Physio AI Companion

Faculty Advisor: Ben Ochoa, bochoa@ucsd.edu
Domain Advisor: Mr. Kevin Duenas MS Exercise Physiology, TPI, PES, CES, FMT from Dynamic Fitness; Mr. Brian Richardson MS Kinesiology, CPL3, PES, CES, FMT from Dynamic Fitness; Dr. Malerie Franco, DPT, PT
Team Members : Vaaruni Desai v1desai@ucsd.edu, Laben Fisher lffisher@ucsd.edu, Zufeshan Imran zuimran@ucsd.edu, Sagar Jogadhenu sjogadhenu@ucsd.edu,Prakhar Shukla prshukla@ucsd.edu

Abstract

Physical Therapists (PT) and Kinesiologists recommend a series of exercises but often face challenges in continuously monitoring individuals performing exercises to ensure correct postures and prevent injury aggravation. This research attempts to address this issue by building a product designed to automate the detection of the incorrect exercises and provide users with timely feedback. The research effort began with a set of curated exercise videos, a set of biomechanical standards as well as developing the Physio AI Core Model to analyze a single exercise - overhead squat. The core model approach consists of three main steps: preprocessing to standardize the videos, creating a 3D pose estimation model and calculating incorrectness scores for each repetition and aggregation based on measured joint angles. The work uses state-of-the-art computer vision models and computational algorithms for a customized solution. The results from the Physio AI Core Model are used to provide feedback to both practitioners and users through visual overlays on the exercise video and graphical presentation of biomechanical measures captured during the exercise.

Introduction

Individuals, clients, seek help from Physical Therapists (PT) and Kinesiologists for many reasons, such as medical rehabilitation, athletic improvements, mobility concerns, etc. One of the first activities that is performed by the PT and Kinesiologist is to assess clients to determine biomechanical issues are observed and to set goals. A set of evaluation exercises are performed to conduct the assessments. The PT and Kinesiologist do not have an effective way to capture the clients movement and show the information to the clients with immediate feedback and currently need to use demonstrations and/or manual manipulation. PT and Kinesiologist also do not have an easy way to generate quantitative measurements associated with these assessments resulting in subjective assessments based on the experience of the PT. The current approach requires motion capture systems that are both expensive and require a dedicated space, less expensive systems require the client to wear/attach sensors to specific parts of their body and do not perform very well due to lack of low accuracy and precision measurement sensors.

Once the evaluation/assessment is complete, PT and Kinesiologist prescribe a series of exercises that clients will go through over a period of time to achieve the desired outcome or goal. While a portion of the exercises are performed at the practitioner's facilities, likely some exercising at the clients home or at another facility will occur. Full time observation does not occur in the office or gym and little to no observations occur while the client is at home. Compounding these issues is the increased demand for telehealth/remote sessions that expanded rapidly during COVID-19 pandemic and continues post pandemic. Without direct supervision or some effective method of sharing clients results from home, some clients might not follow through with the exercises affecting their goals and outcomes. Even more importantly, incorrect biomechanics might go unnoticed increasing the potential for an injury.

The PT, Kinesiologist, and clients would like to have an easy way to document and visualize progress overtime (e.g., show progress in biomechanics, strength, flexibility, or mobility). There currently lacks an easy way to capture the clients performing the exercises and capturing quantitative data to provide the necessary feedback and history beyond manually typed notes entered in by the PT/Kinesiologist. This lack of frequent feedback can lead to demotivation by the client and pursuing less effective routines/programs.

Based on the above understanding of the problem the following key series of questions arises as the focus of this application level research effort.

- As the Client:
    - How can I know if I am doing the prescribed exercise correctly (aka correct biomechanics)?
    - How can I more effectively perform my exercises at home by enabling an accountability partner to see progress visually and quantitatively?
- As the PT and Kinesiologist:
    - How can I automate the gathering of quantitative metrics during the assessment with a client?
    - How can I help ensure my clients are performing exercises with correct biomechanics while I am not able to watch them perform their exercises?
    - How can I track my client's progress with visual and quantitative data on a regular basis and provide necessary updates (with or without requiring an in-person office visit)?
    - How can I analyze my practice to see what are the trends and what recovery or training plans have been most effective?

