International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 55 No. 1 January 2026, pp. 107-120

A Stochastic Game Theory Framework For Multi-Agent Decision-Making In Clinical Healthcare Settings

UDOK, Udeme Vincent

University of Uyo Teaching Hospital, Uyo. Akwa Ibom State
Corresponding Author: Udeme Vincent. E-mail: udemevin@gmail.com

Abstract: The study focused on a stochastic game theory framework for multi-agent decision-making in clinical healthcare settings, driven by advancement in a clinical and simulation-based environment. Despite advancements in single-agent models, there remains a notable knowledge gap in incorporating multi-agent strategic interactions within stochastic frameworks that adequately addressed uncertainty. The objectives of the study were to: develop a stochastic game-theoretic framework for modeling dynamic, interactive decision-making in clinical healthcare settings, incorporate Stochastic game models based on Markov Decision Processes (MDPs) and Partially Observable Markov Decision Processes (POMDPs), and evaluate outcomes under realistic clinical constraints, including limited resources, diagnostic uncertainty, and time-sensitive interventions, using simulation-based analysis. This study addressed patient progression through distinct health states (critical, serious, stable and recovered), influenced by healthcare interventions and a simulated patient summary table. The study considered the patients as the principal agents; characterized by initial severity (Mild, Moderate, Severe) and risk (Low, Medium, High) which directly influenced their initial states and potential health progression. The study adopted simulation-based analysis framework which was implemented in Python. The computed value function, expected reward for each health state, derived via the Bellman equation. The simulated-based analysis was conducted using (4×4) transition probability matrices on table 2 and table 4 with a discount factor of 0.95. The method of data analysis was based on Markov Decision Process and Partially Observable Markov Decision Processes. Two (4×4) matrices were created from simulated-based data for transition probability on "treat" and "wait" actions where the analysis revealed that all states eventually absorbed into 'recovery with probability 1.0. Based on Markov Decision Process, the expected time to recovery from Critical was 3.25 compared to 6.0 from Serious and 7.75 from Stable. Using a reward structure penalizing critical states (0) and rewarding recovery (+8.5), the expected cumulative reward from each initial state was computed as: Critical = 3.25, Serious = 6.0, Stable = 7.75. Also, using Partially Observable Markov Decision Process on table 9 revealed the values of belief update of 20 stimulated patients which ranged approximately from 152.87 to 167.59, showing moderate variability across patient beliefs The relative closeness of patient value of belief, (v(b)) on table 9 indicated that each patient health state had the potential to recover with an average payoff of 6.8, improving recovery odds by 15-20%. However, the study recommended that the healthcare agents and systems should improve clinical decision-making under uncertainty by applying Markov Decision Processes and Partially Observable Markov Decision Processes to minimize patient times spent in critical or serious health states, delays and costs of care in order to ensure evidenced-based support and overall system performance.

Keywords: Stochastic game theory, Markov decision processes, Partially observable Markov decision processes, Multi-agent decision-making, Healthcare settings.

Introduction

A Stochastic game theory framework, also known as a Markov Game, provides a sophisticated approach to model complex, dynamic decision-making scenarios in clinical settings involving multiple interacting agents with the consequences of diverse actions. In healthcare, uncertainties may arise from the unpredictable progression of diseases, variability in patient responses to treatment, and dynamic changes in healthcare policies and resource availability over time. The framework allows for the analysis of strategic

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 55 No. 1 January 2026, pp. 107-106

interactions where each agent's decisions do not only yield immediate rewards or costs but also influence the probability of transitioning to future states of the system, typically representing the evolving health status of patients or the operational conditions of a healthcare system. Stochastic game theory provides a formal structure for analyzing such scenarios by modeling agents' interactions as probabilistic state transitions influenced by their collective strategies (Basar & Olsder, 1999). This makes it particularly suitable for modeling real-world healthcare problems such as patient flow optimization, treatment planning, and resource allocation. Partially Observable Markov Decision Processes (POMDPs) also provide a structured approach to derive optimal treatment policies, enhancing decision-making efficiency and leading to improved-patient outcomes opposed to heuristic or rule based approaches commonly used in older studies. Game theory started to gain popularity in medical decision-making, doctor–patient interactions, organ transplant management, resource planning, and training, (McFadden et al., 2012). Zhu et al. (2016) developed a framework based on evolutionary game theory that combines group phenotypic composition with ecological interactions. Their framework specifically maps Quantitative Trait Loci (QTLs) for population demographics and evolution. Archetti (2013) employed evolutionary game theory to study the collective interaction between cancer cells, analyzing the dynamics of these cells' growth factors and treatment effectiveness in reducing the cell population. Game theory has shown successful application in medical resource management and training, yielding favourable outcomes. McFadden and Tsai et al. (2012) applied game theory in complex operating room system management, resulting in positive effects on the environment and benefiting all stakeholders. Blake & Carroll (2016) proposed using game theory in medical training and practice to encourage better recognition of competing priorities and adjustment of approaches when one's preferred outcome is unlikely

Clinical healthcare settings are inherently complex systems involving a multitude of interacting agents, including patients, physicians, nurses, specialists, hospitals, payers, and policymakers. These agents make sequential decisions under conditions of significant uncertainty, where the outcomes of their choices, such as treatment effectiveness, disease progression, and resource utilization, are often probabilistic (Folland et al., 2017). Therefore, this study aims to fill this gap by developing a stochastic game theory framework for multi-agent decision-making in clinical healthcare settings, providing MDP and POMDP theoretical underpinnings and practical insights for enhancing health outcomes.

