This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2022.3187950

Multi-Objective Reinforcement Learning Based Healthcare Expansion Planning Considering Pandemic Events

Salman Sadiq Shuvo, Hasan Symum, Md Rubel Ahmed, Yasin Yilmaz, José L. Zayas-Castro
University of South Florida, Tampa, FL, USA
{*salmansadiq, hsymum, mdrubelahmed, yasiny, josezaya*}*@usf.edu*

***Abstract*—Hospital capacity expansion planning is critical for a healthcare authority, especially in regions with a growing diverse population. Policymaking to this end often requires satisfying two conflicting objectives, minimizing capacity expansion cost and minimizing the number of denial of service (DoS) for patients seeking hospital admission. The uncertainty in hospital demand, especially considering a pandemic event, makes expansion planning even more challenging. This work presents a multi-objective reinforcement learning (MORL) based solution for healthcare expansion planning to optimize expansion cost and DoS simultaneously for pandemic and non-pandemic scenarios. Importantly, our model provides a simple and intuitive way to set the balance between these two objectives by only determining their priority percentages, making it suitable across policymakers with different capabilities, preferences, and needs. Specifically, we propose a multi-objective adaptation of the popular Advantage Actor-Critic (A2C) algorithm to avoid forced conversion of DoS discomfort cost to a monetary cost. Our case study for the state of Florida illustrates the success of our MORL based approach compared to the existing benchmark policies, including a state-of-the-art deep RL policy that converts DoS to economic cost to optimize a single objective.**

***Index Terms*—multi-objective optimization, deep RL, markov decision process, healthcare capacity expansion, denial of service.**

I. Introduction

Healthcare is a universal need that includes health promotion, prevention, treatment, rehabilitation, and palliative care. The distribution, management, operation, augmentation, and demand of healthcare facilities have become a delicate and crucial reality for our existence in this world. Pandemic events such as COVID-19 have highlighted the lack of a resilient and sustainable augmentation plan for healthcare facilities, even in developed countries. High population growth and the nationwide increase of median age in the US [1] indicate the need for widespread healthcare facilities.

The dynamics of hospital bed demand results from a wide range of stochastic variables, making it very challenging to model the future demand and augmentation scenarios. Emergency department crowding, natural disasters, and humanitarian crises are often not adequately addressed in the current annual development plans. Augmentation plans based on the yearly demand statistics can often be misleading [2]. The number of beds is often increased by observing the local population's needs, known as the Certificate of Needs (CON). In many states in the US, the hospital bed capacity is regulated based on the CON [3]. This method aims to maintain a target occupancy level of hospital capacity to minimize expenditures. The large number of casualties caused by the COVID-19 pandemic proved that this method has limitations in forecasting future bed demands. Lack of treatment often causes irreparable damage to the patients and families, physically and psychologically. Therefore, hospital bed demand forecasting and facility augmentation planning need meticulous attention from the planners, administrators, and research community to ensure sustainable and accessible healthcare for all.

This study aims to address this research gap in hospital capacity expansion planning, especially under pandemic events like COVID-19. We include critical features in forecasting hospital bed demand for making augmentation plans. Firstly, different age-based population groups (e.g., infants, older people) require different levels of hospitalization, hence the age distribution is a primary factor in hospital occupancy forecasting. Secondly, Disease Burden (DB), which represents the hospital dependency of the residents in a region for critical diseases, is another major factor for hospitalization requirement. Moreover, the Social Vulnerability Index (SVI) of a region represents the vulnerability of its residents to diseases. Finally, a pandemic event is another factor that shapes the hospitalization need.

Beside the human health factors, the economics of maintaining hospitals should also be considered in the augmentation plan as the demand and supply in this sector is non-trivial. Maintaining an enormous capacity to meet uneven demand is not economically sustainable. Furthermore, cost of goods and services vary region to region because of transportation costs, tariff/taxes, or other reasons. Different administrative regions control the prices of goods/services in different ways, which can be summarized by the regional price parity index (PPI). PPI measures the cost of goods and services compared to the national average, making it a good regional cost indicator for establishing, expanding, and operating a hospital. Beside serving the health needs of community, for-profit and non-profit hospitals are significant revenue and employment providers locally and nationally. An oversight to these critical health and economic factors while devising an augmentation plan can significantly harm a region's health and economy. A robust, dynamic, and detailed hospital augmentation plan can benefit both the government and private parties, underscoring the scope of this work for

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2022.3187950

the policymakers.

For a sustainable solution to this highly stochastic problem [4], [5], we utilize the important factors discussed above in a systematic artificial intelligence (AI) framework, deep reinforcement learning (RL). We propose a Multi-Objective Reinforcement Learning (MORL) method based on deep neural networks to satisfy two objectives: minimize augmentation cost and Denial of Service (DoS) to the patients. In our preliminary work [6], we proposed an RL approach that converts the patients' discomfort caused by DoS into monetary cost through fixed coefficients. However, defining fixed coefficients for different places and periods is not feasible, causing a practical challenge for the applicability of the work [6]. To this end, in this work, motivated by the Pareto-optimal Q-learning (PQL) method [7] we propose multi-objective actor-critic method to avoid the forced conversion of DoS discomfort to monetary cost. Since we need to deal with high-dimensional state and action spaces for hospital augmentation planning, we utilize deep neural network based approximations for the MORL task, similar to [8], [9].

In the proposed method, the healthcare authority is the MORL agent that selects a region for hospital augmentation at each decision time (e.g., annually) by considering several important factors, the age-partitioned population, DB, SVI, PPI, and the existing hospital capacity for all regions. As a result of its augmentation actions, the agent observes the DoS and expansion costs. We modify Advantage Actor-Critic (A2C) [10], a popular deep RL algorithm, to address the considered MORL problem. The contributions of this work can be summarized as follows.

- A novel Markov decision process (MDP) formulation based on important health and economic factors, such as DB, SVI, and PPI, is proposed to learn the optimal hospital capacity expansion policy which minimizes the expansion cost and DoS.
- A novel deep MORL algorithm is developed based on the actor-critic framework.
- An extensive case study is performed for the state of Florida using real data to evaluate the proposed MORL approach.

The rest of the paper is arranged as follows. Section II presents literature review for hospital capacity expansion planning. We present the MDP formulation in Section III, and the proposed deep MORL algorithm in Section IV. The experimental setup, results, and discussions are given in Section V, Section VI, and Section VII, respectively, and the paper is concluded in Section VIII.

II. Related Work

Proactive planning to address hospital bed occupancy problems and future expansion decisions under population changes and emergencies have been a critical problem for hospitals and care providers. The challenges in hospital bed management and expansion decision have been approached by several researchers based on the different understanding of the problem [11], [12]. The hospital bed occupancy and expansion decision literature can be divided into two major areas: (1) bed occupancy management and allocations within a hospital and (2) capacity planning and allocation of the hospital beds within a region. In the first type of study, researchers typically proposed a mathematical framework addressing systemic issues such as overcrowding within the hospital settings focusing on optimum use of healthcare resources that maximize bed usage and reduce boarding time. These studies include forecasting hospital bed occupancy and resources, healthcare personnel and critical resource allocation, and patient allocation and ambulance diversion [13], [14]. Prior studies in this area are widely varied by hospital division (e.g., psychiatric, emergency medicine, and maternity ward), care delivery setting (e.g., trauma hospital, children hospital, and specialty care), forecasting horizon (e.g., one hour to seven days), hospital resources (e.g., ICU bed, ventilation equipment, and physicians), patient case-mix setting (e.g., children, elderly, and pregnancy) and data-source (e.g., EMR, EHR, and clinical data) [15]–[18]. However, the majority of these forecasting and resource allocation-based studies focused on supporting optimal use of crucial healthcare resources within the hospital setting rather than long-term bed expansion planning. Given the importance of long-term hospital bed capacity and geographical allocation, we focus our study on models intended to address hospital bed expansion within a region considering the increased demand, shifts in population demographics, and emergencies such as COVID-19.

