Project Report

Evaluating the Performance of Double Deep Q-Network (DDQN) Reinforcement Learning in Heart Disease Prediction: A Study on Algorithmic Ineffectiveness

Submitted by

Silvia Sifath
Class Roll:09-001-04
Course: DSP-01
Department of Computer Science & Engineering,
University of Barishal

Submitted to

Dr.Tania Islam
Assistant Professor
Department of Computer Science & Engineering,
University of Barishal

Department of Computer Science and Engineering
University of Barishal
10 November,2024

Evaluating the Performance of Double Deep Q-Network (DDQN) Reinforcement Learning in Heart Disease Prediction: A Study on Algorithmic Ineffectiveness*

Abstract—**In recent times, reinforcement learning (RL) algorithms, especially Double Deep Q-Networks (DDQNs), have shown notable success in fields that require sequential decision-making and control. Nevertheless, their use in predicting structured tabular data, such as heart disease classification, has not been thoroughly investigated. This research assesses the performance of the DDQN method in forecasting heart disease using a clinical dataset. We executed the DDQN algorithm to measure its effectiveness in this area, employing standard evaluation metrics. Our findings indicate that DDQN has difficulty achieving competitive levels of accuracy and efficiency when compared to conventional supervised learning models, implying it may not be ideal for tasks that necessitate precise classification over strategic sequential actions. The study explores possible explanations for this discrepancy in performance and proposes alternative algorithms that may be more appropriate for medical prediction tasks.**

Index Terms—**Heart Disease Prediction, Double Deep Q-Network (DDQN), Reinforcement Learning, Machine Learning in Healthcare, Algorithmic Ineffectiveness, Supervised vs. Reinforcement Learning, SHAP (SHapley Additive exPlanations**

I. Introduction

Cardiovascular disease continues to be one of the most common and lethal health issues worldwide, leading to a substantial amount of mortality and morbidity. Accurately predicting heart disease is essential for early detection, prompt interventions, and enhanced patient outcomes. As the incidence of heart disease rises, especially among aging populations and in areas with limited healthcare access, the need for effective predictive models has grown. In light of this, both researchers and healthcare practitioners have increasingly adopted machine learning methods, utilizing extensive datasets and advanced algorithms to pinpoint individuals at an elevated risk and guide clinical decision-making [1], [5].

. Conventional machine learning models, including logistic regression, support vector machines (SVMs), decision trees, and ensemble techniques, have been frequently used for heart disease prediction. These models typically demonstrate strong performance in terms of accuracy, offering valuable insights from the structured data commonly found in medical records [1], [4].

In recent times, advancements in artificial intelligence (AI) and machine learning have led to the development of more sophisticated models that can identify complex patterns within data. Among these advancements are reinforcement learning algorithms, which have proven to be highly effective in settings that demand adaptive decision-making [2].

. Specifically, reinforcement learning, especially in the form of Deep Q-Networks (DQNs) and their derivatives, has played a significant role in areas like robotics, gaming, and autonomous systems, where decision-making occurs in a sequence and learning is based on interactions with an environment. Reinforcement learning has demonstrated the ability to enhance intricate decision-making tasks by learning from experiences in order to maximize overall rewards over time [3].

. A distinctive aspect of reinforcement learning is its interaction with the environment via a feedback loop, whereby the agent learns to perform actions that lead to positive results. This feature renders reinforcement learning particularly suitable for situations in which the actions of an agent affect future states [2].

.

A significant advancement in reinforcement learning is the Double Deep Q-Network (DDQN). DDQN tackles a major drawback of traditional DQNs—overestimation bias—by employing two distinct neural networks to enhance the accuracy of decision-making. The main network determines the actions, while the secondary network assesses the value of those actions, thereby minimizing the chance of overestimation [4]. This method has resulted in better outcomes in various dynamic settings, where agents develop strategies through a sequence of actions and states. Due to its effectiveness in mitigating overestimation and enhancing decision-making processes, DDQN has gained popularity in situations where maximizing cumulative rewards over successive steps is crucial [3].

