Enhancing Medical Training Through Learning From Mistakes by Interacting With an Ill-Trained Reinforcement Learning Agent

Yasar C. Kakdas, Sinan Kockara, Tansel Halic, and Doga Demirel, *Member, IEEE*

***Abstract*—This article presents a 3-D medical simulation that employs reinforcement learning (RL) and interactive RL (IRL) to teach and assess the procedure of donning and doffing personal protective equipment (PPE). The simulation is motivated by the need for effective, safe, and remote training techniques in medicine, particularly in light of the COVID-19 pandemic. The simulation has two modes: a tutorial mode and an assessment mode. In the tutorial mode, a computer-based, ill-trained RL agent utilizes RL to learn the correct sequence of donning the PPE by trial and error. This allows students to experience many outlier cases they might not encounter in an in-class educational model. In the assessment mode, an IRL-based method is used to evaluate how effective the participant is at correcting the mistakes performed by the RL agent. Each time the RL agent interacts with the environment and performs an action, the participants provide positive or negative feedback regarding the action taken. Following the assessment, participants receive a score based on the accuracy of their feedback and the time taken for the RL agent to learn the correct sequence. An experiment was conducted using two groups, each consisting of ten participants. The first group received RL-assisted training for donning PPE, followed by an IRL-based assessment. Meanwhile, the second group observed a video featuring the RL agent demonstrating only the correct donning order without outlier cases, replicating traditional training, before undergoing the same assessment as the first group. Results showed that RL-assisted training with many outlier cases was more effective than traditional training with only regular cases. Moreover, combining RL with IRL significantly enhanced the participants' performance. Notably, 90% of the participants finished the assessment with perfect scores within three iterations. In contrast, only 10% of those who did not engage in RL-assisted training finished the assessment with a perfect score, highlighting the substantial impact of RL and IRL integration on participants' overall achievement.**

***Index Terms*—Agent-based, donning and doffing, interactive reinforcement learning (IRL), medical training, reinforcement learning (RL).**

I. Introduction

In recent years, the integration of 3-D simulations, video games, and virtual reality based applications into educational frameworks has enhanced learning paradigms, assisting students in improving their skills, knowledge, and motivation to learn [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16]. The 3-D simulations are computer-generated representations that mimic the real world by creating virtual environments, offering visually immersive experiences. An approach known as simulation-based learning leverages these 3-D simulations as educational tools, enabling individuals to experiment, immerse, make decisions, and learn from the consequences in a risk-free setting [17], [18], [19], [20].

Especially in safety-focused industries, such as aviation and the military, the widespread acceptance of 3-D simulations and simulation-based education has been notable for decades [18], [21], [22]. Furthermore, with rapid technological advancements, 3-D simulations have gained significant traction in medical training by offering immersive and safe learning experiences [21], [23], [24], [25], [26], [27], [28], [29], [30].

Traditional medical training faces significant challenges, including high costs associated with creating a physical training environment, purchasing equipment, the extensive time required for preparations, limitations in demonstrating all possible scenarios, and ensuring the safety of both patients and healthcare personnel [21], [31]. The global spread of COVID-19 has further necessitated a shift to remote platforms, posing additional challenges. However, it remains crucial for medical trainees to gain hands-on experience and visualize medical procedures before performing them on actual patients. Considering these challenges, 3-D immersive training experiences have become more critical than ever [32], [33], [34].

This article introduces an innovative approach to overcome the constraints associated with conventional training methodologies. It presents a solution leveraging advanced 3-D simulation tools for comprehensive learning and instruction. The focal point of our work involves the design and implementation of an agent-based learning approach, integrating principles from

Manuscript received 31 August 2023; revised 19 January 2024 and 21 February 2024; accepted 28 February 2024. Date of publication 4 March 2024; date of current version 21 March 2024. This work was supported in part by the National Institutes of Health (NIH) Awards NIBIB R01EB005807, NIAMS R44AR075481, and NHLBI 1R01EB025241, and in part by Arkansas INBRE Program NIH/NIGMS P20 GM103429. *(Corresponding author: Doga Demirel.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by IRB Florida Polytechnic University under Application No. 23-007.

Yasar C. Kakdas and Doga Demirel are with the Department of Computer Science, Florida Polytechnic University, Lakeland, FL 33805 USA (e-mail: ykakdas@floridapoly.edu; ddemirel@floridapoly.edu).

Sinan Kockara is with the Department of Computer Science, Rice University, Houston, TX 77005 USA (e-mail: skockara@rice.edu).

Tansel Halic is with Intuitive Surgical, Peachtree Corners, GA 30092 USA (e-mail: tansel.halic@intusurg.com).

This article has supplementary downloadable material available at https://doi.org/10.1109/TLT.2024.3372508, provided by the authors.

Digital Object Identifier 10.1109/TLT.2024.3372508

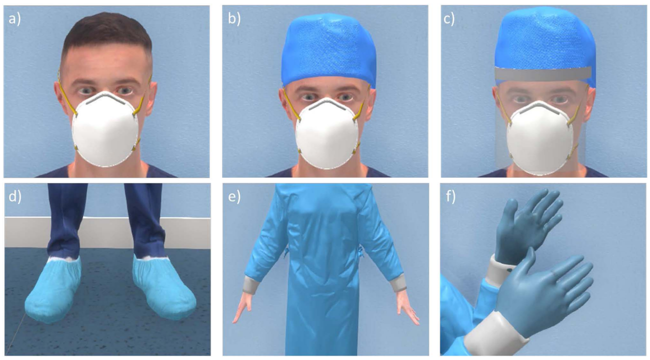


Fig. 1. Donning order of the PPE. (a) N95 respirator. (b) Surgical scrub cap. (c) Visor. (d) Shoe covers. (e) Surgical gown. (f) Surgical gloves.

both classical reinforcement learning (RL) [35] and interactive RL (IRL) [36]. The application of this approach is demonstrated in the simulation of healthcare procedures, specifically the process of donning and doffing personal protective equipment (PPE) [37]. The standards and procedures for using PPE have become crucial, particularly with the rapid spread of the COVID-19 pandemic [38] to safeguard the health of healthcare workers and patients. The donning and doffing procedures describe the correct order of equipping PPE and its proper usage to provide maximum safety [39]. The PPE items and donning procedure utilized in our simulator are illustrated in Fig. 1.

The primary objective of this study is to create an engaging 3-D simulation that accurately teaches medical trainees the donning procedure while evaluating their performance. The simulation includes two distinct modes: tutorial and assessment. In the RL-assisted tutorial mode, a computer-based, ill-trained RL agent attempts to learn the correct sequence of donning PPE through trial and error. The agent starts with initial knowledge comprising random and misleading information.

Trainees observe the RL agent's learning trajectory, wherein it undergoes a series of iterations to acquire the correct sequence of PPE. The RL agent's learning journey includes the instances of errors and mistakes, showcasing outlier scenarios and false sequences. Therefore, we named it an ill-trained RL agent. This nuanced approach introduces a dimension not typically encountered in traditional PPE learning procedures, providing a richer and more realistic training experience. In the assessment mode of our 3-D simulation, instead of the RL agent correcting its mistakes by itself, trainees provide external feedback and intervene in the agent's actions using IRL. At the end of the assessment, trainees receive a score based on the correctness of interventions and the time taken to teach the RL agent the correct sequence. Both modes include a feedback user interface (UI) that alerts trainees when the RL agent makes a mistake during training or when the trainee provides incorrect guidance during the assessment phase. Using this feedback feature, we aim to replace the traditional training methods with learning from the mistakes made by ill-trained RL agents.

II. Literature Review

Traditional training for donning and doffing PPE in an operating room typically involves theoretical instruction in a classroom setting or video lessons. Christensen et al. [40] suggest that there was no significant difference in donning and doffing scores between instructor-led and video lessons for training in PPE donning and doffing. This finding implies that video training could be a fast and resource-efficient method of training, particularly in responding to the COVID-19 pandemic, where there may be a need for rapid dissemination of training to a large number of healthcare professionals. However, it is crucial to acknowledge that, although conventional training methods may be effective for specific aspects of PPE training, real-time interaction and the opportunity to learn from mistakes remain essential for ensuring proficiency in donning and doffing procedures.

In recent years, there have been numerous studies that incorporate artificial intelligence (AI) and machine learning (ML) into healthcare applications [41], [42], [43], [44], [45], [46], [47]. A comprehensive literature review [48] delved into the recent healthcare studies utilizing a broad range of ML techniques, such as RL, artificial neural networks, and natural language processing in surgeries. The study concludes that AI plays a vital role in surgeries and assisting surgeons. Another study [49] explores the use of AI in nanomedicines, especially in oncology, and finds promise in cancer management through AI-assisted nanotechnology.

Despite the widely recognized benefits of AI and ML, several issues persist. Predominant among these are data privacy concerns, ethical considerations, and limitations in datasets, such as improper labeling or bias. Multiple studies examining AI and ML applications in healthcare [50], [51], [52], [53], [54], [55], [56], [57] agree on the advantages of AI in clinical practice while highlighting challenges in current applications.

RL has been instrumental in medical simulations to address diverse challenges [58], including determining appropriate treatment and medication for patients [59]. Each patient may require different treatment strategies based on their medical history, making it a critical decision-making process. Dynamic treatment regime (DTR) establishes rules for providing appropriate and optimal treatments. Numerous DTR systems [60], [61], [62], [63], [64], [65] have adopted an RL-based approach, providing valuable insights into conditions, such as epilepsy [66], lung cancer [67], and sepsis [68]. These studies report promising results in predicting treatments for various illnesses. However, as the RL-based DTR approach relies on training pre-existing clinical data, it necessitates dependable data sources and preprocessing operations.

