COVID-19 Vaccine Distribution Policy Design with Reinforcement Learning (Unpublished Version)

Pu Tan*

Abstract : The COVID-19 pandemic has put immense strain on healthcare systems globally, with vaccines seen as the most viable way to curb the outbreak. Given the limited resources available for vaccine distribution, there is an urgent need for a data-driven approach to optimize allocation. This study proposes a dual-model pipeline, focusing specifically on the city of San Diego. Initially, a Random Forest model predicts daily new COVID-19 cases using vaccine distribution data as inputs. Following this, a Deep Q-Network (DQN) model designs a daily allocation strategy for three different types of vaccines with the objective of minimizing new infections. For robustness and generalizability, we employ hyperparameter tuning and feature engineering techniques. Our model was validated using real-world datasets from San Diego, confirming its effectiveness in reducing new confirmed cases. This work contributes to current understandings of optimal vaccine distribution and offers a reliable model for policy decision-making.

Key words: coronavirus; COVID-19; multi-class classification; random forest; machine learning, Reinforcement Learning, Vaccine, computational methodologies

1 Introduction

The COVID-19 pandemic reached the U.S. in early 2020, with 35.6 million confirmed cases nationally and 4.1 million in California as of August 2021. The spread of the virus is exacerbated by high population density, especially in urban areas. Vaccination remains the primary method to combat the pandemic, with varying degrees of effectiveness—86% for fully vaccinated individuals according to a study by the University of California at Berkeley.

California administers three types of vaccines: Moderna, Pfizer-BioNTech, and Johnson & Johnson's Janssen, each with different efficacies and costs. As of August 2021, California has received 50.9 million doses and administered 46.1 million. The state faces a challenge in optimizing vaccine distribution to meet local preferences, affecting supply-demand dynamics and potentially worsening the pandemic. In San Diego, California, despite robust vaccine production, limitations in transportation and medical staffing create a bottleneck in vaccine delivery. This paper aims to optimize the distribution of three types of COVID-19 vaccines: Moderna, Pfizer-BioNTech, and Johnson & Johnson’s Janssen, given a fixed daily supply.

The study focuses on minimizing new COVID-19 cases from a California governmental perspective, disregarding industrial production costs. Previous research using machine learning and reinforcement learning has shown promise in COVID-19 forecasting but is generally not tailored to local conditions in California or specific vaccine types. Moreover, existing models often overlook variables like political decisions and community mobility, which can vary from county to county.

Objectives of This Study:

(1) Optimize the distribution strategy of three authorized COVID-19 vaccines—Moderna, Pfizer-BioNTech, and Johnson & Johnson's Janssen—in San Diego, California.

(2) Accurately predict the daily new confirmed COVID-19 cases using a Random Forest prediction model.

(3) Design an effective daily allocation ratio for the three types of vaccines with the aim of minimizing new confirmed COVID-19 cases, using a Deep Q-Network (DQN) model.

(4) Compare the effectiveness of the hybrid DQN-Random Forest model with a baseline strategy of uniform distribution to gauge improvements.

2 Datasets

2.1 Methodology and Dataset Characteristics

The study focuses on the spread of the COVID-19 virus, a novel coronavirus responsible for causing respiratory illnesses with varying severity. To better understand the impact and efficacy of different vaccine distribution strategies, we sourced our dataset from multiple reliable repositories. The dataset used in this research is focused on San Diego, California, and encompasses various parameters related to the COVID-19 pandemic.

The dataset contains 2342 samples covering the period from December 15, 2020, to July 23, 2021, and has been derived from both The California Department of Public Health (CDPH) and the San Diego Association of Governments. Table 1 provides an overview of the attributes or features considered in this study, while a snapshot of the dataset is presented in Table 2.

2.2 Feature Selection

Feature selection is a pivotal step in model building to ensure the selection of the most relevant predictors, thereby reducing the model's complexity. We employed the Random Forest Importance Algorithm, implemented in R, to select the most significant attributes for prediction.

Initially, the dataset contained multiple features such as "Observation Date," "Time," "State/Union Territory," "Confirmed Indian National," and "Confirmed Foreign National." However, for this study, we narrowed our focus to a carefully curated subset of features pertinent to San Diego, California. These are included in Table 1:

Some features were omitted as they were only relevant to the early stages of the pandemic when international transmission was a key factor. As the scope of our study is localized to California and primarily internal transmission dynamics, these variables were not considered.