The list of questions above were discussed with the domain advisors and the following three use cases were identified to guide the research:

1. Increasing the effectiveness of health assessments
2. Increasing the effectiveness of unsupervised exercise
3. Increasing the practitioner's visibility on exercise programs' effects on their client base

To address these questions research was conducted to understand the state-of-the-art in 3D pose estimation, computational techniques vs computer vision approaches for repetition identification and calculating incorrectness scores along with some analysis on how best to visualize this data to most effectively communicate the results as they apply to the problems listed above.

Related Work

The initial step in the process is to understand the current state of the art for each of these relevant research categories. For each topic, here is a list of key research papers and references that reflect the current state of the art:
Related works in the field of self monitored and home based exercise monitoring, biomechanical analysis, and computer vision-based pose estimation includes various approaches aimed at enhancing the effectiveness and safety of physical rehabilitation and exercise regimes. Studies such as Clark et al.'s research[4] on real-time posture monitoring using Kinect sensors, which demonstrates the potential of video-based pose estimation for quantitative evaluation of human motion.
Reviews such as Zheng et al.'s comprehensive overview[5] of deep learning-based human pose estimation and Stenum et al.'s survey[6] about applications of pose estimation in human health and performance across lifespan provide valuable insights into the state-of-the-art techniques relevant to the development of automated monitoring systems.
Webster et al.'s systematic review[7] on human pose estimation for physical rehabilitation further emphasizes the growing interest in leveraging deep learning for automated exercise monitoring and feedback systems, aligning with the objectives of our project, providing a foundational understanding of current methodologies and innovations in the field.

Team Roles and Responsibilities

- Laben Fisher: Project Manager, Stakeholder Engagement, Data Engineer
- Sagar Jogadhenu: Solution Architect, Test Lead, User Interface Developer, DevOps
- Zufeshan Imran: Budget Manager, Methods Expert, ML engineer
- Vaaruni Desai: Methods Expert, Data Analysis, ML Engineer
- Prakhar Shulka: Data Scientist, User Interface Developer, QA Engineer
- All: Data Collection, Record Keeper, Report Writer, Peer Reviewer

Data Acquisition

The data acquisition phase was critical to gather the necessary raw input for subsequent analysis and model training. This section provides a comprehensive overview of the data sources utilized, the systematic steps taken for data collection, and the detailed data pipeline involved.

Data Sources

Our raw data source is a set of videos of people performing exercises. For this research project, the exercise set is limited to an overhead squat. During the initial phase of the project videos from YouTube as well as a small set of videos of exercises performed by the team members were used (Set1 and Set2 in table 1 below). As the initial proof concept proved the product to be viable, additional public domain videos were collected from a gym (Set3 in table 1) as well as team members performing exercises on a regular basis (Set4 intable 1).

Dataset Name	Source	Destination	Data Size	Notes
Exercise Videos (Set1)	Videos from YouTube	Google Drive	20 MB	Publicly available data
Exercise Videos (Set2)	Device captured videos from a single phone camera	Google Drive	126 MB	Releasability pending IRB approval
Curated Videos (Set3)	1 video from Set1 3 videos from Set2 30 videos from gym	Google Drive	288 MB	New data provided by gym, not releasable
Exercise Videos (Set4) Final Dataset	24 videos from Set3 videos captured from product web interface	Amazon Web Services (AWS) S3	606 MB	Developed for testing purposes, data not releasable

Table 1: Raw data sources

Data Collection

For this research effort there were several datasets used. This section will discuss the initial datasets, the EDA performed on the initial dataset, and then how we expanded our data with the development of a web application.

Initial Datasets

During the course of this project data was collected in several phases, initially from youtube for early exploratory data analysis (EDA) (Set1), then further expanded the datasets with additional data collected from a variety of single phone's cameras (both iPhone and Android cameras) (Set2), then from a camera setup that consisted of three synchronized cameras (Set3).

Datasets Set1, Set2 and Set3 were initially stored on Google Drive. The gym dataset (Included in Set3) was collected at a local gym facility using a three camera setup as shown in Figure 1.This setup facilitated the capture of videos from 3 different viewing angles.