AI and deep reinforcement learning (DRL) are frequently employed in medical image analysis [69], [70], [71]. Medical image analysis is crucial in terms of medical diagnoses and prognoses. In one study [72], DRL facilitated the automated detection of breast lesions from the dynamic contrast-enhanced magnetic resonance volumes, reducing inference time. Similarly, another study [73] used RL to localize brain tumors on 2-D contrast-enhanced MRI brain images, demonstrating efficiency even with small training sets. In addition, Chen et al. [74] used RL to diagnose and predict COVID-19 cases from computer tomography images, showcasing effective and robust detection and prediction.

Beyond medical decision-making and image analysis, RL is applied to robot-assisted surgeries [75], [76], [77], [78]. For instance, Chiu et al. [79] focused on using RL for grasping suture needles during robot-assisted minimally invasive surgeries, significantly reducing time compared with other methods. This RL approach is model free and adaptable to various operative robots. RL is also integrated into surgical training, such as robot-assisted training in laparoscopy [80], where the combination of experts' experiences and RL agent's training outcomes is demonstrated on a robot-assisted device. However, the lack of communication between the RL agent and the trainee restricts to shape the agent's behavior during training and limits visibility into potential trainee decision outcomes.

Apart from RL, several studies incorporate human intervention in the RL agent's learning process to enhance solutions to diverse problems. In a study [81], researchers employed human feedback to fine-tune and align language models. According to the resulting model, using RL with human feedback improved the language model's accuracy. Shuvo and Yilmaz [82] proposed using AI-controlled smart home appliances to optimize electricity use through the home energy recommendation system, integrating DRL with human feedback and activity data to optimize costs.

Moreover, AI has found extensive applications in education, enhancing learning outcomes and catering to students' needs without direct human intervention [83], [84], [85], [86]. The intelligent tutoring system (ITS) framework offers prompt, personalized feedback to learners, with numerous AI and RL-based models [87], [88]. One of the proposed ITS models was designed to provide a one-on-one learning approach tailored to the needs of autistic students [89]. Another study aimed to enhance e-teaching and e-learning using agent-based ITS [90]. RL-based ITS models effectively improve the learning rate by providing customized training strategies based on learners' prior behaviors.

III. Method

A. Three-Dimensional Simulation and Virtual Environment

Our designed 3-D simulation begins within a dashboard where trainees can choose between two game modes: an RL-assisted tutorial or an IRL-based assessment. The simulation unfolds within the same virtual environment regardless of the selected mode. This virtual environment, as depicted in Fig. 2, comprises a room featuring PPE arranged on a table, a cabinet containing new and clean PPE, and a rigged humanoid avatar functioning as an RL agent. The implementation of avatar walking animations and interactions with objects is facilitated by inverse kinematics (IK) [91]. IK calculates positions and rotations for each bone, configuring joints to replicate realistic humanoid movements. In addition, predefined hand poses have been designed based on the bone structure of the avatar's hand model, examples of which can be seen in Fig. 3.

Fig. 2. Virtual environment: A room, a 3-D humanoid avatar, PPE on the table, and a cabinet for clean PPE.

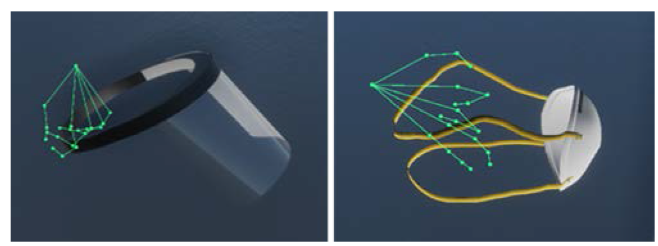


Fig. 3. Hand poses for visor and N95 face mask.

The environment's creation utilized the Markov decision process (MDP). The MDP state is updated each time the RL agent interacts with the environment and performs an action. A reward is obtained from a pre-established rewards table (applicable in RL-assisted tutorial mode) or user feedback (utilized in IRL-based assessment mode) to determine the agent's subsequent move. The Q-learning algorithm then processes the acquired reward, and the Q-value table is updated accordingly. Subsequently, the simulator signals the RL agent for the next move or terminates the simulation upon meeting specified conditions. In the case of assessment mode, once the simulation concludes, participants receive their scores.

B. Markov Decision Process

MDP [92], [93] is a mathematical framework used for solving problems dependent on the decision-making process. Fig. 4 depicts the state diagram representing the MDP utilized within the simulation.

An MDP is a four-tuple model $(S,\ A,\ T,\ R)$ defined as follows.

1) $S = \{s_1, \ldots, s_n\}$ is a finite set of states existing within the environment. These states include clean–dirty pairs for each PPE. Initially, the environment is comprised solely of clean PPE items. As the RL agent interacts with

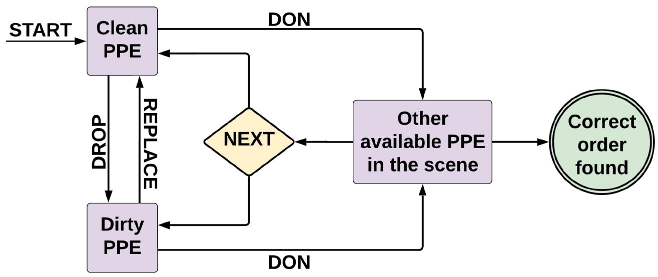


Fig. 4. State diagram of the MDP.

the environment and takes action, states can transition between clean and dirty for each PPE based on the action performed.

2) $A = \{\text{DON, DROP, REPLACE}\}$ is a predetermined set of actions available to the agent in each state.
3) $T: S \times A \to S$ is a transition function. Transition function T defines how actions performed within a state lead the RL agent to another state. For instance, when the agent executes the action DON within a state from the state space S, the agent transitions to another PPE state in S. Similarly, the DROP action will move the agent from a clean PPE state to a dirty PPE state.
4) $R: S \times A \to R$ is an immediate reward function operating with the domain combinations of $S \times A$ and outputs rewards based on the performed action within a given state-action pair.

C. RL and IRL

RL, a training method rewarding desired behaviors to form sequential decisions, is employed for the training aspect of the simulation. In RL, the RL agent engages with its environment, makes decisions, and learns through trial and error.

One benefit of RL is its ability to operate without reliance on pre-existing data for teaching the RL agent. Instead, the data are accumulated progressively as the agent works toward understanding the goal. In addition, IRL is incorporated into the assessment portion of the simulation. In the assessment mode, trainees attempt to instruct ill-trained RL agents on the correct sequence for donning PPE through their interactions. The primary difference between IRL and RL lies in the reward mechanism. In IRL, an external trainer provides feedback to the RL agent during its action selection phase, a technique also recognized as policy shaping [94]. External feedback resources aim to shape the agent's policy, guiding it along the correct path based on the provided feedback.

Both RL and IRL involve the RL agent taking actions based on a policy, denoted as π. The *Q*-function [95] serves as an action-value function for a policy π, defined as $Q^\pi(s,\ a)$, where a represents an action taken in the state s under the policy π. The solution to an RL problem involves finding the optimal policy, denoted as $\pi*$. An optimal policy indicates that the agent can determine the correct path by selecting the best actions for each state, thereby maximizing the accumulated rewards.

Mathematically, an optimal policy aims to achieve the highest expected cumulative reward over time, as shown in the following equation:

$$Q^*(s,\ a) = \max_\pi Q^\pi\ (s,\ a). \tag{1}$$

Furthermore, the optimal action-value function Q^* satisfies the Bellman optimality equation

$$Q^*(s,\ a) = E\left[\, R(s,\ a,\ s') + \gamma \max_{a'} q^*(s', a')\right]. \tag{2}$$

This equation calculates the expected value for any action-state pairs $(s,\ a)$ for optimal policy.

The following is a detailed analysis of the components comprising (2).

1) $R(s,\ a,\ s')$ represents the reward the agent receives for performing an action a in the state s and transitioning to state s'.
2) $\gamma\ \max_{a'} q^*(s', a')$ signifies the maximum expected discounted return. After an action a is performed in state s, transitioning the agent to state s', this calculates the expected value of all potential outcomes originating from state s' through any action a'. The discount factor γ dictates the weight given to future rewards relative to present ones, where $\gamma \in [0, 1]$.

D. Q-Learning

Q-learning [95] stands as one of the algorithms reliant on action-value functions (1) and the Bellman (2). This algorithm aims to determine the optimal policy for a given RL problem. Each action a performed in state s corresponds to a value denoted as $Q(s,\ a)$. These *Q*-values undergo updates using the following equation:

$$Q(s,a) \leftarrow Q(s,a) + \alpha\left[r + \gamma \max_{a'} Q(s',a') - Q(s,a)\right]. \tag{3}$$

Here, α represents the learning rate, determining the weight given to new information, while γ signifies the discount factor, determining the influence of future rewards, s denotes the current state, a represents the action taken in state s, s' is the resulting state from s when a is performed, and a' indicates any action that can be taken in state s'.

Simplifying (3) reveals that the new *Q*-value relies on three factors.

1) $(1 - \alpha)\, Q(s,a)$ represents the current *Q*-value weighted by the learning rate. When the learning rate is closer to 1, the agent primarily focuses on the new values and ignores the previous ones.
2) αr denotes the reward weighted by the learning rate. Agent receives this reward when the action a is performed in the state s. In training mode, rewards come from predefined tables, while in IRL-based assessment mode, an external trainer provides them.
3) $\alpha\gamma \max_{a'} Q(s', a')$ is the maximum *Q*-value achievable from the transitioned state s'. This value is influenced by both the learning rate and discount factor.