There are a few existing studies that considered capacity and expansion decisions for the medium or long-term planning horizon. These studies implemented various forecasting methods, including the simple ratio method, formula method, Michigan's bed need model, and usage projection model to predict estimates at different regional settings (e.g., county, city) [12], [19]. Implementing these methods into long-term hospital expansion decision-making might lead to several critical limitations [11]. First, they do not consider the importance of complex interactions between the hospital bed need and population demographics changes, which might play a fundamental role in determining decisions [20]. Second, most of the studies faced challenges in forecasting accuracy, model fitting (e.g., over and underestimating), and incorporating geographical and hospital administration variations. It is suggested that an alternative robust decision support model incorporating uncertainty might have the potential to reliably predict hospital capacity planning and bed extension decisions for the regions with rapidly changing demographics and patient case-mix population [5], [21].

A promising alternative approach for hospital expansion decision consists of modeling with data-driven approaches and constrained optimization in the decision-making framework [19], [22]. A few studies showed potential with improved prediction performance for forecasting bed occupancy in various hospital settings and geographical regions through the data-driven forecasting approach [23], [24]. These studies used various statistical and machine learning (ML) methods, such as linear regression models. However, most of these studies are limited to forecasting, considering only

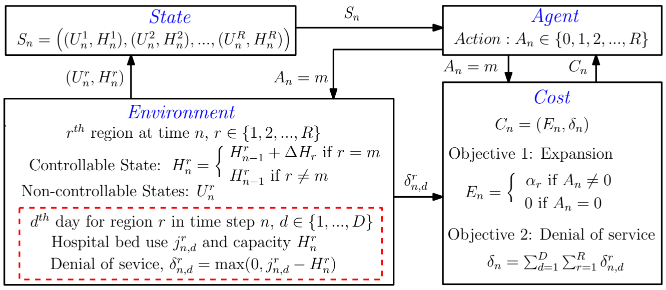

Fig. 1. Proposed multi-objective MDP (MOMDP) model.

the patient volume. A few studies utilized several neural network-based algorithms in forecasting intensive and critical care bed usage, surgical room prediction, and overall bed capacity estimations [24], [25]. However, only a handful of studies implemented ML-based methods to investigate hospital expansion planning at the regional and state level.

With the recent theoretical and technical achievements in RL approaches, the application of deep RL methods can potentially integrate prediction models with optimizing conflicting multiple objectives. Therefore, RL-based methods have been widely used in various applications areas, including robotics, virtual reality, finance, communications, and transportation [26]–[28]. The applications of RL-based methods in the healthcare domain are mainly focused on adverse outcome predictions, rather than healthcare policy-related decision making [29]. Based on the RL-based studies in non-healthcare settings, RL-based algorithms have the potential to improve the hospital augmentation design with capabilities of incorporating multiple decision criteria and critical covariates under the same framework [30]. The works in [31], [32] utilized RL to determine the optimal size of hospital capacity augmentation; however, these methods neglect the fact that the capacity expansion usually happens in bulk numbers (e.g., 120-bed extension unit) [33], [34]. Furthermore, they did not consider interactions between covariates (e.g., patient case-mix and changes in demographics) and appropriate health administration division, which may significantly influence the hospital bed augmentation decisions [35]. Also, these studies assumed a single isolated objective and demand targets that are not necessarily main factors for hospital bed expansion decisions [20], [36]. Unlike previous approaches, our study deals with multiple decision criteria for hospital expansion decision making. In particular, our method aims to simultaneously minimize the capacity expansion cost and the number of service denials.

III. Proposed Decision Model

The proposed MORL method is based on a Markov decision process (MDP) specifically designed for the considered healthcare expansion planning problem. The proposed MDP formulation follows the Markov property: transition to the next state depends only on the agent's current state and action. Fig. 1 shows our multi-objective MDP (MOMDP) model, where the healthcare administration is the agent which manages the healthcare facilities in R regions. The system state at time n, S_n, is defined by the non-controllable state variables U_n^r and the controllable state (hospital bed capacity) H_n^r of each region r. Every time step n, the agent takes action $A_n = m$, which means selecting the m^{th} region for a capacity expansion of ΔH_m beds at the cost of $E_n = \alpha_m$. The agent can decide to decline expansion ($A_n = 0$), resulting in a total of $R + 1$ decision options. The number of patients requiring hospitalization for a day can be greater than the region's capacity. In those days, some patients face with denial of service (DoS). $\delta_{n,d}^r$ denotes the number of DoS for region r in day $d \in \{1, 2, ..., D\}$ of time step n, where D is the total number of days in a time step. The agent has two conflicting objectives:

- Minimize the cumulative capacity expansion cost, $\sum_n E_n$,
- Minimize the total DoS, $\sum_n \delta_n = \sum_n \sum_d \sum_r \delta_{n,d}^r$.

While the expansion actions ($A_n \neq 0$) incur monetary cost, they also reduce the future number of DoS. The agent aims to simultaneously minimize the monetary cost and DoS (i.e., find an optimum balance between them) over a finite time horizon by taking optimal actions. Before explaining our MORL solution to this problem, we next elaborate the state variables and the cost function.

A. State, $S_n = (U_n^r, H_n^r)$

1) Non-controllable States, U_n^r: Selecting appropriate variables for the state definition is a critical task in MDP formulation. The agent gathers essential information from the environment by observing several state variables to inform its actions. The variables which are not directly affected by the agent's actions are called the non-controllable states in our model. The required hospital capacity has a strong correlation with the following factors, which we choose as non-controllable states for our model.

- The work [6] shows that the number of hospital admissions is better captured by age-grouped population data, consistent with the general understanding that some age groups require more medical attention (especially children and older people). The age-partitioned population of the r^{th} region at time n, $p_n^r = [p_n^{r1}, \ldots, p_t^{rG}]$, is a vector of G age groups. We separate the population among 4 age groups: 0-18, 18-44, 44-65, and 65+ years for our case study.
- Disease Burden, DB_n^r, represents the age adjusted death rate per 100,000, which ranges between 450 and 1600 for the regions in our experiment.
- Social Vulnerability Index, SVI_n^r, represents the vulnerability of the population towards diseases and ranges between 0 and 1. SVI is a surrogate measure of the potential negative effects on communities caused by external stresses on human health [37]. Another relevant measure Health Deprivation Index (HDI) is available primarily at the census block group level, which can be used for fine-grained modeling at the neighborhood level. Since larger regions for healthcare administration are considered in this work, we prefer SVI, which is available at the county or census tract level.

- Price Parity Index, PPI_n^r, represents the cost of living in a region normalized by the national average.
- During pandemics the healthcare system allocates part of its capacity to deal with those pandemic-affected patients, which significantly changes the environment. Therefore, we include it in the non-controllable states as a single binary variable $\text{Pand}_n \in \{0, 1\}$. This pandemic flag may also cover other humanitarian crises due to natural disasters or other catastrophic events.

The agent only observes these states from the environment, but cannot control them. In the experiments in Section V, we explain how to reliably estimate these state variables using real-world data. We include variance in the estimated values for these non-controllable states to simulate a realistic environment in the case study.

2) Controllable States, H_n^r: The agent's action controls the hospital bed capacity for each region, which is the only controllable state in this setup. The current hospital capacity for the r^{th} region at time n is given by

$$H_n^r = H_{n-1}^r + \Delta H_n^r = H_0^r + \textstyle\sum_{\tau=1}^n \Delta H_\tau^r,$$

where $\Delta H_n^r = \Delta H_r$ if the region is selected for capacity expansion ($A_n = r$), otherwise $\Delta H_n^r = 0$. The expansion size ΔH_r may vary among the regions. H_0^r is the initial hospital capacity for the region at the beginning of the study.

B. Cost, C_n

Since we have two objectives, the cost in this MOMDP setup is a vector $C_n = (E_n, \delta_n)$.

1) Expansion cost, E_n: The agent can implement capacity expansion by building a new hospital or augmenting an existing facility. The different capacity expansion size ΔH_r for each region incurs the expansion cost α_r. We assume the healthcare authority has the proper understanding to determine these parameters in practice, as demonstrated in our case study. Hence, the capacity expansion cost $E_n = \alpha_r$ varies for different actions $A_n = r$, and $E_n = 0$ for the no capacity expansion decision.

