

AI Fitness Trainer Using Human Pose Estimation

Abhinand G^{#1}, Mohammed Anas^{#2}, Naveen Kumar B^{#3}, Radha G^{#4}, Varsha Jituri^{#5}

Computer science and engineering, Visvesvaraya Technological University, Sri Krishna Institute Of Technology, Chikkabanavara, Bangalore 560090

Abstract - The COVID-19 pandemic has forced many people to work out at home, resulting in difficulties in accessing professional trainers to validate exercise postures. To address this challenge, we make use of Mediapipe, a machine learning and computer vision solution, and BlazePose, a real-time pose estimation model, to analyze exercise movements and provide real-time feedback to users. Our model enables safer and more effective home workout routines by validating posture and offering corrective suggestions, while increasing accessibility and reducing costs associated with professional physical trainers. This paper presents our proposed model, which has the potential to revolutionize the way people exercise at home.

Index terms - Mediapipe, BlazePose, Machine learning, Computer vision, Human pose estimation

I. INTRODUCTION

Physical exercise is a crucial part of maintaining a healthy mind and body. It is a form of discipline that keeps us active and sharp, and there are various ways in which people can engage in exercise. Some prefer to go for a morning walk or hit the gym, while others invest in equipment to exercise at home. However, the ongoing COVID-19 pandemic has made it increasingly challenging for individuals to access gyms and exercise facilities, limiting their options for staying active. The COVID-19 pandemic and associated stay-at-home orders have led to a significant increase in home fitness, as well as a rise in the number of people starting their fitness journey[1].

Regular exercise is essential to prevent muscle weakness, poor cardiovascular activity, and mental distress, which

can ultimately lead to a poor quality of life[2]. Specifically, according to the National Safety Council (NSC), there were 468,000 injuries in 2019 related to exercise [1]. The importance of physical exercise has been recognized globally, as people spend more time indoors, it becomes even more important to find ways to exercise regularly.

The AI Workout trainer proposed in this paper is a solution to this problem. The project's goal is to develop an AI algorithm that utilizes pose estimation to assist individuals in exercising more effectively and efficiently within their homes. It has also been shown that older adults are interested in using mobile technology to assist them at home [11]. This AI-based trainer would determine the quantity and quality of repetitions, allowing individuals to improve their workout routine[3]. It provides an automated and cost-effective means of exercising without the need for human intervention. Unlike traditional fitness devices such as workout watches or fitness bands, the AI Workout trainer does not have any limitations, making it an ideal solution for individuals who need flexibility in their exercise routines.

The model is based on human pose estimation and can be used by anyone with a webcam. The trained model that can use image-based observations to recover the pose of an articulated body, which comprises both joints and rigid parts[4]. It calculates the angles made by a person's body parts during exercise and compares them to the predefined angle range for each exercise. If the person follows the range, the system provides positive feedback, but if they do

not, it offers corrective feedback to ensure that they exercise correctly. The user can define the number of repetitions they wish to perform, or the model has a predefined count. The complete application comprises two primary components that are capable of capturing a video of an exercise and offering feedback to the user[5].

In this paper, we present the AI Workout trainer as an innovative solution to overcome the challenges of exercising during the pandemic as well as to make home workouts easier. The model's unique features and capabilities make it an ideal choice for individuals who want a cost-effective and accurate means of exercising at home.

II. LITERATURE REVIEW

The research paper [6] "Vyayam: Artificial Intelligence based Bicep Curl Workout Tracking System" presents a novel system that uses artificial intelligence to track and analyze bicep curl workouts. This system aims to provide a more accurate and efficient way of monitoring bicep curl exercises, compared to traditional methods. General technologies involved are Human Pose Estimation, AI Trainer, Virtual Skeleton, OpenCV, Mediapipe

The paper starts by discussing the importance of tracking workouts, which helps to measure progress, set goals, and avoid injury. However, traditional tracking methods, such as manual record-keeping or video analysis, can be time-consuming and prone to human error. This is where the Vyayam system comes in, as it uses machine learning algorithms to automatically detect and analyze the movement patterns of the bicep curl exercises.

