Appendix

Proof of the Belief Update Equations (Eq.(9))

We establish the notions of the belief that the state is positive π_t^{P}, negative π_t^{N} and evolving π_t^{Ev} at time t as follows:

$$\begin{aligned}\pi_t^{\mathrm{P}} &:= \Pr(s_t = \mathrm{P} \mid \mathcal{F}_t),\\ \pi_t^{\mathrm{N}} &:= \Pr(s_t = \mathrm{N} \mid \mathcal{F}_t),\\ \pi_t^{\mathrm{Ev}} &:= \Pr(s_t = \mathrm{Ev} \mid \mathcal{F}_t),\end{aligned} \tag{A.1}$$

where $\mathcal{F}_t$ is the σ-algebra that contains all past observations until time t and actions until time $t-1$: $\mathcal{F}_t = \{z_1, k_1, a_1, z_2, \cdots, z_{t-1}, k_{t-1}, a_{t-1}, z_t, k_t\}$.

At the beginning of the next time index $t+1$ and before receiving observation z_{t+1}, the posterior changes to $\hat{\pi}_{t+1}^{\mathrm{P}}$ based on the transition matrix is:

$$\begin{aligned}\hat{\pi}_{t+1}^{\mathrm{P}} &:= \Pr(s_{t+1} = \mathrm{P} \mid \mathcal{F}_t)\\ &= \sum_{s\in\{\mathrm{P,N,Ev}\}} \Pr(s_{t+1} = \mathrm{P} | s_t = s, \mathcal{F}_t)\Pr(s_t = s|\mathcal{F}_t)\\ &= \sum_{s\in\{\mathrm{P,N,Ev}\}} \Pr(s_{t+1} = \mathrm{P} | s_t = s)\Pr(s_t = s|\mathcal{F}_t)\\ &= \pi_t^{\mathrm{P}} + \lambda_3 \pi_t^{\mathrm{Ev}},\end{aligned} \tag{A.2}$$

which follows the law of total probability. And similarly, we have:

$$\hat{\pi}_{t+1}^{\mathrm{N}} = \pi_t^{\mathrm{N}} + (1-\lambda_2-\lambda_3)\pi_t^{\mathrm{Ev}}, \qquad \hat{\pi}_{t+1}^{\mathrm{Ev}} = \lambda_2 \pi_t^{\mathrm{Ev}}. \tag{A.3}$$

As a double-check, we have the following:

$$\hat{\pi}_{t+1}^{\mathrm{P}} + \hat{\pi}_{t+1}^{\mathrm{N}} + \hat{\pi}_{t+1}^{\mathrm{Ev}} = \pi_t^{\mathrm{P}} + \lambda_3\pi_t^{\mathrm{Ev}} + \pi_t^{\mathrm{N}} + (1-\lambda_2-\lambda_3)\pi_t^{\mathrm{Ev}} + \lambda_3\pi_t^{\mathrm{Ev}} = \pi_t^{\mathrm{N}} + \pi_t^{\mathrm{P}} + \pi_t^{\mathrm{Ev}} = 1. \tag{A.4}$$

Based on the posterior, we have the belief update equations (Eq.(9)) as follows based on the Hidden Markov Model filter (Krishnamurthy 2016):

$$\begin{aligned}\pi_{t+1,a_t}^{\mathrm{P}} &= \Pr(s_{t+1} = \mathrm{P} \mid \mathcal{F}_{t+1}, a_t)\\ &= \frac{\Pr(s_{t+1} = \mathrm{P} \mid \mathcal{F}_t)\Pr(z_{t+1} \mid s_{t+1} = \mathrm{P}, \mathcal{F}_t, a_t)}{\Pr(z_{t+1} \mid \mathcal{F}_t, a_t)}\\ &= \frac{\hat{\pi}_{t+1}^{\mathrm{P}}\gamma_{a_t}(z_{t+1} \mid z_t)}{\Pr(z_{t+1} \mid \mathcal{F}_t, a_t)}\\ &= \frac{\hat{\pi}_{t+1}^{\mathrm{P}}\gamma_{a_t}(z_{t+1} \mid z_t)}{\sum_{s\in\{\mathrm{P,N,Ev}\}}\Pr(s_{t+1} = s \mid \mathcal{F}_t)\Pr(z_{t+1} \mid s_{t+1} = s, \mathcal{F}_t, a_t)}\\ &= \frac{\hat{\pi}_{t+1}^{\mathrm{P}}\gamma_{a_t}(z_{t+1}|z_t)}{\hat{\pi}_{t+1}^{\mathrm{P}}\gamma_{a_t}(z_{t+1} \mid z_t) + \hat{\pi}_{t+1}^{\mathrm{N}}\beta_{a_t}(z_{t+1} \mid z_t) + \hat{\pi}_{t+1}^{\mathrm{Ev}}\alpha_{a_t}(z_{t+1} \mid z_t)}\\ &:= B_{a_t}^{\mathrm{P}}(\pi_t^{\mathrm{P}}, \pi_t^{\mathrm{N}}, z_{t+1}, z_t),\end{aligned} \tag{A.5}$$

where $\gamma_{a_t}, \beta_{a_t}, \gamma_{a_t}$ is defined in the observation model.

