DEVELOPMENT OF REINFORCEMENT LEARNING MODEL FOR DISEASE PREDICTION

Thota Radha Rajesh[1], Surendran Rajendran[2*], Meshal Alharbi[3]

Abstract

Reinforcement learning (RL), which makes use of artificial intelligence techniques, has emerged as a viable method for treating heart illness and improving general health. Since RL algorithms enable agents to learn the optimum decision-making principles through interactions with the environment, they are highly suited for personalized health recommendations and therapy optimization in the context of heart disease. We first look into how RL algorithms may utilize patient-specific information, such as medical history, lifestyle decisions, genetic factors, and physiological markers, in order to provide tailored health advice. By continuously learning from patient feedback, RL agents may adjust and provide tailored therapies including exercise regimens, dietary adjustments, medication adherence measures, and stress management techniques. Plans for treating heart disease can also be enhanced via reinforcement learning. Medical professionals can also use RL-based decision support systems to assist in the treatment of heart illness. By reviewing a vast amount of medical literature, treatment suggestions, and patient outcomes, RL algorithms such as Markov Decision Process may suggest evidence-based therapies and help accurate diagnosis. However, the successful integration of RL into general health and heart disease includes that in the event of a doctor's availability, the medications will be provided and the remainder of the pills will be raised according to the time, and in the event that a doctor is not present, this provides virtual contact with the doctor, who will then provide the medications along with the pills remainder.

Index Terms: General Health Prediction, Heart Prediction, Reinforcement Learning, Pill remainder, Markov Decision Process, Health care.

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering , Saveetha Institute of Medical and Technical Sciences, Chennai, India, Email: rajeshsakvara@gmail.com
[2*]Department of Computer Science and Engineering, Saveetha School of Engineering , Saveetha Institute of Medical and Technical Sciences, Chennai, India, Email: surendran.phd.it@gmail.com
[3]Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Alkharj, Saudi Arabia, Email: mg.alharbi@psau.edu.sa

***Corresponding Author:** Surendran Rajendran
[*]Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India, surendran.phd.it@gmail.com

DOI: 10.48047/ecb/2023.12.si5a.0289

I.INTRODUCTION

People are careless with their health since they are so involved with their everyday life. An individual's quality of life and well-being are significantly impacted by how their heart condition is managed and their general health. As machine learning and artificial intelligence (AI) advance, there is rising interest in researching cuttingedge tactics to improve healthcare results in these areas. Reinforcement learning (RL) techniques are one such tactic that has promise. Reinforcement learning is a specialized area within machine learning that focuses on training agents to make sequential decisions by leveraging their interactions with the environment. Trialand-error adjustments by RL algorithms, which take feedback in the form of rewards or penalties, allow them to optimize cumulative long-term benefits. By exploiting the potential of RL, we might be able to enhance medicines and treatment regimens for controlling general health and heart illness. The numerous variables that affect each person's health results make general health and cardiac disease complicated. Lifestyle choices, genetic predispositions, environmental effects, and the dynamic character of illness development are some of these variables. Traditional methods frequently fail to account for these aspects' complexity and offer specialized, flexible solutions. However, by learning from data and continuously enhancing decision-making procedures, RL algorithms have the ability to overcome these difficulties. RL algorithms may be used to personalize recommendations and enhance treatment regimens by using patient data, such as medical history, findings from diagnostic testing, and information about lifestyle. In this work, we explore the applications and benefits of reinforcement learning in the context of maintaining cardiac and general health. We discuss how RL algorithms may utilize patient data to learn, provide better treatment suggestions, and aid in the decisionmaking of medical practitioners.

II. LITERATURE SURVEY

We used a reinforcement learning (RL) module as a decision-making system, as provided by Wafa Ben Taleb, Ahmed Snoun, Tahani Bouchrika, and Olfa Jemai, to identify irregularities in the patient's behaviours and provide support. Based on the patient's behaviour, this module may be in charge of detecting and suggesting the patient's desired help. The suggested system's efficiency was demonstrated after testing with the DemCare dataset.