The authors of the paper conducted a study to evaluate the accuracy and effectiveness of the Vyayam system. The study participants performed bicep curl exercises using this smart fitness model, and the results were compared to manual records. The results showed that the Vyayam system was able to accurately detect and analyze the bicep curl exercises, with a high level of accuracy. The study participants also reported that the real-time feedback provided by the system was helpful in improving their form and performance.

In conclusion, the "Vyayam: Artificial Intelligence based Bicep Curl Workout Tracking System" presents a promising solution for tracking bicep curl workouts. The system's use of artificial intelligence to accurately detect

and analyze the movement patterns of the exercises provides a more efficient and accurate way of monitoring progress compared to traditional methods. The study results also suggest that the system can help users improve their form and performance

The research paper [7] "BlazePose: On-device Real-time Body Pose tracking" her goal was to develop BlazePose, a

mobile-optimized lightweight convolutional neuralnetwork architecture for predicting human posture. On Pixel 2 smartphones, the network creates 33 bodykeypoints (see figure) for one person during inference and operates at over 30 frames per second. This makes it ideal for real-time applications such as fitness tracking. Two of our most significant contributions are a novel body posture monitoring method and a lightweight body pose prediction neural network. Both approaches use heatmaps and regression to find the points. They built a robust technique to estimate the posture using BlazePose, which uses CNN and a dataset with up to 25K photos demonstrating distinct body endpoints, enhancing the accuracy. On a mobile CPU, this model runs in near real-time, and on a mobile GPU, it can run in super real-time.

The given algorithm of 33 keypoint topology is efficient with BlazeFace and BlazePalm. In this paper, the authors have developed a system for majorly upper body key points. A solution that shows lower-body analysis of pose will also be integrated.

The research paper [4] "Smart Gym Trainer Using Human Pose Estimation" talks about Human pose estimation, which involves predicting the position of human joints in an image or video, has been a popular research area in computer vision and has applications in various fields, including sports and fitness. In this context, the authors of this paper present a novel approach to building a smart gym trainer using human pose estimation.

The paper begins by providing an overview of the state-of-the-art in human pose estimation and its applications in the sports and fitness domain. The authors then present their proposed system, which involves using a combination of computer vision algorithms, such as convolutional neural networks (CNNs) and optical flow, to estimate the human pose in real-time.

The authors also discuss the challenges involved in building a smart gym trainer, such as ensuring accurate and

robust pose estimation, dealing with variations in human body shapes and movements, and providing feedback to users in real-time. They present a detailed evaluation of their proposed system on a benchmark dataset and demonstrate its performance compared to existing approaches.

Overall, this paper provides a valuable contribution to the field of computer vision and its applications in the sports and fitness domain. The authors present a novel approach to building a smart gym trainer using human pose estimation, and demonstrate its effectiveness through a thorough evaluation. The paper highlights the challenges and opportunities in this area and provides a starting point for further research and development.

There are multiple computer vision algorithms used in the healthcare and tracking domains. For example, [13] created a chained multi-stream network for action detection and classification. Following, [14] validated an OpenPose-based markerless skeletonization algorithm for gait analysis. Next, [15] considered a real-time video sequence for human detection and motion tracking with framewise displacement and recognition via a skeletal model and a Deep Neural Network (DNN). Finally, [16] also examined real-time skeletonization and gesture recognition via the Star Skeletonization algorithm.

III. METHODOLOGY

The proposed methodology involves using computer vision and machine learning techniques to analyze human body movements during exercise, providing real-time feedback to users and enhancing the effectiveness of home workouts.

A. Image Acquisition

The first step of our methodology involves acquiring high-quality images of the user performing the workout routine. This is accomplished using a camera, which can be either built-in or an external module. The camera acts as the input component of the system, capturing real-time images of the user's movements during the exercise session. Apart from using a camera or webcam to capture real-time images of the user, the system also supports the use of recorded videos as input. In the system, the videos are fed into BlazePose module to extract the pose estimation coordinates for all the key points of the body. BlazePose module returns 33 pose landmarks, which are represented as 3D coordinates (x, y, z) in the video frame, where the z-coordinate indicates the depth of the landmark from the

camera. Along with the pose landmarks, BlazePose also provides a visibility score, which represents the likelihood that the keypoint is visible in the frame (i.e., keypoint "in-frame likelihood"). This information can be further utilized for analyzing the user's exercise form and providing feedback to improve their posture and movements[10]. This data input is critical for the success of the system and ensures that the feedback provided to the user is accurate and effective. We have tested our system with different types of cameras, including webcams and mobile cameras, and found that they are all suitable for our purpose. The continuous capture of images throughout the workout session allows for a detailed analysis of the user's movements, enabling our system to provide comprehensive feedback.

