Journal of Engineering Science and Technology
Special Issue on SIET2022, May (2022) 157 - 171

EXERCISE QUALITY ANALYSIS USING AI MODEL AND COMPUTER VISION

CHEW ZHEN YUE, LAU CHEE YONG*, LAI NAI SHYAN

School of Engineering, Asia Pacific University of
Technology and Innovation, Kuala Lumpur 57000, Malaysia
*Corresponding Author: laucheeyong@apu.edu.my

Abstract

In 2020 where the Covid-19 virus starts spreading around the world, many countries suffered and are forced to go through lockdowns. This creates a new norm where people are encouraged to stay at home and only go out for the purpose of getting necessities. However, in this condition, people are still urged to keep an active lifestyle by doing exercises indoors. This results in the increase of people working out at home with reference to trainer videos found online. Having said so, these trainers cannot advice or look after the people 'mimicking' them through their screens. Therefore, an idea was sparked to create an exercise assistant program to compare the posture of the user against the video referenced. With this program, it can evaluate the performance and accuracy of exercise done by the user and provide feedback to the user to improve in their future attempt on the same exercise. With the help of the MediaPipe pose estimation model, the Joint Coordinates of the users can be obtained. These data will then be converted into 12 of the main human body Joint Angles before going through a dynamic comparison method called the DTW (Dynamic Time Warping). Thus, the similarity of each joint of the user and the referenced video can be successfully acquired. Other than that, both user's and trainer's Joint Angles are also plotted in graphs at the same time to provide a visual representation of the differences found. This system is believed to have the potential to be integrated into the educational and healthcare world in the future as a method to constantly monitor and notice what is necessary to improve.

Keywords: Dynamic time warping, Exercise assistant, MediaPipe, Motion recognition, Pose estimation.

1. Introduction

As the Covid-19 virus continues to spread around the world, many countries suffered and are forced to go through lockdowns. In Malaysia, the form of lockdown is called Movement Control Order (MCO) which was announced on 16 March 2020 by the Prime Minister to be implemented on 18 March 2020 due to a sudden surge of the Covid-19 active cases. This condition restricts people from leaving their homes to prevent the virus from spreading. The restriction measure was then extended to different versions including the Conditional Movement Control Order (CMCO), Enhance Movement Control Order (EMCO) and Recovery Movement Control Order (RMCO). However, in all versions of the restriction measures, the sports and recreational sector was always limited in reopening due to the direct and indirect contacts including through physical contacts, sweat or even respiration gases that occur regularly with them [1]. With these said, many platforms, including the World Health Organization (WHO) have released articles to urge people to stay physically active while staying at home. It is said so as individuals' health, well-being, and quality of life can all be harmed by sedentary and insufficient levels of physical activity. In said article, WHO advised 150 minutes of moderate-intensity exercise or 75 minutes of vigorous-intensity exercise per week, or a combination of the two [2]. Besides, it also improves the overall mental health of the community with the ability of exercising to not only reduce negative mood and overcome common mental health issues like depression and anxiety but also improves one's self esteem and cognitive function [3]. Even without actual gym equipment or a large amount of room, these recommendations can be carried out at home. Besides, many people posting their workout videos or photos on social media also helped push this agenda. This peer pressure will eventually encourage many people to start doing physical exercises at home with reference to the thousands of workout videos available online.

This situation creates a problem where many people without much experience will try to replicate what is being shown in their screens themselves. In fact, even people who regularly hit the gym may not be able to perform all the exercises accurately. These errors or incorrect techniques due to the lack of professional guidance will not only have an impact on the effectiveness of the activities but if worse will also lead to long-term injuries [4]. Therefore, this project aims to design and develop a low-cost exercise assistant program that can guide and correct its user to perform accurate exercise posture and present analysis done through a dashboard.

The aim of this project is to design and develop a low-cost exercise assistant program that can guide and correct its user to perform accurate exercise posture and present analysis done through a dashboard. The detail objectives are as follow:

- To develop an Exercise assistant program using Machine Vision with the help of the AI or machine vision algorithm.
- To design a Post Exercise analysis dashboard as a feedback system for the user.
- To evaluate the performance and accuracy of exercise done by the user.

2. Methodology

Figure 1 presents the block diagram for Exercise Pose Correction System. As shown, the system requires its user to provide the input either by uploading or downloading the interested video file into the device. Then in the system, user is required to declare

the identity of the video file for the system to know which function to execute. By uploading a trainer and a user's video file, the system can go through a series of instructions to finally come up with a scoring system to grade the similarity of the user's movement in the video against the trainer's one, hence fulfilling the first and third objective proposed. Finally, 12 graphs of each user's joint angles will be presented in a dashboard to provide a clearer view on the cause of low scores in certain joints. With this, the second objective can be achieved.

