Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 12 March 2026
Revised 16 April 2026
Accepted 20 May 2026

Natural Resources for Human Health 2026, Vol. 6 (2s): 20–29

KEYWORDS:
Explainable AI; deep reinforcement learning; healthcare resource allocation; epidemic preparedness; public healthcare.

An Explainable Deep Reinforcement Learning Approach for Intelligent Resource Allocation and Epidemic Preparedness in Public Healthcare Systems

[1]S Jagadeesan, [2]Dr Ezhil Muthalvan A, [3]Dr. Pandiyan K.R, [4]Prof. Dr. Johncyrani Rajagopal, [5]R. Naveenkumar, [6]Dr. Sumit Kushwaha, [7]Baljinder Kaur, [8]Sai Krishna Edpuganti,

[1]RCI Licensed Clinical Psychologist & Clinical Director, Rooted Connections. Email: jagadeesan@rootedconnections.in
[2]Senior Resident, Department of Community Medicine Saveetha Medical College Hospital, SIMATS, Chennai, Tamilnadu Mail id - ezhilmuthalvana.smc@saveetha.com ORCID ID: 0009-0001-5046-1337
[3]Professor, Community Medicine, Meenakshi Medical College Hospital & Research Institute, Meenakshi Academy of Higher Education and Research, Enathur, Kanchipuram, Tamil Nadu 631552
[4]Principal, Apex College of Nursing, Varanasi, India. Email: jancir86@gmail.com
[5]Dept of CSE, School of Engineering and Technology, CGC University Mohali-140307, Punjab India.Email: drnk1983@gmail.com 0000-0001-9033-9400
[6]Associate Professor, Department of Directorate of Research, Innovation, and Consultancy (DRIC), Chandigarh University Uttar Pradesh, Unnao - 209859, Uttar Pradesh, India. Email: sumit.kushwaha1@gmail.com ORCID: 0000-0002-3830-1736
[7]Assitant Professor, Faculty of Computing, Guru Kashi Universiity, Bathinda, Punjab, India E-mail:- jind2005@gmail.com Orcid I'd:- https://orcid.org/0009-0007-0846-9996
[8]Assistant Professor, Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur District, Andhra Pradesh, India. saikrishnaedpuganti@gmail.com https://orcid.org/0009-0000-7966-2100

ABSTRACT: Epidemics often lead to poor, delayed, and uneven distribution of public health resources, thereby pushing public healthcare systems to the limits of their capacities. This research introduces a "beds, healthcare personnel, testing kits, vaccines, medicines, and emergency funds" distribution model using explainable deep reinforcement learning for 20 public healthcare facilities that serves 50,000 adolescents and includes realistic 14-day demand forecasting using an LSTM network. Epidemiological, operational and demographic drivers of allocation decisions are captured through SHAP explanations, and constraints on capacity, budget, emergency-reserve and geographic-equity ensure transparent planning. The analyses of performance are done based on demand satisfaction, resource utilization, preparedness, unmet critical demand, vaccine wastage, epidemic peak demand, waiting time, rural–urban service disparity, explanation fidelity. The following reference values are considered: 92% satisfied with the demand; 90% efficiency of the resources; 85/100 preparedness; 5% of unmet critical demand; 3% of vaccine wastage; and 90% explanation fidelity. The method is conducive to efficient, equitable and interpretable epidemic-preparedness decision making.

1. INTRODUCTION

Public healthcare services are key to pandemic prevention and protection of population health, but they can be overwhelmed by the sudden surge in infections, hospitalizations and supply needs. The number of hospital beds, health staff, testing kits,

eISSN: 2583-1194

vaccines, medicines, and emergency funds that can be made available in time are key elements in effective preparedness. This challenge is especially critical for facilities that provide care to adolescents, as health care services could be affected by the shift of resources during a public health emergency.

Previous research has used compartmental epidemic models, machine learning, mathematical optimization and rule-based decision systems to predict disease spread and guide the management of health resources. Recent work shows that deep learning models, especially long short-term memory models, can be used to predict temporal demand, and reinforcement learning models have been applied to sequential resource-management decisions. To make the prediction more explainable, techniques like SHAP have been introduced to the healthcare domain in the form of explainable artificial intelligence (XAI).

