Frontiers in Health Informatics
ISSN-Online: 2676-7104
www.healthinformaticsjournal.com
2024; Vol 13: Issue 2(FEB)
Open Access

Applying Reinforcement Learning to Optimize Healthcare Insurance Premium Pricing

Lakshmi narasimhan srinivasagopalan
senior project manager

Cite this paper as: Lakshmi narasimhan srinivasagopalan (2024). Applying Reinforcement Learning to Optimize Healthcare Insurance Premium Pricing. *Frontiers in Health Informatics*, *13 (2) 11282-11293*

ABSTRACT
In recent years, reinforcement learning (RL) has garnered increasing attention for its applications in various domains, including finance, robotics, and healthcare. One critical area in healthcare where RL has shown potential is the optimization of insurance premium pricing. Traditional pricing models in healthcare insurance often rely on a combination of actuarial calculations and statistical methods. However, these methods can be limited by a lack of dynamic adaptability to the constantly changing nature of healthcare risks. This article explores the potential for using RL to optimize healthcare insurance premium pricing, offering a detailed review of existing models, the challenges in the domain, and the fundamentals of reinforcement learning. Additionally, it provides insights into how RL can be integrated into pricing strategies, discusses various methodologies, and highlights the potential improvements in accuracy, efficiency, and adaptability.
KEYWORDS
Reinforcement Learning, Healthcare Insurance, Premium Pricing, Actuarial Models, Markov Decision Process, Policy Optimization, Dynamic Pricing, Pricing Optimization Models.

1. INTRODUCTION

This poses many challenges the industry faces when creating accurate and fair premium estimates for people. While traditional methods of pricing suit most fields, they usually involve low flexibility with a heavy reliance on history and may result in less than perfectly consistent premium calculations[1]. These models typically include risk assessments like demographic factors, historical claims data, and lifestyle factors; however, they fail to integrate changes in health care expenses or differences in patient behavior. Price-based approaches are also incapable of adjusting to the new trends of health, effects of new medical technologies, or other unforeseen public health emergencies like pandemics. With healthcare expenditures continually on the rise and new factors being introduced, there is thus a need for newer insurance strategies to maintain competitive advantage in presenting premiums as equitable yet fair representation of a risk evaluation by an individual.

It's a category in the overall field of machine learning, a method that works well within this framework in solving these types of problems that allow for ongoing learning and instantaneous decision-making[2]. Against the conventional model that relies largely on predefined rules and fixed assumptions, the dynamic formation of pricing mechanisms allows the RL model to change or improve with time over new additions of data into the system. This approach assimilates data from diverse sets of variables, for example, a patient's health information, a trend in healthcare market, or behaviors of a policyholder, and may help better optimize premium pricing in the field of healthcare insurance. The adaptability of such a method gives insurers the leeway to continually improve their methods of pricing at the same time as improving upon the risk management process, ensuring a more nuanced experience for the policyholder.

This paper investigates the role of reinforcement learning in strengthening the ability of health insurance premium pricing to make decisions[3]. It discusses the limitations of the existing pricing models, assesses the feasibility of using reinforcement learning within pricing models, and provides an overall approach for developing reinforcement learning-based systems. This article explores the potential of RL in solving inherent complexities of healthcare insurance pricing and provides insight into how this state-of-the-art technology can make the industry even more accurate, fair, and flexible in its pricing models.

2. BACKGROUND AND LITERATURE REVIEW

2.1 Traditional Healthcare Insurance Pricing Models

The traditional modes of pricing health insurance products depend mostly on the actuarially-based models calculating possible risk levels based on past history and hence charging premiums accordingly[4]. It forms a model based on the risk factors like age, gender, medical history, and lifestyle. Another common method in operation is experience rating, which means premiums are calculated based on past medical claims of an individual or their expected future health care needs. A correct risk related to a given person is presented, but it can overcharge those who already suffer from other illnesses or diseases due to preexisting conditions. Hence, they might not draw in healthier citizens.

A commonly used approach is community rating. There, it groups people by similarity of characteristics—either geographical location or occupation—and then applies an identical premium rate for all people in the group[5]. This provides greater equity, wider coverage but fails to incorporate risk on individual lines, with healthy members who don't demand significant medical expenses having to pay a higher price for that protection. To address some of these inadequacies, models for risk adjustment have been developed wherein premiums are computed based on the specific risk profile of each policyholder. It uses health risk indicators like age, medical history, and other demographic factors to re-calculate premiums with more precision. Although after the implementation of risk adjustment, these traditional models remain relatively inelastic and are unable to respond to the quick changes that are occurring in the health care environment, such as new medical technologies, changes in healthcare policy, or public health emergencies.

