Dynamic Elderly Care Resource Allocation using SHBM-Tuned Deep Q-Networks (SHBM-RDQN)

Hua Jiang
School of social work, China Civil Affairs University, Beijing, 102600, China
E-mail: jianghua_jh630@outlook.com

Keywords: elderly care services, resource allocation, staff optimization, health management systems, dynamic honeybees mating-tuned resource based Deep Q-Network (DHBM-RDQN)

Received: August 29, 2025

The growing demand for elderly care services necessitates intelligent, efficient, and adaptive resource management strategies. This research presents a dynamic allocation strategy for optimizing elderly care resources using reinforcement learning (RL), addressing the complexity of nursing home resource management. This research proposes a Dynamic Honeybees Mating–tuned Resource-based Deep Q-Network (DHBM-RDQN) for elderly care resource allocation. The approach models the care environment as a Markov Decision Process (MDP), where states capture patient acuity levels, staff availability, daily admissions, and care workload, while actions correspond to dynamic allocation of nursing staff and support resources. The system uses a comprehensive Elderly Care Staffing & Quality Dataset, comprising time-series records at daily or shift-level granularity from multiple long-term care facilities. Preprocessing includes Z-score normalization and missing value imputation, while feature extraction applies Independent Component Analysis (ICA) and Discrete Wavelet Transform (DWT) to capture latent health patterns and temporal fluctuations. The DHBM-RDQN employs a Deep Q-Network with two fully connected layers (256 and 128 neurons) and ReLU activations, trained using the Adam optimizer. The Honeybee Mating optimization layer dynamically tunes learning rate, exploration parameters, and reward weights to prevent premature convergence. Experimental evaluation were implemented in python. Results show 96.5% accuracy, 90.5% resource efficiency, 1.0 s response time, and an adaptability score of 0.925, demonstrating robust adaptability under fluctuating patient demand and staffing. This research introduces a novel RL framework combining deep learning and bio-inspired optimization to achieve superior performance, rapid decision-making, and improved care quality in dynamic long-term elderly care environments.

Povzetek: Raziskava predstavlja dinamično strategijo razporejanja virov v domovih za starejše z okrepljevalnim učenjem (RL) in bio-navdihnjeno optimizacijo, ki izboljša učinkovitost, prilagodljivost ter kakovost oskrbe ob nihanjih povpraševanja in razpoložljivosti osebja.

1 Introduction

One of the major problems facing the world today is the aging population particularly in the developing countries whereby by 2050, the numbers of individuals aged 65 years and above are projected to have doubled. This demographic transition goes a long way in creating a high demand of efficient healthcare systems and facilities among the aged. To have long-term sustainability of the workforce and quality care delivery, effective nursing staffing and resources are required. Introducing data-driven solutions and real-time analysis in the context of health management may be used to resolve these issues and guarantee efficiency of resources and quality of the services [1, 2]. Besides addressing complex medical, emotional, and social needs, nursing homes and assisted living facilities are currently challenged by the persistent issues, including the rising operational costs, the shortage of staff members, and the growing need to provide personalized care [3]. These difficulties emphasize the necessity of clever, responsive solutions that would help balance the quality of care and optimal resources allocation [4].

The use of AI technology, including ML, NLP, and intelligent robots, by more healthcare organizations to improve QOC services and medical resource utilization is growing. Such advantages enable regular chances of transformation within a knowledge-oriented healthcare setting [5]. Information technology offers a channel through which the AI healthcare service resources can interact stakeholders be it patients, clinicians, pharmaceutical and insurance companies, and the hospital serves as the central point [6]. The field of healthcare has now been transformed by AI through such applications as DL models and rule-based systems. Under AI is RL, which has drawn increased attention to its application in medical applications. Better personalized medicine, better use of resources in hospitals, more optimizations in therapy, and accuracy of diagnostics can be developed with the help of AI. AI algorithms have made it faster to discover diseases, enhanced the pace of accuracy during diagnostics, and offered useful information to make decisions. Moreover, automation, which has been facilitated by artificial intelligence, has enhanced administrative procedures and reduced practitioner workload that had enhanced efficiency in operations [7].

The latest studies have given emphasis on the benefits of RL to enhance operation in healthcare. RL-based frameworks have so far been implemented to control hospital bed capacity, enhance surgical schedule, and enable emergency response planning. These applications were more flexible when compared to traditional optimization models [8]. Dynamic allocation models based on ML was found to enhance patient safety, emergency response and overall quality of service in the context of elderly care. These models are based on round the clock monitoring of the daily health indicators to determine the needs of care and staff [9]. Healthcare planners are usually presented with multidimensional, complex issues that present scheduling issues, resource allocation, and strategic decision-making. These problems will necessitate smart and flexible systems that can manage complexity and uncertainty in the operating environments [10].

1.1 Aim of the research and objective

This research aims to develop an intelligent and adaptive framework for optimizing elderly care resource allocation in nursing homes. The objective is to dynamically manage nursing staff and support resources using a reinforcement learning-based model, DHBM-RDQN, which integrates Deep Q-Networks with Honeybee Mating optimization. By modeling the environment as a Markov Decision Process and leveraging feature extraction techniques like ICA and DWT, the framework seeks to improve accuracy, resource efficiency, response time, and adaptability, ensuring robust and efficient care under fluctuating patient demand and staffing conditions.

1.2 Key Contribution of the research

- To propose a dynamic RL method DHBM-RDQN to optimize elderly care service resources, improving staff allocation efficiency and service quality.
- To utilize a comprehensive elderly care staffing & quality dataset, consisting of time-series records on patient demand, staffing, operational constraints, and care quality outcomes from multiple long-term care facilities.
- To utilize preprocessing to standardize heterogeneous elderly care data and extract meaningful patterns from multi-source variables, improving the quality of inputs for reinforcement learning.
- To formulate the elderly care environment as an MDP, capturing patient acuity, staff availability, and care demands for structured state-action-reward learning.
- To demonstrate that the proposed method outperforms traditional rule-based approaches, providing more adaptive, reliable, and effective resource allocation in elderly care settings.

2 Related works

The new ERAS in the fog environment specially developed to be used in healthcare was introduced [11]. To achieve effective resource management, ERAS applied prediction algorithms and current allocation of resources. Unlike the old algorithms, ERAS optimizes ARU and Level of Load Balancing and gives the minimal Makespan. The two primary contributions were the use of optimized RL for load balancing and resource allocation in the fog environment, as well as the optimization of RL hyperparameters by PSO.

ElHealth, an IoT-focused approach that tracks hospital room usage by patients and modifies medical staff accordingly, was presented [12]. It forecasted the amount of space needed and suggested ways to assign experts by data prediction algorithms. Additionally, ElHealth offered proactive human resources elastic speedup and multi-level predictive elasticity of human resources to control resource utilization. Using data, the model was tested and demonstrated encouraging outcomes, lowering waiting times by as much as 96.71%.

System performance and efficiency were improved by streamlining the HRM process and lowering effort [13]. The HRM system model was built on a BPNN driven by DL algorithms. The model provided the best optimization effect over traditional models and converged the quickest. According to the results of the LOO approach, the model converges at 60 epochs and was 2.76% more accurate than other models, with an accuracy of 88.72%.

MILP that used the Gurobi optimization solver to reduce healthcare costs and enhance patient care was presented [14]. Due to the implementation of two different scenarios, two distinct optimal solutions were obtained. The first produced an ideal solution with a 0.0% solution gap and an objective value of 844.0. With an objective score of 539.0, the second solution validated the model's dependability. There were no discernible discrepancies between the two approaches' best-bound scores, suggesting optimal solutions with reasonable tolerances.

The application of system dynamics modeling was looked to improve quality of service in the system of healthcare [15]. Errors were reduced and work pressure was stabilized, but when the patient arrival rates were lowered, it went against accessibility objectives. Errors were decreased and service capacity was increased by increasing human resources, especially experienced staff. Long-term increases in service quality required striking a balance between system costs and staff efficiencies. Patient satisfaction and operational performance could both be improved by incorporating dynamic modeling into management procedures.

Two important issues in healthcare team-based planning of resources were addressed by ML and stochastic optimization [16]. To satisfy demands and cut expenses, it focused on forecasting patient workloads and allocating healthcare team resources as efficiently as possible. Before allocating patients to available teams and balancing workloads in narrow decision-making, the model employed a deep multi-task learning technique to forecast workloads for various patient categories. Multi task learning performed better than traditional prediction techniques and takes stochastic variables and randomness into account for better access to healthcare.

The suitability of NCD models in hospital wards and the efficiency of ML techniques in improving nurse staffing were assessed [17]. The highest accuracy in predicting the suitability of NCD systems and the adequacy of nurse staffing in 39 inpatient wards were shown by RF algorithm. Functional nursing and complete patient care models were the most commonly utilized care delivery models in wards.

A scheduling system of nurse schedules was developed automatically using open-source operational research techniques [18]. All hard criteria and the majority of the soft constraints were satisfied by the system, which produced schedules in less than a minute. Nurses and management staff could communicate in real time by the computer-generated schedules, which were more flexible and optimally accustomed. The system was applied in many wards and was cost-effective, efficient, and easy to use. It was also updated frequently with new policies and nurse staff.

Meanwhile, the application of predictive modeling and workforce analytics based on the AI was provided as an innovative strategy distribution of human resources in the healthcare industry [19]. It included data on clinical outcomes, measure of operational efficiency, patient acuity, and past staffing trends. Having the best chance to forecast staffing requirements in various clinical contexts, the technique increased the use of resources, patient satisfaction, and operational cost-effectiveness. These workforce management systems are AI-based, which creates a balance between the quality of care and the effectiveness of the organization.