B. Pose Estimation

The next step in our methodology involves the use of the MediaPipe library for human pose estimation. We have used libraries such as OpenCV and MediaPipe which is a library using ML algorithms along with different numerical and algorithms. The MediaPipe pose estimation tool uses a 33 key points approach wherein it detects the key points and accordingly uses and studies the data set to estimate the pose. It tracks The pose from the real-time camera frame or RGB video by using the blaze pose tool that has a Machine Learning approach in pose detection. We are basically first detecting the landmark positions on the body in the video with the help of MediaPipe[9].

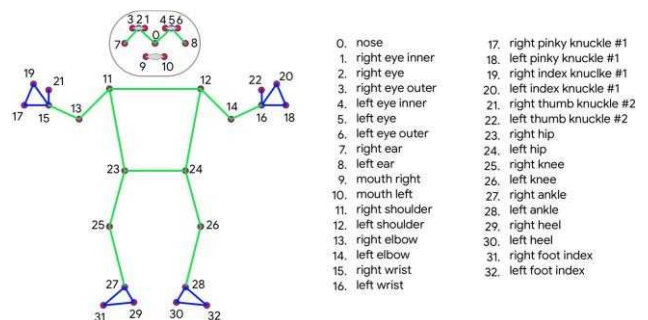


Fig. 1. BlazePose 33 keypoint topology

The 33 landmarks that are output from the MediaPipe library are indexed from 0 to 32, with the first 11 being used for facial landmarking to detect the orientation of the face. The next 11 landmarks, from 11 to 22, are used to detect the upper body, including the shoulders, elbows, wrists, hands, and fingers. The final 11 key-points, from 23 to 32, are used to detect the lower body, including the hips, knees, legs, and feet. Together, these landmarks provide an

estimate of not only the user's body structure but also their orientation in 3D space[3].

MediaPipe's pose estimation tool utilizes a method of detecting 33 key points to estimate a person's pose, which involves studying a dataset to interpret the key point data. The tool tracks the pose from real-time camera frames or RGB videos, utilizing the BlazePose tool, which has a machine learning approach to detecting poses[3]. The use of MediaPipe is crucial to the success of our system, as it enables us to accurately detect and locate the different body parts in real-time. MediaPipe is an open-source, cross-platform, customizable machine learning solution for real-time streaming media such as audio, video, and series data. It has been trained on a large dataset of human poses and can accurately estimate human body poses in different scenarios and environments.

Our system uses the pre-trained model of MediaPipe to detect the user's body pose, and the output is a list of coordinates in the X, Y, and Z-axis for each of the 33 landmarks. These coordinates are then used to construct a 3D skeletal model of the user, which is then compared to an ideal pose for the workout being performed. This comparison forms the basis of the feedback provided to the user, as we are able to identify any deviations from the ideal pose and provide corrective instructions to the user in real-time.

C. Pose Comparison

After the key-points of the user's body are detected using the MediaPipe library, the next step in our methodology is to compare the detected pose with the ideal pose for the specific workout. This comparison enables our system to provide feedback on the accuracy of the user's performance.

The output from the MediaPipe library contains the coordinates of the user's major key-points in the image. However, in order to compare the detected pose with the ideal pose, we need to extract additional information. To achieve this, a function is written in the program to extract the coordinates data and calculate the angles at each joint, such as elbows, shoulders, hips, and knees. Analytic geometry is used to calculate the angle made between the two lines formed by three key-points.

This angle calculation enables our system to provide more detailed feedback on the user's performance, allowing us

to accurately determine the deviation from the ideal pose. With this information, the user can adjust their movements to achieve a more accurate pose and receive real-time feedback on their progress.