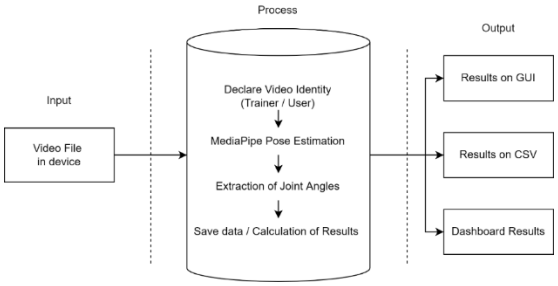

Fig. 1. Overall system block diagram.

2.1. Data acquisition (input)

Data acquisition or collection in this system takes in two distinct inputs for two different purposes. Firstly, there is a 'trainer' dataset where in this system would be any workout video that the user would want to follow. This video uploaded by user will be assumed as the accurate and proper trainer's video where the data obtained from this will be kept and used as reference for comparison usage. The other input to this system will be the 'user' where users will require to upload their workout videos into the device manually. Then, these videos collected will go through the MediaPipe pose estimation model [5] for body key point detection. This process provides 4 values including the X, Y, Z cartesian coordinates and a visibility index for all 33 key points detected as shown in Fig. 2 equating a total of 132 values for each detection. Then, the 11 head key points data as well as the Z-coordinates and visibility index of all 33 key points are removed before storing the data. Therefore, although the MediaPipe model will be generating 132 values per frame, only 21 key points along with their x and y coordinates will be stored in an array resulting in a size of 42 elements only. This step reduces the required storage memory by around 68% ensuring a lightweight system.

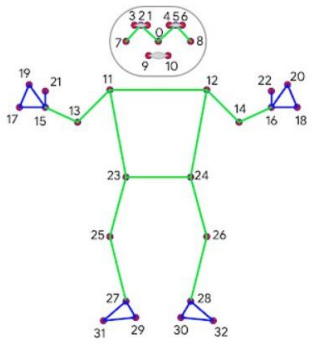

Fig. 2. MediaPipe key points [5].

2.2. Data pre-processing (normalization)

Data pre-processing in this project includes the first step of data transformation in the form of data normalization to achieve translational invariance for both user and trainer's data. By achieving this, the user will not necessarily have to be in the same position as the trainer on screen hence, providing more room for flexibility [6]. The key point normalization as shown in the block diagram will be done to standardize the centre position of the frame detected where in this project will be the nose coordinates. This condition can be achieved by subtracting all coordinates by the nose coordinate hence resulting in the nose coordinate being at (0,0) constantly [7]. Then, these data will be serialized into binary pickle files instead of recording them in the usual CSV file.

2.3. Data pre-processing (joint angles conversion)

Subsequently, these Joint Coordinates obtained will be converted into Joint Angles as seen in Fig. 3 based on the set of limbs pairing tabulated in Table 1 as inspired by [8].

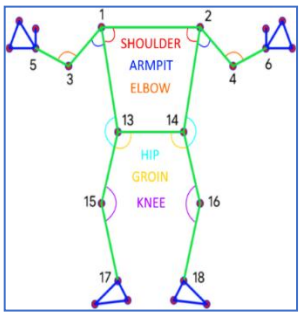

Fig. 3. Body joint angles.

Table 1. Limbs pairing for joint angles.

No.	Angle	Body Part 1	Body Part 2	Body Part 3
1	L Shoulder	R Shoulder, 2	L Shoulder, 1	L Hip, 13
2	R Shoulder	L Shoulder, 1	R Shoulder, 2	R Hip, 14
3	L Armpit	L Elbow, 3	L Shoulder, 1	L Hip, 13
4	R Armpit	R Elbow, 4	R Shoulder, 2	R Hip, 14
5	L Elbow	L Shoulder, 1	L Elbow, 3	L Wrist, 5
6	R Elbow	R Shoulder, 2	R Elbow, 4	R Wrist, 6
7	L Hip	L Shoulder, 1	L Hip, 13	L Knee, 15

$$\theta = tan^{-1}\left(\frac{y_3 - y_2}{x_3 - x_2}\right) - tan^{-1}\left(\frac{y_1 - y_2}{x_1 - x_2}\right) \quad (1)$$

θ = Joint angle at body part 2

$x_{1/2/3}$ = x coordinate of body part 1/2/3

$y_{1/2/3}$ = y coordinate of body part 1/2/3

2.4. Comparison of angles

With the joint angles obtained from both User and Trainer's input videos, these data can move onto the comparison stage using the DTW technique. The DTW technique will help to map the closest matching point even if there is a slight

misalign in timings of the time series as shown in Fig. 4. With this, the discrepancy due to different timings in the two videos can be better avoided than to compare them using the Euclidean Matching.

