr2 C (70)	Scr3 C (88)	NC (4803)	Scr1 C (33)	Scr1.5 C (8)	Scr2 C (1)	Scr3 C (4)	NC (3582)
Baseline nodule size (mm)	19.8 ± 11.7	14.3 ± 18.9	10.7 ± 7.7	7.3 ± 5.9	25.7 ± 14.6	22.1 ± 17.7	$8 \pm .0$	9.3 ± 1.7	6.7 ± 5.6
< 4	0 (0)	37 (52.9)	66 (75)	1936 (40.3)	0 (0)	1 (12.5)	0 (0)	1 (25)	2865 (80)
4	1 (0.8)	3 (4.3)	1 (1.1)	555 (11.6)	0 (0)	0 (0)	0 (0)	0 (0)	182 (5.1)
(4, 5]	2 (1.6)	2 (2.9)	0 (0)	750 (15.6)	0 (0)	1 (12.5)	0 (0)	0 (0)	128 (3.6)
(5, 6]	4 (3.1)	0 (0)	3 (3.4)	464 (9.7)	0 (0)	0 (0)	0 (0)	0 (0)	122 (3.4)
(6, 7]	3 (2.3)	5 (7.1)	4 (4.5)	296 (6.2)	0 (0)	0 (0)	0 (0)	1 (25)	59 (1.6)
(7, 8]	5 (3.9)	5 (7.1)	4 (4.5)	184 (3.8)	1 (3)	0 (0)	1 (100)	0 (0)	40 (1.1)
(8, 9]	3 (2.3)	3 (4.3)	1 (1.1)	139 (2.9)	0 (0)	0 (0)	0 (0)	0 (0)	38 (1.1)
(9, 10]	6 (4.7)	3 (4.3)	1 (1.1)	113 (2.4)	1 (3)	1 (12.5)	0 (0)	1 (25)	29 (0.8)
(10, 11]	8 (6.2)	0 (0)	1 (1.1)	62 (1.3)	3 (9.1)	1 (12.5)	0 (0)	1 (25)	29 (0.8)
$(11 - 27]$	68 (53.1)	10 (14.3)	6 (6.8)	275 (5.7)	15 (45.6)	2 (25)	0 (0)	0 (0)	78 (2.2)
≥ 27	28 (21.9)	2 (2.9)	1 (1.1)	29 (0.6)	13 (39.4)	2 (25)	0 (0)	0 (0)	12 (0.3)
Baseline attenuation									
Soft tissue	100 (78.1)	21 (30)	14 (15.9)	2906 (43.6)	9 (27.3)	2 (25)	0 (0)	1 (25)	1330 (37.1)
Ground-glass	11 (8.6)	5 (7.1)	7 (8)	410 (8.5)	2 (6.1)	2 (25)	0 (0)	0 (0)	303 (8.5)
Mixed	13 (10.2)	5 (7.1)	0 (0)	148 (3.1)	9 (27.3)	3 (37.5)	0 (0)	2 (50)	63 (1.8)
Other	1 (0.8)	0 (0)	1 (1.1)	20 (0.4)	5 (15.2)	0 (0)	0 (0)	0 (0)	18 (0.5)
Missing	3 (2.3)	39 (55.7)	66 (75)	2129 (44.3)	8 (24.2)	1 (12.5)	1 (100)	1 (25)	1868 (52.1)
Baseline margins									
Spiculated (Stellate)	80 (62.5)	14 (20)	4 (4.5)	270 (5.6)	15 (45.5)	1 (12.5)	0 (0)	1 (25)	20 (0.6)
Smooth	14 (10.9)	7 (10)	7 (8)	1902 (39.6)	0 (0)	1 (12.5)	0 (0)	1 (25)	47 (1.3)
Poorly defined	26 (20.3)	10 (14.3)	8 (9.1)	524 (10.9)	5 (15.2)	2 (25)	0 (0)	0 (0)	75 (2.1)
Missing	8 (6.2)	39 (55.7)	69 (78.4)	2107 (43.9)	13 (39.4)	4 (50)	1 (100)	2 (50)	3440 (96)

Table 3.10: Variables used in the Brock2b model on NLST participants for each screening and their distributions. The NLST population was limited to participants with only one nodule. For numeric variables, the values are mean ± standard deviation. For categorical or binary variables, the values are count (%). **C:** cancer; **NC:** non-cancer; **LC:** lung cancer.

Characteristics	NLST (Participant count)					
	Scr1 C (128)	Scr1 NC (4961)	Scr2 C (70)	Scr2 NC (4891)	Scr3 C (88)	Scr3 NC (4803)
Brock2b risk score	0.326 ± 0.217	$0.025 \pm .07$	$0.191 \pm .167$	$0.024 \pm .068$	$0.211 \pm .189$	$0.021 \pm .064$
Age	63.9 ± 5.3	61.6 ± 5	63.5 ± 4.5	62.5 ± 5	64.8 ± 5.2	63.5 ± 5
Gender (% female)	55 (43)	2060 (41.5)	27 (38.6)	2033 (41.6)	36 (40.9)	1997 (41.6)
Fam. hist. of LC (% Yes)	34 (26.6)	1112 (22.4)	23 (32.9)	1089 (22.3)	22 (25)	1067 (22.2)
Hist. of emphysema (% Yes)	14 (10.9)	431 (8.7)	8 (11.4)	423 (8.6)	12 (13.6)	411 (8.6)
Nodule size (mm)	19.8 ± 11.7	4.4 ± 6	13.8 ± 11.7	4.3 ± 5.1	15.5 ± 15.1	4.2 ± 5
Nodule type						
Ground-glass	11 (8.6)	422 (8.5)	4 (5.7)	373 (7.6)	10 (11.4)	403 (8.4)
Part-solid	13 (10.2)	153 (3.1)	9 (12.9)	120 (2.5)	10 (11.4)	92 (1.9)
Nodule location (% Upper lobe)	83 (64.8)	1086 (21.9)	40 (57.1)	1068 (21.8)	62 (70.5)	1042 (21.7)
Spiculation (% Yes)	80 (62.5)	288 (5.8)	31 (44.3)	250 (5.1)	32 (36.4)	193 (4)

CHAPTER 4

modPOMDP2: The Two-Part modPOMDP

4.1 Introduction

In Chapter 3, MODPOMDP (mp) was shown to reach parity with physicians (p) in terms of true positives (mp: 57.33 vs. p: 57.26) and slightly more false positives (mp: 1975.95 vs. p: 1926.45) across three screenings in the NLST dataset. It is nevertheless desirable to reduce false positives even further. The key insight for the research in this chapter is that in general, the developed lung cancer screening POMDPs are aiming to optimize *earlier detection* of a positive finding and is based on mimicking experts' actions, which includes a relatively high false positive rate. Introducing another step, therefore, may keep the "earlier" detection while reducing the false positives.

MODPOMDP2 is designed to reduce the false positives by leveraging classification techniques to disambiguate a POMDP's positive predictions. That is, both true positive and false positive predictions of the POMDP are fed as inputs to a classifier. The goal is to fill in the gap in a POMDP's decision making and capture subtleties that MODPOMDP might have missed. While the two-part model has a long history and worked well in other fields, it does not seem to be utilized much in biomedical informatics. This work shows that chaining a POMDP and commonly used classifying techniques in a two-part framework can largely maintain true positives while decreasing false positives, more so than either technique alone.

The idea of the two-part model has existed for a long time and was used in the 1970's for rainfall prediction [100–102]. The two-part model is often used in the field of health

38

economics today where the task is often to predict healthcare spending and healthcare access. When it comes to healthcare spending, most patients do not spend or spend very little, but a small group can spend a lot. A similar trend is observed in healthcare access as well. This highly 0-skewed data therefore pose a challenge in the prediction task. Within the health economics domain, the two-part model approach is more or less the consensus [103]. Researchers use one model to predict whether a patient will spend or access healthcare, and a second model to predict the amount of spending or access.

Another area of research that employs combining methodologies is artificial intelligence (AI) in games. While Deep Blue has bested human players since 1997 [104], it was not until AlphaGo [105] emerged on the scene in 2016 that the best Go players were defeated by AI. Go is a particularly difficult game for AI because enumerating all possibilities is considered intractable due to its large state and action space. In the original AlphaGo, two neural networks were used to train a policy function and a value function, while a Monte Carlo Tree Search (MCTS) [106] is then used to optimize the policy. Effectively, the neural networks reduced the search space required for the MCTS to calculate optimized policies. Later versions, AlphaGo Zero [107] and AlphaZero, [108] reduced the neural networks down to a single joint policy and value network, but the same idea of reducing the search space followed by optimizing policy remained.

