APPENDIX

APPENDIX A

Q-learning

Q-learning is a popular Reinforcement learning technique used to find the optimal action-selection policy for a given finite Markov decision process (MDP) (Jayaraman et al., 2024). It aims to maximize the cumulative reward an agent can achieve by learning the best actions to take from each state. In the section on the RL Framework, we will discuss it in Appendix B.

1. **Agent and Environment**: The agent interacts with the environment, which is represented by states (s) and actions (a).
2. **Q-values**: The agent maintains a table of Q-values, $Q(s, a)$, which estimates the expected utility (or cumulative reward) of taking action a in state s and following the optimal policy thereafter.
3. **Initialization**: Initially, Q-values are typically set to arbitrary values.
4. **Policy**: The agent selects actions based on its current Q-values, often using an exploration-exploitation strategy (e.g., **ε**-greedy policy) to balance between exploring new actions and exploiting known ones.
5. **Learning**:

    5.1 **Experience**: The agent takes an action a in state s, receives a reward r, and transitions to a new state s',

    5.2 **Update Rule**: The Q-value for the state-action pair (s, a) is updated using the formula:

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \, max \;\; a'Q(s',a') - Q(s,a)] \qquad \textbf{(A.1.1)}$$

where

α is the learning rate (how much new information overrides old information),

γ is the discount factor (how much future rewards are valued compared to immediate rewards),

$max \;\; a'Q(s',a')$ is the maximum estimated Q-value for the next state s'.

6. **Convergence**: Over time, as the agent explores the environment and updates its Q-values, the Q-values converge to the true Q-values, allowing the agent to learn the optimal policy for selecting actions.

Q-learning is model-free, meaning it doesn't require knowledge of the environment's dynamics, making it versatile and widely applicable in various Reinforcement learning problems.

APPENDIX B

RL-based SuraSole Maze Game Level Difficulty Adjustment

The Q-learning method is one of the popular and widely used Reinforcement learning (RL) algorithms. Its goal is to enable an agent to learn the best decisions in various situations through trial and error. It uses the concept of Q-values, which measure the value of taking an action in a given state. These Q-values help the agent determine which actions will yield the highest long-term rewards (Watkins and Dayan, 1992).

B.1 Defining the states, actions and rewards for SuraSole maze game

State: The state refers to the condition of the patient, which in this framework is the COP of the patient or the current situation. In this work, we define the states where the coordinates of the COP are located, i.e., as Quadrant 1, Quadrant 2, Quadrant 3, Quadrant 4, and Quadrant 5. Each quadrant represents a different patient condition based on the COP value as shown in Figure B.1.1.

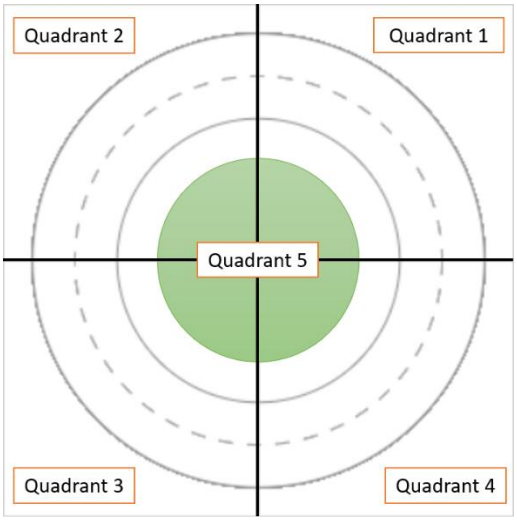
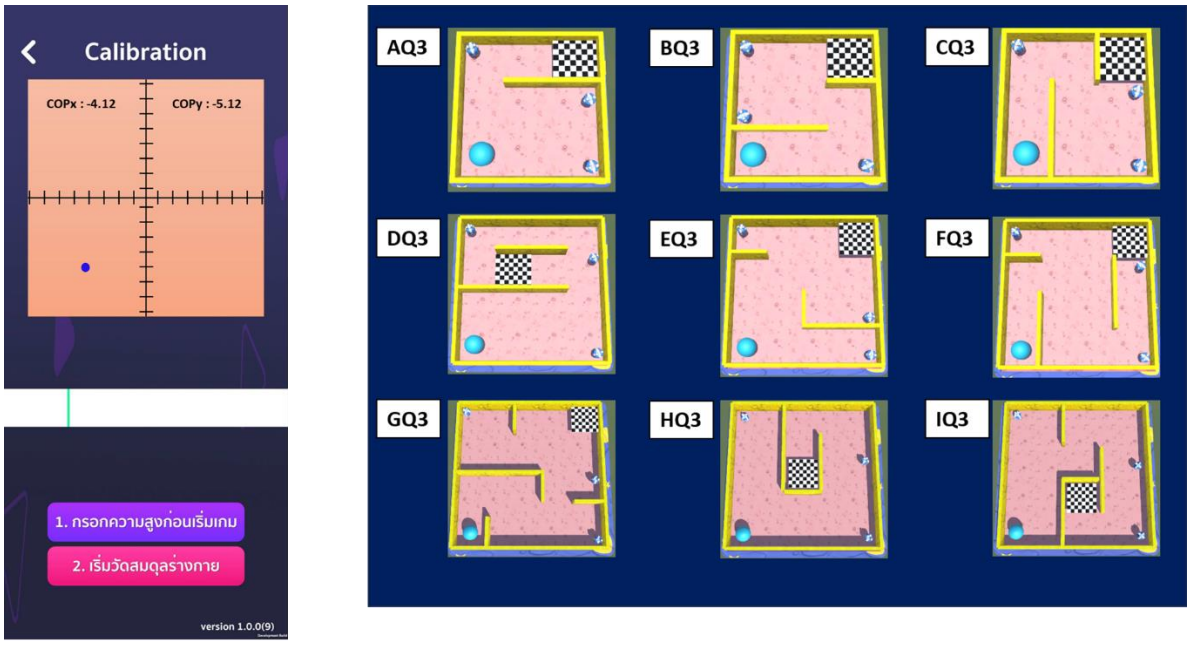