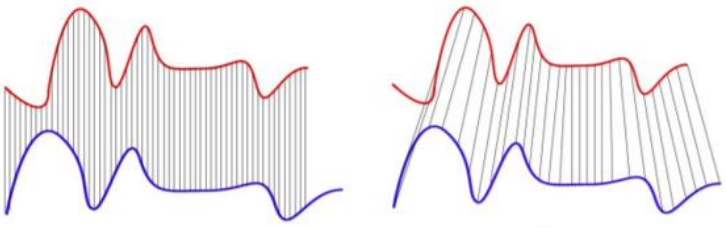

Fig. 4. Euclidean Matching vs DTW [9].

Then from the results in the form of warping distance, it will be converted into percentage form to provide a more straightforward as well as concise results for the users to acknowledge and accept.

2.5. Visualization (output)

Subsequently, the results obtained will be presented in three different forms including in the GUI itself, a pop-up dashboard of 12 graphs and a record of the results in a Microsoft Excel CSV file in the final system. Some sample images of the results will be presented in the following chapters.

2.6. Graphical user interface (GUI)

The GUI of the system is shown in Fig. 5. Multiple input requirements are seen to be required from the user, however not all are necessary when executing the 'Trainer' or 'User' functions. In fact, only the Exercise Activity Name (solid line) is compulsory to be filled for both functions. As for the Reference File Name (dashed line), it will only be required and used as the newly created pickle file name if the 'Trainer' option is selected. On the other hand, if the 'User' function is used, two file dialogs will pop up sequentially for users to select their reference file and video file to be assessed.

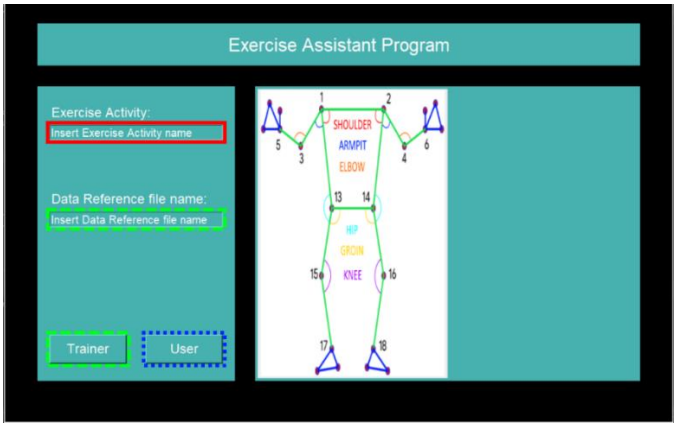

Fig. 5. GUI of the system.

3. Experiments and Results

Before moving to the two experiments done, a sample of the three forms of result visualizations are displayed as shown in Fig. 6.

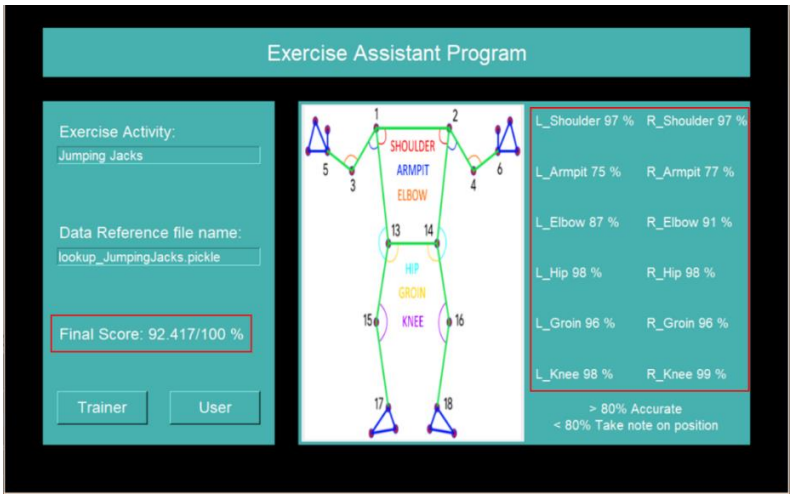

Fig. 6. Results presented in GUI.

