Hand Function Assessment using Computer Vision for Hand Rehabilitation

Kamal[1], Debanga Raj Neog[1†], H Pallab Jyoti Dutta[2], and M. K. Bhuyan[1,2]
[1]*Mehta Family School of Data Science and Artificial Intelligence*
[2]*Department of Electronics and Electrical Engineering*
Indian Institute of Technology Guwahati, Assam, India - 781039
email: {kkamal, dneog, h18, mkb}@iitg.ac.in

***Abstract*—Individuals who suffer a stroke often experience impaired hand function. The normal function can be restored by hand rehabilitation exercises with constant monitoring and assessment of motor recovery progress. However, this poses a challenge to patients due to regular hospital visits and physiotherapy sessions. We address this inconvenience by developing a reliable vision-based hand rehabilitation system that guides patients to perform rehabilitation exercises, such as the Sollerman Hand Function Test and Box and Block Test. The patients are asked to perform exercises based on instructions shown on a computer screen and the system tracks finger and hand to estimate performance scores. This scoring provides feedback to the patients to monitor their progress and improve their hand function gradually. Experiments conducted with ten males and five females highlight that the proposed vision-based hand rehabilitation exercises are effective in improving hand functionalities.**

***Index Terms*—box and block test, computer vision, hand rehabilitation, sollerman hand function test.**

I. Introduction

Hand function impairment is a typical outcome of stroke and it can greatly affect the quality of life and ability to perform daily activities. Traditionally, the Sollerman Hand Function Test (SHFT) [1] and Box and Block Test (BBT) [2] have been used to evaluate hand function and track recovery progress in stroke rehabilitation. However, these tests have limitations in terms of accessibility and requirements, such as skilled assessors, specific equipment, and physical presence in a clinical setting [3].

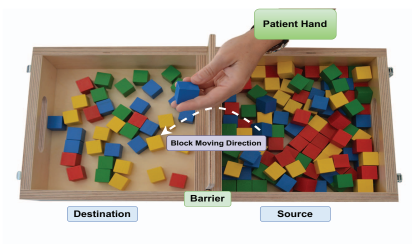

Fig. 1: An example of the original Box and Block Test.

The original BBT is demonstrated in Fig. 1, where the patient tries to move as many blocks as possible from the source to the destination within 60 seconds without touching the barriers. On the other hand, Fig. 2 shows all eight subtests of the SHFT. This work develops a system for hand rehabilitation using computer vision techniques that can serve as a virtual alternative to the SHFT and BBT. The proposed system eliminates the need for exercise-specific hardware by suggesting exercises on a computer screen and accurately evaluating them based on a score. This score provides real-time feedback to patients enabling them to track their rehabilitation progress and improve hand function gradually. This will not only increase the accessibility of rehabilitation services but also enhance the efficiency of healthcare delivery, especially for patients facing difficulties due to regular hospital visits and physiotherapy sessions.

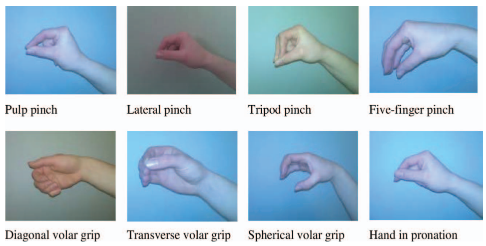

Fig. 2: Subtests of Sollerman Hand Function Test [4].

Integrating computer vision techniques into hand rehabilitation eliminates the need for additional hardware, making it a cost-effective solution that can be easily installed in various clinical and home settings. Furthermore, it increases the efficacy and effectiveness of hand rehabilitation.

Contribution. The main contribution of this paper is summarized below.

- Development of a vision-based alternative of SHFT and BBT using a 3D hand landmarks detection model tracking hand movements. Previous work [5] utilized a 2D framework of hand landmarks detection for this task, which is revamped by our 3D hand model.
- Demonstrated accurate handling of curve-shaped alphabets for the Sollerman writing test, which the 2D model [5] failed to accomplish. This enhancement in our approach contributes to the accuracy and effectiveness of our vision-based hand rehabilitation system.