D. User Feedback

The ability to provide real-time feedback is one of the main advantages of our system. It enables users to correct their posture and maintain proper form throughout the workout routine, which is essential for avoiding injuries and maximizing the benefits of the exercise. By setting a threshold value, our system can quickly detect any deviation from the correct position and alert the user in real-time. Our system ensures that users receive immediate feedback, allowing them to adjust their posture and improve their performance as they progress through their workout routine.

Our system is designed to provide real-time feedback and guidance to individuals performing workout routines. It uses a pose detection model and geometric analysis to detect and correct posture, ensuring accurate exercise performance. The system takes input from a live camera or video dataset and provides feedback to the user, including progress tracking and repetition counting. Here is a flow diagram depicting the design of our system

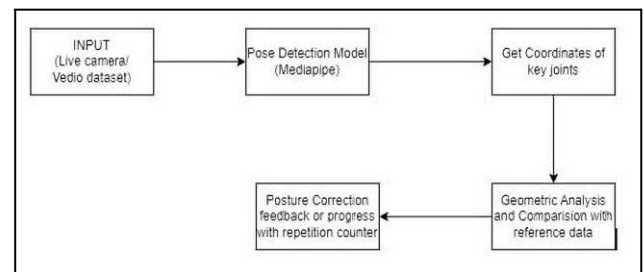


Fig. 2. System design

IV. MODEL ARCHITECTURE

BlazePose is the deep learning model architecture used in our system for human pose estimation. It is a lightweight and efficient architecture designed for real-time pose detection on mobile and web applications. It is a single-shot detector that can process an input image in one forward pass of the neural network.

BlazePose was developed by Google Research and is based on the MobileNet architecture, which is well known for its lightweight and efficient design. The main idea behind BlazePose is to use a multi-stage approach to accurately estimate human poses in real-time. The first stage of the

model known as the detector stage detects the presence of a human body in the input image, and the second stage known as the tracker stage estimates the pose of the detected body.

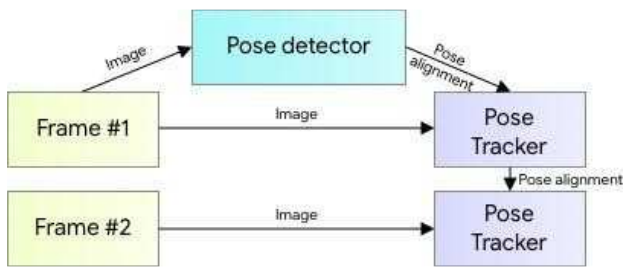


Fig. 3. BlazePose Architecture - Two stage pipeline

Detector stage: This stage involves locating the human body within the image or video frame. BlazePose uses a lightweight neural network to estimate the presence and location of human body key points in the input image. This network produces a heatmap indicating the likelihood of the key points being present at each location in the image.

Tracker stage: Once the human body key points are detected, BlazePose uses a more complex neural network to estimate the exact 3D coordinates of each keypoint, as well as the orientation of the body in 3D space. This network takes the heatmap produced by the detection stage and uses it to refine the estimation of the keypoint locations and the overall pose of the body.

BlazePose uses a combination of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to estimate human poses accurately. It also uses a novel method called "anchor-free regression" to predict the coordinates of the human body's key-points, which helps to reduce the computational cost of the model[7].

The model is trained on a large-scale dataset of annotated images called COCO (Common Objects in Context), which contains over 330,000 images of humans in various poses.

In our system, we used the pre-trained version of the BlazePose model, which was trained on the COCO dataset, to estimate the user's pose accurately. The COCO dataset, which was released by Microsoft, contains a wealth of information related to object detection, image classification, image segmentation, and semantic text description. On the other hand, the poseTracker dataset consists of a series of video clips extracted from the raw video of the MPII dataset. These video clips are composed

of 41-298 adjacent frames that have been selected for analysis[8]. The model takes the image input from the camera or webcam and predicts the 2D key-points of the user's body. The output from BlazePose is then used in the subsequent steps of our methodology, such as pose comparison and feedback to the user.