Table 1 delineates the specific attributes or features utilized for the analyses in this study.

Table 1. Features used in the prediction and vaccine distribution models.

Column	Description
date	It is the date on which how many COVID-19 positive cases have occurred.
confirmed	It is the total number of confirmed COVID-19 cases found in California at the starting of SARS-CoV-2.
county	It is the name of the county in California where COVID-19 cases were found. San Diego is used in this study.
pfizer_doses	Number of Pfizer vaccine doses shipped on a particular date.
cumulative_pfizer_doses	Cumulative number of Pfizer vaccine doses shipped up to a particular date.
moderna_doses	Number of Moderna vaccine doses shipped on a particular date.
cumulative_moderna_doses	Cumulative number of Moderna vaccine doses shipped up to a particular date.
jj_doses	Number of Janssen vaccine doses shipped on a particular date.
cumulative_jj_doses	Cumulative number of Janssen vaccine doses shipped up to a particular date.

2.3 Daily Confirmed Cases and Vaccine Doses: An Overview

Figure 2: Represents daily newly confirmed COVID-19 cases in San Diego.

Figure 3: Displays the daily administration of three authorized vaccines: Moderna, Pfizer-BioNTech, and Johnson & Johnson's Janssen.

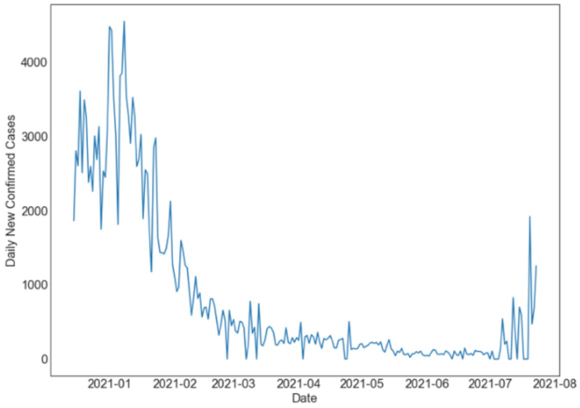

Figure 2. Daily new confirmed cases in San Diego.

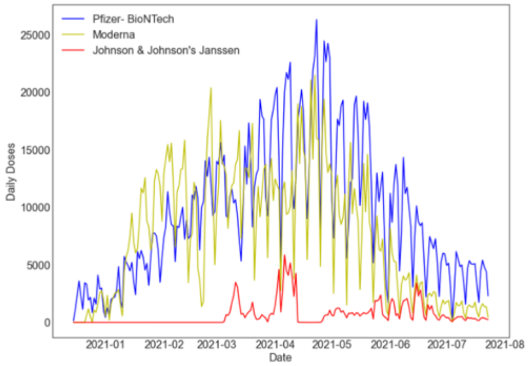

Figure 3. The daily doses of three types of vaccines.

3 Prediction Model

3.1 Predictive Modeling Framework

The predictive modeling framework employed in this study is primarily grounded in Random Forest (RF) algorithms, an ensemble learning method employing decision trees as base estimators. This choice is motivated by the algorithm's robust performance in diverse machine learning tasks, including those that involve nonlinear relations between variables.

Step 1: Dataset Partitioning for Model Training and Testing

The comprehensive dataset, as elaborated in Section 2, is bifurcated into training and test sets according to an 80:20 ratio. Subsequent to this partitioning, both Random Forest and two baseline models—Linear Regression (LR) and AdaBoost—are instantiated for comparative evaluation. These models are implemented using the scikit-learn library, a standard Python package for machine learning applications.

Step 2: Feature Selection

Prior to model instantiation, a feature selection process is executed to minimize model complexity. As delineated in Section 2.2, a Random Forest Importance algorithm is employed for this purpose. The selected features function as input parameters for the predictive model, serving to forecast the following:

$$\text{Confirmed Cases} = f\begin{pmatrix}\text{Date, Confirmed Cases, County,}\\ \text{Three Types of Doses, Cumulative Doses}\end{pmatrix}$$

Step 3: Model Training with Multi-Class Classification

Upon feature selection, the dataset undergoes a training phase employing multi-class classification algorithms. Random Forest, Support Vector Machine (SVM), Decision Tree, Multinomial Logistic Regression, and Neural Networks are applied to the training set, which constitutes 80% of the complete dataset.