However, traditional allocation methods are usually static, based on fixed rules and do not respond well to rapidly changing epidemic conditions. Forecasting and allocation are dealt with separately, and many intelligent solutions maximize resources and do not take into account budgets, capacity, emergency reserves, geographic equity, or clear decision making. These restrictions make an adaptive and accountable system that allows several resources to coordinate over all kinds of public health services necessary.

Therefore, a study to design an explainable deep reinforcement learning solution towards intelligent resource allocation and epidemic preparedness is in order. The proposed solution is a combination of 14-day LSTM model for forecasting demand, constrained Proximal Policy Optimization for determining weekly allocations at 20 health facilities for 50,000 adolescents. Using SHAP-based explanations, epidemiological, operational and demographic factors are characterized for each decision.

The key contributions are: 1) the creation of an integrated epidemic-demand forecasting and sequential multi-resource allocation; 2) the creation of the capacity, budget, preparedness, and geographic-equity constraints; 3) the incorporation of facility-level explanations for allocation decisions; and 4) the creation of an evaluation framework for efficiency, preparedness, fairness, wastage, unmet demand, and interpretability.

2. RELATED WORK

There are always compartmental models used for the estimation of transmission and need for care, such as Susceptible–Infected–Recovered models and Susceptible–Exposed–Infected–Recovered models, which have traditionally been used for epidemic preparedness. These methods have been extended recently by simulation and deep reinforcement learning to find adaptive epidemic-control policies [1]. Reinforcement learning has also been studied in the field of healthcare applications including treatment planning, clinical decision support and sequential intervention selection [2]. There is increased interest in value-based and hybrid approaches to reinforcement learning as evidenced in the systematics, but interpretability, reward definition, data quality, and clinical acceptance are significant limitations [3]. Therefore, before employing reinforcement learning recommendations in healthcare practice, it is crucial to evaluate the situation with care and develop responsible policies [4]. Further, the interactions between disease dynamics, individual behavior, perceived risk and compliance with interventions impacts the course of the epidemic, as illustrated in game-theoretic models [5].

Despite the benefits of representation learning for constructing and modeling meaningful states from complex healthcare data, the performance of policies in these applications is sensitive to the chosen representation [6]. The problem of deriving "Individualized and interpretable Sepsis-Treatment Strategies" from intensive-care information has been solved by applying reinforcement learning techniques [7]. Deep reinforcement learning has been used at the population level to make decisions for the time and level of lockdown and travel restrictions during Covid-19 [8]. Similarly, deep reinforcement learning has been used to solve large-scale epidemic control research to find adaptive intervention policies in the nonlinear disease dynamics [9]. In agent-based environments like Pyfectious, detailed simulation settings of population structure and transmission, and the saturation of the healthcare system, are taken into account for learning containment policies [10].

The use of continuous state space reinforcement learning to maintain physiological information and assist in sequential decisions for sepsis treatment has been explored [11]. But, offline policy selection remains challenging as candidate policies cannot be reliably assessed by allowing the random deployment of policies in the clinic [12]. To overcome this, clinician-in-the-loop approaches propose near-optimal sets of actions, instead of forcing a single automated offered action [13]. Seamless and extensive surveys reinforce the possibility of reinforcement learning for dynamic treatment, diagnosis, scheduling, and healthcare control, as well as identifying the issues of safety, generalizability and interpretability.

In spite of this progress, these factors are infrequently combined in a single public health framework: forecasting, multi-resource allocation, preparedness reserves, geographic equity, operational constraints, and explainability. It is the present approach that fills this gap, by leveraging LSTM-based epidemic-demand forecasting, the constrained version of Proximal Policy Optimization, and explanation via SHAP to facilitate coordinated deployment of health resources.

3. PROPOSED METHODOLOGY

3.1 Overall Framework and Public Healthcare Environment

The proposed system is a public health care network that coordinates the coordination of epidemic preparedness and resources distribution between 20 health care facilities with an adolescent population. There are eight urban, seven semi-urban and five rural facilities, representing a total of 50,000 adolescents ages 10-19 years. A planning horizon of 104 weeks is thought to cover routine healthcare demand, seasonal variation and varying epidemic conditions. The decisions on allocating resources are produced every seven days, the forecast for demand in healthcare is made over the next 14 days. The network configuration is summarised in Table 1.