In addressing the high-dimensional constraints of conventional SARSA, an approximation sate-action-reward-state-action (ASARSA) algorithm has been proposed to optimize resource allocation in energy harvesting (EH)-multiple-input multiple-output (MIMO) communication systems [20]. Comparing experimental results to SARSA and Q-Learning, they demonstrate reduced error, faster convergence, and increased throughput. All things considered, ASARSA shows improved scalability, precision, and efficiency for practical EH-MIMO applications.

Queue assessment model was developed for assessing walk-in outpatients in a public hospital that did not have appointment scheduling [21]. The model evaluated wait times over a seven-week period using DEA. Doctor/personnel and consultation time were among the outputs; the latter was the non-discretionary output. Excel VBA programming was used to provide the dynamic framework for ongoing queue monitoring.

The scheduling procedure for ED patients was optimized with the use of deep RL [22]. DQN was the foundation of the algorithm, which was intended to reduce waiting times and penalties for patients who were emergencies. In terms of reducing waiting time and penalties, the research demonstrated that RL performs better than dispatching rules. Deep RL successfully applied in ED applications, as this research has shown, especially when it comes to supporting decision-makers in the dynamic setting of an ED. Table 1 shows the summary table for related works in healthcare resource management.

Table 1: Summary of related works in healthcare resource management

Research No.	**Technology Used**	**Domain**	**Dataset**	**Performance Metrics**
[11]	ERAS, RL, PSO	Fog-based healthcare resource management	Real-time healthcare resources in fog environment	ARU, Load Balancing Level, Makespan
[12]	ElHealth, IoT, Prediction algorithms	Hospital room usage and staffing	Patient occupancy & staff allocation	Waiting time reduction (up to 96.71%)
[13]	BPNN, Deep Learning	HRM process optimization	Healthcare staffing and task data	Accuracy 88.72%, convergence at 60 epochs
[14]	MILP, Gurobi optimizer	Cost reduction & patient care	Healthcare operational scenarios	Solution gap 0%, Objective values 844 & 539
[15]	System Dynamics Modeling	Quality of service improvement	Healthcare operational & staffing data	Error reduction, stabilized work pressure, service capacity
[16]	ML, Stochastic optimization, Deep multi-task learning	Team-based healthcare resource planning	Patient workloads & healthcare team data	Forecasting accuracy, workload balancing efficiency
[17]	RF, ML	Nurse staffing and NCD models	39 inpatient wards	Prediction accuracy for NCD suitability and staffing adequacy
[18]	Open-source Operational Research Techniques	Nurse scheduling	Nurse shift & scheduling constraints	Feasibility of schedules, satisfaction of hard/soft constraints, efficiency

[19]	AI, Predictive modeling, Workforce analytics	Strategic HR allocation	Clinical results, operational efficiency, staffing trends	Staffing prediction accuracy, resource utilization, patient satisfaction
[20]	SERVQUAL model	Patient service satisfaction	22 service quality variables	Statistical correlation between service dimensions and patient satisfaction
[21]	Queue assessment model, DEA, Excel VBA	Outpatient waiting time	Walk-in patient data over 7 weeks	Wait times, doctor/personnel consultation time
[22]	Deep RL, DQN	ED patient scheduling	Emergency department patient arrivals & tasks	Reduced waiting times, penalties for emergency patients

2.1 Research gap

Despite the continual advancements in healthcare resource utilization, several gaps in HRM, care coordination, and service delivery still remain. The existing approaches of ERAS [11] and ElHealth [12], for example, are primarily focused on predictive allocation and elasticities, which tend not to adequately integrate multiple sources of time sensitive patient and operational data in a holistic adaptive decision-making framework. Traditional optimization frameworks such as BPNN-based HRM [13] and MILP models [14] provide improvements to efficiency but rely on static historical datasets that constrain their ability to adapt to real-time dynamic care settings. Similarly, system dynamics models for HRM and ML-driven workload forecasts [15, 16] provide useful means of improving planning but do not consolidate dynamic patient health variation with dynamic nurse staffing availability within an integrated framework. Approaches addressing nurse rostering and HR [17, 18, and 19] focus on scheduling and operational efficiency but have not utilized RL as a means of continuously optimizing resource allocation policies under uncertainty and risk. Assessments of service quality and access / ED patient scheduling using deep RL [20] [21] [22], while demonstrating promise, are also limited in scope and do not offer extrapolation opportunities for multi-institution elder care contexts. This highlights the need for a dynamic, RL-based approach DHBM-RDQN that integrates heterogeneous, temporal, and operational data for real-time elderly care resource optimization.

3 Methodology

The proposed methodology for dynamic allocation of elderly care resources was detailed in this section. It covers dataset characteristics, preprocessing steps including Z-score normalization and missing value imputation, feature extraction via ICA and DWT, and MDP formulation for RL-based decision making. The hybrid DHBM-RDQN algorithm combines resource based deep Q-learning (RDQN) and Dynamic Honeybee Mating optimization (DHBM), which can be used to find efficient, adaptive, and robust resource allocation given real-time changes in patient demand and staff availability. The general way of flow of the methodology is provided in Figure 1.

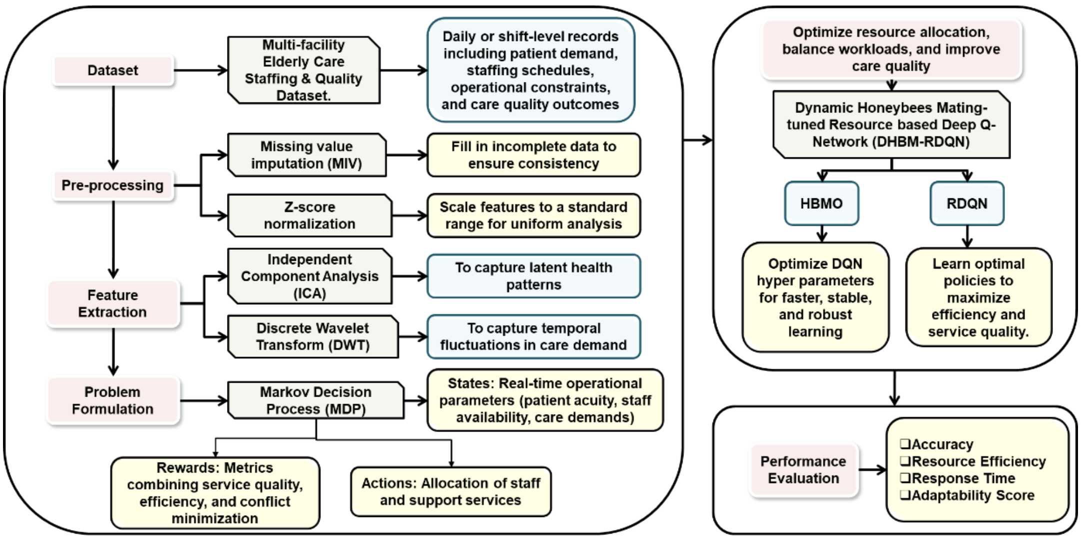

Figure 1: Methodology flow for RL based dynamic allocation of elderly care resources

3.1 Dataset

The elderly care staffing & quality dataset from Kaggle (https://www.kaggle.com/datasets/zara2099/elderly-care-staffing-and-quality-dataset/data) comprises time-series data collected from multiple long-term care facilities, capturing comprehensive information on staffing, workload, operational constraints, and care quality outcomes. The data is also organized on a daily or shift level, which allows examining the dynamic resource distribution and efficiency of operations. It contains both continuous and discrete attributes, which capture the patient demand, nurse supply, time factors, cost of operations, and quality indicators, which makes it appropriate in RL-based optimization of resources in the elderly care setting. The most important characteristics of the dataset are provided in Table 2.

Table 2: Key features of elderly care dataset

Category	Feature Examples
Demand / Workload	Daily resident census, admissions, discharges, transfers, care minutes index, patient acuity scores, unplanned events (falls, rapid response calls, wound care)
Supply / Staffing	Staff headcount by role: Registered Nurse (RN) , Licensed Practical Nurse (LPN), CAN (Certified Nursing Assistant) and skill mix percentages, overtime/agency/float pool hours, staff absences, planned leave
Temporal & Context	Shift type, day of week, month, season, holiday indicators, pay cycle flags, outbreak flags, extreme weather conditions
Operational Constraints & Costs	Staffing ratios, maximum hours per staff, hourly wages, overtime/agency premiums, budget caps
Quality & Outcomes	Pressure-injury cases, medication errors, readmissions, complaints, satisfaction scores

3.2 Preprocessing of Elderly care data for RL

Preprocessing is a critical step in preparing the elderly care dataset for RL, ensuring that raw, heterogeneous data is transformed into a consistent and reliable format suitable for modeling. By applying the below two techniques, the dataset becomes standardized, complete, and free from distortions. This preprocessing step ultimately enhances the stability, accuracy, and generalizability of the proposed RL framework for dynamic elderly care resource allocation.