4.2 Methods

To demonstrate that the MODPOMDP2 method itself can lower false positives, only features previously used in MODPOMDP were used. Table 4.1 summarizes the features and their types. The first nine features are part of the $PLCO_{M2012}$ model [93], which was previously used to inform MODPOMDP's initial cancer state beliefs and showed good predictive power in NLST patients (AUC–ROC: 0.701). However, history of chronic obstructive pulmonary disease (COPD) and smoking status were excluded because an earlier feature importance
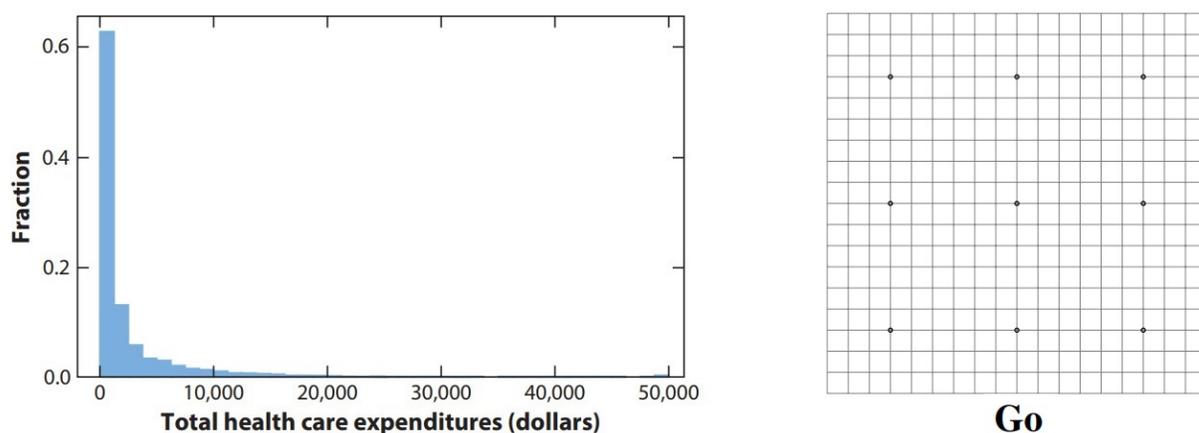
Figure 4.1: Left: figure showing the highly 0-skewed data in healthcare spending. Most patients do not spend or spend very little, while a small group spend a lot; adapted from [103]. Right: figure of an empty Go board, showing the initial large state and action spaces; adapted from [109].

analysis (not shown) showed low importance. Four additional features, gender and three baseline nodule characteristics (baseline nodule size bin, baseline attenuation, and baseline margins) were included to represent the patient's nodule characteristics at the start of screening.

Data preprocessing. The data were processed the same way as was done in Chapter 3.2.1, with additional steps for the classifiers. For categorical variables, one-hot encoding was used. k-1 was used for baseline attenuation and baseline margins because these had missing values (not all nodules appeared at baseline). An earlier analysis (not shown) did not show much difference in tree-based classifier performance whether one-hot encoded or not. Additionally, while baseline nodule size bin is categorical, it was treated as numeric. For `MLPClassifier` [110–113] (MLP), numeric columns (including baseline nodule size bin) were transformed using a standard scaler of mean 0 and variance 1.

Table 4.1: List of features and data types used in MODPOMDP2's classifiers.

Feature	Data Type
Age	Numeric
Race/ethnicity	Categorical
Education	Categorical
BMI	Numeric
Prior non-lung cancer	Binary
Family history of lung cancer	Binary
Cigarettes per day	Numeric
Total years of smoking	Numeric
Smoking quit time	Numeric
Gender	Categorical
Baseline nodule size bin	Categorical (treated as numeric)
Baseline attenuation	Categorical
Baseline margins	Categorical

Classifiers. Different classifiers provided in `scikit-learn` [114], `imbalanced-learn` [115], `XGBoost` [116] were explored for the purpose of implementing the MODPOMDP2 method. Due to the large imbalance in true positives (TPs) vs. false positives (FPs), a variety of techniques were used to address class imbalance. Thresholding [117] was used in conjunction with other methods by all classifiers for maximizing recall. Both `BalancedRandomForest-Classifier` [118] (BRF) and `EasyEnsembleClassifier` [119] (EEC) by default resample the non-minority class. `RandomForestClassifier` [120] (RF) and support vector classifier [121] (SVM) used the `class_weight=balanced` option to weigh the minority class more, while `XGBClassifier` (XGB) used the `scale_pos_weight` option to do the same. For MLP, over-sampling the minority class was tried as MLP does not have a built in class weight option. However, thresholding alone produced higher recall than thresholding + oversampling.

4.2.1 Classifier Selection

A classifier selection study using the NLST dataset (5,089 patients) was conducted to see how each classifier performs. This is both to serve as a baseline for comparisons and for selecting classifiers to combine with MODPOMDP. The dataset was divided into 60% training, 20% validation for threshold tuning, and 20% for testing. Hyperparameter tuning (see Table 4.2) was conducted for each classifier and all runs were over 200 iterations of bootstrap. For each classifier, results for the set of hyperparameters that produced the highest summed recall over three screening time points are shown in Table 4.3. Based on these results, two classifiers were selected for use in MODPOMDP2. The first one is MLP, which had the highest summed recall. The second one is BRF, which was selected for its relatively low FPs while maintaining similar TPs to EEC and XGB. RF and SVM were not considered because of RF's low TP counts and SVM's in-between performance relative to MLP and BRF.

Table 4.2: Classifier hyperparameter tuning.

Classifier	Hyperparameter	Categorical or Lower	Upper	Selected
BRF	n_estimators	100	200	200
EEC	n_estimators	100	200	200
RF	n_estimators	100	200	200
XGB	max_depth	6	10	6
SVM	kernel	linear, rbf, poly, sigmoid		linear
	C	0.1	100	1
MLP	activation	relu, identity, logistic, tanh		relu
	solver	adam, lbfgs, sgd		adam
	max_iteration	200	500	200
	hidden_layer_sizes	(6,)	(36,36)	(24,)

Table 4.3: Exploring classifiers for MODPOMDP2. Classification performance (mean counts ± 95% CI) of various methods over 200 iterations of bootstrap on the testing set. In each iteration, data were divided into 60% training, 20% validation, and 20% testing. Prediction threshold to bias for recall was tuned on the validation set and applied to the testing set. Hyperparameter tuning was performed for each classifier and the set of hyperparameters that gave the highest summed recall over three screening time points was selected. Classifiers are ordered by summed recall.

		TN	FP	FN	TP	Pre.	Rec.	ROC-AUC	PR-AUC
MLP	Scr1	269.17 ± 36.89	724.83 ± 36.89	$0.58 \pm .16$	$25.43 \pm .16$	$0.0464 \pm .0052$	$0.9779 \pm .0063$	$0.6179 \pm .0055$	$0.3778 \pm .0115$
	Scr2	34.02 ± 13.38	945.98 ± 13.38	$0.17 \pm .09$	$13.83 \pm .09$	$0.0145 \pm .0002$	$0.9879 \pm .0064$	$0.5000 \pm .0000$	$0.0446 \pm .0050$
	Scr3	44.31 ± 13.62	917.69 ± 13.62	$0.18 \pm .10$	$17.82 \pm .10$	$0.0193 \pm .0003$	$0.9900 \pm .0056$	$0.5002 \pm .0004$	$0.0387 \pm .0025$
	Sum	347.50	2588.50	0.93	57.08				
SVM	Scr1	470.82 ± 43.55	523.19 ± 43.55	$0.86 \pm .16$	$25.15 \pm .16$	$0.0637 \pm .0050$	$0.9671 \pm .0062$	$0.8787 \pm .0046$	$0.3282 \pm .0099$
	Scr2	77.00 ± 12.78	903.01 ± 12.78	$0.31 \pm .10$	$13.70 \pm .10$	$0.0151 \pm .0002$	$0.9782 \pm .0075$	$0.6116 \pm .0090$	$0.0323 \pm .0022$
	Scr3	108.63 ± 15.39	853.38 ± 15.39	$0.45 \pm .12$	$17.56 \pm .12$	$0.0204 \pm .0003$	$0.9753 \pm .0068$	$0.6347 \pm .0074$	$0.0328 \pm .0015$
	Sum	656.44	2279.57	1.61	56.40				
EEC	Scr1	556.64 ± 20.67	437.36 ± 20.67	$0.83 \pm .17$	$25.17 \pm .17$	$0.0622 \pm .0038$	$0.9681 \pm .0065$	$0.8711 \pm .0038$	$0.3608 \pm .0117$
	Scr2	204.79 ± 13.77	775.21 ± 13.77	$0.82 \pm .16$	$13.18 \pm .16$	$0.0169 \pm .0002$	$0.9414 \pm .0117$	$0.6047 \pm .0077$	$0.0287 \pm .0016$
	Scr3	159.48 ± 14.03	802.52 ± 14.03	$0.89 \pm .17$	$17.12 \pm .17$	$0.0211 \pm .0003$	$0.9508 \pm .0095$	$0.6289 \pm .0072$	$0.0446 \pm .0038$
	Sum	920.91	2020.54	2.54	55.47				
BRF	Scr1	584.38 ± 32.85	409.62 ± 32.85	$0.90 \pm .18$	$25.11 \pm .18$	$0.0779 \pm .0058$	$0.9656 \pm .0067$	$0.8910 \pm .0040$	$0.3530 \pm .0110$
	Scr2	257.28 ± 15.03	722.72 ± 15.03	$0.78 \pm .16$	$13.23 \pm .16$	$0.0182 \pm .0003$	$0.9446 \pm .0111$	$0.6364 \pm .0081$	$0.0300 \pm .0018$
	Scr3	227.78 ± 19.03	734.22 ± 19.03	$0.92 \pm .16$	$17.09 \pm .16$	$0.0233 \pm .0005$	$0.9492 \pm .0089$	$0.6622 \pm .0075$	$0.0380 \pm .0018$
	Sum	1069.44	1866.56	2.59	55.42				
XGB	Scr1	250.26 ± 29.85	743.75 ± 29.85	$0.87 \pm .17$	$25.13 \pm .17$	$0.0362 \pm .0019$	$0.9665 \pm .0067$	$0.7705 \pm .0064$	$0.3330 \pm .0110$
	Scr2	114.79 ± 14.34	865.21 ± 14.34	$0.98 \pm .18$	$13.03 \pm .18$	$0.0149 \pm .0002$	$0.9304 \pm .0129$	$0.5157 \pm .0046$	$0.0220 \pm .0011$
	Scr3	146.66 ± 14.82	815.35 ± 14.82	$0.96 \pm .17$	$17.04 \pm .17$	$0.0207 \pm .0003$	$0.9467 \pm .0096$	$0.5367 \pm .0057$	$0.0354 \pm .0023$
	Sum	511.70	2424.30	2.81	55.20				
RF	Scr1	724.93 ± 12.45	269.08 ± 12.45	$1.35 \pm .18$	$24.66 \pm .18$	$0.0920 \pm .0038$	$0.9483 \pm .0071$	$0.5254 \pm .0029$	$0.3422 \pm .0104$
	Scr2	442.51 ± 8.19	537.50 ± 8.19	$3.71 \pm .27$	$10.30 \pm .27$	$0.0189 \pm .0004$	$0.7354 \pm .0190$	$0.5000 \pm .0000$	$0.0230 \pm .0010$
	Scr3	393.46 ± 7.48	568.55 ± 7.48	$3.87 \pm .28$	$14.13 \pm .28$	$0.0243 \pm .0004$	$0.7850 \pm .0155$	$0.4999 \pm .0000$	$0.0326 \pm .0015$
	Sum	1560.89	1375.12	8.92	49.08				

