IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 17 March 2023, accepted 30 March 2023, date of publication 11 April 2023, date of current version 20 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3266245

APPLIED RESEARCH

STPT: Spatio-Temporal Polychromatic Trajectory Based Elderly Exercise Evaluation System

RIAD AHMED, RAIYAAN ABDULLAH, AND LAFIFA JAMAL, (Senior Member, IEEE)
Department of Robotics and Mechatronics Engineering, University of Dhaka, Dhaka 1000, Bangladesh

Corresponding author: Lafifa Jamal (lafifa@du.ac.bd)

This work was supported by a Special Grant in Research from the Information and Communication Technology Division, Government of the People's Republic of Bangladesh and the University of Dhaka.

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT This paper introduces an elderly exercise evaluation system. To determine the quality of a performed exercise, the authors propose a novel system to generate and use Spatio-Temporal Polychromatic Trajectory (STPT) images. Usually, the elder people need to perform some exercises or take physiotherapy in order to stay healthy both physically and mentally. It becomes difficult to evaluate the quality of their exercise routine without the aid of a trained physiotherapist. The system aims to overcome this problem by allowing elders to record their exercise videos using an easy-to-use Graphical User Interface and evaluate the results. A dataset of 109 subjects performing four types of shoulder exercises several times was created. The videos are labelled as correct or incorrect and an STPT image is generated from each video. Using our newly introduced method, the movement of the elder person is projected into an image which can be input to a Convolutional Neural Network (CNN). The dataset is further augmented to increase accuracy. Using our proposed method, the best model achieved an F1 Score over 90% in three of the four exercises. The CNN is trained based on these clips and the models are added to the backend of the interface. The proposed system requires only an ordinary camera and a computer with an entry level GPU allowing it to be deployed at a large scale.

INDEX TERMS Elderly assistance, exercise quality evaluation, deep learning, spatio-temporal polychromatic trajectory.

I. INTRODUCTION

The number of people aged 60 or above, more commonly referred to as senior citizens, is on the rise globally. Currently, they consist of around 9.3% of the world population [1]. World Health Organization projects that the elderly population will rise to 22 percent by 2050 [2]. Senior citizens become weaker as they age, and their body strength deteriorates swiftly if they do not maintain a proper exercise routine. It becomes necessary for a caregiver to look after them or aid them in certain tasks. As the number of elderly persons is increasing, it will be more problematic to find people to assist them. If they perform exercise regularly, it will aid them both physically and mentally. Many of them also require physiotherapy to maintain their vigour. However, it is difficult for them to evaluate if they have performed the exercise or taken physiotherapy correctly.

Usually, a trained physiotherapist is required to evaluate the quality of an exercise routine. To do this, the elderly person will have to visit a physiotherapy centre and pay a fee. He/She also has to take the hassle of commuting which is difficult at this age. If the person wants to hire a physiotherapist at home, that also comes with a huge cost. Also, it is not guaranteed that there will be a sufficient number of physiotherapists to cater to the needs of all the elderly persons in a particular area. The problem of evaluating exercise performance without the presence of a trained professional needs to be resolved. This can be achieved by training

The associate editor coordinating the review of this manuscript and approving it for publication was Senthil Kumar.

an intelligent system which can provide a basic form of detection. The system should be affordable and able to be manufactured at scale. As seen in the work [3], a home device can be very useful to elders. If we develop a home device to monitor and evaluate the exercise of elderly people, it will greatly benefit them.

Intelligent systems consist of technologies such as the Internet of Things [4] (IoT), Artificial Intelligence [5] and Computer Vision [6]. This work aims to deploy a system that can combine these technologies and deploy a device capable of detecting the quality of the exercise. The proposed system requires just an ordinary camera (e.g. webcam, phone camera) and a computer with minimal GPU acceleration. To create such a detector, a good number of sample videos of elderly persons exercising need to be collected first. For activity recognition, the body's skeletal joints need to be detected first. From the movement of the joints, it would be possible to look further into the quality of the performed exercise. In the proposed method, the relevant keypoints of the body are detected, and a Spatio-Temporal Polychromatic Trajectory (STPT) is plotted using the points. For example, if someone moves their hand from a downward position to extending it on one side and later to the top, a semicircular motion is seen in the image. The system can be trained to differentiate between correct and incorrect exercises. The proposed method uses Convolutional Neural Network (CNN) [7] to differentiate between these STPT images, thus turning this into a Computer Vision problem. The system also contains a graphical interface, enabling the elderly person to easily select which exercise he/she wants to perform and follow on-screen instructions. The software will take his video footage, extract keypoints, generate the STPT image and evaluate his exercise allowing the elderly person to evaluate his exercise without putting much effort.

The main contributions of this work are:

- **Determination of suitable exercises for elders**
  There are many types of exercises. Usually, physiotherapists suggest four general shoulder exercises, which are easy to perform by an elderly person and have less risk of injury.
- **Creating a dataset for correct and incorrect exercises**
  To the best of our knowledge, there is no dataset for the mentioned shoulder exercises that evaluates the exercise's quality. Thus, a dataset was constructed from scratch, which can be used for further research in this area.
- **Training an intelligent system to distinguish the type of exercise**
  Using the STPTs, a Convolutional Neural Network (CNN) is trained to detect if the performed exercise is correct or incorrect.
- **Creating an application using the trained model**
  For ease of usage, the trained model is deployed with an interface that can be operated easily to perform the exercise and receive the results.

The rest of the paper is organized as follows - Section II discusses the existing works and the motivation for the proposed system. Then the methodology of the proposed system is discussed in Section III. The results are discussed in Section IV. Finally, Section V concludes the paper with a comprehensive discussion and future scopes.

II. LITERATURE REVIEW

Elderly people become weak as they age. Physical activity can contribute both to physical and mental health and keep them strong. It can also prevent many diseases and make their daily life better. However, it is tough for elderly people to evaluate whether they have performed their exercises properly. This is why activity recognition is required to evaluate whether an exercise is correct or incorrect. Some of the related works are discussed in this section.

In the work [8], Zhi et al. developed a computer vision system that automatically detects compensatory motions by users of robotic stroke rehabilitation therapy. They worked with nine stroke survivors and ten healthy patients. They worked on three compensatory motions - Shoulder elevation, Trunk rotation and Lean forward. Their approach requires a depth camera (Microsoft Kinect) for detection purposes.

Lee et al. propose an automated method [9] to access stroke rehabilitation exercises both qualitatively and quantitatively. They also used a Kinect sensor to take the joint positions and extract features such as angle, trajectory, speed, acceleration etc. Then these features are used by binary classifiers to detect the movement quality and generate a performance score. These were evaluated later by professional therapists. However, these approaches are specially designed for stroke rehabilitation patients but not for exercises performed by a general elderly person.

Chen and Yang developed a Pose Trainer [10]. The trainer's job is divided into two stages. The first one detects the body pose of the person correctly, and the next task determines how well the poses are. OpenPose [11] is used to detect the body pose. The OpenPose model returns 18 keypoints of the body & a corresponding confidence score, and then the list of keypoints is fed into the machine learning model to estimate the pose. The model tells how good the pose is. For demonstration purposes, they fed a trimmed video clip into the model and got the result correctly. A normalization process is introduced to remove the distance from camera ambiguity. Their posture evaluation is designed for exercises that are more common for younger people. Also, in their second approach, they worked with time-series data using dynamic time warping (DTW). This is prone to noisy keypoints.

The detailed explanation of Dynamic Time Warping (DTW) can be found in [12]. In this book, many algorithms and applications are mentioned that are obtained from DTW. The scopes and modeling are also presented which correspond to motion retrieval. All of the body keypoints are not equally important in making the decision in action recogni-

TABLE 1. Comparison between various research.

Work	Types of motions/exercises	Sensor	Target User	No. of participants in dataset	Classifying Method
[8]	4	Kinect Sensor	Post-stroke survivors	19	SVM, RNN
[9]	3	Kinect Sensor	Post-stroke survivors	26	DT, LR, SVM, NN, LSTM
[10]	4	Ordinary Camera	Young people	Not mentioned	Geometric Evaluation, DTW
Proposed Method	4	Ordinary Camera	Old people	109	CNN

tion. In this paper [13], the authors used feature weighting in conjunction with DTW to determine which joint is most important in which action. DTW is used to measure the difference in features returned by two models.

The paper [14] focuses on recognizing specific actions such as waving, punching and clapping using an exemplar-based sequential signal-layered approach that utilizes Dynamic Time Warping (DTW). To improve recognition rates, a depth camera is used to track body parts and recover joint information in 3D real-world coordinates. This paper [15] proposes a method for recognizing human actions using 3D skeletal models. They used a feature called CF, based on relative angles between body parts, which is refined by OBRF (Observational Reduction Feature) and TVRF (Time Variance Reduction Feature). This study [16] describes a method for recognizing actions using skeletal data trained from depth images. The proposed approach treats motion sequences as multi-dimensional time series and uses the Dynamic Time Warping (DTW) metric to map the time series to a vector space. In order to categorize human activities utilizing wearable sensors on both actual data and synthetic data, this research [17] developed a template selection strategy based on Dynamic Time Warping (DTW). The author of this work [18] put up a technique for identifying human activity based on the self-similarities of activities and dynamic time warping (DTW). The proposed approach first calculates self-similarities between adjacent action sequence frames, chooses a reference sequence that perfectly captures the action's essential shape and motion, applies DTW to compare each reference sequence to the test sequence, and finally uses DTW as an action recognition feature. The study [19] suggested a technique for identifying human activity that combined Dynamic time warping and Laban movement analysis (LMA). In this paper [20], the authors have used Convolutional Neural Networks (CNN) alongside Support Vector Machine (SVM) on the time series data for exercise performance evaluation.

Based on the works discussed, it can be observed that there are many approaches to recognizing motion and assessing the quality of exercises or detecting compensation. However, there seems to be no dataset that specifically assesses the quality of exercise performed by elderly persons. Also, some of the works [8], [9] depend on an external sensor like Kinect. Our goal was to make the system widely available; this would be possible by ensuring the system would work on entry-level webcams or cheap IP cameras. As long as the video has a resolution of at least 1280×720 pixels, sufficient data will be available for processing. This setting is now supported in most types of cameras found in the market. Furthermore, it was decided to use Openpose [11] similar to Pose Trainer [9]. After taking the keypoint data from Openpose, the system will generate our proposed Spatio-Temporal Polychromatic Trajectory (STPT) images. Instead of using time series data, we explore this alternate method of training a CNN model using our newly introduced STPT images. This allows for augmentation of the existing dataset and is more robust to noisy keypoints. A comparison of the above-mentioned works is shown in Table 1.

The proposed method is discussed in detail in the following section.

