

# Automatic Feedback For Physiotherapy Exercises Based On PoseNet

Hussein Ayman Hassan

*Computers and Artificial intelligence*  
(Computer Science Department)  
Helwan University, Cairo, Egypt  
husseinayman2@fci.helwan.edu.eg

Basant Hussein Abdallah

*Computers and Artificial intelligence*  
(Computer Science Department)  
Helwan University, Cairo, Egypt  
basant 20160107@fci.helwan.edu.eg

Asmaa Ahmed Abdallah

*Computers and Artificial intelligence*  
(Computer Science Department)  
Helwan University, Cairo, Egypt  
Asmaa 20160069@fci.helwan.edu.eg

Rana Osama Abdel-Aal

*Computers and Artificial intelligence*  
(Computer Science Department)  
Helwan University, Cairo, Egypt  
ranaosama263@fci.helwan.edu.eg

Rewas Reda Numan

*Computers and Artificial intelligence*  
(Computer Science Department)  
Helwan University, Cairo, Egypt  
rewasreda2015@gmail.com

Aya Khaled Darwish

*Computers and Artificial intelligence*  
(Computer Science Department)  
Helwan University, Cairo, Egypt  
Aya 20160099@fci.helwan.edu.eg

Wessam H. El-Behaidy\*

*Computers and Artificial intelligence*  
(Computer Science Department)  
Helwan University, Cairo, Egypt  
W\_behaidy@fci.helwan.edu.eg

**Abstract**—This project helps patients who undergo physiotherapy exercises. It can help patients to do right poses and prevent them getting injured during the exercises. The home-based physiotherapy exercise (HPTE) dataset is used. The dataset consists of eight exercises which focus on shoulder, and knee. It is performed by five different actors with different shapes and sex. Our system uses seven exercises with two sides for more variety and better user experience. The main idea of the system is based on detecting the body key points using one of pose estimation techniques. The key points are extracted from different exercises and build our pickle file. Then, those pickle file's key points are compared to the key points of the patient using dynamic time warping. The system is implemented as a desktop application with a simple user interface and a 3D avatar is used to guide the patients to perform the exercises correctly. The system gives them a score on the exercise and a feedback on which parts of their body have incorrectly posture. Our system offers high accuracy pose estimation and pose comparison with a minimum accuracy of 80%.

**Index Terms**—pose estimation; physiotherapy exercises; key points extraction; automatic feedback

## I. INTRODUCTION

Physical therapy (or physiotherapy) is a medical science that deals with the diagnosis and treatment of patients who have problem in performing functional activities due to injuries or other problems [1]. Physiotherapy exercises are usually the

solution for these patients. These exercises should be regularly performed in a controlled manner to adjust therapy parameters. Before or/and during the exercise sessions, the therapists give some oral instructions, and physical guidance [2]. So, patients can enhance their ability to detect and correct errors by repeating the required exercise. Physiotherapy treatments are usually long running as it takes time to restore a person's movement. One must perform the same set of exercises for a few months every day with the correct posture to regain the movement [3]. The problem here when the patient must go to the hospital or may be travelled to receiving in-clinic treatment. Therapy availability and accessibility are the major problems that restrict the sessions between the patients and their therapists. In the other hand, hiring a private therapist is not a good idea because the cost is very expensive and not everyone can afford that. For that, some patients will have no choice but doing it by them-self at their home. But this is so dangerous and not reliable. Machine learning based motion recognition research has recently performed major advances in building low-price, accurate, and stable action recognition systems from video data for controlled environments. In this paper, a proposed system is built to guide, provide instant feedback and act as a personal virtual trainer that would help people do exercises by themselves.

Section II reviews some of the recent related research that deliver similar systems. The implemented system and dataset

\* Corresponding author

are introduced in Section III. Section IV clarifies the user interface and results obtained. Finally, the conclusion is stated in last section.

## II. RELATED WORK

This section covers different work that deliver similar systems.