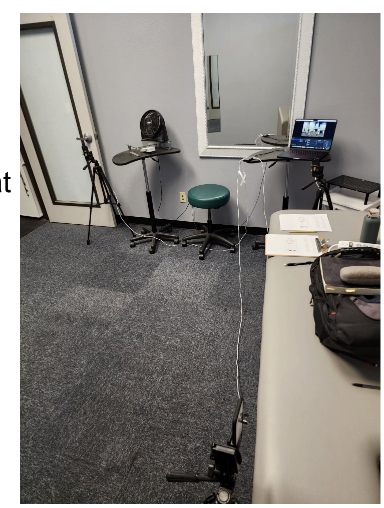

Figure 1: Three Camera Setup

EDA of Initial Dataset

The initial input data consists of the video data from Set1 and Set2. The initial EDA of the video data sets the following findings listed below and shown in Figure 2.

- The raw dataset is comprised of 63 videos
- Each video has 45 parameters and only a subset of them such as frame count, frame width, frame height, frame rate are needed for further analysis
- The majority of the videos are 10 seconds or less in duration with ~300-500 frames per video
- The frame rate for majority of the videos is standard 30 frames per second
- Approximately 50% of videos are recorded at the 1080x1920 resolution
- 81% of the videos are oriented in portrait mode and the remaining are in landscape mod
- Videos with a resolution below 640x480 or 480x640 are not used

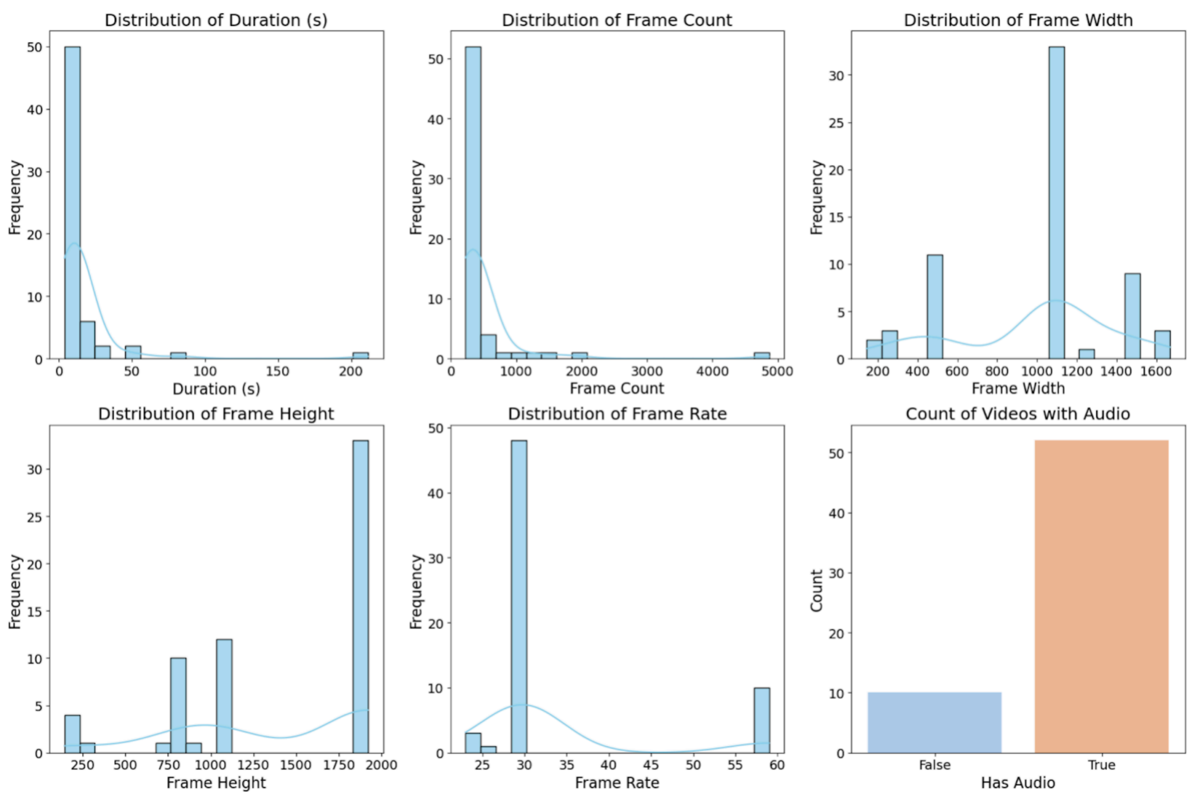

Figure 2: EDA for input data

With the EDA completed, we established the preprocessing steps needed to modify raw video into a condition suitable for the subsequent steps in the data pipeline, as shown below.