Aim and Objectives of the Study

The aim of this study is to present the application of a stochastic game theory framework for multi-agent decision-making in clinical healthcare settings**.**

The specific objectives of this research are to:

i. develop a stochastic game-theoretic framework for modeling dynamic, interactive decision-making in clinical healthcare settings.

ii. incorporate Stochastic game models based on Markov Decision Processes (MDPs) and Partially Observable Markov Processes (POMDPs) within the developed framework.

iii. evaluate outcomes under realistic clinical constraints, including limited resources, diagnostic uncertainty and time-sensitive interventions, using simulation-based analysis.

The theoretical and conceptual framework serves as the foundation upon which this study is grounded. For a comprehensive understanding of the implementation of Multi-Agent Systems (MAS) within healthcare, particularly under conditions of uncertainty, it is necessary to draw from multiple theories and conceptual underpinnings. These include Decision Theory, Game Theory (particularly Stochastic Game Theory), Complex Adaptive Systems Theory, Agent-Based Modeling, and a derived conceptual framework that guides the empirical implementation of the study. Game theory studies interactions among rational decision-makers. It is highly applicable in MAS where multiple agents, each with their own objectives, must interact, cooperate, or compete. In healthcare, game theory has been applied to model negotiations between stakeholders, resource allocation, and treatment planning (Osborne & Rubinstein, 1994).

Myerson (2004), extended classical models with incomplete information to Bayesian games, providing a foundation for decision-making under uncertainty. In healthcare, these models are used to optimize the interactions between healthcare providers and patients (or among providers) in cost-sharing, prescription behavior, or compliance monitoring (Mendonça et al., 2020). These models allow decision-makers to plan over time while considering future consequences, making them suitable for chronic disease management and hospital admission strategies. A study by Adida et al. (2018) employed a stochastic game model to analyze patient adherence and physician effort in managing coronary heart disease, considering behavioral factors, where the models highlighted how patient choices and provider interventions dynamically influence the patient's health trajectory.

Healthcare decision-making, an important process in which the best action to achieve the desired goals is chosen, largely determines the quality of care, patient safety, and the possibility of future complications, (Stubbings et al.,2012). As an essential part of the professional duties of the medical personnel, clinical decision-making consists of analysis of information, making decisions, and taking action based on those decisions to accomplish the desired objective, (Wu et al., 2016). In other words, game theory deals with mathematical models of cooperation and conflicts between rational decision-makers. Decision-making situations can be viewed as games in examining strategic behaviours and interactions (Scharpf, 1997). Stochastic models have been used to analyze vaccination behavior, quarantine strategies, and treatment adoption in epidemics. For instance, Bauch and Earn (2004) used game theory to show how individual choices in vaccination uptake can lead to suboptimal population-level immunity.

Game theory provides a powerful mathematical tool to model and analyze strategic interactions between rational decision-makers. In healthcare, it has been applied to problems such as organ allocation, insurance design and resource competition. Conversely, stochastic models, including Markov Decision Processes (MDPs), allow for the modeling of uncertainty in disease evolution but typically involve a single decision-maker and fail to capture the strategic interplay between multiple agents. The stochastic game theory framework which merges game theory with stochastic processes, offers a promising solution by enabling the analysis of multi-agent strategic decision-making in uncertain, dynamic environments. Yet, its application in clinical decision-making remains minimal (Schelling, 2010).

A core strength of stochastic game theory lies in its capacity to delineate and analyze the interplay between multiple decision-makers. Healthcare agents, including patients, doctors, nurses, and administrators, often possess distinct utility functions and information sets, leading to a blend of cooperative and competitive dynamics. Acuna et al. (2021) developed a stochastic game theory framework for multi-agent decision-making in clinical healthcare settings offers a powerful and comprehensive approach to understanding and optimizing complex interactions. By integrating the dynamic nature of patient health states, the interplay of cooperative and competitive agent behaviors, and the pervasive presence of uncertainty through MDPs. The reliance on simulation-based analysis further ensures that the insights generated are robust and relevant to the practical constraints of real-world clinical environments. Reinforcement learning, a branch of AI related to stochastic game theory, is increasingly used to personalize care pathways, particularly in intensive care and oncology (Komorowski et al., 2018). With the increasing complexity of healthcare delivery, involving multiple stakeholders such as physicians, nurses, patients, caregivers, and administrative personnel has made decision-making evolved into a multi-agent process. Multi-Agent Systems (MAS) involve autonomous entities (agents) that interact to achieve individual or collective goals, and are increasingly used to model dynamic, decentralized healthcare environments (Jennings, et al, 1998). It refers to a collection of autonomous, intelligent agents that interact with one another within an environment to achieve individual or collective goals. In healthcare, these agents may represent software systems, robots, clinical decision-support tools, or human professionals like doctors and nurses.