Chen et al. [4] made a Pose Trainer application. The system receives the recording video of the user for a specific exercise (Bicep Curl, Front Raise, Shoulder Shrug and Shoulder Press), and ends with providing personalized feedback on specific exercise improvements. They use deep convolution neural networks (CNNs) to label RGB images for pose estimation. This model is composed of a multi-stage CNN with two branches, one to learn the confidence of a KeyPoint on an image, and the other to learn the part affinity fields. They choose to use OpenPose pre-trained model for pose detection. After that they are approaching the evaluation part using dynamic time warping (DTW) with a nearest neighbor classifier.

Mayorquin et al. [5] implement a pose recognition assistant YogAI which gives a corrective advice on users' yoga posture. YogAI guides the user through the workout, checks the current pose of the user and give some advice for the user if his pose is not correct. Firstly, they built up a corpus labeled "Yoga poses" that collects images of common yoga positions. For each posture, the system localizes body key points using OpenPose. The resulting key points are then passed as features to XGBoost model to classify the user pose, if it is correct or not. If it is not, the system guides the user through a conversational interface until the pose is correctly performed. Finally, it guides the user to the next pose until the yoga session is completed.

Jain [1] creates a system that can guide, provide instant real-time feedback and act as a personal virtual trainer that would help people do physiotherapy exercises. This system takes a live video for the user while doing the exercise. During the exercise, the key points appear on top of user's skeleton. The user starts to perform a movement and hold on for seconds. Between these movements, the system gives an instant written feedback as instructions to the user. For pose estimation, he used open-pifpaf and tf-pose-estimation to have the highest FPS (frames per second) rates. But open-pifpaf gave more accurate results in cases when the full body wasn't visible, or the user was lying sideways. Then, he used cosine similarity to provide a score on how close the motion was to the perfect motion reference. Also, the system considers the variance in the duration of videos as it is not guaranteed that both videos would be of the same length and movements with same speed.

## III. METHODS

This section clarifies the dataset used, and front-end and back-end of implemented system.

Our system extends the work of Raj et al. [6] to be a desktop application instead of command line and to deal with more

than one video for the same exercise. Our contribution goes further to integrate a 3D avatar to guide the user through the exercise. In addition, it calculates the range that each key point in each exercise to provide an appropriate feedback to the user about the different parts of his body.

### A. Dataset

The dataset used is the home-based physiotherapy exercise (HPTE) dataset [1]. This dataset focused on shoulder and knee using eight exercises: arm lifting, shoulder abduction, single leg extension, static triceps extension, swing arm, circle arm, seated leg and seated hamstring. It has been performed by five different actors with different shapes and sex. Each actor repeats each exercise six times with different positions and variation in time. The videos are taken by Microsoft Kinect sensor, which contains a depth and RGB camera. Both RGB and gray-level of these videos are stored as well as their depth, with total of 480 videos.

In our system, only seven exercises are used. The single leg extension exercise is excluded because its key points were not clear to work on them, which affect the accuracy. Only 210 videos are used from the dataset, after excluding the exercise and gray-level videos, with 30 videos for each exercise.

### B. Front End System

The front end of our application is simple interface that the user can easily interact with it. When the user starts the application, he firstly logins if he has an account or registers to perform an account. Then, the user chooses an exercise to perform and the side of body (i.e, left side or right side) if exist in the exercise. At this moment, the camera will open for 30 seconds and take the user's live exercise video. During the exercise, an avatar starts to do the exercise in front of the user to guide him. After the exercise is finished, the feedback and user's score appear on screen.

### C. Back End System

During the exercise, many processes are performed in the back end of our system as shown in Figure 1. The user's live exercise video is passed as input to PoseNet model. PoseNet extracts the key points and saves them in an array. This array is normalized and compared to the dataset pickle file using the dynamic time warping algorithm. Finally, a feedback is constructed based on the comparison scores. In this section, these processes are clarified.

