future internet

MDPI

Article

ReHAb Playground: A DL-Based Framework for Game-Based Hand Rehabilitation

Samuele Rasetto [1], Giorgia Marullo [2], Ludovica Adamo [1], Federico Bordin [1], Francesca Pavesi [1], Chiara Innocente [2], Enrico Vezzetti [2] and Luca Ulrich [2,*]

[1] Biomedical Engineering, Politecnico di Torino, C.so Duca degli Abruzzi, 24, 10129 Torino, Italy; s333832@studenti.polito.it (S.R.); s328576@studenti.polito.it (L.A.); s337573@studenti.polito.it (F.B.); s330729@studenti.polito.it (F.P.)

[2] Management and Production Engineering, Politecnico di Torino, C.so Duca degli Abruzzi, 24, 10129 Torino, Italy; giorgia.marullo@polito.it (G.M.); chiara.innocente@polito.it (C.I.); enrico.vezzetti@polito.it (E.V.)

* Correspondence: luca.ulrich@polito.it

Abstract

Hand rehabilitation requires consistent, repetitive exercises that can often reduce patient motivation, especially in home-based therapy. This study introduces ReHAb Playground, a deep learning-based system that merges real-time gesture recognition with 3D hand tracking to create an engaging and adaptable rehabilitation experience built in the Unity Game Engine. The system utilizes a YOLOv10n model for hand gesture classification and MediaPipe Hands for 3D hand landmark extraction. Three mini-games were developed to target specific motor and cognitive functions: Cube Grab, Coin Collection, and Simon Says. Key gameplay parameters, namely repetitions, time limits, and gestures, can be tuned according to therapeutic protocols. Experiments with healthy participants were conducted to establish reference performance ranges based on average completion times and standard deviations. The results showed a consistent decrease in both task completion and gesture times across trials, indicating learning effects and improved control of gesture-based interactions. The most pronounced improvement was observed in the more complex Coin Collection task, confirming the system's ability to support skill acquisition and engagement in rehabilitation-oriented activities. ReHAb Playground was conceived with modularity and scalability at its core, enabling the seamless integration of additional exercises, gesture libraries, and adaptive difficulty mechanisms. While preliminary, the findings highlight its promise as an accessible, low-cost rehabilitation platform suitable for home use, capable of monitoring motor progress over time and enhancing patient adherence through engaging, game-based interactions. Future developments will focus on clinical validation with patient populations and the implementation of adaptive feedback strategies to further personalize the rehabilitation process.

Keywords: gesture recognition; hand rehabilitation; home-based rehabilitation; serious game

check for updates

Academic Editors: Shun Zhang, Feng Gao and Mingyang Ma

Received: 14 October 2025
Revised: 5 November 2025
Accepted: 14 November 2025
Published: 17 November 2025

Citation: Rasetto, S.; Marullo, G.; Adamo, L.; Bordin, F.; Pavesi, F.; Innocente, C.; Vezzetti, E.; Ulrich, L. ReHAb Playground: A DL-Based Framework for Game-Based Hand Rehabilitation. *Future Internet* **2025**, *17*, 522. https://doi.org/10.3390/fi17110522

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The purpose of exercise rehabilitation is not only to restore physical function after surgery, such as in muscles, joints, or conditions related to chronic pain, fatigue, neurological, or metabolic disorders, but also to support the recovery of psychological well-being. Although traditional rehabilitation is effective, exercises can be mundane and boring due

to their repetitive nature. Furthermore, the fatigue and pain during the traditional exercise rehabilitation process can also exacerbate the player's negative psychology [1].

Stroke is the leading cause of long-term disability worldwide [2], and upper-limb paresis is one of the most persistent impairments following a stroke. Most stroke survivors have impaired upper-limb motor function and body balance problems because one side of the body is stronger than the other [3,4]. Rehabilitation of the impaired limb and body balance is essential to help stroke survivors accelerate the recovery process and regain their independence. Only up to 34% of stroke patients fully recover upper-limb function, which appears to be necessary for using the affected arm in daily tasks [5].

Rehabilitation usually begins in the hospital after an injury. As therapy progresses, patients are often required to travel to specialized outpatient units for supervised treatment. Eventually, home-based programs, sometimes supported by visiting professionals, allow patients to continue developing their skills in a familiar environment, while also reducing travel costs to clinics or hospitals. This approach is especially beneficial for patients recovering from hand-related conditions such as stroke, arthritis, and tendinopathies. These disorders often impair fine motor skills and require repetitive, targeted exercises to restore function [1]. However, maintaining engagement and ensuring the intensity and quality of rehabilitation outside clinical environments remain significant challenges [6].

In this context, serious games, namely video games specifically designed with therapeutic goals while remaining engaging and enjoyable [7], have emerged as a promising approach, particularly in the rehabilitation of upper limbs [8–10]. By incorporating key game design principles such as real-time feedback, rewarding incentives, adaptive difficulty, and goal-oriented tasks, serious games can significantly enhance user motivation and engagement [11]. Real-time feedback allows players to track their performance and skill progression, while adaptive difficulty adjusts the challenge level based on in-game performance and therapeutic constraints, ensuring both safety and personalization. The user interface should remain intuitive and easy to navigate so that players can focus on gameplay rather than on learning how to use the system [12]. Moreover, continuously informing patients about their progress toward therapeutic goals helps sustain motivation throughout the rehabilitation process [13].

Games designed for upper-limb motor rehabilitation typically require users to point to, reach for, touch, grasp, and move game elements. The positions of these elements can be adjusted to make them easier or more challenging to reach, depending on the player's capabilities [6].

This study aims to develop ReHAb Playground, a non-immersive virtual reality (VR) serious game for hand rehabilitation in individuals with impaired upper-limb mobility due to illness or trauma. The primary objective of the study is to develop an artificial intelligence-based solution to support hand rehabilitation, with the broader goal of reducing recovery times and providing more effective and accessible care for patients. Based on a benchmarking analysis of existing methods, the system combines real-time gesture recognition with 3D hand tracking to provide an engaging and adaptable home-based therapy experience. Using a YOLOv10n [14] model for gesture classification and MediaPipe Hands [15] for 3D landmark extraction, the platform precisely monitors hand movements. The research question guiding this study is whether an AI-driven system combining real-time gesture recognition and 3D hand tracking can effectively support home-based hand rehabilitation by ensuring sufficient accuracy, responsiveness, and adaptability to therapeutic needs. For this purpose, three mini-games targeting motor and cognitive functions were implemented, with tunable parameters such as repetitions, time limits, and gestures to match therapeutic protocols. Designed for modularity and scalability, the system can integrate new exercises, gesture sets, and adaptive difficulty mechanisms to

support ongoing growth and evolution. Preliminary tests with healthy participants showed learning effects, improved gesture control, and enhanced engagement, demonstrating the system's potential as a low-cost, home-based rehabilitation tool. The main contributions of this paper can be summarized as follows: (1) evaluation of the accuracy and responsiveness of gesture recognition in real-time interactions, (2) establishment of baseline performance ranges in healthy participants, and (3) examination to determine if there is learning rate improvement among task repetitions.

The article is divided into sections: Section 2 discusses state-of-the-art rehabilitation methods supported by technological tools and hand gesture approaches; Section 3 describes the system developed for the serious game, including chosen gestures, deep learning architectures, and mini-games; Section 4 presents and discuss the obtained results; Section 5 reports the study's conclusions.

2. Related Works

The development of serious games for hand rehabilitation has evolved rapidly over the past decade, driven by the need to make therapy more engaging, adaptive, and accessible outside clinical environments. Early approaches focused on sensor-based systems capable of translating natural movements into interactive exercises, while recent advances in computer vision and lightweight deep learning models have enabled markerless, AI-assisted solutions that enhance precision and user experience. This evolution reflects a continuous effort to bridge clinical effectiveness with technological usability.

One of the early milestones in this field was the StableHand VR game proposed by Pereira et al. [16], which exploited the Meta Quest 2 VR headset to create an immersive rehabilitation environment. The system leveraged Unity SDK (v12) and its built-in hand tracking API, enabling users to manipulate virtual elements through sequential open-hand and multi-finger pinch gestures. This approach demonstrated how VR could simulate motor exercises while maintaining patient engagement. However, the cost of VR headsets (approximately USD 300–500) and the potential for motion sickness or disorientation in elderly or neurologically impaired users may limit their accessibility for home-based rehabilitation programs [17,18]. Building upon this foundation, Bressler et al. [19] enhanced the therapeutic dimension by introducing a narrative-driven farm scenario that combined adaptive visual feedback with dynamic difficulty adjustment. By calibrating the range of motion and tracking daily progress, the system provided personalized motivation and continuous challenge.

Parallel to VR-based systems, several studies have explored motion capture using affordable sensor technologies. Trombetta et al. [20] introduced Motion Rehab AVE 3D, an exergame for upper-limb motor and balance recovery after mild stroke, utilizing Kinect sensors. Kinect-based platforms provide robust full-body tracking suitable for gross motor rehabilitation and balance training, but their depth sensors are less accurate for fine finger movements and demand substantial computational resources. The system delivered real-time feedback through a "virtual mirror", allowing patients to visualize and correct their movements [21], building upon Fiorin et al. [22], who initially developed a 2D rehabilitation game later expanded into a 3D Unity environment with customizable avatars and difficulty levels. Similarly, Elnaggar et al. [23] developed a web-based rehabilitation platform using the Leap Motion controller, a low-cost, markerless device capable of detecting hand and finger movements with high precision. Leap Motion-based systems offer sub-millimeter accuracy at close range, making them particularly effective for fine motor control exercises; however, their limited field of view, sensitivity to lighting conditions, and cost (approximately USD 100–150) restrict their applicability in uncontrolled home environments and resource-limited settings. Their iterative design process involved therapists and patients to

ensure clinical relevance, resulting in gamified exercises modeled on familiar entertainment mechanics. By integrating customizable thresholds and performance feedback, the system addressed both motivational and accessibility challenges inherent to conventional hand therapy.