Table 1. Configuration of the public healthcare environment

Component	Configuration
Healthcare facilities	20
Urban facilities	8
Semi-urban facilities	7
Rural facilities	5
Population represented	50,000 adolescents
Age range	10–19 years
Planning period	104 weeks
Allocation interval	7 days
Historical observation window	28 days
Forecasting horizon	14 days

The framework involves public-health data acquisition, epidemic-demand forecasting, constructing a healthcare state, deep reinforcement learning for allocation, and explanation of decision by SHAP, as shown in Figure 1. This data-acquisition component gathers facility-level epidemiological, demographic, operational, resource and environment factors. The forecasting part takes the previous 28 days of data and feeds it into two layers of long short-term memory (LSTM) network to forecast the demand for the next 14 days. The construction of healthcare-states incorporates all these forecasts, existing stocks, capacity levels, vulnerability of the population to the impacts of a healthcare state, geographical features, available budgets, and past decisions on healthcare-state configuration.

The outcome is fed into a limited Proximal Policy Optimization (PPO) agent that makes weekly bed, staff, testing kit, vaccine, medicine and emergency cash allocation decisions. A constraint-validation mechanism checks inventory, capacity, budget, critical-care, minimum-reserve and geographic-access requirements prior to acceptance of actions. The explanation component in the SHAP is then used to estimate the contribution of each input in the recommended allocation and the main reasons for the decisions at the facility level are presented. Unlike epidemic forecasting and resource allocation, it is an architecture of predictive modelling and adaptive and interpretable resource planning.

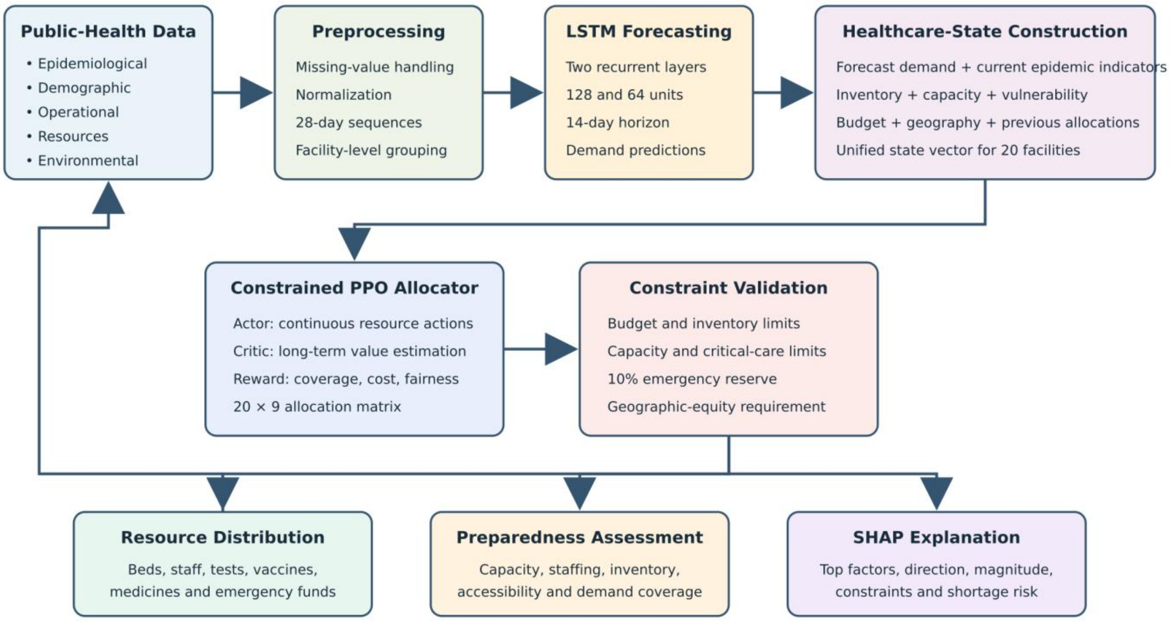

Figure 1. Overall architecture of the proposed explainable DRL framework.

3.2 Input Processing and Epidemic-Demand Forecasting

The framework adapts five sets of input variables. Confirmed cases, suspected cases, test-positivity rate, effective reproduction number, hospitalization rate, ICU admission rate, recovery rate, and vaccine coverage are examples of epidemiological variables. Demographic inputs consist of age distribution, prevalence of chronic conditions, socioeconomic vulnerability, population density, population geography, and adolescent population. Operational variables are outpatient volume, percentage of bed occupancy, occupancy of Intensive Care Units, delay in the waiting time, workload of staff, delay in delivery of supplies, and interfacility travel time. Resource variables are the number of beds, doctors, nurses, tests, vaccines, drugs, ambulances and emergency money available. The environmental factors are the temperature, humidity, mobility and school-activity indices.