1) *PoseNet*: PoseNet [7] is a pre-trained model used to pose estimation. PoseNet is chosen for its ease of installation and use for end-users and its lightweight. There are two architectures in PoseNet: Resnet architecture and MobileNet architecture. The implemented architecture in our system is the MobileNet architecture. One of the important factors in choosing PoseNet with MobileNet architecture is its high performance. The output from PoseNet is heatmaps with its confidence score, which is the probability that a part of that key points type exists in that position. The second output is the offset vectors which is the location of the heatmaps. Then, the

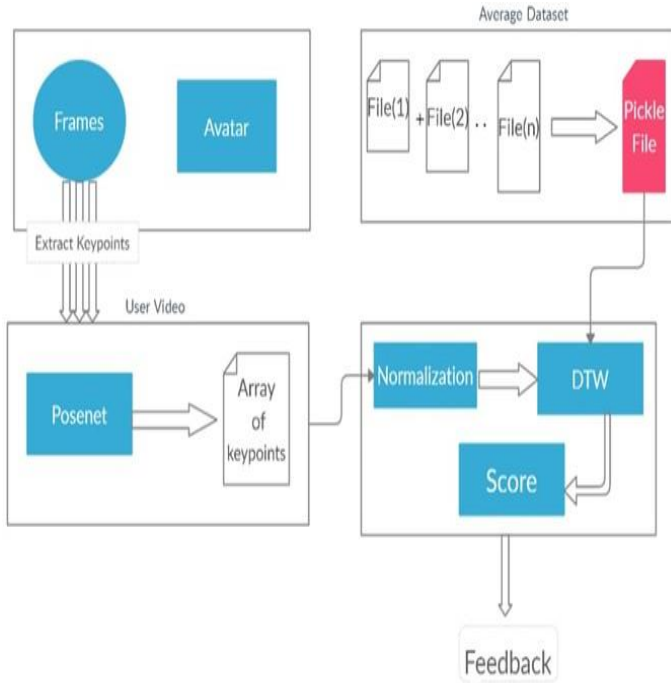


Fig. 1. A detailed diagram of Back end system

model apply  $\text{argmax}()$  function on the confidence score to take the highest score to be the exact key-point, and the exact offset vector is the corresponding location of this highest confidence score. PoseNet can detects 17 key points for the different parts of body.

2) *Key points Normalization*: The users who use our system may differ in shape, weight, and tall, as well as some of them may be close to the camera and others may be far away. All these pervious factors have a great effect on the key points scores and accuracy [8]. For this reason, L2-Normalization type is applied on key points got from PoseNet, to make the sum of squares of the key points equals one.

3) *Comparison*: To give a feedback to the user, a comparison is needed between the user's key points and dataset's key points. Here, the key points of the dataset are stored in a pickle file and the dynamic time warping (DTW) is used for comparison.

a) *Pickle file*: Using PoseNet, the key points got from the 30 videos of each exercise are extracted and saved into pickle files. In this work, two modes are used. The first mode (Mode 1), each exercise has 30 pickle files will be used in comparison. The second mode (Mode 2), the key points got from the 30 videos of each exercise are averaged. The same process is repeated for the seven exercises. The average pickle file is generated by putting these average key points in a single pickle file format for all these exercises.

b) *Dynamic Time Warping (DTW)*: DTW [9] is a metric used to measure the nonlinear similarity between two time series. It can overcome the problem that the two similar sequences are phase-shifted (i.e. shifted in the time dimension). It develops a one-to-many match so that the troughs and peaks

with the same pattern are perfectly matched, and there is no left out for both curves. DTW used in many fields like computer vision, sound Recognition, and stock market. Here, DTW is used to compare the user exercise keypoints to the pickle file of same exercise.

4) *Feedback*: After the user has finished the exercise, the model provides the user with a score and a feedback on which part of the body is wrongly done. The given score is the average percentage of comparing the key points of user's body with the generated pickle file. To provide this feedback, many comparisons are made on the videos of the dataset with different duration, actors, and position of the actors. These comparisons are used to specify the minimum percentage of each key point. Each key point of user's body should be within the specified minimum for this key point and 100% to consider the posture of this key point right. If, it is not in the specified range, the feedback is an advice for the user to correctly posture this part of his body.

