**Reinforcement Learning for Optimizing Delivery Path in a Hospital Setting**

**Anna Nardelli1, Luke Shao2, Brandon Hu3**

**1 Department of Computer Science and Software Engineering, Monmouth University ​**

**2 Department of Computer Science, Cornell University**

**3 Biotechnology High School**

**Faculty Advisor: Dr. Jiacun Wang**

**Department of Computer Science and Software Engineering**

**Funding Source:**

**School of Science**

Abstract

Reinforcement learning (RL) is a type of machine learning that has many applications in real-world industry. RL intends to “teach” a model best-decision practices through exploration and trial-and-error. This project explores an application of RL in the medical field, creating a model that can efficiently navigate a space and complete tasks in a sample hospital environment. We created a simulation of a hospital floor and programmed an agent to learn the fastest route to a destination while completing tasks and avoiding obstacles. We envision our agent to be a robot that can optimally pick-up and deliver supplies/medication to specific rooms on the floor, which would make hospital practices more efficient and keep patients happier and healthier. Our model was created through a Python program and OpenAI’s Gym library. We modified a Gym environment called GridWorld, adding custom obstacles, actions, and desired locations to reach. The Python program was created to model our RL mechanism which is fundamental to the success of the project. The program creates a state-transition system using Q-learning, simulating all of the possible movement decisions the agent can make. For each of these movements, the agent receives a “reward” of some numerical value. During training, the agent is allowed to explore the environment, calculating a Q-value for each movement, which is then stored in its own matrix. The Q-value is a mathematical estimation of the immediate and long-term value of an action. The model thus learns which actions produce the highest reward through the Q-table’s analysis, and the agent learns to take the most efficient path to complete its tasks by seeking the highest possible long-term reward. Ultimately, the potential applications of this RL model are endless, with significance across industries in decision-making systems to simulate optimal outcomes.