Dynamic healthcare risk is usually hard to capture in the traditional way because methods relying exclusively on past information may be found wanting under great uncertainty about future events or needs of care[6]. As a consequence, there has been an increased interest in alternatives to traditional actuarial pricing for health care and, specifically, in using approaches like reinforcement learning (RL) that adapt quickly to reflect individualized complex interactions and thus promise much more precise personal premium settings.

2.2 Recent Advances in Reinforcement Learning Applications

Over the past few years, reinforcement learning has witnessed significant growth especially in dynamic decision-making where the adaptation needs to occur with respect to the real-time streams of information[7]. Adaptability and data-driven decision making have been areas where it showed promising results, whether in finance, healthcare, or robotics. Therefore, models of reinforcement learning can set the best strategy by interacting with their environment; thus, there is a very promising application area in healthcare insurance pricing.

In the health sector, RL has been applied in everything from in-hospital resource optimization to personalized medicine. For example, hospitals have begun to apply RL in scheduling and managing patient flow so that there is an optimal deployment of medical staff and resources[8]. Applications of RL algorithms to personalized medicine tailored the treatments according to the patient based on learning from response to different therapies and changing future treatment plans accordingly. Examples of the above applications demonstrate how flexible and adaptable RL can be in dynamic data-driven environments.

Though applications of RL in healthcare are still at the nascent phase, there is increasing interest in utilizing its capabilities toward making better insurance price quotes[9]. In a dynamic fashion, very broad health data can be learned from continuously and available toward more personalized and accurate premium pricing. Unlike traditional actuarial models, the technology of RL has a better opportunity to consider more inputs into the analytical model, such as real-time patient behavior, emerging trends in healthcare, and policy changes. In addition, in light of the role of evolution and transformation in the evolution of the healthcare landscape, similar frameworks can be applied to create pricing models that more closely reflect the true risk profiles of policyholders—thus providing a solution that is considerably more flexible and responsive than the traditional static pricing models. This is the most significant leap forward in addressing the various shortcomings of the latter, thus enabling an insurance premium pricing methodology that is truly data-driven and responsive.

3. REINFORCEMENT LEARNING FUNDAMENTALS

3.1 Key Concepts of Reinforcement Learning

Reinforcement learning is considered one of the more specific classifications within the larger machine learning paradigm. It outlines a process by which an agent learns to act to maximize a reward signal through interactions with the environment[10]. An agent acts in this example and receives afterward feedback on the actions taken. The agent changes its action based on the total reward achieved after such feedback has been received. The basic building blocks of reinforcement learning include the agent, the environment, the states, the actions, and the rewards.

This would be the decision-maker interacting with the environment. In healthcare insurance, this could be the market or pool of policyholders. States refer to conditions or configurations of the environment at any one time. Healthcare insurance might view this as a person's health status, demographics, or even claims history[11]. These actions are in reference to the options that the agent has access to, including calculating the premium rate or providing coverage in any of its variants. After performing each action, the agent receives a reward based on which the agency decides whether their decision-making is effective or not. In an insurance context, reward-influencing factors are profitability, retention of a customer, or accuracy of pricing.

This is the kind of feedback loop that the RL algorithms use to improve their strategies about decision-making on the go[12]. In other words, exploration and exploitation are the most significant trade-off characteristic of RL: while exploration is a discovery of new, potentially better strategies by trying new actions, exploitation is about leveraging the known actions with the highest rewards in the past. This balance is particularly important in rapidly changing contexts, like the healthcare sector, where emerging variables and uncertainties may encourage the model to explore other pricing strategies while using the data it has gathered.