III. METHODOLOGY

Older people need to do regular physical exercise to stay fit. Some of them try to perform exercises but are unable to evaluate their results. The proposed system can take their exercise videos as input and detect if it is correct or incorrect. The main challenge was to select exercises appropriate for older persons. A dataset comprising exercise videos with perfect and imperfect postures was created. As seen in Figure 1, from the frames of these videos in 1(a), the body keypoints needed to be extracted 1(b). We propose a novel system to generate Spatio-Temporal Polychromatic Trajectory (STPT) 1(c) image from these frames. The images are put through the detection network 1(d) to determine the exercise quality 1(e). The steps are:

A) Representation of body key point data
B) Exercise selection & data collection
C) Process of creating Spatio-Temporal Polychromatic Trajectory
D) Analyzing Spatio-Temporal Polychromatic Trajectory for different exercises
E) User interface for exercise feedback

A. REPRESENTATION OF BODY KEY POINT DATA

A convenient way is required to make the computer understand the change of body positions in each frame of the input video. The first approach that came to mind is feeding the video frames as images into a Convolution Neural Network. But in this case, as multiple frames need to be input at once to understand the sequential data, the input matrix will be huge and computationally expensive. The body positions need to be expressed using less memory. As a result, a model named OpenPose [11] is used, which can be considered the state-of-the-art approach for real-time human pose estimation. The

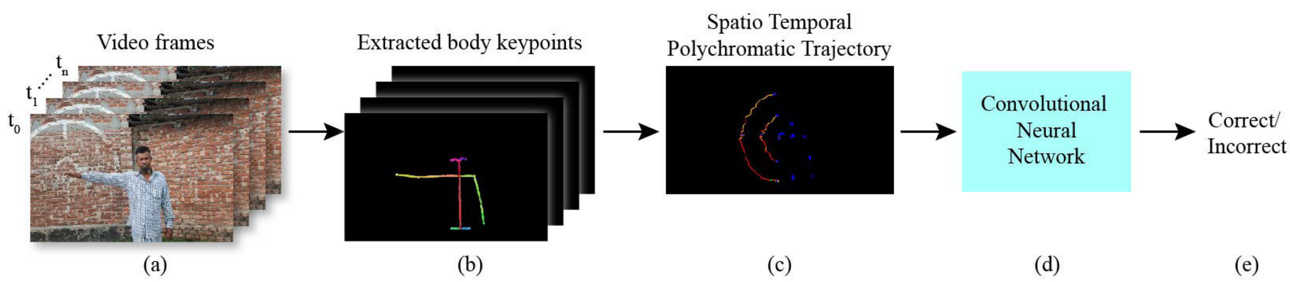


FIGURE 1. Pipeline for processing a video.

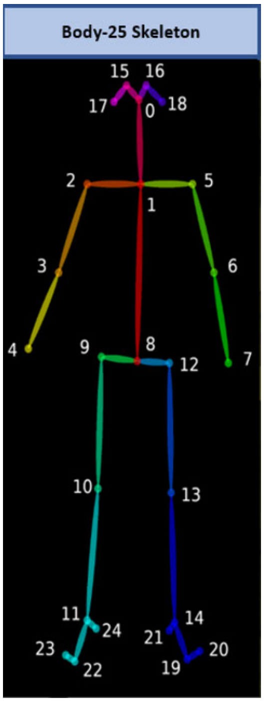


FIGURE 2. Keypoints for Body_25.

code base is well-documented and open-sourced on Github. Openpose is originally written in C++ and Caffe. It creates a skeleton representation by detecting keypoints of the body in an image.

The Body_25 output format by OpenPose [11] was selected for this task. In this format, there are 25 different body keypoints. Figure 2 shows the different body parts with their assigned ID in Body_25. They are described in Table 2.

This output format is used in the proposed system, which enables pose detection. Some examples are given in Figure 3 of running Openpose on the images obtained from both our video feed and from online to demonstrate its capability on complex backgrounds and varying lighting conditions.

B. EXERCISE SELECTION AND DATA COLLECTION

The authors visited The Physical Therapy and Rehabilitation Centre (PTRC) in Uttara, Dhaka and Dhaka City Physiotherapy Hospital in Jigatola, Dhaka, to determine which exercise routines would be most suitable for elderly people. After discussing with various physiotherapists, four shoulder-related exercise routines were selected:

TABLE 2. Body_25 output format.

Body part	Index	Body part	Index
Nose	0	Left Knee	13
Neck	1	Left Ankle	14
Right Shoulder	2	Right Eye	15
Right Elbow	3	Left Eye	16
Right Wrist	4	Right Ear	17
Left Shoulder	5	Left Ear	18
Left Elbow	6	Left Big Toe	19
Left Wrist	7	Left Small Toe	20
Mid Hip	8	Left Heel	21
Right Hip	9	Right Big Toe	22
Right Knee	10	Right Small Toe	23
Right Ankle	11	Right Heel	24
Left Hip	12	Background	25

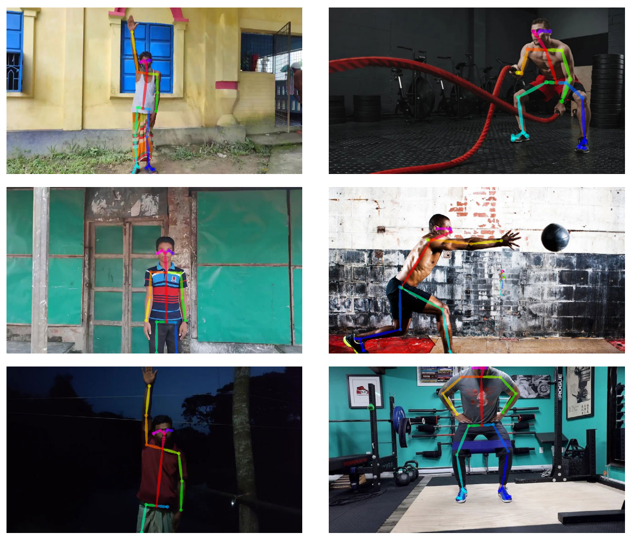


FIGURE 3. Openpose results on various lighting and environments.

1) **Arm flexion and extension**
   Steps:
   a) Person extends hand forward
   b) Raises hand 180 degrees upward

(a) Extending arm in the forward direction (b) Reaching 180 degrees upward

FIGURE 4. Demonstration of Arm flexion and extension exercise.

(a) Extending hand sideways (b) Reaching 180 degrees upward

FIGURE 5. Demonstration of Arm abduction and adduction exercise.

   c) Retracts back down
   d) Repeats steps a-c three times
2) **Arm abduction and adduction**
   Steps:
   a) Person extends hand sideways parallel to the ground
   b) Raises hand 180 degrees upward
   c) Retracts back down
   d) Repeats steps a-c three times
3) **Arm lateral and medial rotation**
   Steps:
   a) Person sets the forearm forward parallel to the ground
   b) Rotates hand 45 degrees to the side
   c) Retracts back in the same way
   d) Repeats steps a-c three times
4) **Arm circumduction**
   Steps:
   a) Person rotates the hand 360 degrees keeping it straight and perpendicular to the ground
   b) Repeats step a three times

The physiotherapists demonstrated these four types of exercises as seen in Figures 4, 5, 6 & 7 and instructed how to perform them correctly. They also showed the possible incorrect ways an elderly person might do these exercises.

(a) Set forearm forward (b) Rotate 45 degrees to the side

FIGURE 6. Demonstration of Arm lateral and medial rotation exercise.

(a) Circumventing hand 360 degrees - i (b) Circumventing hand 360 degrees - ii

FIGURE 7. Demonstration of Arm circumduction exercise.

TABLE 3. Age category of subjects.

Age category	Number of subjects
Below 25 years	12
25-60 years	54
Above 60 years	43

Along with the examples demonstrated by them, similar exercises performed by 109 other subjects were recorded. As it is rare to find lots of people aged 60 or above who are willing to perform exercises, middle-aged men and young boys were also requested to do these exercise steps. The number of people in each age category is seen in Table 3. All of the subjects of the dataset are males. They provided their consent to have their recorded video used in our dataset. Many females were contacted as well. However, they were hesitant to perform exercises in front of the camera due to social norms. Since we could not obtain their consent, we did not include them in our dataset. The length of the videos ranges from 1 to 19 seconds. These clips are categorized as "Correct" or "Incorrect". In our practical experience, old people contributed more to the incorrect clips; further proving our motivation.

A snapshot of the dataset is provided in Figure 8.

C. PROCESS OF CREATING SPATIO-TEMPORAL POLYCHROMATIC TRAJECTORY

After capturing the videos, all of them were organized, trimmed and labelled correctly. They are all converted to

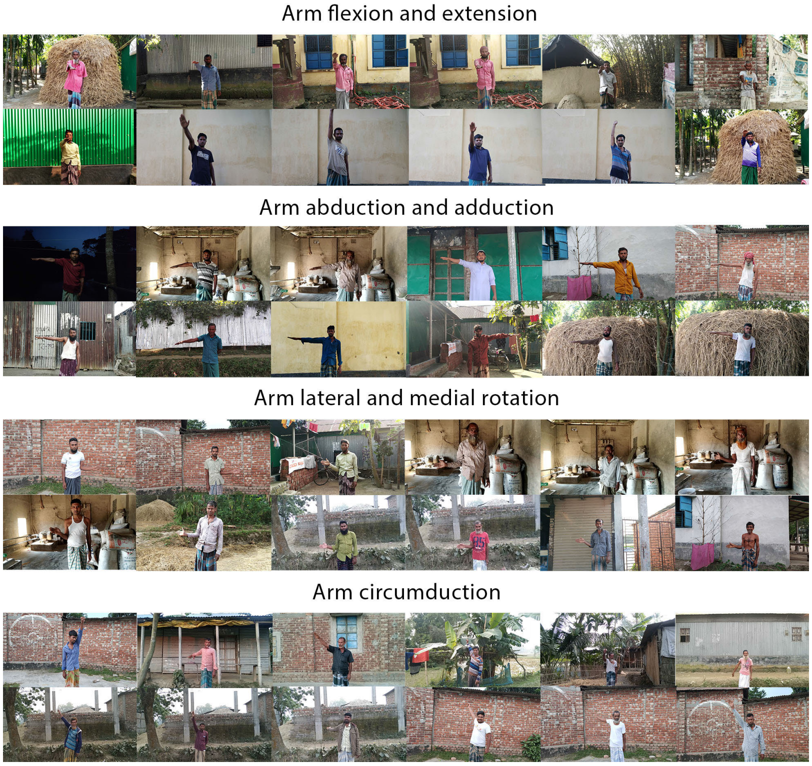


FIGURE 8. Exercise performed by various subjects.