Figure B.1.1 Division of quadrants in various states.

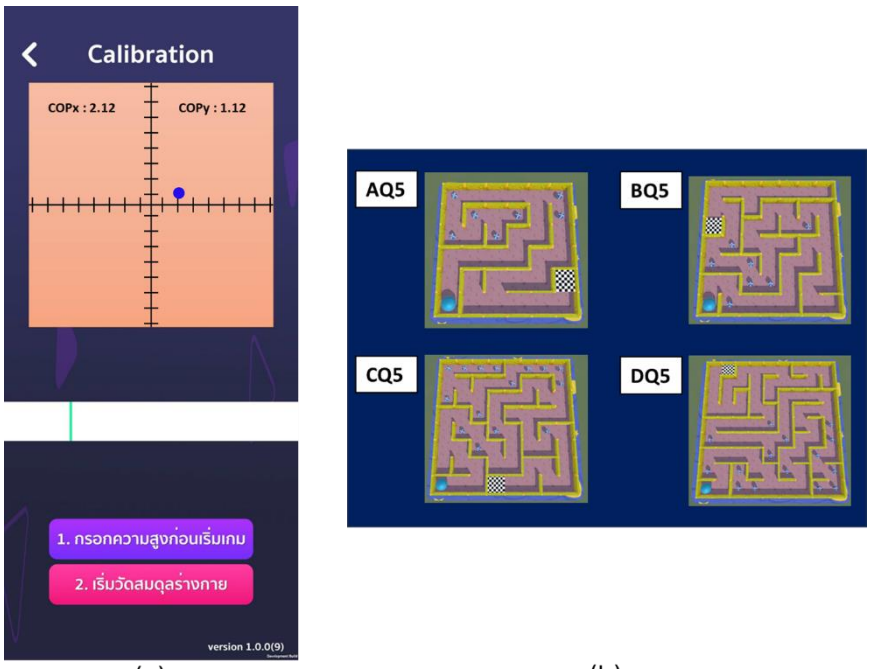
Action: The action refers to the decisions made by the agent to improve the current state of the player. In this framework, actions are defined as the 9 maze levels, each of which is designed based on the (Baranyi et al., 2013) template and corresponds to the COP state. For example, if a participant's COP is located in Quadrant 3, the RL system will select one of the 9 levels specifically designed for Quadrant 1 to match the player's state. For Quadrant 5, which is designed for individuals with normal COP conditions, the RL system will select from only 4 levels, as illustrated in Figures B.1.2 and B.1.3.

Additionally, to illustrate the maze designs of other Quadrants, we have included Figures B.1.4, B.1.5, and B.1.6, which present the maze layouts for Quadrants 1, 2, and 4, respectively. Each maze is designed to provide challenges that correspond to the COP positions of players within those specific Quadrants, enabling the RL system to select levels that are well-matched to the player's abilities and progress with precision.