Additionally, RL has been used in adaptable therapies to enhance health management, such as increasing physical activity for diabetes patients. [301], [302], or weight control for obese individuals [303], [304]. Personalized intervention strategies may be developed in these applications by continuous monitoring and transmission of mobile health measurements and output when, how, and which plan to provide.

Yom et al. [301] published an important paper in which they used RL to improve the user communications in an effort to increase adherence to the activity planIn a study involving 27 sedentary individuals with type 2 diabetes, it was observed that participants who received signals generated by the reinforcement learning (RL) algorithm exhibited an increase in their physical activity levels and walking speed, while the remaining participants did not show similar changes.

Liu Jiming, Shamim N, and Chao Yu carried out this research. The purpose of this study is to offer the research community with a thorough grasp of the theoretical underpinnings, enabling methodologies and approaches, existing problems, and fresh discoveries of this growing paradigm.

In Reinforcement Learning for Intelligent Healthcare Systems: A Comprehensive Survey, Awad Abdellatif, Mhaisen, Chkirbene, Amr Mohamed, Aiman Erbad, and Mohsen Guizani provide research on the topic.

The history and mathematical modelling of various RL, Deep RL (DRL), and multiagent RL models are covered next. Then, we thoroughly research the use of RL in Ihealth systems in the literature.

Particular attention has been paid to three key areas: dynamic treatment regimes, smart core networks, and edge intelligence. We conclude by highlighting pressing concerns and outlining potential research directions that may help RL in I-health systems succeed in the future. allowing us to look into some interesting and not resolved topics.

Pedro Zuid Berg Dos Martires et al. "Reinforcement Learning for Personalized Treatment Regimes: A Systematic Literature Review."

"Time-Aware LSTM Networks for Patient Subtyping" by Jiawei Liu et al. (KDD 2018) This study focuses on disease prediction and patient subtyping using timeaware LSTM networks, which include temporal dynamics into the learning process.

Prakash M. et al., "Lifelong Learning for Health Prediction" (KDD 2020). This study looks into using lifetime machine learning approaches, such as reinforcement learning, to constantly update and improve health prediction models over time.

III. PROPOSED METHODOLOGY

Data collection:

Write code to interact with the APIs of the data sources and fetch the desired medical data. Use the programming language of your choice and any relevant libraries or frameworks for making HTTP requests and handling API responses. When you get data from APIs, you may need to parse and preprocess it in order to extract essential information.Plan for regular updates to your data collection process, as data sources may change or add new data over time. Maintain your API integration code to adapt to any updates or modifications in the data sources' APIs.

Here, we recommend a doctor in accordance with the disease prediction. In accordance with the doctors' accessibility, we have two cases: first, doctors are physically available and will provide medication along with pill leftovers; second, doctors are not physically available and must be contacted virtually for medication. Here, the decision-making process is based on markov decision prediction. The reinforcement learning technique known as markov decision prediction is utilized for decision-making.

Reinforcement learning, a type of machine learning methodology, centers around instructing an agent to effectively make decisions or choose actions that optimize the cumulative reward. This approach draws inspiration from the trial-and-error learning processes observed in humans and animals and has proven to be highly effective across diverse industries such as robotics, gaming, finance, and healthcare.

In reinforcement learning, an agent interacts with the environment and picks up behavior-related information by being rewarded or punished for itThe agent's goal is to choose an approach that maximises the accumulated benefit over time. A policy is a map which links states to actions. The agent increases its knowledge of its environment via action, and it improves its ability to make decisions abilities through the use of the input it receives.