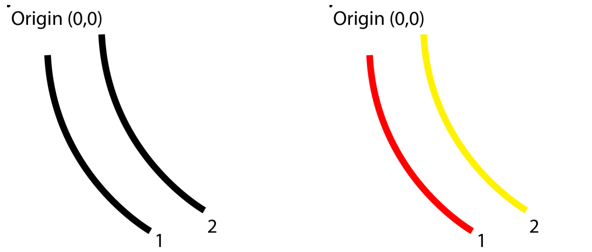
the dimensions 1280 × 720 to achieve consistency. Then the videos are fed into Openpose [11]. Openpose analyzes each video frame and outputs the body key point positions (pixel value on X-axis and Y-axis) into a JSON (JavaScript Object Notation) file. The number of JSON files for each video depends on its duration (number of frames).

After getting all of the JSON files from Openpose [11], the important keypoints of the video were placed in an image creating a Spatio-Temporal Polychromatic Trajectory (STPT). The body keypoints were used to extract features that would be helpful for the detection of quality. Though Openpose returns 25 body keypoints, it can be observed that only the keypoints of the head, neck, mid-hip, shoulders, elbows and wrists are enough for arm-related movements. The shoulder exercises are performed mainly by the body's upper portion, and the lower portion does not move. So only these points are used to evaluate the exercises.

When a person performs the exercise, his limbs create a series of motions. In the case of exercises related to the arms, the arm has to be moved from starting point to the ending point following specific guidelines. When it moves, the key point also moves in each frame. Since each exercise has a different set of movements for the body joints, the motion of the keypoints perfectly reflects the task of the exercise. The proposed method aims to draw the movements of limbs and use this as a learning technique for the neural network to detect the exercise quality perfectly. For example, the movements of joint 1 are drawn in Figure 12. From the first frame, the location of the joint is received. In the second frame, the new location of the same joint is received and so forth. After getting the keypoints frame-by-frame of a specific joint, a curve can be drawn from the movement.

More information about the movement can be provided to the network by incorporating direction. However, if the curve consists of one colour, it is difficult to interpret the direction of the movement, as seen in Figure 9(a). Curves 1 and 2 both seem similar. The curves can be made polychromatic by defining different colours for the arm moving in the bottom-right direction (Curve 1) and top-left direction (Curve 2) as seen in (b). The top-left corner of the screen is denoted by (0,0) so that the positive x-direction is to the right, and the positive y-direction is at the bottom. The four possible cases are demonstrated in Table 4 with their respective colours. Instead of a single coloured line, this provides a better intuition in the image of how the motion has occurred. Thus polychromatic trajectories help the network better understand the quality of the movement.

It can be observed from Figure 10(a), the subject started his movement of joint 1 from (x_0, y_0). According to the instructions, the hand should remain straight and within the same plane throughout the exercise. When the hand is horizontal position with respect to the ground, the position of joint 1 is indicated as (x_1, y_1). As the origin(0,0) is at top left corner,



(a) Monochromatic Trajectory (b) Polychromatic Trajectory

FIGURE 9. Difference between monochromatic and polychromatic trajectory.

TABLE 4. Direction colour chart.

Direction	x value	y value	colour
bottom-right	+	+	red
top-right	+	-	green
bottom-left	-	+	orange
top-left	-	-	yellow

here: $x_0 > x_1, x = x_1 - x_0 =$ negative. $y_0 > y_1, y = y_1 - y_0 =$ negative. According to Table 4, if $x =$ negative, $y =$ negative; then colour = yellow and direction = top-left.

In Figure 10(b), the starting point is (x_0, y_0) which is the horizontal position with respect to the ground. The ending position is (x_1, y_1). Here: $x_0 < x_1, x = x_1 - x_0 =$ positive. $y_0 > y_1, y = y_1 - y_0 =$ negative. According to Table 4, if $x =$ positive, $y =$ negative; then colour = green and direction = top-right.

In some specific exercises, stopping the hand in a particular location is mandatory. It is also required to recognize that only the wrist and elbow joints are moving. The remaining joints should be standstill. However, it is not possible to determine the halting points by using single-coloured trajectories. In our trajectory generation algorithm, we determined the halt points accurately and plotted those points with a separate distinctive colour (blue). In Figure 11, the monochromatic trajectories look similar, although there were some stopping positions within the trajectory. But on the right side, it can be noticed that the halt points are detected correctly and coloured blue. And there is a colour changing from yellow to green because the colours change with the direction, for instance, transitioning from top-left to top-right.

The process followed by the system is given below:

i) Creates a blank black image
ii) Draws a line on the image using two adjacent points returned by Openpose from the exercise video. The colour of the line depends on the direction of the movement.
iii) If a key point stays around the same area for a series of frames then marks that point with a blue circle.

D. ANALYZING SPATIO-TEMPORAL POLYCHROMATIC TRAJECTORY FOR DIFFERENT EXERCISES

As different exercises have different series of motions, these exercises will generate different-looking STPTs. For the "Arm lateral and medial rotation" exercise, joint 1 is allowed to move. The forearm should be kept straight. Joint 2 is allowed to move as well for the other exercises. The hand should remain straight for those. It is imperative to keep the other body parts still. When the person is able to maintain this and perform the exercise while keeping a camera in front of him, his STPT image corresponds to the correct samples. But if he/she does not maintain the guideline properly, the STPT images have irregular and arbitrary shapes, as seen in the incorrect samples. In those images, many random movements of the hand and body can be seen, which are prohibited for exercise. (The STPT images demonstrated below have been cropped to highlight the regions of interest.)

1) ARM FLEXION AND EXTENSION

In this exercise, the hand moves in the forward direction of the body. Figure 12 is showing two movement lines coloured by yellow (top-left and top direction) and green (top-right direction) which is the first part of the exercise. The still points are blue dots where the hand (keypoints of wrist, elbow and shoulder) remained in the same area for some time.

Sample STPT images of the exercise can be seen in Figures 13 and 14. In the correct samples, there is a vertical line from the starting point to the ending point, while in the incorrect samples, there are deviations.

2) ARM ABDUCTION AND ADDUCTION

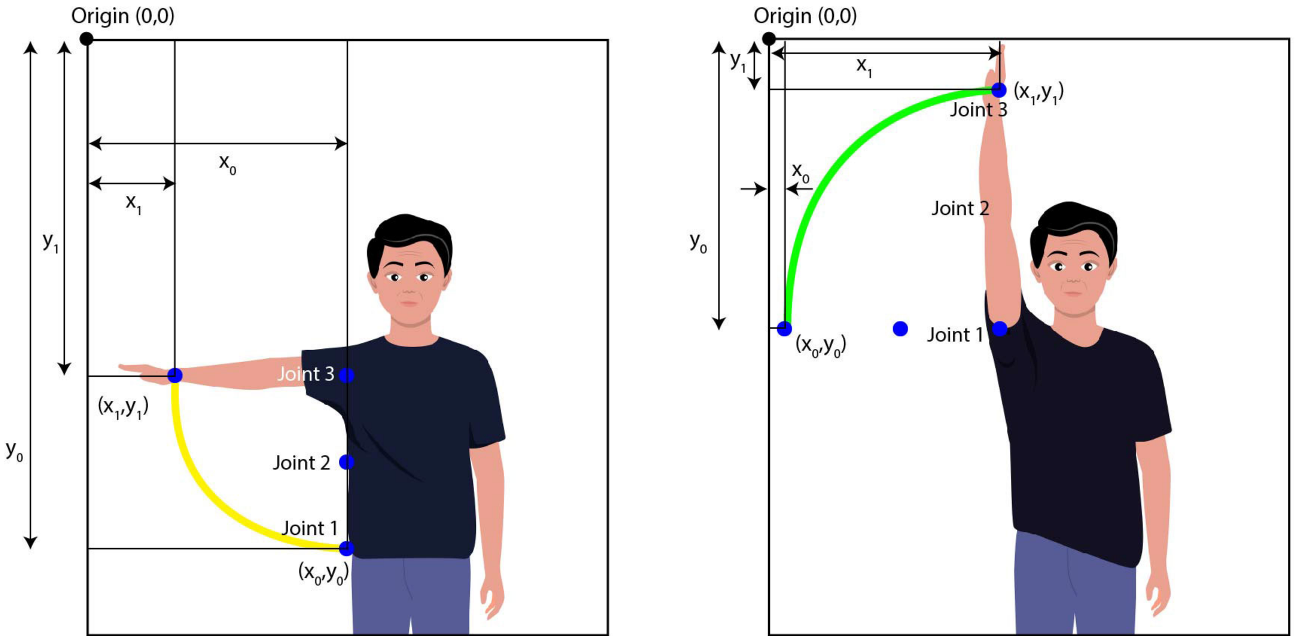
In this exercise it can be observed from Figure 15 that there are two movement curves coloured by yellow (top-left direction) and green (top-right direction). The keypoints marked with blue dots remained still for some time. Figure 15(c) represents the end of the first stage of this exercise as the hand reaches the top. From the image samples demonstrated in Figures 16 and 17, it can be seen if the hand is not properly moved sideways, the image exhibits irregular lines and abrupt patterns.

3) ARM LATERAL AND MEDIAL ROTATION

This exercise mostly concerns the movement of the forearm. Figure 18 shows three movement lines. In Figure 18(b) it is coloured yellow (top-left direction). Figure 18(c) is coloured orange (bottom-left direction) and finally Figure 18(d) is shown in green (top-right direction). As seen in Figures 19 and 20, in the correct samples, a vertical line is first created towards the top, and then a horizontal line forms sideways. But irregular curving lines are displayed in the incorrect samples if the movement is not done properly.

4) ARM CIRCUMDUCTION

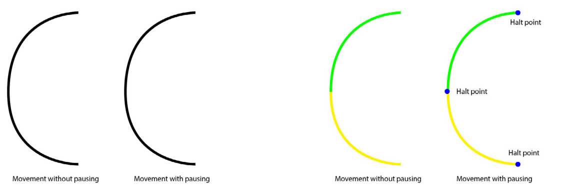
For the last exercise type, the hand might rotate either from the front of the person to the back or in the opposite path. It should be kept straight, and the rotation should maintain a proper angle perpendicular to the ground at every point, as observed in Figure 21. A well-shaped oval is generated in the STPT image for a perfect rotation, as seen in the correct



(a) Moving arm in the top-left direction

(b) Moving arm in the top-right direction

FIGURE 10. Demonstration of direction calculation.



(a) Monochromatic trajectory of arm movement without stopping and with stopping

(b) Polychromatic trajectory of arm movement without stopping and with stopping

FIGURE 11. Demonstration of Stopping points.

samples in Figure 22. Otherwise, the circle will be distorted, showing incorrect exercise characteristics in Figure 23.

From the discussion, it can be concluded that each STPT image contains the features of the videos, which will be beneficial for analyzing the movement of the videos. After getting all of the images of correct and incorrect exercise videos, these images were input into a Convolutional Neural Network (CNN) [7] to extract the features from the image and classify each clip.