On the right side of the interface, the 12 different accuracy score of the Joint Angles are illustrated with the average of them all being the designated 'Final Score' and presented on the left side of the interface. Then, these 13 values will be recorded in a Microsoft Excel CSV file as shown in Fig. 7.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Date Time	Activity	L_Shoulde	R_Should	L_Armpit	R_Armpit	L_Elbow	R_Elbow	L_Hip	R_Hip	L_Groin	R_Groin	L_Knee	R_Knee	Total Score	
71	Mon Mar 14 16:10:39 2022	Jumping Jacks	96	96	81	81	93	93	96	95	97	97	96	96	93.083	
72	Mon Mar 14 16:19:06 2022	Jumping Jacks	83	66	45	36	78	81	80	77	61	73	64	63	67.25	
73	Mon Mar 14 21:12:03 2022	Jumping Jacks	100	100	100	100	99	96	99	96	100	100	99	99	99	
74	Mon Mar 14 21:21:31 2022	Jumping Jacks	100	100	100	100	100	100	100	100	100	100	100	100	100	
75	Mon Mar 14 21:23:25 2022	Jumping Jacks	97	97	86	86	92	91	97	98	97	97	98	98	94.5	
76	Mon Mar 14 22:11:51 2022	Jumping Jacks	97	97	82	82	82	74	98	98	97	98	98	99	91.833	
77	Mon Mar 14 22:13:03 2022	Jumping Jacks	97	97	75	77	87	91	98	97	96	96	98	98	92.25	
78	Mon Mar 14 22:34:36 2022	Jumping Jacks	97	98	86	86	86	81	99	98	98	98	99	99	93.75	
79	Mon Mar 14 22:38:57 2022	Jumping Jacks	93	91	68	68	87	90	86	88	91	90	88	86	85.5	
80	Mon Mar 14 22:43:42 2022	Jumping Jacks	97	97	86	86	93	94	97	97	98	97	98	98	94.833	
81	Tue Mar 15 14:09:34 2022	Jumping Jacks	99	99	97	97	97	97	99	99	98	99	99	99	98.25	
82	Tue Mar 15 14:10:43 2022	Jumping Jacks	99	99	96	95	95	94	99	99	99	99	99	99	97.667	
83	Tue Mar 15 14:22:04 2022	Jumping Jacks	98	98	94	94	95	95	98	98	97	98	98	98	96.75	
84	Tue Mar 15 14:24:48 2022	Jumping Jacks	97	97	86	86	92	91	97	97	97	97	97	97	94.25	
85	Tue Mar 15 14:26:47 2022	Jumping Jacks	97	97	86	86	92	91	97	97	97	97	97	97	94.25	
86	Tue Mar 15 14:28:07 2022	Jumping Jacks	96	95	85	85	88	89	97	96	94	96	97	97	92.917	
87	Tue Mar 15 14:30:02 2022	Jumping Jacks	97	96	88	88	90	91	97	96	95	97	97	97	94.083	
88	Tue Mar 15 14:36:11 2022	Jumping Jacks	97	97	78	81	86	90	98	97	96	95	98	98	92.583	
89	Sun Mar 20 15:01:49 2022	Jumping Jacks	98	98	96	95	97	96	99	97	98	98	99	99	97.5	
90	Mon Mar 21 16:08:04 2022	Jumping Jacks	97	97	75	77	87	90	98	97	96	96	98	99	92.25	
91	Mon Mar 21 16:24:55 2022	Jumping Jacks	97	97	75	77	87	90	98	97	96	96	98	99	92.25	

Jumping Jacks Results

Fig. 7. Results recorded in CSV file.

Then, a dashboard (Fig. 8) with the graphs of all 12 sets of joint angles tracked from both user's and trainer's video will pop-up. Its main function is to help users to easily determine what are the errors causing their accuracy score to lower than expected.

Throughout the two experiments and results collection, 2 volunteers including a female and a male participant have contributed with an exercise video each which are the Mountain Climber and the Jumping Jacks exercise. These two volunteers are represented as the 'users' while the 'trainers' in this case are represented by

Pamela Reif, a famous female Youtuber with 8.59M subscribers at the point of writing and a male trainer from Allblanc TV, with 1.86M subscribers.

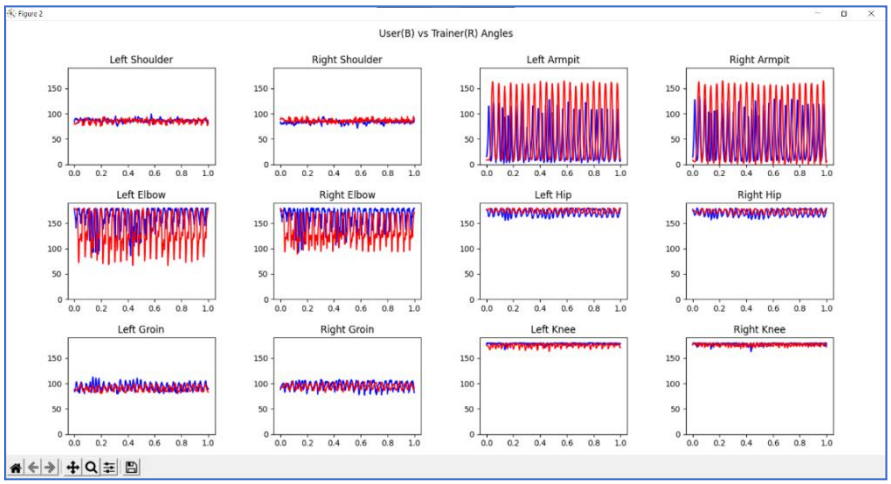

Fig. 8. Results presented in dashboard.