Step 4: Model Testing and Evaluation

The testing phase encompasses the remaining 20% of the dataset. Performance metrics, specifically Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), are employed for model evaluation. The mathematical formulations of these metrics are as follows:

$$\text{RMSE} = \sqrt{\frac{1}{\text{n}}\sum_{\text{i}=1}^{\text{n}}(\text{x}_\text{i} - \widehat{x_i})^2}$$

$$\text{MAE} = \frac{\sum_{i=1}^{N}|x_i - \widehat{x_i}|}{N}$$

Where N is the number of samples, x_actual are the true values, and x_predicted are the predicted values.

MAE serves as the primary evaluation metric for comparing the effectiveness of all considered models. Upon evaluation, it was ascertained that the Random Forest model yielded the minimum MAE, thereby outperforming the baseline models.

In conclusion, the Random Forest model demonstrates superior predictive accuracy and is thus selected as the most suitable algorithm for forecasting COVID-19 case numbers, informing the optimal vaccine distribution strategy.

Table 2. Performance of the machine learning models for prediction.

Model	Random Forest	Linear Regression	AdaBoost
RMSE	404.52	1026.07	374.90
MAE	193.50	984.23	227.89

3.2. Random Forest Model

The Random Forest (RF) algorithm [12], an ensemble machine learning approach utilizing decision trees as base estimators, is selected for this study. The model operates on the principle of bagging, a parallel ensemble technique. Decision trees are trained on various subsets of the training dataset and are subsequently aggregated—typically through voting or averaging—to produce a final output. This method has demonstrated efficacy in numerous machine learning applications. In the context of this study, the RF model is employed to predict the daily number of new COVID-19 cases based on historical vaccine data. Specifically, the model uses the last five days' confirmed cases to forecast the subsequent day’s case count. This forecasted data serves as the input for the Deep Q-Network (DQN) model and informs the parameters for the reward function in vaccine distribution optimization.

3.3. Data Partition and Model Implementation

The comprehensive dataset delineated in Section 2 is partitioned into training and testing subsets with an 80:20 split ratio. Alongside the RF model, two baseline models—Linear Regression (LR) [13] and AdaBoost [14]—are also implemented utilizing the Python package, scikit-learn. The LR model, although straightforward, lacks the capacity to encapsulate nonlinear relationships between input features and the prediction target. In contrast to the bagging method employed in RF, AdaBoost operates through boosting. Boosting trains base estimators sequentially, adjusting for the errors made by preceding models.

3.4. Performance Evaluation Metrics

Upon training, the models are evaluated using two key metrics: Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). The performance outcomes of these machine learning models are summarized in Table 2, with specific definitions accompanying the metrics.

4 Vaccine Distribution Model

4.1 DQN Model Overview

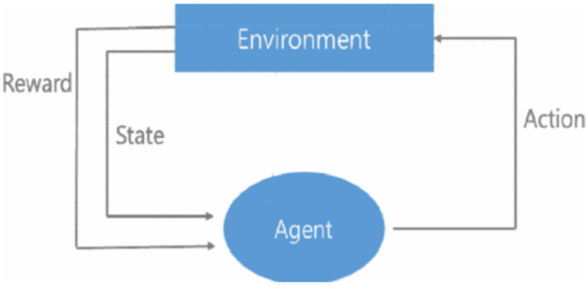

Figure 4. The scheme of reinforcement learning problems.

To develop an optimal vaccine distribution strategy for three distinct types of vaccines, we have employed a Deep Reinforcement Learning (DRL) algorithm, specifically the Deep Q-Networks (DQN), utilizing a deep neural network as its underlying architecture. This DQN model was then juxtaposed with a baseline equal distribution model to evaluate its effectiveness.

The DQN model implemented in our research is structurally akin to those delineated in prior studies, as referenced in [9]. Figure 4 encapsulates the essential workflow of our reinforcement learning algorithm. At its core, the DQN model operates on the principles of a Markov Decision Process

(MDP), wherein the cumulative return value is formally described by the Q-value function. Mathematically, this is expressed as:

$$Q(s,a) = E\left[r(s,a) + \gamma \max_{a'} Q(s',a')\right]$$

Here, (s) represents the state, (a) signifies the action undertaken by the agent, and (γ) is the discount factor. The iterative updating of the Q-value occurs according to the following equation:

$$Q_{i+1}(s,a) = Q_i(s,a) + \alpha\left[r_i(s,a) + \gamma \max_{a'} Q_i(s',a') - Q_i(s,a)\right]$$

In this equation, $(\alpha)$is the learning rate parameter that governs the speed of convergence. A higher value of (α) promotes faster convergence, whereas a lower value aids in more comprehensive exploration. Finally, a greedy strategy $\alpha^* = \arg\max Q^*(s,\alpha)$ is used to decide the optimal action for each state,

Our empirical analysis subsequently generates and compares optimal distribution strategies for the aforementioned vaccines within the San Diego District. The aim is to establish a robust, data-driven approach to vaccine distribution, leveraging advanced machine learning techniques to navigate the complexities inherent to public health logistics.