The *Q*-learning algorithm progresses step-by-step, updating and storing *Q*-values in a table. The initial *Q*-value table reflects the RL agent's foundational knowledge. One of the primary objectives of this study is to assess how effectively trainees teach the agent to accomplish the task. We implemented a fully randomized action selection process to expand the range of scenarios. Moreover, to establish distinct training levels based on the agent's initial knowledge, three techniques were employed to generate initial conditions for the *Q*-learning algorithm.

1) *Zero Initialization*: Set all initial *Q*-values to 0. This indicates that the agent starts with no pre-existing knowledge of the objective and takes random actions. While this type of agent's policy is straightforward to develop, it is time-consuming. This model is utilized in training mode, where each run creates a distinct scenario due to randomness. This variability enables trainees to observe various actions that might not be accessible in conventional educational models.
2) *Arbitrary Initialization*: Initialize *Q*-values with arbitrary values. Here, the agent has some understanding of the objective, although it might be inaccurate or misleading due to the randomness. This agent type serves well in generating distinctive scenarios for assessments.
3) *Partial Training*: Run the training mode for a specific number of iterations without allowing the agent to learn the objective entirely. This model, used in assessment mode, creates specific evaluation scenarios and assesses trainees' ability to detect and correct mistakes.

E. RL-Assisted Tutorial Mode

In contrast to the conventional training models, the tutorial mode presents a unique educational approach utilizing RL. The provided pseudocode, Algorithm 1, outlines the step-by-step process for training an RL agent to interact with PPE. The algorithm initializes crucial parameters, such as the number of states (N), which represents the whole state space, including both dirty and clean PPE states, and the total number of actions (M), consisting of actions DON, DROP, and REPLACE for handling the equipment. In addition, it sets up a *Q*-table to store *Q*-values for different state-action pairs, initializing these values to zeros, which symbolize the agent's initially empty knowledge about the environment and actions. Furthermore, the learning rate ($\alpha = 0.01$) and discount factor ($\gamma = 0.9$) are determined through internal tests via hyperparameter tuning, significantly influencing the agent's learning process within the environment. The RL agent begins by initializing its policy, empty at the beginning, and engaging with the environment. At each step, it takes an action based on the current policy and receives its corresponding reward. The agent's received rewards come from a predetermined rewards table, indicating whether an action is correct or wrong. Simultaneously, participants receive immediate feedback via a UI to indicate the correctness of each action. This feedback UI, as depicted in Fig. 5, briefly displays red and green transparent lights on the screen without disrupting the scene's content.

Throughout the iterative process, the agent's actions are guided by a policy derived from *Q*-values. These *Q*-values are continuously updated based on the received rewards and the estimation of future rewards, following the *Q*-learning algorithm. The agent updates its policy after no item is left to interact within the environment. The agent's ultimate goal is to maximize its cumulative rewards and achieve proficiency in correctly handling the PPE, thereby reaching the optimal policy. As long as the optimal policy is not achieved, a new iteration commences with the most current policy, perpetuating the afore-mentioned process. Fig. 6 depicts the system model of the designed tutorial mode.

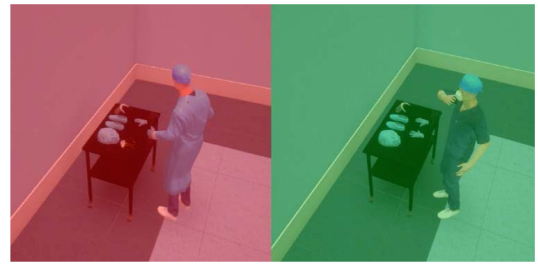


Fig. 5. Feedback UI. Green flashlight when action is correct and red flashlight when action is wrong.

Algorithm 1: RL-Assisted Tutorial Mode With *Q*-Learning.

```
Initialize the number of states N ← 16
Initialize the number of actions M ← 3
Initialize Q-table Q[N x M] ← 0
Initialize learning rate α ← 0.01, discount factor γ ← 0.9
Define current policy π
Get optimal policy π*
Get initial state s
Get rewards R[N x M]
repeat
  Spawn an RL-agent
  repeat
    Take action a in state s under policy π, Q^π(s, a)
    Define current Q-value Q^old ← Q^π(s, a)
    Perform action a, receive new state s'
    Receive available actions a' in state s'
    Immediate reward r ← R(s, a, s')
    Q^new ← Q^old + α[r + γ max_a' Q(s', a') − Q^old]
    Q(s, a) ← Q^new
    s ← s'
  until s is terminal (all PPE donned)
  Get current policy π based on updated Q
until π equals π*
```

F. IRL-Based Assessment Mode

The primary goal in the assessment mode is to appraise participants' performance by assessing their ability to provide correct information and evaluating their capability to rectify errors within a limited timeframe (10 min). Unlike the training mode's use of predetermined rewards with RL, this mode employs external intervention via IRL for the rewarding mechanism. The game flow is similar to the training mode. However, in this mode,

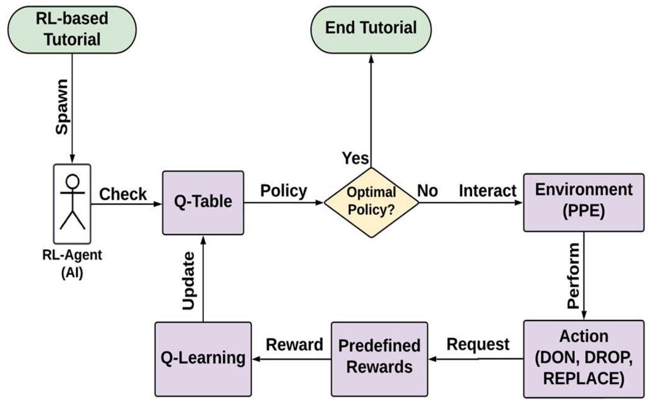


Fig. 6. System model of the training mode.

Fig. 7. Example of interactive mode from the simulation.

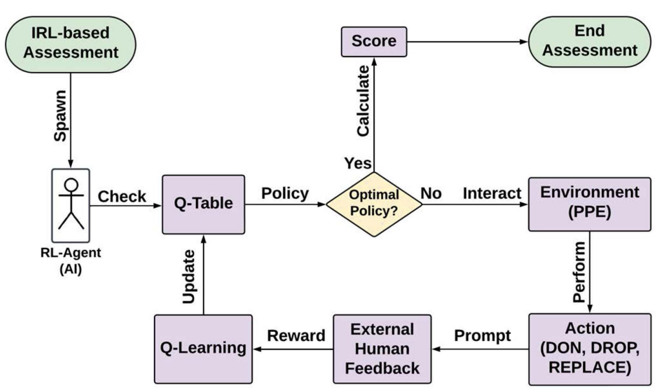


Fig. 8. System model of the assessment mode.

the RL agent begins with a base knowledge comprising 30% incorrect information, leaving the remaining 70% of knowledge empty. Consequently, 70% of actions are randomly picked. Even if the 70% portion of randomization results in correct actions, having 30% of incorrect information ensures that the participants observe wrong decisions made by the RL agent. Hence, completing the assessment necessitates participants correct the RL agent's mistakes.

The pseudocode for the assessment mode (Algorithm 2) shares parameters identical to the tutorial mode given in Algorithm 1, including the number of states (*N*), representing

Algorithm 2: IRL-Assisted Tutorial Mode With *Q*-Learning.

Initialize the number of states $N \leftarrow 16$
Initialize the number of actions $M \leftarrow 3$
Initialize 10-min timer
Initialize *Q*-table $Q[N\ x\ M]$ with 30% ill cases
Initialize learning rate $\alpha \leftarrow 0.01$, discount factor $\gamma \leftarrow 0.9$
Define current policy π
Get optimal policy π^*
Get initial state s
repeat
  Spawn an RL-agent
  repeat
    Take action a in state s under policy π, $Q^{\pi}(s,\ a)$
    Define current *Q*-value $Q^{\text{old}} \leftarrow Q^{\pi}(s,\ a)$
    Perform action a, receive new state s'
    Receive available actions a' in state s'
    Prompt user feedback, get reward r
    $Q^{\text{new}} \leftarrow Q^{\text{old}} + \alpha[r + \gamma \max_{a'} Q(s', a') - Q^{\text{old}}]$
    $Q(s, a) \leftarrow Q^{\text{new}}$
    $s \leftarrow s'$
  until s is terminal (all PPE donned)
  Get current policy π based on updated Q
until π equals π^* or timer runs out
Calculate the score and show

clean and dirty PPE states, and actions (*M*), comprising DON, DROP, and REPLACE for handling the equipment. On the other hand, the *Q*-table initialization differs, as it is prepared with the metadata reflecting 30% ill cases, instead of being zeroed initially in the training mode. In addition, the learning rate (α) and discount factor (γ) are set as determined during internal tests via hyperparameter tuning, mirroring the settings of the tutorial mode.

During each iteration, the agent engages with the environment, performing actions until no PPE remains. Following an action, the game pauses, prompting the trainee to provide positive or negative feedback. Positive feedback indicates that the agent's action was observed as correct, while negative feedback means that the agent's action was perceived as wrong. The interactive process is displayed in Fig. 7 and operates in a loop until the agent attains an optimal policy, or the designated timer expires (e.g., 10 min). Upon completion of the assessment, the trainee's score is determined based on the number of incorrect guidance instances and the time taken to complete the assessment. Fig. 8 illustrates the system model for the assessment mode.