3.1. Consistency of the system

To demonstrate that the system can cater to different trainers' videos including both male and female trainers and even some inverted videos. Three different videos of trainers doing the Mountain Climber exercise (as shown in Fig. 9) were incorporate to prove this point. These include a video from Pamela Reif (Female trainer in Fig. 10), Allblanc TV (Male trainer in Fig. 11) and the inverted version of the female trainer's video as shown in Fig. 12. The goal of this experiment is to check when the same user video is used to be compared in all three cases via the system, can similar results be obtained.

- User's Video:

Fig. 9. User mountain climber.

3.1.1. Comparison against female trainer

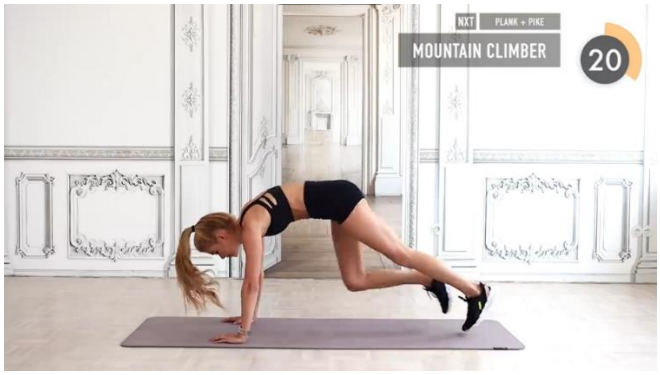

Fig. 10. Female trainer mountain climber [10].

3.1.2. Comparison against male trainer

Fig. 11. Male trainer mountain climber [11].

3.1.3. Comparison against inverted trainer video

Fig. 12. Inverted female trainer mountain climber [10].

The results obtained were relatively positive as the differences in accuracies for all Joint Angles when compared to the 3 trainers of the user are not significant. The

largest discrepancies as shown are the left hip and left groin in Case 1 and Case 2 with a difference of 13% and 14% respectively. However, these inconsistencies can easily be ignored in this case due to the posture of person being side view causing the shoulders, groins and the right sided part of the body except the knees to be mostly out of sight. Having said so, in the performance of the Mountain Climber exercise, the most important parts of the body to take note of are the armpits or elbows, hips and knees of the person. In this case, the consistencies of these results among all three cases are proven to be very high with a maximum of 5% differences across the 8 sets of data. The result is shown in Fig. 13.

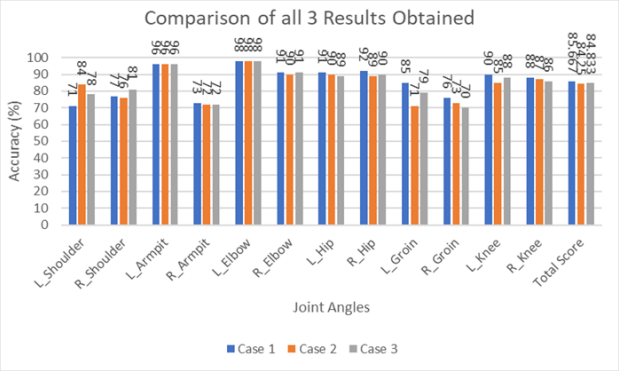

Fig. 13. Comparison of results from all 3 cases.

3.2. Detection of error in posture

In this test, the second objective of the project in providing feedback to the user using data is assessed. In preparation for this experiment, the male user recorded a set of normally executed Jumping Jacks exercise (as shown in Fig. 14) and another where his arms are only raised to the shoulder height. These two videos will then be compared to a single trainer's video. To highlight on the contrast, the focus of this experiment will be on the armpit angles for both sets of the exercise done. Then, using the comparison between the two cases, the data presented on the dashboard can be judged whether they are informative enough. The outcome is shown in Figs. 15 and 16.

- Trainer's Video:

Fig. 14. Trainer's jumping jacks [12].

3.2.1. Case 1: correctly performed jumping jacks

Fig. 15. Trainer's jumping jacks.

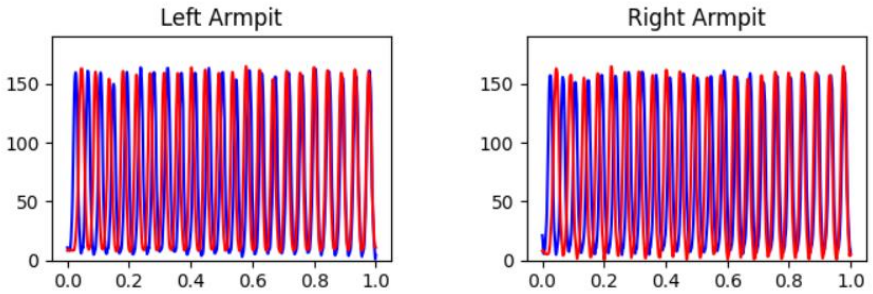

Fig. 16. User correctly performed jumping jacks dashboard results, user (b) trainer (r).