2) DoS, δ_n: The per capita (per 1000 people) hospital bed capacity varies widely among countries, e.g., Japan has 13 hospital beds per capita while Mali has only 0.1 [38]. The US has a moderate per capita of 2.5, where South Dakota leads the chart with 4.8 in comparison with Oregon's 1.6 per capita hospital bed [39]. For any region in the world, the actual hospital admission on a given day can be more than the available capacity, especially during pandemic times such as COVID-19. Since it is not financially feasible to maintain capacity capable of providing healthcare for all scenarios, the healthcare authority tries to maintain a reasonable capacity. However, the patients living in lower per capita capacity regions are prone to more frequent DoS. The DoS for the r^{th} region is

$$\delta_{n,d}^r = \max(0, j_{n,d}^r - H_n^r),$$

where $j_{n,d}^r$ is the hospital admission requirement for day $d \in \{1, 2, ..., D\}$ at time step n. So, the total number of DoS for the n^{th} time step is

$$\delta_n = \sum_{r=1}^R \sum_{d=1}^D \delta_{n,d}^r.$$

IV. Solution Approach

A. Multi-Objective Reinforcement Learning

In a Reinforcement Learning (RL) setup, the agent takes an action that changes the environment, and the environment responses by providing an immediate reward/cost. In the standard setting, the goal of the RL agent is to maximize the discounted cumulative reward $R_N = \sum_{n=0}^N \gamma^n R_n$ by taking optimal actions over a time horizon of N steps. The discount factor $\gamma \in (0, 1)$ determines the weight of future rewards/costs relative to the immediate one for the RL agent. The traditional way to obtain a scalar reward from the multiple costs present in the original objectives (i.e., expansion cost and DoS in our case) is to combine them using a conversion parameter, $R_n = -E_n - \beta\delta_n$. There is a significant challenge in setting the conversion parameter β to a realistic value since it is in general not obvious what the conversion rate should be. Specifically, DoS is a health-related discomfort cost for the patients, which is not easy to convert into a monetary cost like the expansion cost. Although one can find studies that try to assign economic value to such an important discomfort cost, there is no unique and optimum way of doing this. Avoiding such a forced conversion, we treat each objective in a natural way through a deep MORL algorithm.

To this end, instead of a single value function used for the scalarized cost in the traditional RL approach, we define two value functions for the expansion cost and DoS, which are given by the Bellman equations [10]

$$\begin{aligned} V_E(S_n) &= \max_{A_n} \left\{ \mathsf{E}\left[-E_n + \gamma_E V_E(S_{n+1}) | S_n, A_n\right] \right\}, \\ V_\delta(S_n) &= \max_{A_n} \left\{ \mathsf{E}\left[-\delta_n + \gamma_\delta V_\delta(S_{n+1}) | S_n, A_n\right] \right\}. \end{aligned} \tag{1}$$

The value functions $V_E(S_n)$ and $V_\delta(S_n)$ represent the maximum expected reward at a certain state achievable by taking the optimum actions in the current time step and in the future.

B. Deep MORL

Our MOMDP model consists of $8R$ states and $R + 1$ actions for R regions. This high-dimensional state-action space requires neural network (NN) based approximations to learn the value functions in Eq. (1). The NN-based RL approaches are called deep RL, and Advantage Actor-Critic (A2C) is a popular deep RL technique. A2C is known to be more successful for high-dimensional action space than its most prominent alternative Deep Q Network (DQN) [10] and thus is suitable for our problem. A2C uses a function called the advantage function for policy update to address the high variance problem of its predecessor, the REINFORCE algorithm [10]. We propose a multi-objective A2C algorithm for the considered MORL problem, following the Pareto

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2022.3187950

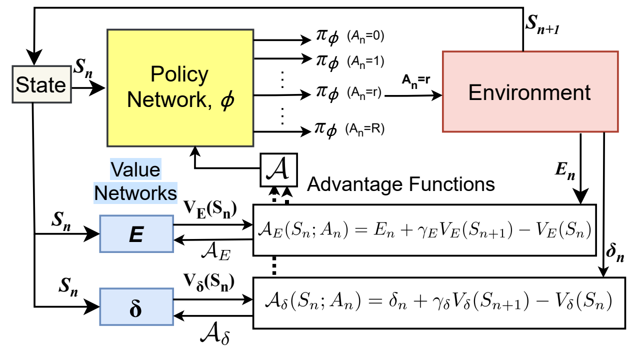

Fig. 2. Proposed multi-objective A2C architecture.

optimality approach [9]. Fig. 2 shows the proposed multi-objective A2C architecture. The pseudo code is also given in Algorithm 1. A2C uses two different type of networks, the actor network and the critic network. In our multi-objective A2C architecture, while there is a single actor network for action decisions, we utilize two critic networks for the two objectives, as explained next.

1) Critic (Value) networks: The two *value networks* for the two objectives aim to learn the value functions $V_E(S_n)$, and $V_\delta(S_n)$ for a given state S_n. Based on the agent's action A_n, the target values are estimated from the immediate cost and the value function for the next state S_{n+1}. Then, the advantage functions for the state action pair (S_n, A_n) are calculated as the difference between the target values and the predicted values:

$$\begin{aligned} \mathcal{A}_E(S_n; A_n) &= -E_n + \gamma_E V_E(S_{n+1}) - V_E(S_n), \\ \mathcal{A}_\delta(S_n; A_n) &= -\delta_n + \gamma_\delta V_\delta(S_{n+1}) - V_\delta(S_n). \end{aligned} \tag{2}$$

The value networks use the advantage functions as the temporal difference error for gradient descent update through backpropagation. They are called *critic networks* as they guide the policy network about the quality of its actions through the advantage functions.

2) Actor (Policy) network: The *policy network* outputs probability $\pi_\phi(A_n = r)$ for each action through a softmax function, i.e., $\sum_{r=0}^{R} \pi_\phi(A_n = r) = 1$. It aims to maximize the expected return $J(\pi_\phi)$ by performing gradient ascent with respect to the weights ϕ of the NN through the following equation:

$$\nabla_\phi J(\pi_\phi) = \mathsf{E}_{\pi_\phi}[\nabla_\phi \log(\pi_\phi(A_n|S_n))\mathcal{A}(S_n; A_n)]. \tag{3}$$

While updating the critic networks by their corresponding advantage functions is straightforward in this multi-objective setup, we define the following advantage function for the actor network

$$\mathcal{A} = \begin{cases} w_E \mathcal{A}_E + w_\delta \mathcal{A}_\delta, & \text{if } |\mathcal{A}_E + \mathcal{A}_\delta| = |\mathcal{A}_E| + |\mathcal{A}_\delta| \\ 0, & \text{otherwise,} \end{cases} \tag{4}$$

where $w_E + w_\delta = 1$. The coefficients w_E and w_δ reflect the priority of the policymaker for the two objectives mentioned above. Such a flexibility is missing in [9]. Notably, the actor network is updated only when both advantage functions

Algorithm 1 Multi-objective A2C algorithm (Fig. 2)

Input: discount factors γ_E and γ_δ, objective weights w_E and w_δ, learning rate α.
Initialize policy network with random weights ϕ and the value networks with random weights E and δ.
for episode $= 1, 2, \ldots$ **do**
  Initialize the MOMDP, obtain the initial state S_0;
  for n $= 1, 2, \ldots, N$ **do**
    Sample action A_n, from probability distribution generated by the actor network ϕ.
    Execute action A_n, and observe reward vector $R_n = [-E_n, -\delta_n]$ and next state S_{n+1}.
  end for
  Calculate the advantage functions for the value networks from Eq. (2).
  Update policy network ϕ using the advantage function $\mathcal{A}$ (Eq. (4)) in gradient ascent (Eq. (3)).
  Update value networks E and δ using their advantage functions $\mathcal{A}_E$, and $\mathcal{A}_\delta$.
end for

have the same sign (both positive or both negative), as seen in Eq. (4). This intuition is in line with the Pareto optimality discussed in [9], which prescribes to update only when the gradient ascent directions (advantage functions) corresponding to all objectives are the same. Updating in the same gradient ascent direction will discover new undominated points on the Pareto front. On the contrary, different gradient ascent directions for different objectives indicate the discovery of dominated points since they do not increase all objectives concurrently. Hence, we do not update the actor network when the advantage functions have different directions.