(a) (b)

Figure B.1.2 COP states and RL-selected levels for Quadrant 3, (a) SuraSole maze game interface showing COP in Quadrant 3 before the game (b) SuraSole maze game level for a sample COP in Quadrant 3



(a) (b)

Figure B.1.3 COP states and RL-selected levels for Quadrant 5, (a) SuraSole maze game interface showing COP in Quadrant 5 before the game, (b) SuraSole maze game level for a sample COP in Quadrant 5

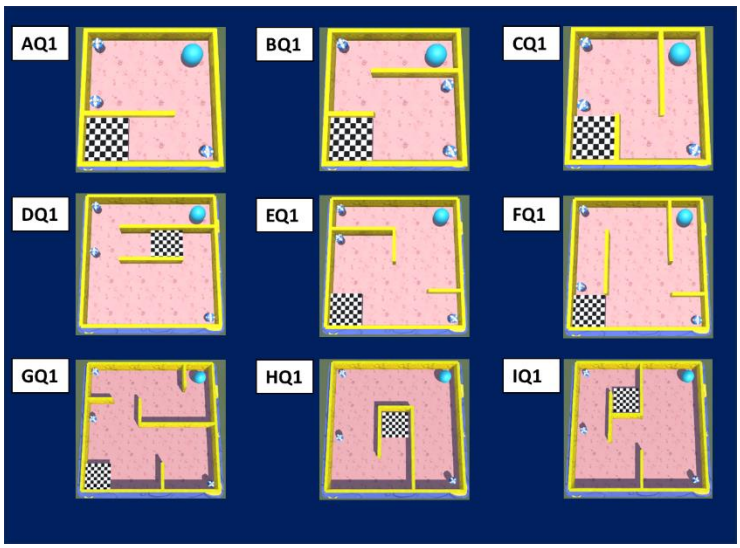


Figure B.1.4 SuraSole maze game level for a sample COP in Quadrant 1

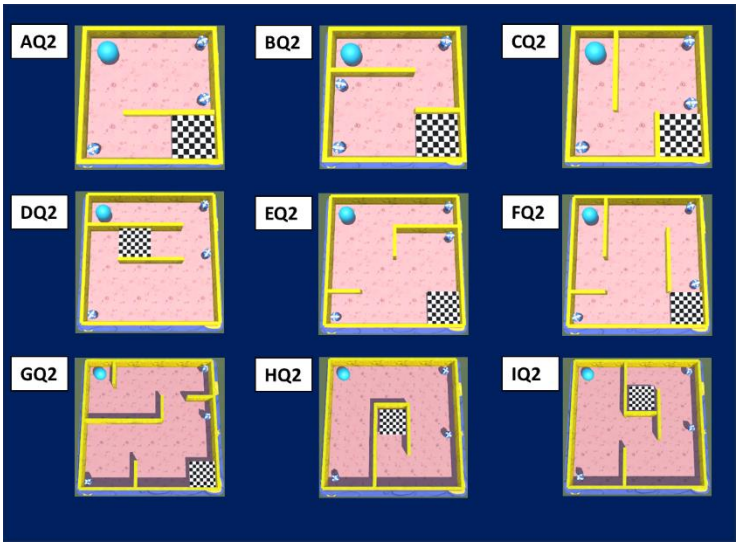


Figure B.1.5 SuraSole maze game level for a sample COP in Quadrant 2

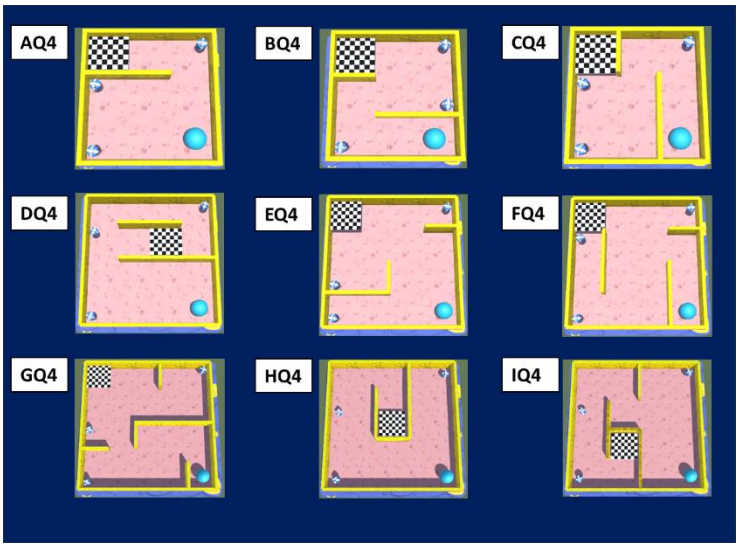


Figure B.1.6 SuraSole maze game level for a sample COP in Quadrant 4

Reward: In this framework, the reward is given as Reward = $\frac{1}{C}$, where C is the distance between the COP coordinate (X, Y) and the origin, calculated using the formula

$$C = \sqrt{X^2 + Y^2} \quad \textbf{(B.1.1)}$$

where, X is the COP value on the x-axis and Y is the COP value on the y-axis.