3.2.2. Case 2: correctly performed jumping jacks

From the accuracy results obtained from the system presented in Fig. 17, the expected outcome from the system was achieved. The standout information obtained are the relatively lower scoring Joint Angles which are the left armpit and right armpit. The justifications of these scores are as shown in Figs. 16 and 18. In the former image, it can clearly be seen that the blue and red lines were only slightly misaligned in the initial stage. However, they were very much in sync afterwards with the relatively same amplitudes. On the other hand, Fig. 19 illustrated the significant difference in amplitude for the red and blue lines. With this, user can understand the factor causing the decrease in similarity scores is the amplitude of the angles. More specifically, the user (blue lines) will have to increase his armpit angles by around 30° – 40° more to perform the similar movement done by the trainer. Therefore, the information on the dashboard is proven to be capable to act as feedback system for the user hence fulfilling the second objective of the project.

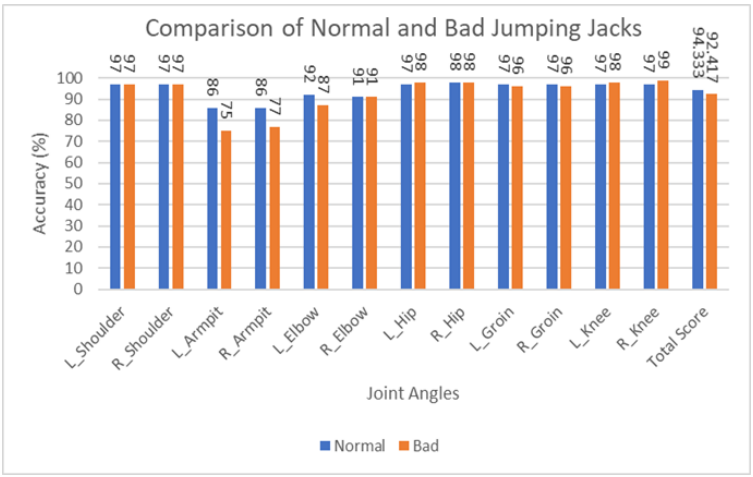

Fig. 17. Comparison of results from both jumping jacks cases.

Fig. 18. User badly performed jumping jacks.

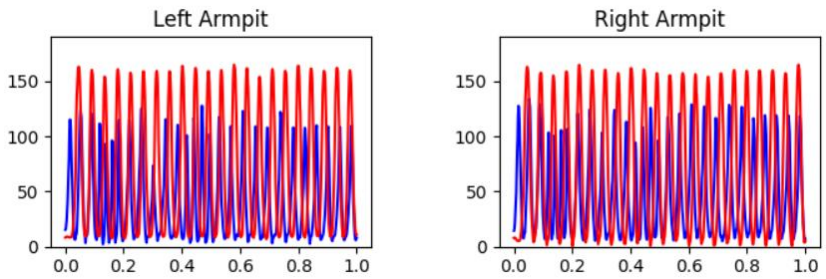

Fig. 19. User badly performed jumping jacks dashboard results, user (b) trainer (r).

3.3. Comparison with previous research

In this section, a similar project found on GitHub uploaded by Krishna Raj (https://github.com/kr1210/Human-Pose-Compare), was used to compare on the accuracy results obtained. In Krishna's project, the pose estimation model used is PoseNet. PoseNet is a free and open-source application that is available on

Tensorflow announced by Dan Oved et al. in a collaboration with Google Creative Lab. PoseNet generates poses, pose confidence scores, key point placements, and key point confidence scores based on the model outputs. More specifically, it functions to extract 17 essential points on human body detected [13].

Other than that, he also used the Joint Coordinates instead of converting them to the Joint Angles as done in this project. To test on these two systems, two cases were formulated including comparing the Trainer's Jumping Jacks video (21 seconds) to itself with the expectation of achieving 100% similarity. The comparison result is shown in Figs. 20 and 21. In the second case, the 'user' video used will be the original video but edited to end 1 second earlier (20 seconds). Therefore, in the second case, the similarity results are expected to be near to 100% as well. The result is shown in Figs. 22 and 23.