V. Experimental Setup

Having warm tropical weather, Florida is an attractive retirement home for an increasing number of older people in the US. Older people are more prone to medical care and longer stays in hospital. The high population growth in both infant and older age groups requires robust planning and expansion of healthcare facilities in Florida. So, we assess our MORL policy for Florida, where the Agency for Healthcare Administration (AHCA) can represent the MORL agent. AHCA grouped the 67 counties of the state in $R = 11$ regions or health districts, as shown in Fig. 3.

A. Data Generation

The Bureau of Economics and Business Research provides Florida's county-wise population history and projections up to the year 2045 [40]. We extracted historical hospital admission, Disease Burden (DB), and Social Vulnerability Index (SVI) data between 2010-2019 from State Inpatient Databases (SID) [41].

Although the US Bureau of Economic Analysis (BEA) publishes the state-wise Price Parity Index (PPI) [42], county-wise PPI data is yet to be published. Since PPI has a strong

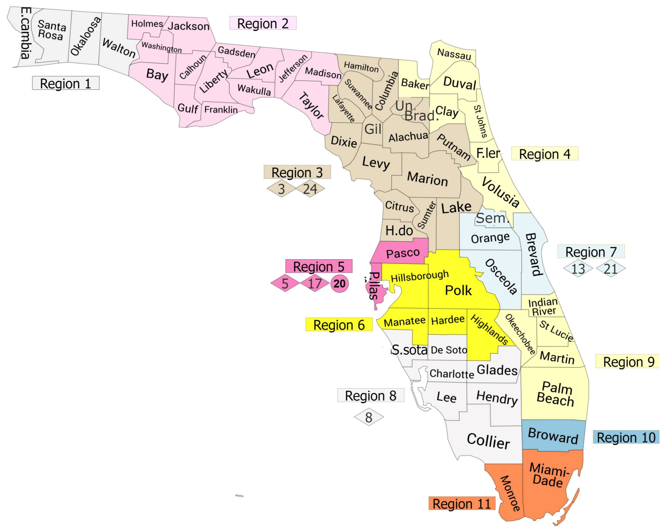

Fig. 3. Map of the 11 health regions of Florida for the case study. Diamond shape with number n next to the region label indicates capacity expansion decisions for that region by the proposed MORL in year n for both scenarios, whereas round shape indicates expansion for the pandemic scenario only.

correlation with the household income of a region [42], we generate the PPI data for the r^{th} region in our case study as follows

$$PPI_n^r = \frac{HI_n^r}{Med(HI_n)}.$$

HI_n^r is the household income for the region, and $Med(HI_n)$ is the median of household incomes for all 67 Florida counties. We found that Region 6 is the costliest in Florida, which closely matches the BEA's [42] map of real personal income and regional price parity map of the major metropolitan areas in the US. Hence, establishing and extending hospital facilities in Region 6 will be the costliest in Florida. We devise yearly decisions in each policy to expand the capacity of a region with $\Delta b = 120$ hospital beds. We set the cost of adding 120 hospital beds with a normal distribution of mean $\mu = 50$ and variance $\sigma^2 = 3$ M USD for Florida [33], [34]. So, the expansion cost at the n^{th} time step depends on the PPI of the selected region as in

$$E_n = \alpha_r = PPI_n^r \times \mathcal{N}(\mu, \sigma^2).$$

Instead of projecting PPI values, we follow the PPI data of the year 2019 for the simulation period, i.e., $PPI_n^r = PPI_{2019}^r, \forall n$.

B. Hospital Occupancy Forecasting

Historical hospital admission for the regions was obtained from the Florida State's healthcare website [41]. Although the hospital admission requirement for an area depends on multiple factors, we hypothesize that the elements in our MOMDP state space U_n^r (except PPI) be sufficient to predict future hospital bed requirements. In this regard, Harrison et al. [5] shows the suitability of Poisson distribution in predicting hospital admission. Data shows higher hospital admission on weekdays than weekends on average [41].

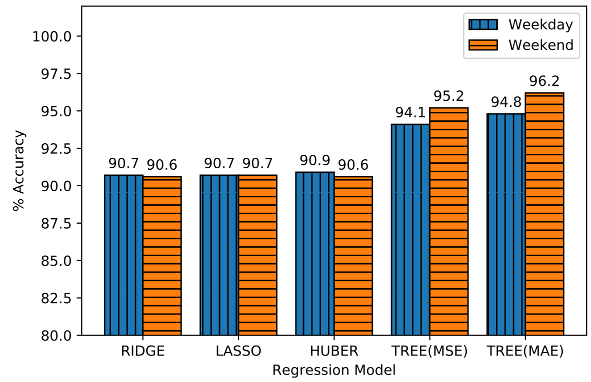

Fig. 4. Average accuracy for different regression models in predicting hospital admission for weekdays and weekends based on data from [41].

So we fit the historical hospital admission data using the features (age-partitioned population, disease burden, social vulnerability index) within separate models for weekdays and weekends. Prediction accuracy above 90% with different regression algorithms shows the appropriateness of input features. Fig. 4 shows the prediction accuracies for different regression models, where data from 2010-2016 forms the training set, and 2017 data is used as the test set. We choose the best regressor (decision tree with Mean Absolute Error) to predict the Poisson distribution mean for weekdays and weekends in each region. Based on our observation of the hospital admission data, to better account for the day-to-day variation, we add 20% Gaussian variance around the predicted value from the regression model to obtain λ_n^r. Finally, the number of beds required for a particular day in the r^{th} region is modeled as

$$j_n^r \sim \text{Poisson } (\lambda_n^r). \tag{5}$$

As the average length of stay per admission is 4.7 days throughout the US [39], we set the number of hospital bed requirements as 4.7 times the random number j_n^r generated in Eq. (5).

C. Scenarios

During the COVID-19 outbreak, the healthcare system allocated part of its capacity to deal with the pandemic-affected patients, decreasing the regular healthcare capacity. Hence, the healthcare authority needs to include pandemic scenarios in its policymaking. We define two scenarios with no pandemic and a 3-year long pandemic (between $20^{th}-22^{nd}$ year) event within the 30 year decision time horizon (year 2021-2050). During the peak period of the COVID-19 pandemic, the average hospitalization in Florida was 12250, which is around 20% of the hospital beds in Florida [23], [43], which we integrate into this case study. Specifically, in the pandemic years, 80% of beds will be available for regular healthcare, keeping the rest 20% reserved to handle the pandemic. The pandemic scenario may also cover other humanitarian crises due to natural disasters or other catastrophic events.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2022.3187950

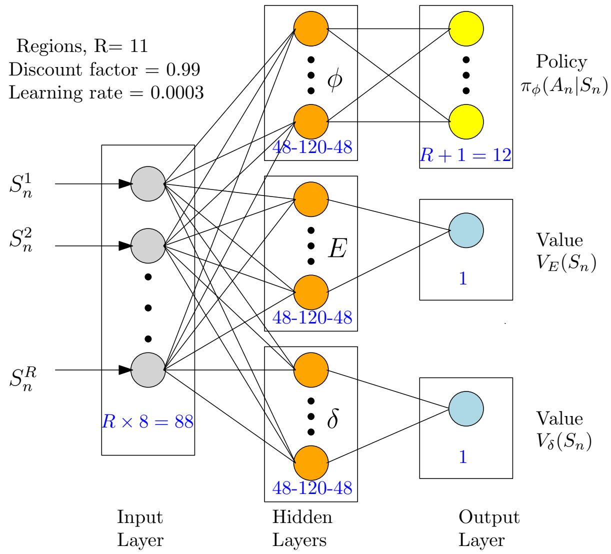

Fig. 5. Neural network architecture for the proposed multi-objective deep RL-based policy.