#### IV. RESULTS

The front-end of our system is a user-friendly interface which makes the user feels comfortable when he uses it. Firstly, the user registers or login to do the required exercise. After login, the user chooses which exercise he wants to perform from the **seven available exercises**, as Figure 2. As well as, he chooses which side (**Left Side or Right Side**) to train, if it is available in the exercise.

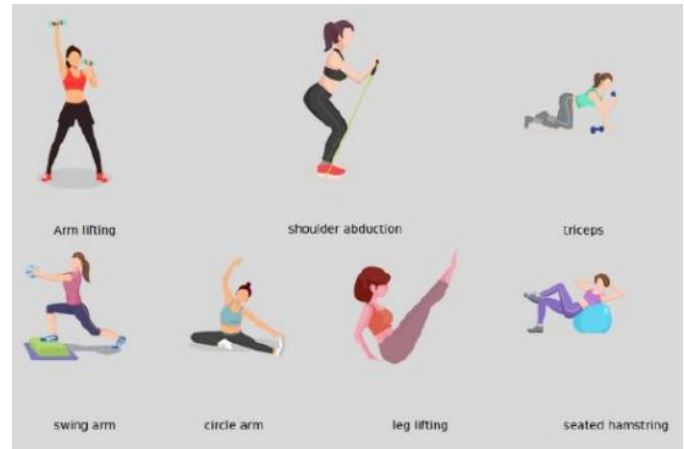


Fig. 2. The exercises Menu.

At that moment, the camera is opened for 30 seconds for user to perform the exercise. At the same time, a 3D avatar will appear to guide the user how to perform the exercise. While the user is training, the frames of user's live video is passed to the back-end system. The PoseNet detects the 17 key points in each frame and returns an array of key points and their offsets. Key points are indexed by ids from 0 to 16, as Table I.

These extracted key points of user's body will be normalized, as in sec. III-B, then compared to the pickle file of the dataset using DTW as in sec. III-C.

TABLE I  
POSENET KEYPOINTS ID

ID	Part	ID	Part
0	Nose	9	Left Wrist
1	Left Eye	10	Right Wrist
2	Right Eye	11	Left Hip
3	Left Ear	12	Right Hip
4	Right Ear	13	Left Knee
5	Left Shoulder	14	Right Knee
6	Right Shoulder	15	Left Ankle
7	Left Elbow	16	Right Ankle
8	Right Elbow		

To provide an appropriate feedback to the user, many comparisons are made on the videos of the dataset with different duration, actors, and position of the actors. These comparisons are used to specify the minimum percentage of each key point for each exercise. Table II shows 11 comparisons made for Circle Arm Left exercise. This exercise is based on only 7 interested key points. To calculate the minimum percentage, the values for each key points are ordered ascendingly. The minimum percentage is specified as the minimum mode (i.e. the smallest of most repeated list) within these values. If the minimum value was not within the 5 smallest values, the median of the smallest five is taken. The final minimum percentage calculated for all key points for each exercise are shown in Table III and IV.

TABLE II  
SCORE LIST OF COMPARISONS MADE ON CIRCLE ARM LEFT EXERCISE TO SPECIFY THE MINIMUM PERCENTAGE FOR EACH KEY POINT IN THE EXERCISE. NOTE: L. FOR LEFT AND R. FOR RIGHT

Part of Body	Comparisons											Min.
	1	2	3	4	5	6	7	8	9	10	11	
Nose	65	65	69	79	63	70	70	66	65	70	58	65
L. Shoulder	63	66	66	76	66	78	75	70	73	75	69	66
R. Shoulder	63	64	61	72	66	75	72	73	70	74	72	69
L. Elbow	78	78	79	85	77	86	84	83	83	85	73	78
L. Wrist	63	67	66	69	74	75	68	71	71	69	75	67
R. Wrist	59	67	59	63	73	57	64	71	71	69	73	59

After comparison, the user gets a feedback after 3 seconds from finishing his exercise. If all his key points are within the specified ranges of Table 3 and 4 upon the exercise, the user got a feedback message as Figure 3. Whereas, if any of his key points is out of specified range, the system will specify the part that was out of right posture. Figure 4 shows a feedback of later case, both the left elbow and right wrist were out of right posture.