Note that the closer (X, Y) is to the origin, the better the balance and the larger the reward which depicts the desired outcome of improved balance.

B.2 Q-learning framework for SuraSole maze game

Q-table: The Q-table stores Q-values for every state-action pair. In this case, we have a total of 5 states corresponding to the 5 defined quadrant. Quadrant 1 to 4 has 9 possible actions (levels) whereas Quadrant 5 has 4 possible actions (levels). The learning Rate (α) is set to α = 0.001, The exploration rate is set to 0.8 during training, The discount factor (γ) is set to γ = 0.5.

Exploration and Exploitation: In the early stages of training, the agent will use exploration to try out different actions. With an exploration rate of 0.8, the agent will randomly choose actions 80% of the time to find the best way to proceed, and 20% of the time, the agent uses exploitation, selecting actions with the highest known Q-values at that particular state.

Data Collection and Processing: Actions and their outcomes are recorded and continuously used to update the Q-table. Training data is processed to improve the accuracy of Q-values, which are then used to select appropriate maze levels for each patient. Block diagram as shown in Figure B.2.1.

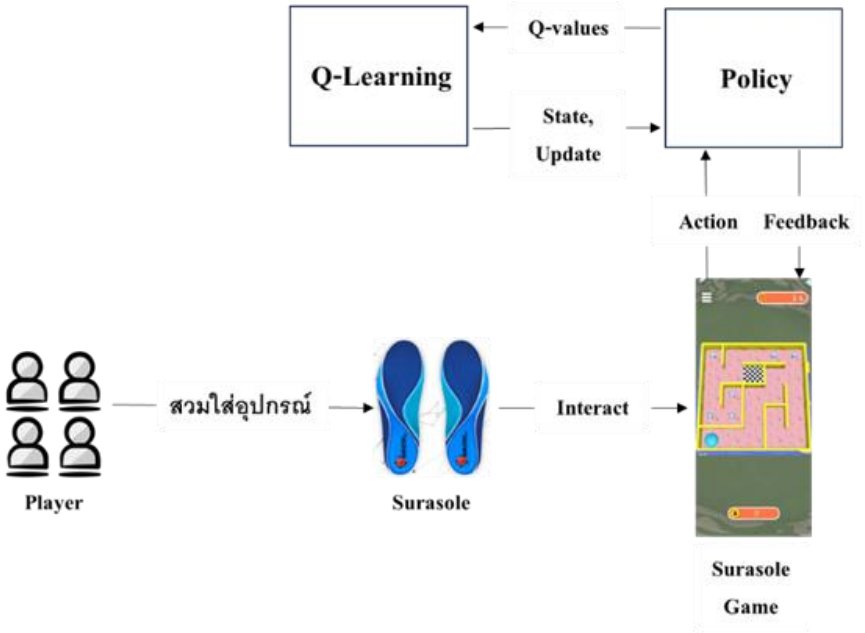


Figure B.2.1 Block diagram of the user interaction with the game

B.3 Model training

Since we initially did not have any patient data to train the model, we simulated data using the state transition diagram technique to simulate the COP transition of patient states.

State Transition Diagram: This diagram is a state-action probability transition model that shows state transitions, consisting of nodes representing different states (Wikipedia, 2024). Here, each node represents a patient's state, and transitions are represented by arrows from one state to another, with each arrow having a probability associated with it, ensuring that the total probability equals 1.

Example of state transitions

When the player is in State 1 or Quadrant 1, state transitions occur as shown in Figure B.3.1.

The state transitions for Quadrants 2, 3, and 4 are similar to those of Quadrant 1.

For Quadrant 5, or the optimal state, state transitions occur as shown in Figure B.3.2.

The goal is to create a game that can select appropriate maze levels for each patient's COP.