A model developed by klein, (1993) demonstrates how agents can be programmed to coordinate test results, drug administration, and ventilator adjustments based on continuously updated patient data. The MAS system reacts dynamically to patient deterioration and alerts human caregivers, improving response time and reducing mortality risk. Markov Perfect Equilibrium (MPE) is a refinement of the Nash equilibrium used in dynamic settings, particularly stochastic games. In MPE, players' strategies depend only on the current state of the game, not the full history of play. In healthcare, MPE can model decision-making over time where the system's state (e.g., patient condition, hospital congestion) evolves. In treating chronic diseases, decisions on medication dosages or interventions depend on the current health state of a patient.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 55 No. 1 January 2026, pp. 107-106

MPE identifies optimal treatment strategies that adapt dynamically over time, maximizing patient outcomes under probabilistic health state transitions (Hauskrecht, 2000). Multi-Agent Systems (MAS) and Game Theory are two prominent frameworks that have seen growing application in healthcare for facilitating decision-making, resource allocation, diagnostics, and treatment planning. One of the most commonly used stochastic models in healthcare is the Markov Decision Process (MDP).

MDPs are used to model decision-making in situations where outcomes are partially random and partially under the control of the decision-maker. In the context of healthcare, an MDP provides a structured approach to model healthcare decisions, where the system evolves from one state to another with certain probabilities, and each state is associated with a reward or cost (Puterman, 2005). An MDP is defined by states, actions, transition probabilities and rewards. Where, states (S) represent the different possible conditions of the system (e.g., different stages of a patient's disease or health condition). Actions (A), represent the decisions or interventions that can be taken (e.g., different treatment options or procedures). Transition Probabilities (P) represent the probabilities of moving from one state to another after taking an action. The immediate reward or cost associated with being in a particular state and taking a specific action (e.g., the cost of a treatment, the improvement in patient health). MDPs are particularly useful in chronic disease management, where a patient's condition evolves over time, and medical interventions can either improve or worsen the patient's health. For example, MDPs can model decisions in managing diabetes, where actions (such as medication or lifestyle changes) lead to transitions between health states (e.g., controlled vs. uncontrolled diabetes) with associated costs and benefits. MDPs and POMDPs can also incorporate value iteration or policy iteration algorithms to compute the optimal policy and a sequence of actions that maximizes long-term benefits or minimizes long-term costs for the patient, (Filar & Vrieze, 1997).

Partially Observable Markov Decision Processes (POMDPs) allow for situations where the agent does not directly observe the true state of the system but must infer it based on noisy or incomplete observations (Monahan, 1982). Integrating POMDP elements into a multi-agent stochastic game allows for a more realistic portrayal of information asymmetry and the need for agents to make decisions under conditions of significant ambiguity, thereby addressing the objective of incorporating uncertainty and dynamic patient health states. Healthcare operates under severe limitations, including finite resources, inherent diagnostic uncertainty, and the often time-sensitive nature of interventions. Simulation-based analysis emerges as a crucial tool for evaluating the outcomes of these frameworks under such realistic constraints (Powell, 2011).

Simulations allow for the exploration of a wide range of scenarios. For instance, discrete-event simulations can model patient flow through a hospital, incorporating stochastic processes for arrivals, service times, and transitions between departments, while also embedding the game-theoretic decision rules of various agents (e.g., administrators allocating beds, doctors prioritizing patients), (Jacobson & Hall, 2012). This allows for the evaluation of system-level outcomes such as average length of stay, readmission rates, and resource utilization under various policy configurations or agent strategies. Sensitivity analyses through simulation can also identify critical bottlenecks or leverage points within the system.

Acuna et al. (2021) developed a stochastic game theory framework for multi-agent decision-making in clinical healthcare settings that offers a powerful and comprehensive approach to understanding and optimizing complex interactions. By integrating the dynamic nature of patient health states, the interplay of cooperative and competitive agent behaviors, and the pervasive presence of uncertainty through POMDPs, allow for situations where the agent does not directly observe the true state of the system but must infer it based on noisy or incomplete observations (Monahan, 1982). Integrating POMDP elements into a multi-agent stochastic game allows for a more realistic portrayal of information asymmetry and the need for agents to make decisions under conditions of significant ambiguity, thereby addressing the objective of incorporating uncertainty and dynamic patient health states. Healthcare operates under severe limitations, including finite resources, inherent diagnostic uncertainty, and the often time-sensitive nature of interventions. Simulation-based analysis emerges as a crucial tool for evaluating the outcomes of these frameworks under such realistic constraints (Powell, 2011).