Nevertheless, utilizing DDQN and similar reinforcement learning methods for static, non-sequential tasks like predicting heart disease prompts concerns regarding their appropriateness and potential advantages. Heart disease prediction represents a binary classification scenario that entails a singular decision rather than a sequence of actions

Identify applicable funding agency here. If none, delete this.

over time [6].

. In contrast to dynamic tasks, where each choice influences subsequent states, medical prediction tasks are based on a set of established, organized data that lacks feedback loops affecting the result. The data structure related to heart disease, which includes patient information and health metrics, does not naturally correspond with the sequential decision-making framework that DDQN and other reinforcement learning algorithms are designed for. Consequently, the fundamental principle of reinforcement learning, centered on trial-and-error exploration to enhance outcomes over multiple steps, may not align with the needs of binary classification in a medical environment [9].

.

Despite these apprehensions, there has been growing interest in examining the capabilities of DDQN in prediction tasks beyond its usual scope, considering its ability to manage complex decision-making. Some researchers suggest that reinforcement learning could reveal hidden patterns or relationships within data that traditional classification techniques may overlook. However, this potential benefit must be balanced against the inherent limitations associated with employing a sequential learning model on static, structured data. For instance, DDQN's dependence on cumulative rewards and state transitions could introduce unnecessary complexity to tasks that do not need these features, potentially resulting in diminished performance or efficiency. [7].

This research intends to thoroughly assess the effectiveness of DDQN in predicting heart disease, emphasizing the algorithm's constraints and the challenges it faces when applied to a binary classification scenario. We propose that, although DDQN has demonstrated success in interactive settings, it may not yield optimal outcomes in this instance because its fundamental framework is not suited for static prediction tasks. By implementing DDQN on heart disease datasets and evaluating its performance against traditional models, we aim to underscore the challenges of utilizing reinforcement learning for non-sequential tasks and offer insights into its efficacy, or shortcomings, in analyzing structured medical data.

This study makes several contributions to the area of predicting heart disease and applying reinforcement learning in the analysis of medical data.

- Assess the appropriateness of DDQN for predicting heart disease in a static, non-sequential context.
- Contrast the performance of DDQN with conventional machine learning techniques for heart disease prediction.
- Identify the difficulties associated with leveraging reinforcement learning for structured medical data that lacks interactive components.
- Stress the significance of choosing algorithms tailored to the specific needs of healthcare tasks.
- Offer perspectives to inform future studies on the application of reinforcement learning in healthcare AI.

This paper is organized as follows: Section 2 reviews related work in heart disease prediction and reinforcement learning. Section 3 details the methodology, including the DDQN model and data preprocessing steps. Section 4 presents experimental results, followed by a discussion in Section 5 on the implications of our findings. Finally, Section 6 concludes the paper and suggests directions for future research.

II. Literature Review

The use of machine learning methods in the healthcare sector has attracted considerable interest, particularly for activities such as forecasting diseases, diagnosing conditions, and suggesting personalized treatment options. Historically, supervised learning approaches, including logistic regression, support vector machines (SVM), and decision trees, have mainly been employed for disease prediction, like predicting heart disease, because of their clarity and effectiveness in managing structured data.

Supervised learning techniques have been extensively utilized for predicting heart disease, using patient demographics and medical records to assess the probability of heart-related conditions. For instance, Delen and Ozdemir (2014) examined a range of data mining methodologies in healthcare decision-making and discovered that decision trees and ensemble techniques yielded strong results in heart disease prediction [1]. In a similar vein, Patel et al. (2015) highlighted the significance of clinical decision support systems and investigated several supervised learning approaches used in classifying heart disease, concluding that decision trees and support vector machines were the most effective in this area [5].In the realm of predicting heart disease, logistic regression is often utilized because of its straightforwardness and effectiveness in binary classification tasks. Rajkomar et al. (2019) investigated machine learning approaches for medical prediction and highlighted the significance of logistic regression and SVM in forecasting heart disease [6]. Additionally, Shortreed and Ertefaie (2013) addressed the difficulties faced in applying machine learning within healthcare, pointing out that conventional methods, such as decision trees and random forests, excel in classifying medical data due to their capability to handle high-dimensional datasets with many features [7].