†Corresponding author

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00129

II. Background

A. Literature Review

Recent advances in assessment tools and technologies have opened up new possibilities for more efficient and accessible evaluation of patients recovering from conditions such as tetraplegia and stroke in upper limb rehabilitation [1]. The Digital Box and Block Test is a noteworthy innovation that automates the traditional Box and Block Test, allowing for assessments to be conducted in residential spaces and results to be sent remotely to healthcare professionals [6]. The Virtual Box and Block Test provides a virtual alternative to hand, finger, and grasping assessments, which can allow unsupervised assessments and home-based virtual rehabilitation [7]. Computer Vision Box & Block Test (CV-BBT) is another significant development, integrating computer vision technologies, such as MediaPipe Hands, to create an interactive virtual assessment that requires minimal additional hardware and can be deployed in various settings [8]. These digital and virtual approaches are designed to reduce the time and clinic dependency associated with traditional assessments, making rehabilitation more accessible and efficient [9].

Immersive technologies, such as BBT Virtual Reality [10] and computer vision-based upper-limb rehabilitation suites, offer a unique assessment experience [5]. The BBT-VR, a virtual reality version of the Box and Block Test, is a viable and accurate assessment option, particularly for stroke patients. It allows remote monitoring and evaluation, eliminating the need for physical presence during assessments. The computer vision-based upper limb rehabilitation suite provides virtual alternatives to traditional assessment tools, such as the Sollerman hand function test, addressing restrictions on specialized materials and the presence of the therapist, which can be time-consuming and clinic dependent [8].

Furthermore, systems like the Hand Measurement System, based on haptic and vision devices, provide a unified approach to measuring essential hand parameters. These technologies aim to improve precision in post-stroke patients' rehabilitation by enabling accurate measurements of hand size, wrist, finger range of motion, and even force measurement [11]. To expand the horizon further, the Computer-Vision Based Upper-Limb Rehabilitation Suite is developed, aiming to eliminate the need for specialized materials and therapist presence. Offering virtual alternatives to traditional assessment tools increases accessibility and reduces the time constraints of upper limb rehabilitation assessments.

B. Hand Rehabilitation Assessment Methods

We have implemented the BBT and SHFT hand rehabilitation assessment methods using computer vision techniques to increase the efficacy and effectiveness of hand rehabilitation.

1) Box and Block Test: Our proposed Computer Vision-based BBT enhances the original BBT by virtualizing every step of the procedure, using only a mid-range PC and camera without any specialized computer peripherals or high computational resources [5]. We implemented the BBT using various computer vision and image processing packages and libraries. To ensure optimal results, we placed the camera approximately fifty centimeters away from the hand and made sure that there was enough natural light in the testing environment. Our system is most effective at tracking the hand and fingers of the patient in these circumstances. An example of our virtual BBT is shown in Fig. 3.

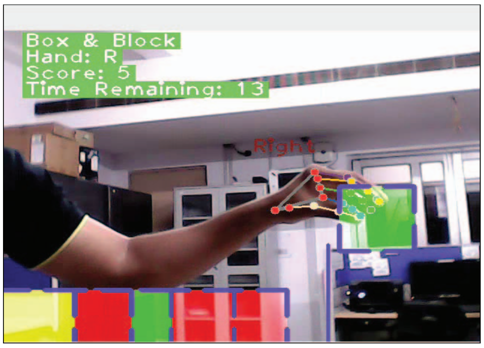

Fig. 3: Example of the virtual Box and Block Test. Here the user picks a box from the right side and drops it to the left side using his right hand.

2) Sollerman Hand Function Test: The SHFT is used to evaluate the function of the hand and provides an understanding of the quality and capabilities of the hand in daily activities. In the field of hand surgery, this test is very important because it helps determine how well different treatments work. The SHFT [12] consists of eight subtests for activities of daily life (ADL). Each subtest is allocated a percentage of use in ADL shown in Table I.