3.3.1. Case 1: testing on a 100% identical video

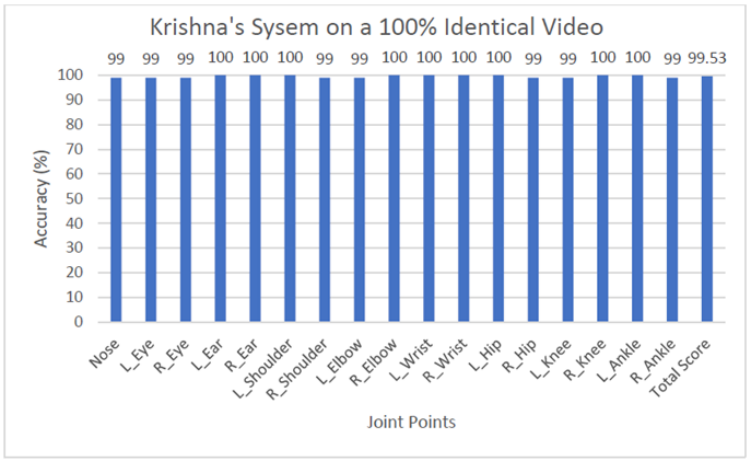

Fig. 20. Krishna's system on a 100% identical video.

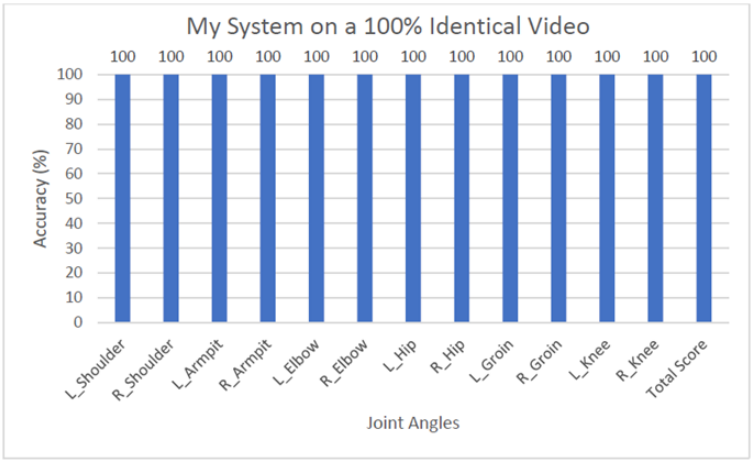

Fig. 21. My system on a 100% identical video.

3.3.2. Case 2: testing on the same video but edited to be 1s shorter

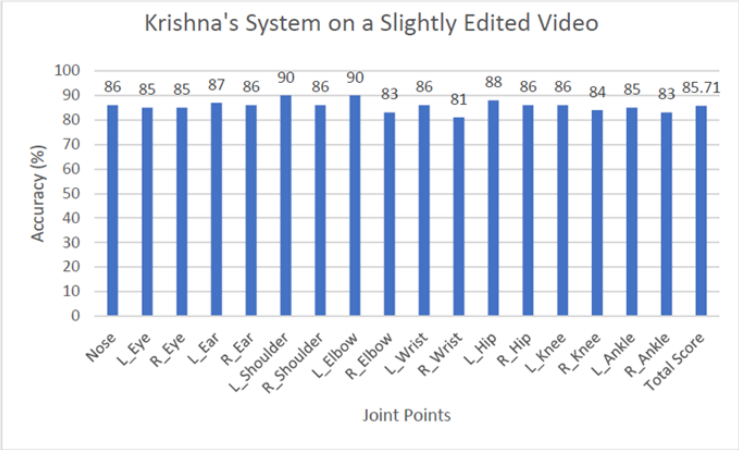

Fig. 22. Krishna's system on a slightly edited video.

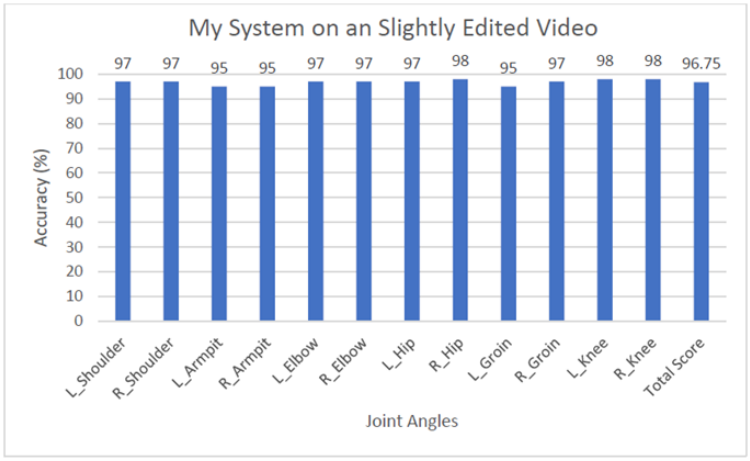

Fig. 23. My system on a slightly edited video.