With the growing maturity of deep learning (DL), research attention shifted from sensor-based setups to computer-vision-driven interaction. Convolutional and graph convolutional networks (CNNs and GCNs) have proven particularly effective for real-time gesture recognition [24,25], facilitating markerless and portable rehabilitation systems. Chen et al. [26] proposed a lightweight residual GCN that processed 3D hand-joint coordinates for gesture classification within an interactive "air-writing" rehabilitation game, where users control an "invisible pen" to draw characters, receiving real-time feedback. The model achieved over 99% accuracy while remaining compact enough for low-resource devices, demonstrating the feasibility of AI-assisted rehabilitation in home environments.

More recently, Husna et al. [27] investigated gesture intuitiveness in rehabilitation exergames using MediaPipe v0.10.21 for real-time hand tracking. Their study evaluated ten gestures captured via webcam and assessed usability with sixteen participants using the system usability scale (SUS) and user experience questionnaire (UEQ). Results identified a subset of gestures offering superior accessibility, comfort, and user experience. In a related study, Xiao et al. [28] developed a web-based exergame employing MediaPipe for tenosynovitis prevention, implementing three gesture sets tested among students. The system demonstrated positive usability scores and measurable improvements in wrist flexibility.

Additional developments include systems for wrist motion classification using mobile sensors to support adaptive exergames for children with motor impairments [29] and augmented reality solutions that integrate real-time kinematic assessment via HoloLens 2 for upper-limb rehabilitation tasks [30].

Together, these studies illustrate the ongoing transition from sensor-dependent rehabilitation platforms to markerless, vision-based systems that involve artificial intelligence for accurate hand tracking. The current trend points toward hybrid architectures that integrate gesture recognition, adaptive feedback, and engaging interaction design within accessible, low-cost environments suitable for daily therapeutic use.

ReHAb Playground involves standard RGB webcams ubiquitously available in laptops and affordable external cameras, combined with MediaPipe Hands for landmark detection and YOLOv10n for gesture classification. This approach positions our system as a complementary solution that prioritizes accessibility and scalability over precision, accepting the trade-off of MediaPipe's relative depth estimation limitations in exchange for zero hardware barriers beyond equipment most users already possess.

3. Materials and Methods

The serious game presented in this study enables hand rehabilitation through the execution of gestures during interaction with a Unity-based environment. The platform includes a set of mini-games, namely Cube Grab, Coin Collection, and Simon Says, that allow the patient to achieve micro-objectives aimed at improving their clinical condition. Figure 1 illustrates the workflow of the entire process, which is executed cyclically until the patient reaches an optimal functional state.

The process begins with a meeting between the physician and the patient to assess the patient's baseline condition and define an appropriate therapeutic plan. Next, the game developer sets the therapy parameters discussed during the meeting, such as the number of repetitions, gesture mappings, and time limits for completing each mini-game. The patient is then instructed to run a Python script that includes two neural networks: one for gesture recognition (Section 3.1) and another one for hand marker tracking (Section 3.2).

The resulting data are transmitted to Unity via two separate sockets. The patient proceeds with their training session within the game environment. This cycle is repeated periodically according to the prescribed therapy to gradually improve the patient's performance. The game also provides feedback to support continuous improvement over time.

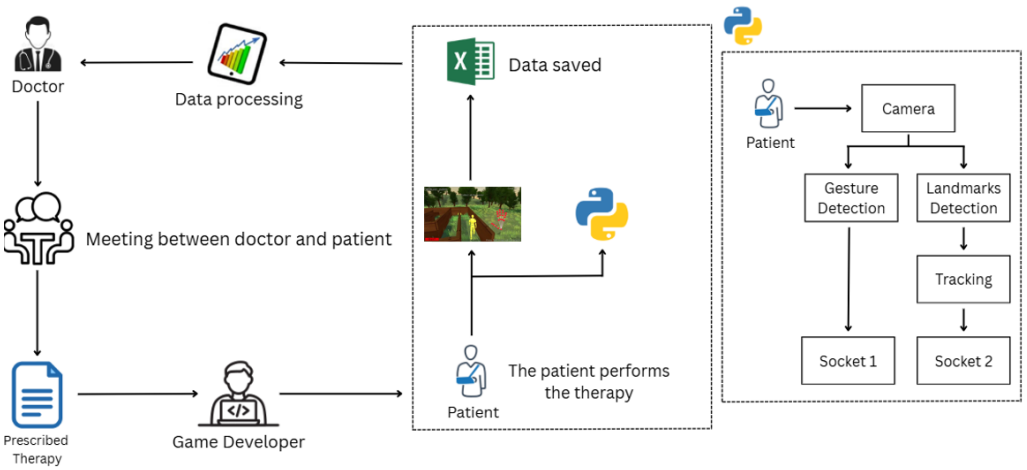

Figure 1. Process overview of the ReHAb Playground workflow. The figure illustrates the interaction between the doctor, patient, and game developer throughout the rehabilitation process. After defining the personalized therapy plan, the developer encodes the parameters into the game. The patient performs the exercises using a Python-based system that detects gestures and tracks hand landmarks in real time, while performance data are recorded and later analyzed by the clinician to monitor progress and adjust therapy.

All the chosen and analyzed gestures, as shown in Figures 2 and 3, play a significant role in hand rehabilitation by improving joint mobility in both the fingers and wrist, helping to prevent stiffness and maintain an optimal range of motion. These movements also enhance coordination and dexterity, which are essential for performing fine motor tasks such as writing, typing, and grasping objects. Furthermore, they promote fine motor control by encouraging the isolation of individual or grouped finger movements.

Hand Gesture						
Therapy	Fingers Abduction	Fingers Adduction	Clenched Fist	Fingers Selective Extension/ Flexion	Fingers Selective Extension/ Flexion	Fingers Selective Extension/ Flexion Wrist Rotation
Map Movement	Stop	Move Forward	Move Backward	Jump	Turn Right	Turn Left
Body Part	Fingers	Fingers	Fingers	Fingers	Fingers	Fingers, Wrist

Figure 2. Modeling and mapping of physical therapy gestures used for map movement. The figure illustrates the set of hand gestures employed to control avatar navigation within the game environment. Each gesture corresponds to a specific therapeutic movement, such as finger abduction, adduction, or wrist rotation, designed to enhance motor control and joint mobility while providing intuitive in-game actions.

Gesture selection is guided by evidence-based rehabilitation principles to ensure that each movement corresponded to specific therapeutic goals. Finger abduction and adduction activate the intrinsic hand muscles, particularly the interossei and lumbricals, responsible for grip stability and fine motor coordination, and are frequently included in hand therapy protocols [31,32]. Fist clenching primarily engages the finger flexors, including the flexor digitorum profundus and superficialis, and supports recovery of power grip strength

essential for object manipulation and daily activities [33,34]. Finger flexion and extension promote activation of the extensor digitorum muscles, whose recovery is strongly associated with improved upper-limb outcomes after neurological injury [35,36]. Finally, wrist rotation tasks enhance pronation and supination range of motion, promoting forearm mobility and functional reach, which are essential for performing everyday tasks such as turning or grasping objects [37,38]. Altogether, these gestures reflect evidence-based exercises widely applied in clinical rehabilitation, ensuring that gameplay not only fosters engagement and motivation but also supports meaningful motor recovery and functional improvement of the upper limb.

Hand Gesture					
Therapy	Selective Fingers Extension/ Flexion	Selective Fingers Extension/ Flexion	Index-Thumb Plier	Fingers Abduction	Fingers Flexion
Minigame	Simon Says	Simon Says	Simon Says	Simon Says	Grab The Cube/ Simon Says
Body Part	Fingers	Fingers	Fingers	Fingers	Fingers

Figure 3. Modeling and mapping of physical therapy gestures related to the specific mini-games. This figure shows the gestures integrated into the mini-games, each associated with targeted rehabilitation goals. The selected movements, such as finger extension, flexion, and thumb–index pinching, support fine motor recovery and dexterity improvement through interactive gameplay.

The Hand Gesture Recognition, Hand Landmarks Tracking pipelines, and the mini-games will be deepened in Sections 3.1, 3.2, and 3.3, respectively.

3.1. Hand Gesture Recognition

YOLO (you only look once) is a family of real-time object detection models that perform detection by framing it as a single regression problem, directly predicting bounding boxes and class probabilities from full images in one evaluation [39]. Unlike traditional two-stage detectors (e.g., R-CNN), YOLO networks are single-stage detectors that offer a favorable trade-off between speed and accuracy. YOLOv10 [14] introduces an end-to-end design that removes the need for non-maximum suppression (NMS) and incorporates spatial–channel decoupled down-sampling with a simplified detection head, improving inference speed and computational efficiency while maintaining detection accuracy. These optimizations allow YOLOv10 to achieve higher mean average precision (mAP) and lower latency than previous YOLO versions, such as YOLOv5 and YOLOv8, across multiple benchmarks, confirming its suitability for real-time applications and resource-constrained environments [40–42].