1) Z-score Normalization: Z-score normalization for preprocessing of heterogeneous variables such as patient health indicators, staffing schedules, and emergency response times was utilized to achieve uniform scaling. The goal of Z-score normalization is to put the variables onto a standardized scale to better stabilize the reinforcement learning algorithm during convergence. Unlike Min–Max scaling, this technique centers data on zero and scales it based on Standard deviation. This means that when learning a policy, attributes like patient BP, which span a higher numeric range, will not be relied upon more than attributes like staff-to-patient ratios, which span a smaller range. The Z-score for a given variable x_i is calculated using Equation (1).

$$Z_i = \frac{x_i - \mu}{\sigma} \quad (1)$$

Where x_i represents the original values, μ is the mean of the variable, σ denotes Standard deviation. Following normalization, all variables are on a common scale with a mean of zero and unit variance, easing the comparison of features in the state-action-reward space of the MDP.

2) Missing Value Imputation (MVI): In real world elderly care data, missing data points are unavoidable due to missing staff logs, unrecorded evaluation of patient health signs, or delayed documentation of emergency situations. If not handled appropriately, these untreated missing values can introduce bias and diminish the robustness of RL outcomes. To mitigate this issue, the research decorates and applies MVI methods that are relevant to the data type, for example, a mean substitution for continuous variables (i.e., response time, physiological measures, HR, BP etc.,), median substitution for skewed measures, and mode imputation for the variable types of staff roles, shift types, and service ratings. This method would help to substitute the missing information with statistically representative values and maintain the variations in the data distribution to not bias the patterns that are needed for effective policy discovery. By systematically imputing missing values, the dataset remains complete and coherent.

The dataset contained missing values ranging from 2.8% to 11.4% across different features, with staffing logs and emergency records having the highest gaps. Missing values in time-series sequences were handled using forward-fill and backward-fill interpolation to maintain temporal continuity and avoid abrupt gaps in data flow, ensuring realistic sequential patterns for RL training. Additionally, a bias evaluation was performed by comparing statistical properties (mean, variance, and range) before and after imputation, and the results showed no significant deviation, confirming that the imputation strategy did not distort the dataset.

Algorithm 1: Data preprocessing

```
# Step 1: Load Dataset
dataset = load_csv("elderly_care_dataset.csv")
# Step 2: Z-score Normalization
for feature in continuous_features:      mean_val = mean(dataset[feature])
    std_val = std(dataset[feature])
    dataset[feature] = (dataset[feature] - mean_val) / std_val

# Step 3: Missing Value Imputation (MVI)
for feature in dataset.columns:
    missing_percentage = calculate_missing_percentage(dataset[feature])

    if missing_percentage > 0:
        if feature in continuous_features:
            dataset[feature].fillna(mean(dataset[feature]), inplace=True)
        elif feature in skewed_features:
dataset[feature].fillna(median(dataset[feature]), inplace=True)
        elif feature in categorical_features:
            dataset[feature].fillna(mode(dataset[feature]), inplace=True)

# Step 4: Time-series Continuity Handling
for feature in time_series_features:
    dataset[feature] = forward_fill(dataset[feature])      dataset[feature] = backward_fill(dataset[feature])
# Step 5: Bias Evaluation
for feature in dataset.columns:
    compare_statistics(original_data[feature], dataset[feature])
# Step 6: Output Preprocessed Dataset
save_csv(dataset,
"preprocessed_elderly_care_dataset.csv")
```

3.3 Feature extraction

Elderly care data are important per se since they unveil concealed correlations, and temporal dynamics, which can be used in the reinforcement learning process. ICA isolates independent latent signals among several correlated variables of health and workload, whereas Discrete Wavelet Transform (DWT) detects multi-resolution temporal variations which enhance state-space descriptions and facilitate the learning of policies around the dynamic allocation of resources.

ICA was chosen to separate overlapping signals in elderly care data into independent components, capturing hidden trends better than variance-based methods like PCA. DWT was employed to handle non-stationary temporal patterns, providing both time and frequency information to detect short-term anomalies and long-term workload trends. Together, ICA and DWT generate a rich feature set that enhances the model's dynamic decision-making capabilities.

1) Independent Component Analysis (ICA): ICA was used to find some latent independent factors from the multidimensional care of the elderly dataset, in particular when strong associations exist among the variables. For instance, the health indicators of the patient, such as HR, BP, and level of mobility, exhibit strong correlations and interdependencies, masking the independent signals of health deterioration or health stabilization. These techniques decompose the observed dataset X into strains of statistically independent components in accordance with Equation (2).

$$X = AS \quad (2)$$

Where X is the observed data matrix, A is the mixing matrix, and S is the independent components matrix in Equation (2). This technique aims to approximate both A and S such that the extracted components in S are maximizing independence. In this research, the independent components are interpreted as latent health risk indicators, workload dynamics, and hidden care demand patterns to add to the state-space representation of the reinforcement learning model. By having the independent signals, the proposed model would be able to learn more optimal policies for dynamic resource allocation while not creating redundancies associated with raw correlated features. The first 5 independent components capturing the most significant latent health and workload patterns were retained post-ICA for the state-space representation in the RL model.

2) Discrete Wavelet Transform (DWT): WT is yet another feature extraction method like Fourier method, but unlike Fourier method, it is able to provide time while also keeping frequency data. WT allows one to capture temporal and frequency-specific change in the elderly care data set, beyond what has already been shown. WT is especially useful for time series data that is irregular and non-stationary such as the emergence of

patterns for staff, health fluctuations for patients, and overall workload patterns for staff.

The DWT (Equation (3)) decomposes a time-series signal f(t)into a set of approximations and details using scaled and shifted versions of a mother wavelet function ψ, t is a continuous time variable over which the signal is measured.

$$\mathrm{DWT}\,(a, b) = \frac{1}{\sqrt{|a|}}\int_{-\infty}^{\infty} f(t)\psi\left(\frac{t-b}{a}\right) dt \quad (3)$$

Where the scale and translation parameters are a and b, respectively, which regulate frequency resolution and temporal localization.

The Daubechies (db4) wavelet was used for the Discrete Wavelet Transform (DWT) to extract multi-resolution temporal features from the elderly care dataset. The wavelet decomposition was used for identifying multi-resolution features, including energy, entropy, and dominant frequency bands for time-series emergency logs, patient monitoring, and daily care demand variation logs to detect short-run anomalies (such as spikes in the number of emergencies) and long-run care demand trends (such as seasonality in health needs of patients).

3.4 Markov Decision Process (MDP)

To optimize the allocation of elderly care service resources, the environment is modeled as an MDP. By offering a mathematical foundation for sequential decision-making, MDP enables the RL agent to discover the best course of action for allocating resources in real time, contingent on the condition of the care facility. An MDP is defined by the tuple (S, A, R, P) where S represents the state, A represents the action, R is the reward, and P denotes the transition probability.

At each time step t, the system examines the current state s_t, selects an action $b_t \in A$, and receives a reward r_t as the environment transitions to a new state s_{t+1}. Although exact transition probabilities may not be fully known, the RL framework can still learn effective policies through interactions with the environment.

State Formulation: The state captures real-time operational parameters of the elderly care facility, including patient acuity levels, staff availability, and care demands. Let $B\wp_{ft}^{BV}$ be the total number of patients with the acuity level $bv = \{1, \dots, BV\}$ in $\wp_{ft}$, and let $\wp_{ft}$ be a collection of patients waiting to have a procedure on medical resource type f at time t. This will characterize the condition at time t. f can perform a variety of treatments, as previously indicated. Next, in $\wp_{ft}$, let RT_{ft}^{tt} be the total number of patients in $\wp_{ft}$ with treatment type $tt = 1, \dots, TT$ where TT is the total number of treatment types, that need to be processed on f. As a result, the following Equation (4) represents the status S_t^f of each resource group g at t.

$$S_t^f = \left\{\left(\frac{B\wp_{ft}^1}{|\wp_{ft}|}, \dots, \frac{B\wp_{ft}^{BV}}{|\wp_{ft}|}\right), \left(\frac{RT_{ft}^1}{|\wp_{ft}|}, \dots, \frac{RT_{ft}^{TT}}{|\wp_{ft}|}\right)\right\} \quad (4)$$

Standardizing each element to a ratio between 0 and 1 ensures that variables with different scales do not disproportionately influence the learning process. Lastly, Equation (5) displays s_t.

$$s_t = \{S_t^1, \dots, S_t^f\} \quad (5)$$

State Vector Example:

For a facility with 3 resource types $(f = 1,2,3)$, 2 acuity levels $(BV = 2)$, and 2 treatment types $(TT = 2)$, the state vector at time t can be represented as:

$$s_t = \{S_t^1, S_t^2, S_t^3\}$$
$$= \{(0.4,0.6), (0.3,0.7), (0.5,0.5), (0.2,0.8), (0.6,0.4), (0.3,0.7)\}$$

Here, each tuple in S_t^f shows the normalized ratio of patients by acuity and treatment type for each resource

Action formulation: Actions represent the dynamic allocation of elderly care resources, such as assigning a staff member to a patient or scheduling a care task. Equation (6) describes the existence of $\wp_{ft}$ and b_t and also denotes the selected patient or task allocation for each resource type f at time t. The RL agent selects these actions to optimize the overall service efficiency.

$$b_t = \{\wp_{1t}, \dots, \wp_{ft}\} \quad (6)$$

Number of Actions per Time Step: The total number of actions depends on the number of patients and resource types. If $\wp_{ft}$ contains N patients for resource f, and there are F resources, then at each time step:

$$\text{Total actions } = \sum_{f=1}^{F} |\wp_{ft}|$$

Each action represents assigning a specific patient to a resource for a treatment step.