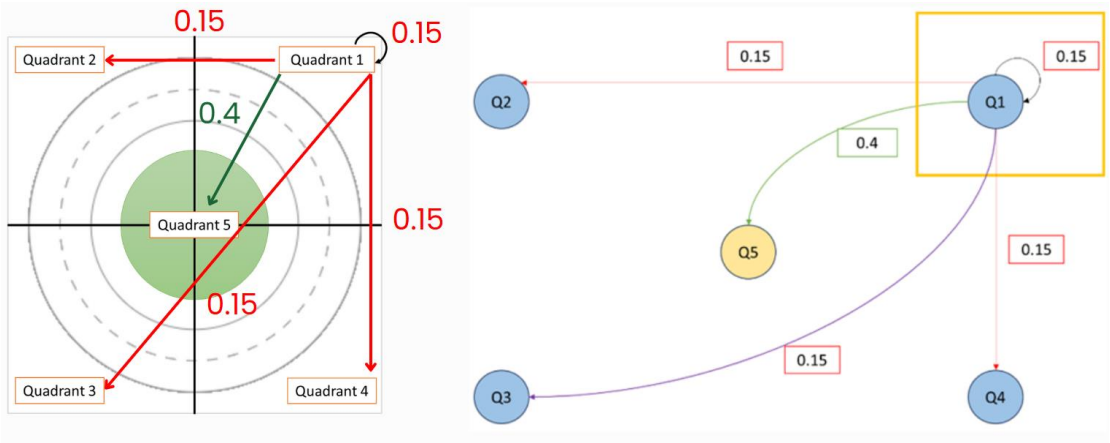


Figure B.3.1 State transitions of Quadrant 1 to various states.

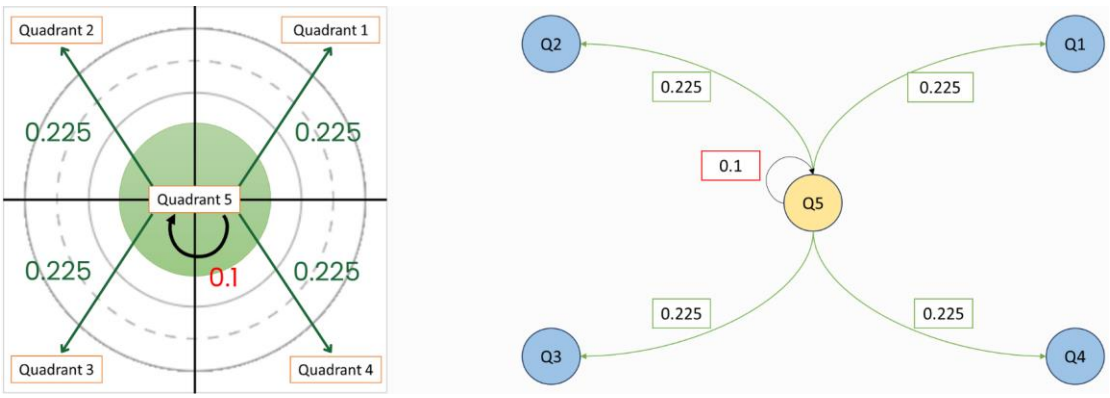


Figure B.3.2 State transitions of Quadrant 5 to various states.

B.4 Model validation

We validated the model's performance by simulating patient COP performance that could potentially help patients reach a favorable state. In this experiment, we assumed that the maze level F is the most suitable level (optimum action) for COPs in Q1, Q2, Q3, and Q4. In particular, level F uses a state transition diagram with a probability of 0.6 for transitioning to a favorable state Q5 and 0.1 for transitioning to other states (Q1, Q2, Q3, and Q4). An example of the state transition diagram for maze level F in the state of Quadrant 1 is shown in Figure B.4.1. For other levels other than level F, the state transitions remain consistent with the original pattern, as shown in Figure B.3.1.

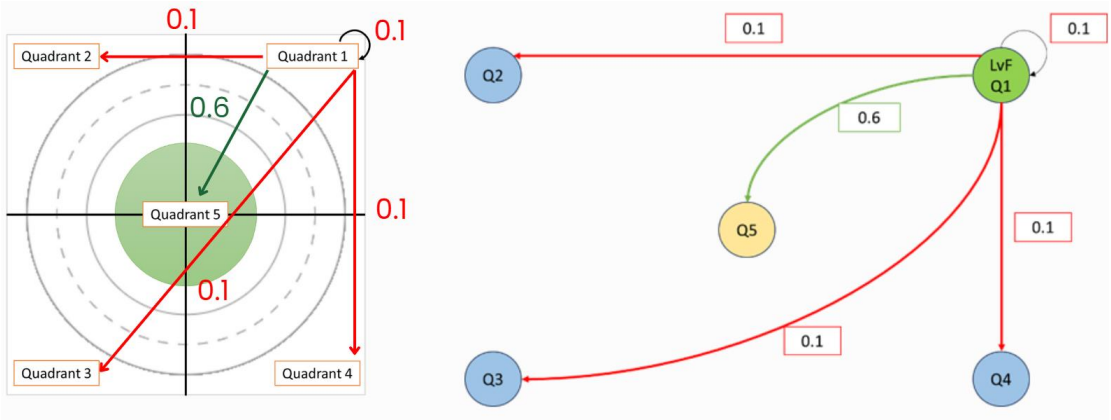


Figure B.4.1 State transitions of maze level F in Quadrant 1 to various states.