And similarly, we have

$$\begin{aligned}\pi_{t+1,a_t}^{\mathrm{N}} &= \frac{\hat{\pi}_{t+1}^{\mathrm{N}}\beta_{a_t}(z_{t+1} \mid z_t)}{\hat{\pi}_{t+1}^{\mathrm{P}}\gamma_{a_t}(z_{t+1} \mid z_t) + \hat{\pi}_{t+1}^{\mathrm{N}}\beta_{a_t}(z_{t+1} \mid z_t) + \hat{\pi}_{t+1}^{\mathrm{Ev}}\alpha_{a_t}(z_{t+1} \mid z_t)}\\ &:= B_{a_t}^{\mathrm{N}}(\pi_t^{\mathrm{P}}, \pi_t^{\mathrm{N}}, z_{t+1}, z_t).\end{aligned} \tag{A.6}$$

Proof of Theorem 1

Recall the POMDP formulation's cost objective is:

$$\mathbb{E}_{\{s_0,z_0,s_1,z_1,\cdots\}}\Big[C_{\mathrm{D}}(\delta_T) + \sum_{t=1}^{T-1} C_{\mathrm{I}}(z_t, a_t, k_t)\Big] \overset{(a)}{=} \mathbb{E}_T\Big[\mathbb{E}_{\{s_0,z_0,s_1,z_1,\cdots\}}\Big[C_{\mathrm{D}}(\delta_T) + \sum_{t=1}^{T-1} C_{\mathrm{I}}(z_t, a_t, k_t) \mid T\Big]\Big], \tag{A.7}$$

where T is a stopping time and (a) follows the smoothing property of conditional expectations.

For a fix T and we first analyze the misdiagnosis cost $C_{\mathrm{D}}(\delta_T)$. Here we simplify $\mathbb{E}_{\{s_0,z_0,s_1,z_1,\cdots\}}[\cdot]$ as $\mathbb{E}[\cdot]$

$$\begin{aligned}\mathbb{E}\Big[C_{\mathrm{D}}(\delta_T)\Big] &\overset{(a)}{=} \mathbb{E}_{\mathcal{F}_T}\Big[\mathbb{E}\Big[C_{\mathrm{D}}(\delta_T) \mid \mathcal{F}_T\Big]\Big] \\ &= \mathbb{E}_{\mathcal{F}_T}\Big[\mathbb{E}\Big[c_{\mathrm{d}1}\mathbb{1}_{\{\delta_T=\mathrm{N},s_T=\mathrm{P}\}} + c_{\mathrm{d}2}\mathbb{1}_{\{\delta_T=\mathrm{N},s_T=\mathrm{Ev}\}} + c_{\mathrm{d}3}\mathbb{1}_{\{\delta_T=\mathrm{P},s_T=\mathrm{N}\}} + c_{\mathrm{d}4}\mathbb{1}_{\{\delta_T=\mathrm{P},s_T=\mathrm{Ev}\}} \mid \mathcal{F}_T\Big]\Big] \\ &\overset{(b)}{=} \mathbb{E}\Big[c_{\mathrm{d}1}\mathbb{1}_{\{\delta_T=\mathrm{N}\}}\pi_T^{\mathrm{P}} + c_{\mathrm{d}2}\mathbb{1}_{\{\delta_T=\mathrm{N}\}}\pi_T^{\mathrm{Ev}} + c_{\mathrm{d}3}\mathbb{1}_{\{\delta_T=\mathrm{P}\}}\pi_T^{\mathrm{N}} + c_{\mathrm{d}4}\mathbb{1}_{\{\delta_T=\mathrm{P}\}}\pi_T^{\mathrm{Ev}}\Big],\end{aligned} \tag{A.8}$$

where (a) follows the smoothing property of conditional expectations and (b) follows the definition of the beliefs (A.1).

For the delay-diagnosis cost $C_{\mathrm{I}}(z_t, a_t, k_t)$:

$$\begin{aligned}\mathbb{E}\Big[\sum_{t=1}^{T-1} C_{\mathrm{I}}(z_t,a_t,k_t)\Big] &= \mathbb{E}_{\mathcal{F}_T}\Big[\mathbb{E}\Big[\sum_{t=1}^{T-1} C_{\mathrm{I}}(z_t,a_t,k_t) \mid \mathcal{F}_T\Big]\Big] \\ &= \mathbb{E}_{\mathcal{F}_T}\Big[\mathbb{E}\Big[\sum_{t=1}^{T-1} c_{a_t} + c_{\mathrm{m}}\lambda(z_t,k_t)\mathbb{1}_{\{s_{\mathrm{F}}=\mathrm{P}\}} \mid \mathcal{F}_T\Big]\Big] \\ &= \mathbb{E}\Big[\sum_{t=1}^{T-1} c_{a_t} + c_{\mathrm{m}}\lambda(z_t,k_t)\pi_T^{\mathrm{P}}\Big].\end{aligned} \tag{A.9}$$

Moreover, $\pi_t^{\mathrm{N}}, \pi_t^{\mathrm{P}}, z_t, k_t$ will be a sufficient statistic for $\mathcal{F}_T$ (Krishnamurthy 2016). Therefore, with the new state defined as $\theta_t = (\pi_t^{\mathrm{N}}, \pi_t^{\mathrm{P}}, z_t, k_t)$ in Theorem 1, the POMDP is converted to a continuous state fully observable MDP problem.