If an observation is missing for a continuous variable, it is filled with the median value of that variable for the same facility and from the previous time window; if an observation is missing for a categorical variable, then it is filled with the mode for that variable for that facility and the previous time window. When there is no data, a missingness indicator is preserved if the absence of data is indicative of operational disruption. Min–max scaling is used for normalizing continuous variables:

$$x'_{j,t} = \frac{x_{j,t} - x_j^{min}}{x_j^{max} - x_j^{min} + \epsilon}, \qquad (1)$$

where $x_{j,t}$ is the value of variable j at time t, x_j^{min} and x_j^{max} are its minimum and maximum reference values, $x'_{j,t}$ is the normalized value, and ϵ is a small constant that prevents division by zero.

For each facility, a 28-day sequence is used, with the use of two-layer LSTM. The first layer is composed of 128 hidden units, and the second layer of 64 hidden units. The dropout rate of 0.20 prevents the recurrent layers to overfit. The proposed configuration sets the learning rate to be 0.001, batch size to be 64, and sets the number of training epochs to be 100 iterations. The state transition of the LSTM can be represented as follows:

$$h_{i,t}, c_{i,t} = LSTM\left(x_{i,t}, h_{i,t-1}, c_{i,t-1}; \theta_L\right), \qquad (2)$$

where $x_{i,t}$ denotes the normalized input vector of facility i at time t, $h_{i,t}$ is the hidden state, $c_{i,t}$ is the cell state, and θ_L represents the trainable LSTM parameters. The multi-output demand forecast is defined as:

$$\hat{d}_{i,t+1:t+14} = g\left(h_{i,t}^{(2)}\right), \text{\_\_\_\_\_\_\_\_\_\_\_\_\_\_} (3)$$

where $h_{i,t}^{(2)}$ is the final hidden representation produced by the second LSTM layer, g(·) is the output transformation, and $\hat{d}_{i,t+1:t+14}$ contains the predicted infections, outpatient visits, hospital admissions, ICU admissions, testing demand, vaccine demand, and medicine demand for the next 14 days. These predicted quantities become part of the state observed by the allocation agent.

3.3 Constrained PPO-Based Resource Allocation

Resource allocation is formulated as a constrained Markov decision process:

$$\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma, \mathcal{G} \rangle, \text{\_\_\_\_\_\_\_\_\_\_\_\_\_\_} (4)$$

where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{P}$ denotes the state-transition dynamics, $\mathcal{R}$ is the reward function, γ is the discount factor, and $\mathcal{G}$ represents the set of operational and equity constraints. The state of facility i at allocation time t is represented by:

$$s_{i,t} = \left[e_{i,t}, \hat{d}_{i,t}, q_{i,t}, c_{i,t}, v_{i,t}, b_t, a_{i,t-1}\right], \text{\_\_\_\_\_\_\_\_\_\_\_\_\_\_} (5)$$

here $e_{i,t}$ contains current epidemic indicators, $\hat{d}_{i,t}$ represents forecast demand, $q_{i,t}$ denotes available inventory, $c_{i,t}$ describes facility capacity, $v_{i,t}$ represents demographic and socioeconomic vulnerability, b_t is the remaining budget, and $a_{i,t-1}$ contains the previous allocation. For each decision interval, the agent produces an action matrix:

$$A_t = \left[a_{i,k,t}\right] \in \mathbb{R}^{20\times 9}, \text{\_\_\_\_\_\_\_\_\_\_\_\_\_\_} (6)$$

where $a_{i,k,t}$ is the quantity of resource k assigned to facility i. The 20 rows are healthcare facilities and the nine columns are general beds, isolation beds, ICU beds, physicians, nurses, testing kits, vaccine doses, medicine courses and emergency funds. Table 2 provides the resource pool and principal values of the PPO that will be considered to establish the allocation environment.