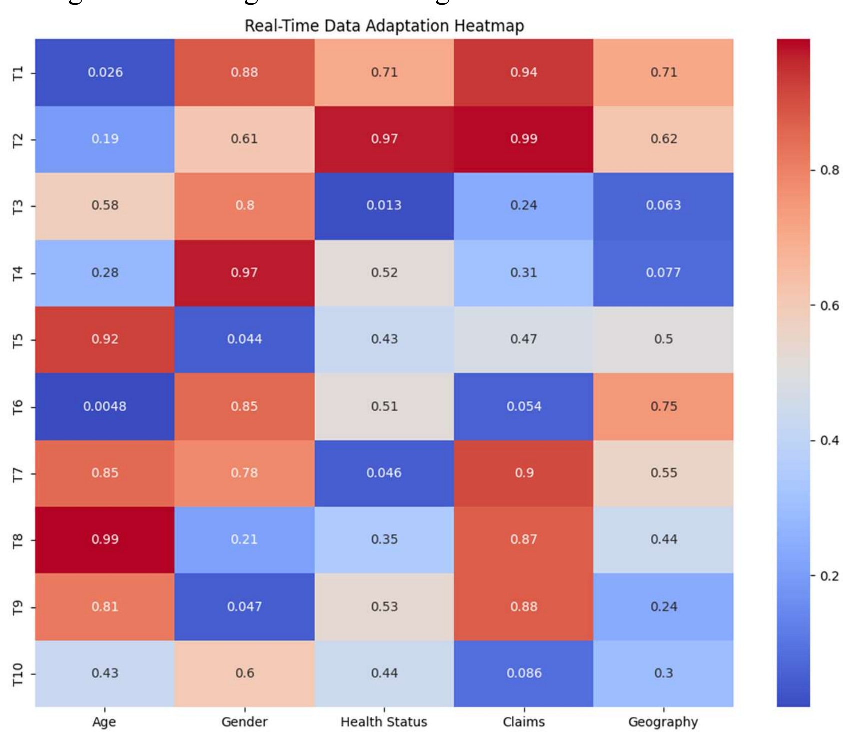

Figure 1 Heatmap of real-time data adaptation and its effect on dynamic insurance pricing (Source: Research, 2024).

3.2 Exploration vs. Exploitation in Reinforcement Learning

This dilemma of exploration vs. exploitation represents one of the major challenges in the area of reinforcement learning. Exploration involves the act by which an agent tries multiple actions, although they are less sure about maximizing their immediate reward. This means the agent gets exposed to more innovative ways and expands their experiences[13]. Exploration, in this sense, within the context of health insurance would be trying different price strategies: integration of behavioral health data within premium computation and even exploring alternate risk groups.

On the contrary, exploitation entails activities that, in the past, have been giving a high return to the agent. With regard to health care insurance, it is when premium strategies that previous information has shown to be successful are adopted[14]. The right balance between these two dimensions is important, as excessive exploration may lead to inefficiencies or unwarranted risks, while excessive exploitation may prevent the model from recognizing potentially superior strategies that emerge due to shifts in market conditions or new trends.

This balance between exploration and exploitation is really hard to achieve in the field of healthcare insurance, taking into account the different unpredictable factors which influence this area, such as the heterogeneity of patients' behavior, innovations in medical technology, and modifications in government rules[15]. RL models can be trained to change the balance over time, shifting gradually more towards exploitation, and more confident in strategies used yet leaving room for exploration and trying new things when new risks or uncertainties arise.

4. HEALTHCARE INSURANCE PREMIUM PRICING OPTIMIZATION

4.1 Factors Influencing Insurance Premium Pricing

It ranges from individual and systemic levels in factors that affect the pricing of insurance premiums. On the individual level, several demographic features such as age, gender, and occupation will tend to determine the risks an individual would be characterized with for health conditions and therefore the related costs of the premiums[16]. More importantly, one thing that is fundamental is the health status of a person; those individuals who have chronic conditions or even experience a history of substantial medical expenses may result in higher rates of premium. A claims history, which basically tracks the past healthcare usage of an individual, is another determinant based on expected future claims for an insurer.

On a macro-level, economy determines general levels of premiums while at times sensitive to inflation in general health expenditure as well as to new changes in policies; for instance the recently developed policy known as affordable care in US will influence and determine changes of a pool for any risk affecting insurance[17]. Besides, unexpected factors like pandemics or new medical breakthroughs can create unpredictable shifts in healthcare risk factors, and thus conventional price models cannot be expected to react

.

Figure 2 Radar chart showing the impact of different factors on insurance premium pricing (Source: Research, 2024).

These factors are, therefore, interlinked in very complex ways, making it complicated for the insurers to predict premiums precisely[18]. Since these variables are dynamic, traditional models based on historical information often prove to be a static assumption. Reinforcement learning, on the other hand is relatively more flexible and reactive with respect to premium pricing as it can learn from immediate feedback.

Table 1: Factors Affecting Healthcare Insurance Premium Pricing

Factor	Description	Impact on Premium Pricing
Age	A key factor affecting risk, with older individuals typically facing higher premiums.	Older individuals generally have higher healthcare needs.
Gender	Gender-related risk factors can influence healthcare needs, though the impact varies.	Gender-based pricing can differ due to varying healthcare risks.
Health Status	Chronic conditions or medical history can drastically raise premiums.	Individuals with high medical costs or chronic conditions often face higher premiums.
Claims History	Past insurance claims provide insight into future medical needs.	Frequent claims increase the likelihood of future claims, leading to higher premiums.