4.2.2 The modPOMDP2 Framework

Figure 4.2 shows the experimental setup of MODPOMDP2. At each time point, the positive predictions (both TPs & FPs) from the POMDP training set were separated in a 75%/25% split where the 75% were used in a classifier's training and 25% were used to tune the decision

threshold towards biasing recall. Similar to the POMDP training data, positive predictions were also retrieved from the POMDP testing set. The classifier's decision threshold is applied to POMDP's testing positives and the combined model's confusion matrix calculations are summarized in Table 4.4. In short, the classifier's positives are the final positives and MOD-POMDP's negatives are adjusted by the classifier's negatives. For each instance of POMDP bootstrap, 100 instances of machine learning (ML) were performed. At each screening time point, the 95% confidence interval is calculated over $10,000$ sets of numbers.
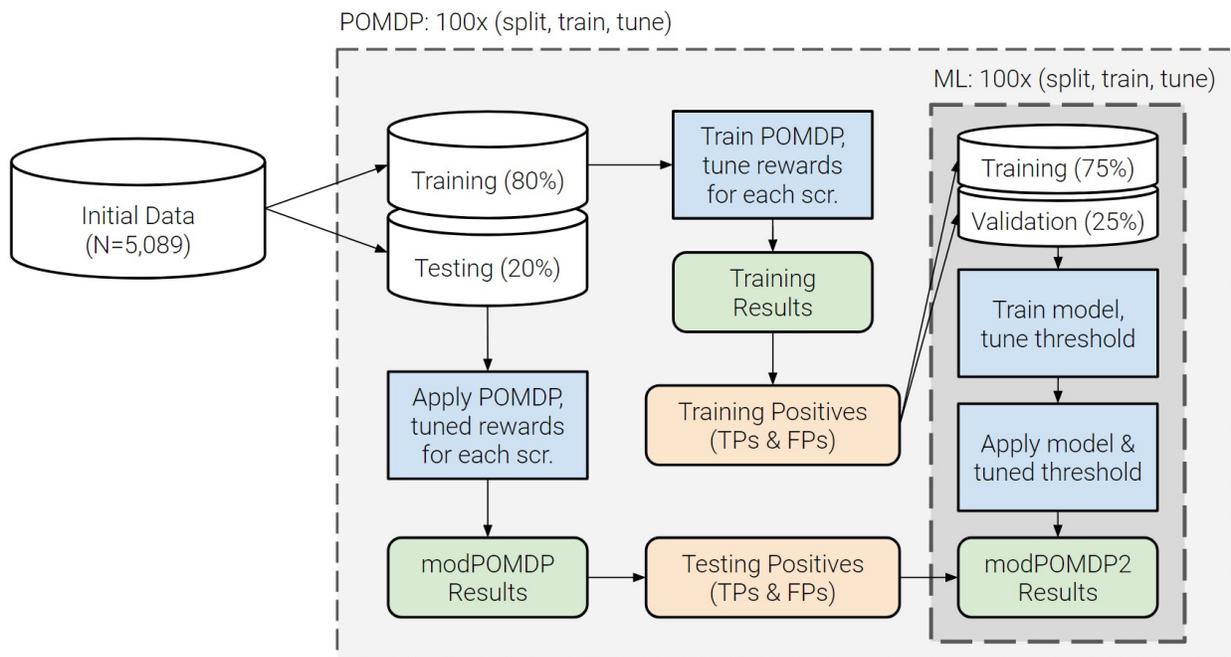
Figure 4.2: Experimental setup of MODPOMDP2. An extension of Figure 3.4, the positive predictions from the training results are used as inputs in the classifiers. For external validation, the `Testing (20%)` box is replaced with UCLA data. **ML:** machine learning.

4.2.3 The Brock Model

The Brock model was previously described in Section 3.2.4. Overall, Brock2b (B2b) performed similarly to BRF in TPs, but much lower FPs. The Brock model was therefore

Table 4.4: Confusion matrix calculation in the combined MODPOMDP2 model. **ML:** machine learning.

TN = POMDP TN + ML TN	TP = ML TP
FN = POMDP FN + ML FN	FP = ML FP

included in the analysis with MODPOMDP2 and chained similarly as the other classifiers. While other classifiers used baseline characteristics, Brock2b used information from the time of screening, with screening-adjusted ages. Similar to other classifiers, MODPOMDP positive predictions were split in two portions of 75% and 25%. As there is no training involved, only the 25% were used for threshold tuning and biased for recall. This step was done 100 times per POMDP instance. The tuned threshold was applied to the positive predictions from the POMDP testing set.

4.3 Results

Table 4.5 summarizes MODPOMDP2 results, with physician and MODPOMDP results reproduced from Table 3.5 and Table 3.6 for comparison purposes. The area under the curve values were calculated from the classifier/B2b portions only and therefore not available for the physicians and MODPOMDP. All three chained models, MODPOMDP + MLP (mp-MLP), MODPOMDP + BRF (mp-BRF), and MODPOMDP + B2b (mp-B2b), lowered FPs and TPs from MODPOMDP.

NLST. In the NLST testing set, the results are more or less mirroring classifier performance from the exploratory study. That is, mp-MLP had both more TPs and FPs than mp-BRF. When looking over all three screenings and compared to MODPOMDP, mp-MLP decreased TPs and FPs by 1.16 and 167.53 respectively, while mp-BRF decreased TPs and FPs by 2.62 and 512.43 respectively. These results indicate that in NLST, chaining MLP

can decrease FPs without affecting TPs by too much. Chaining BRF however, performed similarly to Scr3-O in TPs (mp-BRF: 54.71 vs. Scr3-O: 54.39), but worse in FPs (mp-BRF: 1463.52 vs. Scr3-O: 903.78). However, MODPOMDP2 achieved better results than MODPOMDP or individual classifiers alone. mp-B2b had the lowest TPs (53.14), but also lowered FPs substantially (697.40).

UCLA. When testing on the UCLA dataset, mp-BRF did better than mp-MLP, where mp-BRF achieved both higher TPs (mp-BRF: 44.24 vs. mp-MLP: 43.32) as well as lower FPs (mp-BRF: 1794.07 vs. mp-MLP: 2111.44). Both classifiers were able to decrease the FPs from MODPOMDP while maintaining similar TPs. When looking over all five screenings and compared to MODPOMDP, mp-MLP decreased TPs and FPs by 1.68 and 393.61 respectively, while mp-BRF decreased TPs and FPs by 0.76 and 710.98 respectively. These results indicate that chaining BRF can decrease FPs without affecting TPs by too much.

4.3.1 Error Analysis

To investigate the false negative predictions from the classifier, a study was conducted on a prior implementation of MODPOMDP2 using BRF on a larger cohort of $5,402$ NLST patients. In this study, 23 cases most likely to be BRF false negatives out of the $10,000$ classifier bootstrap instances were selected, along with a random sample of 6 cancer cases and 11 non-cancer cases as control. Two radiologists read all patient images[1] and reached consensus on Lung-RADS categories, whether nodule was mismatched, and any additional context or comments. Nodule mismatch was defined in one of two categories: a) the imaging-identified nodule was not the same as the procedure-identified (e.g., biopsy) nodule or b) the procedure-identified nodule was not detected in imaging. A total of 12 out of 23 false negative predictions were mismatched cases, suggesting a fairly large contribution towards misclassifi-

---

[1]Drs. Denise Aberle and Ashley Prosper read and interpreted patient LDCTs. Professor William Hsu, Yannan Lin, and Rina Ding assisted with image and data retrieval.

cations. Following this study, a comparison between imaging-identified nodule locations and procedure-identified nodule locations was done for all cancer patients, and mismatched cases were removed. However, three mismatched cases remained in the dataset, as their nodule locations did not differ between imaging and procedure. This shows that for some patients, nodules in the same lobe over time were actually different nodules. Unfortunately this is unrectifiable in the data. A summary of the findings can be found in Table 4.6.