To test and evaluate our system, a user performs arm lifting exercise using our application. Using mode 1, the comparison between user's extracted key points with the 30 pickle files of the exercise takes from 2 to 3 minutes to provide the feedback to the user. Whereas using mode 2 for the same exercise, the comparison between user's extracted key points with the average pickle file only takes from 30 to 40 seconds to have feedback. But mode 1 has higher accuracy than mode 2 at first. Mode 1 gives 70.1% accuracy, whereas mode 2 gives 61% as

TABLE III  
THE FINAL MINIMUM PERCENTAGE CALCULATED FOR ALL KEY POINTS FOR THE THREE EXERCISES SHOULDER ABDUCTION, SEATED HAMSTRING AND STATIC TRICEPS EXTENSION FOR BOTH SIDES.

Key points	Shoulder Abduction Left	Shoulder Abduction Right	Seated Hamstring Left	Seated Hamstring Right	Static Extension Left	Static Extension Right
Nose	63	57	-	-	-	-
Left Eye	-	46	-	-	-	-
Right Eye	-	-	61	-	-	-
Left Shoulder	88	77	80	-	72	76
Right Shoulder	66	72	-	68	53	61
Left Elbow	80	83	71	-	68	50
Right Elbow	64	71	-	55	50	68
Left Wrist	69	72	65	-	62	57
Right Wrist	64	64	-	69	50	70
Left Hip	-	75	79	-	-	-
Right Hip	69	-	-	60	-	-
Left Knee	-	-	63	-	-	-
Right Knee	-	-	-	73	-	-
Left Ankle	-	-	64	-	-	-
Right Ankle	-	-	-	62	-	-

TABLE IV  
THE FINAL MINIMUM PERCENTAGE CALCULATED FOR ALL KEY POINTS FOR THE TWO EXERCISES CIRCLE ARM, SWING ARM AND LEG LIFTING EXTENSION FOR BOTH SIDES PLUS ARM LIFTING.

Key points	Circle Arm Left	Circle Arm Right	Swing Arm Left	Swing Arm Right	Arm Lifting	Leg Lifting Ext. Right	Leg Lifting Ext. left
Nose	74	65	-	-	-	65	55
Left Ear	-	-	-	64	-	-	-
Right Ear	-	-	61	-	-	-	-
Left Shoulder	70	66	71	63	75	74	72
Right Shoulder	67	64	63	53	61	51	56
Left Elbow	78	78	66	72	77	70	66
Right Elbow	76	71	60	50	60	49	50
Left Wrist	75	67	50	70	58	72	66
Right Wrist	73	59	60	50	41	50	52
Left Hip	-	-	-	-	71	70	66
Right Hip	-	-	-	-	70	55	56
Left Knee	-	-	-	-	72	64	62
Right Knee	-	-	-	-	70	55	50
Left Ankle	-	-	-	-	67	-	-
Right Ankle	-	-	-	-	66	-	-

shown in Table V. The degradation of accuracy of mode 2 was a cause of taking the average while there is a variation in lengths of video frames. After solving this problem, mode 2 gives for the same exercise 87.2% accuracy which is higher than mode 1. For that, mode 2 is the recommended mode. Our system is tested using another 2 exercises performed compared using mode 2 (i.e. the fastest and high-performance mode), as shown in Table VI. The final accuracy of Shoulder abduction and Swing arm exercises is 81.7% and 82.8%, respectively, which are pretty good. By taking the average accuracy of Arm lifting, Shoulder abduction and Swing arm exercises, their average is 83.9%. For that, we can say that the average accuracy of our system is around 80%.