Proof of Theorem 2

Following the proof of Theorem 1, and combining Eq. (A.9) and Eq. (A.8), as well as $\pi_T^{\mathrm{P}} + \pi_T^{\mathrm{N}} + \pi_T^{\mathrm{Ev}} = 1$, we will get our Belief MDP formulation's cost objective as in Theorem 2:

$$\inf_{T,\delta_T\in\mathcal{A}_{\mathrm{D}},a_t\in\mathcal{A}_{\mathrm{C}}} \mathbb{E}\left[C_{\mathrm{D}}(\theta_T,\delta_T) + \sum_{t=1}^{T-1} C_{\mathrm{I}}(\theta_t,a_t)\right]. \tag{A.10}$$

We have for all T:

$$\begin{aligned}\inf_{\delta_T\in\mathcal{A}_{\mathrm{D}}} \mathbb{E}\Big[C_{\mathrm{D}}(\delta_T)\Big] &= \inf_{\delta_T\in\{\mathrm{P},\mathrm{N}\}} \mathbb{E}\Big[c_{\mathrm{d}1}\mathbb{1}_{\{\delta_T=\mathrm{N}\}}\pi_T^{\mathrm{P}} + c_{\mathrm{d}2}\mathbb{1}_{\{\delta_T=\mathrm{N}\}}\pi_T^{\mathrm{Ev}} + c_{\mathrm{d}3}\mathbb{1}_{\{\delta_T=\mathrm{P}\}}\pi_T^{\mathrm{N}} + c_{\mathrm{d}4}\mathbb{1}_{\{\delta_T=\mathrm{P}\}}\pi_T^{\mathrm{Ev}}\Big] \\ &= \mathbb{E}\left[\inf\Big\{(c_{\mathrm{d}1}-c_{\mathrm{d}2})\pi_T^{\mathrm{P}} + c_{\mathrm{d}2}\big(1-\pi_T^{\mathrm{N}}\big), (c_{\mathrm{d}3}-c_{\mathrm{d}4})\pi_T^{\mathrm{N}} + c_{\mathrm{d}4}\big(1-\pi_T^{\mathrm{P}}\big)\Big\}\right].\end{aligned} \tag{A.11}$$

To simplify, we define:

$$g(\pi_T^{\mathrm{P}},\pi_T^{\mathrm{N}}) = \inf\Big\{(c_{\mathrm{d}1}-c_{\mathrm{d}2})\pi_T^{\mathrm{P}} + c_{\mathrm{d}2}\big(1-\pi_T^{\mathrm{N}}\big), (c_{\mathrm{d}3}-c_{\mathrm{d}4})\pi_T^{\mathrm{N}} + c_{\mathrm{d}4}\big(1-\pi_T^{\mathrm{P}}\big)\Big\}. \tag{A.12}$$

Replacing Eq. (A.11) into Eq. (A.10), we have:

$$\inf_{T,\delta_T\in\mathcal{A}_{\mathrm{D}},a_t\in\mathcal{A}_{\mathrm{C}}} \mathbb{E}\left[C_{\mathrm{D}}(\theta_T,\delta_T) + \sum_{t=1}^{T-1} C_{\mathrm{I}}(\theta_t,a_t)\right] = \inf_{T,a_t\in\mathcal{A}_{\mathrm{C}}} \mathbb{E}\left[g(\pi_T^{\mathrm{P}},\pi_T^{\mathrm{N}}) + \sum_{t=1}^{T-1} C_{\mathrm{I}}(\theta_t,a_t)\right], \tag{A.13}$$

which has the same formulation as Snell envelop in the optimal stopping problem setting.

Proof of Theorem 3

The Concavity of the Value Function and the State Action Function

To begin, we will first establish the convexity of the state action function $Q(\theta, a)$ through an induction process. Subsequently, proving the convexity of the value function $V(\theta)$ will become a simpler task. Throughout this section, we will employ the superscript ^{i} to indicate the i-th iteration of the standard value iteration process, and $'$ will denote the characteristic of the

next time step. The proof process has been divided into three steps, with the first two steps focusing on demonstrating some preliminary equations. These initial steps aim to simplify the main proof presented in **Step 3**.

By the contraction property of the standard value iteration process, we initialize $V^1(\pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z, k) = 0$ and $Q^1((\pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z, k), a) = 0$ for all $\pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z, k$, and a. Consequently, they are both concave. Suppose at iteration $i-1$, $Q^{i-1}((\pi^{\mathrm{N}}, \pi^{\mathrm{P}}, z, k), a)$ and $V^{i-1}(\pi^{\mathrm{N}}, \pi^{\mathrm{P}}, z, k)$ are all concave. Consider $\pi_1, \pi_2 \in \Pi(X)$ where $\pi_1 = [\pi_1^{\mathrm{P}}, \pi_1^{\mathrm{Ev}}, \pi_1^{\mathrm{N}}]$, $\pi_2 = [\pi_2^{\mathrm{P}}, \pi_2^{\mathrm{Ev}}, \pi_2^{\mathrm{N}}]$ and define $\pi_3 = \rho\pi_1 + (1-\rho)\pi_2$ where $\rho \in [0,1]$. For a fixed z, k, a and at iteration i:

Step 1 We show that $\Pr(z' \mid \pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z, a)$ is affine with respect to π.

$$\begin{aligned}\Pr(z' \mid \pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z, a) &= \sum_{s'} \Pr(z'|z, a, s')\Pr(s' \mid \pi^{\mathrm{P}}, \pi^{\mathrm{N}}) \\ &\overset{(a)}{=} \alpha_a(1 - \hat{\pi}^{\mathrm{P}} - \hat{\pi}^{\mathrm{N}}) + \beta_a\hat{\pi}^{\mathrm{N}} + \gamma_a\hat{\pi}^{\mathrm{P}} \\ &\overset{(b)}{=} (\alpha_a\lambda_2 + \beta_a\lambda_1 + \gamma_a\lambda_3)(1 - \pi^{\mathrm{N}} - \pi^{\mathrm{P}}) + \beta_a\pi^{\mathrm{N}} + \gamma_a\pi^{\mathrm{P}},\end{aligned} \tag{A.14}$$

where (a,b) comes from definition Eq. (A.2) and α_a, β_a, γ_a is as defined in the observation model. Here, we omit the input $(z' \mid z)$ of functions $\alpha_a(z' \mid z)$, $\beta_a(z' \mid z)$, $\gamma_a(z' \mid z)$ when there is no obscure. We can easily check that:

$$\Pr(z' \mid \pi_3^{\mathrm{P}}, \pi_3^{\mathrm{N}}, z, a) = \rho\Pr(z' \mid \pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z, a) + (1-\rho)\Pr(z' \mid \pi_2^{\mathrm{P}}, \pi_2^{\mathrm{N}}, z, a). \tag{A.15}$$