D. Objective Priority

In the current practice, following the certificate of need (CON) process, the healthcare authority sets a threshold on the occupancy level to make the expansion decision, which implicitly represents their priority levels for the DoS and expansion cost objectives [11]. We reflect the healthcare authority's priority levels for the two objectives in its healthcare expansion policy by explicitly considering varying weights for the actor network's advantage function:

$$(w_E, w_\delta) = \Big((0.1, 0.9), (0.2, 0.8), ..., (0.9, 0.1)\Big). \quad (6)$$

These weight pairs respectively represent a range of policies from service-centric to cost-centric.

E. Neural Network Architecture and Computation Time

We have one policy (actor) and two value (critic) networks in the A2C architecture for our MORL-based policy presented in Section IV. Fig. 5 shows the NN architecture of our method. We use a learning rate of 0.0003 and a discount factor of 0.99 for all three networks. Although the input state is the same for all 3 deep NNs, they have separate hidden layers to output the policy and value estimates. The hidden layers have 48, 120, and 48 neurons for all 3 deep NNs. The two value networks output one value for each value function estimate. However, the policy-network outputs $R + 1 = 12$ action probabilities for the given state. For each combination of weights (w_E, w_δ) in Eq. (6), both objectives converge within 3,000 episodes, i.e., the agent learns the optimal policy after 3,000 runs. Table I shows the computation time for the proposed method. It takes 5 minutes for data processing, including hospital occupancy forecasting by using an Intel® Core i7, 3.60 GHz, 16 GB RAM computer. The MORL algorithm needs 510 minutes to perform the 3,000 episodes for convergence. Notably, the computational time for each decision is 0.33 seconds; negligible compared to our approach's policy-making steps (i.e., 1 year).

Table I: Computational details for the experiments.

Hardware	Software	Task	Computation time
Intel® Core i7 3.60GHz 16 GB RAM	Python 3.7 Pytorch 1.8.1 sklearn 0.23.2	Data Preparation	5 min
		MORL Convergence	510 min
		MORL Decision	0.33 sec

VI. Results

A. Proposed Deep MORL-based Policy

Our method provides a set of trade-off solutions for the healthcare authority. Depending on the objective weight range from Eq. (6), Fig. 6(d) shows the cumulative expansion cost (left y-axis) and DoS (right y-axis) for the 30-year timeline obtained from the proposed deep MORL-based optimal policy. For the most service-centric healthcare authority ($w_E = 0.2 < w_\delta = 0.8$), the expansion cost is 990 and 1550 Million USD, respectively, for the non-pandemic and pandemic scenarios over a 30-year period. The cumulative DoS for the regions is 220 and 1187 thousand, respectively, for the two scenarios. Although pandemic occurrence makes both costs worse, the obtained DoS is a lot more acceptable than the actual situation during the COVID-19 pandemic. With more emphasis on cost minimization ($w_E > 0.2$), the DoS number goes up, and the expansion cost goes down, as shown in Fig. 6(d). For the most cost-centric policy in our setup ($w_E = 0.8, w_\delta = 0.2$), the healthcare authority makes no investment actions and endures 1470 and 3143 thousand DoS, respectively, for the two scenarios. The source codes are available at GitHub [1].

B. Benchmark Policies

We compare our deep MORL-based policy with a myopic policy from [3], a Recurrent Neural Network (RNN)-based policy from [24], and a single objective RL-based policy from [6] for a 30-year scheme. We selected decision thresholds to incur investment costs ranging from maximum to minimum for every policy to make a head-to-head comparison with our proposed MORL method.

1) Target Occupancy Level Based Policy [3]: Historically, many states regulated the number of hospital beds by the certificate of need (CON) process, under which hospitals could only expand under state review and approval. The CON process follows a target occupancy level of hospital beds as the decisive factor [3]. We select this method as a baseline policy where the region with the maximum percentile occupancy on the previous year is selected for augmentation. No augmentation action is selected if target occupancy for each region is lower than a threshold τ_{occu}. We sweep the threshold τ_{occu} over a range to represent priority over the objectives, as shown in Fig. 6(a). For lower thresholds ($\tau_{occu} = 60\%$), the expansion cost is high, but DoS is low (service-centric) and vice versa (cost-centric) for higher thresholds ($\tau_{occu} = 90\%$). The DoS is higher for the pandemic scenario (dashed lines) than no-pandemic scenario over the entire range. However, the expansion cost is higher only for $\tau_{occu} > 80\%$.

[1] https://github.com/Secure-and-Intelligent-Systems-Lab/MORL-Based-Healthcare-Expansion-Planning

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2022.3187950

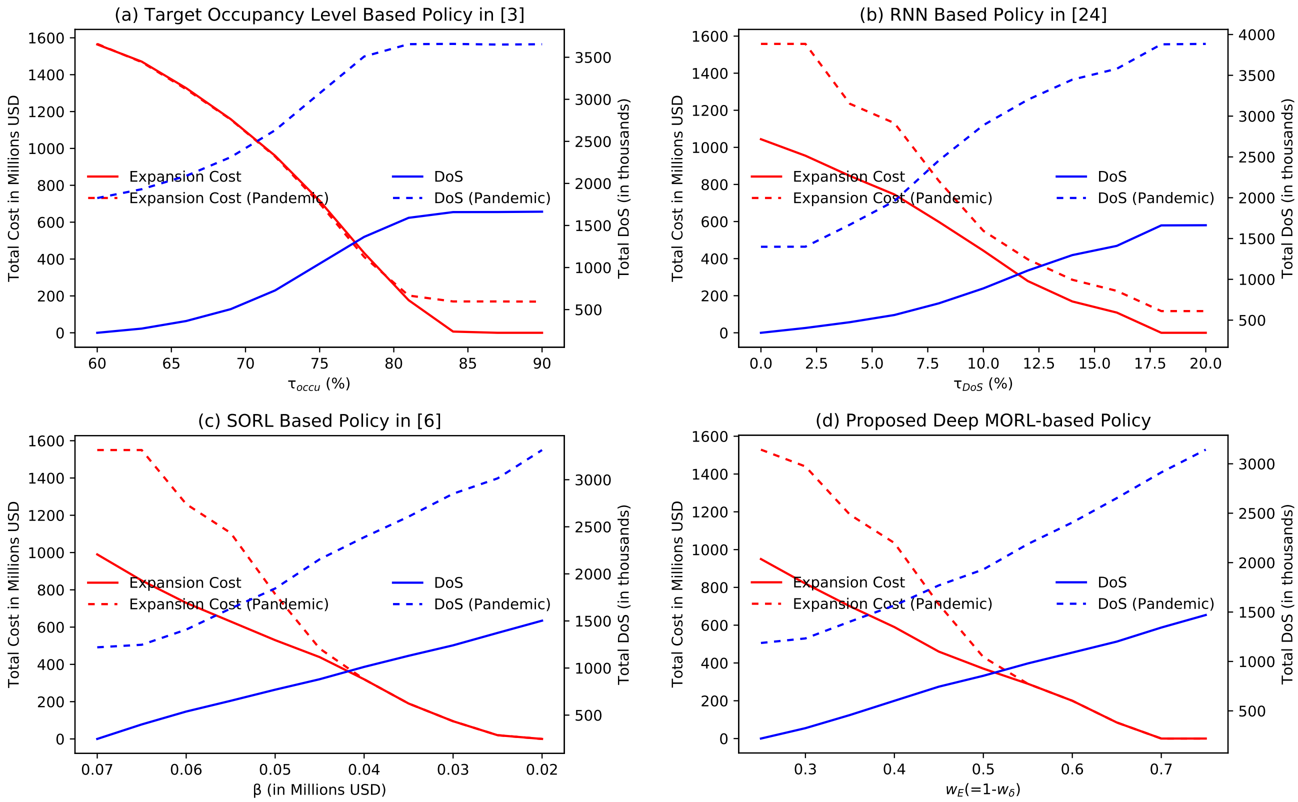

Fig. 6. Episodic (30-year) total cost and DoS for healthcare authority under different objective priorities for the policies. Objective priorities are represented by (a) occupancy threshold levels for the target occupancy level based policy in [3], (b) DoS threshold levels for the RNN based method in [24], (c) β values for each DoS for the SORL based policy in [6], and (d) objective weights w_E(=1-w_δ) for the proposed MORL based policy. Dashed lines represent the pandemic scenario.