Fig. 3. A feedback message, if all the user’s key points are within the specified ranges



Fig. 4. A feedback message, if some of user’s key points are not within the specified ranges

V. CONCLUSION

In this paper, a desktop system is implemented to provide a feedback on physiotherapy exercises. The user can choose the exercise type and body side (i.e., left or right), then the avatar is guiding the user to perform the exercise correctly. While the user is performing the exercise for 30 seconds, the back-

TABLE V

RESULT OF TWO MODES FOR ARM LIFTING EXERCISE. THE RESULTS OF MODE 2 ARE SHOWN BEFORE AND AFTER SOLVING THE PROBLEM OF VARIATIONS IN LENGTHS OF VIDEO FRAMES.

Interested key-points	Mode 1 Before	Mode 2 Before	Mode 2 After
Left shoulder	79	62	84
Right shoulder	63	67	80
Left elbow	81	70	91
Right elbow	65	58	92
Left wrist	57	59	75
Right wrist	45	50	95
Total Accuracy	70.1%	61%	87.2%

TABLE VI

RESULT OF MODE 2 AFTER SOLVING THE PROBLEM OF IN LENGTHS OF VIDEO FRAMES FOR SHOULDER ABDUCTION AND SWING ARM EXERCISES

Key Points	Shoulder Abduction	Swing Arm
Nose	68	90
Left Eye	72	88
Right Eye	62	90
Left Ear	80	89
Right Ear	61	90
Left Shoulder	90	60
Right Shoulder	84	84
Left Elbow	81	75
Right Elbow	86	84
Left Wrist	74	74
Right Wrist	73	81
Left Hip	94	82
Right Hip	95	73
Left Knee	93	93
Right Knee	93	78
Left Ankle	91	94
Right Ankle	92	83
Total Accuracy	81.7%	82.8%

end system detects the key-points of user for each frame and compares them with average pickle file using DTW approach. In less than three seconds after finishing the exercise, our system gives the user a feedback with a final score and some advises. Our system got an overall 80% for giving right feedback.

REFERENCES

- [1] I. ar and Y. Akgul, *A Monitoring System for Home-Based Physiotherapy Exercises*, 10 2013, pp. 487–494.
- [2] D. P. M. van Vliet Research Fellow and G. Wulf, “Extrinsic feedback for motor learning after stroke: What is the evidence?” *Disability and Rehabilitation*, vol. 28, no. 13-14, pp. 831–840, 2006, pMID: 16777770. [Online]. Available: <https://doi.org/10.1080/09638280500534937>
- [3] Samkit, Mar 2020. [Online]. Available: [https://medium.com/@\\_samkitjain/physio-pose-a-virtual-physiotherapy-assistant-7d1c17db3159](https://medium.com/@_samkitjain/physio-pose-a-virtual-physiotherapy-assistant-7d1c17db3159)
- [4] S. Chen and R. R. Yang, “Pose trainer: Correcting exercise posture using pose estimation,” 2020.
- [5] Salma and Terry, “Yogai: Smart personal trainer,” Feb 2019. [Online]. Available: <https://www.hackster.io/yogai/yogai-smart-personal-trainer-f53744>
- [6] K. R. R, “Human pose comparison and action scoring using deep learning,opencv python,” Medium, 04 2020. [Online]. Available: <https://medium.com/analytics-vidhya/human-pose-comparison-and-action-scoring-using-deep-learning-opencv-python-c2bdf0ddecba>
- [7] Dan, “Real-time human pose estimation in the browser with tensorflow.js,” May 2018. [Online]. Available: <https://blog.tensorflow.org/2018/05/real-time-human-pose-estimation-in.html>
- [8] Jason, Feb 2018. [Online]. Available: <https://machinelearningmastery.com/gentle-introduction-vectors-machine-learning/>
- [9] H. Li, “On-line and dynamic time warping for time series data mining,” *International Journal of Machine Learning and Cybernetics*, vol. 6, 02 2014.