To identify hand gestures, a YOLOv10n neural network was employed, pre-trained on the HaGRIDv2 dataset (https://github.com/hukenovs/hagrid, accessed on 27 May 2025). HaGRIDv2 is a large-scale dataset comprising over 1,000,000 RGB images collected from approximately 66,000 unique individuals, encompassing a wide range of demographic variability. The dataset includes annotations for 33 distinct hand gestures, 7 of which involve the use of both hands. According to official benchmarks, the YOLOv10n model achieves a mean average precision (mAP) of 88.2 on this dataset, indicating robust performance in gesture classification tasks. After testing various lighting conditions, the optimal detection range from the camera was determined to be between 0.4 and 1.5 m.

The "n" variant, YOLOv10n model, for gesture detection was empirically evaluated on a custom dataset comprising 140 manually annotated images, with 14 samples per gesture class. The evaluation was restricted to gesture classification performance, yielding

an excellent F1-score of 0.993. For comparison, the larger YOLOv10x model was also assessed, achieving the same performance. Consequently, the nano variant was selected for subsequent development, as it offers substantially higher inference speed while preserving prediction accuracy, thus rendering it more suitable for real-time applications. Table 1 shows the results after testing.

Table 1. YOLOv10 model comparison.

Model	F1-Score	Mean fps	Max fps
YOLOv10n	0.993 ± 0.014	12.59 ± 5.94	29.72
YOLOv10x	0.993 ± 0.014	7.05 ± 3.98	12.82

3.2. Hand Landmark Tracking

The MediaPipe Hands framework [15] is an open-source, cross-platform framework developed by Google, offering fast and robust hand tracking through the detection of 21 3D landmarks per hand, as shown in Figure 4, using a single RGB camera (Intel RealSense SR305, Intel Corporation, Santa Clara, USA).

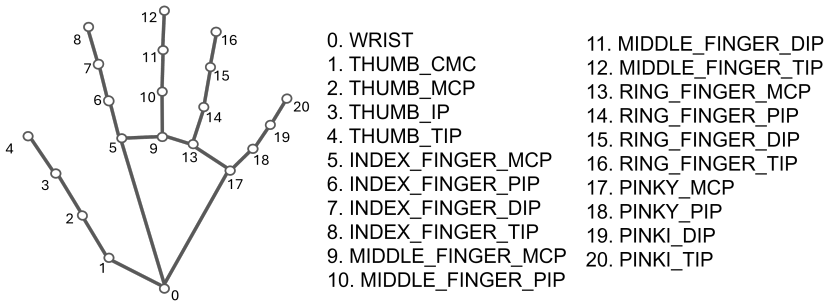

Figure 4. MediaPipe Hands' landmarks positioning. Illustration of the 21 three-dimensional landmarks identified by MediaPipe on the right hand. These keypoints are used for gesture recognition and motion tracking, enabling real-time interaction within the rehabilitation game environment.

Mediapipe Hands was selected for hand landmark tracking due to its high accuracy, real-time performance, and robustness. Indeed, it achieves 95.7% palm detection accuracy at 120+ FPS on desktop GPUs, with a mean 3D pose error of 1.3 cm [43]. Compared with Meta Quest and Leap Motion, it maintains low error at close range (12.4–21.3 mm) and reliable tracking at distances up to 2.75 m (4.7 cm median error) [44]. Robustness evaluations report 95% recall and 98% precision, with superior performance under challenging conditions relative to other frameworks [45], and joint-angle errors typically range from 9–15° (2–3 mm spatial error), reducible through multi-camera triangulation [46]. Validation against motion-capture standards further confirms consistent temporal tracking and improved spatial accuracy with depth-enhanced methods [47]. Recent comparative studies highlight that MediaPipe Hands achieves higher landmark localization accuracy and lower latency than lightweight full-body pose estimators, such as BlazePose [48] and OpenPose [49], while maintaining lower computational cost and on-device efficiency. These features establish MediaPipe Hands as a reliable and efficient framework for precise hand tracking across a range of experimental conditions.

To provide real-time feedback to the user during gameplay, a live 3D visualization of the hand landmarks was integrated into the game's visuals. This was accomplished by extending the Python script responsible for gesture detection with hand landmark extraction. The coordinates of Mediapipe Hands landmarks are transmitted to Unity, where each is rendered as a red sphere connected to the others with a white line within the 3D

environment. To determine the most suitable depth estimation strategy, three different approaches were evaluated:

1. MediaPipe: This method offers the most straightforward implementation, as MediaPipe directly provides a relative Z-coordinate for each landmark. However, its accuracy is limited. MediaPipe estimates depth using a technique known as scaled orthographic projection, which combines orthographic projection with depth scaling to approximate perspective. This approach assumes a uniform distance from the camera for all points and typically uses the wrist as a reference to compute relative depth values. While convenient, it does not yield true 3D spatial positioning, remains sensitive to occlusions and variations in camera pose, is not calibrated, and does not correspond to real-world units like centimeters or meters.
2. RGB-D Camera: As an alternative, an Intel RealSense SR305 RGB-D camera [50,51] was integrated into the system. RGB-D cameras provide synchronized RGB images and depth maps by actively projecting infrared patterns and analyzing their distortion, employing either structured light or time-of-flight techniques. This enables more accurate and absolute depth measurements. Although this method significantly enhances depth reliability, RGB-D cameras are not commonly available in standard home environments, making them less suitable for at-home rehabilitation scenarios.
3. Depth Prediction Models: The third approach involves estimating depth from monocular RGB images using a pre-trained neural model, such as Depth-Anything V2 (https://github.com/DepthAnything/Depth-Anything-V2, (accessed on 13 November 2025)). These models generate dense depth maps by leveraging large-scale training on diverse datasets. Although the resulting depth estimates surpass MediaPipe's heuristic values in accuracy, they remain less precise than measurements from RGB-D sensors. Furthermore, the high computational load associated with these models may hinder real-time performance, particularly on low-power devices.

To quantitatively compare the three depth estimation approaches, their performance was evaluated in terms of frame rate, latency, and depth accuracy (Table 2). Tests were conducted using 30 hand acquisitions at known distances (0.25 m, 0.5 m, and 0.75 m), comparing the detected depth at the index fingertip landmark. Results indicate that the Intel RealSense camera achieved the best overall performance, with the lowest latency (7.48 ± 4.80 ms), highest frame rate (31.05 ± 3.82 FPS), and most accurate depth estimation (0.014 ± 0.009 m). MediaPipe showed comparable frame rates (29.30 ± 3.99 FPS) and latency (7.86 ± 5.31 ms), with only a modest reduction in depth accuracy (0.023 ± 0.012 m), sufficient to maintain fluid and responsive gameplay. Conversely, DepthAnything Light exhibited lower frame rates (12.67 ± 9.24 FPS), a slightly higher latency (9.91 ± 6.04 ms), and reduced depth precision (0.113 ± 0.038 m), confirming its lower efficiency for real-time rehabilitation tasks. Although the RealSense solution may provide the most accurate and responsive depth estimation, its reliance on an external RGB-D camera limits practicality in home environments, where cost and setup complexity are major considerations. For this reason, a hybrid strategy was adopted, combining the technical advantages of RealSense with the accessibility of MediaPipe. The RealSense-based implementation is intended for clinical or laboratory settings, where precise and absolute depth data are needed for detailed motion analysis. In contrast, the MediaPipe-based configuration enables at-home rehabilitation using only a standard webcam, providing adequate responsiveness and gesture fidelity without additional hardware.

Table 2. Quantitative comparison of the three evaluated depth estimation models: Intel RealSense, MediaPipe, and DepthAnything Light. Metrics include average frame rate (FPS, frames per second), maximum frame rate, end-to-end latency (in milliseconds), and depth accuracy (in meters). Latency was measured as the time between the transmission of a gesture from Python and the corresponding action executed in Unity. Depth accuracy was computed as the mean absolute difference between the detected and true fingertip depth across 30 acquisitions at known distances (0.25 m, 0.5 m, and 0.75 m). Values are expressed as mean ± SD, where SD denotes standard deviation.

Model	FPS (Mean ± SD)	Max FPS	Latency [ms] (±SD)	Depth Accuracy [m] (±SD)
RealSense	31.05 ± 3.82	35.14	7.48 ± 4.80	0.014 ± 0.009
MediaPipe	29.30 ± 3.99	32.08	7.86 ± 5.31	0.023 ± 0.012
DepthAnything Light	12.67 ± 9.24	28.73	9.91 ± 6.04	0.113 ± 0.038

A visual comparison between MediaPipe and RealSense estimated depth is presented in Figure 5. As shown, the depth values provided by MediaPipe can be negative, indicating that the analyzed landmark is positioned closer to the camera compared to the wrist landmark. This difference arises because MediaPipe outputs relative distances with respect to the wrist, whereas RealSense provides absolute distances from the camera. This dual configuration enables the serious game to function effectively across both professional and domestic environments, promoting accessibility and scalability without compromising core functionality.