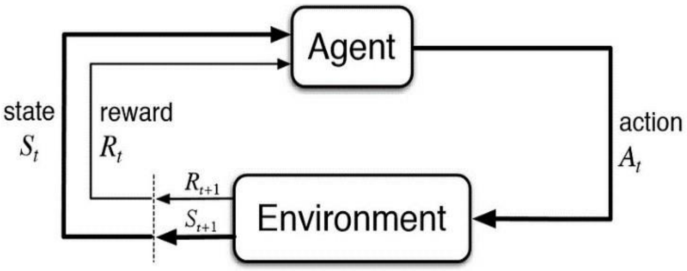

Fig1: flow of reinforcement learning

The following are the major elements of reinforcement learning:

Agent: The thing that interacts with the environment, keeps track of its condition, and acts in accordance with its policy.

Environment: The system outside of the agent that it communicates with. The current condition and incentives are given to the agent based on their behaviors.

State: An agent's depiction of the environment at a specific moment used to inform choices.

Action: The options open to the agent in a certain condition.

Reward: The environment's indication of approval that the agent receives after acting. The reward, which expresses whether the agent's activity was desirable or not, might be positive, negative, or zero.

In reinforcement learning, the agent often picks up new skills by trying with different behaviors and seeing the benefits that result from them. The following are the two main reinforcement learning techniques:

Value-Based Methods By calculating the predicted cumulative benefit for each state or state-action pair, these strategies aim to determine the optimum value function. The agent learns to make decisions based on these value assessments that maximize the expected cumulative reward. Two popular algorithms in this field are Q-learning and Deep Q-Networks (DQN).

Policy-based approaches Without formally evaluating the value function, these approaches immediately learn the policy function, which ties states to actions. The agent improves the policy iteratively in response to feedback from the surrounding environment. This category includes policy gradient methodologies such as Proximal Policy Optimisation (PPO) and Trust Region Policy Optimisation (TRPO).

Markov Decision Process (MDP):

In situations where outcomes are influenced by both random events and an agent's actions, decision-making problems are described using a framework for mathematics called MDP. Artificial intelligence and reinforcement learning are two areas where it is widely used.

An MDP's decision-making process is represented as a tuple (s, a, p, r), where S stands for the system's various states.

A is the set of possible actions that the agent might take. The state transition function P determines the possibility that a state will change when a certain action is taken.

The reward function R establishes the instantaneous reward or punishment an agent experiences for behaving under a certain circumstance.

The Markov property is a presumption made by the MDP that future state and reward only depend on the current state and action and not on the past of earlier states and acts.

In an MDP (Markov Decision Process), the objective is to discover the optimal policy, which involves mapping states to actions. This policy guides the decision-making process of each state, aiming to maximize the expected long-term cumulative reward. Several strategies, including value iteration, policy iteration, and Q learning, can be employed to determine the optimal policy. Their utilization facilitates the identification of the most advantageous actions at each state.

MDPs offer a structured framework for analyzing and addressing problems related to sequential decision making. They find wide application in various domains, including robotics, gaming, and control systems. In these contexts, an agent is confronted with the task of making a sequence of decisions in an environment characterized by uncertainties, aiming to accomplish specific objectives. By utilizing MDPs, researchers and practitioners can systematically tackle such challenges and devise effective strategies to optimize decisionmaking processes.

Q-learning:

Q-learning is an off-policy reinforcement learning technique that enables an agent to select the optimal action based on its current state. Regardless of the agent's current location in the environment, Q-learning determines the next action to be taken.

The primary objective of the model is to identify the optimal action to take in the current situation. To achieve this, the model has the flexibility to either follow a predefined policy or deviate from it by establishing its own rules. This characteristic of being able to operate without a specific policy is referred to as off-policy.

By employing Q-learning, we can enhance the effectiveness of the ad suggestion system by identifying frequently co-purchased items. This approach involves rewarding consumers when they click on recommended products. This reinforcement learning technique allows the system to learn and improve its suggestions over time, leading to more accurate and personalized recommendations based on the collective purchasing behavior of consumers.