1. Extract metadata from each video
2. Depending on the orientation (i.e., portrait/landscape) of video, resize frames to be 480x640 or 640x480 for standardization of input. The purpose of this resizing is to reduce storage requirements for our products
3. Discarded videos less than 640x480 or 480x640 resolution
4. Exclude videos below a minimum threshold based on EDA finding
5. Convert RGB frames to grayscale frames to reduce compute load
6. Write processed frames to a new video file and store them in the a shared folder inGoogle Drive

Data Collection - Web Application

After validation of the hypothesis during the initial stages of this work an additional capability was needed that could capture video while reducing the amount of unnecessary movement in the video (e.g., walking back and forth from the camera). The Capture web application UI was developed to address this issue and enable users to capture videos for themselves. This interface not only captures exercise videos but also integrates the input video into the data pipeline for processing. The functionality and design of Capture UI are shown in Figure 3.

On the Capture UI web page a red border is presented with the request that the user place their whole body within the border. A countdown timer enables the user to get into place after pushing the start (red) button. To reduce the amount of video that captures the user walking to the camera, the video capture automatically stops after a set amount of time (30 seconds). This user interface reduces the amount of motion that is captured not related to the actual exercise being performed, increasing the accuracy of the results that will be generated and presented to the user.

Dataset Set4 was created using this data as well as curracted data from Datasets Set1-3 and stored in AWS S3.

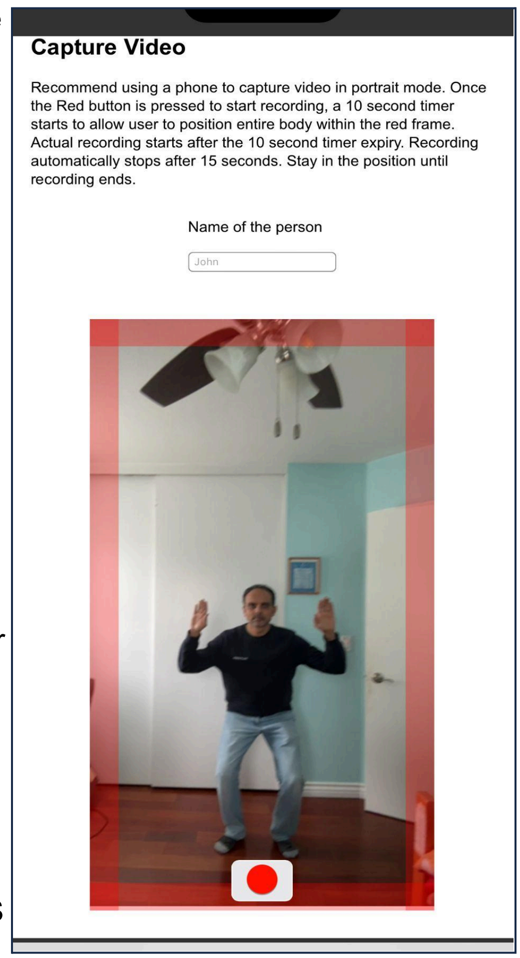

Figure 3: Physio AI Companion Capture User Interface

Data Pipelines

Based on the objectives of the research activity and the understanding of the data conducted during the EDA process, a data pipeline was generated as depicted in Figure 4. Datasets 1-3 were manually fed into the pipeline to develop and test the data pipeline in the early stages of the research. Once the product developed to capture video directly, the pipeline was automatically fed video data.