Socioeconomic Factors	Income, occupation, and education levels affect access to care and risk.	Individuals in lower socioeconomic groups may face higher premiums due to increased health risks.
Geographic Location	Regional differences in healthcare costs and risk factors.	Regional health risk factors can drive premiums up or down.
Healthcare Policy Changes	Government regulations and policies, such as the introduction of the Affordable Care Act.	Can directly affect premium rates by adjusting risk pools and coverage mandates.
New Medical Technologies	Emerging medical technologies that reduce or increase healthcare costs.	Premium adjustments may be necessary to cover new treatments or therapies.
External Events (e.g., Pandemics)	Large-scale public health crises (e.g., COVID-19) can create dramatic shifts in risk.	Increased premiums may be necessary to cover additional health risks and insurance claims.

4.2 Challenges in Traditional Pricing Models

Traditional healthcare insurance pricing models, being the back bones of the entire industry, possess several drawbacks. The most crucial one is the fact that it is too rigid. These depend on past statistics and predetermined suppositions that, in most cases, tend to be inflexible to the healthcare environment's drastic changes[19]. In this case, the traditional actuarial models may ignore some sudden health crisis, like a pandemic, that would alter the risk profile of the insured population altogether.

Another problem is the lack of granularity in traditional models. Most actuarial methods rely on broad categories or age bands, general health status, which fail to capture many facets of an individual's health journey. Therefore, at times, some people are overpriced and others underpriced when they don't strictly fit into those categories.

The traditional models also cannot account for new or emerging risks. That is, when there are new medical treatments or technologies, there might not be a reflection on how it impacts healthcare costs and insurance premiums, and therefore, no immediate adjustment of pricing[20]. This may mean that premiums may not reflect the real risk or value of coverage, which hurts both insurer profitability and customer satisfaction.

It could be applied much more dynamically with real-time data in learning its approach to fine-tune the price strategy. In fact, the models can adapt RL to change its preference as it emerges. A greater possibility will exist toward the arrival at more accurate, timely, and fairer structures for price structures as the changes of the landscape of health risks evolve.