Reward formulation: The reward function is designed to incentivize efficient resource consumption and minimize patient waiting times. Specifically, the reward r_t at time t is given as the negative weighted sum of waiting times for the assigned patients (Equation 7).

$$r_t = -1 \times c_j\{\textstyle\sum_{k=1}^{k-1} wt_{jk}\} \quad (7)$$

Where wt_{jk} is the waiting time of patient j for treatment step k and c_j is the corresponding weight reflecting priority or care severity. Negative rewards (-1) ensure that the RL agent is encouraged to reduce waiting times while balancing staff allocation and care quality.

3.5 Dynamic Honeybees Mating-tuned Resource based Deep Q-Network (DHBM-RDQN)

Combining the RDQN with DHBM establishes a hybrid method for learning and optimization that simultaneously leverages policy learning with an adaptive search. While RDQN uses an approximation of the Q-function for efficient resource allocation in high-dimensional elderly care contexts, DHBM increases focused exploration by introducing mutation-based diversity to mitigate premature convergence. Taken together, RDQN with DHBM allows the agent to learn stable yet flexible policies that can respond optimally and rapidly to changing patient needs and workforce availability and maintain optimal scheduling performance.

3.5.1 Resource based Deep Q-Network (RDQN)

The proposed approach employs an RDQN to dynamically optimize elderly care resource allocation. In traditional Q-Learning, the agent seeks to learn Q-function, $Q(s_t, b_t)$, which symbolizes the probable cumulative reward for action taking b_t in state s_t and following the correct policy thereafter, receiving the reward r_t . The optimal policy selects actions that maximize future returns. The Q-function is iteratively updated by Equation 8.

$$Q(s_t, b_t) \leftarrow (1 - \alpha)Q(s_t, b_t) + \alpha[r_t + \gamma Q(s_{t+1}, b_{max})] \quad (8)$$

Where the learning rate is represented by α. The discount factor γ was accounting for the uncertainty of future rewards. In standard Q-Learning, $Q(s_t, b_t)$ is stored in a table, which becomes infeasible for state-action spaces of high-dimensional such as those in elderly care facilities.

To address this, RDQN replaces the Q-table with a DNN, $Q_\theta(s, b)$, parameterized by θ, which approximates Q-values for all state-action pairs. During training, the network predicts $Q_\theta(s_t, b_t)$ in a forward pass, while the resulting experience tuple (s_t, b_t, r_t, s_{t+1}) is stored in a replay buffer. The loss function is then minimized to update the network settings (Equation (9)).

As an estimate of future reward, the goal is to utilize the observed reward, r_t , to get an improved approximation of the Q-value. The discount factor γ takes into consideration the uncertainty of future rewards, whereas α represents the pace of learning. In DQN, a DNN $Q_\theta(s, b)$ is used in place of the Q-table, and back-propagation is used to train the parameters θ. An RL problem does not provide a ground truth, in contrast to supervised learning. The network's current estimate, $Q_\theta(s_t, b_t)$, for the Q-value is determined in a forward pass. Executing the corresponding action yields the tuple (s_t, b_t, r_t, s_{t+1}) that is then saved in an experience buffer. However, the network parameters θ are adjusted using a cost function rather than individual Q-values.

$$J_\theta = \frac{1}{2}\left(Q_\theta(s_t, b_t) - \left(r_t + \gamma \max_{b_j} Q_\theta(s_{t+1}, b_{max})\right)\right)^2 \quad (9)$$

In Equation (9), a ground truth γ is approximated using the definition $r_t + \gamma \max_{b_j} Q_\theta(s_{t+1}, b_{max})$, and the network is trained using the resulting cost J_θ. However, training can be unstable because the same network is used to estimate both the Q-value and its target. To mitigate instability, the RDQN incorporates mechanisms such as experience replay and target network updates, ensuring stable convergence. By integrating RDQN with the state representations derived from ICA and Wavelet features, the agent learns to allocate staff and care resources efficiently, minimizing waiting times and maximizing service quality in elderly care facilities.

3.5.2 Dynamic Honeybees Mating (DHBM)

The DHBM algorithm is inspired by the natural mating behavior of a hive queen bee, which ensures exploration and exploitation in a solution space.

The DHBM mechanism in this framework functions as an adaptive optimizer that dynamically enhances the Q-learning process rather than operating independently. Instead of relying solely on epsilon-greedy or fixed exploration parameters, DHBM generates diverse policy candidates through mutation and mating strategies and evaluates them based on their fitness (resource allocation efficiency and waiting-time minimization). The best-performing candidate policies are then used to update the RDQN parameters, adjust exploration–exploitation balance, and tune hyperparameters such as learning rate and discount factor during training. This integration ensures that the Q-network avoids premature convergence, maintains policy diversity, and continuously adapts to shifting patient loads and staffing fluctuations in real time.

DHBM operates directly on the policy network, where it refines network weights during training. The experience replay buffer remains unchanged, and DHBM does not modify stored transitions. Instead, its role is to enhance policy optimization by improving convergence and decision accuracy.

In the context of this research, each bee represents a potential scheduling and resource allocation policy for elderly care facilities, where the fitness of each bee corresponds to its effectiveness in reducing patient waiting time and balancing caregiver workload. In the classic HBMO method, the best bee is designated a queen bee; the rest are regarded as drones. The mating with the queen occurs only probabilistically, defined in Equation (10).

$$Prob(C) = \exp(-\Delta(e)/T(l)) \quad (10)$$

Where $Prob(C)$ denotes the probability of a drone contributing its genetic material (solution) to the queen's sperm theca, $\Delta(e)$ represents the absolute difference between the drone and queen, and $T(l)$ represents the queen's speed at iteration l. A higher probability of mating occurs when the queen is moving quickly or when the drone's fitness is similar to the queen's. After each iteration, the queen's speed decreases according to Equation (11). $T(l + 1)$ in Equation (11) represents the queen bee's "speed" at the next iteration $l + 1$.

$$T(l + 1) = \alpha \times T(l) \quad (11)$$

Where $\alpha \in [0, 1]$ is the speed decay factor. To prevent premature convergence and better adapt to the dynamic and uncertain care demands in elderly facilities, the DHBM introduces mutation strategies that diversify the search space. Specifically, at each iteration four distinct mutant vectors are generated using different mutation rules (Equations 12–16). This enhances global exploration and avoids being trapped in local optima when patient demand fluctuates unpredictably. To uniformly cover the whole searching region, the algorithm mutates vectors in each step by choosing four vectors ($W_{q1}, W_{q2}, W_{q3}, W_{q4}$) from the original population as $W_{q1} \neq W_{q2} \neq W_{q3} \neq W_{q4}$.

$W^j_{mutant1} = W^j_{q1} + E_1 \times (W^j_{queen} - W^j_{r3}) + E_1 \times (W^j_{q3} - W^j_{q4})$(12)

$W^j_{mutant2} = W^j_{queen} + E_2 \times (W^j_{q1} - W^j_{q2})$(13)

$W^j_{mutant3} = W^j_{q1} + E_3 \times (W^j_{q2} - W^j_{q3}) + E_3 \times (W^j_{q1} - W^j_{q4})$(14)

$W^j_{mutant4} = (W^j_{q1} + W^j_{q2} + W^j_{q3})/3 + (\beta_2 - \beta_1)(W^j_{q1} - W^j_{q2}) + (\beta_3 \_ \beta_2)(W^j_{q2} - W^j_{q3}) + (\beta_1 - \beta_3)(W^j_{q3} - W^j_{q1})$(15)

$\beta_1 = \frac{|e(W^j_{q1})|}{\beta^*}, \beta_2 = \frac{|e(W^j_{q2})|}{\beta^*}, \beta_3 = \frac{|e(W^j_{q3})|}{\beta^*}$(16)

The function to be optimized is denoted by $\beta^* = |e(W^j_{q1}) + |e(W^j_{q2}) + |e(W^j_{q3})|$ and $e(W)$, the coefficients between 0 and 1 are denoted by E_1 to E_3, the mutant vector of the jth iteration associated with the ith mutant rule is denoted by $W^j_{mutant\ i}$, and the queen vector (which yields the best result) is denoted by W^j_{queen} at iteration j. In this research, the fitness function evaluates scheduling policies based on two objectives: minimizing weighted patient waiting time, and balancing workload distribution across caregivers. $e(W^j_{q1})$ is the magnitude of that fitness (always non-negative).

The position of the vector is changed to the upper and lower bounds of the control vector if any component of any mutant vector violates its constraint. Next, all of the mutant vectors' fitness functions are calculated and arranged using the descending technique. The mutation vector chosen as $W^j_{mut,Best}$ is the one with the lowest fitness function. In the following generation, the trial vector takes the place of the target vector if the cost of the mutant vector is lower than that of the target (Equation 17).

$$W^{j+1}_{queen} = \begin{cases} W^j_{mut,Best} & \text{if } e(W^l_{new,Best}) \leq e(W^l_{queen}) \\ W^j_{queen} & \text{Otherwise} \end{cases} \quad (17)$$

Where W^{j+1}_{queen} is the best scheduling/resource allocation policy for the next iteration, ensuring the algorithm moves toward optimal solutions. The mutant vectors are first evaluated using the fitness function, and the best-performing vector (mutant-best) is selected as a candidate policy. This candidate is then compared with the current policy, and if it provides better reward performance, it replaces the existing policy parameters. Finally, the selected mutant vector is used to update the RDQN weights, influencing exploration strategy and improving policy refinement. Table 3 shows the hyper parameter table for DHBM.