G. User Study Design

A user study was conducted using our agent simulation to evaluate two key research questions: first, how effective is RL compared to traditional training? and second, what are the benefits of using IRL with feedback mechanisms instead of classical RL for learning and assessment?

The study involved 20 participants, with an average age of 24.05 years (ranging from 18 to 32). Among them, there were

15 males and 5 females. None of the participants had prior experience in medical fields, categorizing them as novices in the donning PPE procedure. The participant pool consisted of ten individuals with undergraduate and ten with graduate degrees. It was approved by the IRB committee at Florida Polytechnic University (Approval #23-007), with all participants providing consent. The participants were randomly divided into two groups. Randomization ensured that each group, Group-1 and Group-2, had an equal distribution of undergraduate and graduate degree holders.

During the simulation, the participants' screen was recorded, capturing participants' interactions while experimenting. In addition, in-simulation statistics, such as intermediate Q-value tables after each iteration and participants' scores in the assessment mode, were captured. The recordings facilitated the analysis of participant performance, encapsulating information about the current policy, actions taken for each state, and participant actions, illustrating policy evolution throughout the study.

Group-1 was comprised of ten participants, focusing on evaluating IRL over RL. The experiment involved two sessions. The first session featured observation of an RL-based training, where participants witnessed an RL agent attempting to learn correct PPE donning. Each participant observed different policy developments due to randomization, impacting the training duration (averaging around 5 min). Although the policy progression differs, all participants start with the same empty policy without influencing each other's sessions. Subsequently, participants engaged in an interactive assessment, guiding an ill-trained RL agent for 10 min, scoring based on feedback correctness and completion time. Three iterations of interactive mode were conducted, resulting in 30 recorded data instances for Group-1.

Group-2 consisted of ten participants, assessing RL against the traditional training methods. The experiment was carried out in two sessions. Participants watched a video showcasing an RL agent correctly performing the donning procedure in the training session, differing from Group 1's self-taught RL observation and aiming to simulate the traditional training. Recall that one of the conventional approaches to PPE training involves the use of video lectures, which demonstrate the correct sequence of procedures. Then, they participated in an interactive assessment for three iterations, similar to Group-1. This also resulted in 30 recorded data instances for Group-2.

H. Performance Evaluation and Scoring Metrics

In this study, a comprehensive scoring metric from [96] was employed to evaluate participants' performance in guiding an RL agent through the process of donning PPE. This weighted scoring model, ranging from 0 to 100 in increments of 10, quantified the efficiency of task completion by incorporating two fundamental criteria: accuracy and time efficiency. $\text{Accuracy}\ (A)$ serves as a measure of the correctness of participants' guidance concerning the number of correct feedback provided (C) in relation to the total number of actions (N) the RL agent performed during the experiment. It was calculated using the following equation:

$$A = \text{Round}\left(\frac{C \times 10}{N}\right) \times 10. \tag{4}$$

Time Efficiency (TE), crucial in assessing the teaching pace within the allocated maximum time ($t_{\max}$) of 10 min, employed a decile-based scale relative to the benchmark time ($t_{\text{benchmark}}$) of 5 min. Benchmark time, determined during the internal tests, indicates the optimal maximum time required to complete the assessment. This criterion, evaluated within the range of 0–100, was calculated as follows:

$$\text{TE} = \max\left(0, \min\left(1, \frac{t_{\max} - t}{t_{\max} - t_{\text{benchmark}}}\right)\right) \times 100. \tag{5}$$

Here, t denotes the actual time taken by the participants to complete the assessment. Participants achieving task completion within or under the benchmark time received a perfect score of 100 for the time efficiency portion of the total score, reflecting optimal efficiency. This score linearly diminished in deciles down to 0 as participants approached the maximum limit of 10 min.

Both $\text{Accuracy}\ (A)$ and Time Efficiency (TE) were pivotal in evaluating participant performance, contributing equally to the overall score, shown as follows:

$$\text{Score} = w_1 \times A + w_2 \times \text{TE}. \tag{6}$$

Here, w_1 and w_2 represent the weights assigned to accuracy and time efficiency, respectively. A balanced assessment of both criteria was adopted for this study, assigning equal weights of 0.5 to w_1 and w_2.

It is important to note that our scoring metric, although utilized with equal weights for these two criteria in this study, can accommodate additional criteria by tuning the weights accordingly. This adaptability allows its applicability across diverse scenarios, enabling the inclusion of further performance measures without being limited to only two criteria equally weighted.

IV. Results

To ensure a fair comparison between the two groups, identical scenarios were utilized despite potential alterations caused by randomization and participant interactions within the simulation. During the assessment phase, agents were equipped with initial knowledge comprising 30% of misinformation regarding the donning of PPE. The intentional inclusion of misleading information aimed to evaluate participants' abilities in teaching the agent to unlearn and rectify its existing knowledge. Additionally, the exposure to misleading information presented potential outlier cases to participants. This created a learning environment that highlights knowledge acquisition through identifying and rectifying mistakes—a concept integral to the learning-from-mistakes paradigm.

The duration of the experiment varied due to randomization. For Group-1, the mean time taken by the RL agent to learn the correct sequence in the tutorial mode is 5.587 min (sd = 1.278). In contrast, the mean time to complete the assessment is 2.91 min (sd = 1.967) for Group-1 and 4.954 min (sd = 2.172) for Group-2. This observation suggested that the human

TABLE I
GROUP-1 EXPERIMENTAL RESULTS

Participant	Scores		
	1st Iteration	*2nd Iteration*	*3rd Iteration*
P1	90	20	100
P2	10	80	100
P3	80	100	100
P4	100	90	100
P5	0	90	100
P6	70	100	100
P7	50	70	100
P8	70	80	100
P9	100	100	100
P10	0	0	80

TABLE II
GROUP-2 EXPERIMENTAL RESULTS

Participant	Scores		
	1st Iteration	*2nd Iteration*	*3rd Iteration*
P1	10	0	0
P2	0	40	40
P3	20	60	90
P4	20	0	30
P5	50	70	80
P6	0	0	50
P7	0	60	70
P8	90	40	100
P9	0	0	70
P10	0	50	90

intervention during RL expedited the learning process, which is evident in the shorter assessment duration compared to the tutorial. Similarly, comparing iterations required for the RL agent to grasp the objective indicated an average of 8 in the tutorial mode versus 3.60 in the assessment mode. These results highlight that human intervention accelerates RL policy shaping. However, Group-2 required more time than Group-1 to complete the assessment, emphasizing the positive impact of commencing with an RL-based tutorial rather than a traditional educational model.

The experimental results of both groups are presented in Tables I and II. Nearly all participants from both groups (95%) managed to increase or maintain their scores by the third iteration. Specifically, 80% of Group-1 and 90% of Group-2 improved their scores, with 20% of Group-1 maintaining a perfect score of 100 throughout their sessions. This suggests the effectiveness of both versions of the simulation for learning purposes. Nonetheless, a noticeable distinction was noted in the scores achieved by participants in the two respective groups. As illustrated in Fig. 9, Group-1 exhibited higher starting and ending scores than Group-2, with a mean score of 98 for the third iteration in Group-1 compared to 62 in Group-2. These findings indicate Group-1's superior success in learning the donning process in three iterations, emphasizing RL as a more effective tutorial method compared with the traditional approaches.

Upon opting for RL as an educational model, our focus shifted to evaluating its potential as an independent assessment tool. To investigate this, the assessment mode employed IRL, allowing participants to provide feedback on the RL agent's actions. The analysis concentrated on Group-1, as they were exposed to both

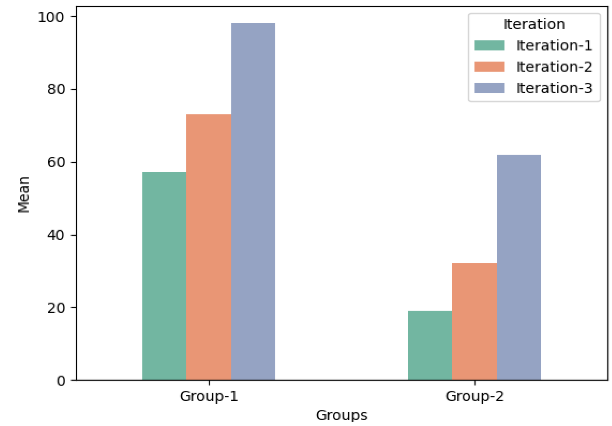


Fig. 9. Mean scores of each iteration of both groups.

RL (training) and IRL (assessment). Despite participants being trained by observing the RL process, the initial mean score was 57 out of 100, indicating that RL alone lacks the potency to replace traditional training methods. However, subsequent iterations in the interactive assessment mode showcased improvement. The mean scores for the second and third iterations were 73 out of 100 and 98 out of 100, respectively. These results underscore the effectiveness of the interactive learning method, where participants learn while teaching RL agents. Participants adjusted the RL agent's policy through interaction, leading to score enhancements. This suggests that IRL, when coupled with RL training, serves as a powerful assessment tool, empowering individuals to improve their performance significantly.

V. DISCUSSION

The experiments were conducted with 20 participants, revealing the efficiency of RL in contrast to the traditional methods. RL facilitated participant observation of the agent's mistakes and subsequent corrections, offering insights into outlier cases often missed by traditional models. Furthermore, the inclusion of IRL, incorporating human intervention, proved more effective than RL alone. However, the objective of this study is not solely to observe that repeated training sessions can lead to learning but also to pioneer a new educational model enabling rapid and precise learning.