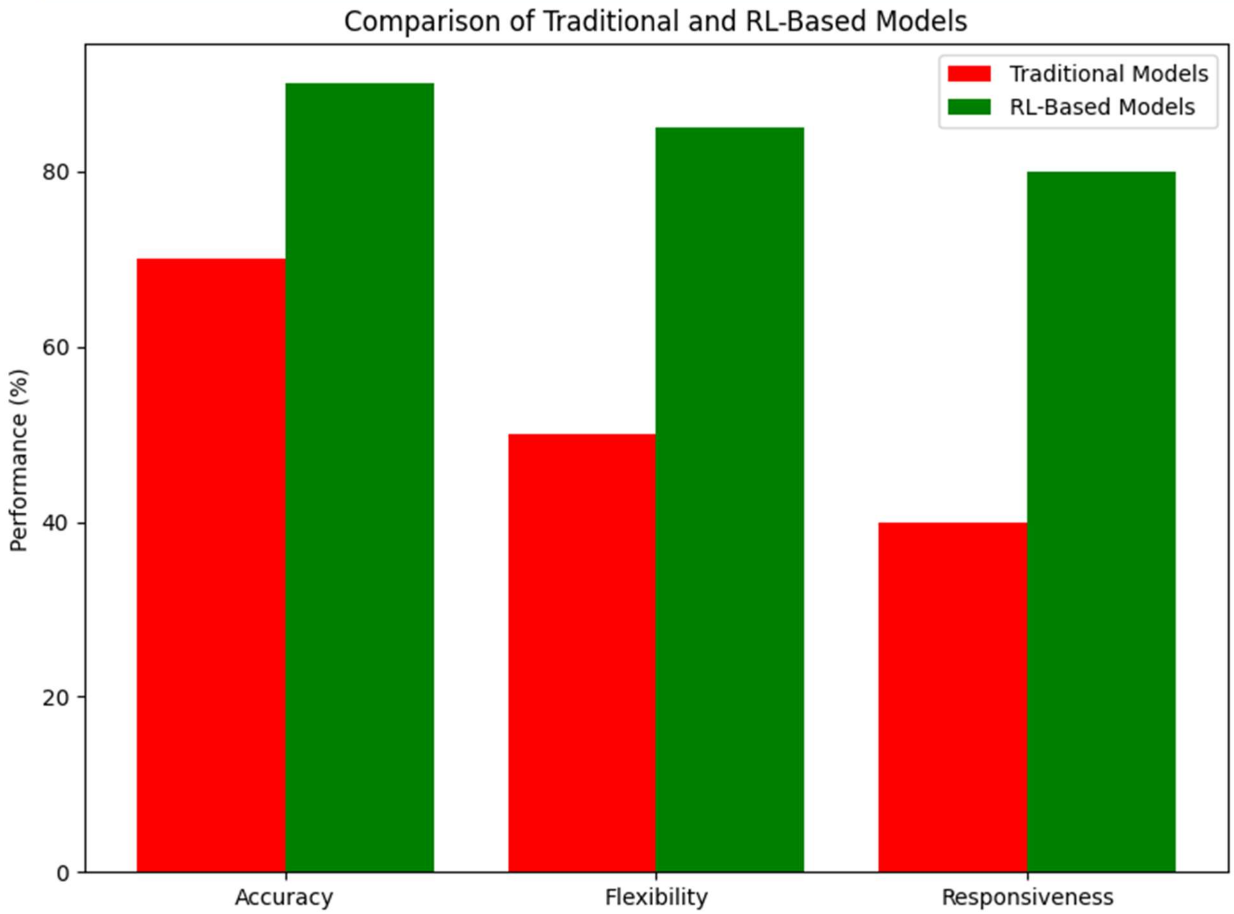

Figure 3 Comparison of traditional and RL-based pricing models in terms of key performance metrics (Source: Research, 2024).

5. METHODOLOGY FOR INTEGRATING REINFORCEMENT LEARNING IN PRICING

5.1 Setting Up the Problem as a Markov Decision Process

The first step to achieve the successful implementation of RL into healthcare insurance premium pricing is by modeling the problem as an MDP. Indeed, the framework of MDP is highly effective for decision-making tasks with an uncertain outcome as well as results dependent on earlier actions and environment states. This for the purpose can be said as the environment means the pool of insured people that have different states reflecting health status, age, history of diseases and claims and any other pertinent characteristics.

The core building blocks of an MDP include states, actions, and rewards. For example, a set of variables that describes the current state of an individual might define the state space related to their medical claims, health risks, or socioeconomic factors. The action space provides various pricing strategies that an insurer can adopt; these include adjusting premiums related to certain categories of risks, introducing new coverage options, or changing the existing payment structures.

The reward function is the guide that helps through the learning process. In this context, there can be any one of multiple goals for rewards associated with profit maximization, better customer retention, or making the insurance rates more competitive as compared to that of other competitors. Long-term objectives may be considered in terms of improving the quality of risk pool or making sure that there are minimum chances of adverse selection, avoiding better people from getting coverage or those seeking cheaper ones.

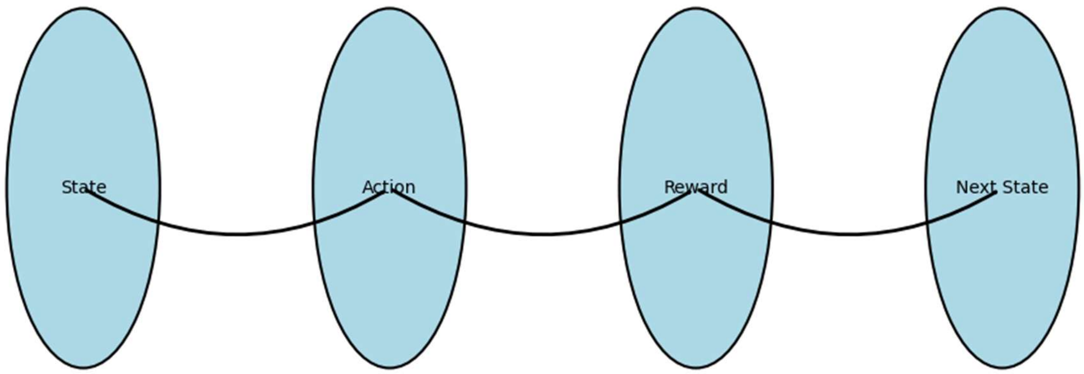

Figure 4 Illustration of Markov Decision Process for healthcare insurance pricing optimization (Source: Research, 2024).

The conceptual formulation of the problem as an MDP really offers a structured framework for the optimization of decision-making by RL agents, hereby aiming to achieve maximization of cumulative rewards over time. Thus, any action the agent makes might cause a modification in the environment, such as premium adjustments and customer retention; through continuous interaction with this environment, the RL agent will successively enhance its policy for premium pricing.