To further evaluate the overall performance of the model, we calculated the Average reward per round using Equation (B.4.1), which serves as an indicator of the

model's performance consistency. This Average reward reflects the cumulative success of the model in guiding participants toward favorable states across multiple rounds of testing. The results from the experiment demonstrated that the model could consistently select actions that lead to favorable states i.e. level F, as shown in Figure B.4.2.

$$Avg\_Reward = \frac{\sum_{i=1}^{n} C_i^{-1}}{n} \quad \text{(B.4.1)}$$

where C_i^{-1} is the reciprocal of the distance from the center of the COP circle in test round i, which serves as a measure of postural stability. A smaller distance indicates greater stability, n is the total number of test rounds.

Figure B.4.2 Number of action selections for each state

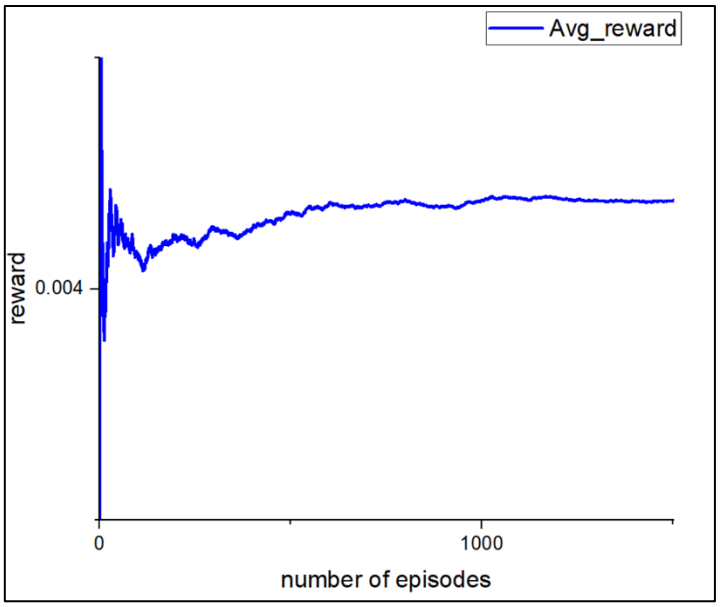


Figure B.4.3 Sample of average reward of the RL model depicting consistent improvement