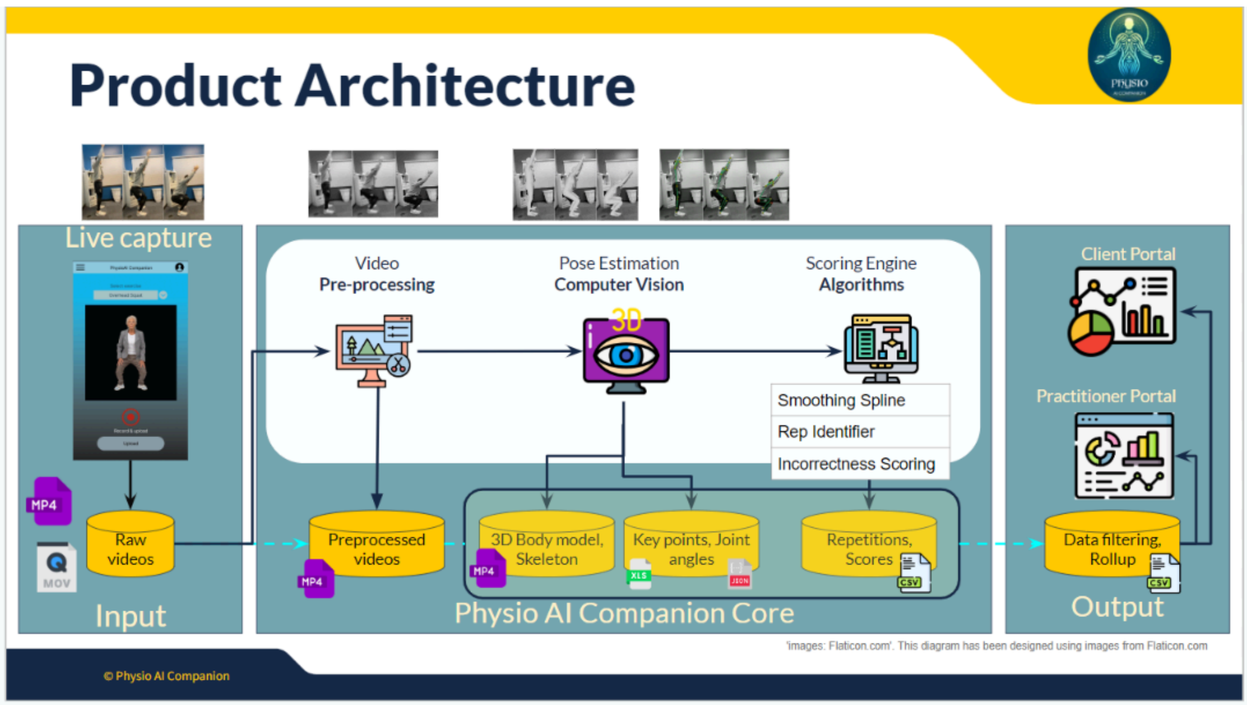

Figure 4: Physio AI Companion Pipeline

The key components of the data pipeline are as follows:

Step 1 - Data Collection: Videos of users performing exercises are acquired
Step 2 - Preprocessing: Videos were preprocessed as described above
Step 3 - Pose Estimation: The World-grounded Humans with Accurate 3D Motion (WHAM)[1] model generates 3D body models and key points
Step 4 - Scoring Engine - Repetition Identification: Joint angle data and smoothing algorithms are used to count repetitions
Step 5 - Scoring Engine - Incorrectness Scoring: Joint movements were assessed against predefined thresholds to score exercise correctness
Step 6 - Data Visualization: A portal for Practitioners and Clients to view video, receive feedback, and other visual representations of the data output

Data Environment

The early exploration phase used Google Drive as a means to store input videos collected from publicly available datasets. The model and the data pipeline were implemented as a Jupyter notebook executed on a graphics processing unit (GPU) capable computer. Intermediate results from the notebook were first stored on the local computer and manually uploaded to Google Drive upon completion.

As the solution matured, the solution migrated to the AWS environment and the current implementation is fully run on AWS end-to-end. The data collection happens through a web server implemented on AWS. Captured videos are automatically published to an S3 bucket called Physioai-invideo on AWS. The Jupyter notebook runs on an AWS Sagemaker GPU instance. Intermediate results are stored on Sagemaker and final results including 3D body model, key points, joint angle measurements and scores are uploaded to another S3 bucket called Physioai-outvideo. Output data from this bucket is processed and presented to the end users via visualization pages served by the web server.

Data Preparation

This research project has several different data sets. Each of these data sets has some quality issues with them that needed to be addressed. Here is a list of each data set, where it was acquired, and the quality issues associated with it:

- Set1 videos collected from YouTube suffered from loss of sharpness and detail due to heavy compression, low resolution, variable frame rates and reduced bitrate.
- Set2 videos captured using a single mobile device (android and iPhones were used) contained some videos with the following various issues: stabilization issues due to shakiness and motion blurring; background noise; variable frame rates; low resolution; occlusions; various lighting and environments.
- Set3 contained cameras synchronized and collected via Open Broadcaster Software (OBS) apps on mobile devices that covered 3 different perspectives were relatively higher resolution and stable videos but some of the videos had occlusions or reflections (e.g., a wall mirror) of the user.
- Set4 contained videos captured from the Physio AI Companion Capture UI with variable resolution and frame rates depending upon the device being used. These videos often have background noise and motion blurring.