V. RESULTS

The images presented depict individuals performing push ups correctly, captured from a video dataset titled "Pushup Exercise-Correct". Each image shows the subject in a different stage of the pushup movement, starting from the starting position and ending in the fully extended arm position. The dataset contains various videos that show a variety of individuals demonstrating proper pushup form, showcasing the exercise's correct technique. Push Ups are an effective bodyweight exercise that work multiple muscle groups simultaneously, including the chest, shoulders, triceps, and core. However, improper form can lead to injury or reduced effectiveness of the exercise.

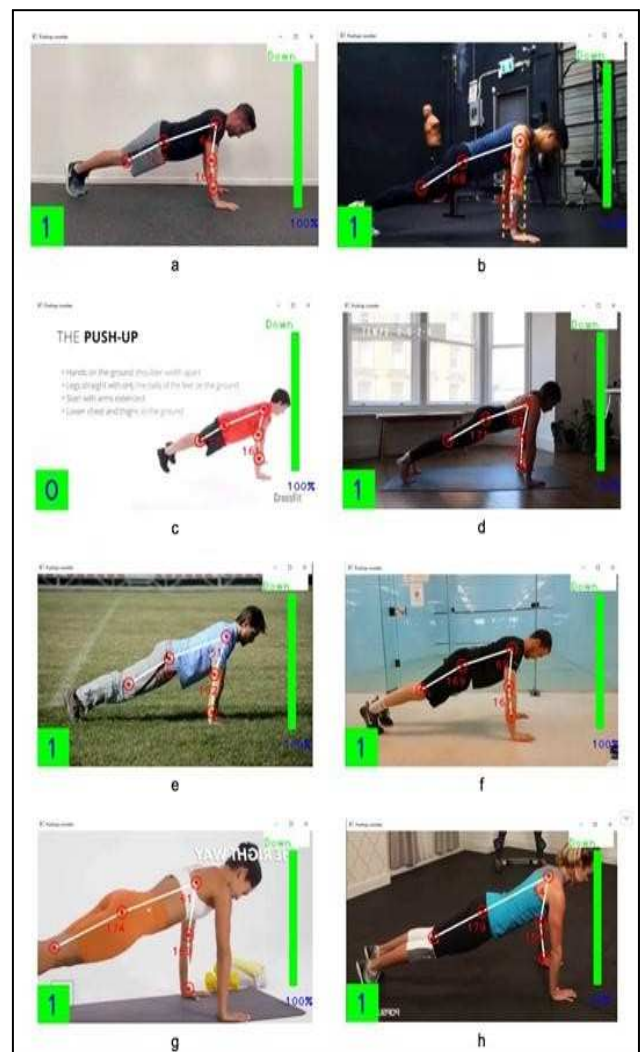


Fig. 4. People performing push up correctly in upward position

The up position of a pushup exercise performed correctly is typically characterized by a straight line from the head to the heels with the arms fully extended and the hands placed shoulder-width apart or slightly wider. The shoulders should be directly over the hands. In this position, the body should be lifted off the ground, with the chest, shoulders, and triceps muscles activated.

Below results are obtained from the above Fig. 4.

Standard angles for Push up exercise UP position-Correct					
Fig. no	Elbow >160	Shoulder >40	Hip >160	Correct/Incorrect	Position
a	164	64	173	Correct	Up
b	167	59	168	Correct	Up
c	161	56	172	Correct	Up
d	169	61	179	Correct	Up
e	162	51	170	Correct	Up
f	167	69	169	Correct	Up
g	169	61	174	Correct	Up
h	160	54	179	Correct	Up