Linear regression analysis was employed to correlate the participants' scores in each iteration and assess learning progression based on gathered data. Fig. 10 displays the regression between Iteration-1 and Iteration-2 for each experiment group, indicating that playing only two iterations is insufficient for mastering the objective, which is evident in the prediction model's poor performance. Conversely, Fig. 11 depicts the regression between Iteration-2 and Iteration-3, revealing a significant improvement in Group-1 based on the prediction model, contrasting Group-2's limited progress. Fig. 12 reinforces this by showing the regression between Iteration-1 and Iteration-3, highlighting Group-1's faster learning outcome than Group-2.

The precision of Group-1's prediction model, aligning more accurately with data points than Group-2, suggests Iteration-1's

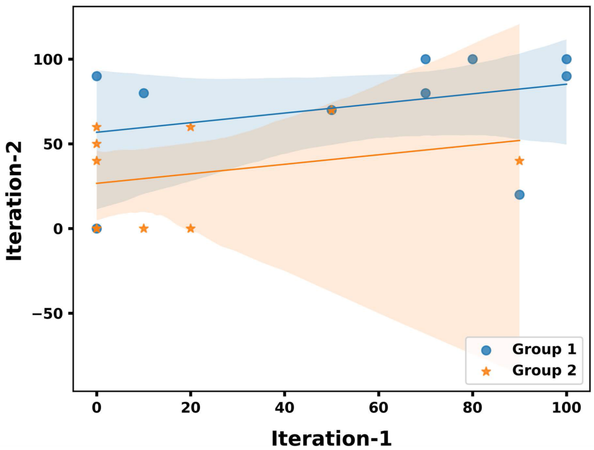


Fig. 10. Linear regression of Iteration-1 and Iteration-2.

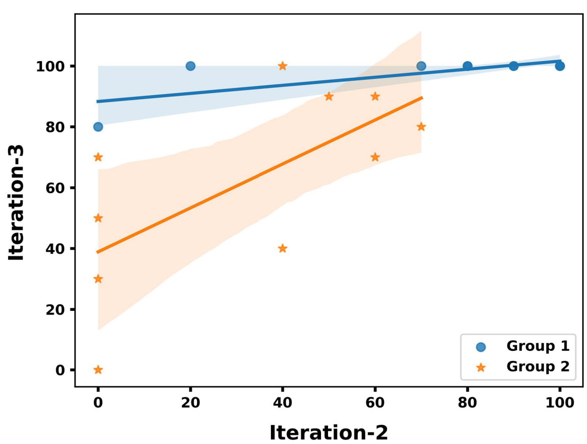


Fig. 11. Linear regression of Iteration-2 and Iteration-3.

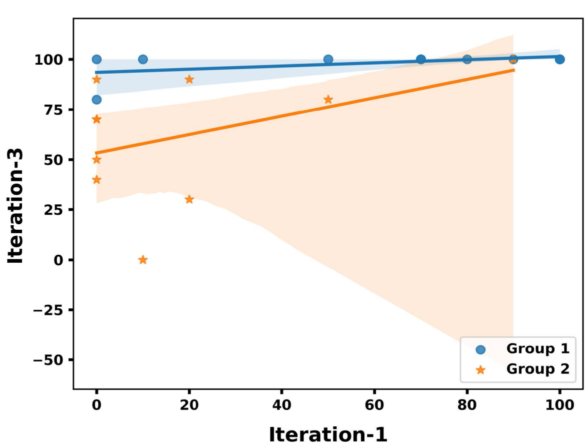


Fig. 12. Linear regression of Iteration-1 and Iteration-3.

stronger positive impact on Iteration-3 for Group-1. Overall, Group-1's iterations exhibit a better correlation, positively influencing subsequent iterations, approaching near-perfect fit after two iterations. In contrast, weak correlations between iterations were observed in Group-2, indicating a requirement for

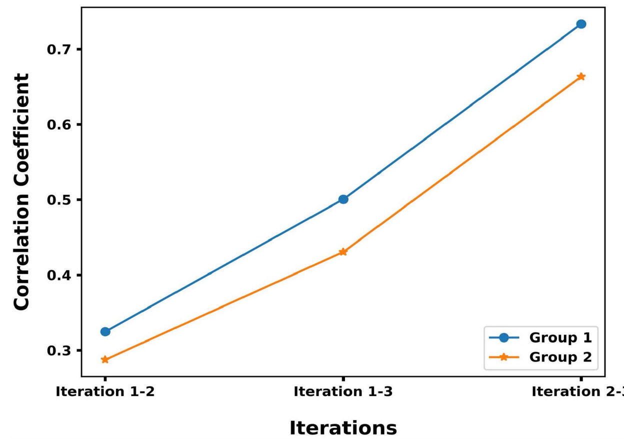


Fig. 13. Correlation coefficient between iterations.

more than three iterations to master the objective. Correlation coefficients shown in Fig. 13 also corroborate these claims, particularly noting the strong correlation between Iteration-2 and Iteration-3 for Group-1 (correlation coefficient = 0.73). Moreover, confidence intervals represented by blue and orange regions around the regression lines in Figs. 9, 10, and 12 validate this claim. Group-1 exhibits narrower intervals, signifying greater certainty in the positive impact of iterations on learning.

However, outliers in the data (e.g., P1 for both groups) necessitate further exploration. Our experiment involved donning various PPE items, some of which comprised pairs, such as gloves and shoe covers, with distinctions between left and right. Participants were briefed that the order of donning the left or right side did not influence correctness; both sequences were considered equally valid. Whether starting with the left or right side first was not a factor in determining the correctness of their actions. However, the subsequent actions after donning pairs were critical. Despite these clarifications, some participants encountered confusion regarding handling paired PPE items. This confusion likely contributed to the emergence of outliers in the data. Further analysis of the screen recordings provided insights into the reasons behind these outliers. This highlights the importance of offering explicit and comprehensive instructions in future simulations to minimize confusion and ensure participants' more transparent comprehension of the task requirements.

Despite the seemingly small sample size of 20 participants in the user study, including three iterations per participant substantially augmented the statistical power of our analysis. Our power analysis, employing a two-sample t-test with an alpha level of 0.05, revealed a power of 0.989, indicating a high probability of detecting a significant difference between Group-1 (mean = 76 and $sd = 34.40$) and Group-2 (mean = 37.66 and $sd = 34.30$).

Furthermore, although the simulator aims to enhance medical training, all participants involved were novices without any medical background. This deliberate choice was made to ensure unbiased results. By engaging individuals unfamiliar with medical practices, we aimed to eliminate potential biases that might have been present among participants with medical expertise. This approach allowed for a more explicit evaluation of the intervention's effectiveness, as the procedure was unknown to all participants, ensuring a more objective assessment of the study outcomes.

As a part of future work, the proposed RL and IRL-based educational model aims to expand its applications beyond the scope of PPE donning and doffing. This adaptable approach, suitable for any operation modeled as MDPs, intends to develop a generic framework usable across various medical procedures. This advancement could significantly benefit medical training by providing versatile educational tools for diverse scenarios. Although the exclusive focus on novice participants assists in revealing the potential of our proposed novel educational approach, future research should engage a broader participant demographic, including experienced medical students. Furthermore, extending simulation scenarios to cover more complex medical situations would fortify the applicability and robustness of our educational approach.

VI. Conclusion

Our research explored transforming medical learning by combining 3-D simulations as adaptable tools for training and evaluation. Utilizing RL in tutorials and IRL in assessments, we introduced a novel educational approach that promotes engaged learning and immediate feedback, specifically for PPE. The application of RL as a tutorial demonstrated its effectiveness, surpassing conventional educational methods. Remarkably, human interaction facilitated by IRL during the assessment phase yielded substantial improvements in learning outcomes and participants' performance.

In our user study involving 20 participants, both Group-1 and Group-2 demonstrated improvements in their mean scores, depicting the success of IRL employed in both groups. For instance, in Group-1, after undergoing an RL tutorial followed by an IRL assessment, mean scores escalated from an initial 57 out of 100 to 98 out of 100 after three iterations. Similarly, Group-2, utilizing traditional training methods and IRL assessments, exhibited substantial progress, elevating their mean scores from 19 out of 100 to 62 out of 100 throughout the three iterations. However, it is noteworthy that Group-1 displayed a higher initial mean score compared to Group-2, emphasizing the effectiveness of RL as a tutorial method in surpassing traditional training approaches. The distinct advantage of starting from a higher mean score in Group-1 indicates RLs potential to accelerate the learning process significantly.

The success of IRL, evident in both groups, underlines the effectiveness of interactive learning environments, enabling participants to actively engage in teaching while learning. In specialized fields, such as medicine, interactive learning within immersive environments proves to be more efficient than traditional classroom settings confined by time and location. Our approach offers solutions to these limitations by seamlessly blending immersive 3-D virtual simulations with education. These discoveries highlight the promise of RL and IRL in educational contexts, particularly within medical training. The effectiveness of our proposed RL-based training and IRL-based assessment indicates a potential shift from conventional approaches to immersive simulation-based training. Moving forward, expanding simulations to cover a wider range of medical scenarios will strengthen this transformative educational model, potentially revolutionizing education across various domains.

References

[1] R. Elmoazen, M. Saqr, M. Khalil, and B. Wasson, "Learning analytics in virtual laboratories: A systematic literature review of empirical research," *Smart Learn. Environ.*, vol. 10, no. 1, Mar. 2023, Art. no. 23, doi: 10.1186/s40561-023-00244-y.

[2] A. Alam, "Leveraging the power of 'modeling and computer simulation' for education: An exploration of its potential for improved learning outcomes and enhanced student engagement," in *Proc. Int. Conf. Device Intell., Comput. Commun. Technol.*, 2023, pp. 445–450, doi: 10.1109/DICCT56244.2023.10110159.