Step 2 We define:

$$\eta = \frac{\rho\Pr(z' \mid \pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z, a)}{\Pr(z' \mid \pi_3^{\mathrm{P}}, \pi_3^{\mathrm{N}}, z, a)} = \frac{\rho\Pr(z' \mid \pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z, a)}{\rho\Pr(z' \mid \pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z, a) + (1-\rho)\Pr(z' \mid \pi_2^{\mathrm{P}}, \pi_2^{\mathrm{N}}, z, a)}. \tag{A.16}$$

Note that for simplicity, we omit the input of function η, α_a, β_a, γ_a when there is no obscure.

In the following, we verify that:

$$\begin{aligned}&\eta B_a^{\mathrm{P}}(\pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z', z) + (1-\eta)B_a^{\mathrm{P}}(\pi_2^{\mathrm{P}}, \pi_2^{\mathrm{N}}, z', z) \\ &= \frac{\rho\Pr(z' \mid \pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z, a)B_a^{P}(\pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z', z)}{\rho\Pr(z' \mid \pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z, a) + (1-\rho)\Pr(z' \mid \pi_2^{\mathrm{P}}, \pi_2^{\mathrm{N}}, z, a)} + \frac{(1-\rho)\Pr(z' \mid \pi_2^{\mathrm{P}}, \pi_2^{\mathrm{N}}, z, a)B_a^{\mathrm{P}}(\pi_2^{\mathrm{P}}, \pi_2^{\mathrm{N}}, z', z)}{\rho\Pr(z' \mid \pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z, a) + (1-\rho)\Pr(z' \mid \pi_2^{\mathrm{P}}, \pi_2^{\mathrm{N}}, z, a)} \\ &= \frac{\rho\Pr(z' \mid \pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z, a)B_a^{\mathrm{P}}(\pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z', z) + (1-\rho)\Pr(z' \mid \pi_2^{\mathrm{P}}, \pi_2^{\mathrm{N}}, z, a)B_a^{\mathrm{P}}(\pi_2^{\mathrm{P}}, \pi_2^{\mathrm{N}}, z', z)}{\rho\Pr(z' \mid \pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z, a) + (1-\rho)\Pr(z' \mid \pi_2^{\mathrm{P}}, \pi_2^{\mathrm{N}}, z, a)} \\ &\overset{(a)}{=} \frac{\rho\gamma_a\hat{\pi}_1^{\mathrm{P}} + (1-\rho)\gamma_a\hat{\pi}_2^{\mathrm{P}}}{\Pr(z' \mid \pi_3^{\mathrm{P}}, \pi_3^{\mathrm{N}}, z, a)} \\ &= \frac{\rho\gamma_a(z' \mid z)(\pi_1^{\mathrm{P}} + \lambda_3(1 - \pi_1^{\mathrm{P}} - \pi_1^{\mathrm{N}})) + (1-\rho)\gamma_a(z' \mid z)(\pi_2^{\mathrm{P}} + \lambda_3(1 - \pi_2^{\mathrm{P}} - \pi_2^{\mathrm{N}}))}{\Pr(z' \mid \pi_3^{\mathrm{P}}, \pi_3^{\mathrm{N}}, z, a)} \\ &= \frac{\gamma_a\hat{\pi}_3^{\mathrm{P}}}{\Pr(z' \mid \pi_3^{\mathrm{P}}, \pi_3^{\mathrm{N}}, z, a)} \\ &\overset{(a)}{=} B_a^{\mathrm{P}}(\pi_3^{\mathrm{P}}, \pi_3^{\mathrm{N}}, z', z),\end{aligned} \tag{A.17}$$

where $B_a^{\mathrm{P}}(\pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z', z)$ as defined in (A.6) and (a) based on Eq. (A.14):

$$\begin{aligned}B_a^{\mathrm{P}}(\pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z', z) &= \frac{\hat{\pi}^{\mathrm{P}}\gamma_a(z' \mid z)}{\hat{\pi}^{\mathrm{P}}\gamma_a(z' \mid z) + \hat{\pi}^{\mathrm{N}}\beta_a(z' \mid z) + \hat{\pi}^{\mathrm{Ev}}\alpha_a(z' \mid z)} \\ &= \frac{\hat{\pi}^{\mathrm{P}}\gamma_a(z' \mid z)}{\Pr(z' \mid \pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z, a)}.\end{aligned} \tag{A.18}$$

In the same way, we also have:

$$\begin{aligned}\eta B_a^{\mathrm{N}}(\pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z', z) + (1-\eta)B_a^{\mathrm{N}}(\pi_2^{\mathrm{P}}, \pi_2^{\mathrm{N}}, z', z) &= \frac{\rho\beta_a\hat{\pi}_1^{\mathrm{N}} + (1-\rho)\beta_a\hat{\pi}_2^{\mathrm{N}}}{\Pr(z' \mid \pi_3^{\mathrm{P}}, \pi_3^{\mathrm{N}}, z, a)} \\ &\overset{(a)}{=} B_a^{\mathrm{N}}(\pi_3^{\mathrm{P}}, \pi_3^{\mathrm{N}}, z', z),\end{aligned} \tag{A.19}$$

where $B_a^{\mathrm{N}}(\pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z', z)$ as defined in (A.7) and (a) comes from:

$$\begin{aligned} B_a^{\mathrm{N}}(\pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z', z) &= \frac{\hat{\pi}^{\mathrm{N}}\beta_a(z' \mid z)}{\hat{\pi}^{\mathrm{P}}\gamma_a(z' \mid z) + \hat{\pi}^{\mathrm{N}}\beta_a(z' \mid z) + \hat{\pi}^{\mathrm{Ev}}\alpha_a(z' \mid z)} \\ &= \frac{\hat{\pi}^{\mathrm{N}}\gamma_a(z' \mid z)}{\Pr(z' \mid \pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z, a)}. \end{aligned} \tag{A.20}$$