Table II: Parameter, cost, and DoS comparison among the policies for different objective priorities (non-pandemic scenario).

Proposed Deep MORL			Target Occupancy [3]			RNN based [24]			SORL [6]		
$w_E = 1 - w_\delta$	Exp. cost	DoS (K)	τ_{occu}	Exp. cost	DoS (K)	τ_{DoS}	Exp. cost	DoS (K)	β	Exp. cost	DoS (K)
0.25	\$950 M	220	60 %	\$1564 M	224	0 %	\$1044 M	345	0.07	\$990 M	247
0.35	\$700 M	458	66 %	\$1327 M	364	4 %	\$846 M	475	0.06	\$730 M	538
0.45	\$460 M	745	72 %	\$960 M	728	8 %	\$598 M	706	0.05	\$530 M	770
0.55	\$290 M	980	78 %	\$433 M	1364	12 %	\$279 M	1108	0.04	\$320 M	1012
0.65	\$85 M	1201	84 %	\$6 M	1659	16 %	\$109 M	1410	0.03	\$95 M	1241
0.75	\$0 M	1470	90 %	\$0 M	1664	20 %	\$0 M	1663	0.02	\$0 M	1503

2) RNN Based Policy [24] : Kutafina et al. [24] provide an RNN-based hospital occupancy forecasting method. They achieved an accuracy rate of 93.76 % on eight validation sets from a German hospital's 13-year (2002-2015) hospital records data set. They included the day of the week, day of the year, public holidays, and school holidays as the features for the RNN. We include their method for the comparative analysis with the following adaptations:

- We include age-based population vector, DB, SVI, and pandemic flag on top of the features used in [24].
- We select a region for expansion that is predicted to have the most DoS based on the RNN forecast for the next step.
- We select no expansion if the DoS of the selected region is less than a threshold τ_{DoS}; otherwise, that region gets the capacity expansion.

The τ_{DoS} indicates how much percentile DoS the healthcare allows, i.e., it does not make any expansion if all of the regions' DoS is below that threshold. We do a sweep search for τ_{DoS} between 0-20% that characterizes a range between service-centric to cost-centric healthcare authority as shown in Fig. 6(b). This method has a one-step look ahead benefit compared to the Target Occupancy Level method of [3]. Hence, it incurs higher expansion cost and less increase in DoS for the pandemic scenario throughout the threshold range compared to the Target Occupancy Level method.

3) Single Objective RL (SORL) Based Policy [6]: In our preliminary work [6], we converted the DoS into monetary cost by assigning DoS cost for each region each day as in

$$c^r_{n,d} = \begin{cases} \beta(j^r_{n,d} - H^r_n), & \text{if } j^r_{n,d} - H^r_n > 0 \\ 0, & \text{otherwise,} \end{cases}$$

where we selected β = \$0.04M, which represents the monetary cost equivalent of the discomfort of an unattended patient, based on the study [44]. This cost is summed up over

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2022.3187950

all regions as the DoS cost for the time step n as

$$E_n^{DoS} = \sum_{r=1}^{R} \sum_{d=1}^{D} c_{n,d}^{r}. \tag{7}$$

The sum of the expansion cost and the DoS cost is used as the negative reward, $R_n = -(E_n + E_n^{DoS})$ for the single objective RL approach in [6]. This method does not provide a handle over preference between the two objectives. Furthermore, the monetary equivalence for DoS is an abstract idea, and setting a universal value for β is impossible. In fact, this value can represent the mindset of the healthcare authority about how much it cares about the population. So, we use a range of values $\beta = (0.07, 0.065, ..., 0.02)$ that represents from service-centric to cost-centric policies with the decreasing value of β as seen in Fig. 6(c). The pandemic scenario incurs higher cost and DoS; however, the expansion actions are similar when the agent puts less value on $\beta < 0.045$.

C. Comparative Analysis

We conduct a comparative analysis among the policies mentioned above in terms of the total expansion cost and total DoS for the 30-year timeline (Fig. 7). The x-axis represents the objective priority that we generalized as cost-centric, moderate, and service-centric (from left o right) based on the decision process range discussed and shown in Fig. 6. In particular, Fig. 7 puts all the policies' outcomes in a single frame to better understand the benefit of our proposed MORL-based policy. The target occupancy level based policy in [3] performs worst among the policies as it is always one step behind, i.e., its decision is based on the previous year's experience. The RNN forecasting based policy in [24] performs better mainly due to the inclusion of the observable state's data in the forecasting method. However, this policy in [24] lacks the RL mechanism to minimize the cumulative costs from all future states. The single objective RL-based policy in [6] and our proposed MORL perform better than the other two policies. However, our MORL based policy outperforms the SORL in [6] by utilizing more data (information) in its state definition. Especially with the PPI data, our method picks a less expensive region for expansion when there is a tie between two regions of different PPI levels. This better performance of MORL is more emphasized in the pandemic scenario, as shown in Fig. 7.

Fig. 8 provides a synonymous view of the Pareto front of trade-off solutions discussed in [7] for the initial state (i.e., year 2021). The x-axis represents expansion cost, and the y-axis represents the number of DoS achieved from the initial state (i.e., $n = 0$) for all the policies considering equal (moderate) priority between the two objectives. The ideal Pareto front would be a curve that no other policy can go under, i.e., no other points can decrease expansion cost without increasing DoS and vice versa. Many works [7]–[9] focus on obtaining the ideal Pareto front for multi-objective optimization tasks, yet we can only approximate for our high dimensional state and action space problem. In Fig. 8, for the deep MORL based policy, each point on the curve is a non-dominated solution among the compared policies, which means none of the competing policies achieves better optimization for the corresponding objective priority level. Furthermore, our method provides an easy-to-use and natural way to control the trade-off between the objectives through setting simple weights between zero and one (w_E and w_δ), whereas the SORL policy requires further studies to strike a desired balance between the objectives.

The comparative analysis is summarized in Table III for a healthcare authority that puts equal (moderate) priority for both objectives. Our MORL based policy provides the lowest cost and DoS among the other policies. To compare the policies in more detail, their selected actions for the 30-year period are also shown in Table III. The different actions in the pandemic scenario are given in parentheses. The target occupancy level based policy in [3] takes the same actions under both scenarios. The RNN based policy in [24] takes similar actions like in [3]; however, with its prediction capability it takes those actions early to prevent higher costs. SORL in [6] and the proposed MORL policy perform better as they are even more proactive in making preventive expansions. Both policies expand in the early years to keep the system in a balanced state in the future. Region 5 gets the most expansions, suggesting this region expects higher DoS in the future. However, the proposed MORL also selects Region 8 for expansion in the early years, which is one of the significant differences from the other policies. Since it is a larger region with a high population, as a result of that expansion, the overall DoS goes down significantly. The expansion decisions of the proposed MORL policy are also shown in Fig. 3.

VII. Discussions

The key insights from the conducted study are

- The historical patient data based methods (e.g., target occupancy policy [3]) focus on the trend and lacks the root cause analysis (RL states) for future patient estimation.
- The RNN based policy [24] predicts the future hospital needs well; however, it lacks prescriptive analysis. It requires a dynamic DoS threshold for making decisions to adapt to different situations.
- Deep RL-based prescriptive analysis is suitable for the task as evident in experimental results for SORL [6] and our deep MORL method. As the SORL policy converts the DoS into a monetary cost with the help of a coefficient (β here), it requires literature to support spatial and temporal generalization for a suitable value of β. Our deep MORL approach addresses this issue and provides an easy handle to the authority to set the relative weights for the two objectives.