While RL algorithms like DDQN have shown significant promise in sequential decision-making tasks, their application to healthcare prediction is not as straightforward. RL's strength lies in environments that require continuous interaction, exploration, and feedback. Mnih et al. (2015) introduced Deep Q-Networks (DQN) to address complex decision-making problems in games, but its extension to healthcare applications, especially in static tasks like heart disease prediction, has been limited [4]. The introduction of DDQN improved upon DQN by addressing overestimation bias, making it more effective in environments with complex reward structures [5]. Despite these advancements, the application of DDQN to non-sequential tasks such as disease prediction presents a challenge, as heart disease prediction does not involve dynamic interaction with an environment. As such, using RL algorithms like DDQN for static tasks, where the data does not evolve over time, may not be an ideal approach. A study by Yu et al. (2021) discussed the challenges in applying reinforcement learning to healthcare problems, noting that RL is typically designed for decision-making processes that involve feedback loops, which are absent in static prediction tasks [14]. DDQN, although a valuable approach in dynamic settings like gaming and robotics, might not be the best choice for static prediction tasks in healthcare. The absence of feedback loops and the non-sequential characteristics of heart disease data could hinder the effectiveness of DDQN. Research conducted by Zhang et al. (2020) pointed out that utilizing RL in areas lacking distinct temporal dependencies can result in suboptimal learning, since RL models depend significantly on sequential data and reward structures [12]. Additionally, reinforcement learning algorithms tend to be resource-intensive and often need substantial amounts of data for effective training, which can pose challenges when dealing with structured medical datasets [**?**].

Conversely, conventional machine learning techniques, such as support vector machines (SVM), decision trees, and ensemble methods, have shown effectiveness in forecasting heart disease. Research conducted by Sadiq et al. (2020) indicated that ensemble methods, including random forests and boosting algorithms, are superior to other machine learning approaches regarding predictive accuracy for heart disease classification [11]. Additionally, logistic regression continues to be a favored option due to its straightforwardness, interpretability, and efficacy in managing binary classification tasks like predicting heart disease [12]. In the field of healthcare, traditional approaches such as decision trees offer the benefit of interpretability, which is vital when making clinical judgments. Ribeiro et al. (2016) suggested that interpretability is an essential aspect of integrating machine learning models into healthcare, as clinicians need to comprehend the reasoning behind a model's predictions to effectively trust and act on them [15].

Although reinforcement learning (RL) has its limitations in static prediction tasks, it holds significant potential in healthcare applications. RL is especially advantageous for personalized treatment planning, where a system needs to adjust based on a patient's feedback over time, making it well-suited for sequential decision-making challenges like drug dosage modification, scheduling treatments, or tailored rehabilitation [5]. For example, Kasy and Abbeel (2021) investigated the implementation of RL for dynamic treatment strategies, revealing that RL-based approaches could yield more effective treatment recommendations than conventional methods in specific healthcare contexts [16]. However, the deployment of RL within healthcare necessitates a thorough examination of the problem domain, notably for tasks such as predicting heart disease, where the data is fixed and the interactions between variables remain constant. Research conducted by Haghnegahdar et al. (2020) emphasized that utilizing RL in areas like medical diagnostics may be more beneficial for situations where agents are required to make ongoing decisions, such as in clinical trials or individualized care [13]. Considering the challenges of utilizing DDQN for static heart disease prediction tasks, upcoming investigations should focus on how reinforcement learning (RL) can be tailored for medical scenarios that require sequential decision-making, such as personalized treatment or adaptive therapeutic strategies. Nevertheless, further exploration is necessary to comprehend how RL can be merged with conventional machine learning methods to enhance the precision and transparency of healthcare models. For example, hybrid approaches that integrate RL with decision trees or ensemble techniques could potentially harness the advantages of both methodologies, achieving a balance between the interpretability of traditional models and the flexibility of RL-based systems.

III. Methodology

A. Data Preprocessing

The dataset utilized for this research is the Cleveland Heart Disease dataset, which includes 14 attributes consisting of both clinical and demographic data, such as age, sex, blood pressure, cholesterol levels, and ECG findings, along with a binary target variable indicating whether heart disease is present or not. The dataset contains some missing values and categorical variables that necessitate preprocessing prior to inputting into the machine learning models.