Step 3 For fixed observation z, index k, and iteration i, if $a \in \mathcal{A}_{\mathrm{D}}$, $Q^i(\theta, a) = g(\pi^{\mathrm{P}}, \pi^{\mathrm{N}})$ is linear with respective to π based on it's formulation. If $a \in \mathcal{A}_{\mathrm{C}}$, then according to the Bellman equation, the standard value iteration process can be written as:

$$Q^i(\theta, a) = C_{\mathrm{I}}(\theta, a) + \sum_{z',k'} \Pr(z' \mid \pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z, a) V^{i-1}\Big(B_a^{\mathrm{P}}(\pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z', z) B_a^{\mathrm{N}}(\pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z', z), z', k'\Big). \tag{A.21}$$

Since $C_{\mathrm{I}}(\theta, a)$ is affine with respect to π, therefore we only need to analyze the second term in the above equation. We define $\hat{V}^{i-1}(\theta) = \sum_{z',k'} \Pr(z' \mid \pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z, a) V^{i-1}\Big(B_a^{\mathrm{P}}(\pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z', z), B_a^{\mathrm{N}}(\pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z', z), z', k'\Big)$ with $a = \arg\min_a Q^{i-1}(\theta, a)$.

Combining the results from **step 1** and **step 2**, we have:

$$\begin{aligned} &\rho\hat{V}^{i-1}(\pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z, k) + (1-\rho)\hat{V}^{i-1}(\pi_2^{\mathrm{P}}, \pi_2^{\mathrm{N}}, z, k) \\ &= \sum_{z',k'} \rho\Big[\Pr(z'|\pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z, a) V^{i-1}\Big(B_a^{\mathrm{P}}(\pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z', z), B_a^{\mathrm{N}}(\pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z', z), z', k'\Big)\Big] \\ &\quad + \sum_{z'} (1-\rho)\Big[\Pr(z'|\pi_2^{\mathrm{P}}, \pi_2^{\mathrm{N}}, z, a) V^{i-1}\Big(B_a^{\mathrm{P}}(\pi_2^{\mathrm{P}}, \pi_2^{\mathrm{N}}, z', z), B_a^{\mathrm{N}}(\pi_2^{\mathrm{P}}, \pi_2^{\mathrm{N}}, z', z), z', k'\Big)\Big] \\ &\overset{(a)}{=} \sum_{z',k'} \Pr(z' \mid \pi_3^{\mathrm{P}}, \pi_3^{\mathrm{N}}, z, a)\Big[\eta V^{i-1}\Big(B_a^{\mathrm{P}}(\pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z', z), B_a^{\mathrm{N}}(\pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z', z), z', k'\Big) \\ &\quad + (1-\eta)V^{i-1}\Big(B_a^{\mathrm{P}}(\pi_2^{\mathrm{P}}, \pi_2^{\mathrm{N}}, z', z), B_a^{\mathrm{N}}(\pi_2^{\mathrm{P}}, \pi_2^{\mathrm{N}}, z', z), z', k'\Big)\Big] \\ &\overset{(b)}{\leq} \sum_{z',k'} \Pr(z' \mid \pi_3^{\mathrm{P}}, \pi_3^{\mathrm{N}}, z, a)\Big[V^{i-1}\Big(\big[\eta B_a^{\mathrm{P}}(\pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z', z) + (1-\eta)B_a^{\mathrm{P}}(\pi_2^{\mathrm{P}}, \pi_2^{\mathrm{N}}, z', z)\big], \\ &\quad \big[\eta B_a^{\mathrm{N}}(\pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z', z) + (1-\eta)B_a^{\mathrm{N}}(\pi_2^{\mathrm{P}}, \pi_2^{\mathrm{N}}, z', z)\big], z', k'\Big)\Big] \\ &\overset{(c)}{=} \sum_{z',k'} \Pr(z' \mid \pi_3^{\mathrm{P}}, \pi_3^{\mathrm{N}}, z, a)\Big[V^{i-1}\Big(B_a^{\mathrm{P}}(\pi_3^{\mathrm{P}}, \pi_3^{\mathrm{N}}, z', z), B_a^{\mathrm{N}}(\pi_3^{\mathrm{P}}, \pi_3^{\mathrm{N}}, z', z), z', k'\Big)\Big] \\ &= \hat{V}^{i-1}(\pi_3^{\mathrm{P}}, \pi_3^{\mathrm{N}}, z, k), \end{aligned} \tag{A.22}$$

where equality (a) is due to the definition of η in Eq. (A.16). Inequality (b) is due to the concavity assumption of $V^{i-1}(\pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z, k)$ and equality (c) is due to equation Eq. (A.17), Eq. (A.19). In our current implementation, $k' = k + 1$ due to the limitation of our dataset. We will explore more kinds of medical examinations with different $k' - k$ in future works.