While significantly improving the state-of-the-art, the proposed method also has certain limitations. For instance, it does not consider selecting multiple regions concurrently for expansion. Also, our method does not provide a mechanism for selecting different capacity expansion sizes for different regions in this current form. This research can be further

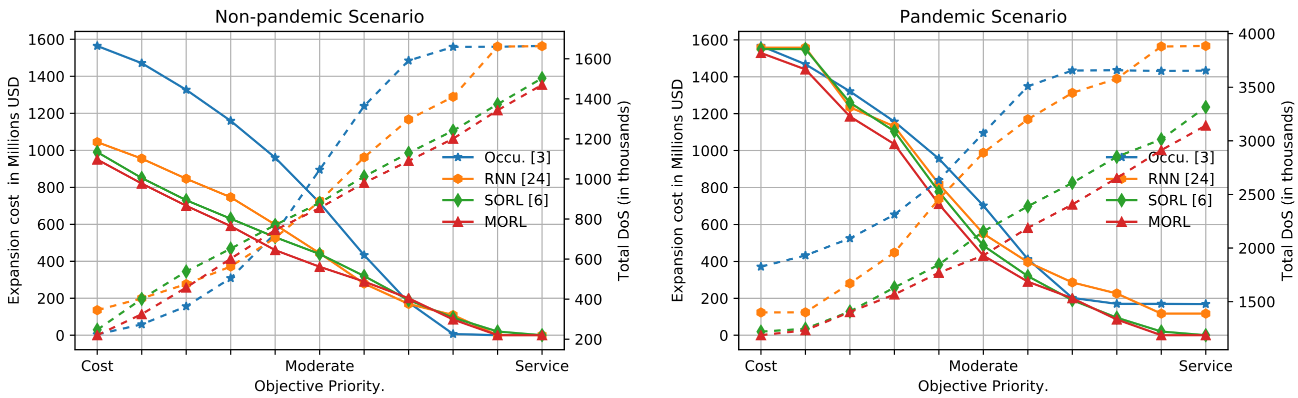

Fig. 7. Episodic (30-year) total cost (solid lines) and DoS (dashed lines) for healthcare authority under different objective priorities for the policies for non-pandemic (left) and pandemic (right) scenarios.

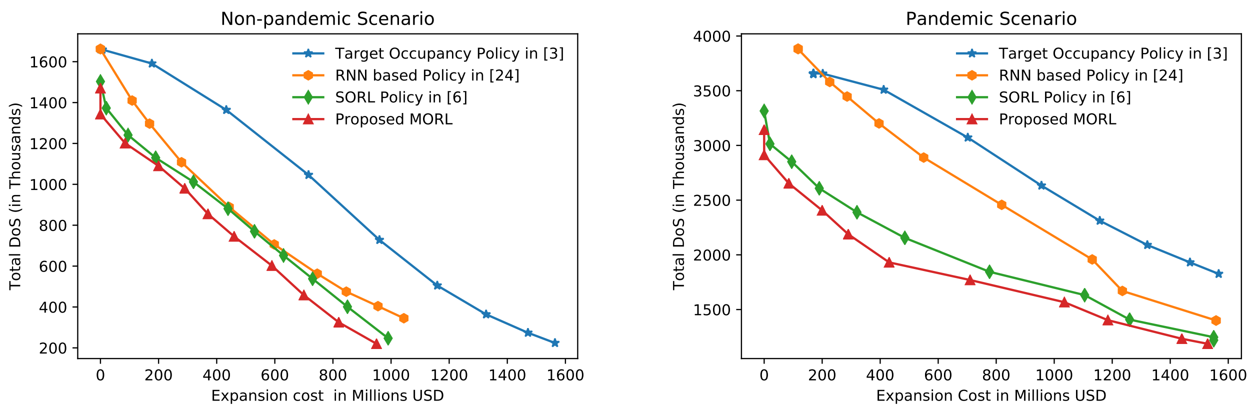

Fig. 8. Trade-off solutions for each policy for the non-pandemic (left) and pandemic (right) scenarios with equal (moderate) objective priority. The points in the curves refer to episodic (30-year) total cost and DoS for the healthcare authority at the initial state (i.e., $n = 0$).

Table III: Expansion cost, DoS, and selected sequential actions for each policy over a 30-year period for equal priority on the two objectives. Different actions in the pandemic scenario are shown in the parentheses.

	Non-pandemic		Pandemic		
Policy	Exp. cost	DoS (K)	Exp. cost	DoS (K)	Selected Sequential Actions (in year 1 to 30)
Target Occupancy [3]	\$716 M	1046	\$716 M	3072	0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 5 , 0, 5, 5, 7, 5, 7, 3, 2, 5, 2, 7, 5, 3
RNN Based [24]	\$443 M	889	\$550 M	2889	0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 5, 0 , 7, 5, 5, 0 (7), 0 (6), 7, 3, 2, 5, 2, 0, 0, 0
SORL Based [6]	\$439 M	881	\$485 M	2155	0, 0, 0, 0, 0, 0, 0, 5, 0, 0, 0, 0, 7, 0, 0, 3, 5 , 0, 2, 5, 7, 5, 0 (7), 3, 0, 0, 0, 0, 0, 0
Proposed MORL	\$370 M	855	\$425 M	1931	0, 0, 0, 0, 5, 0, 0, 8, 0, 0, 0, 0, 7, 0, 0, 3, 5 , 0, 2, 0 (5), 7, 0, 0, 3, 0, 0, 0, 0, 0, 0

utilized for allocating human resources such as physicians and healthcare personnel for a region. Private organizations often provide significant healthcare, hence including them in policymaking can provide the basis for a multi-agent RL setup. In that context, the reward function for the private organizations may include their financial benefit, and the central RL agent (healthcare authority) may consider the private facilities as buffer capacity to accommodate emergencies. Addressing these limitations and new scopes can provide several future research directions.

VIII. Conclusions

We proposed a multi-objective reinforcement learning (MORL) framework to develop a healthcare expansion plan and demonstrated its efficacy in a case study for Florida. The MORL method enables the user to conveniently set different weights for its two objectives, minimization of expansion cost and number of Denial of Service (DoS), in a natural way by only setting their priority percentages. Our data-driven approach is suitable for coping with the dynamic behavior of the region's healthcare needs over a long period, especially to deal with emergency scenarios like pandemic events. We significantly improved our preliminary work in [6], which follows a single objective RL approach through converting DoS into monetary cost, by developing a multi-objective solution to enable intuitive objective priority setting; including Disease Burden (DB) and Social Vulnerability Index (SVI), apart from the age-partitioned population, for hospital occupancy prediction and making expansion decisions; utilizing Price Parity Index (PPI) to accommodate different expansion costs for different regions.

References

[1] Worldometer, "United states demographic," Accessed: 2021-09-19. [Online]. Available: https://www.worldometers.info/demographics/us-demographics/#median-age

[2] D. DeLia, "Annual bed statistics give a misleading picture of hospital surge capacity," *Annals of Emergency Medicine*, vol. 48, no. 4, pp. 384–388.e2, 2006.

[3] L. V. Green, "How many hospital beds?" *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, vol. 39, no. 4, pp. 400–412, 2002.

[4] L. Seematter-Bagnoud, S. Fustinoni, D. Dung, B. Santos-Eggimann, V. Koehn, R. Bize, A. Oettli, and J.-B. Wasserfallen, "Comparison of different methods to forecast hospital bed needs," *European Geriatric Medicine*, vol. 6, no. 3, pp. 262–266, 2015.

[5] G. W. Harrison, A. Shafer, and M. Mackay, "Modelling variability in hospital bed occupancy," *Health Care Management Science*, vol. 8, no. 4, pp. 325–334, 2005.

[6] S. S. Shuvo, M. R. Ahmed, H. Symum, and Y. Yilmaz, "Deep reinforcement learning based cost-benefit analysis for hospital capacity planning," in *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2021, pp. 1–7.

[7] K. Van Moffaert and A. Nowé, "Multi-objective reinforcement learning using sets of pareto dominating policies," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 3483–3512, 2014.