A POMDP involves the belief states and observational probabilities in belief states, instead of knowing the exact state of the system, the decision-maker has a belief about the state based on partial observations (e.g., test results, symptoms). Also, in observable probabilities the likelihood of observing certain outcomes or symptoms given the current state of the system. POMDPs are computationally challenging due to their complexity, but they provide a more accurate representation of real-world clinical decision-making, where information is often partial, and uncertainty plays a significant role, (Hauskrecht, 2000).

The reliance on simulation-based analysis further ensures that the insights generated are robust and relevant to the practical constraints of real-world clinical environments. This interdisciplinary approach promises to yield actionable strategies for improving patient outcomes, enhancing resource efficiency, and fostering more effective collaborative decision-making within the intricate landscape of healthcare. (Mendonça et al., 2020). In healthcare, stochastic game theory is particularly useful when considering situations where multiple agents, such as doctors, patients, and healthcare systems, must make decisions under uncertainty (Stubbings et al.,2012).

The stochastic game-theoretic approach is particularly apt for modeling such uncertainties, making it an invaluable tool in analyzing strategic interactions among healthcare providers, patients, policymakers, and insurers. Uncertainty in healthcare manifests in multiple dimensions such as clinical uncertainty in which variability in disease diagnosis, treatment outcomes, and side effects. Behavioral uncertainty where unpredictable patient adherence to treatment plans and physician choices. Environmental uncertainty such as pandemics, disasters, or geopolitical events affecting healthcare delivery. (Rasmussen, 2007) Players choose actions to optimize their expected utility over time, considering both immediate payoffs and future consequences, while accounting for the randomness inherent in state transitions (Altman, 1999).

Multi-agent systems have found broad applications across the healthcare domain including patient monitoring and alerts, resource allocation and scheduling, and diagnostic decision support. A model developed by klein, (1993) demonstrates how agents can be programmed to coordinate test results, drug administration, and ventilator adjustments based on continuously updated patient data. The MAS reacts dynamically to patient deterioration and alerts human caregivers, improving response time and reducing mortality risk. However, by integrating uncertainty and time into the decision-making process, stochastic games provide a more realistic and flexible approach to optimizing decisions over time. Healthcare systems are inherently multi-agent environments where multiple agents (doctors, nurses, patients, insurers, etc.) interact with one another, each with different goals, preferences, and information, (Adida et al., 2018).

Materials and Methods

This study adopts a simulation-based research design to investigate decision-making in healthcare systems using stochastic game theory and reinforcement learning. The model-developing approach focuses on constructing a framework that represents the dynamic nature of healthcare decisions, incorporating time-dependent variables. Model implementation is done in Python. The MDP is defined by the tuple (S, A, P, R, γ), where:

- **States** S: The patient health states are $S = \{0,1,2,3\}$, corresponding to Critical (0), Serious (1), Stable (2), and Recovered (3).
- **Actions** A: The decisions are $A = \{\text{Treat,Wait}\}$. "Treat" represents active intervention
- **Transition Probabilities** P: $P(s' \mid s, a)$ is the probability of transitioning from state s to state s' under action a.
- **Discount Factor** γ: A value like $\gamma = 0.95$ discounts future rewards.

Rewards R: $R(s, a)$ is the immediate reward for taking action a in state s. Reward Function ($R(s, a)$): The immediate reward received for taking action a in state s is defined by a function $R: \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$. The values were derived for 'Treat' and hypothesized for 'Wait' as: $R(s,\text{Treat}) = \{1.0,2.0,3.0,0.0\}$ for states $s \in \{\text{Crit, Ser, Stab, Rec}\}$. $R(s,\text{Wait}) = \{-5.0,0.5,4.0,0.0\}$ for states $s \in \{\text{Crit, Ser, Stab, Rec}\}$. A Bellman Equation for Policy Evaluation: The Value Function $V^{\pi}(s)$, representing the expected cumulative reward for a given policy π starting from state s, is defined by the Bellman Equation: $V^{\pi}(s) = R(s, \pi(s)) + \gamma \sum_{s' \in S} P\,(s'|s, \pi(s)) V^{\pi}(s')$

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
SCHOLAR AI Be Smart
Vol. 55 No. 1 January 2026, pp. 107-106

The MDP component models the patient progression under the assumption of full observability of the patient's true health state. Transition Probability ($P(s'|s,a)$): The probability of transitioning from state s to state s' after taking action a is defined by two 4×4 matrices: The belief state is a probability distribution over the unobservable health states, $b_t(s) = P(s|o_1, a_1, \dots, o_t, a_t)$. The belief state is a vector $\mathrm{b}_t \in \Delta(\mathrm{s})$, where $\Delta(\mathcal{S})$ is the probability simplex over the state space. In a POMDP, the policy $\pi(\mathrm{b})$ is a function of this belief state. Belief Update Equation: The belief state is updated after an action a is taken and an observation o is received, using a Bayesian update rule: $b_{t+1}(s') = \eta \cdot Z(o|s') \cdot \sum_{s \in S} P(s'|s,a) b_t(s)$ where η is a normalization constant to ensure the new belief state b_{t+1} remains a valid probability distribution.