5) TRAINING THE STPT IMAGES USING CNN

After generating the STPT image dataset by using all of the exercise videos, these images are fed into a CNN to classify which image is for the correct exercise and which one is incorrect. The tests were run on the CNN architectures - ResNet50_V2 [21], MobileNetV2 [22] and EfficientNetB0 [23].

As the dataset is relatively small and these images still have a complicated outlook because of the features, they caused some overfitting in the initial training. Data augmentation was applied to solve this problem.

List of augmentations:

1) **Horizontal flip:** Makes sure the dataset is balanced with both left and right-hand exercise footage.
2) **Random Zoom:** Takes into account the various distances of the user from the camera.
3) **Rotation:** Mimics the subtle changes in camera angle.

Adding the augmented images to the dataset increased its size by six-eight times and helped achieve better training results. The details of the training are discussed in the "Results section".

E. USER INTERFACE FOR EXERCISE FEEDBACK

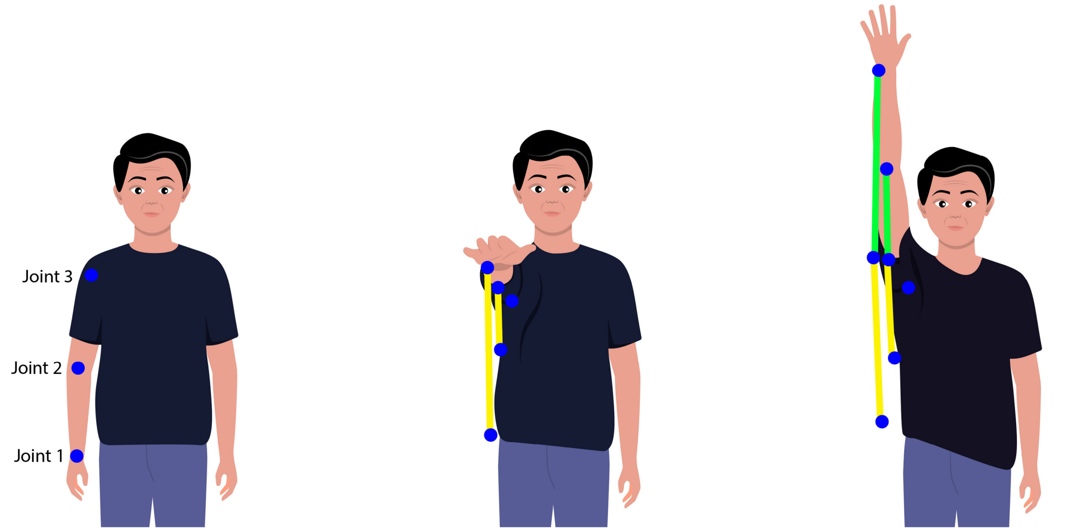
After training the models, the next step is to deploy them in a central system for exercise quality inference. A Graphical User Interface is created using PyQt5. The screens are discussed below:

1) MAIN SCREEN

As seen in Figure 24, the four exercises are shown side by side along with GIF animations of them being performed so that the elderly person can easily recall what to do in each exercise. The user can choose which exercise he/she wants to perform and click on "Start Exercise".

2) EXERCISE INSTRUCTIONS

Each exercise can consist of one or more steps, and the user is instructed on what to do in the first step. Figure 25 shows the instruction for "Arm circumduction".



(a) Initial position of hand (b) 90-degree position of hand (c) 180-degree position of hand

FIGURE 12. Motion curve shown over subject performing Arm flexion and extension exercise.

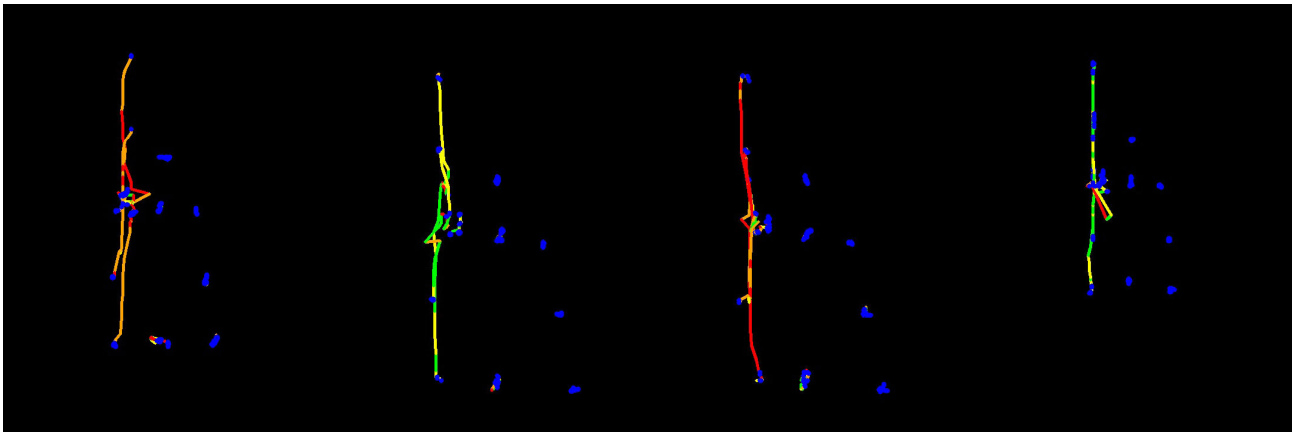


FIGURE 13. STPT images of correct samples of Arm flexion and extension exercise.

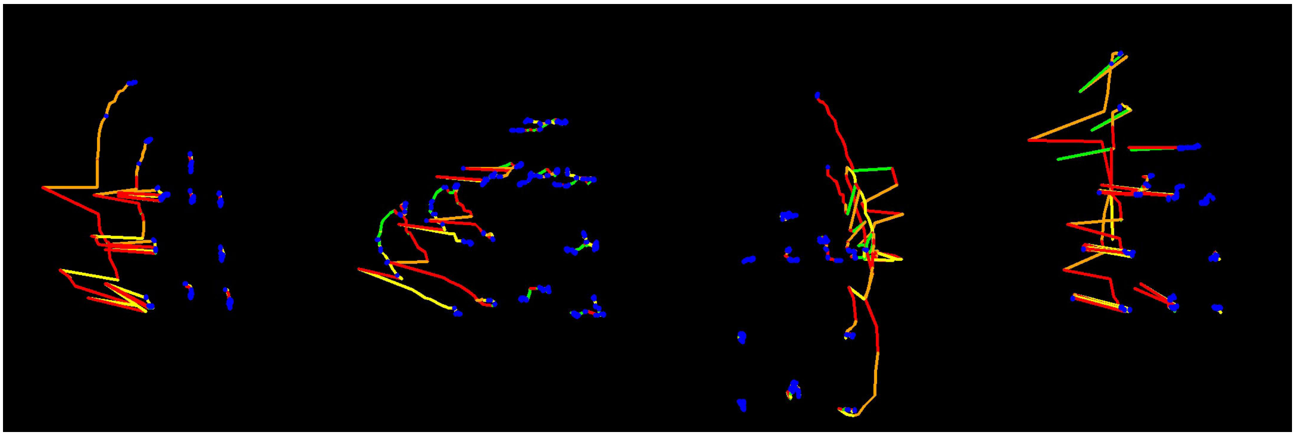
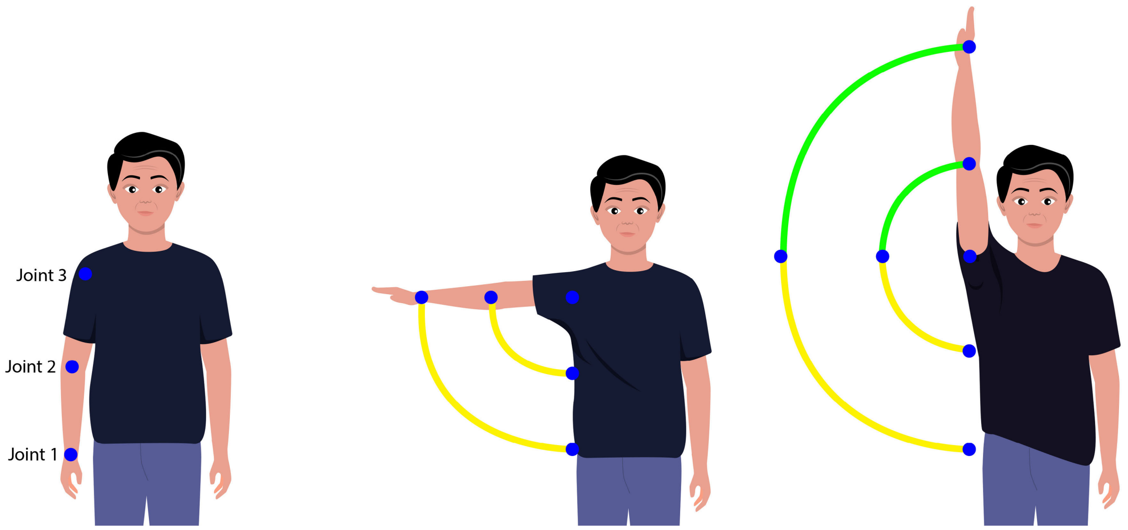


FIGURE 14. STPT images of incorrect samples of Arm flexion and extension exercise.

After reading the instruction, the user can click "Start". Then the system performs the steps shown in Figure 26. It uses the Pyttsx3 library for voice commands. The system calculates if $|y_{\text{left-wrist}} - y_{\text{left-wrist}}|$, $|x_{\text{left-shoulder}} - x_{\text{left-wrist}}|$ and $|x_{\text{right-shoulder}} - x_{\text{right-wrist}}|$ are under a threshold value to determine if the user's pose is similar to Figure 27. Once



(a) Initial position of hand (b) 90-degree position of hand (c) 180-degree position of hand

FIGURE 15. Motion curve shown over subject performing Arm abduction and adduction exercise.

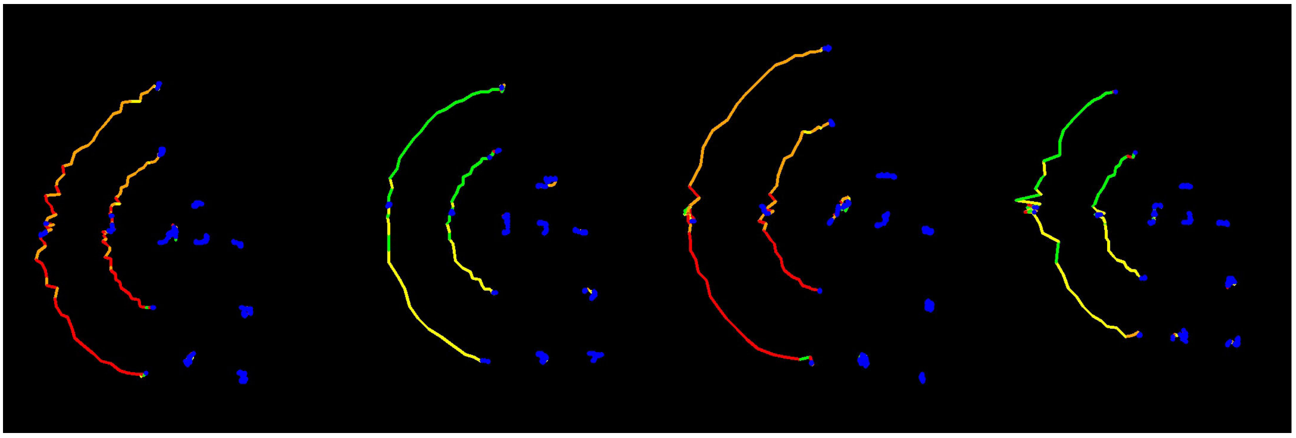


FIGURE 16. STPT images of correct samples of Arm abduction and adduction exercise.