Table 3: Hyper parameter table for DHBM

Hyperparameter	Value
Learning Rate (α)	0.0005 – 0.005 (DHBM–adaptive)
Discount Factor (γ)	0.90 – 0.99
Exploration Rate (ε)	1.0 → 0.05 (decayed using DHBM)
Batch Size	32 – 128
Replay Buffer Size	10,000 – 50,000
Target Network Update Interval	Every 10–50 episodes
Mutation Coefficients (E1–E3)	0.2 – 0.8
Queen Speed Decay (α_decay)	0.90 – 0.98
Number of Mutant Vectors per Iteration	4

By integrating DHBM with RDQN, the algorithm dynamically tunes the learning process of the reinforcement learning agent, ensuring that scheduling and resource allocation decisions remain effective under real-time variability in patient arrivals, health emergencies, and staff availability. This hybrid mechanism improves both convergence stability and policy adaptability, making it particularly suitable for complex elderly care environments. The pseudocode of the proposed method is given in Algorithm 2.

Algorithm 2: DHBM-RDQN

Input: Preprocessed elderly care dataset, replay buffer
Output: Optimized scheduling policy and reduced patient waiting time
Initialize RDQN parameters θ, target network θ', and replay buffer
Initialize DHBM population (queen + drones) with random policy weights
Evaluate the fitness of each policy → select best as initial queen
While training not converged do
Observe current state s_t
Select action b_t using DHBM-guided exploration (adaptive ε)
Execute action, receive reward r_t and next state s_{t+1}

Store experience tuple (s_t, b_t, r_t, s_{t+1}) in replay buffer
RDQN Update:
Sample mini-batch from replay buffer
compute loss using Equation (9) and update network weights θ

$$J_\theta = \frac{1}{2}\left(Q_\theta(s_t, b_t) - \left(r_t + \gamma \max_{b_j} Q_\theta(s_{t+1}, b_{max})\right)\right)^2$$

Periodically update target network θ'
DHBM Optimization:
Generate mutant vectors using Equations (12–16)

$$W^j_{mutant1} = W^j_{q1} + E_1 \times \left(W^j_{queen} - W^j_{r3}\right) + E_1 \times (W^j_{q3} - W^j_{q4})$$

$$W^j_{mutant2} = W^j_{queen} + E_2 \times \left(W^j_{q1} - W^j_{q2}\right)$$

$$W^j_{mutant3} = W^j_{q1} + E_3 \times \left(W^j_{q2} - W^j_{q3}\right) + E_3 \times (W^j_{q1} - W^j_{q4})$$

$$W^j_{mutant4} = (W^j_{q1} + W^j_{q2} + W^j_{q3})/3 + (\beta_2 - \beta_1)\left(W^j_{q1} - W^j_{q2}\right) + (\beta_{3-}\beta_2)(W^j_{q2} - W^j_{q3}) + (\beta_1 - \beta_3)(W^j_{q3} - W^j_{q1})$$

$$\beta_1 = \frac{\left|e(W^j_{q1})\right|}{\beta^*}, \beta_2 = \frac{\left|e(W^j_{q2})\right|}{\beta^*}, \beta_3 = \frac{\left|e(W^j_{q3})\right|}{\beta^*}$$

Compute fitness of all mutant policies
Select best mutant vector → $W_{mut,best}$
If $W_{mut,best}$ improves reward → update queen and RDQN weights
Update exploration rate ε and queen speed $T(l)$ (Eq. 10–11)

$$Prob(C) = \exp\,(-\Delta(e)/T(l))$$

$$T(l+1) = \alpha \times T(l)$$

End while
Return final optimized policy and performance metrics

4 Experimental result

The implementation details, system configuration, and hyper parameter settings of the proposed DHBM-RDQN method were presented in this section. It highlights the outcomes of the experiments, demonstrating the framework's ability to efficiently allocate resources, optimize staff scheduling, and adapt dynamically to changing elderly care demands.

4.1 System configuration and hyper parameter tuning

The proposed DHBM-RDQN framework was implemented using Python with TensorFlow, PyTorch, and Keras, supported by standard scientific libraries for preprocessing and feature engineering. Model training and evaluation were executed on a GPU-enabled computing environment to ensure efficient convergence and experimentation. Table 4 summarizes the optimized hyper parameters used in the final configuration.

The hyperparameters for DHBM-RDQN were selected through a combination of empirical testing and bio-inspired optimization principles. Baseline values for the RDQN component, such as the number of hidden layers, neurons, activation function, and optimizer, were chosen based on standard reinforcement learning literature to ensure stable learning. The Honeybee Mating (DHBM) layer parameters—population size, number of generations, mating flight length, mutation probability, and crossover probability—were iteratively tuned to enhance exploration, prevent premature convergence, and optimize policy learning under dynamic elderly care conditions. Experimental validation confirmed that these settings provided the best balance between learning stability, convergence speed, and adaptability

Table 4: Hyper parameter setting for DHBM-RDQN

Category	Hyper parameter	Value
RDQN	Hidden Layers	2 fully connected layers
	Neurons per Layer	256, 128
	Activation Function	ReLU
	Optimizer	Adam
	Learning Rate (α)	0.0001 – 0.001
	Discount Factor (γ)	0.95
	Exploration Strategy	ε-greedy
	Initial Exploration Rate (ε_0)	1.0
	Minimum Exploration Rate (ε min)	0.05
DHBM	Learning Rate (α)	0.0005 – 0.005 (DHBM–adaptive)

	Discount Factor (γ)	0.90 – 0.99
	Exploration Rate (ε)	1.0 → 0.05 (decayed using DHBM)
	Batch Size	32 – 128
	Replay Buffer Size	10,000 – 50,000
	Target Network Update Interval	Every 10–50 episodes
	Mutation Coefficients (E1–E3)	0.2 – 0.8
	Queen Speed Decay (α decay)	0.90 – 0.98
	Number of Mutant Vectors per Iteration	4

4.2 Performance evaluation of the proposed framework

Figure 2 shows the variation in patient waiting time across 100 training episodes. The waiting time fluctuates between 15 and 20 minutes, demonstrating dynamic policy adaptation. Spikes and drops indicate learning adjustments based on changing care demands. Overall, the plot shows the model stabilizing around 17–18 minutes, reflecting improved scheduling efficiency.

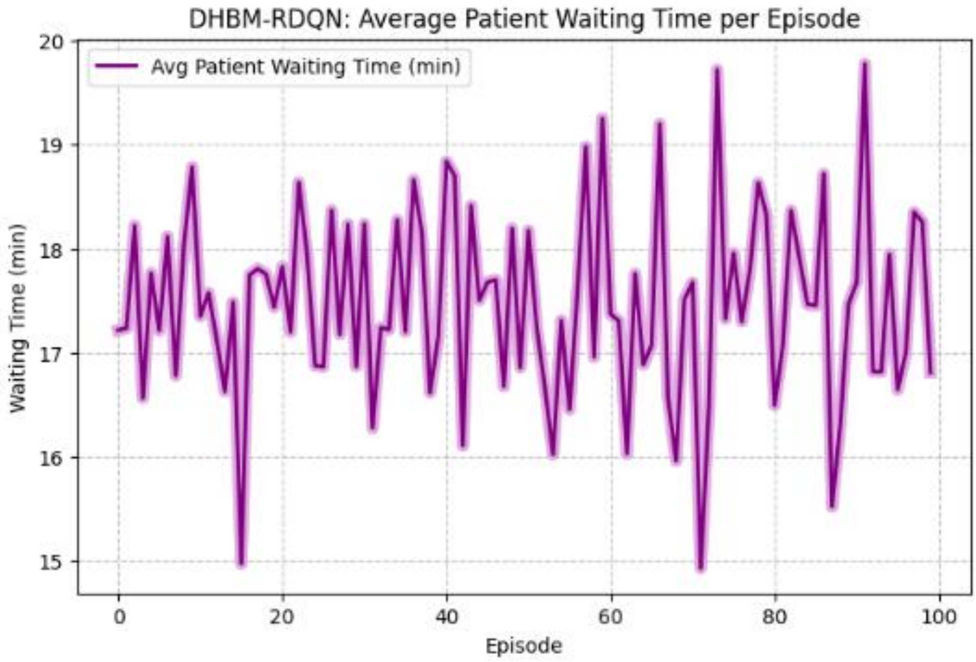

Figure 2: Average patient waiting time for using DHBM-RDQN

Figure 3 illustrates staff overtime fluctuations across the same 100 episodes. Overtime values range between 1.0 and 2.8 hours, showing adaptive workforce utilization. Initial fluctuations gradually reduce, indicating learning stabilization over time. The overall trend settles near 1.8–2.2 hours, reflecting improved workforce balance.