The optimal Value Function $V^*(s)$, which provides the maximum possible expected cumulative reward starting from state s by choosing the optimal action, is defined by the Bellman Optimality Equation: $V^*(s) = \max_{a \in \mathcal{A}}(R(s,a) + \gamma \sum_{s' \in S} P(s'|s,a) V^*(s'))$

$$P(s'_t \,|s_t, a) = P(s'_t \,|s_t, a_1, a_2 \dots, a_m) = \frac{N_i}{N} \quad (1)$$

$$R(s,a) = r(s) + c(a) \quad (2)$$

$$r(s) = (10 - severity\ score \times risk\ score) \quad (3)$$

Bellman optimality equation:

$$V_{k+1}(s) = \max_{a \in \{\text{Wait,Treat}\}} [R(s,a) + \gamma \sum_{s' \in S} P(s' \mid s,a) V_k(s')] \quad (4)$$

$$R(s,a) = \sum_{s'} P(s'|s,a)\, \mathrm{r}(s,a,s') \quad (5)$$

$$V(s_t) = \sum_{s_{t+1} \in S} P(s_{t+1} \mid s_t, Treat)[r_{s_t}, s_{t+1} + V(s_{t+1})] \quad (6a)$$

$$V(s_t) = \sum_{s_{t+1} \in S} P(s_{t+1} \mid s_t, Wait)[r_{s_t}, s_{t+1} + V(s_{t+1})] \quad (6b)$$

$$V(s) = R(s, Treat) + \sum_{s' \in S} P(s' \mid s, Treat) V(s') \quad (7a)$$

$$V(s) = R(s, Wait) + \sum_{s' \in S} P(s' \mid s, Wait) V(s') \quad (7b)$$

$$V(s) = R(s, Treat) + P_{s,C} V(C) + P_{s,S} V(S) + P_{s,B} V(B) + P_{s,R} V(R) \quad (8a)$$

$$V(s) = R(s, Wait) + P_{s,C} V(C) + P_{s,S} V(S) + P_{s,B} V(B) + P_{s,R} V(R) \quad (8b)$$

$$V(C) = R(C,Treat) + PC,SV(S) + PC,BV(B) \quad (9a)$$

$$V(C) = R(C,Wait) + PC,SV(S) + PC,BV(B) \quad (9b)$$

$$V(S) = R(S,Treat) + PS,BV(B) + PS,RV(R) \quad (10a)$$

$$V(S) = R(S,Wait) + PS,BV(B) + PS,RV(R) \quad (10b)$$

$$V(B) = R(B,Treat) + PB,BV(B) + PB,RV(R) \quad (11a)$$

$$V(B) = R(B,Wait) + PB,BV(B) + PB,RV(R) \quad (11b)$$

$$V(R) = R(R,Treat) + PR,RV(R) \quad (12a)$$

$$V(R) = R(R,Wait) + PR,RV(R) \quad (12b)$$

$$\bar{b}(s') = \sum_s P_{\text{treat}}\,(s' \mid s)\, b_0(s). \quad (13a)$$

$$\bar{b}(s') = \sum_s P_{\text{wait}}\,(s' \mid s)\, b_0(s). \quad (13b)$$

$$V(b) = \max_a \sum_s b\,(s) R(s,a) \quad (14)$$

Results

This section presents and interprets the results of the stochastic game model implemented to simulated clinical decision-making involving multiple agents (patients, doctors and healthcare systems). To develop a stochastic game-theoretic framework for modeling dynamic, interactive decision-making in clinical healthcare settings, the key variables are: condition severity: mild, moderate, severe, risk level: low, medium, high and payoff where: Severity Score is grouped into Mild = 1, Moderate = 2, Severe = 3 while Risk Score is categorized into: Low = 1, Medium = 2, High = 3. Each case is assessed across two categorical variables—Severity and Risk—which are quantified into numeric scores (Severity Score and Risk Score respectively). A derived metric, payoff, serves as an index of overall prognosis or expected outcome. Interpretations are provided to guide potential actions or decisions based on these scores. The table below summarizes the outcomes of the classification model:

Table 1: Simulated Summary of Each Patient Result

ID	Severity	Risk	Severity score	Risk Score	Payoff	Interpretation
P001	Moderate	Low	2	1	8	Medium condition, low risk – manageable case
P002	Mild	Medium	1	2	8	Mild condition, some risk – likely recoverable
P003	Mild	High	1	3	7	Mild issue, high risk – may need close monitoring
P004	Moderate	Medium	2	2	6	Medium severity and risk – moderately concerning