Table 2: Example of a Markov Decision Process for Healthcare Insurance Pricing

State	Action	Reward	Next State
Policyholder with chronic illness	Increase premium by 10%	High reward if the increased premium leads to profitability	Policyholder remains in the same state, but with higher premium
Young, healthy policyholder	Offer lower premium with additional wellness incentives	Small reward if the policyholder remains insured	Transition to a state where the policyholder stays insured long-term
Policyholder with recent claims	Increase premium by 15%	Moderate reward for balancing risk and profitability	Higher probability of retention but with increased premium risk
Healthy, low-risk policyholder	No premium adjustment, maintain current rate	Low reward due to low-risk nature	Policyholder stays in a low-risk state, ensuring long-term retention

5.2 Algorithmic Approaches for Policy Optimization

Once the problem is framed as an MDP, the next step is the choice of appropriate RL algorithm for effective policy optimization. Several algorithms are now available with their strengths and weaknesses. The choice also depends on the complexity of the pricing environment, as well as the nature of the state and action spaces.

Q-learning is perhaps one of the most intuitive and most widely used RL algorithms in cases where the action space is discrete. In Q-learning, an agent learns about the value of every state-action pair (Q-value) and updates these iteratively towards converging towards the optimal policy. In the realm of health care insurance, pricing decisions are always discrete but in some cases will include more than one possible modification; therefore Q-learning can become useful for decision-making in directing the pricing model, which would learn from previously taken actions along with their associated rewards.

For continuous action spaces, that is, premium rates are not categorically discrete and take continuous values, more powerful methods are employed. Deep Q-Networks extend Q-learning naturally by using deep neural networks to approximate the Q-values, which can handle much larger and more complex state-action spaces. DQNs can be used effectively in scenarios involving continuous variables for premium pricing decisions requiring models of high-dimensional states, for instance, the fusion of demographic and medical data.

A whole sub-class of policies gradient methods operate by optimizing parameters governing the process of decision via updating the same. These works exceptionally well in domains involving continuous action. These proceed by iterating upon a policy - improving it a little at every step by finding the gradient of the expected reward with respect to policy parameters. In healthcare insurance, with pricing perhaps along a continuum of premium rates, policy gradient methods enable the RL agent to scan through a number of possible premium policies and learn how best to behave to achieve higher profitability or higher customer satisfaction.

Lastly, Actor-Critic Methods, which is the integration of value-based and policy-based approaches, can provide more stability and efficiency in training. The actor (which selects the actions) and the critic (which evaluates the actions) collaborate to enhance the agent's decision-making process. Such a combination can be particularly useful where the agent is making both discrete and continuous decisions, such as when adjusting both categorical factors (such as risk groups) and continuous factors (such as specific premium amounts).