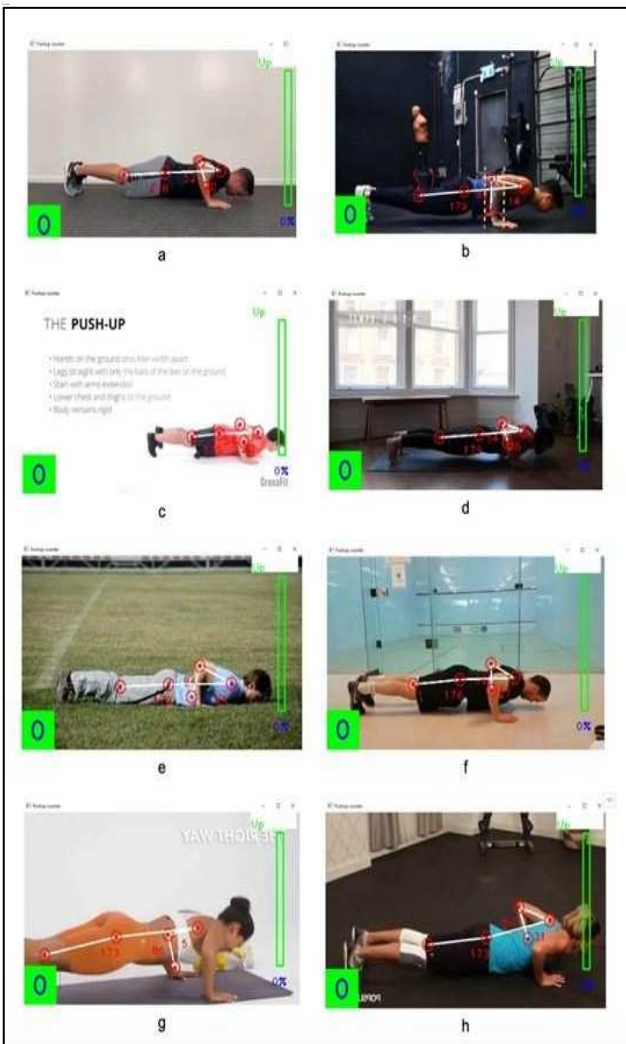


Fig. 5. People performing push up correctly in downward position

In the down position of a pushup exercise performed correctly, the body is lowered toward the ground while

maintaining a straight line from the head to the heels. The arms are bent at the elbows, with the forearms and hands supporting the body weight. The chest should be close to the ground, and the elbows should be tucked in close to the body. The core muscles should be engaged to maintain stability and proper alignment throughout the movement.

Below results are obtained from the above Fig. 5.

Standard angles for Push up exercise DOWN position-Correct				
Fig. no	Elbow <=90	Hip >160	Correct/Incorrect	Position
a	52	178	Correct	Down
b	82	179	Correct	Down
c	54	175	Correct	Down
d	54	179	Correct	Down
e	84	178	Correct	Down
f	62	176	Correct	Down
g	86	173	Correct	Down
h	50	173	Correct	Down

The below images include people performing push ups incorrectly, the images were taken from several videos from a video dataset named “Pushup Exercise-Incorrect”.

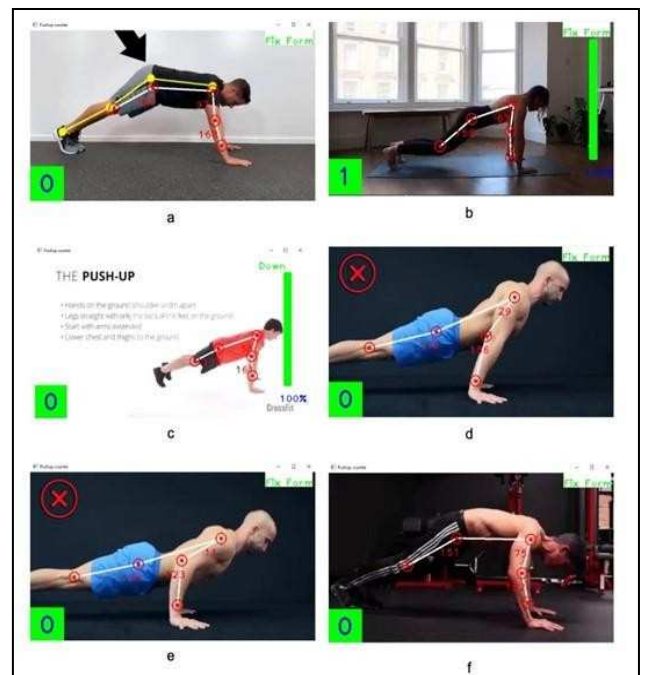


Fig. 6. People performing push up incorrectly

Above Fig. 6, represents the position of pushup exercise performed incorrectly. However, in general, an incorrect pushup position can take many forms, including: Drooping or sagging of the lower back: This can put excessive strain on the lower back muscles and increase the risk of injury, Allowing the hips to drop or pick up: This can cause instability in the body and reduce the effectiveness of the exercise, Hunching the shoulders: This can cause

unnecessary tension in the neck and shoulders, and reduce the effectiveness of the exercise.