TABLE I: SHFT Sub-Tests

Hand Pinch and Grips	**Percentage of Use in ADL**
Pulp Pinch	20%
Lateral Pinch	20%
Tripod Pinch	10%
Five-Finger Pinch	15%
Diagonal Volar Grip	15%
Transverse Volar Grip	14%
Spherical Volar Grip	4%
Extension Grip	2%

Here are different types of grips that are used to hold objects with our hands:

a) Pulp pinch: This grip involves using the thumb and index, middle, or both fingers to hold the object. Examples of this grip include the Zip Open-Close Test, putting the key into a Yale lock and turning it 90 degrees, and more.

b) Lateral pinch: In this grip, the thumb and radial side of the index finger are used to hold the item. Examples of this grip include putting the key into a Yale lock and turning it 90 degrees, opening/closing a purse, taking coins from a purse, and performing a Yale lock unlock test.

c) Tripod pinch: This grip involves using the thumb, index, and middle fingers to cover the item, which might come into contact with the thumb's web. Examples of this grip include writing with a pen, picking up nuts and putting them on bolts, cutting modeling clay with a knife and fork, and more.

d) Five-finger pinch: This grip involves using the thumb and all four joined fingers to hold the item, with the palm not in contact with it. Examples of this grip include lifting wooden cubes over the edge (5 cm in height), folding paper and putting it into an envelope, pouring water from a 1-L paper milk package, and more.

e) Diagonal volar grip: In this grip, the thumb rests against the four fingers as the object is held, with it being in contact with the palm and its axis perpendicular to the hands. Examples of this grip include turning a screw with a screwdriver, picking up a telephone receiver, putting it to the ear, and more.

f) Transverse volar grip: This grip involves the thumb resting against the four fingers while holding the item, with its axis transverse to that of the hand and making contact with the palm. Examples of this grip include lifting the iron over the edge (5 cm in height), turning the door handle 30 degrees, pouring water from a jug, and more.

g) Spherical volar grip: This grip involves the thumb, four other fingers, and the palm being in contact with the object. An example of this grip is unscrewing the lid of jars.

h) Extension grip: In this grip, the object is held between the thumb and the four fingers, which are extended in the interphalangeal joints. It has no contact with the palm.

III. METHODOLOGY

We have created two assessment tools for hand rehabilitation using advanced computer vision techniques. The BBT evaluates manual dexterity by counting the number of blocks that an individual can move from one compartment to the other within 60 seconds. On the other hand, SHFT subtests provide a comprehensive assessment of various hand functions by providing a detailed understanding of the overall hand performance. These two methods together allow for accurate and efficient hand rehabilitation assessments through computer vision techniques. Fig. 4 shows the proposed architecture for our models. First, we capture a video using the webcam. Then, we perform hand pose estimation for the rehabilitation task and measure the performance of each test through the score.

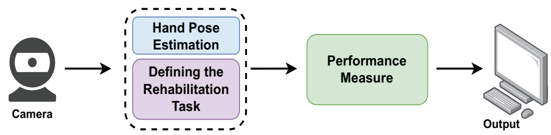

Fig. 4: The block diagram of the proposed work.

A detailed description of the proposed architecture's functionalities is given below:

Hand Pose Estimation. The webcam captures a video of the hand movement of the user during the experiment. The process starts by horizontally flipping the video frame and then shifting it horizontally. Subsequently, an absolute difference is calculated to verify if there is vertical-axis symmetry. After this, a pre-trained hand pose estimation (HPE) model is modified to detect 3D hand landmarks. Finally, essential elements, such as a timer, score counter, and partition divider are added to the video frame. The horizontal flipping of the video frame is represented in Eq. 1 :

$$I_{flipped}(x, y) = I_{original}(W - x - 1, y) \tag{1}$$

where, $I_{flipped}(x, y)$ represents the coordinates of the flipped video frame, and W represents the width of the frame. Eq. 2 represents the horizontal shifting of the video frame:

$$I_{shifted}(x, y) = I_{flipped}(x - shift, y) \tag{2}$$

where, $I_{shifted}(x, y)$ represents the coordinates of the shifted video frame, $shift$ represents how much the flipped frame is shifted horizontally.

The absolute difference between the original and shifted frame is represented by Eq. 3:

$$\text{Diff}(x, y) = |I_{original}(x, y) - I_{shifted}(x, y)| \tag{3}$$

where, Diff (x, y) represents the absolute difference between the values at coordinates (x, y), $I_{original}(x, y)$ represents the original coordinates, and $I_{shifted}(x, y)$ represents the shifted coordinates.