As a result of the study, 51 cancer patients (Scr1: 17, Scr2: 18, Scr3: 16) out the original 5, 402 cohort were removed from the dataset. The match to mismatch ratio is summarized in Table 4.7. Note that the "Match" column does not match the number of cancers in Table 3.3 because of other exclusion criteria applied after this. The post-screen cohort were patients who did not develop cancer during the NLST screening period, but during the up to five years post-screening follow-up period. Unsurprisingly, these patients had a higher chance of nodule mismatches. In this dissertation, the post-screen patients were grouped under non-cancer patients as they did not develop cancer following a screen.

Within the 13 cases of mismatched cases identified in the BRF false negative study, A closer look at the categories of mismatched cases showed differences in cancer type distributions. When a mismatch is due to imaging and procedure location differences, the cancer types are more varied. When the nodule was not detected during screening, 50% were the fast-growing small-cell lung cancer (SCLC). These cancers would be less likely to be captured by LDCT and consequently by MODPOMDP and the classifier. As none of the three mismatched cases that remained in the dataset were SCLC, a slight improvement to classification is likely. The cancer type distributions are shown in Figure 4.3.

4.4   Discussion

This chapter described the design and implementation of a two-part model of modularized POMDP + classifiers and demonstrated that the combined model, MODPOMDP2, can
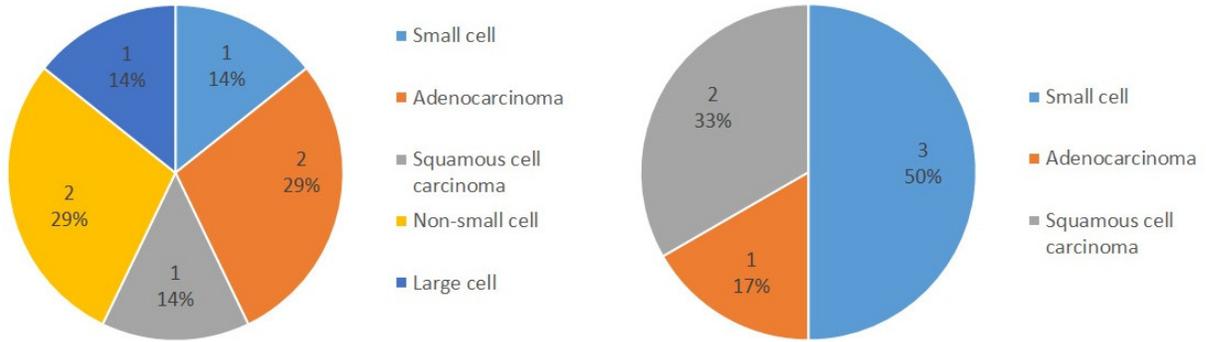
(a) Imaging-procedure locations are different.   **(b)** Nodule was not detected during imaging.

Figure 4.3: Distribution of cancer types for different nodule mismatch categories.

perform better than each of the two parts alone. A MODPOMDP + Brock2b model was developed for comparison purposes. MODPOMDP2 was trained and tested on the NLST dataset, and externally validated on the UCLA dataset. Statistics such as Youden's J statistic [122] and F-scores [123] were explored; however, high scores in these metrics did not result in higher true positives, as these statistics balance both positives and negatives.

Often times, modelers report areas under the curve results and say clinicians can pick a suitable decision point. However, picking this point is a very difficult choice. In this work, the classifiers' prediction thresholds were optimized for recall, as the cost of missing cancer cases is potentially very high for the patients screened. When compared to MODPOMDP or classifiers alone, the combined model can maintain similar true positives while further reducing false positives. This suggests that using MODPOMDP to filter out a portion of the negative samples is helpful to the classification task. Screen 2 of NLST appears especially difficult for the classifiers and the Brock model, mirroring both the physicians and MODPOMDP, as precision is lowest among the three screenings. Using baseline patient and nodule characteristics to predict cancer at different screening time points remain a challenge, as shown by the classifiers missing positive cases in both the classifier selection study and

when chained with MODPOMDP.

The MODPOMDP + Brock2b model having the lowest true positives out of the three chained models is somewhat surprising. First, Brock2b alone achieved very similar true positives as BRF alone (B2b: 55.52 vs. BRF: 55.42). Second, Brock2b uses data, including nodule characteristics, from screens 2 and 3. Third, Brock2b was tuned to bias for recall for each screen. It appears that these steps could not optimize true positive predictions as well as classifiers using baseline characteristics could. Previous research [99] performed external validation of the Brock model on the NLST dataset and found that the model performance can be improved by calibrating some of the beta coefficients. Adapting the recalibrated beta values may be one way to improve the MODPOMDP + Brock2b model in this work.

There are several key differences between the NLST and UCLA populations. The NLST RCT was a nationwide study and drew participants from different populations while UCLA patients are likely more localized. Overall, UCLA patients at baseline tended to be older, higher $PLCO_{M2012}$ risk score, higher proportion with scores ≥ 0.0151 (higher screening efficiency threshold [124]), higher education attainment, more minorities, much higher proportion with history of COPD, and much higher proportion with prior non-lung cancers. For smoking history, UCLA patients smoked fewer cigarettes per day, fewer current smokers, but also less smoking quit time. These results suggest MODPOMDP2 trained on the NLST dataset can generalize to a higher risk population. Baseline participant and patient characteristics can be found in Table 3.8.

There are several limitations to this work and improvements to the classifiers can be made in several ways. While a fairly extensive hyperparameter tuning experiment was done for MLP and SVM, a more thorough hyperparameter tuning step can be done for some of the other classifiers. In XGB for example, `eta` and `booster` were identified as important hyperparameters whose tuning can lead to gains in performance [125]. Another limitation is the use of baseline patient and nodule characteristics. It can be seen from Table 3.9 that in NLST, large number of cancer patients after screen 1 did not have a suspicious finding

at baseline (Scr2: 52.9%, Scr3: 75%). This is due to many cancer nodules not having appeared or grown large enough yet and likely contributing to the large number of false positives experienced by the classifiers. The indication for this is that the false positive rate (not shown) of NLST's screen 1 is consistently the lowest for all classifiers in Table 4.3. Additionally, MODPOMDP + Brock2b, which calculates Brock risk scores at each screening using data from those respective screens, has the lowest false positives out of the three chained models at only 697.40 across three screenings. Therefore, a feature selection process that incorporates nodule features from the time of the screen would likely help lower false positives significantly in screens 2 and 3. For screen 2 cancer patients at screen 2, 95.7% had nodule characteristics with an average nodule size of 14.4 ± 11.7 mm. For screen 3 cancer patients at screen 3, 97.7% had nodule characteristics with an average nodule size of 15.8 ± 15.2 mm. Lastly, only positive MODPOMDP predictions are used as input to the classifiers. While filtering out negative cases may help the classification task as previously mentioned, this also means that positive cases missed by the POMDP will never make it to the classifier. A separate two-part model that incorporates both positive and negative MODPOMDP predictions can be used to study the relative benefits of including each.

Table 4.5: Results for modPOMDP2 on both NLST and UCLA dataset. Asterisks (*) indicate results reproduced from Table 3.5 and Table 3.6. For modPOMDP2, model performance (mean counts ± 95% CI) is calculated over $10,000$ instances of bootstrap on the testing set and the area under the curve values are from the classifier/B2b portions only.