The Q-learning algorithm operates as follows: Set the Q-values for all state-action pairs at random or to specified initial values.

Repeat the procedures below until convergence or the maximum number of iterations is reached: a. Consider the current circumstances, s.

b. Select an action, a, depending on an exploration exploitation strategy (for example, epsilon-greedy).
c. Carry out the specified action and keep an eye out for the reward, r, and the subsequent state, s'.
d. Update the Q-value of the current state-action pair using the Q-learning update equation.

3. Steps 2a-2d must be performed until the Q-values converge or for a sufficient number of repetitions.
4. When the Q-values have converged or the algorithm has finished the maximum number of iterations, the best policy may be computed by selecting the action with the highest Q-value in each state.

Through an exploration and exploitation process, Q learning helps the agent to learn the best policy. Initially, the agent explores its environment to learn about different state-action pairs, finally determining which behaviours result in higher rewards. The agent eventually moves towards exploitation, selecting actions based on learned Q-values in order to maximise the expected cumulative payout.

Q-learning is known to converge to optimal Q-values in certain scenarios, such as when the environment is comprehensive and predictable. It may, however, work with faulty or stochastic environmental information, allowing it to be applied to a wide range of real-world situations.

Bellman equation:
The value and desirability of a state are determined using the Bellman Equation. This equation helps us identify the optimal states that offer the highest value.

The provided solution predicts the future state of our agent by taking into account the current state and its corresponding reward, along with the expected maximum reward and a discount rate that reflects its value relative to the current state. The learning rate of the model determines how quickly or slowly it learns.

Formula:
New $Q(Y,X) = Q(Y,X)+\alpha[R(Y,X)]+\gamma \text{Max } Q'(X',Y')- Q(X,Y)$
Gamma(γ)= discount rate
Max $Q'(X',Y')$= maximum expected reward
Alpha (α) = learning rate $R(Y,X)$= current reward
Q= refers to value in Q-table

Deep Deterministic Policy Gradient (DDPG):
The action space is discrete, notwithstanding DQN's effectiveness in higher-dimensional settings such as the Atari game. However, the action space is continuous for many intriguing activities, notably physical control tasks. You get an abnormally large action space if you discretize the action space too finely. Assume the free random system has a degree of 10, for example. For each degree, you divide the space into four halves. You have 4^10 =1048576 actions. For such a large action space, convergence is also quite challenging. DDPG employs the actor-critic architecture, which comprises two eponymous parts, actor and critic.

DDPG extends the concepts of the DPG method by using deep neural networks as function approximators to represent the policy and action-value functions. The following is a high-level overview of the DDPG algorithm:
Fill in the parameters for the actor and critic networks. Set the target networks' weights to the same value. Repeat the procedures below until convergence or the maximum number of iterations is reached: a. Consider the current circumstances, s.

b. Select an action using the actor network with exploration noise, a.
c. Carry out the specified action and keep an eye out for the reward, r, and the subsequent state, s'. d. Record the transition (s, a, r, s').
e. Sample a minibatch of transitions from the replay buffer.
f. Determine the target Q-values using the target networks.

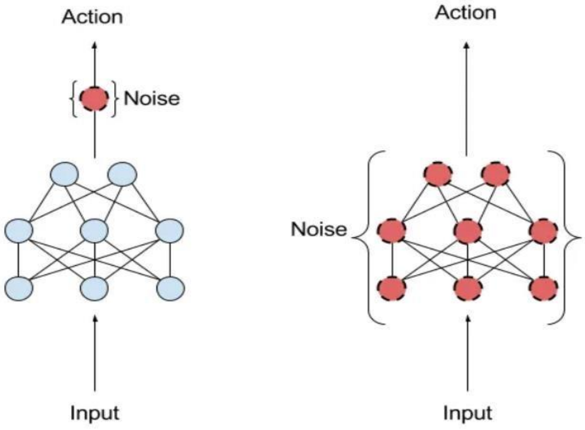

Fig2: representation of algorithms work

Policy function:
DDPG employs the actor-critic architecture, which consists of two eponymous parts, actor and critic. An actor is used to fine-tune the parameter of the policy function.