Now, for getting the 3D hand landmarks L we are applying the HPE function to the input video frame I, by using the coordinates (x_i, y_i, z_i) of each landmark. The calculation of landmarks L using HPE is shown in Eq. 4.

$$L = \text{HPE}(I) \tag{4}$$

Finally, each landmark can be represented by a vector $p_i = (x_i, y_i, z_i)$, where i ranges from 1 to 21:

$$L = [p_1, p_2, \ldots, p_{21}] \tag{5}$$

Defining the Rehabilitation Task. Now, after successful pose estimation, we can define the task for the computer vision-based BBT and SHFT.

Performance Measures. For the Box and Block test we have measured the performance based on the score out of 150 (there are 150 blocks in the source compartment). For the SHFT subtests performance measures, we have followed Table II.

A. Box and Block Test

The three stages of CV-BBT implementation are as follows: establishing and configuring the computer vision environment; identifying and tracking the hand and fingers; and putting the operational logic for every frame that is captured into practice.

1) Virtual Environment Construction: To create the computer vision environment, we use a single RGB camera that captures frames continuously and feeds them into the processing algorithm. During the testing phase, we use a camera running at 30 fps with a resolution of 480p connected to our host PC. We generate each of the 150 blocks one by one, new blocks appear only when the patient moves a block from source to destination. We assign a maximum limit of eight successfully moved blocks on each frame.

2) *Hand Detection and Operation Control:* We use a pre-trained hand pose estimation model based on Mediapipe for hand tracking and gesture detection. We reduce the need for data augmentation techniques by reusing the bounding box from the previous frame for predicting the locations. To start the "grabbing" behavior, we estimate the hand's pose and determine whether the thumb and index fingers are closed. We establish the hand's orientation with respect to the camera and whether it is inside the block's region. We continuously use the patient's hand movements to determine where the virtual block will appear next on each frame. To make sure the moving block cannot go through the partition divider, we have added collision checking. A virtual block falls into the appropriate compartment of the box when it is released from the patient's hand and crosses the solid divider.

B. Sollerman Hand Function Tests

We have implemented two SHFT subtests, which are Pulp pinch, and Tripod pinch. According to [12], there are eight ADLs that we have described in Table I. Also, we can see that the Pulp pinch and lateral pinch are the most widely used, and the Tripod pinch is the third most widely used ADL. Therefore, we decided to implement these three ADLs by using computer vision techniques to help the patient recover hand function. The evaluation of each sub-test is scored based on Table II.

TABLE II: Scoring system of the SHFT Sub-Tests

Score	Action
0	If the patient is unable to perform the task
1	If the task is partially performed within 60 seconds
2	If the task is completed but with great difficulty or takes between 40 to 60 seconds
3	If the task is completed with slight difficulty or takes between 20 to 40 seconds
4	If the task is carried out smoothly within 20 seconds

1) *Wallet Zip Open/Close:* The first sub-test of SHFT involves using computer vision to test a patient's ability to open and close a wallet zip, which is part of the Pulp pinch test typically used in ADL. To make the process easier, a virtual wallet with a zip that can be opened and closed was created. The patient grabs the zip by closing their thumb and index finger and drags point A to point B to open the zip, and then reverses the action to close it. The virtual element logic techniques for hand recognition are similar to those used in CV-BBT. The test also includes collision detection mechanisms to ensure proper function. An example of a wallet zip open/close test is shown in Fig. 5 and a simple collision detection mechanism is given by Eq. 6.

$$(x_2 - x_1)^2 + (y_2 - y_1)^2 \leq (r_1 + r_2)^2 \qquad (6)$$

where (x_1, y_1) and (x_2, y_2) are the centers of the circles A and B, respectively, with r_1 and r_2 being their respective radii. The collision between circles A and B is detected when the centers of each circle satisfy Eq. 6 [5].

Fig. 5: Example of the SHFT subtest wallet zip open/close. Here a user moves his hand from point A to point B for opening the zip and point B to Point A for closing the zip.