Table 2. Resource availability and PPO parameters

Category	Parameter	Value
Resource	General beds	1,000
Resource	Isolation beds	300
Resource	ICU beds	100
Resource	Physicians	150
Resource	Nurses	400
Resource	Testing kits per week	15,000
Resource	Vaccine doses per week	12,000
Resource	Medicine courses per week	8,000
Resource	Emergency budget per week	USD 500,000
PPO	Actor learning rate	0.0003
PPO	Critic learning rate	0.001
PPO	Discount factor	0.99
PPO	Clipping coefficient	0.20
PPO	Batch size	64
Constraint	Emergency reserve	10%

The composite reward balances healthcare coverage, containment, efficiency, preparedness, utilization, and fairness:

$$R_t = 0.30S_t + 0.20C_t + 0.15E_t + 0.15P_t + 0.10U_t + 0.10F_t - \lambda V_t, \text{\_\_\_\_\_\_\_\_\_\_\_\_\_\_} (7)$$

where S_t denotes normalized demand satisfaction, C_t represents containment of epidemic-related pressure, E_t denotes cost efficiency, P_t is the preparedness score, U_t represents resource utilization, F_t measures allocation fairness, $V_t$and λ controls the violation penalty. Each positive component is normalized to the interval [0,1] before aggregation. The allocations must remain within the available resource pool:

$$\sum_{t=1}^{20} a_{i,k,t} \leq Q_{k,t}(1-\rho_k) \quad k=1,\dots,9, \_\_\_\_\_\_\_\_\_\_\_\_(8)$$

where $Q_{k,t}$ is the available quantity of resource k, and ρ_k is the reserve proportion. A value of ρ_k=0.10 retains 10% of critical resources for unexpected demand. The budget constraint is expressed as:

$$L^{CLIP}(\theta) = \mathbb{E}_t\left[\min\left(r_t(\theta)\hat{A}_t, clip(r_t(\theta), 1-\varepsilon, 1+\varepsilon)\hat{A}_t\right)\right], \_\_\_\_\_\_\_\_\_\_\_\_(9)$$

where $r_t(\theta)$ is the probability ratio between the updated and previous policies, $\hat{A}_t$ is the estimated advantage, ε=0.20 is the clipping coefficient, and θ denotes the actor-network parameters. Clipping restricts excessively large policy changes and supports stable sequential allocation decisions.

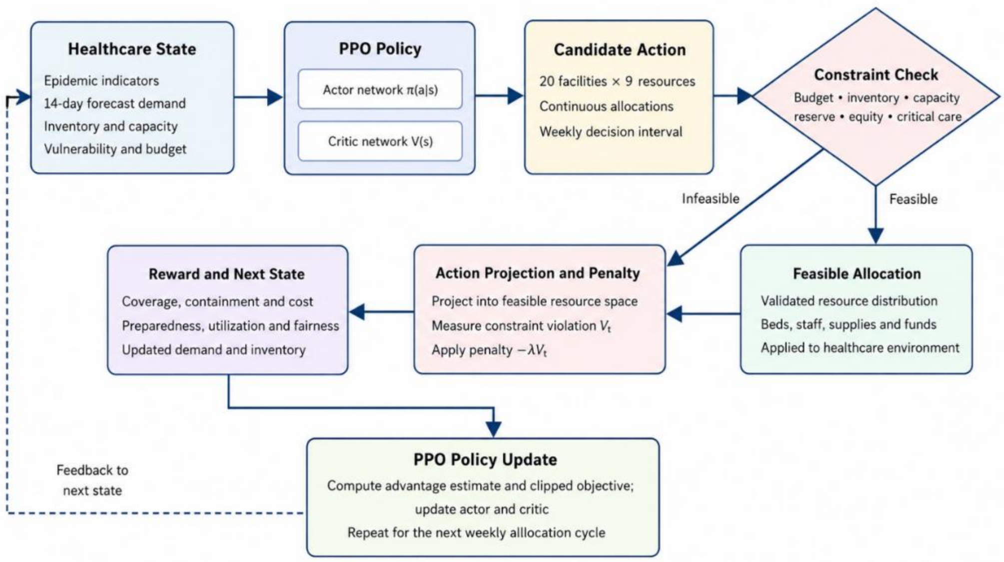

Figure 2. Constrained PPO-based resource-allocation workflow**.**

The environment state is outlined in the first column of Figure 2, based on forecast demand and conditions at the facilities. The creation of a candidate allocation by the actor network and the assessment of the network critic on the probable long term value of the allocation. The constraint layer verifies capacity, inventory, budget, reserve, critical-care and equity requirements. A feasible action is executed on the healthcare environment, while an infeasible action is thrown into the feasible allocation space, where it gets penalized using V_t. This resulting healthcare state then becomes the next healthcare state used for the next weekly decision cycle.