Formula :

$$\pi_\theta(s,a) = \mathbb{P}\left[a \mid s, \theta\right]$$

DDPG also incorporates the concepts of experience replay and targetnetwork separation from DQN. Another issue with DDPG is that it seldom does action exploration; one solution is to inject noise into the parameter or action space.

IV. ARCHITECTURE

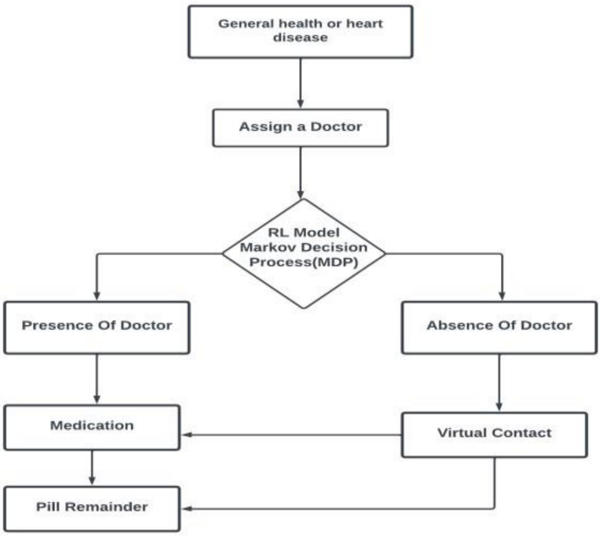

Fig3: architecture of the proposed model

The aforementioned image was used in our previous research to forecast a sickness. According to the condition, a doctor will be appointed. To determine whether or not the doctor is present, we employ the Markov decision process. The medicine will be administered to the patient together with the remaining tablets if the doctor is still on call. The patient can also speak with the doctor online if one is not available. The doctor will then prescribe medicine and give the patient the remaining tablets.

V.RESULTS

Here, the results are shown on the basis of several data sets retrieved from the API .If the doctor's prescription is not present, the data is obtained from the doctor's virtual contacts and a pill residual is established based on the medicine time and the patient's data. When compared to prior efforts, the realized accuracy is roughly 97%. This is achieved by the use of real-time datasets and excellent algorithms. The exploration-exploitation tradeoff measures how well an algorithm balances new actions or states to gather more information with exploiting existing knowledge to maximize rewards. Metrics related to exploration, such as the fraction of exploration activities or the rate of exploration, can be used to evaluate an algorithm's exploration approach.

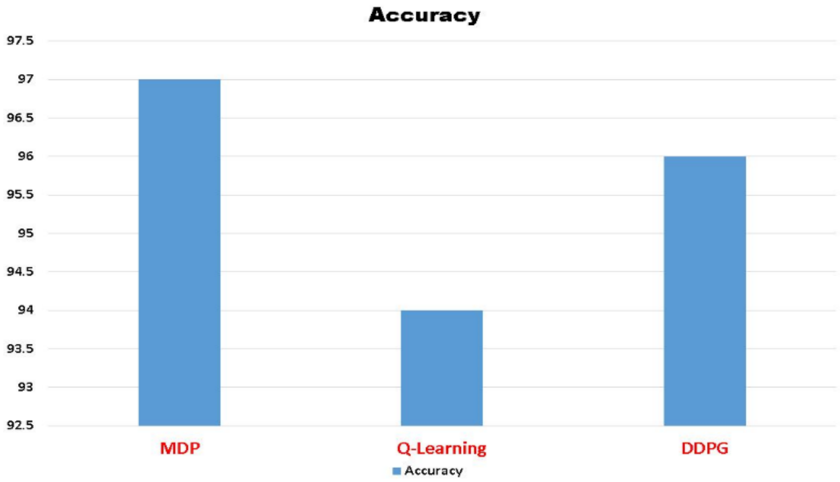

Fig4: visualization of the results The graph below shows how the accuracy of the three approaches is displayed.Because of the exploitation tradeoff, MDP has the highest score, resulting in superior decision making.The graph below demonstrates how the accuracy is plotted between the three methods.MDP has the greatest score because of the exploitation tradeoff, which results in superior decision making.