Due to these quality issues, several additional preprocessing steps were involved to ensure the videos met the quality and storage requirements of the model using them. The videos were preprocessed using Fast Forward Moving Picture Experts Group (FFmpeg) commands to convert them into a correctly encoded version with grayscale background and appropriate resolution that offers compression capabilities with removal of audio, as well as standardizing video outputs from the preprocessing step into a single codec, H.264. The additional preprocessing steps result in the following:

- **High-Quality Compression**: Compression reduces video file sizes significantly, which is crucial for storage management and compute processing times.
- **Consistent Aspect Ratio**: While we attempted to conform to a single aspect ratio, that produced anomalies in processing for other components of the pipeline. So instead, the preprocessing step ensures that the original aspect ratio is maintained through the transformation processes in the preprocessing step.
- **Audio Removal**: Removed unnecessary audio, saving on storage size.
- **Consolidating to a single codec**: The preprocessing ensured videos were encoded into a single codec, reducing complexity of the solution.

The data sets were further reduced to specific videos that did not contain aspects that were outside the bounds of the research activity. This research was not focused on handling multiple people in a single video as that is not the normal use case as described by the domain advisors. This research did not focus on how to track video with occlusions and other factors that would be important to object detection and tracking of objects through video frames. The research team discarded videos that contained multiple people or too many occlusions that would hamper the processing of 3D pose estimation or repetition identification.

Preprocessing Results

After developing the preprocessing steps as described above, a series of limited tests generated the following results.

- Storage Improvements: This test compared the size of original videos with the size of preprocessed videos, which showed a reduction of 80% the storage size
- Processing Time Improvements: A test was conducted to measure the processing time of a subset of six videos. While the Pose Estimation and Scoring Engine processes took 1 hour on average to process the original video, the preprocessed videos were completed in 5 min, showing a 92% improvement in the processing time
- Quality Improvements: To help understand if the preprocessing steps reduced the quality of the Physio AI Core Model, a test comparing the results of the model using the original videos vs the preprocessed videos was conducted. For this test, 6 videos were selected out of the first three data sets. The comparison was done to understand if the preprocessing steps reduced the quality of the Physio AI Core Model. The test captured the variation in the peak values (angle of a key point) generated by the Pose Estimation Model. Additionally, the test captured the alignment of the Physio AI Core Model output with the Domain Advisor’s labels (our ground truth for this effort) to determine how accurate the models were with the original video as input vs the preprocessed videos as input.The results are described in Table 2 below, showing an overall accuracy improvement of 39%

Video	Body Model Peak Value Change (Original Peak Value - Preprocessed Peak Value)	Original Video Model Alignment with Domain Advisors	Preprocessed Video Model Alignment with Domain Advisors	Percent Change of Alignment
6 Sample videos	9.81% lower peak values on average	108 out of 240 reps	150 out of 240 reps	39% Improvement
	23.91% absolute value change in peak values on average			

Table 2: Preprocessing Improvement Results

Analysis Methods

Primary input data for this project consists of videos of individuals performing exercises. The goal of the project is to compute the incorrectness of the individual's biomechanics from these videos by extracting key points and joint angles. At this point in the process EDA has been completed and a general understanding of the video datasets has been achieved. The next step is to review the data at a more granular level for each video and label the videos for correctness/incorrectness for important key points in the body namely hip, knee and ankle.

The data labeling activity involved the researchers manually reviewing each video to label the environmental conditions: occlusions, lighting, number of people, camera angle, etc. The next labeling activity involved reviewing each video and counting the number of repetitions and making an assessment of the incorrectness of the measures the Physio AI Core Model is targeting to automatically identify. This information was then used by the research team during the initial development phases of the Physio AI Core Model. During the course of the development of the Physio AI Core Model, the domain advisors became available and provided their professional assessments of the Set 3 videos. A sample of the data labels are provide in Figure 5.