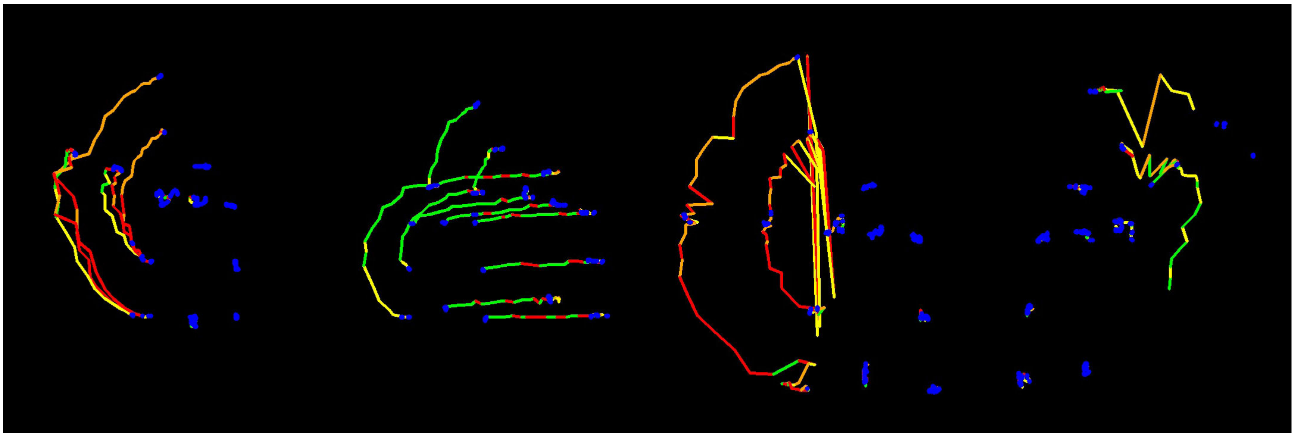
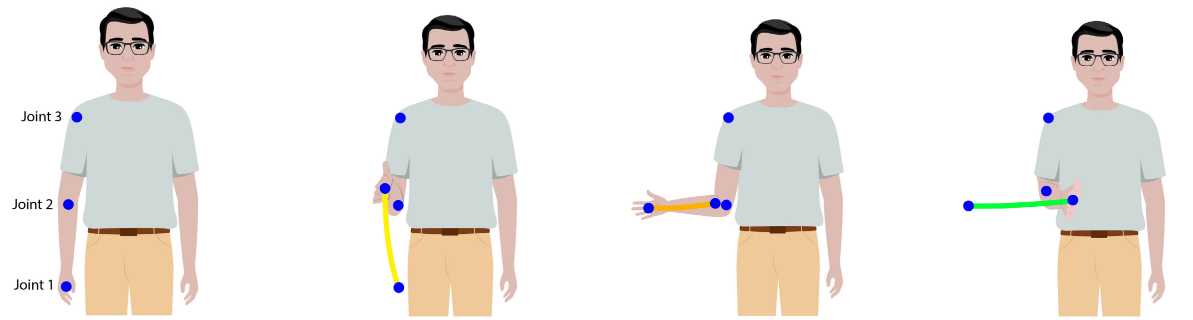


FIGURE 17. STPT images of incorrect samples of Arm abduction and adduction exercise.

this is confirmed, it starts the exercise and then creates Spatio-Temporal Polychromatic Trajectory (STPT) images. Finally, it infers the images to detect the exercise quality and provides feedback.

3) RESULTS SCREEN

If the user performs the exercise well and gets a correct score above 70% from the model, he/she is congratulated, mentioning his/her high score as shown in Figure 28 (a), while if



(a) Initial position of hand (b) Forearm set forward (c) Moved 45 degrees sideways (d) Forearm rotated back

FIGURE 18. Motion curve shown over subject performing Arm lateral and medial rotation exercise.

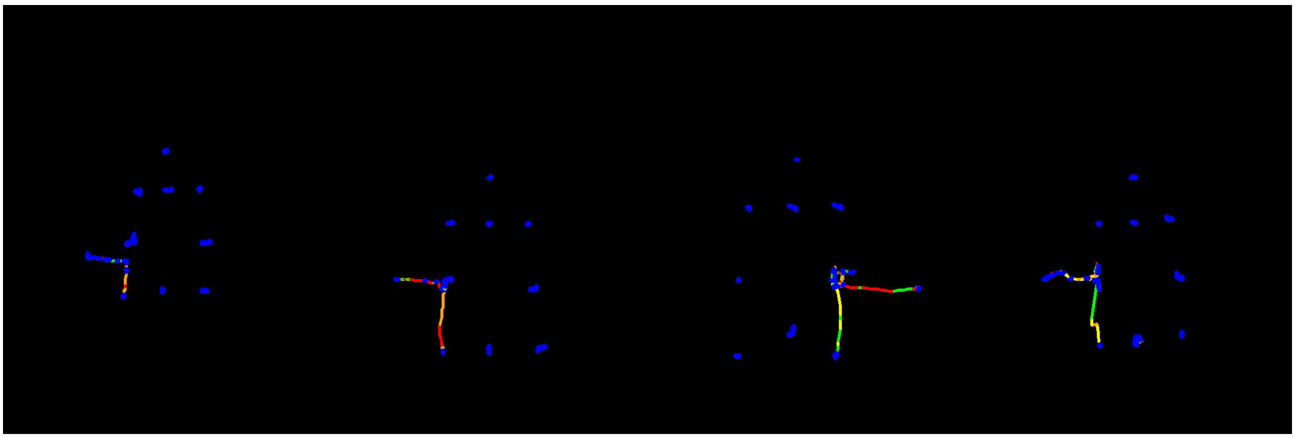


FIGURE 19. STPT images of correct samples of Arm lateral and medial rotation.

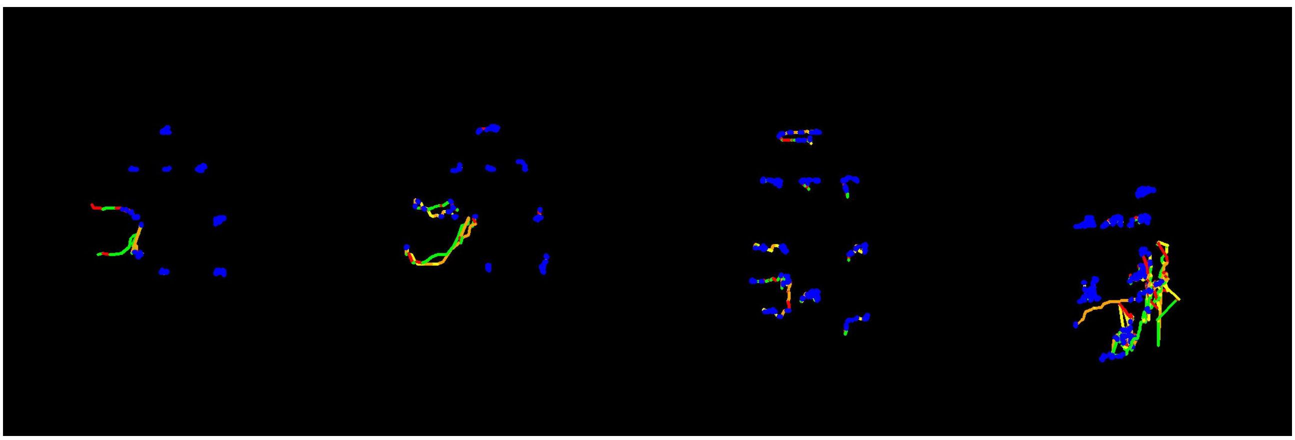
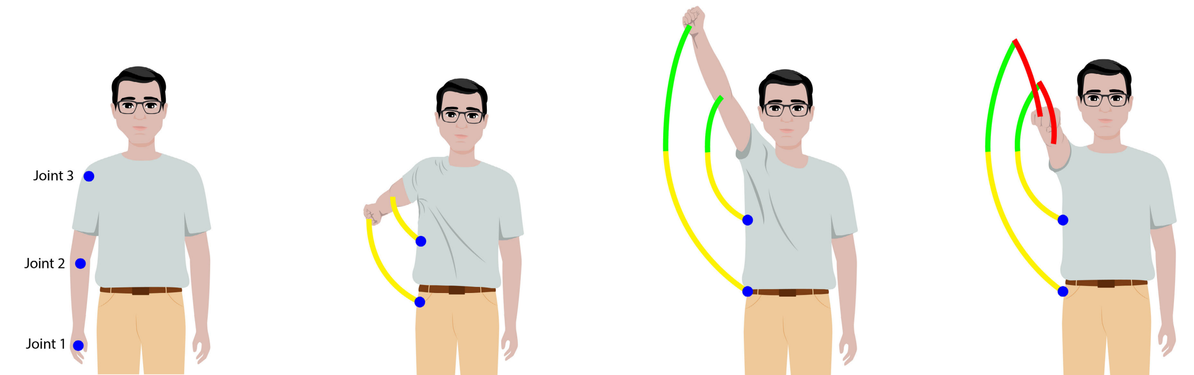


FIGURE 20. STPT images of incorrect samples of Arm lateral and medial rotation.

the exercise was not properly performed, it is mentioned that he/she got a low score and is given tips about how to improve it in (b).

4) INCORRECT EXERCISE METRICS

To further aid the person in amending his incorrect exercise, the JSON data of the keyframes are taken and the system uses these to calculate hand-crafted features to provide the user feedback. This helps in understanding what issues might be causing the user to get a low score. Based on these metrics, the system provides suggestions in the GUI at the end of the exercise on how to do well on the next try. However, these hand-crafted features are not defined explicitly in the STPT image and the calculation does not directly impact the



(a) Initial position of hand (b) Arm circumduction a (c) Arm circumduction b (d) Arm circumduction c

FIGURE 21. Motion curve shown over subject performing Arm circumduction exercise.