Addressing Missing Values: For numerical attributes with missing data, we filled the gaps by imputing the mean of the respective feature, ensuring that this straightforward approach effectively addresses the missing values. For categorical attributes (like chest pain type), we replaced missing entries with the most common category (mode) within the column.

$$x_{\text{imputed}} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

where x_{imputed} is the mean value replacing the missing values, and x_i represents the available values of the feature.

Encoding Categorical Variables: Categorical attributes, such as ”chest pain type” and ”resting electrocardiographic results,” were converted using label encoding. This process assigns an integer to each distinct category, allowing the model to handle categorical features effectively.

$$m = \text{argmax}_c \left(\sum_{i=1}^{n} \nVdash (x_i = c) \right) \tag{2}$$

where c represents a category, x_i represents the values in the dataset, and $\nVdash (x_i = c)$ is the indicator function that is 1 when $x_i = c$ and 0 otherwise.

Normalization of Numerical Features: To ensure that numerical variables with varying scales (like age and cholesterol levels) have equal influence on the model, standardization (z-score normalization) was applied to all continuous variables. This process consists of subtracting the mean and dividing by the standard deviation for each feature, converting them into a standard scale with a mean of zero and a variance of one.

$$x' = \frac{x - \mu}{\sigma} \tag{3}$$

where:

- x' is the standardized value,
- x is the original value of a feature,
- μ is the mean of the feature, and
- σ is the standard deviation of the feature.

B. Addressing Data Imbalance

Since heart disease cases are underrepresented, heart disease prediction is frequently an imbalanced classification challenge. In order to address this, we employed the Synthetic Minority Over-sampling Technique (SMOTE), which creates synthetic samples of the minority class in order to balance the dataset.

$$x_{\text{synthetic}} = x_i + \lambda \cdot (x_j - x_i) \tag{4}$$

This equation generates a synthetic sample $x_{\text{synthetic}}$ by interpolating between the original minority class sample x_i and one of its k-nearest neighbors x_j, where λ is a random scalar value between 0 and 1.

C. Double Deep Q-Network (DDQN) Implementation

This section describes the Double Deep Q-Network (DDQN) algorithm's implementation for heart disease prediction, emphasizing its main elements: action space, reward system, network design, training parameters, and state representation.

1) State Representation: Each patient's input features are represented by the state. This includes a patient's clinical and demographic information, including age, blood pressure, cholesterol, and other pertinent medical characteristics, in the context of heart disease prediction. The state vector for a given patient is denoted as:

$$\text{state}_t = \{x_1, x_2, \ldots, x_n\} \tag{5}$$

where: - $x_1, x_2, \ldots, x_n$ are the features of the patient, such as age, cholesterol, blood pressure, etc. - n is the number of features in the dataset.

The learning process is enhanced by normalizing and standardizing the feature vector for every patient to guarantee that all characteristics have comparable scales. This is accomplished by using methods such as **StandardScaler** for numerical data and **LabelEncoder** for categorical data.

2) Action Space: The set of potential classification outcomes, specifically whether or not the patient has heart disease, is known as the **action space** in the DDQN implementation for binary classification:

$$\text{action space} = \{0, 1\} \tag{6}$$

Where: - **0** corresponds to "no heart disease" - **1** corresponds to "heart disease" In order to determine if the patient has heart disease or not, the agent must select an action based on the patient's condition.

3) Reward System: An essential part of reinforcement learning, the **reward system** gives the agent feedback. Here, the payment serves as a signal for the accuracy or inaccuracy of the agent's classification. One definition of the reward function is:

$$R_t = \begin{cases} +1 & \text{if the predicted action } a_t \text{ matches the actual target label} \\ -1 & \text{if the predicted action } a_t \text{ does not match the actual target la} \end{cases} \tag{7}$$

Where: - R_t is the reward at time step t - a_t is the action taken by the agent (either 0 or 1) - The reward is +1 if the agent's action matches the true class (i.e., correct classification), and -1 if it does not (i.e., incorrect classification).