Combining Eq. (A.22) with Eq. (A.21), the concavity of $Q((\pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z, k), a)$ is proved as:

$$\begin{aligned} \rho Q^i((\pi_1^{\mathrm{P}}, \pi_1^{\mathrm{N}}, z, k), a) + (1-\rho)Q^i((\pi_2^{\mathrm{P}}, \pi_2^{\mathrm{N}}, z, k), a) &\leq C_{\mathrm{I}}((\pi_3^{\mathrm{P}}, \pi_3^{\mathrm{N}}, z, k), a) + \hat{V}^{i-1}(\pi_3^{\mathrm{P}}, \pi_3^{\mathrm{N}}, z, k) \\ &= Q^i((\pi_3^{\mathrm{P}}, \pi_3^{\mathrm{N}}, z, k), a). \end{aligned} \tag{A.23}$$

Since the $V^i(\pi^P, \pi^N, z, k)$ is the minimum of concave functions, itself will also be concave. Let $i \to \infty$ which proved the concavity of both $Q((\pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z, k), a)$ and $V(\pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z, k)$.

The Concavity of the Stopping Region

We now present another structure result: the stopping region's convexity. Recall that we define set $\mathcal{R}_{\mathrm{P}} \in \Pi(X)$ as the set of belief states for which the diagnosis action $\delta_T = \mathrm{P}$ is the optimal action. The same, define $\mathcal{R}_{\mathrm{N}} \in \Pi(X)$ as the set of belief

states for which the diagnosis action $\delta_T = \mathbf{N}$ is the optimal action and $\mathcal{R}_{\mathrm{C}} \in \Pi(X)$ for belief states with continuous action (more follow-up exam) as the optimal action.

Pick $\pi_1, \pi_2 \in \mathcal{R}_{\mathrm{N}}$. For any $\rho \in [0, 1]$:

$$\begin{aligned} V\Big(\rho\pi_1 + (1-\rho)\pi_2, z, k\Big) &\overset{(a)}{\geq} \rho V(\pi_1, z, k) + (1-\rho)V(\pi_2, z, k) \\ &\overset{(b)}{=} \rho Q(\pi_1, z, k, a = \mathrm{N}) + (1-\rho)Q(\pi_2, z, k, a = \mathrm{N})) \\ &\overset{(c)}{=} Q\Big(\rho\pi_1 + (1-\rho)\pi_2, z, k, a = \mathrm{N})\Big) \\ &\overset{(d)}{\geq} V\Big(\rho\pi_1 + (1-\rho)\pi_2, z, k\Big), \end{aligned} \tag{A.24}$$

where (a) comes from the concave of function V, (b) satisfied since $\pi_1, \pi_1 \in \mathcal{R}_N$, (c) comes from the linear property of $Q(\pi, z, k, a = \mathrm{N})$ since $a \in \mathcal{A}_{\mathrm{D}}$ and (d) satisfied since V is the optimal value function. Thus all the inequalities above are qualities, which means $\rho\pi_1 + (1-\rho)\pi_2 \in \mathcal{R}_{\mathrm{N}}$. The same result can be obtained for $\pi_1, \pi_2 \in \mathcal{R}_{\mathrm{P}}$ using a similar way.

Clinical Background Appendix

Lung Nodule Dynamic Model

Tumor progression dynamics, specifically characterized by magniligant nodules, have conventionally been subjected to modeling through the utilization of ordinary differential equations. Following a comprehensive analysis involving diverse mathematical frameworks encompassing exponential growth models, logistic growth models, and the Von Bertalanffy equation (Brown, Rothery et al. 1993), a judicious selection has been made in favor of the Gompertz model (Kirkwood 2015). This selection has been predicated upon its optimal alignment with the observed growth dynamics and the associated parameter domains inherent to our specific case of lung tumor.

Let d_t denote the diameter of a nodule (in mm) at time t, then the value of the nodule equal to $\frac{d_t^3 \pi}{6}$ and the number of cells in the nodule equal to $p_t = \frac{d_t^3 \pi}{6} / (5.42 \times 10^{-13})$ (de Pillis et al. 2013).

Gompertz growth model can be represented as:

$$\frac{\partial p}{\partial t} = r log(\frac{K}{p_t}) p_t, \tag{A.25}$$

where K represents the carrying capacity, i.e. the maximum size that can be reached with the available nutrients. r is the constant related to the proliferative ability of the cells. In our experiments, $r = 0.034$ and $K = 2.84 \times 10^{21}$.

Based on the Gompertz growth model, and with $a_t \in \mathcal{A}_{\mathrm{C}}$ in our dataset represents 1-year follow-up LDCT scan, the observation model is defined as:

$$\begin{aligned} \alpha_{a_t}(p_{t+1}|p_t) &= \mathcal{N}(p_{t+1} \mid p_t, \omega_1), \\ \beta_{a_t}(p_{t+1}|p_t) &= \mathcal{N}(p_{t+1} \mid p_t, \omega_2), \\ \gamma_{a_t}(p_{t+1}|p_t) &= \mathcal{N}(p_{t+1} \mid p_t + \frac{\partial p}{\partial t}, \omega_3), \end{aligned} \tag{A.26}$$

where $\mathcal{N}(p \mid \mu, \sigma)$ represent the probability of p follows normal distribution with mean μ and standard deviation ω. During the implement, $\omega_1 = 3$ since the nodule under a evolving state, $\omega_2 = 1$ and $\omega_3 = 1$.

The Lung-RADS and the Brock model

In the experiment section, a comparative analysis is conducted between the EarlyStop-RL model and the Lung CT Screening Reporting and Data System (Lung-RADS) (McKee et al. 2016; Ardila, Kiraly et al. 2019), as well as the Brock model (McWilliams et al. 2013). The outcomes of the Lung-RADS and Brock model were computed utilizing essential risk factors, encompassing pertinent patient demographic information and radiological descriptions of nodules. These risk factors were directly extracted from the National Lung Screening Trial (NLST) dataset. More specific implementation details can be found in (Pinsky et al. 2015) for lung-RADS and (Winter, Aberle, and Hsu 2019) for the Brock model.