[3] X. Chen, D. Zou, H. Xie, and F. L. Wang, "Metaverse in education: Contributors, cooperations, and research themes," *IEEE Trans. Learn. Technol.*, vol. 16, no. 6, pp. 1111–1129, Dec. 2023, doi: 10.1109/TLT.2023.3277952.

[4] W.-Y. Hwang, R. Nurtantyana, S. W. D. Purba, and U. Hariyanti, and A. Suprapto, "Augmented reality with authentic GeometryGo app to help geometry learning and assessments," *IEEE Trans. Learn. Technol.*, vol. 16, no. 5, pp. 769–779, Oct. 2023, doi: 10.1109/TLT.2023.3251398.

[5] M. Teplá, P. Teplý, and P. Šmejkal, "Influence of 3D models and animations on students in natural subjects," *Int. J. STEM Educ.*, vol. 9, no. 1, Oct. 2022, Art. no. 65, doi: 10.1186/s40594-022-00382-8.

[6] S. S. Oyelere, N. Bouali, R. Kaliisa, G. Obaido, A. A. Yunusa, and E. R. Jimoh, "Exploring the trends of educational virtual reality games: A systematic review of empirical studies," *Smart Learn. Environ.*, vol. 7, no. 1, Oct. 2020, Art. no. 31, doi: 10.1186/s40561-020-00142-7.

[7] T. H. Laine and R. S. N. Lindberg, "Designing engaging games for education: A systematic literature review on game motivators and design principles," *IEEE Trans. Learn. Technol.*, vol. 13, no. 4, pp. 804–821, Oct. 2020, doi: 10.1109/TLT.2020.3018503.

[8] N. Pellas, A. Dengel, and A. Christopoulos, "A scoping review of immersive virtual reality in STEM education," *IEEE Trans. Learn. Technol.*, vol. 13, no. 4, pp. 748–761, Oct. 2020, doi: 10.1109/TLT.2020.3019405.

[9] S. Zhao et al., "An intelligent augmented reality training framework for neonatal endotracheal intubation," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2020, pp. 672–681, doi: 10.1109/ISMAR50242.2020.00097.

[10] D. Vlachopoulos and A. Makri, "The effect of games and simulations on higher education: A systematic literature review," *Int. J. Educ. Technol. Higher Educ.*, vol. 14, no. 1, Jul. 2017, Art. no. 22, doi: 10.1186/s41239-017-0062-1.

[11] M.-B. Ibáñez, Á. Di-Serio, and C. Delgado-Kloos, "Gamification for engaging computer science students in learning activities: A case study," *IEEE Trans. Learn. Technol.*, vol. 7, no. 3, pp. 291–301, Jul. 2014, doi: 10.1109/TLT.2014.2329293.

[12] C.-P. Dai and F. Ke, "Educational applications of artificial intelligence in simulation-based learning: A systematic mapping review," *Comput. Educ., Artif. Intell.*, vol. 3, Jan. 2022, Art. no. 100087, doi: 10.1016/j.caeai.2022.100087.

[13] F. Ke, Z. Dai, C.-P. Dai, M. Pachman, R. S. Chaulagain, and X. Yuan, "Designing virtual agents for simulation-based learning in virtual reality," in *Cognitive and Affective Perspectives on Immersive Technology in Education*. Hershey, PA, USA: IGI Global Publishing, 2020, pp. 151–170, doi: 10.4018/978-1-7998-3250-8.ch008.

[14] J. Dials, D. Demirel, R. Sanchez-Arias, T. Halic, S. De, and M. A. Gromski, "Endoscopic sleeve gastroplasty: Stomach location and task classification for evaluation using artificial intelligence," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 31, Jan. 2024, Art. no. 38212470, doi: 10.1007/s11548-023-03054-2.

[15] A. Jalalimanesh, H. Shahabi Haghighi, A. Ahmadi, and M. Soltani, "Simulation-based optimization of radiotherapy: Agent-based modeling and reinforcement learning," *Math. Comput. Simul.*, vol. 133, pp. 235–248, Mar. 2017, doi: 10.1016/j.matcom.2016.05.008.

[16] Y. Bi, Z. Jiang, Y. Gao, T. Wendler, A. Karlas, and N. Navab, "VesNet-RL: Simulation-based reinforcement learning for real-world us probe navigation," *IEEE Robot. Autom. Lett.*, vol. 7, no. 3, pp. 6638–6645, Jul. 2022, doi: 10.1109/LRA.2022.3176112.

[17] J. Zeng, S. Parks, and J. Shang, "To learn scientifically, effectively, and enjoyably: A review of educational games," *Hum. Behav. Emerg. Technol.*, vol. 2, pp. 107–195, Apr. 2020, doi: 10.1002/hbe2.188.

[18] A. Shah, C. L. Mai, R. Shah, and A. I. Levine, "Simulation-based education and team training," *Otolaryngologic Clin. North Amer.*, vol. 52, no. 6, pp. 995–1003, Dec. 2019, doi: 10.1016/j.otc.2019.08.002.

[19] T. Ören, C. Turnitsa, S. Mittal, and S. Y. Diallo, "Simulation-based learning and education," in *Guide to Simulation-Based Disciplines*, S. Mittal, U. Durak, and T. Ören, Eds. Cham, Switzerland: Springer, 2017, pp. 293–314, doi: 10.1007/978-3-319-61264-5_13.

[20] F. Lateef, "Simulation-based learning: Just like the real thing," *J. Emergencies, Trauma, Shock*, vol. 3, no. 4, pp. 348–352, 2010, doi: 10.4103/0974-2700.70743.

[21] H. Y. So, P. P. Chen, G. K. C. Wong, and T. T. N. Chan, "Simulation in medical education," *J. Roy. College Physicians Edinburgh*, vol. 49, no. 1, pp. 52–57, Mar. 2019, doi: 10.4997/jrcpe.2019.112.

[22] M. Aebersold, "The history of simulation and its impact on the future," *AACN Adv. Crit. Care*, vol. 27, no. 1, pp. 56–61, Feb. 2016, doi: 10.4037/aacnacc2016436.

[23] G. Offiah et al., "Evaluation of medical student retention of clinical skills following simulation training," *BMC Med. Educ.*, vol. 19, no. 1, Jul. 2019, Art. no. 263, doi: 10.1186/s12909-019-1663-2.

[24] A. H. Al-Elq, "Simulation-based medical teaching and learning," *J. Fam. Community Med.*, vol. 17, no. 1, pp. 35–40, 2010, doi: 10.4103/1319-1683.68787.

[25] N. D. Salimova, M. S. Salaeva, and H. K. Boltaboev, "Simulation training in medicine," *J. Modern Educ. Achievements*, vol. 3, pp. 138–142, 2010.

[26] D. Demirel et al., "Virtual airway skills trainer (VAST) simulator," *Stud. Health Technol. Inform.*, vol. 220, pp. 91–97, 2016.

[27] D. Demirel, A. Yu, T. Halic, and S. Kockara, "Web based camera navigation for virtual pancreatic cancer surgery: Whipple surgery simulator (VPanSS)," in *Proc. IEEE Innov. Technol. Conf.*, 2014, pp. 1–8, doi: 10.1109/InnoTek.2014.6877375.

[28] B. Cetinsaya et al., "Design of virtual endoluminal surgery simulator (VESS): Colorectal endoscopic submucosal dissection training module: 809," *Official J. Amer. College Gastroenterol.*, vol. 112, Oct. 2017, Art. no. S452.

[29] J. Farmer et al., "Virtual rotator cuff arthroscopic skill trainer: Results and analysis of a preliminary subject study," in *Proc. 2020 4th Int. Conf. Inf. Syst. Data Mining*, 2020, pp. 139–143, doi: 10.1145/3404663.3404673.

[30] D. Qi et al., "Surgeons with five or more actual cricothyrotomies perform significantly better on a virtual reality simulator," *J. Surg. Res.*, vol. 252, pp. 247–254, Aug. 2020, doi: 10.1016/j.jss.2020.03.021.

[31] I. Motola, L. A. Devine, H. S. Chung, J. E. Sullivan, and S. B. Issenberg, "Simulation in healthcare education: A best evidence practical guide. AMEE guide no. 82," *Med. Teacher*, vol. 35, no. 10, pp. e1511–e1530, Oct. 2013, doi: 10.3109/0142159X.2013.818632.

[32] E. Herrera-Aliaga and L. D. Estrada, "Trends and innovations of simulation for twenty first century medical education," *Front. Public Health*, vol. 10, 2022, Accessed on: Dec. 11, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpubh.2022.619769

[33] D. Gill, C. Whitehead, and D. Wondimagegn, "Challenges to medical education at a time of physical distancing," *Lancet*, vol. 396, no. 10244, pp. 77–79, 2020, doi: 10.1016/S0140-6736(20)31368-4.

[34] P. A. Guze, "Using technology to meet the challenges of medical education," *Trans. Amer. Clin. Climatol. Assoc.*, vol. 126, pp. 260–270, 2015.

[35] R. S. Sutton and A. G. Barto, *Reinforcement Learning*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[36] K. Subramanian, C. L. I. Isbell, and A. L. Thomaz, "Exploration from demonstration for interactive reinforcement learning," in *Proc. 15th Int. Conf. Auton. Agents Multiagent Syst.*, 2016, pp. 1–10.