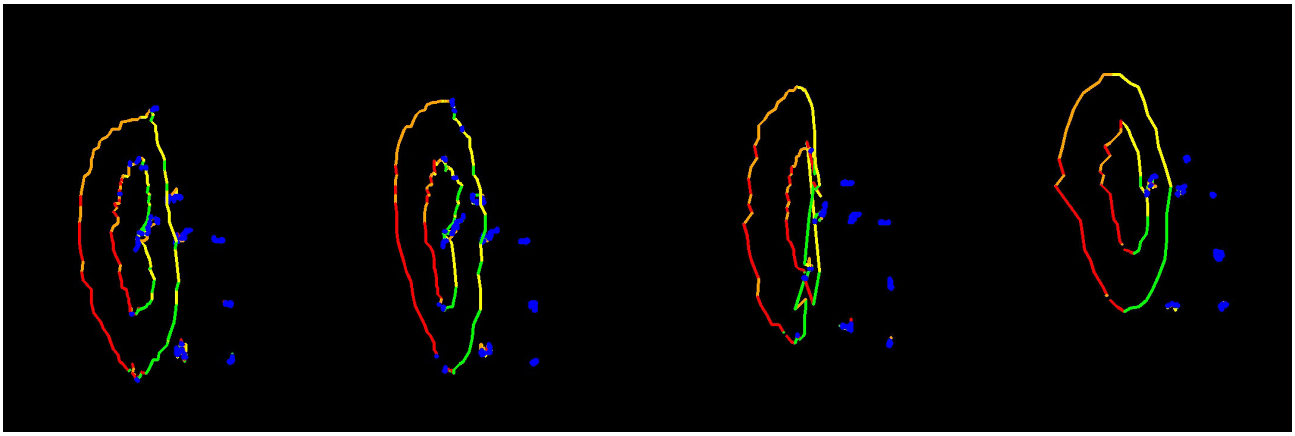


FIGURE 22. STPT images of correct samples of Arm circumduction exercise.

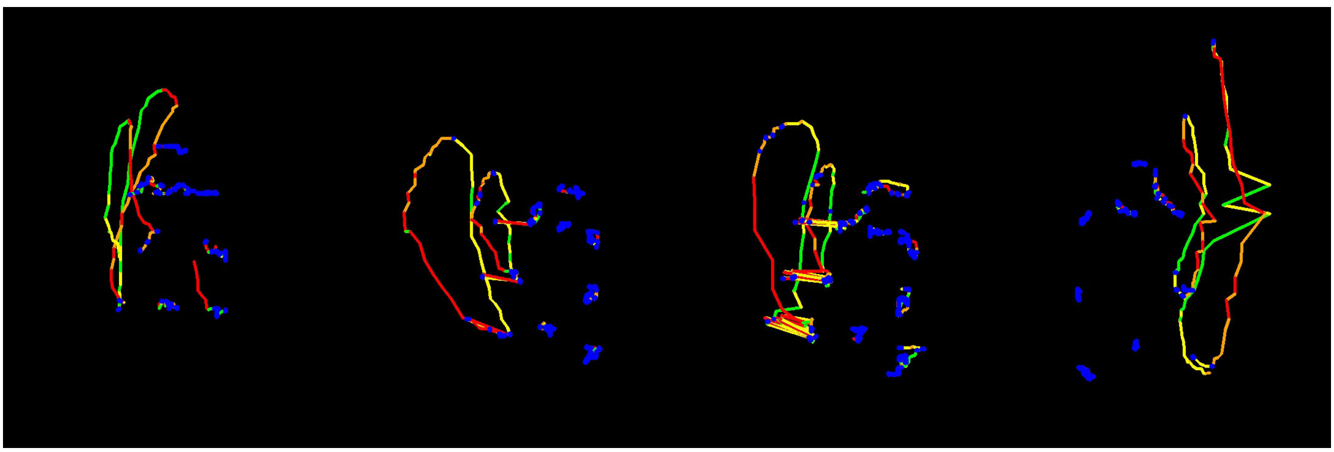


FIGURE 23. STPT images of incorrect samples of Arm circumduction exercise.

trained model's evaluation. The incorrect exercise metrics are analyzed as follows:

For "Arm flexion and extension", "Arm abduction and adduction" and "Arm circumduction" exercises, the system checks if the user's hand was bent by calculating how many frames were the points of the wrist, elbow and shoulder joints on the same line. The steps performed by the system are:

1) Takes the values x_{wrist}, y_{wrist}, x_{elbow}, y_{elbow}, x_{shoulder} and y_{shoulder}.
2) Calculates $\mathrm{Area}_{\mathrm{triangle}}$ formed by these points.

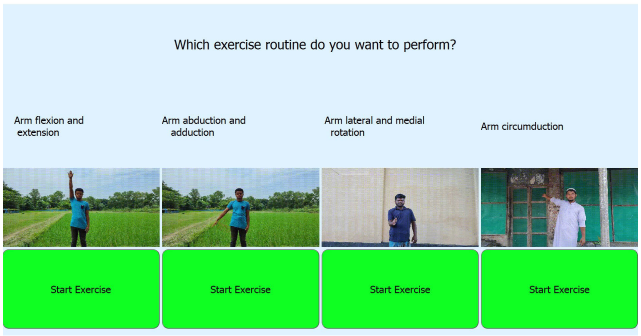


FIGURE 24. Welcome page of Exercise GUI.

FIGURE 25. Instructions on how to perform the first part of an exercise.

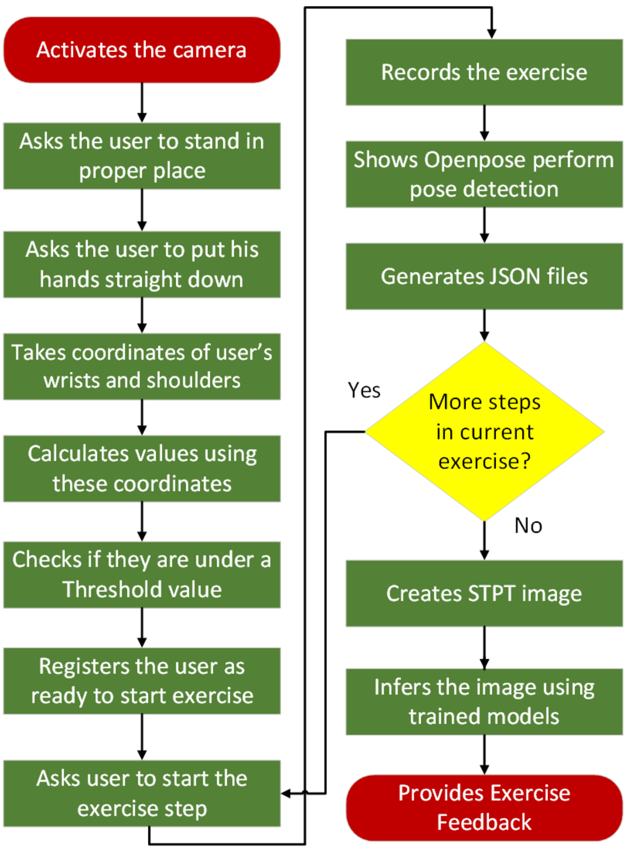


FIGURE 26. Exercise instruction steps.

3) Checks if $\text{Area}_{\text{triangle}}$ is below the threshold.
4) Calculates steps 1-3 for each hand in each frame and checks if the number of frames where the hand deviated is below the threshold.

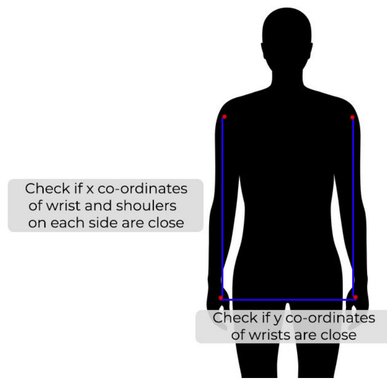


FIGURE 27. Checking if the user is in position to start exercise.

For the "Arm abduction and adduction" exercise, the system checks if the user's hand is parallel to the ground in the middle part of the exercise by calculating the difference in Y-axis values of the hand, elbow and shoulder for those frames. The steps are:

1) Takes the values y_{wrist}, y_{elbow} and y_{shoulder}.
2) Calculates $\text{Total\_difference} = |y_{\text{wrist}} - y_{\text{elbow}}| + |y_{\text{elbow}} - y_{\text{shoulder}}|$
3) Checks if Total_difference is below the threshold.
4) Calculates steps 1-3 for each hand in each frame and checks if the number of frames where the hand was parallel to the ground is above the threshold.

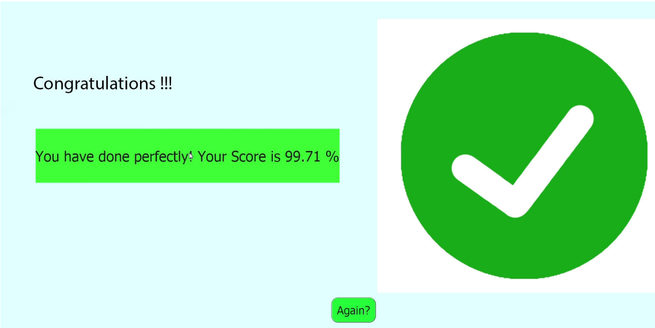
Similarly, for the "Arm lateral and medial rotation" exercise, the system also checks if the user's forearm (not the entire hand) is parallel to the ground. This is checked for the whole exercise instead of just the middle part like "Arm abduction and adduction" exercise. The steps are:

1) Takes the values y_{wrist} and y_{elbow}.
2) Calculates $|y_{\text{wrist}} - y_{\text{elbow}}|$ and checks if value is below the threshold.
3) Calculates steps 1 and 2 for each hand in each frame and checks if the number of frames where the hand was parallel to the ground is above the threshold.

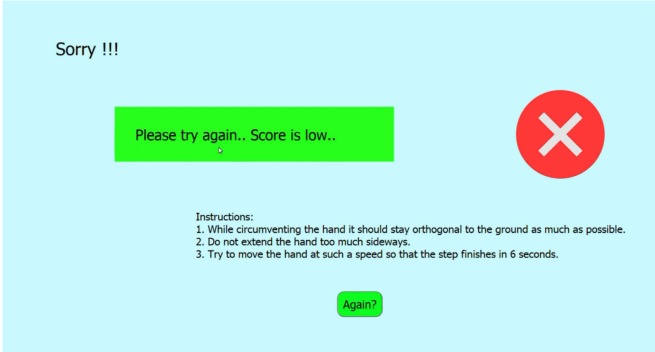
For the "Arm lateral and medial rotation" exercise, the system also checks if the user's elbow is close to the body (by checking difference between the midpoint of the elbow and wrist with the midpoint of the mid hip and neck) and orthogonal to the ground during the whole exercise. The steps are:

1) Takes the values x_{elbow}, x_{shoulder}, x_{neck} and $x_{\text{mid hip}}$.
2) Calculates $\text{Total\_difference} = |x_{\text{wrist}} - x_{\text{elbow}}| + |\frac{x_{\text{wrist}}+x_{\text{elbow}}}{2} - \frac{x_{\text{mid hip}}+x_{\text{neck}}}{2}|$
3) Checks if Total_difference is below the threshold.
4) Calculates steps 1-3 for each hand in each frame and checks if the number of frames where the forearm deviated is below the threshold.