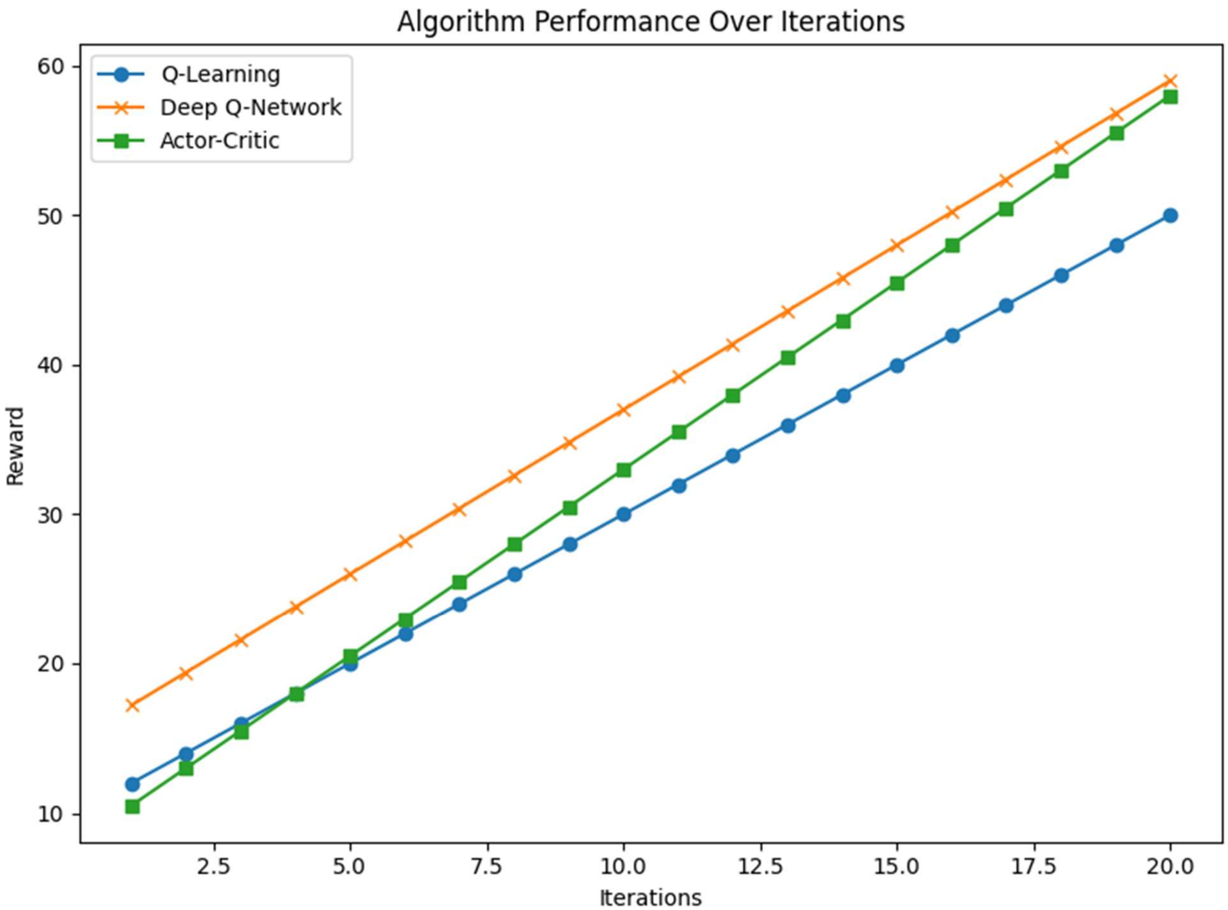

Figure 5 Heatmap of real-time data adaptation and its effect on dynamic insurance pricing (Source: Research, 2024).

The right choice of RL algorithm could be applied in training agents on making optimal premium pricing decisions in real-time, which would continue improving over time based on new data. The algorithms, through iterative learning, could adapt to change in market conditions, emerging health risks, and new data patterns, ensuring the premium prices stay accurate and competitive.

Table 3: Comparison of RL Algorithms for Healthcare Insurance Pricing Optimization

Algorithm	**Action Space**	**Strengths**	**Weaknesses**	**Best Use Case**
Q-Learning	Discrete	Simple to implement and understand; effective for small action spaces.	Limited by the size of the action space; struggles with high-dimensional spaces.	Optimal for discrete action spaces (e.g., fixed premium categories).
Deep Q-Networks (DQN)	Discrete	Scalable to high-dimensional spaces; works well with complex environments.	Computationally intensive; requires large amounts of training data.	Best for environments with a large state-action space, such as complex risk categories.
Policy Gradient Methods	Continuous	Direct optimization of the policy; works well for continuous action spaces.	High variance in training; may require careful tuning.	Effective for continuous pricing adjustments (e.g., exact premium amounts).

Actor-Critic Methods	Continuous	Combines value-based and policy-based approaches; more stable learning process.	Requires more computational resources; may be slower to converge.	Useful for complex environments where both discrete and continuous decisions are involved.

6. CONCLUSION

It would hence alter the health care dynamics as far as risk measurement by the insurance providers goes and therefore, premiums. While these models developed such approaches, they failed in the face of such static models alongside reliance on preset assumptions; it cannot adjust to these changes in the health care dynamics transformations. As compared to it, reinforcement learning is much more flexible and adaptable and changes accordingly with real-time data so as to enable insurance companies to be able to tune their pricing better.

Another major plus of RL is the very fact that RL can easily model complex, changing factors into its pricing model having fluctuating costs of health care, emerging health trends, or shifts in consumption behaviors. This being a dynamic health care environment, the RL-based systems can have a real-time change in their decision-making parameters to be the market trend leaders with better pricing and provide the reinsurers with benefits based on the risk profiling of that individual. It is particularly crucial in an industry as unpredictable and complex as healthcare where new risks, medical breakthroughs, and policy changes can impact the individual premium and overall pricing structures dramatically.

Other than the balance carried out in RL for exploration and exploitation, the experience built up to now could be leveraged on besides the best new pricing by innovative new routes for even more effective policy or cost-effectiveness. This is how that experimentation in RL about novel factors of danger, which most likely are modelled using some alternative structural forms of structure of pricing sometime deliver novel elements of information, which may actually be used for building an inference, that actual traditional models fails in which aspects. This, in turn leads to a process of continuous improvement in the management of premium price based on exploration and hence generates a significantly more responsive and dynamic system for both the insurer and the policyholder.

Although these advantages are promising, challenges are faced when they are applied to RL in healthcare insurance. There is a data privacy requirement and a need for quality data in huge quantities to train the agent. An appropriate reward function must be defined according to the business goals of the insurer. The conventional pricing model will require investment in technology, training, and systems integration into the RL-based systems.