More specific procedure is attached in the appendix.

4.2 Epidemiological Assumptions and Objectives

Epidemiological efficiency is assessed based on the trends in newly confirmed cases. An increase in such cases is indicative of lower effectiveness of the implemented anti-epidemic measures, while a decrease signifies higher efficiency.

4.3 Reward Function Formulation

4.3.1 Mathematical Representation

The reward function is mathematically defined as

$$Reward = 100 \times e^{-(\frac{confirmed}{\alpha})}$$

, where α is a tunable parameter, affecting the sensitivity of the reward to changes in the number of new cases.

Behavior at Extremes

1) Zero Newly Confirmed Cases: In a scenario where the number of newly confirmed cases is zero, the reward function reaches its upper limit. Mathematically, as the number of new cases approaches zero, the exponential term approaches one, making the reward 100×1=100. This signifies the highest efficiency of the anti-epidemic measures in place and encourages the model to pursue actions leading to this outcome.
2) Infinite Newly Confirmed Cases: Conversely, if the number of newly confirmed cases tends toward infinity, the exponential term approaches zero. Therefore, the reward function would yield a value close to zero. This serves as a penalty, indicating the complete failure of the anti-epidemic measures and discouraging the model from taking similar actions in the future.

By setting such extreme values, the reward function is designed to sharply distinguish between highly effective and ineffective anti-epidemic measures, thus guiding the DQN model toward optimal decision-making.

4.3.2 Parameter Optimization

The choice of the α parameter is crucial for the optimal functionality of the DQN network. A comparative analysis of reward curves with α values of 100, 1000, 3000, and 4000 was undertaken. An α value of 3000 exhibited the most significant variations in the reward curve relative to the number of confirmed cases and was thus selected.

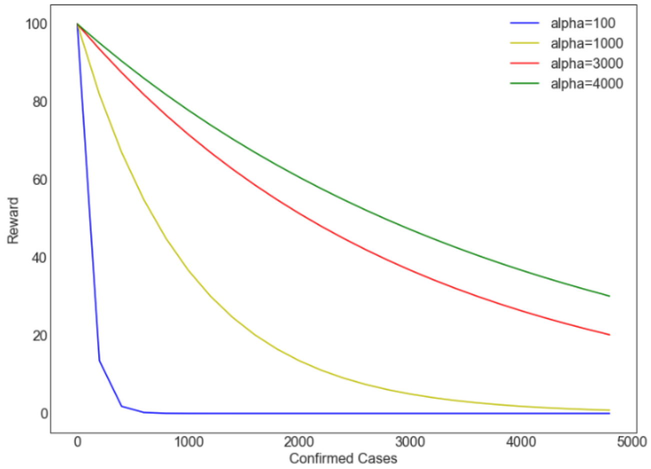

Figure 5. The comparison of the value choices of the parameter alpha in the reward function.

4.4 Model Inputs and Actions

The input layer of the neural network utilizes prior states related to the cumulative doses of Pfizer, Moderna, and Johnson & Johnson vaccines, along with the action taken, current states, and previously computed rewards. Actions are represented in terms of vaccine allocation ratios, denoted in units of 10%.

For example, the allocation plan of [0.0, 0.0, 1.0] represents all vaccine doses are J&J doses; the allocation plan of [0.0, 0.9, 0.1] represents all vaccine doses are distributed in the proportion of 0% Pfizer doses, 10% Moderna doses, and 90% J&J doses; the allocation plan of [0.0, 0.2, 0.8] represents all vaccine doses are distributed in the proportion of 0% Pfizer doses, 20% Moderna doses, and 80% J&J doses.

4.5 Neural Network Architecture and Training

The study employs a fully-connected neural network, the details of which are depicted in Figure 6.