P005	Mild	Low	1	1	9	Mild condition and low risk – very favorable
P006	Severe	Low	3	1	7	Serious condition but low risk – likely stable
P007	Severe	Low	3	1	7	Serious condition but low risk – likely stable
P008	Severe	High	3	3	1	Very severe, high risk – critical and high alert
P009	Moderate	Medium	2	2	6	Medium severity and risk – moderately concerning
P010	Mild	Low	1	1	9	Mild condition and low risk – very favorable
P011	Moderate	High	2	3	7	Medium severity and high risk – moderately concerning
P012	Mild	High	1	3	7	Mild condition, high risk – needs close monitoring
P013	Moderate	Low	2	1	8	Mild condition, high risk – manageable case
P014	Severe	High	3	3	6	Very severe and high risk – critical and high alert
P015	Severe	Low	3	1	7	Very severe but mild risk – likely stable
P016	Mild	Low	1	1	9	Mild condition and low risk – very favorable
P017	Moderate	Medium	2	2	6	Medium severity and risk – moderately concerning
P018	Mild	Medium	1	2	8	Mild condition and risk – likely recoverable
P019	Severe	Low	3	1	7	Very severe and low risk – likely stable
P020	Mild	Low	1	1	9	Mild condition and low risk – very favorable

Source: Simulation-Based Data

Table 2: Transition Matrix for "Treat''

From \ To	Critical	Serious	Stable	Recovered
Critical	0	17	3	0
Serious	0	0	15	5
Stable	0	0	5	15
Recovered	0	0	0	20

Source: Simulation-Based Data

Table 3 Summary of the Reduced Transition Matrix for ''Treat'' based on Eqn (1)

From \ To	Critical	Serious	Stable	Recovered
Critical	0	0.85	0.15	0
Serious	0	0	0.75	0.25
Stable	0	0	0.25	0.75
Recovered	0	0	0	1

Table 4 Transition Matrix for ''Wait''

From \ To	Critical	Serious	Stable	Recovered
Critical	16	2	2	0
Serious	6	10	4	0
Stable	2	4	12	2
Recovered	0	0	0	20

Source: Simulation-Based Data

Table 5 Summary of the Reduced Transition Matrix for ''Wait'' based on Eqn (3.1)

From \ To	Critical	Serious	Stable	Recovered
Critical	0.8	0.1	0.1	0
Serious	0.3	0.5	0.2	0
Stable	0.1	0.2	0.6	0.1
Recovered	0	0	0	1

Table 6 Health States Value Function for MDP

State	Value
Critical (0)	155.52
Serious (1)	161.92

State	Value
Stable (2)	165.39
Recovered (3)	168.99

Table 7 Treat and Wait Probabilities for Health State

State	Prob Wait	Prob Treat
Critical (0)	0.046	0.954
Serious (1)	0.035	0.965
Stable (2)	0.139	0.861
Recovered (3)	0.881	0.119

Table 8 Reward Function for Health State on Wait and Treat Actions $(\boldsymbol{R}(\boldsymbol{s},\boldsymbol{a}))$:

State s	R(s,Wait)	R(s,Treat)
0 (Critical)	153.50	156.53
1 (Serious)	159.62	162.93
2 (Stable)	164.57	166.39
3 (Recovered)	170.00	168.00

Table 9 Patient Value of Belief

Patient	$V(b)$
P001	164.73
P002	165.85
P003	164.30
P004	162.94
P005	167.59
P006	162.47
P007	162.47
P008	157.51
P009	162.94
P010	167.59

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 55 No. 1 January 2026, pp. 107-106

Patient	$V(b)$
P011	158.55
P012	166.48
P013	157.22
P014	162.47
P015	160.07
P016	152.87
P017	154.88
P018	166.47
P019	158.55
P020	164.32

Discussion

Table 1 provides information for 20 (P001 to P020) patients states and a derived payoff or reward associated with that state / risk profile with the following column:

ID: Unique identifier for each patient. Severity: Categorical description of the patient's condition (Mild, Moderate, Severe). Risk: Categorical description of the patient's risk level (Low, Medium, High). Severity Score: A numerical score assigned to Severity (1 for Mild, 2 for Moderate, 3 for Severe). Risk Score: A numerical score assigned to Risk (1 for Low, 2 for Medium, 3 for High). The payoff is a numerical value that is derived from the Severity Score and Risk Score.

Table 2 shows the empirical frequencies of the “treat” simulation data. This table supports the numerical validity of table 1. It confirms that the treat transition probability matrix was empirically derived from simulation-based data. Hence, the transition matrix assumes treatment shifts probability mass towards better states, while reducing the poor health states.

Table 3 displays the reduced frequencies of the stimulated data for “Treat’’ decision on table 3 by applying equation (1). The reduced data is obtained by dividing the frequency of the patient’s state over the total number of patients. This illustrates that when critical patients are treated, 85% of them improve to serious state and 15% of them become stable and none recovers immediately. 75% of serious patients become stable and 25% recovers directly. Stable patients largely transitioned to recovered (75%) confirming that treatment accelerated full recovery, recovered patients remain recovered (absorbing state). This analysis confirms that “treat” decision is highly effective

Table 4 demonstrates the empirical frequencies of the “wait” simulation counts. The high counts on the transition frequency from diagonal $(\text{critical} - \text{critical}) = 16, (\text{serious} - \text{serious}) = 10$. It implies that without treatment patients tend to persist in critical state or worsen from serious to critical states. The recovered state still acts as absorbing state. This transition matrix explains that with “wait” intervention, critical and serious state have a high chance of persistence or worsening while the stable state may slowly improve.