From the 4 bar charts results obtained, it shows that Krishna's System was not able to achieve the first expectation of reading a 100% similar video. This may be caused by the method of normalization used by Krishna where he takes the average against a single image resulting in this drawback. Other than that, in case 2, when a slight change was done to the length of the video, the results dropped by around 15% in Krishna's system while in this project, the drop in accuracy only happened by less than 4% which is around the percentage decrease is length of the video. To conclude from this, the normalization and comparison algorithm done in this project is more accurate in achieving the goal of human pose comparison and scoring.

4. Conclusion

All in all, the aim of the research to design and develop an exercise assistant program to guide and correct its user in performing accurate exercise posture and

present the data through a dashboard was proven to be a success. This was justified by the achievements of all three Research Objectives proposed through the implementation of the MediaPipe AI model to extract the relevant data, display of accuracy results on the GUI with further details of it in the dashboard and a accuracy grading system that was validated to be reliable.

One of the limitations of the developed project include inability to function in real time. The disadvantage brought by this issue include user will only be able to know their mistakes after the exercise session when they upload their videos into the system. It may be inefficient to some people. However, the upside of this is that people with lower hardware specification can be accommodated as the computational power required to get accurate results in real time is very high and will only be affordable for very less of the population. The other limitation of this system is the requirement of the trainer's videos being taken from a constant angle to be compatible. This requirement removes many of the trainer's video found on YouTube as most of them have a videographer who will move around to show different angles of the same exercise. Having said so, Pamela Reif who currently has 8.55M subscribers and is one of the most famous workout trainers on YouTube uploads most of her videos in the format where its suitable for the system.

One of the recommendations to be made for the project done is to add a function to identify the important joints of the exercise done before calculating the overall score. Two possible approaches include the incorporation of machine learning to educate the system to recognize the exercises then identify the significant joints. From this, the least significant joints can be adjusted to have lower weightage on the final accuracy score. The second method which does not require machine learning is to take the visibility score given from the MediaPipe model into account to lower the weightage of that specific joint if it not visible by the AI model. Other than that, an additional feature to allow users to record their video from the system straight away can also be a plus point to the system. With this, user can save time and effort from uploading their videos post workout hence, increasing the efficiency.

References

1. Elengoe, A. (2020). COVID-19 outbreak in Malaysia. *Osong Public Health and Research Perspectives*, 11(3), 93-100.
2. Sohrabi, C.; Alsafi, Z.; O'Neil, N.; Khan, M.; Kerwan, A.; Al-Jabir, A.; Losifidis, C.; and Agha, R. (2020). World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *International Journal of Surgery*, 76, 71-76.
3. Sharma, A.; Madaan, V.; and Petty, F.D. (2006). Exercise for mental health. *The Primary Care Companion to The Journal of Clinical Psychiatry*, 8(2).
4. Nyenhuis, S.M.; Greiwe, J.; Zeiger, J.S.; Nanda, A.; and Cooke, A. (2020). Exercise and fitness in the age of social distancing during the COVID-19 pandemic. *The Journal of Allergy and Clinical Immunology: In Practice*, 8(7), 2152-2155.
5. Bazarevsky, V.; Grishchenko, I.; Raveendran, K.; Zhu, T.; Zhang, F.; and Grundmann, M. (2020). Blazepose: On-device real-time body pose tracking. *Computer Vision and Pattern Recognition*, 1-4.

6. Schneider, P.; Memmesheimer, R.; Kramer, I.; and Paulus, D. (2019). *Robocup 2019: Robot world cup XXIII*. Chapter: Gesture recognition in RGB videos using human body keypoints and dynamic time warping. Springer. 281-293.
7. Eivindsen, J.E.; and Kristensen, B.Y. (2020). *Human pose estimation assisted fitness technique evaluation system*. Master thesis, Norwegian University of Science and Technology.
8. Nagarkoti, A.; Teotia, R.; Mahale, A.K.; and Das, P.K. (2019). Realtime indoor workout analysis using machine learning and computer vision. *Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Berlin, Germany, 1440-1443.
9. Li, H.; Liu, J.; Yang, Z.; Liu, R.W.; Wu, K.; and Wan, Y. (2020). Adaptively constrained dynamic time warping for time series classification and clustering. *Information Sciences*, 534, 97-116.
10. Reif, P. (2020). 12 min happy sweat workout - good mood cardio workout / including HIIT I Pamela Reif. Retrieved January 5, 2022, from https://www.youtube.com/watch?v=06cwQlikBEA.
11. Allblanc TV. (2020). No gym full body workout (feat. 5 min tabata). Retrieved January 5, 2022, from https://www.youtube.com/watch?v=Tz9d7By2ytQ, 2020.
12. Reif, P. (2019). 20 min full body workout - beginner version // no equipment i Pamela Reif. Retrieved January 5, 2022, from https://www.youtube.com/watch?v=UItWltVZZmE, 2019.
13. Dan, O.; Alvarado, I.; and Gallo, A. (2018). Real-time human pose estimation in the browser with tensorflow.js. *TensorFlow Blog*.