After all, reinforcement learning would change the face of health insurance premium pricing, no matter how big an obstacle that is. RL brings the industry one step closer to more personalized, equitable, and fairer pricing for everyone. With better profitability also comes better customers' satisfaction and confidence. As a result, much further maturity in reinforcement learning will then enable healthcare insurers to design even more precise products and assist in managing even more dynamic risks in their health care operations. This should have a positive effect on the insurance market's development, making it highly responsive, efficient, and appropriately aligned with the changes occurring in health care.

REFERENCES

[1] C. Yu, J. Liu, S. Nemati, and G. Yin, "Reinforcement learning in healthcare: A survey," *ACM Comput. Surv.*, vol. 54, no. 5, 2021.

[2] J. Perumalsamy, C. Althati, *et al.*, "Advanced AI and machine learning techniques for predictive analytics in annuity products: Enhancing risk assessment and pricing accuracy," *J. Artif. Intell.*, vol. 49, no. 2, 2022.

[3] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.

[4] G. Y. Ban and C. Rudin, "The big data newsvendor: Practical insights from machine learning," *Oper. Res.*, vol. 67, no. 3, pp. 656–673, 2019.

[5] G. A. Susto, A. Schirru, S. Pampuri, *et al.*, "Machine learning for predictive maintenance: A multiple classifier approach," *IEEE Trans. Ind. Inform.*, vol. 11, no. 3, pp. 812–820, 2014.

[6] M. G. Kibria, K. Nguyen, G. P. Villardi, O. Zhao, *et al.*, "Big data analytics, machine learning, and artificial intelligence in next-generation wireless networks," *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 2316–2331, 2018.

[7] J. Waring, C. Lindvall, and R. Umeton, "Automated machine learning: Review of the state-of-the-art and opportunities for healthcare," *Artif. Intell. Med.*, vol. 104, p. 101822, 2020.

[8] R. Xiong, J. Cao, and Q. Yu, "Reinforcement learning-based real-time power management for hybrid energy storage system in the plug-in hybrid electric vehicle," *Appl. Energy*, vol. 211, pp. 538–548, 2018.

[9] S. T. Boppiniti, "Machine learning for predictive analytics: Enhancing data-driven decision-making across industries," *Int. J. Sustain. Dev. Comput. Eng.*, vol. 1, no. 1, pp. 18–24, 2019.

[10] T. Panch, P. Szolovits, and R. Atun, "Artificial intelligence, machine learning and health systems," *J. Glob. Health*, vol. 8, no. 2, p. 020303, 2018.

[11] S. T. Boppiniti, "Big data meets machine learning: Strategies for efficient data processing and analysis in large datasets," *Int. J. Creat. Res. Comput. Sci. Technol.*, vol. 4, no. 2, pp. 14–21, 2020.

[12] A. E. Khandani, A. J. Kim, and A. W. Lo, "Consumer credit-risk models via machine-learning algorithms," *J. Bank. Finance*, vol. 34, no. 11, pp. 2767–2787, 2010.

[13] Z. Ahmed, K. Mohamed, S. Zeeshan, and X. Q. Dong, "Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine," *Database*, vol. 2020, Article ID 1489146, 2020.

[14] S. Gu, B. Kelly, and D. Xiu, "Empirical asset pricing via machine learning," *Rev. Financ. Stud.*, vol. 33, no. 5, pp. 2223–2273, 2020.

[15] S. Gu, L. Yang, Y. Du, G. Chen, F. Walter, J. Wang, *et al.*, "A review of safe reinforcement learning: Methods, theory and applications," *arXiv preprint arXiv:2205.09762*, 2022.

[16] B. Recht, "A tour of reinforcement learning: The view from continuous control," *Annu. Rev. Control Robot. Auton. Syst.*, vol. 2, no. 1, pp. 253–281, 2019.

[17] S. Athey, "Beyond prediction: Using big data for policy problems," *Science*, vol. 355, no. 6324, pp. 483–485, 2017.

[18] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future," in *BioApps: Automation of Decision Making*, Springer, 2018, pp. 1–27.

[19] L. Brunke, M. Greeff, A. W. Hall, Z. Yuan, *et al.*, "Safe learning in robotics: From learning-based control to safe reinforcement learning," *Annu. Rev. Control Robot. Auton. Syst.*, vol. 5, pp. 1–25, 2022.

S. Fujimoto and S. S. Gu, "A minimalist approach to offline reinforcement learning," in *Advances in Neural Information Processing Systems*, vol. 34, pp. 201