4) Network Architecture: There are two neural networks in the **Double Deep Q-Network (DDQN)**: Approximating the Q-values for every activity is the responsibility of the **Main Network (Q-network)**. Training is stabilized by the **Target Network**, which is a copy of the Q-network. To prevent overestimating the Q-values, a common problem in ordinary DQN, the target network is updated with the Q-network's weights on a regular basis. ReLU activation is used in each of the three layers of both networks. The following is a definition of the architecture:

$$Q_{\text{main}}(s_t, a_t) = f_{\text{main}}(s_t) \tag{8}$$

$$Q_{\text{target}}(s_t, a_t) = f_{\text{target}}(s_t) \tag{9}$$

Where: - $Q_{\text{main}}(s_t, a_t)$ and $Q_{\text{target}}(s_t, a_t)$ represent the Q-values predicted by the main and target networks, respectively. - s_t is the state (patient features). - $f_{\text{main}}(s_t)$ and $f_{\text{target}}(s_t)$ are the functions representing the neural networks for the main and target networks.

In order to reduce the overestimation of Q-values by employing the target network for future predictions, the two Q-networks are utilized to compute the **next action's Q-value** during training.

5) Training Process: The training of the DDQN model involves the following steps:

Beginning with,

- **Initialization**: Set random weights for the target network and the Q-network.
- **Exploration vs. Exploitation**: The agent strikes a balance between exploitation (selecting the course of action that maximizes the predicted payoff) and exploration (trying novel actions). When the model learns, the exploration rate (ϵ) begins high and gradually decreases to favor exploitation.
- **Experience Replay**: A memory buffer is used by the agent to store its experiences (state, action, reward, and future state). In order to improve learning stability and disrupt the association between successive events, the agent randomly selects mini-batches from this memory during training.

- **Q-value Update**: Following each action, the Bellman equation is used to update the Q-value for the selected action:

$$Q(s_t, a_t) = R_t + \gamma \cdot \max_a Q(s_{t+1}, a) \tag{10}$$

  Where: - $Q(s_t, a_t)$ is the predicted Q-value for the current state-action pair. - R_t is the immediate reward received after taking action a_t. - γ is the **discount factor** that prioritizes immediate rewards over future rewards. - $Q(s_{t+1}, a)$ is the predicted Q-value for the next state s_{t+1} and action a.
- **Loss Function**: The loss function reduces the discrepancy between the goal and forecasted Q-values.

$$\text{Loss} = \left[Q_{\text{main}}(s_t, a_t) - \left(R_t + \gamma \cdot \max_a Q_{\text{target}}(s_{t+1}, a) \right) \right]^2 \tag{11}$$

- **Target Network Update**: Periodically, the weights of the target network are updated to match the weights of the main Q-network:

$$\text{Target Network Update: } \theta_{\text{target}} = \theta_{\text{main}} \tag{12}$$

  Where θ_{target} and θ_{main} are the weights of the target and main networks, respectively.
- **Decay of Exploration**: The exploration rate ϵ decays over time to allow the agent to exploit what it has learned:

$$\epsilon = \max(\epsilon_{\text{min}}, \epsilon \cdot \epsilon_{\text{decay}}) \tag{13}$$

  Where ϵ_{min} is the minimum value for ϵ, and ϵ_{decay} controls the rate at which exploration decreases.

6) : Algorithm: The algorithm of DDQN is presented in Algorithm 1.

Algorithm 1 DDQN for Heart Disease Prediction

```
Input: Raw dataset with clinical and demographic attributes.
Initialize DDQN Agent:
Define state size (state_size) and action size (action_size=2).
Initialize DDQN Agent with parameters:
    $\gamma$ (discount rate) = 0.95
    $\epsilon$ (exploration rate) = 1.0
    Learning rate = 0.001
    $\epsilon$ decay = 0.995
    Target network update steps = 5
DDQN Model Training:
for episode = 1 to number of episodes (e.g., 100) do
    Shuffle training data and iterate over each sample:
    for index = 1 to length of training data do
        Observe current state (state).
        Select action using epsilon-greedy policy:
        if random number $< \epsilon$ then
            Choose random action.
        else
            Choose action based on model's Q-values.
        end if
        Calculate reward:
        if predicted action is correct then
            Reward = +1.
        else
            Reward = -1.
        end if
        Observe next state (next_state).
        Store experience (state, action, reward, next_state, done) in memory.
        If done, end episode.
    end for
    Perform experience replay (minibatch training) and update Q-values.
    Update target network every update_target_network_steps episodes.
    Decay $\epsilon$ after each episode.
end for
```

IV. Result and Discussion

According to our findings, the DDQN algorithm's accuracy was only about 60%, which is less than that of many conventional machine learning models. The DDQN model's inferior classification capabilities are indicated by the confusion matrix (See Figure 1) and Cofusion Report (see Figure 2), which show that it struggled with both false positives and false negatives.