Below results are obtained from the above Fig. 6.

Standard angles for Push up exercise DOWN position-Incorrect					
Fig. no	Elbow >160	Shoulder >40	Hip >160	Correct/Incorrect	Incorrect Angle
a	168	97	149	Incorrect	Hip
b	170	63	158	Incorrect	Hip
c	157	79	150	Incorrect	Hip
d	170	29	156	Incorrect	Shoulder, Hip
e	123	11	174	Incorrect	Elbow, Shoulder
f	152	75	151	Incorrect	Elbow, Hip

In the images below, you'll see individuals demonstrating the bicep curl exercise in both the upward and downward positions. These images were selected from a video dataset named "Bicep Up/Down"

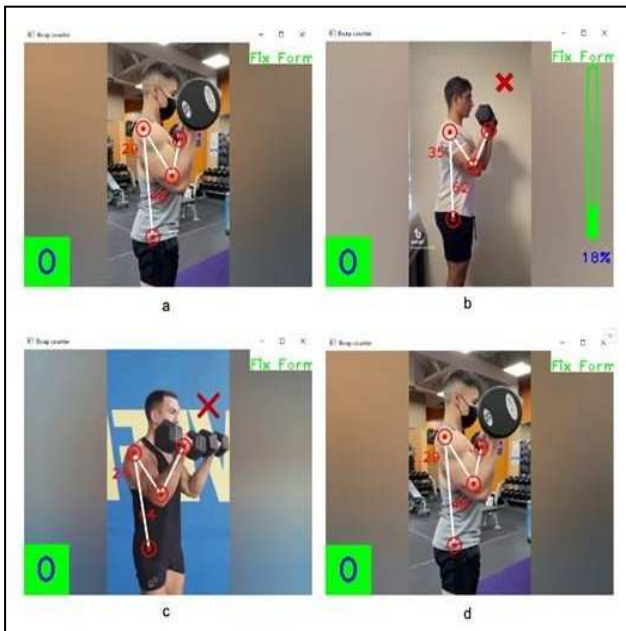


Fig. 7. People performing Bicep Curl - Up position

Above, Fig. 7 represents the up position of bicep curl exercise in the upward position. However, To perform a bicep curl correctly, stand straight with your feet hip-width apart, hold dumbbells in each hand with palms facing up, and curl the dumbbells towards your shoulders while keeping elbows close to your sides. Lower the dumbbells back to starting position slowly.

Below results are obtained from the above Fig. 7.

Standard angles for Bicep Curl exercise DOWN position				
Fig no.	Elbow >160	Shoulder <= 10	Correct/Incorrect	Incorrect angle
a	173	3	Correct	-
b	126	24	Incorrect	Shoulder, Elbow
c	161	4	Correct	-

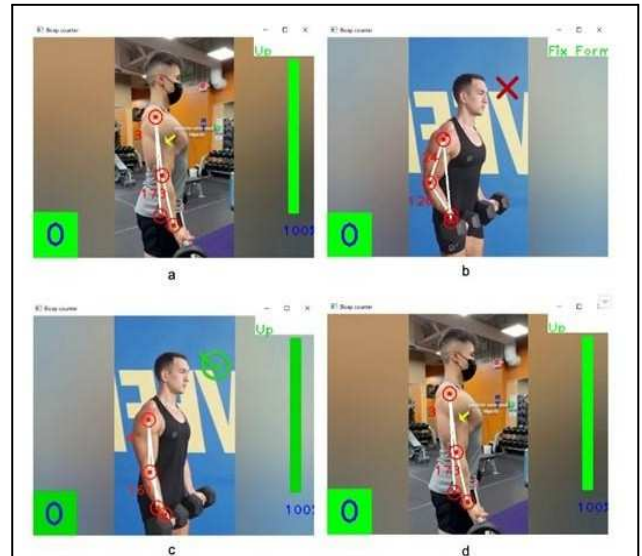


Fig. 8. People performing Bicep Curl - Down position

Below results are obtained from the above Fig. 8

Standard angles for Bicep Curl exercise DOWN position				
Fig no.	Elbow >160	Shoulder <= 10	Correct/Incorrect	Incorrect angle
a	173	3	Correct	-
b	126	24	Incorrect	Shoulder, Elbow
c	161	4	Correct	-