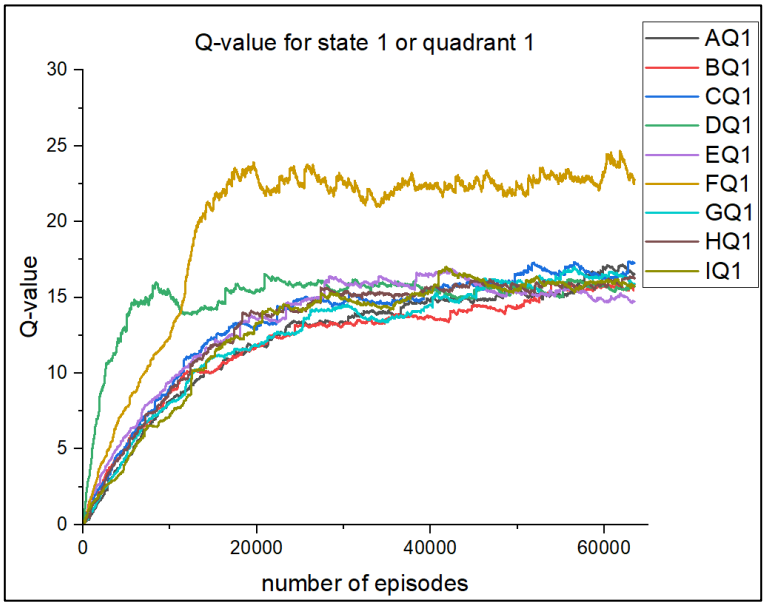


Figure B.4.4 Q-values for state 1 or Quadrant 1

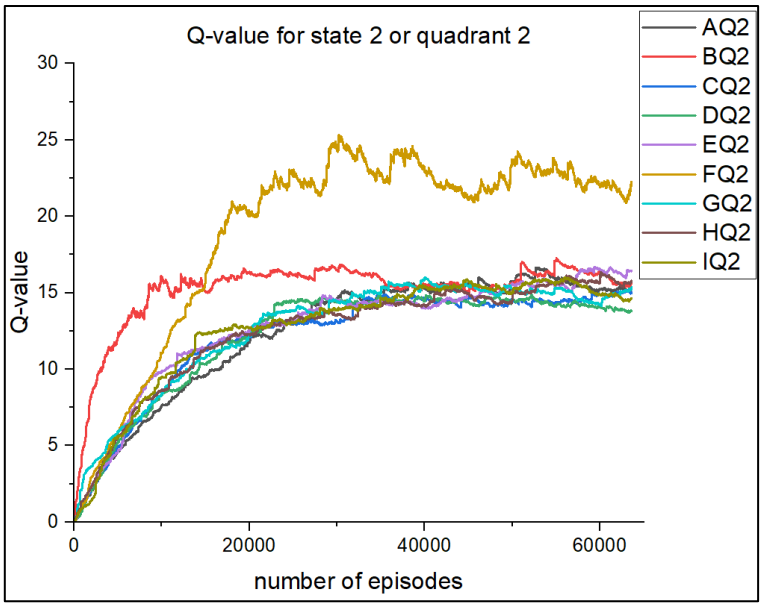


Figure B.4.5 Q-values for state 2 or Quadrant 2

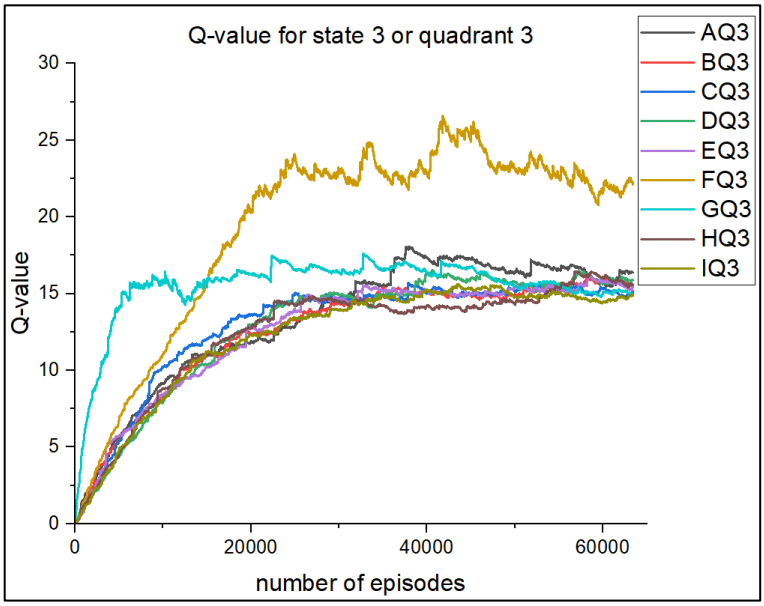


Figure B.4.6 Q-values for state 3 or Quadrant 3

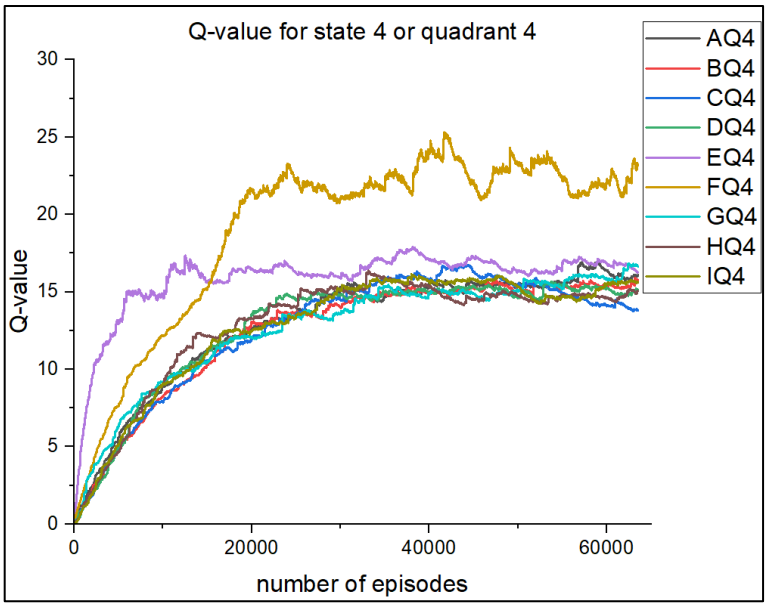


Figure B.4.7 Q-values for state 4 or Quadrant 4

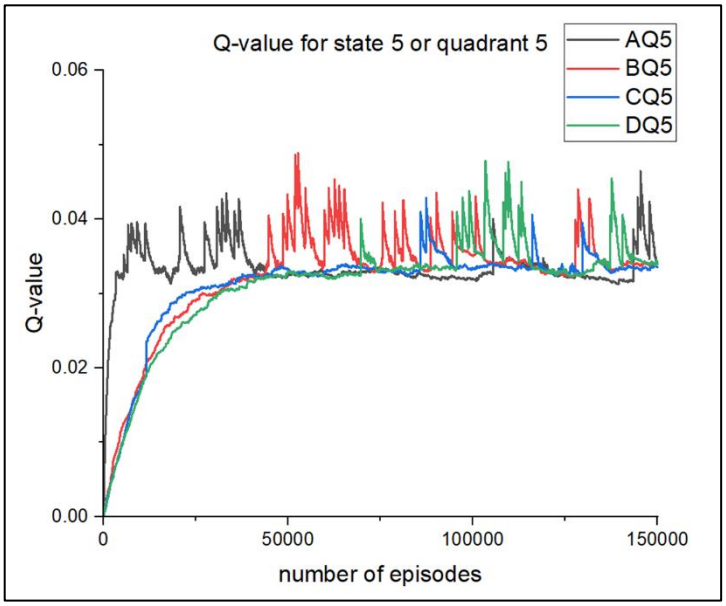


Figure B.4.8 Q-values for state 5 or Quadrant 5

APPENDIX C

Assessment questionnaire

This section compiles information from questionnaires designed to evaluate the opinions and satisfaction of participants across various aspects, including Game safety, Game principles, and the Patient survey. Each section contains questions that address key topics such as the safety of the equipment, the appropriateness of game difficulty levels, the comfort during use, and the motivation to incorporate the game as part of the rehabilitation process. The data gathered from these questionnaires plays a crucial role in analyzing the suitability and effectiveness of the SuraSole maze game, contributing to its further development for more comprehensive and practical applications in the future.

C.1 Patient survey

Patient Survey			
Statement	**Agree**	**Neutral**	**Disagree**
P1) The game was too tiring and did not adjust its difficulty to suit me.			
P2) The exercise (weight shifting) was too easy.			
P3) I would be happy if this game were used more often in my therapy sessions.			
P4) I have no motivation to continue playing.			
P5) The game frustrated me.			
P6) I would play this game at home.			
P7) It was difficult to follow the principles of the game.			
P8) The game was tiring, but manageable.			
P9) I was gradually pushed to my performance limits.			
Remark :			

C.2 Game principles

Game Principles			
Level Creation	Good	Neutral	Needs Improvement
L1) The movement behavior evaluation calculated by the application accurately reflects the patient's balance issues.			
L2) The levels created by the application match the patient's weak directions (not only focusing on the weakest).			