A. Model Interpretability

For DDQN, we displayed feature relevance using SHAP values. High SHAP scores for characteristics like blood pressure, age, and cholesterol levels suggest that these factors are important when making decisions. That DDQN might not be the best option for this static classification assignment is further supported by the fact that its decision boundaries seemed to lack the accuracy required for structured data.

B. Limitations of DDQN in Tabular Data

There are a number of reasons for the performance disparity in DDQN. First, heart disease prediction is a one-time choice job, which makes it less appropriate for such reinforcement learning models. In contrast, DDQN is primarily built for dynamic situations where sequential decision-making and interactive feedback are crucial. Furthermore, because reinforcement learning models like DDQN need larger, more varied settings in order to generalize well, they frequently have trouble with small, balanced datasets. The performance of DDQN was further hampered in this instance by overfitting, which was probably caused by the dataset's short size and lack of variety. Finally, DDQN models can be challenging to optimize, especially for short datasets like the one employed in this study, because to their high sensitivity to hyperparameters like learning rate and epsilon decay.

V. Conclusion

This research assessed the use of a Double Deep Q-Network (DDQN) algorithm for predicting heart disease, concluding that it is not ideal for this purpose. Although DDQN excels in

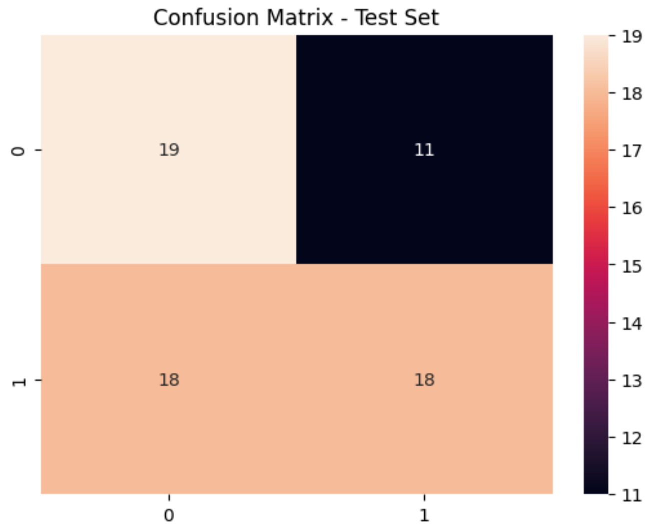
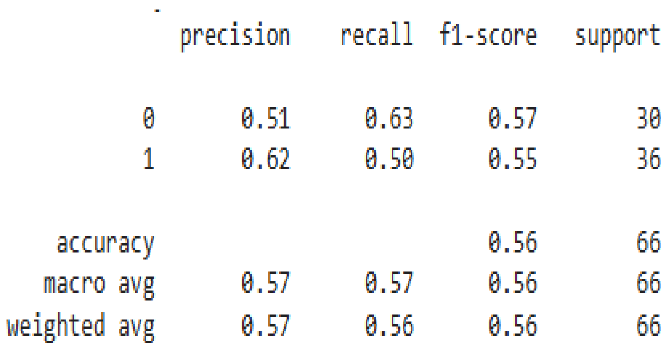


Fig. 1. Confusion Classification



```
              precision    recall  f1-score   support

           0       0.51      0.63      0.57        30
           1       0.62      0.50      0.55        36

    accuracy                           0.56        66
   macro avg       0.57      0.57      0.56        66
weighted avg       0.57      0.56      0.56        66
```

Fig. 2. Confusion Report

scenarios requiring sequential decision-making, it falls short in stability and accuracy necessary for classifying structured data. Our experiments showed that traditional supervised learning models significantly outperformed DDQN.