2) *Sollerman Writing Test:* The proposed SHFT subtests are an effective tool for improving patients' writing abilities, using just a single RGB camera. The system has four segments: an input shape detection algorithm, modeling of a virtual environment, tracking of the patient's hand, and application of the necessary operational logic. Our model is more versatile than the original one developed by [13]. Unlike their model, ours can handle straight lines with curves like the letters S, P, O, etc. An example of the SHFT writing test is shown in Fig. 6 where we have considered two alphabets W and S. Our model also provides the flexibility to the patients to follow either left-to-right or right-to-left writing of the alphabets, rather than assigning predefined paths to follow. An interactive computer vision environment is created by the input shape detection algorithm, which makes it possible to accurately track the patient's hand, determine the proportion of the shape that has been successfully drawn, and calculate the standard deviation of the hand from the shape. The system is set up in a lab environment under normal lighting conditions and the camera is placed about 50 cm away from the patient's hand.

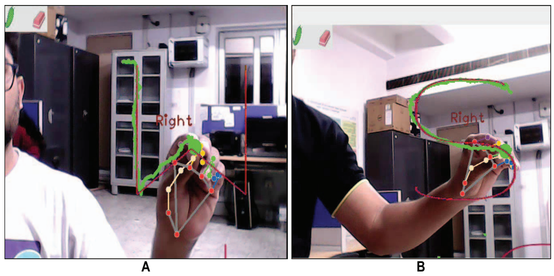

Fig. 6: Example of the SHFT writing test; (A) user writing alphabet W, and (B) user writing alphabet S.

IV. Experimental Results

Our primary objective for conducting this research is to help patients regain hand functionality through vision-based rehabilitation exercises. We conducted experiments on 10 male and 5 female users, with each user performing exercises over a period of 4 weeks using both hands.

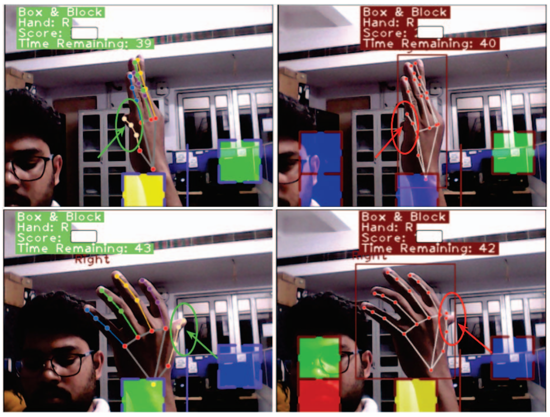

Fig. 7: Qualitative comparison of 3D hand landmark estimation model with respect to 2D model. The 3D model predicts keypoints more accurately (shown with a green oval and a green arrow in the first column) compared to the 2D model (shown with a red oval and a red arrow in the second column).

The effectiveness of the 3D hand landmark estimation model compared to the 2D model [5] is shown in Fig. 7. The 3D model localizes the landmarks in accurate locations (See first column of Fig. 7) compared to the 2D model (See second column of Fig. 7). We can observe that the keypoints within the green oval are accurately localized or estimated by the 3D model. However, the 2D model fails to localize the keypoints on the thumb as indicated by the red oval. Thus, we can conclude that the 3D model has a better keypoint localization capability than the 2D model.

We further considered two tests to observe the performance of the users over a period of 4 weeks. A box plot for the users is shown in Fig. 8, which shows the weekly performance of the users for BBT (row A of Fig. 8), SHFT wallet zip open/close (row B of Fig. 8), and SHFT writing tests (row C of Fig. 8). It is clearly evident that the performance of the users improves with each progressing week, highlighting that the rehabilitation exercises are able to help the users regain normal hand function.

Box and Block Test. Each user performed the exercise five times using both their left and right hand. The comparison of the scores of all male users is shown in Fig. 9 (A). We observed that every user performed less using their left hand than their right hand.

Sollerman Hand Function Tests. The experiment was conducted with the same number of users as BBT to perform each subtest five times. The scoring was based on Table II. The results of wallet zip open/close are shown in Fig. 9 (B).

The scores of the writing test conducted for male and female users in curved-shape alphabet ‘S’ are shown in Fig. 9 (C). Similarly, we can also perform the experiment for other curved alphabets.