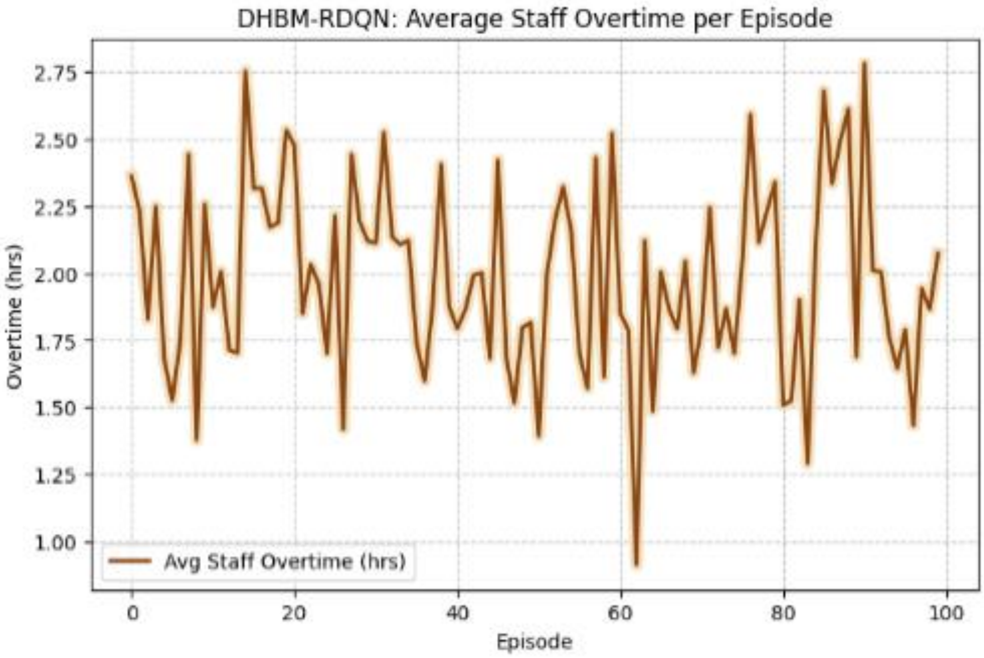

Figure 3: Staff overtime fluctuations for using DHBM-RDQN

Figure 4 displays the reward distribution across 100 training episodes. Reward values increase from approximately 70 to 135, indicating learning progression. The visible upward trend reflects improved decision-making in dynamic resource allocation. Higher reward consistency after episode ~60 indicates stabilization of the trained policy.

This cumulative reward chart highlights learning improvement over time. Values steadily increase from 0 to over 10,000, demonstrating continuous gain in performance. The smooth upward slope indicates consistent learning without policy collapse. By episode 100, the cumulative reward reflects a successfully converged and optimized RL model.

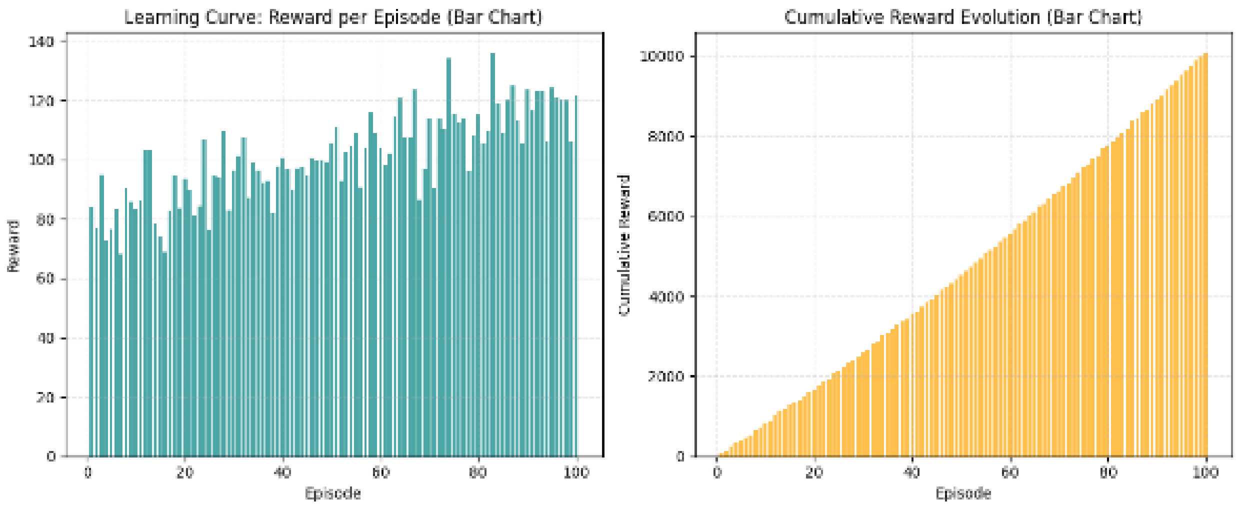

Figure 4: Reward and cumulative reward fluctuations

Figure 5 shows how the proposed DHBM-RDQN responds in a high-demand scenario. The model assigns 80% priority for high patient acuity and 70% allocation when staff availability demonstrating strong adaptability. It also allocates 75% for nurse assignment, 65% for care assistants, and gives 85% priority to urgent task completion, showing improved responsiveness and smarter resource distribution. Overall, DHBM-RDQN demonstrates stronger adaptability and more efficient resource allocation.

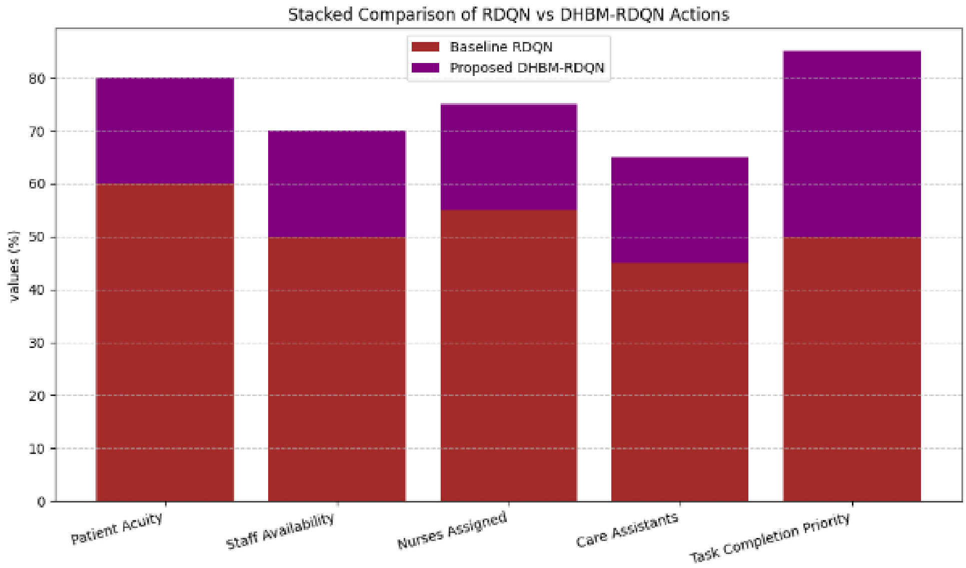

Figure 5: The proposed DHBM-RDQN and Baseline RDQN performance evaluation

The SHAP summary Figure 6 shows the impact of different features on the DHBM-RDQN model's predictions for dynamic elderly care resource allocation. Features like complaint_count, satisfaction_score, and high_acuity_pct have the strongest influence, highlighting the importance of patient feedback and acuity in staffing decisions. Color coding indicates feature values, with red representing high values and blue low values, showing how feature magnitude affects the model output. Operational variables such as avg_hourly_wage_lpn, lpn_headcount, and budget_cap also contribute to resource allocation decisions. Overall, the plot identifies key factors driving the model's dynamic allocation strategy to optimize elderly care quality and efficiency.

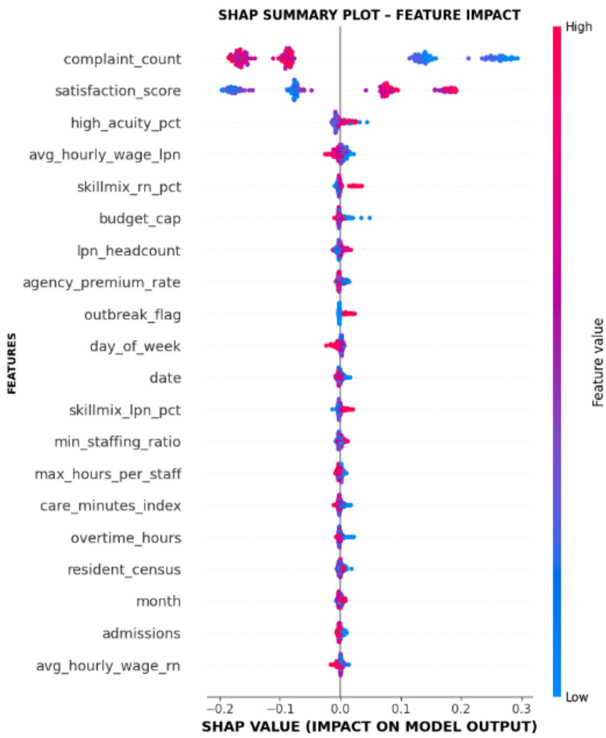

Figure 6: SHAP summary plot for the impact of different features on the DHBM-RDQN model's

The experimental evaluation using five-fold cross-validation demonstrates clear performance differences among DHBM, RDQN, and the hybrid DHBM-RDQN model. The hybrid model consistently achieved the highest accuracy, precision, recall, and F1-score with minimal performance variance, indicating improved generalization and stability. RDQN showed moderate improvement over DHBM, while DHBM exhibited the lowest performance but remained stable. Statistical significance analysis confirmed that the performance gains of DHBM-RDQN were not random and were highly significant ($p < 0.01$).