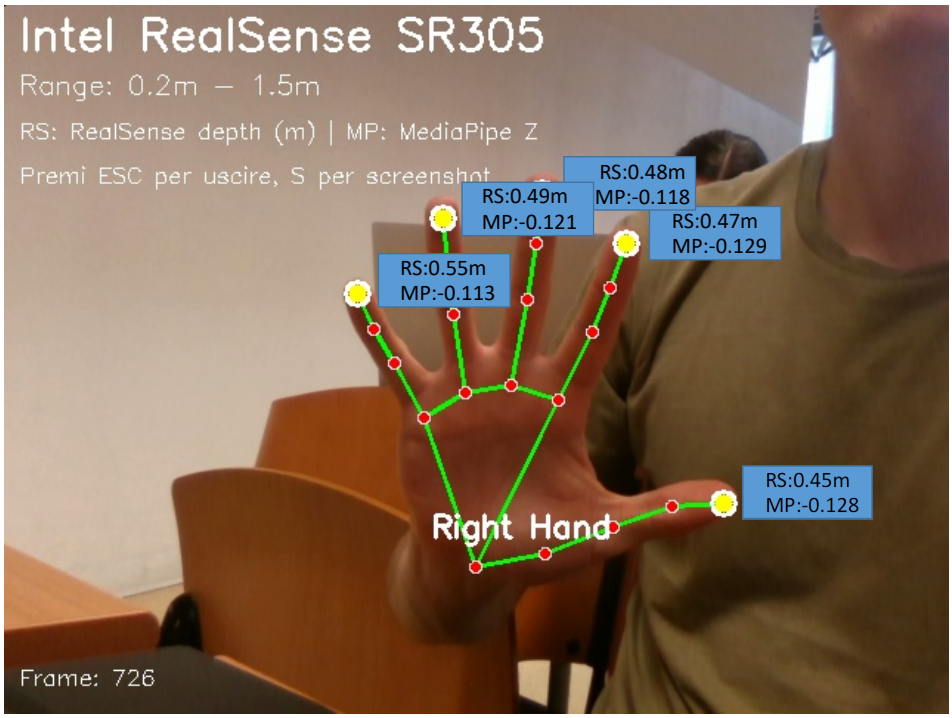

Figure 5. Comparison between MediaPipe and RealSense estimated depth. Example of simultaneous depth estimation using MediaPipe (MP) and Intel RealSense SR305 (RS). MediaPipe provides relative Z-values referenced to the wrist, while RealSense outputs absolute distances from the camera. This comparison highlights the trade-off between accuracy and accessibility in the hybrid configuration adopted for home and clinical use.

3.3. Mini-Games

Within the Unity environment, an avatar can move and reach three different stations, each associated with a specific mini-game. Each game is subject to a configurable time limit within which the target objective must be completed. If the objective is not achieved

within the allotted time, the game is considered complete, and the patient's performance is recorded. This time limit is adjustable through the Unity interface and can vary between games, allowing clinicians to tailor the difficulty according to the therapeutic protocol. The time constraint also ensures that the patient can exit the game in case of inability to complete the task.

A modular approach was adopted for the software architecture: all mini-game scripts communicate with a central manager that regulates the start and end of each game. Minigame parameters are customizable to adapt the therapy to the individual needs of the patient. Additionally, visual feedback is provided on screen to enhance user engagement and track progress, including indicators such as the number of repetitions performed, elapsed time, and the number of coins collected.

To provide a structured overview of the game design, Table 3 summarizes the main features of each mini-game, including their primary and secondary goals, the performance metrics collected during gameplay, and the progression rules governing task adaptation. Specifically, Grab the Cube focuses on finger flexion and extension, simulating grip strengthening exercises. Coin Collection emphasizes wrist mobility and spatial control while engaging visiomotor coordination. Finally, Simon Says integrates cognitive and motor tasks through memory-based gesture sequences. The progression rules were designed to incrementally increase task difficulty by adjusting parameters such as the number of repetitions, completion time, and presence of distractors, ensuring a gradual challenge aligned with user improvement.

Table 3. Overview of mini-games, goals, outcomes, and progression rules. # refers to the number of occurrences.

Mini-Game	Primary Goal	Secondary Goal	Measured Outcomes	Progression Rules
Grab the Cube	Finger flexion and extension	Grip strength simulation	Hold time (s), repetitions completed, gesture stability (SD)	Increase repetitions (5→10→15); Reduce hold threshold (2 s→1.5 s→1 s)
Coin Collection	Wrist mobility and spatial control	Hand–eye coordination	Completion time (s), coins collected (#), collision events (#)	Increase coin count (20→30→40); Narrow maze width; Add time pressure
Simon Says	Cognitive-motor integration	Gesture memory and sequencing	Sequence length (#), transition time (s), error rate (%)	Increase sequence (3→5→7); Reduce display time; Add distractors

In Grab the Cube (Figure 6), the avatar is placed in front of a virtual cube positioned within the interactive environment. Once the avatar approaches the cube, the patient is prompted to perform a specific hand gesture to grasp it. This gesture involves the coordinated movement of fingers and hand muscles to simulate a realistic gripping action. The grabbing motion must be repeated a configurable number of times, according to the parameters set by the clinician, who can adjust the difficulty and intensity of the task based on the patient's rehabilitation needs. Therapeutic benefit: This exercise is particularly effective for promoting the recovery of finger mobility and grasping functionality. By engaging the patient in repetitive yet purposeful movements, it supports the reactivation of fine motor skills required for everyday object manipulation tasks, such as picking up, holding, and releasing small items. Over time, this repetitive interaction can contribute to restoring strength, coordination, and precision in the hand.



Figure 6. Grab the Cube. The player's avatar approaches a virtual cube and performs a hand-closing gesture to grasp it, simulating real finger flexion and extension. The task can be repeated multiple times, with difficulty and repetitions adjusted by the clinician to match the patient's rehabilitation needs and progress.

The **Coin Collection** game (Figure 7) places the avatar inside a maze where several coins are scattered along a predefined path. The objective is to navigate through the maze and collect all the coins while avoiding contact with the maze walls. This requires careful spatial awareness, precise control of hand motion, and continuous adjustment of movement speed and direction. *Therapeutic benefit*: This task helps develop dexterity and coordination by challenging the patient to perform controlled, fine-tuned movements. The alternation between different motion types, such as fast reaching and slow, steady navigation, stimulates fine motor control, particularly in the wrist and finger joints. Moreover, the requirement to avoid collisions encourages the patient to maintain focus and regulate force and timing, thereby reinforcing proprioceptive awareness and precision.

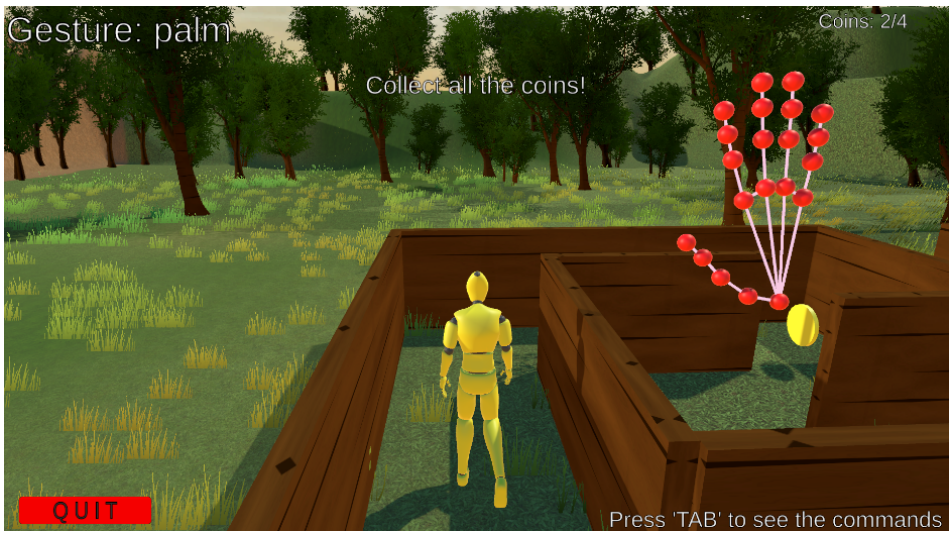

Figure 7. Coin Collection. In this mini-game, the avatar navigates through a maze to collect coins while avoiding wall collisions. The task promotes wrist mobility, spatial coordination, and precision in hand control, supporting the improvement of fine motor and visiomotor skills.

In the **Simon Says** exercise (Figure 8), the avatar stands before a virtual white screen where a sequence of gestures is visually displayed. The patient's goal is to observe and accurately replicate each gesture in the correct order and manner. The exercise gradu-

ally increases in complexity, introducing longer sequences and more intricate gestures as the patient improves. This structure encourages not only motor performance but also cognitive engagement, as the patient must memorize and recall the correct sequence of actions. *Therapeutic benefit*: The game enhances hand–eye coordination and strengthens the connection between cognitive and motor functions. By requiring visual processing, short-term memory, and motor execution to operate in synchrony, it promotes cognitive-motor integration, a crucial aspect of functional recovery. This makes the task especially suitable for patients aiming to regain fluidity, timing, and precision in gesture-based movements used in daily activities.



Figure 8. Simon Says: A cognitive-motor exercise in which the avatar replicates a visual sequence of hand gestures displayed on screen. The sequence length and complexity could increase progressively, encouraging motor coordination, memory, and attention during the rehabilitation process.