L3) The application adjusts the difficulty level according to the patient's skill progression sufficiently.			
L4) The application allows the physiotherapist to progressively customize the patient's level.			
Game Principles	Good	Neutral	Needs Improvement
S1) The difficulty curve of the game does not exceed the patient's capabilities.			
S2) The patient's progression is clearly visible during the game.			
S3) There is no noticeable delay between weight shifts on the SuraSole and the movements in the game.			
S4) Weight-shifting training with SuraSole is an effective practice for balance rehabilitation.			
S5) I can imagine this game being an additional tool for balance rehabilitation in the future.			
Remark :			

C.3 Game safety

Game Safety

Please rate each statement from 1 (Strongly Disagree) to 3 (Strongly Agree).

1. The player feels confident that playing this game is safe and does not pose any harm to the user.
   1 ☐ 2 ☐ 3 ☐
2. During gameplay, the player does not feel at risk of accidents or falling.
   1 ☐ 2 ☐ 3 ☐
3. The difficulty level of the game does not create any risk of harm to the player.
   1 ☐ 2 ☐ 3 ☐
4. The player feels that using the equipment (Surasole Insoles) is safe, does not pose any harm, and does not cause discomfort.
   1 ☐ 2 ☐ 3 ☐
5. The player thinks this game is suitable for individuals with balance or mobility issues.
   1 ☐ 2 ☐ 3 ☐
6. The player feels that the game does not cause any pain or discomfort during gameplay.
   1 ☐ 2 ☐ 3 ☐
7. The difficulty level of the game is not overly pressuring to the point of causing injury to the player.
   1 ☐ 2 ☐ 3 ☐
8. The sensor system and equipment connection function safely without causing any harm to the player.
   1 ☐ 2 ☐ 3 ☐
9. I feel that the game can adjust its difficulty without negatively impacting the player's physical condition.
   1 ☐ 2 ☐ 3 ☐
10. Overall, the player feels that this game is safe to use, does not pose any harm, and is suitable for rehabilitation purposes.
    1 ☐ 2 ☐ 3 ☐