[37] J. H. Kwon et al., "Assessment of healthcare worker protocol deviations and self-contamination during personal protective equipment donning and doffing," *Infection Control Hosp. Epidemiol.*, vol. 38, no. 9, pp. 1077–1083, Sep. 2017, doi: 10.1017/ice.2017.121.

[38] T. M. Cook, "Personal protective equipment during the coronavirus disease (COVID) 2019 pandemic—A narrative review," *Anaesthesia*, vol. 75, no. 7, pp. 920–927, 2020, doi: 10.1111/anae.15071.

[39] L. Yuan, S. Chen, and Y. Xu, "Donning and doffing of personal protective equipment protocol and key points of nursing care for patients with COVID-19 in ICU," *Stroke Vasc. Neurol.*, vol. 5, no. 3, pp. 302–307, Sep. 2020, doi: 10.1136/svn-2020-000456.

[40] L. Christensen, C. S. Rasmussen, T. Benfield, and J. M. Franc, "A randomized trial of instructor-led training versus video lesson in training health care providers in proper donning and doffing of personal protective equipment," *Disaster Med. Public Health Preparedness*, vol. 14, no. 4, pp. 514–520, Aug. 2020, doi: 10.1017/dmp.2020.56.

[41] S. Aminizadeh et al., "The applications of machine learning techniques in medical data processing based on distributed computing and the Internet of Things," *Comput. Methods Programs Biomed.*, vol. 241, Nov. 2023, Art. no. 107745, doi: 10.1016/j.cmpb.2023.107745.

[42] C. Rainey, A. T. Villikudathil, J. McConnell, C. Hughes, R. Bond, and S. McFadden, "An experimental machine learning study investigating the decision-making process of students and qualified radiographers when interpreting radiographic images," *PLoS Digit. Health*, vol. 2, no. 10, Oct. 2023, Art. no. e0000229, doi: 10.1371/journal.pdig.0000229.

[43] P. Manickam et al., "Artificial intelligence (AI) and Internet of Medical Things (IoMT) assisted biomedical systems for intelligent healthcare," *Biosensors*, vol. 12, no. 8, Aug. 2022, Art. no. 562, doi: 10.3390/bios12080562.

[44] J. Bajwa, U. Munir, A. Nori, and B. Williams, "Artificial intelligence in healthcare: Transforming the practice of medicine," *Future Healthcare J.*, vol. 8, no. 2, pp. e188–e194, Jul. 2021, doi: 10.7861/fhj.2021-0095.

[45] G. Rong, A. Mendez, E. Bou Assi, B. Zhao, and M. Sawan, "Artificial intelligence in healthcare: Review and prediction case studies," *Engineering*, vol. 6, no. 3, pp. 291–301, Mar. 2020, doi: 10.1016/j.eng.2019.08.015.

[46] J. He, S. L. Baxter, Jie Xu, J. Xu, X. Zhou, and K. Zhang, "The practical implementation of artificial intelligence technologies in medicine," *Nature Med.*, vol. 25, no. 1, pp. 30–36, Jan. 2019, doi: 10.1038/s41591-018-0307-0.

[47] F. Jiang et al., "Artificial intelligence in healthcare: Past, present and future," *Stroke Vasc. Neurol.*, vol. 2, no. 4, Dec. 2017, Art. no. 2564, doi: 10.1136/svn-2017-000101.

[48] M. M. Dagli, A. Rajesh, M. Asaad, and C. E. Butler, "The use of artificial intelligence and machine learning in surgery: A comprehensive literature review," *Amer. Surgeon*, vol. 89, no. 5, pp. 1980–1988, May 2023, doi: 10.1177/00031348211065101.

[49] P. Tan, X. Chen, H. Zhang, Q. Wei, and K. Luo, "Artificial intelligence aids in development of nanomedicines for cancer management," *Seminars Cancer Biol.*, vol. 89, pp. 61–75, Feb. 2023, doi: 10.1016/j.semcancer.2023.01.005.

[50] X. Gu et al., "Beyond supervised learning for pervasive healthcare," *IEEE Rev. Biomed. Eng.*, vol. 17, pp. 42–62, 2024, doi: 10.1109/RBME.2023.3296938.

[51] N. Meenigea and V. R. K. Kolla, "Exploring the current landscape of artificial intelligence in healthcare," *Int. J. Sustain. Develop. Comput. Sci.*, vol. 5, Mar. 2023, Accessed on: Dec. 12, 2023. [Online]. Available: https://www.ijsdcs.com/index.php/ijsdcs/article/view/285

[52] A. S. Albahri et al., "A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion," *Inf. Fusion*, vol. 96, pp. 156–191, Aug. 2023, doi: 10.1016/j.inffus.2023.03.008.

[53] S. A. Alowais et al., "Revolutionizing healthcare: The role of artificial intelligence in clinical practice," *BMC Med. Educ.*, vol. 23, no. 1, Sep. 2023, Art. no. 689, doi: 10.1186/s12909-023-04698-z.

[54] U. Donins and D. Behmane, "Challenges and solutions for artificial intelligence adoption in healthcare—A literature review," in *Innovation in Medicine and Healthcare*, Y.-W. Chen, S. Tanaka, R. J. Howlett, and L. C. Jain Eds. Singapore: Springer, 2023, pp. 53–62, doi: 10.1007/978-981-99-3311-2_6.

[55] D. Saraswat et al., "Explainable AI for healthcare 5.0: Opportunities and challenges," *IEEE Access*, vol. 10, pp. 84486–84517, 2022, doi: 10.1109/ACCESS.2022.3197671.

[56] H. W. Loh, C. P. Ooi, S. Seoni, P. D. Barua, F. Molinari, and U. R. Acharya, "Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022)," *Comput. Methods Programs Biomed.*, vol. 226, Nov. 2022, Art. no. 107161, doi: 10.1016/j.cmpb.2022.107161.

[57] S. Zhao, X. Zhang, F. Jin, and J. Hahn, "An auxiliary tasks based framework for automated medical skill assessment with limited data," in *Proc. IEEE 43rd Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2021, pp. 1613–1617, doi: 10.1109/EMBC46164.2021.9630498.

[58] A. A. Abdellatif, N. Mhaisen, A. Mohamed, A. Erbad, and M. Guizani, "Reinforcement learning for intelligent healthcare systems: A review of challenges, applications, and open research issues," *IEEE Internet Things J.*, vol. 10, no. 24, pp. 21982–22007, Dec. 2023, doi: 10.1109/JIOT.2023.3288050.

[59] A. Coronato, M. Naeem, G. De Pietro, and G. Paragliola, "Reinforcement learning for intelligent healthcare applications: A survey," *Artif. Intell. Med.*, vol. 109, Sep. 2020, Art. no. 101964, doi: 10.1016/j.artmed.2020.101964.

[60] S. Saghafian, "Ambiguous dynamic treatment regimes: A reinforcement learning approach," *Manage. Sci.*, vol. 31, Oct. 2023, Art. no. 25642, doi: 10.1287/mnsc.2022.00883.

[61] H. Cho, S. T. Holloway, D. J. Couper, and M. R. Kosorok, "Multi-stage optimal dynamic treatment regimes for survival outcomes with dependent censoring," *Biometrika*, vol. 110, no. 2, pp. 395–410, Jun. 2023, doi: 10.1093/biomet/asac047.

[62] C.-Y. Yang, C. Shiranthika, C.-Y. Wang, K.-W. Chen, and S. Sumathipala, "Reinforcement learning strategies in cancer chemotherapy treatments: A review," *Comput. Methods Programs Biomed.*, vol. 229, Feb. 2023, Art. no. 107280, doi: 10.1016/j.cmpb.2022.107280.

[63] S. I. H. Shah, A. Coronato, M. Naeem, and G. De Pietro, "Learning and assessing optimal dynamic treatment regimes through cooperative imitation learning," *IEEE Access*, vol. 10, pp. 78148–78158, 2022, doi: 10.1109/ACCESS.2022.3193494.

[64] L. Wang, W. Zhang, X. He, and H. Zha, "Supervised reinforcement learning with recurrent neural network for dynamic treatment recommendation," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2018, pp. 2447–2456.

[65] Y. Liu, B. Logan, N. Liu, Z. Xu, J. Tang, and Y. Wang, "Deep reinforcement learning for dynamic treatment regimes on medical registry data," in *Proc. IEEE Int. Conf. Healthcare Inform.*, 2017, pp. 380–385, doi: 10.1109/ICHI.2017.45.

[66] J. Pineau, A. Guez, R. Vincent, G. Panuccio, and M. Avoli, "Treating epilepsy via adaptive neurostimulation: A reinforcement learning approach," *Int. J. Neural Syst.*, vol. 19, no. 4, pp. 227–240, Aug. 2009, doi: 10.1142/S0129065709001987.

[67] Y. Zhao, D. Zeng, M. A. Socinski, and M. R. Kosorok, "Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer," *Biometrics*, vol. 67, no. 4, pp. 1422–1433, 2011, doi: 10.1111/j.1541-0420.2011.01572.x.

[68] C. Yu and Q. Huang, "Towards more efficient and robust evaluation of sepsis treatment with deep reinforcement learning," *BMC Med. Inform. Decis. Mak.*, vol. 23, no. 1, Mar. 2023, Art. no. 43, doi: 10.1186/s12911-023-02126-2.

[69] M. Hu, J. Zhang, L. Matkovic, T. Liu, and X. Yang, "Reinforcement learning in medical image analysis: Concepts, applications, challenges, and future directions," *J. Appl. Clin. Med. Phys.*, vol. 24, no. 2, 2023, Art. no. e13898, doi: 10.1002/acm2.13898.