Table 5 displays the reduced frequencies of the stimulated data for “Wait’’ action on table 4 by applying equation (3.1b). The reduced data is obtained by dividing the frequencies of the patient’s states over the total number of patients in the simulation. Table 5 explains the reduced summary transition matrix for “wait” intervention where 80% of critical patients remained critical after "wait" action, 20% of serious patients became stable, and only 10% of stable patients recovered for wait action, while recovered patients remain in an absorbing state.

Table 6 explains the health state value function for MDP showing the values for each state (condition). For state 0 (critical state) is the lowest value, reflecting high immediate penalties, state (1): 161.92 displays a moderate value that indicates a better prospect than critical but still requiring "treat" action, state (2): 165.39 suggests natural recovery potential under with less urgency. This aligns interpretations like "likely stable" in low-risk severe cases, while recovered (3): 168.99 is the absorbing state that enhances full health with positive utilities.

Table 7 represents the optimal action probabilities derived from the policy of the MDP/POMDP model such that when both probability actions are added, it becomes unity .

For critical patients (state 0) implies that "treat" intervention in 95.4% of cases confirming that immediate intervention is optimal since "wait" decision carries a high mortality or deterioration risk or relapse. Serious (State 1) have a high probability of treatment (96.5%) consistent with the need of proactive care to prevent worsening. The recovered (state 3) describes that once recovery is achieved further treatment is usually unnecessary.

Table 8 displays the reward function for health states on wait and treat actions which gives the expected cumulative rewards for each action pair. For critical and serious states "treat" decision yields higher (e.g. 156.53 vs. 153.50) confirming that intervention improves outcomes in high-risk conditions. For stable state "treat" action remains marginally better (166.39 vs. 164.57) showing that proactive treatment still offers effective value.

Table 9 displays the patient value of belief state values of 20 stimulated patients which range approximately from 152.87 to 167.59, showing moderate variability across patients. This table demonstrates how belief-based value functions in POMDPs incorporate uncertainty and enable personalized treatment optimization, even when the true patient state is partially unknown.

Conclusion

This study has successfully developed a stochastic game theoretic framework based on Markov Decision Processes and Partially Observable Markov Decision Proceesses for multi-agent decision-making in clinical healthcare settings. The framework has featured the patient health state (critical, serious, stable and recovered) agent's action being "treat" and "wait" decision. The MDP components models the patient progression under the assumption of full observability of the patient's health state. The probability of transitioning from state s to state s' after taking action a was defined by a (4×4) matrix of $p(t)$ and $p(w)$. The immediate reward received for taking action a is defined by the value function representing the expected cumulative reward which was obtained by Bellman Optimality Equation.

Based on transition matrix in table 2, from a Critical state, a patient never stays Critical or recovers directly. They have a high probability of becoming Serious and a smaller chance of becoming Stable. This indicates effective intervention moving patients out of immediate critical danger. Also, from a Serious state, a patient either improves significantly to Stable or even recovers directly. This suggests the "Treat" action is quite effective here. Again, from a Stable state, a patient has a very high chance of recovering, or a smaller chance of remaining Stable. The recovered state is a very favorable state in a healthcare context in which "Recovered" is an absorbing state. Once a patient is recovered, stays recovered. The POMDP has modeled cases where the patient's true health state is unobservable, and decisions must be made. These cases are then estimated proving that "treat" action accelerates recovery but with some levels of uncertainty, while "wait" action allows for slower natural progression but higher deterioration risk. The study has also defined the patient health states, the decision on "treat" and "wait" actions and the three (3) states of observable outcomes (low risk, moderate risk and high risk) in order to update patient belief. The belief state is updated after an action a is taken and an observation o is received, using a Bayesian update rule. The simulation-based data has been analyzed through the lens of a POMDP where the belief value function is finally determined. Therefore, the relative closeness of the belief value function indicates that each patient health state has the potential to recover if given sufficient time and effectives treatment. The simulation-based analysis has also revealed that observations would be essential for a healthcare agent to update its belief about the patient health state, enabling robust decision making even the patient through health condition is not fully observed.

Recommendations

Based on the findings of this study, the following recommendations are made:

i. The healthcare agents and systems should improve clinical decision-making under uncertainty by integrating Markov Decision Processes (MDPs) and Partially Observable Markov Decision Processes (POMDPs) to minimize patient times spent in critical or serious health states, delays and costs of care in order to ensure evidenced-based support and overall system performance.
ii. The framework's performance should be evaluated using simulation-based analysis to demonstrate quantitative and practical-oriented results, its ability to improve patient outcomes, optimize resource utilization, enhance tangible improvements in care and overall system efficiency in order to ensure practical and measurable outcomes.
iii. Healthcare systems should invest in data-driven decision support tools for real-time insights in order to provide a significant advancement over traditional frameworks.