NLST		**TN**	**FP**	**FN**	**TP**	**Pre.**	**Rec.**	**ROC-AUC**	**PR-AUC**
	Scr1	480.43 ± 2.89	513.57 ± 2.89	$0.00 \pm .00$	$26.00 \pm .00$	$0.0482 \pm .0003$	$1.0000 \pm .0000$	n/a	n/a
	Scr2	322.14 ± 2.83	657.86 ± 2.83	$0.74 \pm .14$	$13.26 \pm .14$	$0.0198 \pm .0002$	$0.9472 \pm .0102$	n/a	n/a
Physician*	Scr3	206.98 ± 2.39	755.02 ± 2.39	$0.00 \pm .00$	$18.00 \pm .00$	$0.0233 \pm .0001$	$1.0000 \pm .0000$	n/a	n/a
	Sum	1009.55	1926.45	0.74	57.26				
	Scr1	376.87 ± 14.23	617.13 ± 14.23	$0.26 \pm .12$	$25.74 \pm .12$	$0.0406 \pm .0010$	$0.9900 \pm .0047$	n/a	n/a
	Scr2	33.06 ± 26.06	946.94 ± 26.06	$0.14 \pm .12$	$13.86 \pm .12$	$0.0148 \pm .0006$	$0.9900 \pm .0083$	n/a	n/a
modPOMDP*	Scr3	550.12 ± 3.51	411.88 ± 3.51	$0.27 \pm .09$	$17.73 \pm .09$	$0.0413 \pm .0002$	$0.9850 \pm .0049$	n/a	n/a
	Sum	960.05	1975.95	0.67	57.33				
	Scr1	481.18 ± 3.48	512.82 ± 3.48	$0.76 \pm .03$	$25.24 \pm .03$	$0.0583 \pm .0007$	$0.9708 \pm .0010$	$0.6114 \pm .0008$	$0.3794 \pm .0015$
modPOMDP	Scr2	66.81 ± 3.07	913.19 ± 3.07	$0.34 \pm .02$	$13.66 \pm .02$	$0.0153 \pm .0001$	$0.9760 \pm .0013$	$0.5000 \pm .0000$	$0.0400 \pm .0005$
+ MLP	Scr3	579.60 ± 0.66	382.40 ± 0.66	$0.74 \pm .02$	$17.26 \pm .02$	$0.0433 \pm .0001$	$0.9589 \pm .0011$	$0.5000 \pm .0000$	$0.0558 \pm .0003$
	Sum	1127.58	1808.42	1.83	56.17				
	Scr1	626.75 ± 3.54	367.25 ± 3.54	$1.19 \pm .03$	$24.81 \pm .03$	$0.0794 \pm .0007$	$0.9541 \pm .0010$	$0.8427 \pm .0007$	$0.3417 \pm .0014$
modPOMDP	Scr2	265.99 ± 2.64	714.01 ± 2.64	$0.99 \pm .03$	$13.01 \pm .03$	$0.0184 \pm .0001$	$0.9294 \pm .0019$	$0.6274 \pm .0012$	$0.0303 \pm .0003$
+ BRF	Scr3	579.73 ± 0.63	382.27 ± 0.63	$1.11 \pm .02$	$16.89 \pm .02$	$0.0424 \pm .0001$	$0.9382 \pm .0014$	$0.5694 \pm .0011$	$0.0595 \pm .0004$
	Sum	1472.48	1463.52	3.29	54.71				
	Scr1	722.99 ± 1.53	271.01 ± 1.53	$1.26 \pm .03$	$24.74 \pm .03$	$0.0907 \pm .0005$	$0.9517 \pm .0011$	$0.9265 \pm .0005$	$0.4447 \pm .0015$
modPOMDP	Scr2	702.27 ± 3.18	277.73 ± 3.18	$2.06 \pm .03$	$11.94 \pm .03$	$0.0528 \pm .0005$	$0.8529 \pm .0021$	$0.8803 \pm .0010$	$0.1435 \pm .0011$
+ B2b	Scr3	813.34 ± 0.70	148.66 ± 0.70	$1.54 \pm .03$	$16.46 \pm .03$	$0.1040 \pm .0004$	$0.9144 \pm .0015$	$0.8798 \pm .0006$	$0.2411 \pm .0014$
	Sum	2238.60	697.40	4.86	53.14				
UCLA		**TN**	**FP**	**FN**	**TP**	**Pre.**	**Rec.**	**ROC-AUC**	**PR-AUC**
	Scr1	3555.00 ± 53.95	14.00 ± 53.95	$0.00 \pm .00$	$33.00 \pm .00$	$0.7021 \pm .0000$	$1.0000 \pm .0000$	n/a	n/a
	Scr1.5	858.00 ± 15.09	246.00 ± 15.09	$0.00 \pm .00$	$8.00 \pm .00$	$0.0315 \pm .0000$	$1.0000 \pm .0000$	n/a	n/a
	Scr2	284.00 ± 10.73	118.00 ± 10.73	$0.00 \pm .00$	$1.00 \pm .00$	$0.0084 \pm .0000$	$1.0000 \pm .0000$	n/a	n/a
Physician*	Scr2.5	91.00 ± 3.15	59.00 ± 3.15	$0.00 \pm .00$	$0.00 \pm .00$	$0.0000 \pm .0000$	n/a	n/a	n/a
	Scr3	33.00 ± 0.89	31.00 ± 0.89	$1.00 \pm .00$	$3.00 \pm .00$	$0.0882 \pm .0000$	$0.7500 \pm .0000$	n/a	n/a
	Sum	4821.00	468.00	1.00	45.00				
	Scr1	2157.50 ± 53.95	1411.50 ± 53.95	$0.00 \pm .00$	$33.00 \pm .00$	$0.0238 \pm .0011$	$1.0000 \pm .0000$	n/a	n/a
	Scr1.5	578.35 ± 15.09	525.65 ± 15.09	$1.00 \pm .00$	$7.00 \pm .00$	$0.0135 \pm .0004$	$0.8750 \pm .0000$	n/a	n/a
	Scr2	13.51 ± 10.73	388.49 ± 10.73	$0.00 \pm .00$	$1.00 \pm .00$	$0.0027 \pm .0002$	$1.0000 \pm .0000$	n/a	n/a
modPOMDP*	Scr2.5	3.95 ± 3.15	146.05 ± 3.15	$0.00 \pm .00$	$0.00 \pm .00$	$0.0000 \pm .0000$	n/a	n/a	n/a
	Scr3	30.64 ± 0.89	33.36 ± 0.89	$0.00 \pm .00$	$4.00 \pm .00$	$0.1089 \pm .0030$	$1.0000 \pm .0000$	n/a	n/a
	Sum	2783.95	2505.05	1.00	45.00				
	Scr1	2441.69 ± 8.91	1127.31 ± 8.91	$0.07 \pm .01$	$32.93 \pm .01$	$0.0400 \pm .0007$	$0.9979 \pm .0002$	$0.6340 \pm .0009$	$0.3383 \pm .0011$
	Scr1.5	675.80 ± 2.97	428.20 ± 2.97	$1.26 \pm .01$	$6.74 \pm .01$	$0.0190 \pm .0002$	$0.8428 \pm .0014$	$0.5739 \pm .0009$	$0.0854 \pm .0006$
modPOMDP	Scr2	19.88 ± 1.12	382.12 ± 1.12	$0.00 \pm .00$	$1.00 \pm .00$	$0.0027 \pm .0000$	$1.0000 \pm .0000$	$0.4999 \pm .0000$	$0.0321 \pm .0006$
+ MLP	Scr2.5	6.18 ± 0.34	143.82 ± 0.34	$0.00 \pm .00$	$0.00 \pm .00$	$0.0000 \pm .0000$	n/a	$0.0000 \pm .0000$	$0.0000 \pm .0000$
	Scr3	34.01 ± 0.11	29.99 ± 0.11	$1.35 \pm .02$	$2.65 \pm .02$	$0.0805 \pm .0006$	$0.6630 \pm .0054$	$0.5002 \pm .0001$	$0.1596 \pm .0014$
	Sum	3177.56	2111.44	2.68	43.32				
	Scr1	2658.28 ± 9.94	910.72 ± 9.94	$0.01 \pm .00$	$32.99 \pm .00$	$0.0542 \pm .0008$	$0.9997 \pm .0001$	$0.9400 \pm .0003$	$0.3446 \pm .0007$
	Scr1.5	753.13 ± 3.42	350.87 ± 3.42	$1.48 \pm .01$	$6.52 \pm .01$	$0.0248 \pm .0003$	$0.8149 \pm .0014$	$0.8065 \pm .0008$	$0.0882 \pm .0008$
modPOMDP	Scr2	38.97 ± 1.12	363.03 ± 1.12	$0.01 \pm .00$	$0.99 \pm .00$	$0.0028 \pm .0000$	$0.9916 \pm .0018$	$0.6076 \pm .0040$	$0.0048 \pm .0002$
+ BRF	Scr2.5	12.35 ± 0.34	137.65 ± 0.34	$0.00 \pm .00$	$0.00 \pm .00$	$0.0000 \pm .0000$	n/a	$0.0000 \pm .0000$	$0.0000 \pm .0000$
	Scr3	32.20 ± 0.09	31.80 ± 0.09	$0.26 \pm .01$	$3.74 \pm .01$	$0.1069 \pm .0004$	$0.9354 \pm .0026$	$0.3854 \pm .0015$	$0.1239 \pm .0014$
	Sum	3494.93	1794.07	1.76	44.24				

Table 4.6: Summary of BRF false negative (FN) analysis.

Type	Screen	Cases	Mismatched Cases	Mismatchs Remained
Cancer	Scr1	2	0	0
	Scr2	2	1	1
	Scr3	2	0	0
Non-cancer	All	11	0	0
BRF FN	Scr1	9	4	0
	Scr2	6	3	1
	Scr3	8	5	1
Total		40	13	3

Table 4.7: Summary of matching and non-matching nodule locations.

	Match (%)	Mismatch (%)	Total
Scr1	135 (83)	27 (17)	162
Scr2	70 (82)	15 (18)	85
Scr3	90 (84)	17 (16)	107
Post-screen	38 (39)	59 (61)	97

CHAPTER 5

Conclusion

This chapter summarizes the results and findings of this dissertation. The potential areas of improvement and future research directions are also described.

5.1 Summary of Research

This dissertation presents methods for optimizing actions under a sequential decision making context in lung cancer screening. Methods were developed to address different aspects of the optimization task in screening while taking into account temporal differences. The optimization task can be thought of as two separate but related optimizations, namely the need to improve true positive predictions while reducing false positive predictions at the same time. The specific contributions of this dissertation are as follows:

1. **An improved POMDP implementation to address the stationarity assumption by optimizing predictions for each decision epoch.** This task was achieved through modularizing larger POMDP models into smaller, time-dependent constituent submodels to account for finer temporal differences. Results in Table 3.5 show the improvements in the NLST. Other improvements were also made to the implementation of a lung cancer screening POMDP. For example in the preprocessing step, patients who had mismatched procedure-identified nodules and screening-identified nodules were identified and removed. Evaluation was conducted in a more rigorous fashion by utilizing hold-out test sets and bootstrapping instead of k-fold cross validation. The UCLA dataset was also introduced

for external validation. Though the approach did not sufficiently generalize to the UCLA data, the trend that modularization can improve results is nevertheless apparent. The differences between the UCLA and NLST populations also highlights the potential need for more tailored modeling.