The threshold values are set empirically. Using these metrics, the user is informed if his/her hand was in the proper position during the exercise.



(a) Correct exercise



(b) Incorrect exercise

FIGURE 28. **Results of a correct and incorrect exercise.**

IV. RESULTS

Videos of 109 subjects performing different exercises have been collected and processed to create Spatio-Temporal Polychromatic Trajectories (STPT). These were used to train and evaluate how accurately the classifier is able to differentiate between good and bad exercises. The results of exercise quality detection are discussed in detail below.

The STPT images generated from the keypoints extracted by Openpose from the videos have a resolution of 1280×720. They were initially resized to 224×224 and fed to the Convolution Neural Network (CNN) classifier. The architectures used for the base model in training are ResNet50_V2 [21], MobileNetV2 [22] and EfficientNetB0 [23]. The top layer was removed, and a global average pooling layer was added, followed by some dense layers, dropouts and batch normalization.

Table 5 shows the number of correct and incorrect clips obtained for each exercise. The STPT images were obtained from the original clips by running through Openpose and then they were augmented by randomly performing horizontal flip, rotation and zoom. For the validation set, 10 correct and 10 incorrect images were kept aside from the original dataset. They were augmented to increase the sizes by six times. For the test set, 20 correct and 20 incorrect images were set aside with no augmentation. The rest were used for training. The training set was augmented, and its size was increased by six or eight times depending on the exercise. The final number of STPT train images including augmentation is kept as such so that the quantity of correct and incorrect images is similar, ensuring a balanced dataset. The number of images distributed to train, validation and test is shown in Table 6.

Each of the three different architectures was taken, and the top layer was removed and modified according to Figure 29 for training on each exercise. A pooling layer and some dense layers were added after the base architecture. Except for the last one, the number and hidden units of dense layers were selected by trial and error until the best possible results were achieved. Dropouts were after some layers to reduce overfitting. Table 7 shows the training parameters for each exercise. Tensorflow 2.6 [24] was used to implement these, and the training device consists of AMD Ryzen 5900× CPU, 64 GB RAM and NVIDIA RTX 3080 GPU.

After the model was trained, the achieved training, and validation accuracy is shown in Table 8. It can be seen that the training accuracy achieved in all four exercises between the three base architectures is equal or above 90%, reaching up to 97%. For validation accuracy, the second, third and fourth exercises ranged from 90% to 93%, while in the first exercise, 94% to 97% was achieved. The best training accuracy was achieved by ResNet_V2 [21] in all the exercises except for "Arm abduction and adduction", where EfficientNetB0 [23] performed the best. The validation accuracy is close for each exercise between the three base architectures.

The models were evaluated with a test set of 20 correct and 20 incorrect STPT images for each exercise. The results are displayed in Table 9. The formula is provided for each of the metrics where TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative:

$$\begin{aligned}\text{Test Accuracy} &= \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}\\ \text{Precision} &= \frac{\text{TP}}{\text{TP}+\text{FP}}\\ \text{Recall} &= \frac{\text{TP}}{\text{TP}+\text{FN}}\\ \text{F1 Score} &= \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}\\ &= \frac{2\times\text{TP}}{2\times\text{TP}+\text{FP}+\text{FN}}\end{aligned}$$

In the first exercise and second exercises, the best F1 Score is achieved by EfficientNetB0 [23] and MobileNetV2 [22]. In the last two exercises, it is achieved by ResNet50_V2 [21]. While in the first, third and fourth exercise, at least one architecture achieved an F1 Score of over 90%, for the "Arm abduction and adduction" exercise the scores ranged only from 0.865 to 0.88.

Comparison With Other Works: The results are first compared with "Pose Trainer: Correcting Exercise Posture using Pose Estimation" [8] as they have similarly extracted the

TABLE 5. Number of video clips.

Exercise name	Correct clips	Incorrect clips	Total Clips
Arm flexion and extension	601	553	1154
Arm abduction and adduction	620	440	1060
Arm lateral and medial rotation	633	468	1101
Arm circumduction	299	506	805

TABLE 6. Number of train, validation and test images.

Exercise name	Train	Multiplier	Train after Augmentation	Validation	Multiplier	Validation after Augmentation	Test
Arm flexion and extension	1094	6	6564	20	6	120	40
Arm abduction and adduction	1000	6	6000	20	6	120	40
Arm lateral and medial rotation	1041	8	8328	20	6	120	40
Arm circumduction	745	8	5960	20	6	120	40

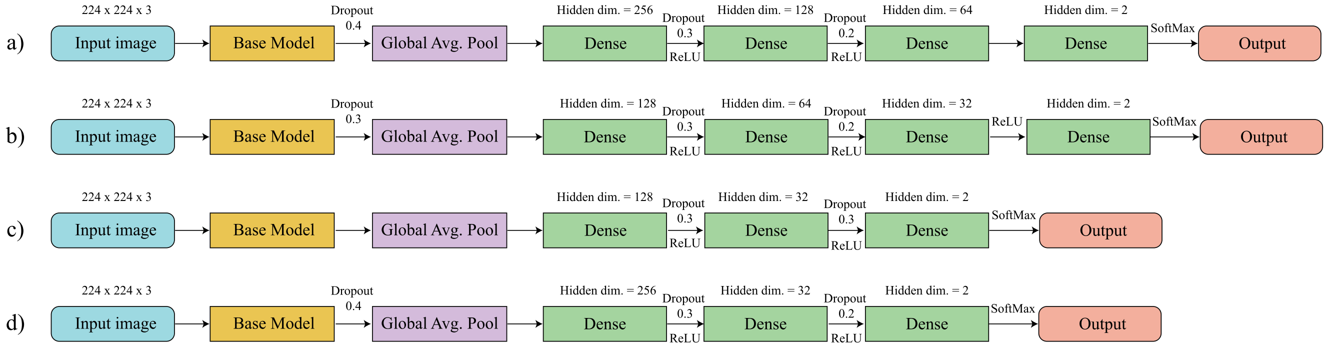


FIGURE 29. Network architectures for a) Arm flexion and extension, b) Arm abduction and adduction exercise, c) Arm lateral and medial rotation and d) Arm circumduction.

TABLE 7. Network parameters for training.

Exercise name	Train data	Validation data	Learning rate	Batch size	Epochs
Arm flexion and extension	6564	120	0.0005	64	45 to 150
Arm abduction and adduction	6000	120	0.00005	64	60 to 100
Arm lateral and medial rotation	8328	120	0.00005	64	40 to 80
Arm circumduction	5960	120	0.0005	64	40

TABLE 8. Training and validation accuracy.

Exercise Name	Base model	Train Acc.	Val. Acc.
Arm flexion and extension	ResNet50_V2 [21]	0.97	0.97
	MobileNetV2 [22]	0.95	0.95
	EfficientNetB0 [23]	0.95	0.94
Arm abduction and adduction	ResNet50_V2 [21]	0.92	0.90
	MobileNetV2 [22]	0.91	0.91
	EfficientNetB0 [23]	0.96	0.92
Arm lateral and medial rotation	ResNet50_V2 [21]	0.97	0.93
	MobileNetV2 [22]	0.94	0.92
	EfficientNetB0 [23]	0.90	0.90
Arm circumduction	ResNet50_V2 [21]	0.97	0.93
	MobileNetV2 [22]	0.95	0.93
	EfficientNetB0 [23]	0.94	0.90

keypoints using Openpose. Openpose only utilizes a simple camera and does not require any separate sensor to obtain 2D body keypoints.

In paper [8], they have detected the quality of four exercises: i) Bicep Curl, ii) Front Raise, iii) Shoulder Shrug, and iv) Shoulder Press. It should be noted that these exercises do

TABLE 9. Test results.

Exercise Name	Base model	Test Acc.	Precision	Recall	F1 Score
Arm flexion and extension	ResNet50_V2 [21]	0.90	0.90	0.90	0.90
	MobileNetV2 [22]	0.85	0.85	0.85	0.85
	EfficientNetB0 [23]	0.925	0.90	0.95	0.927
Arm abduction and adduction	ResNet50_V2 [21]	0.875	0.94	0.80	0.865
	MobileNetV2 [22]	0.875	0.826	0.95	0.88
	EfficientNetB0 [23]	0.875	0.94	0.80	0.865
Arm lateral and medial rotation	ResNet50_V2 [21]	0.95	0.95	0.95	0.95
	MobileNetV2 [22]	0.825	0.76	0.95	0.84
	EfficientNetB0 [23]	0.925	0.90	0.95	0.927
Arm circumduction	ResNet50_V2 [21]	0.975	1.0	0.95	0.974
	MobileNetV2 [22]	0.90	0.90	0.90	0.90
	EfficientNetB0 [23]	0.95	0.95	0.95	0.95

TABLE 10. Comparison of other's work vs proposed method.

Method	Exercise Name	Precision	Recall	F1 Score
S. Chen et al. [8]	Bicep Curl	0.89	0.86	0.85
	Front Raise	1.00	1.00	1.00
	Shoulder Shrug	0.89	0.85	0.85
	Shoulder Press	0.76	0.73	0.73
M. H. Lee et al. [9]	Bring a Cup to the Mouth	-	-	0.9442 ± 0.1
	Switch a Light On	-	-	0.8367 ± 0.2
	Move a Cane Forward	-	-	0.7854 ± 0.2
Y. X. Zhi et al. [10]	SVM healthy groups	0.73	0.78	0.74
	RNN healthy groups	0.74	0.76	0.74
	SVM patient group	0.86	0.88	0.87
	RNN patient groups	0.87	0.9	0.88
Proposed Method	Arm flexion and extension	0.90	0.95	0.927
	Arm abduction and adduction	0.826	0.95	0.88
	Arm lateral and medial rotation	0.95	0.95	0.95
	Arm circumduction	1.0	0.95	0.974

not directly match the ones in the proposed method as they are not suitable for an elderly person. In spite of that, it can be seen in Table 10 that the achieved results are comparable with them. They achieved the best Precision, Recall and F1 Score in the "Front Raise" exercise, in which all are equal to 1.0. While the worst score was from Shoulder Press, being 0.76, 0.73 and 0.73, respectively. The proposed method achieved the best test results from the "Arm circumduction" exercise, which are 1.0, 0.95 and 0.974, respectively. The Precision, Recall and F1 Scores of the next best exercise are 0.95, 0.95 and 0.95. It can be observed that the lowest result is from the "Arm abduction and adduction", which has scores of 0.826, 0.95 and 0.88, respectively. This is still better than their lowest and close to their "Shoulder Shrug" and "Bicep Curl" exercises. This shows that while their method might work well only for the "Front Raise" exercise, the proposed method generalizes well to all types of exercises.