References

[1]. Acuna, J. A., Zayas-Castro, J., Feijoo, F., Sankaranarayanan, S., Martinez, S. R. & Martinez, D. A. (2021). The Waiting Game; How cooperation between public and private hospitals can help reduce waiting lists. Science Business Media.

[2]. Adida, E., Maille, F., & Vial, J. P. (2018). A stochastic game approach to patient-physician interaction in chronic disease management. Operations Research, 66(5), 1223–1241. https://doi.org/10.1287/opre.2017.1704

[3]. Archetti, M. (2013). Evolutionary game theory of growth factor production: implications for tumour heterogeneity and resistance to therapies. British Journal of Cancer, 109, 1056–1062.

[4]. Altman, E. (1999). Constrained Markov Decision Processes. Stochastic modelling. Chapman and Hall/CRC. https://doi.org/10.1201/9780367805669

[5]. Bauch, C. T., & Earn, D. J. (2004). Vaccination and theory of games. https://doi.org/10.1073/pnas.0403823101

[6]. Basar, T. & Olsder, G. J. (1999). Dynamic Noncooperative Game Theory. Society for Industrial and Applied Mathematics. 3, 10A

[7]. Blake, A., & Carroll, B. T. (2016). Game theory and strategy in medical training. Medical Education, 50, 1094–1106.

[8]. Filar, J. A. & Vrieze, O. J. (1997). Competitive Markov Decision Processes Springer Science & Business Media.

[9]. Folland, S., Goodman, A. C., & Stano, M. (2017). The Economics of Health and Health Care (8th ed. Routledge.

[10]. Hauskrecht, M. (2000). Value-function approximations for partially observable Markov decision processes. Journal of Artificial Intelligence Research, 13, 33–94.

[11]. Jacobson, S. H., & Hall, S. N. (2012). A survey of simulation models for hospital operations. Health Systems, 1(1), 1-17.

[12]. Jennings, N. R., Sycara, K., & Wooldridge, M. (1998). A roadmap of agent research and development. Autonomous Agents and Multi-Agent Systems, 1(1), 7–38. https://doi.org/10.1023/A:1010090405266

[13]. Klein, G. A. (1993) A recognition-primed decision (RPD) model of rapid decision making. Decision Making in Action: Models and Methods

[14]. Komorowski, M., Celi, L. A., Badawi, O., Gordon, A. C., & Faisal, A. A. (2018). The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. Nature Medicine, 24(11), 1716–1720.

[15]. McFadden, D., Kadry, B., & Souba, W. W. (2012). Game theory: Applications for surgeons and the operating room environment. Surgery, 152(5), 915–922.

[16]. Mendonça, F. V., Catalão-Lopes, M., Marinho, R. T., & Figueira, J. R. (2020). Improving medical decision-making with a management science game theory approach to liver transplantation.

[17]. Monahan, G. E. (1982). A survey of partially observable Markov decision processes: Theory, models, and algorithms. Management Science, 28(1), 1-16. Omega, 94, 102050. https://doi.org/10.1016/j.omega.2019.102050.

[18]. Myerson, R. (2004) Games with Incomplete Information Played by 'Bayesian' Players, I-III. Part I. Management Science. DOI:10.1287/mnsc.1040.0297

[19]. Osborne, M. J., & Rubinstein, A. (1994). A course in game theory. MIT Press.

[20]. Powell, W. B. (2011). Approximate Dynamic Programming: Solving the Causes of Dimensionality (2nd ed.). John Wiley & Sons.

[21]. Puterman, M. L. (2005). Markov decision processes: Discrete stochastic dynamic programming. Wiley-Interscience.

[22]. Rasmussen, E. (2007). Games and information: An introduction to game theory. (4th ed.). sBlackwell Publishing.

[23]. Scharpf, F. W. (1997). Games real actors play: Actor-centered institutionalism in policy research. Westview Press.

[24]. Schelling, T. C. (2010). Game theory: A practitioner's approach. Economics and Philosophy, 26(1), 27–46.

[25]. Stubbings, L., Chaboyer, W., & McMurray, A. (2012). Nurses' use of situation awareness in decision-making: An integrative review. Journal of Advanced Nursing, 68(7), 1443–1453. https://doi.org/10.1111/j.1365-2648.2012.05989.x

[26]. Tsai, M., McFadden, D., Kadry, B., & Souba, W. W. (2012). Game theory: Applications for surgeons and the operating room environment. Surgery, 152(5), 915–922.

[27]. Wu, M., Yang, M., Liu, L., & Ye, B. (2016). An investigation of factors influencing nurses' clinical decision-making skills. Western Journal of Nursing Research, 38, 974–991. https://doi.org/10.1177/0193945916633458

[28]. Zhu, X., Jiang, L., Ye, M., Sun, L., Gragnoli, C., & Wu, R. (2016). Integrating evolutionary game theory into mechanistic genotype–phenotype mapping. Trends in Genetics, 32, 256–268.