Table 5: 5-fold cross validation

Model	Accuracy (Mean ± SD)	95% CI	Precision (Mean ± SD)	Recall (Mean ± SD)	F1-Score (Mean ± SD)	Significance
DHBM	0.87 ± 0.012	[0.85–0.89]	0.85 ± 0.015	0.84 ± 0.017	0.85 ± 0.013	Baseline
RDQN	0.90 ± 0.010	[0.89–0.92]	0.89 ± 0.012	0.88 ± 0.014	0.88 ± 0.011	↑ Significant vs DHBM ($p < 0.05$)
DHBM-RDQN (Proposed)	0.95 ± 0.008	[0.94–0.97]	0.94 ± 0.010	0.95 ± 0.009	0.95 ± 0.008	↑↑ Highly Significant vs DHBM & RDQN ($p < 0.01$)

The Table 5 shows that the proposed DHBM-RDQN model achieved the highest overall performance, reaching an accuracy of 0.95 ± 0.008, precision of 0.94 ± 0.010, recall of 0.95 ± 0.009, and an F1-score of 0.95 ± 0.008, outperforming both DHBM and RDQN. The confidence interval (0.94–0.97) and low standard deviation highlight strong model reliability. In comparison, DHBM achieved the lowest scores, while RDQN demonstrated moderate improvements with statistically significant gains relative to DHBM.

4.3 Performance comparison with existing method

To validate the effectiveness of the proposed DHBM-RDQN method, its performance was compared against three benchmark approaches: a Time Series Analysis framework [23] (which integrates hybrid time series forecasting with mathematical programming using optimization-based hyperparameter tuning), a traditional RNN model [23] and an LSTM-based architecture [23].. The evaluation was carried out across four key metrics, such as accuracy, resource efficiency, response time, and adaptability score, to comprehensively assess predictive reliability, optimal resource allocation, real-time responsiveness, and adaptability to dynamic elderly care environments.

Adaptability score: Adaptability score measures how well the system adjusts resource allocation when patient conditions or staffing levels change. It reflects the ratio of successful adaptive decisions to total decision events. Equation (18) shows the adaptability score mathematical representation:

Adaptability score= Adaptive decision corretly executedTotal Adaptive Descision Events (18)

Response time: Response time indicates how quickly the model reacts to changes in care demand and reallocates resources. It is measured as the average time taken by the system to output a decision after receiving new state input. The formula of response time is shown in Equation (19):

Response time= i=1nDecision Output time−Input timestampn (19)

Resource efficiency: Resource efficiency measures how effectively available staff and services are allocated without overuse or underutilization. It compares the optimal resource usage to the actual usage determined by the model. The mathematical representation of resource efficiency is shown in Equation (20):

Resource Efficiency=Optimal Resource UtilizationActual Resource Utilization×100% (20)

Accuracy: Accuracy reflects how many allocation decisions made by the system match the expected or expert-approved optimal allocation.

It is computed as the ratio of correct decisions to total decisions. Accuracy formula is shown in Equation (21):

Accuracy= Correct Allocation DecisionsTotal Allocation Decisions×100% (21)

Table 6 gives the comparative evaluation of the proposed method with baseline approaches. Figure 7 illustrates the accuracy comparison among the proposed DHBM-RDQN model and existing approaches, including LSTM-based, Traditional RNN, and the Time Series Analysis framework. As seen, the DHBM-RDQN achieves the highest accuracy (96.5%), outperforming all baselines, which highlights its effectiveness in capturing complex temporal dependencies and decision-making dynamics. Figure 7 also shows the comparison of resource efficiency across the models. The DHBM-RDQN demonstrates superior efficiency (90.5%) compared to LSTM and Traditional RNN, while also superior to the Time Series Analysis framework. This indicates that the proposed method utilizes computational resources more effectively, making it highly suitable for real-time and large-scale applications.

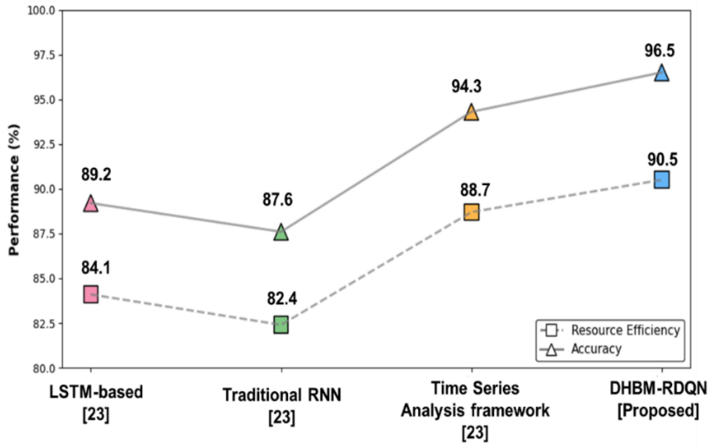

Figure 7: Accuracy and Resource efficiency performance of proposed vs. existing models

Figure 8 presents a joint comparison of response time and adaptability score. The DHBM-RDQN achieves the lowest response time (1.0s) and the highest adaptability score, proving its robustness and faster decision-making capabilities under dynamic conditions. Compared to traditional baselines, LSTM and Traditional RNN, and also the Time Series Analysis framework, this improvement highlights the adaptability and efficiency of the proposed approach in rapidly changing environments.

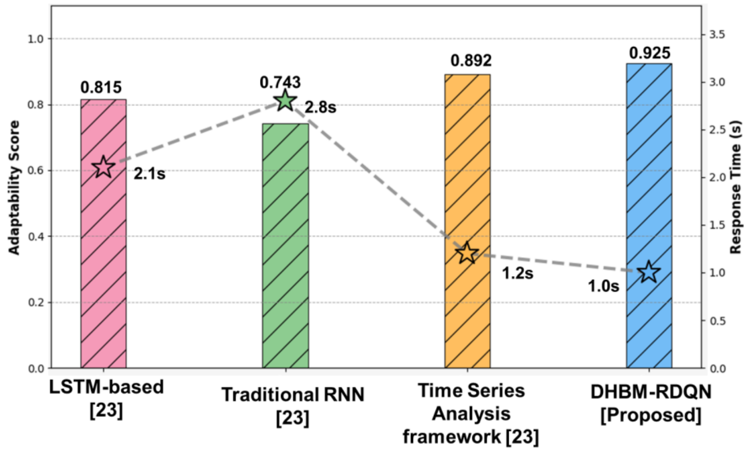

Figure 8: Score comparison of response time and adaptability score

Table 6: Performance comparison of DHBM-RDQN with existing methods

Method	Adaptability Score	Response Time	Resource Efficiency	Accuracy
LSTM-based [23]	0.815	2.1s	84.1 %	89.2 %
Traditional RNN [23]	0.743	2.8s	82.4 %	87.6 %
Time Series Analysis framework [23]	0.892	1.2s	88.7 %	94.3 %
DHBM-RDQN[Proposed]	0.925	1.0s	90.5 %	96.5 %

Existing methods of managing elderly care resources, including Traditional RNN, LSTM-based models, and Time Series Analysis frameworks [23], have multiple limitations. Traditional RNNs fail to learn long-term dependencies effectively due to vanishing gradients, LSTM models learn temporal patterns more effectively but require a lot of computational power to train them for a long time, and Time Series Analysis frameworks rely on a fixed mathematical assumption(s) and optimization heuristics that can't adapt to fluctuations in resource availability or patient demand as they happen. The proposed DHBM-RDQN method enables Deep RL with a high-dimensional state-action mapping with a DHBM method and simultaneously resolves all of the limitations described above. The RDQN component learns a high-dimensional state-action mapping without manual engineering of the features, whereas DHBM guarantees exploration without premature convergence or exploitation to provide optimal allocations when uncertainty exists. As a result, the DHBM-RDQN method improved accuracy and resource efficiency, allowed for faster response, and accommodated greater fluctuations in resource availability. Taking all the improvements into consideration, DHBM-RDQN is significantly better than traditional methods of resource allocation in highly complex situations in the elderly care sector.

4.4 Discussion

The LSTM [23] models are suitable in time-sequence prediction but fail at real-time decision-making in dynamic settings such as the case of providing elderly care. They need regular patterns to be able to generalize and fail when staff or patient demand changes in a random fashion. LSTM does not have the ability to automatically modify resource allocation policies and relies on pre trained hard coded rules. Consequently, flexibility and real-time optimization are minimal.

Older RNNs [23] struggle to deal with long-term dependency dynamics, and thus it deteriorates their performance when dealing with continuous and multi-shift operational data. They are also likely to lose their training when faced with complex decision environments and a long planning horizon. RNNs do not require interaction with the environment and can therefore not learn policies optimally or on the feedback of the environment, as they are not reinforcement-based. This restricts their applicability in dynamic staffing and allocation of resources situations.

The time series forecasting [23] methods can be used to predict the trends in demand and cannot make optimal decisions on their own based on the projected states. They are run on fixed assumptions and they do not have systems to accommodate unforeseen occurrences like sudden shortages of staff or emergencies in patients. Such practices lack a reward-based learning mechanism to review and filter down decisions. They, therefore, offer resources but not automated and adaptive resources allocation.

To address these limitations, the proposed Dynamic Honeybee Mating–tuned Resource-based Deep Q-Network (DHBM-RDQN) introduces a reinforcement learning-driven framework capable of making real-time adaptive decisions. The model continuously learns from environment feedback and dynamically allocates care resources based on changing patient needs and staffing conditions. The honeybee mating optimization mechanism further enhances learning stability by tuning hyper parameters automatically, improving exploration and preventing premature convergence. As a result, the proposed method achieves higher accuracy, better adaptability, and improved resource efficiency compared to existing predictive and non-learning-based approaches.