3.4 Explainability, Fairness, and Preparedness Assessment

SHAP is added to measure the contribution of individual state variable to a specific recommendation for resource allocation. An explanation for an allocation model f(x) is:

$$f(x) = \phi_0 + \sum_{j=1}^{m} \phi_j, \_\_\_\_\_\_\_\_\_\_\_\_(10)$$

where f(x) is the allocation predicted for a selected facility and resource, ϕ_0 is the expected allocation under the reference state, m is the number of input variables, and ϕ_j is the SHAP contribution of variable j. A positive ϕ_j supports an increase in the allocation, whereas a negative value supports a reduction.

Individual site explanations include the five most important variables, their direction and size, the variation they have from the preceding allocation, binding inventory or budget constraints, and the estimated likelihood of there being a shortage. For instance, if the incidence of the disease forecast is high, the proportion of the population vaccinated is low, ICU occupancy is high, and the socioeconomic vulnerability is high, the proportion of the population that is expected to receive the vaccine could increase. Conversely, if there is enough vaccine in stock and demand is low, the proportion of the population to receive the vaccine may be lowered. These explanations are designed to assist healthcare administrators in considering the rationale for the policy, but not to use the DRL output to explain the policy. To measure geographic fairness, the absolute difference between the rural and urban demand coverage is assessed:

$$G_t = \left|D_t^{urban} - D_t^{rural}\right|, \_\_\_\_\_\_\_\_\_\_\_\_\_\_(11)$$

where D_t^{urban} and D_t^{rural} denote the proportions of forecast demand covered in urban and rural facilities. The framework constrains G_t to remain within 10 percentage points while still accounting for differences in epidemic severity and clinical need. The facility preparedness score is calculated as:

$$P_{i,t} = 100\left(0.25K_{i,t} + 0.20H_{i,t} + 0.20I_{i,t} + 0.15R_{i,t} + 0.10A_{i,t} + 0.10D_{i,t}\right), \_\_\_\_\_\_\_\_\_\_\_(12)$$

where $K_{i,t}$ is capacity adequacy, $H_{i,t}$ is staff availability, $I_{i,t}$ is inventory sufficiency, $R_{i,t}$ is emergency-reserve adequacy, $A_{i,t}$ is geographic accessibility, and $D_{i,t}$ is predicted-demand coverage. Each component is normalized between zero and one, producing a final preparedness score from zero to 100.

The decision constraints are set so that only 10% of the services in the rural and urban areas are different, only 3% of vaccines are wasted and only 5% of unmet demand will be in the ICU. An emergency reserve of 10% is provided for critical resources and budget and facility capacity violations are considered unacceptable actions. These values are used to set the operational requirements of the allocation policy, and do not necessarily represent outcomes until accompanied by completed evaluations data.

4. RESULTS AND DISCUSSION

4.1 Evaluation Design and Comparison Methods

The framework will be tested during scenarios of normal, moderate and severe outbreaks. These scenarios will simulate a level of healthcare demand, higher numbers of infections and admissions, and extreme stress with high levels of ICU beds used, shortages of items, and staffing limitations. To allow for a common benchmarking, a consistent set of healthcare network, resource availability, budget, and demand will be followed across all the methods.

The proposed approach will be contrasted to the equal allocation, population-proportional allocation, priority-rule allocation, linear programming, Deep Q-Network and conventional PPO methods. These are optimization-based, static and reinforcement-learning strategies respectively. Demand satisfaction, resource utilization, unmet critical demand, epidemic peak reduction, vaccine wastage, operational cost, waiting time, service gap between rural and urban areas, preparedness score, decision latency, and explanation fidelity will be used as assessment methods for performance. Each epidemic scenario will be analyzed in terms of allocation efficiency, outbreak response preparedness, geographic equity, computational efficiency and interpretability.

4.2 Resource-Allocation and Epidemic-Preparedness Analysis

To explore the distribution of beds, people, testing kits, vaccines, medicines and resources for emergencies in Urban, Semi-Urban, Rural facilities. Predicted infections, hospital occupancy, population vulnerability, inventory availability, and geographic access will be used to look at allocations.

In a normal condition, the allocation stability, emergency reserves and resource wastage will be the focus. Moderate and severe outbreaks will be discussed on the aspects of resource redirection, ICU and staff shortage, demand growth in short period and the budget constraints. The allocation of the resources (people, equipment, etc.) will be compared for the three epidemic cases in Figure 3.