The results are then compared with "Learning to Assess the Quality of Stroke Rehabilitation Exercises" [9]. They have used a separate Kinect sensor which can capture depth information. They also evaluated exercise quality similar to the proposed method. The F1 Score of the proposed method is comparable to their ROM (Range of Motion) as it calculates the amount of movement around a specific joint. For their exercise 1, 2 and 3, these scores are 0.9442 ± 0.1019, 0.8367 ± 0.2874 and 0.7854 ± 0.2405 respectively for their proposed approach as depicted in Table 10. Comparing these to the proposed method's highest F1 Scores for each exercise, it can be seen that the F1 Score for "Arm lateral and medial rotation" & "Arm circumduction" is close to their best result as they are 0.95 and 0.974, respectively. Moreover, the lowest best F1 Score is from "Arm abduction and adduction", which is 0.88, comparable to their second-highest score and much higher than their lowest one.

Furthermore, the proposed method can be compared with "Automatic Detection of Compensation During Robotic Stroke Rehabilitation Therapy" [10]. However, they have performed classification on compensation type, and in this work, exercise quality is classified. They have also used data from Kinect sensors similar to the compared previous work [9]. They classified using both SVM and RNN and got the following results for their three compensations LF (Lean forward), TR (Trunk Rotation), SE (Shoulder elevation) and NC (No compensation). For healthy groups, it can be observed in Table 10 that from SVM, they got a weighted average of Precision, Recall, F1 Score of 0.73, 0.78 and 0.74, and from RNN, they similarly got 0.74, 0.76 and 0.74. For the patient groups, they got better results. From SVM they got 0.86, 0.88 and 0.8, and from RNN, they got 0.87, 0.9 and 0.88. If these are compared with the proposed method's results, in the first, third and fourth exercises, the proposed method got better results than their best result (RNN Patient Group).

The Precision, Recall, F1 Scores are 0.90, 0.95 & 0.927; 0.95, 0.95 & 0.95 and 1.0, 0.95 & 0.974, respectively for these three exercises. The "Arm abduction and adduction" exercise has a similar score to their best score with 0.826, 0.95 & 0.88, respectively.

The proposed method has performed on par or better than other mentioned methods in most exercises ("Arm flexion and extension", "Arm lateral and medial rotation" & "Arm circumduction"). One of the reasons the proposed method performs better is that it took on a sequential problem which is mainly approached using RNNs, and converted it into a computer vision task. The generated Spatio-Temporal Polychromatic Trajectories (STPT) give the machine a clear idea of the motion and can be analyzed easily using CNNs. The data can also be easily augmented, which is more challenging for sequential datasets. Also, the length of the video, in this case, becomes less relevant as the whole video is projected onto an STPT image. The proposed method is observed to perform better if the dataset is expanded by augmentation.

V. CONCLUSION

This paper demonstrates an automated exercise feedback system for elders by taking the keypoints from the body in each frame, generating a newly proposed Spatio-Temporal Polychromatic Trajectory (STPT) image using the data and inferencing it through a CNN model [7]. The aim is to plot the temporal data onto a 2D image to make it easier to classify the quality of the exercise in the video. We designed the STPT image to incorporate both the motion trajectory and the direction, giving a clear overview of the exercise. Some incorrect exercise metrics were also implemented to provide essential feedback to the elderly person, and a graphical user interface was designed to increase the system's usability. The goal is to use this system to perform exercises and get feedback without the help of a doctor or physician, especially for elderly people. The interface is easy to understand, so they can learn to use it intuitively. Though four general-purpose shoulder exercises were demonstrated in this work, the proposed method will incorporate more exercises in the future. Especially exercises that are prescribed for specific diseases. There is scope for improvement by investigating methods to increase the accuracy of exercise routines like "Arm abduction and adduction" in the future. The system can be deployed commercially on a large scale for nursing homes or hospitals.

ACKNOWLEDGMENT

The authors thank the Physical Therapy and Rehabilitation Centre (PTRC), Uttara, Dhaka, and Dhaka City Physiotherapy Hospital, Jigatola, Dhaka, for providing them with guidelines about suitable exercise routines for elderly people. They also thank Johayra Prithula for her illustrations of subjects performing various exercises. They gratefully thank Information and Communication Technology Division for research grant and University of Dhaka for publication grant. *(Riad Ahmed, Raiyaan Abdullah, and Lafifa Jamal contributed equally to this work.)*

REFERENCES

[1] *Population Ages 65 and Above (% of Total Population)*. Accessed: Apr. 20, 2022. [Online]. Available: https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS
[2] *Ageing and Health*. Accessed: Apr. 20, 2022. [Online]. Available: https://ageingasia.org/ageing-population-bangladesh
[3] R. Abdullah, R. Ahmed, and L. Jamal, "A novel IoT-based medicine consumption system for elders," *Social Netw. Comput. Sci.*, vol. 3, no. 6, p. 471, Sep. 2022, doi: 10.1007/s42979-022-01367-8.
[4] K. Ashton, "That 'Internet of Things' thing," *RFID J.*, vol. 22, no. 7, pp. 97–114, Jun. 2009.
[5] P. H. Winston, *Artificial Intelligence*. Reading, MA, USA: Addison-Wesley, 1984.
[6] G. Stockman and L. G. Shapiro, *Computer Vision*. Upper Saddle River, NJ, USA: Prentice-Hall, 2001.
[7] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *Proc. Int. Conf. Eng. Technol. (ICET)*, Aug. 2017, pp. 1–6.
[8] X. Z. Ying, M. Lukasik, M. H. Li, E. Dolatabadi, R. H. Wang, and B. Taati, "Automatic detection of compensation during robotic stroke rehabilitation therapy," *IEEE J. Transl. Eng. Health Med.*, vol. 6, pp. 1–7, 2018.
[9] M. H. Lee, D. P. Siewiorek, A. Smailagic, A. Bernardino, and S. B. I. Badia, "Learning to assess the quality of stroke rehabilitation exercises," in *Proc. 24th Int. Conf. Intell. User Interfaces*. New York, NY, USA: Association for Computing Machinery, Mar. 2019, pp. 218–228, doi: 10.1145/3301275.3302273.
[10] S. Chen and R. Yang, "Pose trainer: Correcting exercise posture using pose estimation," Mar. 2018, *arXiv:2006.11718*.
[11] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172–186, Jan. 2021.
[12] M. Müller, *Information Retrieval for Music and Motion*, vol. 2. Berlin, Germany: Springer, 2007.
[13] M. Reyes, G. Domínguez, and S. Escalera, "Featureweighting in dynamic timewarping for gesture recognition in depth data," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 1182–1188.
[14] S. Sempena, N. Ulfa Maulidevi, and P. Ruswono Aryan, "Human action recognition using dynamic time warping," in *Proc. Int. Conf. Electr. Eng. Informat.*, Jul. 2011, pp. 1–5.
[15] C. H. Pham, Q. K. Le, and T. H. Le, "Human action recognition using dynamic time warping and voting algorithm," *VNU J. Sci., Comput. Sci. Commun. Eng.*, vol. 30, no. 3, pp. 22–30, 2014.
[16] H. Mohammadzade, S. Hosseini, M. R. Rezaei-Dastjerdehei, and M. Tabejamaat, "Dynamic time warping-based features with class-specific joint importance maps for action recognition using Kinect depth sensor," *IEEE Sensors J.*, vol. 21, no. 7, pp. 9300–9313, Apr. 2021.
[17] S. Seto, W. Zhang, and Y. Zhou, "Multivariate time series classification using dynamic time warping template selection for human activity recognition," in *Proc. IEEE Symp. Ser. Comput. Intell.*, Dec. 2015, pp. 1399–1406.
[18] J. Wang and H. Zheng, "View-robust action recognition based on temporal self-similarities and dynamic time warping," in *Proc. IEEE Int. Conf. Comput. Sci. Automat. Eng. (CSAE)*, vol. 2, May 2012, pp. 498–502.
[19] Z. Ramezanpanah, M. Mallem, and F. Davesne, "Human action recognition using Laban movement analysis and dynamic time warping," *Proc. Comput. Sci.*, vol. 176, pp. 390–399, Jan. 2020.
[20] C. Arrowsmith, D. Burns, T. Mak, M. Hardisty, and C. Whyne, "Physiotherapy exercise classification with single-camera pose detection and machine learning," *Sensors*, vol. 23, no. 1, p. 363, Dec. 2022. [Online]. Available: https://www.mdpi.com/1424-8220/23/1/363
[21] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," *CoRR*, vol. abs/1603.05027, pp. 1–15, Mar. 2016.
[22] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
[23] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn.*, vol. 97, K. Chaudhuri and R. Salakhutdinov, Eds., Jun. 2019, pp. 6105–6114. [Online]. Available: https://proceedings.mlr.press/v97/tan19a.html
[24] *TensorFlow*. Accessed: Jun. 21, 2021. [Online]. Available: https://www.tensorflow.org/

RIAD AHMED received the B.Sc. degree from the Department of Robotics and Mechatronics Engineering, University of Dhaka, Bangladesh. He was a Machine Learning Engineer with Headless Technologies Ltd., and a Computer Vision Engineer with Pranon Global Ltd. He was a Research Assistant with the University of Dhaka. His research interests include computer vision, machine learning, the Internet of Things, industrial automation, activity recognition, video analytics, and artificial intelligence in healthcare.

RAIYAAN ABDULLAH received the B.Sc. degree from the Department of Robotics and Mechatronics Engineering, University of Dhaka, Bangladesh, in 2022. He has worked on various projects in the field of machine learning. He is currently a Research Assistant with the University of Dhaka. He is also the Co-Founder and the Development Project Manager of BiTechX, a web development and design agency working in the North American market. His research interests include artificial intelligence, machine learning, computer vision, the Internet of Things, and their applications in various domains, such as healthcare, action recognition, autonomous vehicles, and medical imaging.

LAFIFA JAMAL (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from the Department of Computer Science, University of Dhaka, Bangladesh, and the Ph.D. degree in reversible logic synthesis from the University of Dhaka. She is currently a Professor with the Department of Robotics and Mechatronics Engineering, University of Dhaka. She has published a good number of research papers in international conferences and journals. Her research interests include robotics and intelligent systems, the Internet of Things, reversible logic synthesis, quantum computing, gender and diversity. She is a member of the IEEE Robotics and Automation Society, the IEEE Computer Society, the IEEE Women in Engineering, the American Association for the Advancement of Science (AAAS), the Robotics Society India (RSI), the Internet Society (ISOC), Bangladesh Computer Society (BCS), and the Bangladesh Women in Technology (BWIT).