V. Conclusion

We have developed a vision-based SHFT and BBT, which uses a 3D hand landmark detection model to assist stroke patients in regaining normal hand function. Our model was tested on 10 men and 5 women over a period of 4 weeks to observe the change in the hand function. Results showed that the users gradually regained their hand function as they were able to score higher compared to their previous week’s observation. Our system can be easily installed in clinical or home settings to help patients with hand rehabilitation exercises by eliminating the need for exercise-specific hardware or visiting medical centers for physiotherapy sessions regularly. In the future, we plan to develop a more robust 3D hand landmarks model capable of estimating landmarks even in extreme lighting variations and implement all eight SHFT exercises using vision-based techniques to help patients.

Acknowledgement. The project is supported by Science and Engineering Research Board Startup Research Grant (SRG/2022/001775) and IIT Guwahati Startup Grant (MD-SASUGIITG1329xDRN001).

References

[1] C. Sollerman and A. Ejeskär, “Sollerman hand function test: a standardised method and its use in tetraplegic patients,” *Scandinavian journal of plastic and reconstructive surgery and hand surgery*, vol. 29, no. 2, pp. 167–176, 1995.

[2] V. Mathiowetz, G. Volland, N. Kashman, and K. Weber, “Adult norms for the box and block test of manual dexterity,” *The American journal of occupational therapy*, vol. 39, no. 6, pp. 386–391, 1985.

[3] Q. Cai, J. Li, and J. Long, “Effect of physical and virtual feedback on reach-to-grasp movements in virtual environments,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 2, pp. 708–714, 2021.

[4] T. C. Hall, M. Nixon, J. Dias, T. Graham, and S. Cook, “How accurately does a simulation glove reflect function compared to rheumatoid arthritis sufferers?” *The Annals of The Royal College of Surgeons of England*, vol. 92, no. 7, pp. 605–609, 2010.

[5] O. N. Zestas, “A proposed computer-vision based upper-limb rehabilitation & evaluation suite,” in *Proceedings of the 2nd International Conference of the ACM Greek SIGCHI Chapter*, 2023, pp. 1–5.

[6] C.-P. Hsiao, C. Zhao, and E. Y.-L. Do, “The digital box and block test automating traditional post-stroke rehabilitation assessment,” in *2013 IEEE international conference on pervasive computing and communications workshops (PERCOM workshops)*. IEEE, 2013, pp. 360–363.

[7] S. Cho, W.-S. Kim, N.-J. Paik, and H. Bang, “Upper-limb function assessment using vbbts for stroke patients,” *IEEE computer graphics and applications*, vol. 36, no. 1, pp. 70–78, 2015.

[8] O. N. Zestas, D. N. Soumis, K. D. Kyriakou, K. Seklou, and N. D. Tselikas, “The computer vision box & block test in rehabilitation assessment,” in *2022 Panhellenic Conference on Electronics & Telecommunications (PACET)*. IEEE, 2022, pp. 1–4.

[9] H. P. J. Dutta, M. K. Bhuyan, D. R. Neog, K. F. MacDorman, and R. H. Laskar, “A hand gesture-operated system for rehabilitation using an end-to-end detection framework,” *IEEE Transactions on Artificial Intelligence*, pp. 1–11, 2023.

[10] G. Everard, Y. Otmane-Tolba, Z. Rosselli, T. Pellissier, K. Ajana, S. Dehem, E. Auvinet, M. G. Edwards, J. Lebleu, and T. Lejeune, “Concurrent validity of an immersive virtual reality version of the box and block test to assess manual dexterity among patients with stroke,” *Journal of neuroengineering and rehabilitation*, vol. 19, no. 1, p. 7, 2022.

[11] K. Koter, M. Samowicz, J. Redlicka, and I. Zubrycki, “Hand measurement system based on haptic and vision devices towards post-stroke patients,” *Sensors*, vol. 22, no. 5, p. 2060, 2022.

[12] “Sollerman Hand Function Test-physio-pedia.com,” https://www.physio-pedia.com/Sollerman_Hand_Function_Test, [Accessed 27-11-2023].

[13] O. N. Zestas and N. D. Tselikas, “Sollerman hand function sub-test ldquo;write with a penrdquo;: A computer-vision-based approach in rehabilitation assessment,” *Sensors*, vol. 23, no. 14, 2023.