[70] L. Xu, S. Zhu, and N. Wen, "Deep reinforcement learning and its applications in medical imaging and radiation therapy: A survey," *Phys. Med. Biol.*, vol. 67, no. 22, Nov. 2022, Art. no. 22TR02, doi: 10.1088/1361-6560/ac9cb3.

[71] F. M. Calisto, N. Nunes, and J. C. Nascimento, "Modeling adoption of intelligent agents in medical imaging," *Int. J. Human-Comput. Stud.*, vol. 168, Dec. 2022, Art. no. 102922, doi: 10.1016/j.ijhcs.2022.102922.

[72] G. Maicas, G. Carneiro, A. P. Bradley, J. C. Nascimento, and I. Reid, "Deep reinforcement learning for active breast lesion detection from DCE-MRI," in *Medical Image Computing and Computer Assisted Intervention*, M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, and S. Duchesne Eds. Cham, Switzerland: Springer, 2017, pp. 665–673, doi: 10.1007/978-3-319-66179-7_76.

[73] J. N. Stember and H. Shalu, "Reinforcement learning using deep Q networks and Q learning accurately localizes brain tumors on MRI with very small training sets," *BMC Med. Imag.*, vol. 22, no. 1, Dec. 2022, Art. no. 224, doi: 10.1186/s12880-022-00919-x.

[74] S. Chen et al., "Reinforcement learning based diagnosis and prediction for COVID-19 by optimizing a mixed cost function from CT images," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 11, pp. 5344–5354, Nov. 2022, doi: 10.1109/JBHI.2022.3197666.

[75] A. T. Bourdillon, A. Garg, H. Wang, Y. J. Woo, M. Pavone, and J. Boyd, "Integration of reinforcement learning in a virtual robotic surgical simulation," *Surg. Innov.*, vol. 30, no. 1, pp. 94–102, Feb. 2023, doi: 10.1177/15533506221095298.

[76] A. A. Shahkoo and A. A. Abin, "Deep reinforcement learning in continuous action space for autonomous robotic surgery," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 18, no. 3, pp. 423–431, Mar. 2023, doi: 10.1007/s11548-022-02789-8.

[77] P. M. Scheikl et al., "Sim-to-real transfer for visual reinforcement learning of deformable object manipulation for robot-assisted surgery," *IEEE Robot. Autom. Lett.*, vol. 8, no. 2, pp. 560–567, Feb. 2023, doi: 10.1109/LRA.2022.3227873.

[78] N. D. Nguyen, T. Nguyen, S. Nahavandi, A. Bhatti, and G. Guest, "Manipulating soft tissues by deep reinforcement learning for autonomous robotic surgery," in *Proc. IEEE Int. Syst. Conf.*, 2019, pp. 1–7, doi: 10.1109/SYSCON.2019.8836924.

[79] Z.-Y. Chiu, F. Richter, E. K. Funk, R. K. Orosco, and M. C. Yip, "Bimanual regrasping for suture needles using reinforcement learning for rapid motion planning," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2021, pp. 7737–7743, doi: 10.1109/ICRA48506.2021.9561673.

[80] X. Tan, C.-B. Chng, Y. Su, K.-B. Lim, and C.-K. Chui, "Robot-assisted training in laparoscopy using deep reinforcement learning," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 485–492, Apr. 2019, doi: 10.1109/LRA.2019.2891311.

[81] L. Ouyang et al., "Training language models to follow instructions with human feedback," *Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 27730–27744, Dec. 2022.

[82] S. S. Shuvo and Y. Yilmaz, "Home energy recommendation system (HERS): A deep reinforcement learning method based on residents' feedback and activity," *IEEE Trans. Smart Grid*, vol. 13, no. 4, pp. 2812–2821, Jul. 2022, doi: 10.1109/TSG.2022.3158814.

[83] Y. A. Alshumaimeri and A. K. Alshememry, "The extent of AI applications in EFL learning and teaching," *IEEE Trans. Learn. Technol.*, vol. 17, pp. 653–663, 2024, doi: 10.1109/TLT.2023.3322128.

[84] H. Ma, Z. Huang, W. Tang, H. Zhu, H. Zhang, and J. Li, "Predicting student performance in future exams via neutrosophic cognitive diagnosis in personalized e-learning environment," *IEEE Trans. Learn. Technol.*, vol. 16, no. 5, pp. 680–693, Oct. 2023, doi: 10.1109/TLT.2023.3240931.

[85] G. P. Jain, V. P. Gurupur, J. L. Schroeder, and E. D. Faulkenberry, "Artificial intelligence-based student learning evaluation: A concept map-based approach for analyzing a student's understanding of a topic," *IEEE Trans. Learn. Technol.*, vol. 7, no. 3, pp. 267–279, Jul. 2014, doi: 10.1109/TLT.2014.2330297.

[86] I. Magnisalis, S. Demetriadis, and A. Karakostas, "Adaptive and intelligent systems for collaborative learning support: A review of the field," *IEEE Trans. Learn. Technol.*, vol. 4, no. 1, pp. 5–20, Jan./Mar. 2011, doi: 10.1109/TLT.2011.2.

[87] H. Munir, B. Vogel, and A. Jacobsson, "Artificial intelligence and machine learning approaches in digital education: A systematic revision," *Information*, vol. 13, no. 4, Apr. 2022, Art. no. 203, doi: 10.3390/info13040203.

[88] E. Mousavinasab, N. Zarifsanaiey, S. R. Niakan Kalhori, M. Rakhshan, L. Keikha, and M. Ghazi Saeedi, "Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods," *Interactive Learn. Environ.*, vol. 29, no. 1, pp. 142–163, Jan. 2021, doi: 10.1080/10494820.2018.1558257.

[89] B. H. S. Sarma and B. Ravindran, "Intelligent tutoring systems using reinforcement learning to teach autistic students," in *Home Informatics and Telematics: ICT for The Next Billion*, A. Venkatesh, T. Gonsalves, A. Monk, and K. Buckner Eds. Boston, MA, USA: Springer, 2007, pp. 65–78, doi: 10.1007/978-0-387-73697-6_5.

[90] J. M. Gascueña and A. Fernández-Caballero, "An agent-based intelligent tutoring system for enhancing e-learning/e-teaching," *Int. J. Instructional Technol. Distance Learn.*, vol. 2, no. 11, pp. 11–24, 2005.

[91] K. Grochow, S. L. Martin, A. Hertzmann, and Z. Popović, "Style-based inverse kinematics," in *Proc. 31st Annu. Conf. Comput. Graph. Interactive Techn.*, 2004, pp. 522–531, doi: 10.1145/1186562.1015755.

[92] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Hoboken, NJ, USA: Wiley, 2014.

[93] M. van Otterlo and M. Wiering, "Reinforcement learning and Markov decision processes," in *Reinforcement Learning: State-of-the-Art*, M. Wiering and M. van Otterlo, Eds., Berlin, Germany: Springer, 2012, pp. 3–42, doi: 10.1007/978-3-642-27645-3_1.

[94] S. Griffith, K. Subramanian, J. Scholz, C. L. Isbell, and A. L. Thomaz, "Policy shaping: Integrating human feedback with reinforcement learning," *Adv. Neural Inf. Process. Syst.*, vol. 26, pp. 1–9, 2013.

[95] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, no. 3, pp. 279–292, May 1992, doi: 10.1007/BF00992698.

[96] J. S. Dodgson, M. Spackman, A. Pearman, and L. D. Phillips, "Multi-criteria analysis: A manual," London School of Economics and Political Science, Department of Economic History, Economic History Working Papers 12761, 2009.

Yasar C. Kakdas received the bachelor's degree in computer science from the TOBB University of Economics and Technology, Ankara, Türkiye, in 2020.

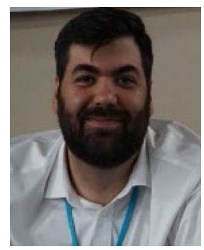
He is a graduate student and a Research Assistant in computer science with Florida Polytechnic University, Lakeland, FL, USA. After he had worked as a mobile application developer for two years, he started working on virtual reality-based simulations and game development. His research interests include computer graphics, virtual reality, artificial intelligence, and mobile applications.

Sinan Kockara received the undergraduate degree in computer engineering from Dokuz Eylul University, Izmir, Türkiye, in 2001, and the Ph.D. degree in applied computing from the University of Arkansas at Little Rock, Little Rock, AR, USA, in 2008.

Subsequently, he joined the University of Central Arkansas, where he later became a Full Professor at the Department of Computer Science and Engineering. He currently holds a faculty position with Computer Science Department, Rice University in Houston, Houston, TX, USA. His research focuses on various areas, including biomedical informatics, medical image processing and analysis, particularly in dermoscopy, the development of minimally invasive surgical simulations for virtual reality environments, parallel computing, and GPGPU.

Tansel Halic received the Ph.D. degree in mechanical engineering from the Department of Mechanical, Aerospace, and Nuclear Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA.

He is currently a Senior Software Engineer with Intuitive Surgical, Inc., Sunnyvale, CA, USA. Before joining Intuitive Surgical, he was an Associate Professor at the University of Central Arkansas and the Director of Virtual Reality, Simulation, Imaging and Modeling Lab.



Doga Demirel (Member, IEEE) received the Ph.D. degree in computer science from the University of Arkansas, Little Rock, AR, USA.

He is currently an Associate Professor with Florida Polytechnic University, Lakeland, FL, USA, where he is the Director of Virtual Reality, Interactive Simulation, and Biomedical Lab. He has worked in the areas of medical simulations for more than 10 years. His research interests include virtual reality, augmented reality, surgical simulators, and haptics.