VI. CONCLUSION

Our proposed model is user-friendly, easy to use, and provides precise results. By using this AI workout trainer, individuals can receive perfect guidance during their workout routine and obtain feedback on every step they take. This is especially important during the pandemic or situations where going to the gym or purchasing exercise equipment can be difficult. The AI workout trainer takes the user's live video as input, generates landmarks on the human body, and provides feedback on posture and settings, making it easy to interact with. We believe that our proposed AI workout trainer can revolutionize the way individuals perform their exercise routines, and we look forward to further exploring its potential.

VII. REFERENCES

- [1] Flores, Asia, et al. "Verum fitness: An AI powered mobile fitness safety and improvement application." *2021 IEEE 33rd international conference on tools with Artificial Intelligence (ICTAI)*. IEEE, 2021.
- [2] Akpan, Aniebi, and Ahmed Aldabbagh. "Remote Body Fitness Monitoring System with Inter-User/Multi-user tracking Software Applications and Social Distancing Warning Sensor." *2020 2nd International Conference on Electrical, Control and Instrumentation Engineering (ICICLE)*. IEEE, 2020.
- [3] Taware, Gourangi, et al. "AI-based Workout Assistant and Fitness guide." *2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA)*. IEEE, 2022.
- [4] Dsouza, Grandel, Deepak Maurya, and Anoop Patel. "Smart gym trainer using Human pose estimation." *2020 IEEE International Conference for Innovation in Technology (INOCON)*. IEEE, 2020.
- [5] Kanase, Rahul Ravikant, et al. "Pose Estimation and Correcting Exercise Posture." *ITM Web of Conferences*. Vol. 40. EDP Sciences, 2021.
- [6] Samhitha, G., et al. "Vyayam: Artificial Intelligence based Bicep Curl Workout Tracking System." *2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICES)*. IEEE, 2021.
- [7] Bazarevsky, Valentin, et al. "Blazepose: On-device real-time body pose tracking." *arXiv preprint arXiv:2006.10204* (2020).
- [8] Chen, Lvcai, Chunyan Yu, and Li Chen. "A multi-person pose estimation with LSTM for video stream." *2019 3rd international conference on electronic information technology and computer engineering (EITCE)*. IEEE, 2019.
- [9] "Composite fields for human pose estimation" by S Kreiss, L Bertoni, and A Alah, *IEEE Conference on Computer Vision and Pattern Recognition* pages 11977–11986, 2019. 1.
- [10] Mroz, Sarah, et al. "Comparing the quality of human pose estimation with blazepose or openpose." *2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART)*. IEEE, 2021.
- [11] J. C. Hughes, T. Banerjee, G. Goodman, and L. Lawhorne, "A preliminary qualitative analysis on the feasibility of using gaming technology in caregiver assessment," *Journal of Technology in Human Services*, vol. 35, no. 3, pp. 183–198, 2017.
- [12] M. Zolfaghari, G. L. Oliveira, N. Sedaghat, and T. Brox, "Chained multi-stream networks exploiting pose, motion, and appearance for action classification and detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2904–2913.
- [13] M. Zago, M. Luzzago, T. Marangoni, M. De Cecco, M. Tarabini, and M. Galli, "3d tracking of human motion using visual skeletonization and stereoscopic vision," *Frontiers in bioengineering and biotechnology*, vol. 8, p. 181, 2020.
- [14] S. Win and T. L. L. Thein, "Real-time human motion detection, tracking and activity recognition with skeletal model," in *2020 IEEE Conference on Computer Applications (ICCA)*. IEEE, 2020, pp. 1–5.
- [15] K. Srijeyanthan, A. Thusyanthan, C. Joseph, S. Kokulakumaran, C. Gu-nasekara, and C. Gamage, "Skeletonization in a real-time gesture recognition system," in *2010 Fifth International Conference on Information and Automation for Sustainability*. IEEE, 2010, pp. 213–218.