The Lung-RADS yields a categorical risk score spanning integers from 1 to 5, while the Brock model generates a continuous risk value ranging from 0 to 1. A higher numerical value in both models signifies an augmented susceptibility to the development of lung cancer. We classify these outcomes into distinct strata of risk for lung cancer, namely low-, medium-, and high-risk. These categories align correspondingly with negative, requiring further follow-up (continuous action), and positive for lung cancer policy in our EarlyStop-RL, respectively.

The groups of low-, medium-, and high-risk are stratified by the application of specified thresholds for Lung-RADS scores ($< 3, = 3, > 3$) that derived from prior studies (Ardila, Kiraly et al. 2019; Huang et al. 2019), and for Brock model ($<$

$0.0117, [0.0117, 0.10], > 0.10)$ that derived from the British Thoracic Society guideline flow chart (Chung et al. 2018; Baldwin et al. 2020).

Parameters for the Problem Formulation

Appendix Table A1 provides the parameter used in the Problem Formulation section.

Table A1: Hyperparameters used in Model

Parameter	Value
λ_1	0.3
λ_2	0.5
λ_3	0.2
$c_{\mathrm{d}1}$	15
$c_{\mathrm{d}2}$	4.1
$c_{\mathrm{d}3}$	12
$c_{\mathrm{d}4}$	3.1
c_a	3
c_{m}	1

And the hazard function used in the delay-diagnosis cost is defined as:

$$\begin{aligned} M &= \frac{z^3\pi}{6}, \\ N &= M + 13M \times 0.034 \times e^{-0.034}, \\ M' &= (\frac{N}{\pi/6})^{\frac{1}{3}}, \\ \lambda(z,k) &= \frac{(M'-M)\times k}{5}. \end{aligned} \tag{A.27}$$

All of the above values and equations are fundamentally predicated upon clinical expertise and opinion.

EarlyStop-RL Training Details

Computing Infrastructure and Training Process

During the implementation, we utilize a three-layer "linear-batch normalization-ReLu" neural network structure with 32 neurons in each layer to represent the Q-function, and use the double deep Q-learning framework (Van Hasselt, Guez, and Silver 2016). The algorithm is developed using Python and PyTorch. Since the NLST database is relatively small, the overall training process can be finished in about 4 hours using one GeForce RTX 2080 Ti Graphics Card. Appendix Table A2 provides the hyper-parameters used in the training process.

Threshold Boundary

In the experiment and deployment phase of our study, subsequent to the completion of the training process and the attainment of the optimal $Q(\theta, a)$ function, we employed a grid approximation technique in conjunction with a support vector machine utilizing a linear kernel, as shown in Algorithm A1, to delineate a threshold boundary within the belief space and divide the belief space into three regions ($\mathcal{R}_{\mathrm{N}}$, $\mathcal{R}_{\mathrm{C}}$, and $\mathcal{R}_{\mathrm{P}}$). We divide $\pi^{\mathrm{P}}, \pi^{\mathrm{N}}$ into grids with N as a proper number, and we use the Python package "sklearn.svm" to achieve the boundary approximation. In this manner, it obviates the necessity of recurrently executing the neural network at each stage during the implementation phase. This attribute renders it more conducive for employment in clinical settings. Fig. A1 shows an example of three regions, with $k = 1$, the diameter of the nodule is 19 mm and grid approximate proper number N = 100.

Table A2: Hyper-parameters used during training

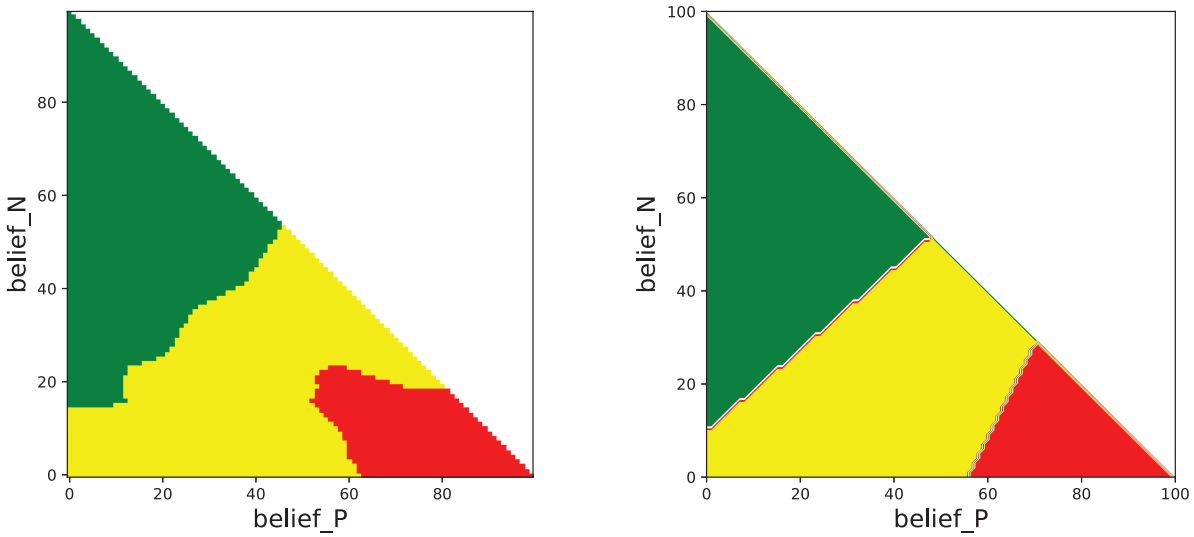
Parameter	Value
Number of layers for Q function	3
Number of neurons for each layer	$32, 32, 32$
Optimizer	Adam
Loss function during Q-learning	Smooth L1Loss
Initial learning rate	0.01
Learning rate schedule	Exponential
Multiplicative factor of learning rate decay	0.9
Batch Size	16