2. **A two-part model approach using classifiers for optimizing positive predictions from the POMDP model.** The two-part model shows that POMDP's positive predictions can be further disambiguated into true positives and true negatives using traditional machine learning techniques. Many classifiers were explored extensively, such as using different strategies for class imbalance and hyperparameter tuning. Results show the two-part approach can decrease false positives at a small cost to true positives.

3. **Demonstrating the validity of the POMDP-based approach for sequential decision making.** This work implemented the Brock model as a comparison to modPOMDP and found that modPOMDP captured more cancer cases over multiple screenings. As capturing cancer cases is an important metric in a lung cancer screening context, this work demonstrated the value and potential of POMDPs for making better consecutive decisions over a conventional (e.g., logistic regression-based) approach. In other words, learning the screening process across multiple time points as a whole is more advantageous than learning on individual time points separately.

5.2 Future Directions

The work in this dissertation has several avenues for improvement. Some of which were previously discussed in Sections 3.4 and 4.4.

modPOMDP setup and components. The POMDP stationarity assumption in this work is addressed mostly through the reward function. However, other components of the POMDP such as transition probabilities can nonetheless be modeled as time-variant. One

possibility is to use DBN implementations that explicitly model transition probabilities as a function of t, such as in non-stationary DBNs. Furthermore, section 3.4 showed that the modPOMDP setup can influence results. Future work can use the following setup in both modPOMDP and modPOMDP2: $I = 2$ where $\sigma_1 = [1, 3)$ and $\sigma_2 = [3, 3]$.

UCLA data update. UCLA data is currently very sparse, with the number of nodules falling very quickly after the first screen. Considering the last data update was September 2021, an update should increase the dataset size and potentially allow more meaningful comparisons. As the UCLA dataset grows, it may be possible for training a new model as well.

UCLA screening interval. Limitations surrounding UCLA screening intervals were previously discussed in 3.4. In brief, the POMDP model was trained on NLST data with interpolated half-year values. The modeling time intervals is therefore six months, but this value is not enforced in the UCLA data. The highly varied screening intervals across UCLA patients may affect the underlying natural state transitions. Future work could model each nodule's screening trajectory individually and allow for sampling at designated time intervals.

Classifier improvements. Limitations and possible improvements to the classifiers were previously discussed in 4.4. Using baseline nodule characteristics for classification beyond screen 1 likely contributed to the high number of false positives. The Brock2b model, using nodule characteristics from screens 2 and 3, substantially lowered the number of false positives. Future work should incorporate nodule characteristics from the same time point as the screen in the classification task. A feature selection process to include more features or even directly extract features from the source images using feature extraction techniques can be done as well. Secondly, the hyperparameter space of classifiers is large and can be further explored, especially as features change. Thirdly, only positive predictions from the POMDP

are used as inputs to the classifier. Incorporating negative predictions can be explored and assessed for contribution to classification.

Generalizability of the model. While the NLST dataset is limited by the NLST inclusion criteria, the UCLA dataset is not similarly restricted. However, the NLST criteria is currently imposed on the UCLA data to get a more comparable population. Future studies should explore relaxing the inclusion criteria to see how MODPOMDP and MODPOMDP2 perform on a wider population. For example, the USPSTF recommends adults between the ages 50-80 and 20 pack-years of smoking history for annual LDCT. If the USPSTF criteria were used, then the UCLA dataset can include both younger and older patients, as well as lighter smokers than the current set.

Explainability of the model. An often cited barrier to physician trust of modeling techniques, including machine learning, is the explainability of the model outcome. That is, the means of justifying clinical decisions [126]. For example, deep learning models are often considered "black boxes" that lack explainability. POMDPs on the other hand, have several advantages. First, POMDP decisions are based on beliefs, which include the belief in the cancer state. This is similar to the probability of cancer from risk models. Second, the optimal policy π^* can be visualized as a *policy graph*, which is effectively a rule-based chart of when to take what action – another easily interpretable construct. Future work can explore extending policy graphs to include information from MODPOMDP and MODPOMDP2 with physician input for deployment in a clinical setting.

REFERENCES

[1] Panayiotis Petousis, Simon X Han, William Hsu, and Alex AT Bui. Generating reward functions using irl towards individualized cancer screening. In *Artificial Intelligence in Health: First International Workshop, AIH 2018, Stockholm, Sweden, July 13-14, 2018, Revised Selected Papers 1*, pages 213–227. Springer, 2019.

[2] Panayiotis Petousis, Audrey Winter, William Speier, Denise R Aberle, William Hsu, and Alex AT Bui. Using sequential decision making to improve lung cancer screening performance. *Ieee Access*, 7:119403–119419, 2019.

[3] Rebecca L Siegel, Kimberly D Miller, Nikita Sandeep Wagle, and Ahmedin Jemal. Cancer statistics, 2023. *CA: a cancer journal for clinicians*, 73(1):17–48, 2023.

[4] Virginia A Moyer and US Preventive Services Task Force*. Screening for lung cancer: Us preventive services task force recommendation statement. *Annals of internal medicine*, 160(5):330–338, 2014.

[5] Alex H Krist, Karina W Davidson, Carol M Mangione, Michael J Barry, Michael Cabana, Aaron B Caughey, Esa M Davis, Katrina E Donahue, Chyke A Doubeni, Martha Kubik, et al. Screening for lung cancer: Us preventive services task force recommendation statement. *Jama*, 325(10):962–970, 2021.

[6] Daniel E Jonas, Daniel S Reuland, Shivani M Reddy, Max Nagle, Stephen D Clark, Rachel Palmieri Weber, Chineme Enyioha, Teri L Malo, Alison T Brenner, Charli Armstrong, et al. Screening for lung cancer with low-dose computed tomography: updated evidence report and systematic review for the us preventive services task force. *Jama*, 325(10):971–987, 2021.

[7] Stuart J Russell. *Artificial intelligence a modern approach*. Pearson Education, Inc., 2010.

[8] William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294, 1933.

[9] Herbert Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535, 1952.

[10] Richard Bellman. A problem in the sequential design of experiments. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 16(3/4):221–229, 1956.

[11] John C Gittins and David M Jones. A dynamic allocation index for the discounted multiarmed bandit problem. *Biometrika*, 66(3):561–565, 1979.

[12] Tze Leung Lai, Herbert Robbins, et al. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

[13] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multi-armed bandit problem. *Machine learning*, 47(2):235–256, 2002.

[14] Eyal Even-Dar, Shie Mannor, Yishay Mansour, and Sridhar Mahadevan. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of machine learning research*, 7(6), 2006.

[15] Michael Woodroofe. A one-armed bandit problem with a concomitant variable. *Journal of the American Statistical Association*, 74(368):799–806, 1979.

[16] Jyotirmoy Sarkar. One-armed bandit problems with covariates. *The Annals of Statistics*, pages 1978–2002, 1991.

[17] John Langford and Tong Zhang. The epoch-greedy algorithm for multi-armed bandits with side information. *Advances in neural information processing systems*, 20, 2007.

[18] Naoki Abe. Learning to optimally schedule internet banner advertisements. In *Proc. of 16th Int. Conf. on Machine Learning*, pages 12–21, 1999.

[19] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670, 2010.

[20] Sofía S Villar, Jack Bowden, and James Wason. Multi-armed bandit models for the optimal design of clinical trials: benefits and challenges. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 30(2):199, 2015.

[21] Linqi Song, William Hsu, Jie Xu, and Mihaela Van Der Schaar. Using contextual learning to improve diagnostic accuracy: Application in breast cancer screening. *IEEE journal of biomedical and health informatics*, 20(3):902–914, 2015.

[22] Ambuj Tewari and Susan A Murphy. From ads to interventions: Contextual bandits in mobile health. In *Mobile Health*, pages 495–517. Springer, 2017.

[23] AJ Schaefer, MD Bailey, SM Shechter, and MS Roberts. Modeling medical treatment using markov decision processes, international series in operations research and management science, 70 (4), 593-612, 2005.

[24] Oguzhan Alagoz, Heather Hsu, Andrew J Schaefer, and Mark S Roberts. Markov decision processes: a tool for sequential decision making under uncertainty. *Medical Decision Making*, 30(4):474–483, 2010.

[25] Jagpreet Chhatwal, Oguzhan Alagoz, and Elizabeth S Burnside. Optimal breast biopsy decision-making based on mammographic features and demographic factors. *Operations research*, 58(6):1577–1591, 2010.

[26] Elizabeth S Burnside, Jagpreet Chhatwal, and Oguzhan Alagoz. What is the optimal threshold at which to recommend breast biopsy? *PloS one*, 7(11):e48820, 2012.

[27] Oguzhan Alagoz, Jagpreet Chhatwal, and Elizabeth S Burnside. Optimal policies for reducing unnecessary follow-up mammography exams in breast cancer diagnosis. *Decision Analysis*, 10(3):200–224, 2013.

[28] Mehrnaz Abdollahian and Tapas K Das. A mdp model for breast and ovarian cancer intervention strategies for brca1/2 mutation carriers. *IEEE journal of biomedical and health informatics*, 19(2):720–727, 2014.

[29] George E Monahan. State of the art—a survey of partially observable markov decision processes: theory, models, and algorithms. *Management science*, 28(1):1–16, 1982.