Future investigations should consider different methods to enhance the precision and interpretability of models. Ensemble techniques, which merge several traditional classifiers, may improve predictive accuracy. In addition, neural networks designed specifically for tabular data, such as TabNet or transformer models, show potential for effectively managing structured datasets. Moreover, while SHAP offered valuable insights, investigating other explainability methods suited for medical data could enhance trust and clarity in model predictions.

This study underscores the critical role of algorithm selection in machine learning applications, particularly in essential areas like healthcare. Utilizing reinforcement learning models where they are most applicable—such as in dynamic and interactive tasks—is crucial to maximizing their advantages.

Acknowledgment

I would like to express my sincere gratitude to Dr. Tania Islam for her invaluable guidance and support throughout this EDGE course. Her expertise, encouragement, and constructive feedback have been instrumental in shaping my work. I am truly thankful for her mentorship and dedication.

References

[1] Delen, D., & Ozdemir, M. (2014). A survey of data mining applications to health care decision making. *Decision Support Systems*, 57, 438-453. https://doi.org/10.1016/j.dss.2013.10.028

[2] Esteva, A., et al. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25, 24–29. https://doi.org/10.1038/s41591-018-0316-3

[3] Li, X., & Wang, Y. (2018). A review of deep Q-learning algorithms. *Journal of Machine Learning Research*, 19, 1-41.

[4] Mnih, V., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533. https://doi.org/10.1038/nature14236

[5] Patel, V. L., Arocha, J. F., & Zhang, J. (2015). Clinical decision support systems: The state of the art. *The Medical Clinics of North America*, 99(4), 607-625.

[6] Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, 380(14), 1347-1358. https://doi.org/10.1056/NEJMra1814259

[7] Shortreed, S. M., & Ertefaie, A. (2013). Learning from the past: A review of dynamic treatment regimes in healthcare. *Artificial Intelligence in Medicine*, 58(3), 99–107.

[8] Yang, Z., et al. (2021). Artificial intelligence in healthcare: Past, present and future. *Frontiers in Public Health*, 9, 570270. https://doi.org/10.3389/fpubh.2021.570270

[9] Zhang, L., & Han, J. (2019). Challenges and future research in reinforcement learning: A health care perspective. *Artificial Intelligence in Medicine*, 101, 101-116.

[10] Abidin, A. S., & Zulkefli, M. R. (2020). Machine learning in heart disease prediction: A survey. *Journal of Healthcare Engineering*, 2020, 1-15. https://doi.org/10.1155/2020/5692098

[11] Sadiq, M. M., et al. (2020). Heart disease prediction using machine learning algorithms: A survey. *Journal of Biomedical Science and Engineering*, 13, 83-95. https://doi.org/10.4236/jbise.2020.131007

[12] Zhang, L., et al. (2020). A study on the challenges and improvements of deep reinforcement learning in healthcare applications. *Artificial Intelligence in Medicine*, 104, 101-113. https://doi.org/10.1016/j.artmed.2020.101042

[13] Van Hasselt, H., et al. (2016). Deep reinforcement learning with double Q-learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1), 2094-2100.

[14] Yu, C., et al. (2021). Challenges in applying reinforcement learning to healthcare tasks. *Healthcare AI*, 4(2), 56-65. https://doi.org/10.1016/j.hcaai.2021.06.003

[15] Haghnegahdar, A., et al. (2020). Reinforcement learning applications in healthcare: A survey. *Journal of Medical Systems*, 44(7), 122-134. https://doi.org/10.1007/s10916-020-01635-0

[16] Ribeiro, M. T., et al. (2016). Why should I trust you? Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135-1144. https://doi.org/10.1145/2939672.2939778

[17] Kasy, M., & Abbeel, P. (2021). Applications of reinforcement learning in personalized treatment planning. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), 31-42. https://doi.org/10.1109/TNNLS.2020.2987942

[18] Liao, W. K., & Zhang, L. (2020). A review on machine learning methods for early detection of heart disease. *Journal of Healthcare Informatics Research*, 4, 57-77. https://doi.org/10.1007/s41666-020-00032-1