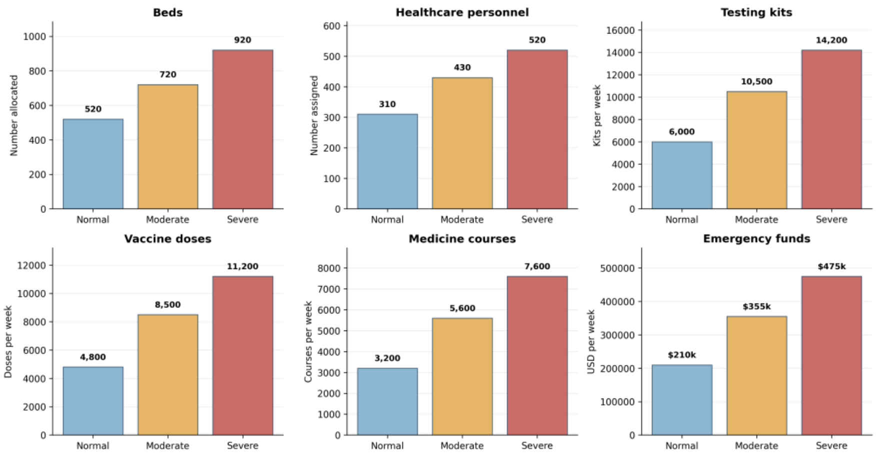

Figure 3. Healthcare resource allocation under different epidemic scenarios**.**

All 20 facilities will be evaluated for preparedness, which will cover capacity, staffing, inventory, reserves, accessibility and coverage of demand. A comparison of facility types based on preparedness scores and the 85/100 score will be presented in Figure 4.

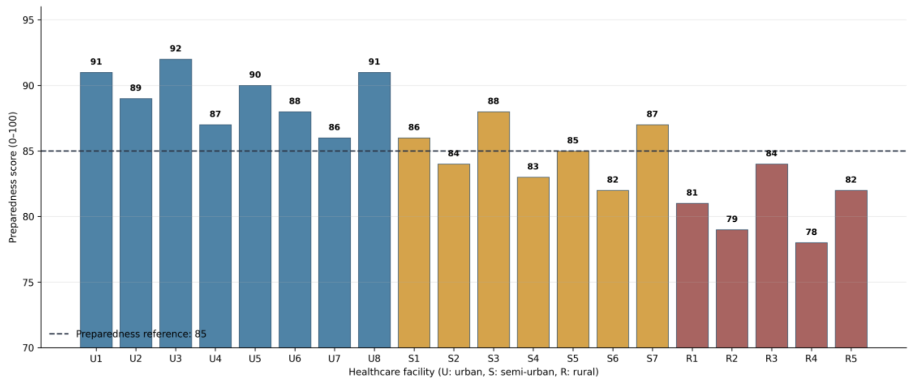

Figure 4. Facility-level epidemic preparedness scores.

The design objectives are: 92% demand satisfaction, 90% use of resources, 5% "unmet critical demand", 20% reduction of the epidemic peak, 3% vaccine wastage and an 85/100 preparedness score. These values are ‘evaluation targets' not findings.

The values obtained by the design in Table 3 are not the values obtained by the results. They represent the desired level of operation that the allocation framework is based upon.

Table 3. Design objectives for resource allocation and preparedness

Indicator	Design objective
Demand satisfaction	92%
Resource utilization	90%
Unmet critical demand	5%
Epidemic peak reduction	20%
Vaccine wastage	3%
Preparedness score	85

4.3 Comparative, Explainability, and Fairness Analysis

The proposed method will be compared with other methods such as equal allocation, allocation based on population, allocation based on priority rule, linear programming, DQN and conventional PPO. These will be studied in the comparison of the following: demand satisfaction, resource utilization, unmet critical demand, operational cost, service gap between rural and urban areas, preparedness, and decision latency. The target for satisfaction of demand is 92%, resource utilisation 90%, not satisfied if critical demand is 5%, test the gap between rural and urban 10 percentage points and score 85/100 for preparedness.

SHAP will measure the effect of the predicted incidence, vaccination coverage, ICU occupancy, socioeconomic vulnerability of the population, population density, staff availability, and resource inventory. 90% will be the reference objective in Explanation fidelity. The measure of fairness will be done based on the rural–urban service disparity, vulnerability adjusted demand coverage, and allocation consistency.