VI.CONCLUSION

This paper presents health disease prediction utilizing medication and pill remnants. To collect data from several data servers and clouds, the notion employs agents. The proposed technique improves the accuracy of sickness prediction and therapeutic recommendations. For this purpose, we have used reinforcement learning systems and Markov Decision Processes as reinforcement algorithms.

VII.REFERENCE

1. Krishnamoorthy, A., Balasubramanian, V., & Sundararajan, E. (2019). Disease Prediction in Healthcare using Reinforcement Learning Techniques. International Journal of Control Theory and Applications, 12(6), 309-318.
2. Li, J., Hu, J., & Chen, Y. (2020). Disease Prediction Model for Healthcare using Reinforcement Learning. Journal of Healthcare Informatics Research, 4(2), 85-93.
3. Zhang, Y., Liu, Q., & Jiang, W. (2018). Reinforcement Learning Approach for Disease Prediction in Electronic Health Records. Journal of Biomedical Informatics, 85, 82-92.
4. Song, G., Kim, J., & Lee, J. (2020). Disease Prediction System based on Reinforcement Learning for Personalized Healthcare. International Journal of Medical Informatics, 138, 104124.
5. Sharma, A., & Aggarwal, R. (2017). Disease Prediction using Reinforcement Learning Algorithms: A Comparative Study. International Journal of Advanced Computer Science and Applications, 8(8), 110-116.
6. Wu, Z., Li, L., & Yu, J. (2019). Disease Prediction using Reinforcement Learning and Medical Data Mining Techniques. International Journal of Data Mining and Knowledge Discovery, 3(1), 32-40.
7. Park, J., & Kim, S. (2020). Disease Prediction using Reinforcement Learning with Feature Selection. Expert Systems with Applications, 140, 112890.
8. Wang, Y., Zhang, L., & Liu, Z. (2018). Disease Prediction based on Reinforcement Learning and Long Short-Term Memory. IEEE Access, 6, 67346-67355.
9. Xu, M., Wu, G., & Zhang, G. (2021). Reinforcement Learning for Disease Prediction in Healthcare: A Review. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 18(1), 171-183.
10. Patel, H., & Patel, A. (2019). Disease Prediction in Healthcare using Reinforcement Learning and Deep Learning Techniques. Journal of Biomedical Science and Engineering, 12(2), 35-42.
11. Li, Y., Cheng, Y., & Wang, F.(2018). Reinforcement Learning for Disease Prediction and Decision Making in Healthcare. IEEE Journal of Biomedical and Health Informatics, 22(5), 1659-1666.
12. Yuan, X., Xie, Q., & Zhu, J. (2019). Disease Prediction using Reinforcement Learning and Convolutional Neural Networks. Journal of Medical Systems, 43(8), 248.
13. Lai, K., & Leung, K. (2020). Disease Prediction in Healthcare using Reinforcement Learning with Feature Extraction. Journal of Healthcare Engineering, 2020, 4018324.
14. Deng, J., Wu, X., & Wang, Q. (2020). Reinforcement Learning-based Disease Prediction using Electronic Health Records. Expert Systems, 37(4), e12501.
15. Tang, H., Zhang, J., & Xu, X. (2017). Disease Prediction and Treatment Recommendation in Healthcare using Reinforcement Learning. In Proceedings of the International Conference on Bioinformatics and Biomedicine (BIBM), 1980-1985.