At the end of each session, an Excel file is generated and stored locally. This file logs the date and time of the session, along with performance metrics for each game. In Grab the Cube, the file records the number of expected repetitions and the number actually performed, as well as the maximum, minimum, and average duration of cube holding. In the Coin Collection game, it stores the maximum number of collectible coins and the number actually collected, along with the minimum, maximum, and average time taken to collect a coin. In Simon Says, the file includes the number of gestures expected to be replicated and the number performed, as well as the minimum, maximum, and average execution time per gesture.

4. Results and Discussion

All tests were conducted on an MSI Katana GF66 (Micro-Star International Co., Ltd. (MSI), Zhonghe District, New Taipei City, Taiwan) equipped with a 12th Gen Intel® Core™ i7-12700H processor (Intel Corporation, Santa Clara, CA, USA) and an NVIDIA GeForce RTX 3050 Ti Laptop GPU (NVIDIA Corporation, Santa Clara, CA, USA). The version of Python used was 3.10, while Unity version 2022.3.61f1 was employed. The initial prototype exhibited an input delay of approximately 0.5 s due to synchronous blocking, where YOLO inference temporarily halted the data flow between Python and Unity. To address this limitation, a multi-threaded asynchronous architecture was implemented. YOLO inference now runs on a dedicated thread operating at 14 Hz, while the main thread performs frame capture, MediaPipe processing, and data transmission at over 100 Hz. Non-blocking queues with a maximum capacity of two elements prevent message accumulation, and buffer

flushing on the Unity client ensures only the most recent data is processed. This solution reduced end-to-end latency from about 500 ms to under 20 ms.

4.1. Performance Evaluation

Performance evaluation results were collected from 36 different healthy subjects, aged between 20 and 60 years, each of whom performed three trials for every mini-game to mitigate the learning curve of a technology and mode of interaction that is not complicated, but unfamiliar to the vast majority of the population.

This choice of conducting the experimental validation exclusively with healthy subjects was aligned with the main objective of the present study, which focused on assessing the technological feasibility, stability, and usability of the proposed system rather than on clinical outcomes. The involvement of healthy participants allowed for the controlled evaluation of interaction mechanisms, gesture recognition accuracy, and overall system responsiveness without the variability typically introduced by clinical populations. This approach ensured that the validation process could concentrate on the robustness and reliability of the developed technologies before proceeding to the more sensitive and heterogeneous context of clinical testing.

Figure 9 shows the boxplots for each attempt across all players. All players completed the required number of repetitions.

As expected, results reveal a downward trend in average completion times across trials, indicating learning effects as users became increasingly familiar with gesture-based control. However, the extent of improvement differs notably across the minigames, reflecting variations in task complexity. The Coin Collection game shows the most pronounced performance gains, with median completion times decreasing substantially from Trial 1 to Trial 3. The relatively wide interquartile range compared to other games highlights greater inter-subject variability, likely due to the task's complexity, which involves continuous spatial navigation, precise gesture execution, and collision avoidance. This steep learning curve suggests that repeated exposure to motor-cognitive integration tasks significantly benefits user performance. The Simon Says game exhibits moderate improvement, with noticeable but less dramatic reductions in completion times. This intermediate learning rate aligns with the task's focus on gesture memory and sequencing rather than continuous spatial control. In contrast, the Grab the Cube game shows minimal changes across trials, with overlapping interquartile ranges and stable median values. The flat learning curve indicates that this simple grasping task quickly reaches a performance plateau, imposing low cognitive demands and requiring minimal motor adaptation.

To quantify the observed learning effects, paired *t*-tests were conducted for all three mini-games across successive trials. The results are reported in Tables 4–6. Coin Collection (Table 4) exhibited the most pronounced improvement. Completion times decreased significantly from Trial 1 to Trial 2 ($p = 0.00334$) and from Trial 1 to Trial 3 ($p = 0.00017$), while the reduction from Trial 2 to Trial 3 was not statistically significant ($p = 0.07430$). Average collection time followed a similar trend, with significant improvement between the first and last trials ($p = 0.00454$). These results suggest a steep learning curve, likely due to the task's complexity, which involves continuous spatial navigation, precise gesture control, and collision avoidance. The wider interquartile range observed in this game further reflects the greater variability in individual learning rates.

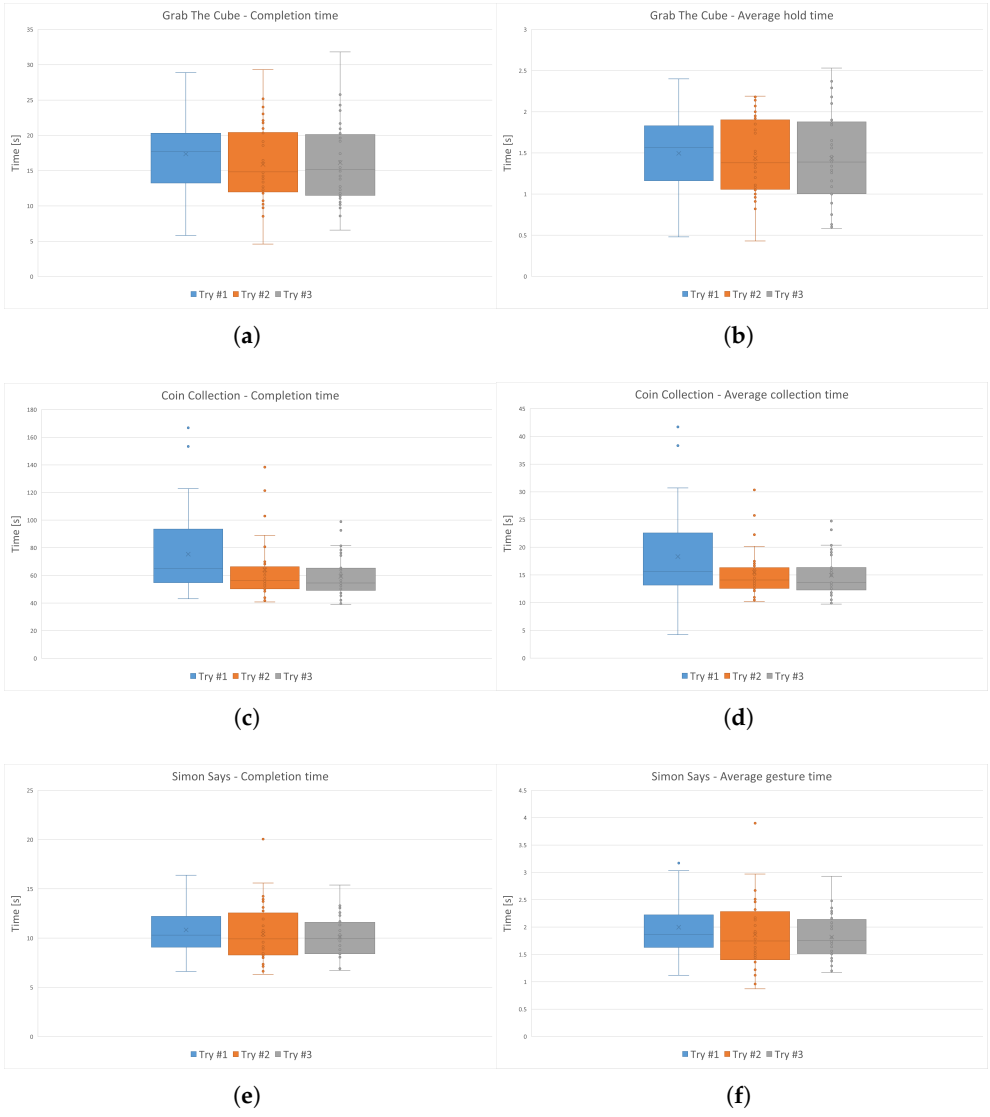

Figure 9. Results for each trial for every variable studied expressed in seconds: (**a**) Grab the Cube—completion time. (**b**) Grab the Cube—average hold gesture time. (**c**) Coin Collection—completion time. (**d**) Coin Collection—average time per collected coin. (**e**) Simon Says—completion time. (**f**) Simon Says—average gesture transition time.

Table 4. Paired *t*-test *p*-values for Coin Collection. Completion times and average collection times across successive trials.

Comparison	Completion Time			Average Collection Time		
	1–2	**2–3**	**1–3**	**1–2**	**2–3**	**1–3**
p-value	0.00334	0.07430	0.00017	0.04349	0.35182	0.00454

Grab the Cube (Table 5) showed a smaller improvement. Completion time was significantly reduced between Trials 1 and 2 ($p = 0.00637$), while later comparisons did not reach significance (1–3: $p = 0.14628$; 2–3: $p = 0.67974$). The average hold time did not show significant changes across trials, indicating that this simpler grasping task quickly reaches a performance plateau. These results align with the low cognitive and motor demands of the task.

Table 5. Paired *t*-test *p*-values for Grab the Cube. Completion times and average hold times across successive trials.

Comparison	Completion Time			Average Hold Time		
	1–2	**2–3**	**1–3**	**1–2**	**2–3**	**1–3**
p-value	0.00637	0.67974	0.14628	0.25144	0.99282	0.44072

Simon Says (Table 6) demonstrated marginal improvement. Only the comparison between Trials 1 and 3 reached significance for both completion time ($p = 0.04760$) and average gesture time ($p = 0.02830$), while intermediate comparisons were not significant. This pattern reflects the task's moderate complexity, focusing on gesture memory and sequencing rather than continuous spatial control, leading to a slower or less consistent learning effect across participants.