[30] Michael L Littman. A tutorial on partially observable markov decision processes. *Journal of Mathematical Psychology*, 53(3):119–125, 2009.

[31] American College of Radiology et al. Breast imaging reporting and data system. *BI-RADS*, 2003.

[32] Turgay Ayer, Oguzhan Alagoz, and Natasha K Stout. Or forum—a pomdp approach to personalize mammography screening decisions. *Operations Research*, 60(5):1019–1034, 2012.

[33] Moshe Leshno, Zamir Halpern, and Nadir Arber. Cost-effectiveness of colorectal cancer screening in the average risk population. *Health care management science*, 6(3):165–174, 2003.

[34] Fatih Safa Erenay, Oguzhan Alagoz, and Adnan Said. Optimizing colonoscopy screening for colorectal cancer prevention and surveillance. *Manufacturing & Service Operations Management*, 16(3):381–400, 2014.

[35] Jingyu Zhang, Brian T Denton, Hari Balasubramanian, Nilay D Shah, and Brant A Inman. Optimization of prostate biopsy referral decisions. *Manufacturing & Service Operations Management*, 14(4):529–547, 2012.

[36] Jingyu Zhang, Brian T Denton, Hari Balasubramanian, Nilay D Shah, and Brant A Inman. Optimization of psa screening policies: a comparison of the patient and societal perspectives. *Medical Decision Making*, 32(2):337–349, 2012.

[37] Athanasios Tsoukalas, Timothy Albertson, Ilias Tagkopoulos, et al. From data to optimal decision making: a data-driven, probabilistic machine learning approach to decision support for patients with sepsis. *JMIR medical informatics*, 3(1):e3445, 2015.

[38] Athanassios Vozikis and John E Goulionis. Medical decision making for patients with parkinson disease under average cost criterion. *Australia and New Zealand health policy*, 6(1), 2009.

[39] John E Goulionis, Athanassios Vozikis, VK Benos, and D Nikolakis. On the decision rules of cost-effective treatment for patients with diabetic foot syndrome. *ClinicoEconomics and outcomes research: CEOR*, 2:121, 2010.

[40] A Vozikis, JE Goulionis, and VK Benos. The partially observable markov decision processes in healthcare: an application to patients with ischemic heart disease (ihd). *Operational Research*, 12(1):3–14, 2012.

[41] Edward J Sondik. The optimal control of partially observable markov processes over the infinite horizon: Discounted costs. *Operations research*, 26(2):282–304, 1978.

[42] Chelsea C White III and William T Scherer. Solution procedures for partially observed markov decision processes. *Operations Research*, 37(5):791–797, 1989.

[43] Craig Boutilier and David Poole. Computing optimal policies for partially observable decision processes using compact representations. In *Proceedings of the National Conference on Artificial Intelligence*, pages 1168–1175, 1996.

[44] Katsushige Sawaki and Akira Ichikawa. Optimal control for partially observable markov decision processes over an infinite horizon. *Journal of the Operations Research Society of Japan*, 21(1):1–16, 1978.

[45] Anthony R Cassandra, Leslie Pack Kaelbling, and Michael L Littman. Acting optimally in partially observable stochastic domains. In *Aaai*, volume 94, pages 1023–1028, 1994.

[46] HT Cheng. Algorithms for partially observable markov decision processes [ph. d. thesis]. *Vancouver Columbia: University of British Columbia*, 1988.

[47] Michael L Littman. The witness algorithm: Solving partially observable markov decision processes. *Brown University, Providence, RI*, 1994.

[48] Michael L Littman, Anthony R Cassandra, and Leslie P Kaelbling. Efficient dynamic-programming updates in partially observable markov decision processes, 1995.

[49] Anthony Cassandra, Michael L Littman, and Nevin L Zhang. Incremental pruning: a simple, fast, exact method for partially observable markov decision processes. In *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence*, pages 54–61, 1997.

[50] Joelle Pineau, Geoff Gordon, Sebastian Thrun, et al. Point-based value iteration: An anytime algorithm for pomdps. In *Ijcai*, volume 3, pages 1025–1032, 2003.

[51] Hanna Kurniawati, David Hsu, and Wee Sun Lee. Sarsop: Efficient point-based pomdp planning by approximating optimally reachable belief spaces. In *Robotics: Science and systems*, volume 2008. Citeseer, 2008.

[52] Guy Shani, Ronen I Brafman, and Solomon Eyal Shimony. Prioritizing point-based pomdp solvers. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 38(6):1592–1605, 2008.

[53] Trey Smith and Reid Simmons. Point-based pomdp algorithms: Improved analysis and implementation. *arXiv preprint arXiv:1207.1412*, 2012.

[54] Matthijs TJ Spaan and N Spaan. A point-based pomdp algorithm for robot planning. In *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004*, volume 3, pages 2399–2404. IEEE, 2004.

[55] Matthijs TJ Spaan and Nikos Vlassis. Perseus: Randomized point-based value iteration for pomdps. *Journal of artificial intelligence research*, 24:195–220, 2005.

[56] Guy Shani, Joelle Pineau, and Robert Kaplow. A survey of point-based pomdp solvers. *Autonomous Agents and Multi-Agent Systems*, 27:1–51, 2013.

[57] Michael L Littman, Anthony R Cassandra, and Leslie Pack Kaelbling. Learning policies for partially observable environments: Scaling up. In *Machine Learning Proceedings 1995*, pages 362–370. Elsevier, 1995.

[58] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. Probalistic robotics. *Kybernetes*, 35(7/8):1299–1300, 2006.

[59] Panayiotis Petousis. *Optimizing cancer screening with POMDPs*. University of California, Los Angeles, 2019.

[60] Marcel AJ Van Gerven, Babs G Taal, and Peter JF Lucas. Dynamic bayesian networks as prognostic models for clinical patient management. *Journal of biomedical informatics*, 41(4):515–529, 2008.

[61] Kevin Murphy et al. The bayes net toolbox for matlab. *Computing science and statistics*, 33(2):1024–1034, 2001.

[62] Nicandro Cruz-Ramirez, Hector Gabriel Acosta-Mesa, Humberto Carrillo-Calvet, Luis Alonso Nava-Fernández, and Rocio Erandi Barrientos-Martinez. Diagnosis of breast cancer using bayesian networks: A case study. *Computers in Biology and Medicine*, 37(11):1553–1564, 2007.

[63] Olivier Gevaert, Frank De Smet, Dirk Timmerman, Yves Moreau, and Bart De Moor. Predicting the prognosis of breast cancer by integrating clinical and microarray data with bayesian networks. *Bioinformatics*, 22(14):e184–e190, 2006.

[64] Jung Hun Oh, Jeffrey Craft, Rawan Al Lozi, Manushka Vaidya, Yifan Meng, Joseph O Deasy, Jeffrey D Bradley, and Issam El Naqa. A bayesian network approach for modeling local failure in lung cancer. *Physics in Medicine & Biology*, 56(6):1635, 2011.

[65] R Marshall Austin and Agnieszka Onisko. Increased cervical cancer risk associated with extended screening intervals after negative human papillomavirus test results: Bayesian risk estimates using the pittsburgh cervical cancer screening model. *Journal of the American Society of Cytopathology*, 5(1):9–14, 2016.

[66] Alexander Stojadinovic, Anton Bilchik, David Smith, John S Eberhardt, Elizabeth Ben Ward, Aviram Nissan, Eric K Johnson, Mladjan Protic, George E Peoples, Itzhak Avital, et al. Clinical decision support and individualized prediction of survival in colon cancer: bayesian belief network model. *Annals of surgical oncology*, 20:161–174, 2013.

[67] Konstantinos P Exarchos, George Rigas, Yorgos Goletsis, and Dimitrios I Fotiadis. Modelling of oral cancer progression using dynamic bayesian networks. In *Data Mining for Biomarker Discovery*, pages 199–212. Springer, 2012.

[68] Micol Sandri, Paola Berchialla, Ileana Baldi, Dario Gregori, and Roberto Alberto De Blasi. Dynamic bayesian networks to predict sequences of organ failures in patients admitted to icu. *Journal of biomedical informatics*, 48:106–113, 2014.

[69] Marion Verduijn, Peter MJ Rosseel, Niels Peek, Evert de Jonge, and Bas AJM de Mol. Prognostic bayesian networks: Ii: An application in the domain of cardiac surgery. *Journal of biomedical informatics*, 40(6):619–630, 2007.

[70] Anthony Costa Constantinou, Mark Freestone, William Marsh, and Jeremy Coid. Causal inference for violence risk management and decision support in forensic psychiatry. *Decision Support Systems*, 80:42–55, 2015.

[71] German Cuaya, Angélica Munoz-Meléndez, Lidia Nunez Carrera, Eduardo F Morales, Ivett Quinones, Alberto I Pérez, and Aldo Alessi. A dynamic bayesian network for estimating the risk of falls from real gait data. *Medical & biological engineering & computing*, 51:29–37, 2013.

[72] Emily W Watt and Alex AT Bui. Evaluation of a dynamic bayesian belief network to predict osteoarthritic knee pain using data from the osteoarthritis initiative. In *AMIA annual symposium proceedings*, volume 2008, page 788. American Medical Informatics Association, 2008.

[73] Pedro Shiguihara, Alneu De Andrade Lopes, and David Mauricio. Dynamic bayesian network modeling, learning, and inference: a survey. *IEEE Access*, 9:117639–117648, 2021.