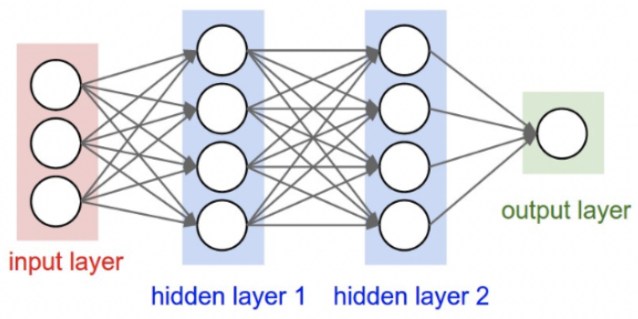

Figure 6. The fully connected neural network.

4.5.1 Computational Framework and Implementation Details

1) Input-Output Configuration: In our model, vaccine-related data serve as the input to a fully-connected neural network. The Q-value is subsequently estimated as the output, corresponding to various potential actions.

2) Iterative Learning and Ratio Adjustment: During each iteration, the vaccine allocation ratios for Pfizer, Moderna, and Johnson & Johnson are dynamically adjusted. The network is trained to minimize loss, using the reward function as feedback to select the action maximizing said reward.

3) Role of Reward Mechanism in Learning: The reward, as derived from the reward function, is utilized as a feedback mechanism for the neural network. This facilitates the automatic selection of the optimal action, identified by the action that maximizes the Q-value.

4) Technical Implementation: Our DQN model was realized using the stable_baselines3 library, a Python package that simplifies the implementation of reinforcement learning algorithms. For model training, 80% of the dataset was utilized, reserving the remaining 20% for the final evaluation of the model's performance.

4.6 Evaluation Metrics and Baseline Comparison

4.6.1 Baseline Allocation Strategy

As a baseline for comparison, we adopted an evenly distributed vaccine allocation strategy, denoted mathematically as [1/3, 1/3, 1/3], given a fixed total number of vaccines.

4.6.2 Model Performance

We evaluated our model by predicting the number of confirmed COVID-19 cases under both the DQN-derived allocation strategy and the baseline strategy. Our findings, illustrated in Figure 7, show a discernible reduction in the number of confirmed cases under the DQN model.

4.6.3 Adaptability of DQN Model

The DQN model is designed to update the vaccine allocation strategy on a daily basis, thereby enhancing its capacity to adapt and optimize the distribution ratios in real-time. This results in a more effective control of the pandemic situation as evidenced by the lower number of confirmed cases.

4.7 Results

As illustrated in Figure 7, the trajectory of confirmed COVID-19 cases under the allocation strategy devised by the Deep Q-Network (DQN) model (represented by the blue curve) consistently demonstrates lower numbers compared to the baseline strategy of evenly-distributed vaccine allocation (represented by the red curve). The inherent flexibility of the DQN model enables daily adjustments to the vaccine allocation ratios, thereby optimizing distribution strategies in real-time. Consequently, the model exhibits enhanced efficacy in controlling the proliferation of confirmed cases, underscoring its superiority over a static, evenly-distributed allocation plan.

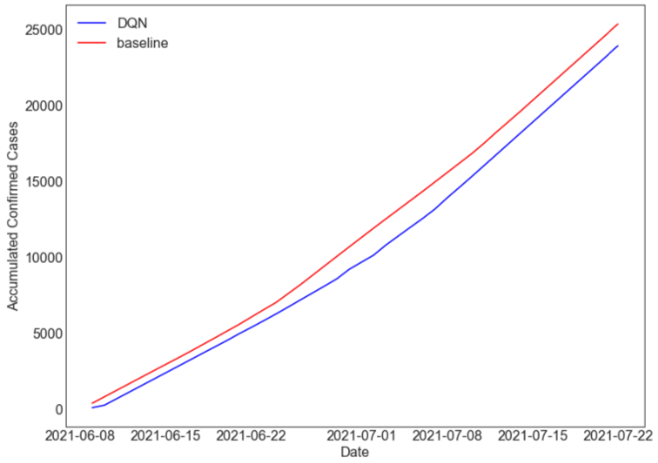

Figure 7. The comparison of accumulated confirmed cases with DQN and the baseline.

5 Conclusion

This research constitutes a multifaceted approach to mitigating the spread of the COVID-19 pandemic through the synergistic application of machine learning and reinforcement learning algorithms. Initially, a predictive model utilizing Random Forest algorithms was constructed to provide accurate forecasts of daily new confirmed cases, employing vaccine distribution data as key predictive variables. Subsequent to this, a Deep Q-Network (DQN) model was developed to optimize the daily allocation ratios of three different types of COVID-19 vaccines—Pfizer, Moderna, and Johnson & Johnson—targeting the minimization of new confirmed cases as the primary objective function.