Table 6. Paired *t*-test *p*-values for Simon Says. Completion times and average gesture times across successive trials.

Comparison	Completion Time			Average Gesture Time		
	1–2	**2–3**	**1–3**	**1–2**	**2–3**	**1–3**
p-value	0.36344	0.44398	0.04760	0.17131	0.67177	0.02830

Overall, the largest improvements occurred between the first and second trials, with smaller gains in subsequent attempts. Secondary metrics (average collection, hold, or gesture times) followed trends similar to the primary completion time, reinforcing the evidence of learning effects. These results indicate that task complexity strongly influences the magnitude and speed of learning: more complex tasks like Coin Collection (Table 4) show steeper learning curves and greater inter-subject variability, while simpler tasks like Grab the Cube (Table 5) plateau quickly.

Given the above considerations, only the data from each user's final attempt was considered, in order to reduce variance caused by inexperience with the system. From these attempts, the mean and standard deviation of the completion times were calculated. The resulting physiological ranges are reported in Table 7.

Table 7. Performance metrics summary. CC: Coin Collection; GtC: Grab the Cube; SS: Simon Says.

Metric	GtC	CC	SS
Completion Time (s)	16.1 ± 5.7	59.3 ± 14.5	10.1 ± 1.9
Average time per collected coin (s)	–	14.9 ± 3.6	–
Average hold gesture time (s)	1.4 ± 0.5	–	–
Average gesture transition time (s)	–	–	1.8 ± 0.4

4.2. Application's Usability and Engagement Evaluation

The application's usability was evaluated through the system usability scale (SUS), a well-established and widely adopted standardized questionnaire for quantifying perceived usability across interactive systems [52]. In this questionnaire, respondents express their level of agreement with each of the ten items on a five-point Likert scale, anchored at "Strongly disagree" (1) and "Strongly agree" (5), and, following the scoring procedure detailed by [53], these raw responses are systematically transformed into individual SUS scores on a 0–100 metric that enables straightforward comparison across users and studies. Consistent with established interpretive benchmarks, a SUS score of 68 or higher is generally regarded as indicative of above-average usability, whereas scores below 68

are interpreted as signaling below-average usability, providing a pragmatic threshold for summarizing overall system performance [54].

The user engagement scale (UES) in its short form [55] was administered to quantify the level of engagement experienced by users during their interaction with the system, and each of the twelve items was rated on a five-point agreement scale, thereby producing item responses that could be aggregated into both total and dimension-specific indicators of engagement. In addition to the overall UES score, four independent sub-scores were computed to capture distinct features of engagement, namely focused attention (FA), describing the extent to which participants felt absorbed in the experience and lost track of time; perceived usability (PU), relating to the sense of control, the effort required, and any negative experiences encountered during interaction; aesthetic appeal (AE), measuring the perceived attractiveness and visual pleasure afforded by the system; reward (RW), capturing feelings of enjoyment, novelty, and motivation to continue using the systems.

Tables 8 and 9 report the item-level and aggregated results. For each questionnaire, the median, minimum, and maximum values were used to summarize central tendency and variability across participants.

Table 8. Questions for the SUS assessment user testing. The median, minimum, and maximum values summarize the questionnaire results. # refers to the number of occurrences.

Item #	Questionnaire Item	Min	Median	Max
1	I think that I would like to use this system frequently	2	4	5
2	I found the system unnecessarily complex	1	2	5
3	I thought the system was easy to use	2	4	5
4	I think that I would need the support of a technical person to be able to use this system	1	2	5
5	I found the various functions in this system were well integrated	3	4	5
6	I thought there was too much inconsistency in this system	1	1	4
7	I would imagine that most people would learn to use this system very quickly	3	4	5
8	I found the system very cumbersome to use	1	2	4
9	I felt very confident using the system	2	4	5
10	I needed to learn a lot of things before I could get going with this system	1	2	4

With regard to the SUS results, participants' scores ranged from a minimum of 57.5 to a maximum of 95, with a median value of 80, and the first and third quartiles at 70 and 87.5, respectively. The mean SUS score was 78.2, which is well above the standard benchmark of 68, indicating a generally high level of usability. A single outlier at 42.5 was identified, likely corresponding to an individual participant who experienced greater difficulty or lower satisfaction with the system. The narrow interquartile range (IQR = 17.5) and the proximity between the median and the upper quartile suggest that the majority of users rated the system's usability highly and consistently.

Regarding user engagement, the results obtained from the UES questionnaire show a similarly positive trend. The short-form UES revealed a generally high level of engagement among participants, with an overall median score of 4.02, a minimum value of 3.17, a first quartile (Q1) of 3.83, a third quartile (Q3) of 4.33, and a maximum value of 4.75, corresponding to an IQR of 0.50. At the subscale level, the median values were 3.15 for FA, 4.37 for PU, 4.23 for AE, and 4.34 for RW. These results indicate that users perceived the system as highly usable, visually appealing, and rewarding, while showing slightly lower and more variable values in the FA dimension. Notably, the lower FA score was strongly influenced

by item FA-1 ("I lost myself in the experience"), which received lower ratings compared to the other items within the same factor. Qualitative feedback suggested that this discrepancy was primarily due to differences in the interpretation of the term "lost", with some participants associating it with a positive sense of immersion, as intended by the questionnaire's authors, and others interpreting it negatively as confusion or disorientation. Despite this semantic ambiguity, the remaining items of the FA dimension, as well as the consistently high PU, AE, and RW scores, confirm that participants remained focused, found the system aesthetically pleasant, and considered the interaction intrinsically engaging and motivating. When considered together, the SUS and UES results portray a complementary relationship between usability and engagement. High SUS and UES-PU scores confirm that ease of use and control are key contributors to positive user experience, while elevated AE and RW values reveal the system's ability to elicit aesthetic pleasure and motivational involvement. The combination of favorable usability ratings and positive affective responses suggests that users were not only able to interact with the system efficiently but also found it enjoyable and motivating, reinforcing the system's potential for sustained use and user satisfaction.

Table 9. Questions for the UES assessment's user testing. The median, minimum, and maximum values are used to display the overall results of the questionnaire.

Item #	Questionnaire Item	Min	Median	Max
FA-1	I lost myself in this experience	1	1	4
FA-2	The time I spent using the Application just slipped away	2	4	5
FA-3	I was absorbed in this experience	2	4	5
PU-1	I felt frustrated while using this Application	1	2	4
PU-2	I found this Application confusing to use	1	1	4
PU-3	Using this Application was taxing	1	1	4
AE-1	This Application was attractive	3	4	5
AE-2	This Application was aesthetically appealing	3	4	5
AE-3	This Application appealed to my senses	3	4	5
RW-1	Using Application was worthwhile	3	4	5
RW-2	My experience was rewarding	3	4	5
RW-3	I felt interested in this experience	3	5	5

4.3. Strengths, Limitations, and Future Directions

Beyond the technological validation, the proposed system holds relevant implications for clinical rehabilitation. ReHAb Playground addresses several key limitations of conventional therapy. First, adherence is enhanced through gamification, which previous studies have shown can increase patient engagement by 40–60% compared to standard exercises [56,57]. Second, accessibility is improved as the system enables home-based rehabilitation, removing travel barriers that can be particularly challenging for patients in rural areas or with limited mobility. Third, objective monitoring is supported through automated metrics, such as completion time and gesture accuracy, allowing clinicians to track patient progress remotely and make more informed decisions [23].

Compared to existing platforms, such as those based on VR headsets [16], Kinect sensors [20], or Leap Motion technology [23], ReHAb Playground offers several clear advantages. It is more cost-effective, as it relies on a standard webcam rather than specialized hardware. Furthermore, it provides a substantially larger gesture library, comprising 33 predefined gestures compared to fewer than 10 in similar systems [16,19,23], while maintaining comparable high usability and high engagement levels. Its modular architecture also facilitates the rapid expansion of exercise sets, enabling adaptation to diverse therapeutic objectives and patient groups. The deployment in clinical settings will require additional

steps, including validation against clinical gold standards, therapist training to tailor system parameters to individual patients, and integration with electronic health records to support longitudinal monitoring. These measures will ensure that ReHAb Playground not only complements traditional rehabilitation but also delivers meaningful, measurable benefits in real-world clinical practice.

The findings presented above provide an initial understanding of ReHAb Playground's performance and user interaction dynamics under controlled conditions. While the results demonstrate encouraging trends in user adaptation and system responsiveness, several aspects require further refinement to ensure greater reliability, scalability, and clinical applicability. Firstly, all healthy subjects were included in the evaluation, thus the system has not yet been validated against clinical gold-standard metrics. This limits the clinical generalizability of the results, as motor impairments can significantly affect gesture execution, movement smoothness, and system responsiveness. Future studies will therefore involve patients with different levels of motor dysfunction to assess therapeutic effectiveness under realistic rehabilitation conditions, implement lighting-robust preprocessing, develop adaptive difficulty algorithms, and conduct controlled trials comparing the system with conventional rehabilitation methods. Particular attention will be paid to measuring clinical outcomes, patient engagement, adherence, and overall satisfaction, to clearly demonstrate the added value of the proposed approach in real rehabilitation contexts.

Testing the updated pipeline across devices showed minimal variation in latency, with average values of 7–10 ms across all models. Such low latency levels guarantee smooth real-time interaction and consistent gesture recognition, eliminating previous desynchronization effects. Future developments will focus on adaptive timing and load balancing to further optimize performance on lower-end hardware.