4.4 Sensitivity, Ablation, and Practical Implications

The level of resources will range from 70% to 130% availability; the reproduction number from 0.5 to 3.0; weekly emergency funding between USD 300,000 and USD 700,000; forecasting uncertainty between 5% and 25% and emergency reserves between 5% and 20%. LSTM forecasting, fairness reward, preparedness term, safety constraints, and SHAP explanations will be removed one by one via an ablation process.

The changes in the demand coverage, shortage, utilization, cost, equity, preparedness, policy stability, and infeasible allocations will be analyzed. Reliable public-health data, privacy protection and administrative oversight, audit mechanisms, region validation and prospective assessment will have to be implemented for practical deployment.

CONCLUSION

This paper proposes an interpretable deep reinforcement learning method for intelligent resource allocation and epidemic preparedness in public health care systems. It unites 14-day forecasting based on LSTM with constrained PPO, a framework that controls beds, health workers, testing kits, vaccines, medicines, and funds in urban, semi-urban and rural health centers. Capacity, budget, emergency-reserve, critical-care and geographic-equity are built into to ensure that allocation decisions are feasible and responsible. SHAP explanations have the useful property of revealing epidemiological, demographic and operational factors driving each recommendation, thus leading to better transparency.

The framework offers a structured approach to assessing the satisfaction of demand, use of resources, critical shortages, preparedness, cost-effectiveness, geographic equitability and fidelity of explanations of the epidemic under various epidemic scenarios. It is not possible to know whether it is effective until it is fully assessed and validated externally. Future research should focus on real public-health data, various disease scenarios and other regional settings, prospective multi-institutional research, and evaluation of privacy, fairness, usability, safety, and regulatory issues before implementation.

REFERENCES

[1] Bushaj, Sabah, et al. "A simulation-deep reinforcement learning (SiRL) approach for epidemic control optimization." Annals of Operations Research 328.1 (2023): 245-277.
[2] Coronato, Antonio, et al. "Reinforcement learning for intelligent healthcare applications: A survey." Artificial intelligence in medicine 109 (2020): 101964.
[3] Frommeyer, Timothy C., et al. "Reinforcement learning and its clinical applications within healthcare: A systematic review of precision medicine and dynamic treatment regimes." Healthcare. Vol. 13. No. 14. MDPI, 2025.
[4] Gottesman, Omer, et al. "Guidelines for reinforcement learning in healthcare." Nature medicine 25.1 (2019): 16-18.
[5] Huang, Yunhan, and Quanyan Zhu. "Game-theoretic frameworks for epidemic spreading and human decision-making: A review." Dynamic Games and Applications 12.1 (2022): 7-48.
[6] Killian, Taylor W., et al. "An empirical study of representation learning for reinforcement learning in healthcare." arXiv preprint arXiv:2011.11235 (2020).
[7] Komorowski, Matthieu, et al. "The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care." Nature medicine 24.11 (2018): 1716-1720.

[8] Kwak, Gloria Hyunjung, Lowell Ling, and Pan Hui. "Deep reinforcement learning approaches for global public health strategies for COVID-19 pandemic." PloS one 16.5 (2021): e0251550.
[9] Libin, Pieter JK, et al. "Deep reinforcement learning for large-scale epidemic control." Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Cham: Springer International Publishing, 2020.
[10] Mehrjou, Arash, et al. "Pyfectious: An individual-level simulator to discover optimal containment policies for epidemic diseases." PLOS Computational Biology 19.1 (2023): e1010799.
[11] Raghu, Aniruddh, et al. "Continuous state-space models for optimal sepsis treatment: a deep reinforcement learning approach." Machine learning for healthcare conference. PMLR, 2017.
[12] Tang, Shengpu, and Jenna Wiens. "Model selection for offline reinforcement learning: Practical considerations for healthcare settings." Machine Learning for Healthcare Conference. PMLR, 2021.
[13] Tang, Shengpu, et al. "Clinician-in-the-loop decision making: Reinforcement learning with near-optimal set-valued policies." International Conference on Machine Learning. PMLR, 2020.
[14] Yu, Chao, et al. "Reinforcement learning in healthcare: A survey." ACM Computing Surveys (CSUR) 55.1 (2021): 1-36.