[8] M. Reymond and A. Nowé, "Pareto-dqn: Approximating the pareto front in complex multi-objective decision problems," in *Proceedings of the adaptive and learning agents workshop (ALA-19) at AAMAS*, 2019.

[9] T. Wang, Y. Luo, J. Liu, and K. Li, "Multi-objective end-to-end self-driving based on pareto-optimal actor-critic approach," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2021, pp. 473–478.

[10] P.-H. Su, P. Budzianowski, S. Ultes, M. Gasic, and S. Young, "Sample-efficient actor-critic reinforcement learning with supervised data for dialogue management," *arXiv preprint arXiv:1707.00130*, 2017.

[11] H. Ravaghi, S. Alidoost, R. Mannion, and V. D. Bélorgeot, "Models and methods for determining the optimal number of beds in hospitals and regions: A systematic scoping review," *BMC Health Services Research*, vol. 20, no. 1, p. 186, 2020.

[12] L. Seematter-Bagnoud, S. Fustinoni, D. H. Dung, B. Santos-Eggimann, V. Koehn, R. Bize, A. Oettli, and J. B. Wasserfallen, "Comparison of different methods to forecast hospital bed needs," *European Geriatric Medicine*, vol. 6, no. 3, pp. 262–266, jun 2015.

[13] J. A. Acuna, J. L. Zayas-Castro, and H. Charkhgard, "Ambulance allocation optimization model for the overcrowding problem in us emergency departments: A case study in florida," *Socio-Economic Planning Sciences*, vol. 71, p. 100747, 2020.

[14] S. Baas, S. Dijkstra, A. Braaksma, P. van Rooij, F. J. Snijders, L. Tiemessen, and R. J. Boucherie, "Real-time forecasting of covid-19 bed occupancy in wards and intensive care units," *Health care management science*, pp. 1–18, 2021.

[15] B. G. Barnes and N. L. Harp, "The U.S. Medicare Disproportionate Share Hospital program and capacity planning," *Journal of Accounting and Public Policy*, vol. 37, no. 4, pp. 335–351, 2018.

[16] P. Landa, M. Sonnessa, E. Tànfani, and A. Testi, "Multiobjective bed management considering emergency and elective patient flows," *International Transactions in Operational Research*, vol. 25, no. 1, pp. 91–110, 2018.

[17] N. Proudlove, K. Gordon, and R. Boaden, "Can good bed management solve the overcrowding in accident and emergency departments?" *Emergency Medicine Journal*, vol. 20, no. 2, pp. 149–155, 2003.

[18] Z. Zhu, B. H. Hen, and K. L. Teow, "Estimating ICU bed capacity using discrete event simulation," *International Journal of Health Care Quality Assurance*, vol. 25, no. 2, pp. 134–144, 2012.

[19] R. B. Tate, L. MacWilliam, and G. S. Finlayson, "A Methodology for Estimating Hospital Bed Need in Manitoba in 2020," *Canadian Journal on Aging / La Revue canadienne du vieillissement*, vol. 24, no. S1, pp. 141–151, 2005.

[20] G. Ma and E. Demeulemeester, "A multilevel integrative approach to hospital case mix and capacity planning," *Computers & Operations Research*, vol. 40, no. 9, pp. 2198–2207, 2013.

[21] R. P. Jones, "Does the ageing population correctly predict the need for medical beds? part two: wider implications," *British Journal of Healthcare Management*, vol. 27, no. 10, pp. 1–9, 2021.

[22] A. Kumar, R. J. Jiao, and S. J. Shim, "Predicting bed requirement for a hospital using regression models," in *IEEE International Conference on Industrial Engineering and Engineering Management*, 2008, pp. 665–669.

[23] J. Heins, J. Schoenfelder, S. Heider, A. R. Heller, and J. O. Brunner, "A scalable forecasting framework to predict covid-19 hospital bed occupancy," *INFORMS Journal on Applied Analytics*, 2022.

[24] E. Kutafina, I. Bechtold, K. Kabino, and S. M. Jonas, "Recursive neural networks in hospital bed occupancy forecasting," *BMC medical informatics and decision making*, vol. 19, no. 1, p. 39, 2019.

[25] J. Schiele, T. Koperna, and J. O. Brunner, "Predicting intensive care unit bed occupancy for integrated operating room scheduling via neural networks," *Naval Research Logistics (NRL)*, vol. 68, no. 1, pp. 65–88, feb 2021.

[26] W. Qiang and Z. Zhongli, "Reinforcement learning model, algorithms and its application," in *2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC)*. IEEE, 2011, pp. 1143–1146.

[27] S. S. Shuvo, Y. Yilmaz, A. Bush, and M. Hafen, "A markov decision process model for socio-economic systems impacted by climate change," in *International Conference on Machine Learning*. PMLR, 2020.

[28] A. Haydari and Y. Yilmaz, "Deep reinforcement learning for intelligent transportation systems: A survey," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[29] C. Yu, J. Liu, S. Nemati, and G. Yin, "Reinforcement learning in healthcare: A survey," *ACM Computing Surveys (CSUR)*, vol. 55, no. 1, pp. 1–36, 2021.

[30] O. Gottesman, F. Johansson, M. Komorowski, A. Faisal, D. Sontag, F. Doshi-Velez, and L. A. Celi, "Guidelines for reinforcement learning in healthcare," *Nature Medicine*, vol. 25, no. 1, pp. 16–18, jan 2019.

[31] S. S. Shuvo, M. R. Ahmed, S. B. Kabir, and S. A. Shetu, "Application of Machine Learning based hospital up-gradation policy for Bangladesh," in *ACM International Conference Proceeding Series*. New York, NY, USA: Association for Computing Machinery, dec 2020, pp. 18–24.

[32] S. S. Shuvo, H. U. Ahmed *et al.*, "Use of machine learning for long term planning and cost minimization in healthcare management," *medRxiv*, 2021.

[33] N. S. Kalman, B. G. Hammill, K. A. Schulman, and B. R. Shah, "Hospital overhead costs: The neglected driver of health care spending?" *Journal of Health Care Finance*, vol. 41, no. 4, 2015.

[34] Fixr, "Cost to Build a Hospital — Hospital Construction Cost." [Online]. Available: https://www.fixr.com/costs/build-hospital

[35] A. E. Cuellar and P. J. Gertler, "How the expansion of hospital systems has affected consumers," *Health affairs*, vol. 24, no. 1, pp. 213–219, 2005.

[36] M. I. Fahmi, M. Zarlisb, and S. E. Tulus, "An optimization model for solving integrated hospital capacity planning problem," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 6, pp. 4816–4824, 2021.

[37] B. E. Flanagan, E. W. Gregory, E. J. Hallisey, J. L. Heitgerd, and B. Lewis, "A social vulnerability index for disaster management," *Journal of homeland security and emergency management*, vol. 8, no. 1, 2011.

[38] WorldBank, "Hospital beds per 1000 people," Accessed: 2021-09-26. [Online]. Available: https://data.worldbank.org/indicator/SH.MED.BEDS.ZS

[39] A. H. Association *et al.*, "Trendwatch chartbook 2018: trends affecting hospitals and health systems, 2018," 2019.

[40] S. Rayer and Y. Wang, "Projections of florida population by county, 2020–2045, with estimates for 2019," *Bureau of Economics and Business Research, Florida Population Stuides, Bulletin 186*, 2020.

[41] "HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2010-2017." [Online]. Available: www.hcup-us.ahrq.gov/sidoverview.jsp

[42] "Real personal income and regional price parities," Accessed: 2021-09-29. [Online]. Available: https://www.bea.gov/system/files/methodologies/RPP2020-methodology_1.pdf

[43] "Coronavirous resource center: Florida," *Johns Hopkins University & Medicine*, Accessed: 2022-01-20. [Online]. Available: https://coronavirus.jhu.edu/region/us/florida

[44] D. Farrell and F. E. Greig, "Paying out-of-pocket: The healthcare spending of 2 million us families," *SSRN Electronic Journal*, 2017.