Algorithm A1: EarlyStop-RL (Experiment and Deployment phase)

Input: Trained $Q(\theta, a)$.

for z,k **do**
  for $\pi^{\mathrm{P}} = 0, \frac{1}{N}, \cdots, 1$ **do**
    for $\pi^{\mathrm{N}} = 0, \frac{1}{N}, \cdots, 1 - \pi^{\mathrm{P}}$ **do**
      $\theta = (\pi^{\mathrm{P}}, \pi^{\mathrm{N}}, z, k)$
      $\mathrm{C_N} = (c_{\mathrm{d1}} - c_{\mathrm{d2}})\pi_T^{\mathrm{P}} + c_{\mathrm{d2}}\big(1 - \pi_T^{\mathrm{N}}\big)$
      $\mathrm{C_P} = (c_{\mathrm{d3}} - c_{\mathrm{d4}})\pi_T^{\mathrm{N}} + c_{\mathrm{d4}}\big(1 - \pi_T^{\mathrm{P}}\big)$
      if $\min_a Q(\theta, a) \geq \inf\{\mathrm{C_N}, \mathrm{C_P}\}$ **then**
        if $\mathrm{C_N} > \mathrm{C_P}$ **then**
          $\theta \in \mathcal{R}_{\mathrm{P}}$
        else if $\mathrm{C_N} \leq \mathrm{C_P}$ **then**
          $\theta \in \mathcal{R}_{\mathrm{N}}$
        end if
      else
        $\theta \in \mathcal{R}_{\mathrm{C}}$
      end if
    end for
  end for
  Perform support vector machine with a linear kernel to get the linear boundary for $\mathcal{R}_{\mathrm{N}}$, $\mathcal{R}_{\mathrm{C}}$, and $\mathcal{R}_{\mathrm{P}}$.
end for

Output: Parameters for the threshold of regions $\mathcal{R}_{\mathrm{N}}$, $\mathcal{R}_{\mathrm{P}}$, $\mathcal{R}_{\mathrm{C}}$, and $Q(\theta, a)$ if needed.



(a) Example of three regions before the threshold approximation

(b) Example of three regions after the threshold approximation

Figure A1: Example of three regions ($\mathcal{R}_{\mathrm{N}}$: green, $\mathcal{R}_{\mathrm{C}}$: yellow, and $\mathcal{R}_{\mathrm{P}}$: red) and the threshold between them.

References

Ardila, D.; Kiraly, A. P.; et al. 2019. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature medicine*, 25(6): 954–961.

Baldwin, D. R.; Gustafson, J.; Pickup, L.; Arteta, C.; Novotny, P.; Declerck, J.; Kadir, T.; Figueiras, C.; Sterba, A.; Exell, A.; et al. 2020. External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules. *Thorax*, 75(4): 306–312.

Brown, D.; Rothery, P.; et al. 1993. *Models in biology: mathematics, statistics and computing.* John Wiley & Sons Ltd.

Chung, K.; Mets, O. M.; Gerke, P. K.; Jacobs, C.; den Harder, A. M.; Scholten, E. T.; Prokop, M.; de Jong, P. A.; van Ginneken, B.; and Schaefer-Prokop, C. M. 2018. Brock malignancy risk calculator for pulmonary nodules: validation outside a lung cancer screening population. *Thorax*, 73(9): 857–863.

de Pillis, L. G.; et al. 2013. Mathematical modeling of the regulatory T cell effects on renal cell carcinoma treatment.

Huang, P.; Lin, C. T.; Li, Y.; Tammemagi, M. C.; Brock, M. V.; Atkar-Khattra, S.; Xu, Y.; Hu, P.; Mayo, J. R.; Schmidt, H.; et al. 2019. Prediction of lung cancer risk at follow-up screening with low-dose CT: a training and validation study of a deep learning method. *The Lancet Digital Health*, 1(7): e353–e362.

Kirkwood, T. B. 2015. Deciphering death: a commentary on Gompertz (1825)'On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1666): 20140379.

Krishnamurthy, V. 2016. *Partially observed Markov decision processes*. Cambridge university press.

McKee, B. J.; Regis, S. M.; McKee, A. B.; Flacke, S.; and Wald, C. 2016. Performance of ACR Lung-RADS in a clinical CT lung screening program. *Journal of the American College of Radiology*, 13(2): R25–R29.

McWilliams, A.; Tammemagi, M. C.; Mayo, J. R.; Roberts, H.; Liu, G.; Soghrati, K.; Yasufuku, K.; Martel, S.; Laberge, F.; Gingras, M.; et al. 2013. Probability of cancer in pulmonary nodules detected on first screening CT. *New England Journal of Medicine*, 369(10): 910–919.

Pinsky, P. F.; Gierada, D. S.; Black, W.; Munden, R.; Nath, H.; Aberle, D.; and Kazerooni, E. 2015. Performance of Lung-RADS in the National Lung Screening Trial: a retrospective assessment. *Annals of internal medicine*, 162(7): 485–491.

Van Hasselt, H.; Guez, A.; and Silver, D. 2016. Deep reinforcement learning with double q-learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30.

Winter, A.; Aberle, D. R.; and Hsu, W. 2019. External validation and recalibration of the Brock model to predict probability of cancer in pulmonary nodules using NLST data. *Thorax*, 74(6): 551–563.