The empirical analysis was carried out on real-world datasets sourced from San Diego County, and the findings substantiate the efficacy of the integrated computational framework. The proposed pipeline effectively amalgamates prediction and optimization processes, thereby leading to optimized vaccine distribution strategies. This not only enhances the adaptability of the healthcare system in response to the rapidly evolving pandemic but also demonstrates a marked reduction in the number of new confirmed cases.

In summary, the research offers a viable and data-driven methodology for not only predicting the trajectory of COVID-19 but also for strategically allocating resources to counter its spread, thereby demonstrating the robustness and versatility of machine learning and reinforcement learning techniques in public health crises.

References

[1] Chimmula V K R, Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks[J]. Chaos, Solitons & Fractals, 2020, 135: 109864.

[2] Shastri S, Singh K, Kumar S, et al. Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study[J]. Chaos, Solitons & Fractals, 2020, 140: 110227.

[3] Zeroual A, Harrou F, Dairi A, et al. Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study[J]. Chaos, Solitons & Fractals, 2020, 140: 110121.

[4] Dairi A, Harrou F, Zeroual A, et al. Comparative study of machine learning methods for COVID-19 transmission forecasting[J]. Journal of Biomedical Informatics, 2021, 118: 103791.

[5] Jiang W, Zhang L. Geospatial data to images: A deep-learning framework for traffic forecasting[J]. Tsinghua Science and Technology, 2018, 24(1): 52-64.

[6] Jiang W. Applications of deep learning in stock market prediction: recent progress[J]. Expert Systems with Applications, 2021: 115537.

[7] Jiang W. Internet Traffic Prediction with Deep Neural Networks[J]. Internet Technology Letters. 2021; e314.

[8] Ghostine R, Gharamti M, Hassrouny S, et al. An extended seir model with vaccination for forecasting the covid-19 pandemic in

saudi arabia using an ensemble kalman filter[J]. Mathematics, 2021, 9(6): 636.

[9] Awasthi R, Guliani K K, Khan S A, et al. VacSIM: Learning Effective Strategies for COVID-19 Vaccine Distribution using Reinforcement Learning[J]. arXiv preprint arXiv:2009.06602, 2020.

[10] https://data.chhs.ca.gov/dataset/vaccine-progress-dashboard

[11] https://sdgis-sandag.opendata.arcgis.com/datasets/vaccine-rate-by-zip-code/explore

[12] Breiman L. Random forests[J]. Machine learning, 2001, 45(1): 5-32.

[13] [Bishop C M. Pattern recognition[J]. Machine learning, 2006, 128(9).

[14] Freund Y, Schapire R E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. Journal of computer and system sciences, 1997, 55(1): 119-139.

Appendix

DQN:

Step-by-Step Method

1. Algorithm Selection: Choose Deep Q-Networks (DQN), a Deep Reinforcement Learning (DRL) algorithm, to develop the vaccine distribution model.

2. Baseline Model: Employ a basic "equal distribution" model as a baseline for comparison with the DQN model.

3. State and Action Definitions: The states include previous data on cumulative doses for Pfizer, Moderna, and J&J vaccines. Actions correspond to the distribution ratios of these vaccines.

4. Reward Function: Define the reward as a function of the change in newly confirmed COVID-19 cases. Use an exponential function with an adjustable parameter $(\alpha)$to compute the reward.

5. Parameter Tuning: Test different values for (α) (100, 1000, 3000, 4000) to find the most effective setting. Based on reward curves, select $(\alpha = 3000)$ as the optimal parameter.

6. Neural Network Design: Use a fully connected neural network to estimate Q-values. The network is trained iteratively to minimize the loss.

7. Training and Testing: Use 80% of the dataset for training and reserve 20% for evaluation. Implement the model using the stable_baselines3 Python package.

8. Performance Evaluation: Compare the DQN model with the equal distribution model by tracking confirmed cases over time.

9. Result Analysis: Assess whether the DQN model outperforms the baseline by reducing the number of confirmed cases more effectively.

We employ a Deep Q-Network (DQN), a specialized form of Deep Reinforcement Learning (DRL), to optimize vaccine distribution for three vaccine types: Pfizer, Moderna, and Johnson & Johnson. As a comparative measure, an equal distribution model is implemented as a baseline.