[74] Oguzhan Alagoz, Turgay Ayer, and Fatih Safa Erenay. Operations research models for cancer screening. *Wiley encyclopedia of operations research and management science*, 2010.

[75] Andrew J Schaefer, Matthew D Bailey, Steven M Shechter, and Mark S Roberts. Modeling medical treatment using markov decision processes. *Operations research and health care: A handbook of methods and applications*, pages 593–612, 2004.

[76] Casey C Bennett and Kris Hauser. Artificial intelligence framework for simulating clinical decision-making: A markov decision process approach. *Artificial intelligence in medicine*, 57(1):9–19, 2013.

[77] Guenter Tusch. Optimal sequential decisions in liver transplantation based on a pomdp model. In *ECAI*, pages 186–190, 2000.

[78] Lisa M Maillart, Julie Simmons Ivy, Scott Ransom, and Kathleen Diehl. Assessing dynamic breast cancer screening policies. *Operations Research*, 56(6):1411–1427, 2008.

[79] Milos Hauskrecht and Hamish Fraser. Planning treatment of ischemic heart disease with partially observable markov decision processes. *Artificial intelligence in medicine*, 18(3):221–244, 2000.

[80] Brian D Ziebart, Andrew L Maas, J Andrew Bagnell, Anind K Dey, et al. Maximum entropy inverse reinforcement learning. In *Aaai*, volume 8, pages 1433–1438. Chicago, IL, USA, 2008.

[81] Stefan Klein, Josien PW Pluim, Marius Staring, and Max A Viergever. Adaptive stochastic gradient descent optimisation for image registration. *International journal of computer vision*, 81:227–239, 2009.

[82] National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *New England Journal of Medicine*, 365(5):395–409, 2011.

[83] Christopher M Bishop. *Pattern recognition and machine learning*. Springer, 2006.

[84] Leslie Pack Kaelbling, Michael L Littman, and Anthony R Cassandra. Planning and acting in partially observable stochastic domains. *Artificial intelligence*, 101(1-2):99–134, 1998.

[85] Aurélien Garivier and Eric Moulines. On upper-confidence bound policies for switching bandit problems. In *Algorithmic Learning Theory: 22nd International Conference, ALT 2011, Espoo, Finland, October 5-7, 2011. Proceedings 22*, pages 174–188. Springer, 2011.

[86] Omar Besbes, Yonatan Gur, and Assaf Zeevi. Stochastic multi-armed-bandit problem with non-stationary rewards. *Advances in neural information processing systems*, 27, 2014.

[87] Robin Allesiardo, Raphaël Féraud, and Odalric-Ambrym Maillard. The non-stationary stochastic multi-armed bandit problem. *International Journal of Data Science and Analytics*, 3:267–283, 2017.

[88] Søren Holbech Nielsen and Thomas D Nielsen. Adapting bayes network structures to non-stationary domains. *International Journal of Approximate Reasoning*, 49(2):379–397, 2008.

[89] Joshua Robinson and Alexander Hartemink. Non-stationary dynamic bayesian networks. *Advances in neural information processing systems*, 21, 2008.

[90] Le Song, Mladen Kolar, and Eric Xing. Time-varying dynamic bayesian networks. *Advances in neural information processing systems*, 22, 2009.

[91] Joshua W Robinson, Alexander J Hartemink, and Zoubin Ghahramani. Learning non-stationary dynamic bayesian networks. *Journal of Machine Learning Research*, 11(12), 2010.

[92] Christophe Gonzales, Séverine Dubuisson, and Cristina E Manfredotti. A new algorithm for learning non-stationary dynamic bayesian networks with application to event detection. In *FLAIRS Conference*, pages 564–569, 2015.

[93] Martin C Tammemägi, Hormuzd A Katki, William G Hocking, Timothy R Church, Neil Caporaso, Paul A Kvale, Anil K Chaturvedi, Gerard A Silvestri, Tom L Riley, John Commins, et al. Selection criteria for lung-cancer screening. *New England Journal of Medicine*, 368(8):728–736, 2013.

[94] Katherine R Tuttle, Radica Z Alicic, O Kenrik Duru, Cami R Jones, Kenn B Daratha, Susanne B Nicholas, Sterling M McPherson, Joshua J Neumiller, Douglas S Bell, Carol M Mangione, et al. Clinical characteristics of and risk factors for chronic kidney disease among adults and children: an analysis of the cure-ckd registry. *JAMA network open*, 2(12):e1918169–e1918169, 2019.

[95] Lydia Chelala, Rydhwana Hossain, Ella A Kazerooni, Jared D Christensen, Debra S Dyer, and Charles S White. Lung-rads version 1.1: challenges and a look ahead, from the ajr special series on radiology reporting and data systems. *American Journal of Roentgenology*, 216(6):1411–1422, 2021.

[96] Paul F Pinsky, David S Gierada, William Black, Reginald Munden, Hrudaya Nath, Denise Aberle, and Ella Kazerooni. Performance of lung-rads in the national lung screening trial: a retrospective assessment. *Annals of internal medicine*, 162(7):485–491, 2015.

[97] Panayiotis Petousis, Simon X Han, Denise Aberle, and Alex AT Bui. Prediction of lung cancer incidence on the low-dose computed tomography arm of the national lung

screening trial: A dynamic bayesian network. *Artificial intelligence in medicine*, 72:42–55, 2016.

[98] Annette McWilliams, Martin C Tammemagi, John R Mayo, Heidi Roberts, Geoffrey Liu, Kam Soghrati, Kazuhiro Yasufuku, Simon Martel, Francis Laberge, Michel Gingras, et al. Probability of cancer in pulmonary nodules detected on first screening ct. *New England Journal of Medicine*, 369(10):910–919, 2013.

[99] Audrey Winter, Denise R Aberle, and William Hsu. External validation and recalibration of the brock model to predict probability of cancer in pulmonary nodules using nlst data. *Thorax*, 74(6):551–563, 2019.

[100] JA Cole and JDF Sherriff. Some single-and multi-site models of rainfall within discrete time increments. *Journal of Hydrology*, 17(1-2):97–113, 1972.

[101] P Todorovic and David A Woolhiser. A stochastic model of n-day precipitation. *Journal of Applied Meteorology (1962-1982)*, pages 17–24, 1975.

[102] Richard W Katz. Precipitation as a chain-dependent process. *Journal of Applied Meteorology (1962-1982)*, pages 671–676, 1977.

[103] Partha Deb and Edward C Norton. Modeling health care expenditures and use. *Annual review of public health*, 39:489–505, 2018.

[104] Murray Campbell, A Joseph Hoane Jr, and Feng-hsiung Hsu. Deep blue. *Artificial intelligence*, 134(1-2):57–83, 2002.

[105] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016.

[106] Rémi Coulom. Efficient selectivity and backup operators in monte-carlo tree search. In *Computers and Games: 5th International Conference, CG 2006, Turin, Italy, May 29-31, 2006. Revised Papers 5*, pages 72–83. Springer, 2007.

[107] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *nature*, 550(7676):354–359, 2017.

[108] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419):1140–1144, 2018.

[109] Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, et al. Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588(7839):604–609, 2020.

[110] Geoffrey E Hinton. Connectionist learning procedures. In *Machine learning*, pages 555–610. Elsevier, 1990.

[111] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feed-forward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.

[112] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[113] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[114] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

[115] Guillaume Lemaître, Fernando Nogueira, and Christos K Aridas. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *The Journal of Machine Learning Research*, 18(1):559–563, 2017.

[116] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[117] Alberto Fernández, Salvador García, Mikel Galar, Ronaldo C Prati, Bartosz Krawczyk, and Francisco Herrera. *Learning from imbalanced data sets*, volume 10. Springer, 2018.

[118] Chao Chen, Andy Liaw, Leo Breiman, et al. Using random forest to learn imbalanced data. *University of California, Berkeley*, 110(1-12):24, 2004.

[119] Xu-Ying Liu, Jianxin Wu, and Zhi-Hua Zhou. Exploratory undersampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(2):539–550, 2008.

[120] Leo Breiman. Random forests. *Machine learning*, 45:5–32, 2001.

[121] John Platt et al. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3):61–74, 1999.

[122] William J Youden. Index for rating diagnostic tests. *Cancer*, 3(1):32–35, 1950.

[123] Nancy Chinchor and Beth M Sundheim. Muc-5 evaluation metrics. In *Fifth Message Understanding Conference (MUC-5): Proceedings of a Conference Held in Baltimore, Maryland, August 25-27, 1993*, 1993.

[124] Martin C Tammemaegi, Timothy R Church, William G Hocking, Gerard A Silvestri, Paul A Kvale, Thomas L Riley, John Commins, and Christine D Berg. Evaluation of the lung cancer risks at which to screen ever-and never-smokers: screening rules applied to the plco and nlst cohorts. *PLoS medicine*, 11(12):e1001764, 2014.

[125] Philipp Probst, Anne-Laure Boulesteix, and Bernd Bischl. Tunability: Importance of hyperparameters of machine learning algorithms. *The Journal of Machine Learning Research*, 20(1):1934–1965, 2019.

[126] Sana Tonekaboni, Shalmali Joshi, Melissa D McCradden, and Anna Goldenberg. What clinicians want: contextualizing explainable machine learning for clinical end use. In *Machine learning for healthcare conference*, pages 359–380. PMLR, 2019.