Several additional factors affect system performance and generalizability. The gesture recognition network may not fully represent impaired hand morphologies, transitional hand positions can cause misclassification, and environmental factors such as non-uniform lighting or shadows can degrade tracking. Hardware heterogeneity, including variations in webcam quality, CPU speed, and GPU availability, further influences latency and reliability. The core code can also be adjusted to redefine gestures directly in Unity, enabling developers to select the most suitable combination from the 33 gestures detected by the recognition network. Furthermore, the depth coordinate is used only for real-time hand visualization. Future mini-games could exploit depth data to detect additional motion patterns, such as wrist flexion and extension, enhancing both gameplay and therapeutic effectiveness. By capturing three-dimensional motion parameters, such as wrist flexion/extension, arm rotation, and postural adjustments, the system could provide precise motion tracking, context-aware feedback, and adaptive difficulty adjustment, increasing both engagement and functional relevance.

These efforts will enhance generalizability, clinical relevance, and robustness, providing quantitative evidence of therapeutic benefit. The gamified environment also offers potential beyond rehabilitation. Interactive and educational activities, such as gesture-based object counting or coordination games, can promote engagement and learning, particularly for children, combining enjoyment with developmental and therapeutic benefits.

5. Conclusions

This study introduces ReHAb Playground, a gesture-controlled serious game framework designed to make upper limb rehabilitation more engaging and accessible. Developed in Unity, it combines virtual reality with real-time markerless hand tracking and gesture recognition to support home-based therapy and encourage user motivation and adherence. The system integrates YOLOv10n-based gesture recognition with MediaPipe tracking

in a modular environment that allows easy customization of therapeutic tasks. Three mini-games, Grab the Cube, Coin Collection, and Simon Says, target specific motor and cognitive functions, while adjustable parameters and performance metrics support clinical assessment. Preliminary tests with healthy participants established reference performance ranges and showed improved execution times over repeated trials, indicating user learning and adaptation. Coin Collection displayed the greatest variability and learning effect due to its higher task complexity. Although limited by the small sample size, these findings provide a foundation for future studies with patients.

Future developments will aim to evolve the system from a validated prototype into a clinically applicable rehabilitation tool, ensuring robustness and adaptability across diverse home environments. Building on its modular architecture, ReHAb Playground can be easily extended with new therapeutic activities and integrated with clinical data management platforms to support personalized, long-term monitoring. The introduction of web-based access and adaptive interaction strategies will further enhance usability, accessibility, and patient engagement, strengthening its potential for large-scale deployment in home-based rehabilitation. This versatility positions it as a promising and scalable platform that bridges technological innovation and clinical practice, offering a concrete step toward more engaging, data-driven, and personalized motor rehabilitation.

Author Contributions: Conceptualization, S.R., L.A., F.B and F.P.; Methodology, S.R., L.A., F.B. and F.P.; Software, S.R., L.A., F.B. and F.P.; Validation, G.M., C.I., E.V. and L.U.; Formal analysis, S.R., L.A., F.B. and F.P.; Investigation, S.R., G.M., L.A., F.B., F.P., C.I., E.V. and L.U.; Resources, E.V. and L.U.; Data curation, S.R., L.A., F.B. and F.P.; Writing—Original draft preparation, S.R., L.A., F.B. and F.P.; Writing—Review and editing, G.M., C.I., E.V. and L.U.; Visualization, S.R., L.A., F.B. and F.P.; Supervision, G.M., C.I. and L.U.; Project administration, E.V. and L.U.; Funding acquisition, E.V. and L.U. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data are available upon request to the authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ning, H.; Wang, Z.; Li, R.; Zhang, Y.; Mao, L. A review on serious games for exercise rehabilitation. *arXiv* **2022**, arXiv:2201.04984. [CrossRef]
2. Tomić, T.J.D.; Savić, A.M.; Vidaković, A.S.; Rodić, S.Z.; Isaković, M.S.; Rodríguez-de Pablo, C.; Keller, T.; Konstantinović, L.M. ArmAssist robotic system versus matched conventional therapy for poststroke upper limb rehabilitation: A randomized clinical trial. *BioMed Res. Int.* **2017**, *2017*, 7659893. [CrossRef]
3. Go, A.S.; Mozaffarian, D.; Roger, V.L.; Benjamin, E.J.; Berry, J.D.; Borden, W.B.; Bravata, D.M.; Dai, S.; Ford, E.S.; Fox, C.S.; et al. Heart disease and stroke statistics—2013 update: A report from the American Heart Association. *Circulation* **2013**, *127*, 143–152. [CrossRef]
4. Norouzi-Gheidari, N.; Levin, M.F.; Fung, J.; Archambault, P. Interactive virtual reality game-based rehabilitation for stroke patients. In Proceedings of the 2013 International Conference on Virtual Rehabilitation (ICVR), IEEE, Piscataway, NJ, USA, 26–29 August 2013; pp. 220–221.
5. Ghaziani, E.; Couppé, C.; Henkel, C.; Siersma, V.; Søndergaard, M.; Christensen, H.; Magnusson, S.P. Electrical somatosensory stimulation followed by motor training of the paretic upper limb in acute stroke: Study protocol for a randomized controlled trial. *Trials* **2017**, *18*, 1–9. [CrossRef]
6. Burke, J.W.; McNeill, M.; Charles, D.K.; Morrow, P.J.; Crosbie, J.H.; McDonough, S.M. Optimising engagement for stroke rehabilitation using serious games. *Vis. Comput.* **2009**, *25*, 1085–1099. [CrossRef]
7. Laamarti, F.; Eid, M.; El Saddik, A. An overview of serious games. *Int. J. Comput. Games Technol.* **2014**, *2014*, 358152. [CrossRef]
8. Proença, J.P.; Quaresma, C.; Vieira, P. Serious games for upper limb rehabilitation: A systematic review. *Disabil. Rehabil. Assist. Technol.* **2018**, *13*, 95–100. [CrossRef] [PubMed]