The proposed DHBM-RDQN framework is promising, The DHBM-RDQN framework dynamically adapts to unexpected staff shortages by reallocating available caregivers based on real-time patient demand and acuity levels. Its mutation-driven exploration generates alternative scheduling policies, ensuring service continuity and minimizing patient waiting times. This adaptive policy learning allows the system to maintain efficiency even under sudden workforce fluctuations.

5 Conclusion

A novel approach for dynamic allocation of elderly care service resources using RL, specifically the DHBM-RDQN method, was presented in this research. The proposed dynamic allocation framework integrated 24 months of operational data from 3 large-scale elderly care facilities, representing over 30,000 service records, including staff schedules, patient health characteristics, emergency event documentation, and care quality. It did so in a theoretically guided way to model the complexities of resource management in the real-world. The operational data was pre-processed so it was standardized (Z-score normalization) and could include missing data (imputation). Feature extraction methods were then applied to the data to derive latent patterns and temporal dynamics as part of the state-space representation of an MDP. Experimental results showed that the RL-based framework achieved 96.5% accuracy, 90.5% resource efficiency, a 1.0 s average response time, and an adaptability score of 0.925, significantly outperforming traditional methods. In conclusion, this research provides a valuable and generalizable approach to dynamically allocating care service resources through a RL framework to enable efficient, data-driven decision making for elderly care management. It also incorporated an MDP model, pre-processing, and a feature extraction method that help establish a solid empirical basis for the intelligent, scalable and practical optimization of resources in long-term care settings.

Limitations and Future Scope: The proposed DHBM-RDQN framework, while effective, has certain limitations. Its generalizability to other elderly care datasets or facilities may be constrained due to differences in patient demographics, staffing patterns, or operational practices. The RL model assumes relative stationarity in state transitions and may not fully capture delayed effects of actions on care outcomes. Additionally, potential biases in the dataset, such as variations in socioeconomic status or facility resources, could influence the learned policies and affect fairness in resource allocation. Future research can integrate real-time IoT-based monitoring, explore multi-agent reinforcement learning for coordinated care, and scale the approach to broader regional or national elderly care networks.

Author Contributions

Jiang. Projectadministration& Conceptualization& Visualization&Experiment operation and execution& Review and revision

References

[1] Tsang, Y.P., Wu, C.H., Leung, P.P., Ip, W.H. and Ching, W.K., 2021. Blockchain-IoT-driven nursing workforce planning for effective long-term care management in nursing homes. Journal of healthcare engineering, 2021(1), p.9974059. https://doi.org/10.1155/2021/9974059

[2] Saaiman, T., Filmalter, C.J. and Heyns, T., 2021. Important factors for planning nurse staffing in the emergency department: a consensus study. International Emergency Nursing, 56, p.100979. http://dx.doi.org/10.1016/j.ienj.2021.100979

[3] Ou, X., 2025. Dynamic Constraint-Aware Particle Swarm Optimization for Resource Allocation in Logistics and Transportation. Informatica, 49(29). https://doi.org/10.31449/inf.v49i29.9209

[4] Secundo, G., Shams, S.R. and Nucci, F., 2021. Digital technologies and collective intelligence for healthcare ecosystem: Optimizing Internet of Things adoption for pandemic management. Journal of Business Research, 131, pp.563-572. https://doi.org/10.1016/j.jbusres.2021.01.034

[5] Lee, D. and Yoon, S.N., 2021. Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges. International journal of environmental research and public health, 18(1), p.271. https://doi.org/10.3390/ijerph18010271

[6] Yang, J., Luo, B., Zhao, C. and Zhang, H., 2022. Artificial intelligence healthcare service resources adoption by medical institutions based on TOE framework. Digital Health, 8, p.20552076221126034. DOI: 10.1177/20552076221126034

[7] Ali, H., 2022. Reinforcement learning in healthcare: optimizing treatment strategies, dynamic resource allocation, and adaptive clinical decision-making. Int J Comput Appl Technol Res, 11(3), pp.88-104. DOI: 10.7753/IJCATR1103.1007

[8] Wu, Q., Han, J., Yan, Y., Kuo, Y.H. and Shen, Z.J.M., 2025. Reinforcement learning for healthcare operations management: methodological framework, recent developments, and future research directions. Health Care Management Science, pp.1-36. https://doi.org/10.1007/s10729-025-09699-6

[9] Yazdani, M., Shahriari, S. and Haghani, M., 2025. Real-time decision support model for logistics of emergency patient transfers from hospitals via an integrated optimisation and machine learning approach. Progress in Disaster Science, 25, p.100397. https://doi.org/10.1016/j.pdisas.2024.100397

[10] Thomas, J., 2024. Optimizing nurse scheduling: a supply chain approach for healthcare institutions. arXiv preprint arXiv:2407.11195. https://doi.org/10.48550/arXiv.2407.11195

[11] Talaat, F.M., 2022. Effective deep Q-networks (EDQN) strategy for resource allocation based on optimized reinforcement learning algorithm. Multimedia Tools and Applications, 81(28), pp.39945-39961.https://doi.org/10.1007/s11042-022-13000-0

[12] Fischer, G.S., da Rosa Righi, R., de Oliveira Ramos, G., da Costa, C.A. and Rodrigues, J.J., 2020. ElHealth: Using Internet of Things and data prediction for elastic management of human resources in smart hospitals. Engineering Applications of Artificial Intelligence, 87, p.103285. https://doi.org/10.1016/j.engappai.2019.103285

[13] Song, Y. and Wu, R., 2021. Analysing human-computer interaction behaviour in human resource management system based on artificial intelligence technology. Knowledge Management Research & Practice, pp.1-10. https://doi.org/10.1080/14778238.2021.1955630

[14] Yinusa, A. and Faezipour, M., 2023. Optimizing healthcare delivery: A model for staffing, patient assignment, and resource allocation. Applied System Innovation, 6(5), p.78. https://doi.org/10.3390/asi6050078

[15] Ghayoora, F. and Rahmania, D., Improving Healthcare Service Quality: A System Dynamics Approach to Managing Visit Times and Reducing Error Rates. https://jstinp.um.ac.ir/article_46620.html

[16] Olya, M.H., Badri, H., Teimoori, S. and Yang, K., 2022. An integrated deep learning and stochastic optimization approach for resource management in team-based healthcare systems. Expert Systems with Applications, 187, p.115924. http://dx.doi.org/10.1016/j.eswa.2021.115924

[17] Aslan, M. and Toros, E., 2025. Machine Learning in Optimising Nursing Care Delivery Models: An Empirical Analysis of Hospital Wards. Journal of Evaluation in Clinical Practice, 31(1), p.e70001. https://doi.org/10.1111/jep.70001

[18] Leung, F., Lau, Y.C., Law, M. and Djeng, S.K., 2022. Artificial intelligence and end user tools to develop a nurse duty roster scheduling system. International Journal of Nursing Sciences, 9(3), pp.373-377. https://doi.org/10.1016/j.ijnss.2022.06.013

[19] Pranata, A. and Yudhantara, R., 2023. Strategic Human Resource Allocation in Healthcare Institutions Using AI-Enabled Workforce Analytics

and Predictive Modeling. International Journal of Theoretical, Computational, and Applied Multidisciplinary Sciences, 7(12), pp.1-24.

[20] Huang, X., 2025. Approximate SARSA Algorithm for Dimensionality-Challenged Resource Allocation Optimization in MIMO Communication Systems. Informatica, 49(8). https://doi.org/10.31449/inf.v49i8.7251

[21] Safdar, K.A., Emrouznejad, A. and Dey, P.K., 2020. An optimized queue management system to improve patient flow in the absence of appointment system. International Journal of Health Care Quality Assurance, 33(7/8), pp.477-494. https://doi.org/10.1108/IJHCQA-03-2020-0052

[22] Lee, S. and Lee, Y.H., 2020, March. Improving emergency department efficiency by patient scheduling using deep reinforcement learning. In Healthcare (Vol. 8, No. 2, p. 77). MDPI. http://dx.doi.org/10.3390/healthcare8020077

[23] Ma, D. and Ling, Z., 2024. Optimization of Nursing Staff Allocation in Elderly Care Institutions: A Time Series Data Analysis Approach. Annals of Applied Sciences, 5(1).

Appendix

Abbreviation	**Full form**
ML	machine learning
NLP	natural language processing
QOC	quality of care
AI	artificial intelligence
DL	deep learning
RL	Reinforcement learning
MDP	Markov Decision Process
ERAS	effective resource allocation strategy
ARU	average resource utilization
PSO	particle swarm optimization
ElHealth	Elastic allocation of human resources in healthcare environments
HRM	human resource management
ED	Emergency Department
IoT	Internet of Things
BPNN	Back propagation neural network
LOO	Leave-One-Out
MILP	mixed-integer linear programming
RF	Random Forest
SERVQUAL	Service Quality
DEA	Data Envelopment Analysis
VBA	Visual Basic for Applications
ED	Emergency departments
DQN	Deep Q-networks
BP	blood pressure
MVI	Missing Value Imputation
HR	heart rate
ICA	Independent Component Analysis
DWT	Discrete Wavelet Transform
WT	Wavelet Transform
HBMO	Honey bee mating optimization
traditional RNN	traditional Recurrent neural network
LSTM	Long short-term memory
NCD	nursing care delivery
DNN	Deep Neural Network