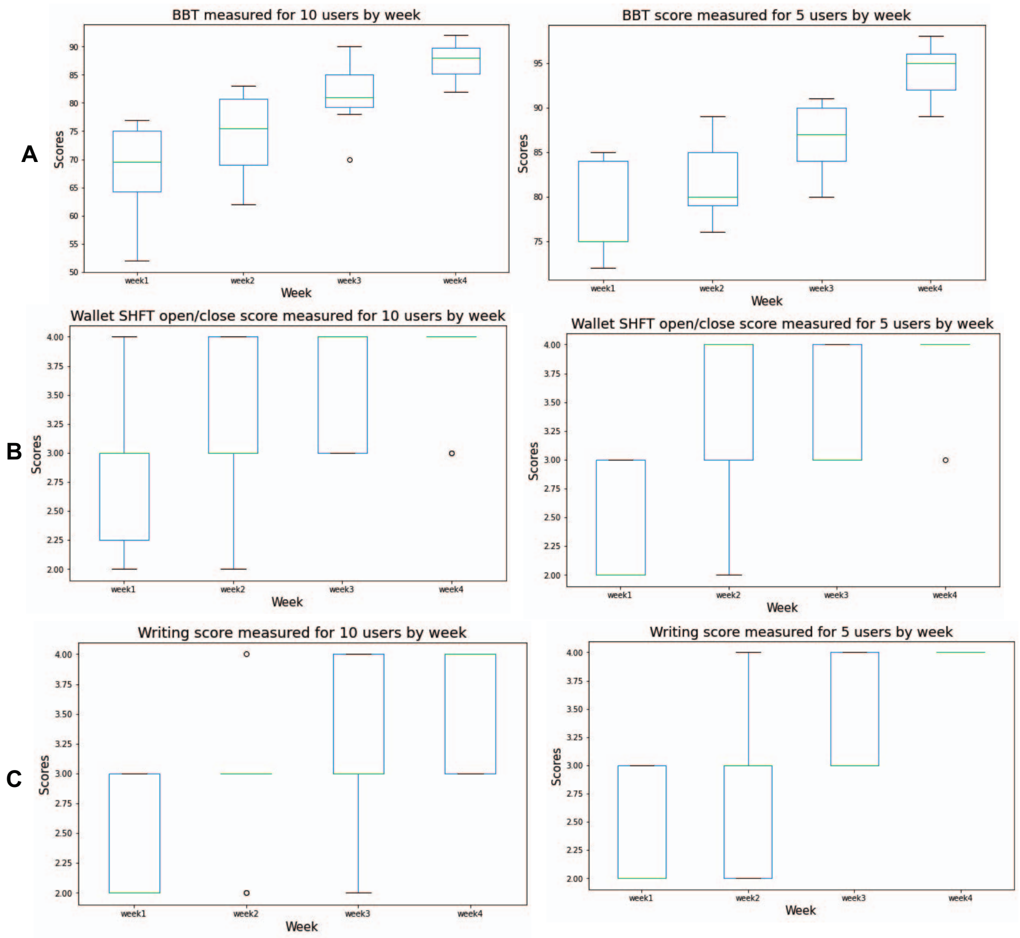

Fig. 8: Box plot for 10 male (1st column) and 5 female (2nd column) users measured for different rehabilitation exercises.

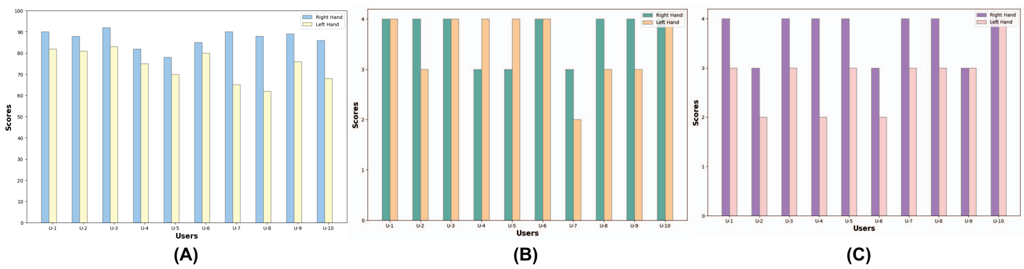

Fig. 9: Bar plots for different rehabilitation exercises showing the performance of the right and the left hand— (A) for BBT, (B) for SHFT wallet zip open/close, and (C) for SHFT writing test.