9. Doumas, I.; Everard, G.; Dehem, S.; Lejeune, T. Serious games for upper limb rehabilitation after stroke: A meta-analysis. *J. Neuroeng. Rehabil.* **2021**, *18*, 100. [CrossRef] [PubMed]
10. Marullo, G.; Innocente, C.; Ulrich, L.; Lo Faro, A.; Porcelli, A.; Ruggieri, R.; Vecchio, B.; Vezzetti, E. Home-based mirror therapy in phantom limb pain treatment: The augmented humans framework. In *Multimedia Tools and Applications*; Springer: Berlin/Heidelberg, Germany, 2025; pp. 1–33.
11. Moradi, M.; Noor, N.F.B.M. The impact of problem-based serious games on learning motivation. *IEEE Access* **2022**, *10*, 8339–8349. [CrossRef]
12. Ulrich, L.; Carmassi, G.; Garelli, P.; Lo Presti, G.; Ramondetti, G.; Marullo, G.; Innocente, C.; Vezzetti, E. SIGNIFY: Leveraging machine learning and gesture recognition for sign language teaching through a serious game. *Future Internet* **2024**, *16*, 447. [CrossRef]
13. Corona, F.; Chiuri, R.M.; Filocamo, G.; Foà, M.; Lanzi, P.L.; Lopopolo, A.; Petaccia, A. Serious games for wrist rehabilitation in juvenile idiopathic arthritis. In Proceedings of the 2018 IEEE Games, Entertainment, Media Conference (GEM), IEEE, Galway, Ireland, 15–17 August 2018; pp. 35–42.
14. Ao, W.; Hui, C.; Lihao, L.; Chen, A.; Lin, Z.; Han, J.; Ding, G. YOLOv10: Real-Time End-to-End Object Detection. *arXiv* **2024**, arXiv:2405.14458.
15. Zhang, F.; Bazarevsky, V.; Vakunov, A.; Tkachenka, A.; Sung, G.; Chang, C.L.; Grundmann, M. Mediapipe hands: On-device real-time hand tracking. *arXiv* **2020**, arXiv:2006.10214.
16. Pereira, M.F.; Prahm, C.; Kolbenschlag, J.; Oliveira, E.; Rodrigues, N.F. A virtual reality serious game for hand rehabilitation therapy. In Proceedings of the 2020 IEEE 8th International Conference on Serious Games and Applications for Health (SeGAH), IEEE, Vancouver, BC, Canada, 12–14 August 2020; pp. 1–7.
17. Song, X.; Ali, N.M.; Mhd Salim, M.H.; Rezaldi, M.Y. A literature review of virtual reality exergames for older adults: Enhancing physical, cognitive, and social health. *Appl. Sci.* **2025**, *15*, 351. [CrossRef]
18. Dore, B.; Gaudreault, A.; Everard, G.; Ayena, J.C.; Abboud, A.; Robitaille, N.; Batcho, C.S. Acceptability, feasibility, and effectiveness of immersive virtual technologies to promote exercise in older adults: A systematic review and meta-analysis. *Sensors* **2023**, *23*, 2506. [CrossRef]
19. Bressler, M.; Merk, J.; Gohlke, T.; Kayali, F.; Daigeler, A.; Kolbenschlag, J.; Prahm, C. A Virtual Reality Serious Game for the Rehabilitation of Hand and Finger Function: Iterative Development and Suitability Study. *JMIR Serious Games* **2024**, *12*, e54193. [CrossRef]
20. Trombetta, M.; Henrique, P.P.B.; Brum, M.R.; Colussi, E.L.; De Marchi, A.C.B.; Rieder, R. Motion Rehab AVE 3D: A VR-based exergame for post-stroke rehabilitation. *Comput. Methods Programs Biomed.* **2017**, *151*, 15–20. [CrossRef]
21. Taylor, M.J.; Lindsverk, H.; Nygaard, M.; Hunt, J.; Shanks, J.; Griffin, M.; Shawis, T.; Impson, R. Comparing the energy expenditure of Wii-Fit-based therapy with that of traditional physiotherapy in an older adult population. *J. Am. Geriatr. Soc.* **2014**, *62*. [CrossRef] [PubMed]
22. Fiorin, M.R.; De Marchi, A.C.B.; Colussi, E.L.; Rieder, R.; Trombetta, M. Motion Rehab: Um jogo sério para idosos com sequelas de Acidente Vascular Encefálico. In Proceedings of the XIV Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS), Foz do Iguaçu, Brazil, 19–22 May 2014. SBC: Rio de Janeiro, Brazil 2014; pp. 1776–1779.
23. Elnaggar, A.; Reichardt, D. Digitizing the hand rehabilitation using serious games methodology with user-centered design approach. In Proceedings of the 2016 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 15–17 December 2016; IEEE: Las Vegas, NV, USA, 15–17 December 2016; pp. 13–22.
24. Yan, S.; Xiong, Y.; Lin, D. Spatial temporal graph convolutional networks for skeleton-based action recognition. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LO, USA, 2–7 February 2018; Volume 32.
25. Zhang, W.; Lin, Z.; Cheng, J.; Ma, C.; Deng, X.; Wang, H. Sta-gcn: Two-stream graph convolutional network with spatial–temporal attention for hand gesture recognition. *Vis. Comput.* **2020**, *36*, 2433–2444. [CrossRef]
26. Chen, J.; Zhao, S.; Meng, H.; Cheng, X.; Tan, W. An interactive game for rehabilitation based on real-time hand gesture recognition. *Front. Physiol.* **2022**, *13*, 1028907. [CrossRef]
27. Husna, R.; Brata, K.C.; Anggraini, I.T.; Funabiki, N.; Rahmadani, A.A.; Fan, C.P. An Investigation of Hand Gestures for Controlling Video Games in a Rehabilitation Exergame System. *Computers* **2025**, *14*, 25. [CrossRef]
28. Xiao, Y.; Funabiki, N.; Anggraini, I.T.; Shih, C.L.; Fan, C.P. A Study of Exergame System Using Hand Gestures for Wrist Flexibility Improvement for Tenosynovitis Prevention. *Information* **2024**, *15*, 622. [CrossRef]
29. Schoorl, K.; Cisneros, T.P.; Salah, A.A.; Schouten, B. Wrist movement classification for adaptive mobile phone based rehabilitation of children with motor skill impairments. *arXiv* **2024**, arXiv:2401.17134. [CrossRef]
30. Luciani, B.; Pedrocchi, A.; Tropea, P.; Seregni, A.; Braghin, F.; Gandolla, M. Augmented reality for upper limb rehabilitation: Real-time kinematic feedback with HoloLens 2. *Virtual Real.* **2025**, *29*, 57. [CrossRef]
31. Li, Z.; Zatsiorsky, V.; Latash, M. Contribution of the extrinsic and intrinsic hand muscles to the moments in finger joints. *Clin. Biomech.* **2000**, *15*, 203–211. [CrossRef]

32. Akbaş, A. Hand-Focused Strength and Proprioceptive Training for Improving Grip Strength and Manual Dexterity in Healthy Adults: A Systematic Review and Meta-Analysis. *J. Clin. Med.* **2025**, *14*, 6882. [CrossRef]
33. Nakayama, H.; Jørgensen, H.S.; Raaschou, H.O.; Olsen, T.S. Recovery of upper extremity function in stroke patients: The Copenhagen Stroke Study. *Arch. Phys. Med. Rehabil.* **1994**, *75*, 394–398. [CrossRef]
34. Vinstrup, J.; Calatayud, J.; Jakobsen, M.D.; Sundstrup, E.; Jørgensen, J.R.; Casaña, J.; Andersen, L.L. Hand strengthening exercises in chronic stroke patients: Dose-response evaluation using electromyography. *J. Hand Ther.* **2018**, *31*, 111–121. [CrossRef]
35. Nijland, R.H.; van Wegen, E.E.; Harmeling-van der Wel, B.C.; Kwakkel, G. Presence of finger extension and shoulder abduction within 72 hours after stroke predicts functional recovery: Early prediction of functional outcome after stroke: The EPOS cohort study. *Stroke* **2010**, *41*, 745–750. [CrossRef]
36. Kim, D.; Baghi, R.; Koh, K.; Zhang, L.Q. MCP extensors respond faster than flexors in individuals with severe-to-moderate stroke-caused impairment: Evidence of uncoupled neural pathways. *Front. Neurol.* **2023**, *14*, 1119761. [CrossRef] [PubMed]
37. Jarque-Bou, N.J.; Vergara, M.; Sancho-Bru, J.L. Understanding forearm muscle activity during everyday common grasps: Insights for rehabilitation, prosthetic control, and human–machine interaction. *Appl. Sci.* **2024**, *14*, 3190. [CrossRef]
38. Anderton, W.; Tew, S.; Ferguson, S.; Hernandez, J. Movement preferences of the wrist and forearm during activities of daily living. *J. Hand Ther.* **2023**, *36*, 580–592. [CrossRef]
39. Alif, M.A.R.; Hussain, M. YOLOv1 to YOLOv10: A comprehensive review of YOLO variants and their application in the agricultural domain. *arXiv* **2024**, arXiv:2406.10139. [CrossRef]
40. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J. Yolov10: Real-time end-to-end object detection. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 107984–108011.
41. Mao, M.; Lee, A.; Hong, M. Efficient Fabric Classification and Object Detection Using YOLOv10. *Electronics* **2024**, *13*, 3840. [CrossRef]
42. Yaseen; Kwon, O.J.; Kim, J.; Lee, J.; Ullah, F. Evaluation of Benchmark Datasets and Deep Learning Models with Pre-Trained Weights for Vision-Based Dynamic Hand Gesture Recognition. *Appl. Sci.* **2025**, *15*, 6045. [CrossRef]
43. Bazarevsky, V.; Zhang, F. *On-Device, Real-Time Hand Tracking with Mediapipe*; Google AI Blog: Columbus, OH, USA, 2019.
44. Reimer, D.; Podkosova, I.; Scherzer, D.; Kaufmann, H. Evaluation and improvement of HMD-based and RGB-based hand tracking solutions in VR. *Front. Virtual Real.* **2023**, *4*, 1169313. [CrossRef]
45. Pu, M.; Chong, C.Y.; Lim, M.K. Robustness evaluation in hand pose estimation models using metamorphic testing. In Proceedings of the 2023 IEEE/ACM 8th International Workshop on Metamorphic Testing (MET), Melbourne, Australia, 14 May 2023; pp. 31–38.
46. Maggioni, V.; Azevedo-Coste, C.; Durand, S.; Bailly, F. Optimisation and comparison of markerless and marker-based motion capture methods for hand and finger movement analysis. *Sensors* **2025**, *25*, 1079. [CrossRef] [PubMed]
47. Amprimo, G.; Masi, G.; Pettiti, G.; Olmo, G.; Priano, L.; Ferraris, C. Hand tracking for clinical applications: Validation of the Google MediaPipe Hand (GMH) and the depth-enhanced GMH-D frameworks. *Biomed. Signal Process. Control* **2024**, *96*, 106508. [CrossRef]
48. Bazarevsky, V.; Grishchenko, I.; Raveendran, K.; Zhu, T.; Zhang, F.; Grundmann, M. Blazepose: On-device real-time body pose tracking. *arXiv* **2020**, arXiv:2006.10204.
49. Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.E.; Sheikh, Y. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 172–186. [CrossRef]
50. Intel. *Intel RealSense Viewer Home Page*; Intel: Santa Clara, CA, USA, 2024.
51. Intel. *Intel RealSense Depth Camera SR300 Series Product Family Datasheet*; Intel: Santa Clara, CA, USA, 2024.
52. Lewis, J.R. The System Usability Scale: Past, Present, and Future. *Int. J. Hum.–Comput. Interact.* **2018**, *34*, 577–590. [CrossRef]
53. Brooke, J. SUS: A 'Quick and Dirty' Usability Scale. In *Usability Evaluation In Industry*; CRC Press: Boca Raton, FL, USA, 1996; p. 6.
54. Sauro, J.; Lewis, J.R. *Quantifying the User Experience: Practical Statistics for User Research*, 1st ed.; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 2012.
55. O'Brien, H.L.; Cairns, P.; Hall, M. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *Int. J. Hum.-Comput. Stud.* **2018**, *112*, 28–39. [CrossRef]
56. Sabaté, E. *Adherence to Long-Term Therapies: Evidence for Action*; World Health Organization: Geneva, Switzerland, 2003.
57. Maclean, N.; Pound, P.; Wolfe, C.; Rudd, A. Qualitative analysis of stroke patients' motivation for rehabilitation. *Bmj* **2000**, *321*, 1051–1054. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.