Filename	Location	Lighting	Camera_Angle	Occlusion	Frame_loss	watermarks	Rep #	ankle_inversion_l	ankle_inversion_r	knee_angle_l	knee_angle_r	hip_rotation_l	hip_rotation_r	head_position	Domain_Advisor_Verified
20240102_161947.mp4	Living Room	day light	0	0	0	0	1	0	0	0	0	0	0	-1	1
20240102_161947.mp4	Living Room	day light	0	0	0	0	2	0	0	0	0	1	1	-1	1
20240102_161947.mp4	Living Room	day light	0	0	0	0	3	1	1	0	0	1	1	-1	1
20240102_161947.mp4	Living Room	day light	45	0	0	0	4	0	0	0	0	0	0	0	1
20240102_161947.mp4	Living Room	day light	45	0	0	0	5	0	0	0	0	0	0	0	1
20240102_161947.mp4	Living Room	day light	90	0	0	0	6	0	0	0	0	0	0	0	1
20240102_161947.mp4	Living Room	day light	90	0	0	0	7	0	0	0	0	0	0	0	1
0.mp4	outside in a parking lot	clear	0	0	0	0	1	0	0	1	1	0	0	0	1
0.mp4	outside in a parking lot	clear	0	0	0	0	2	0	0	1	1	0	0	0	1
2024-04-26 07-04-29 ID01.mp4	Gym	Bright Light	0,45,90	0	0	0	1	1	1	0	0	1	1	1	1

Figure 5: Sample of Data Labels

Figure 6 and Figure 7 capture scope of the manual labeling task which was tedious consuming many hours and reviewed by the domain advisors. The research team reviewed 24 videos identifying and labeling various aspects such as camera orientation, number of repetitions, correctness for each repetition and key point. Based on the data label reviews and further analysis of the performance of the Physio AI Core Model, the following decisions were made:

- Videos out of scope - videos containing the following: Significant occlusions; watermarks; multiple people; reflections of the single person in the frame (ex: from a mirror); significant motion not related to the exercise; children; significant frame loss
- Videos in scope - videos containing: Adults of multiple genders, sizes, and athletic levels (based on visual assumptions and not verified); locations (ex: inside, outside, home, gym); camera angles, variations in incorrectness labels for each of the measures; camera type (ex: iphone, android, laptop)
- Modified the design of the product: Some early videos recorded by team members captured recording setup that resulted in the user pressing the record button and walking back to perform steps. Sometimes the user was not fully within the video frame. Aspects caused the model to produce incorrect results. The product User Interface was modified to incorporate a 15 second timer after pressing the record button to allow the users sufficient time to get ready in position to start the exercise. In addition, the user interface is also enhanced to provide a primitive guideline via a red box to ensure the user's body stays within the frame.

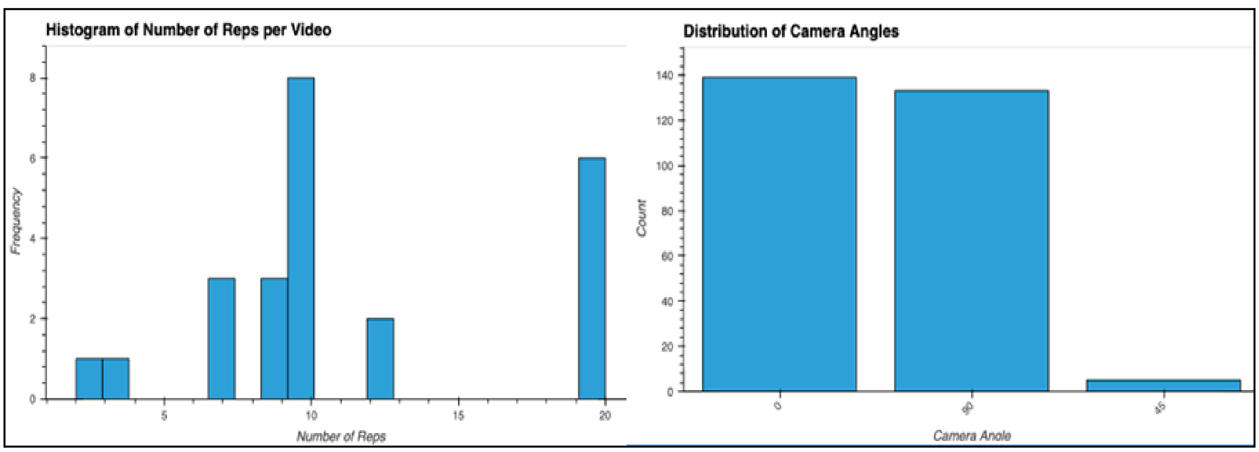

Figure 6: Manual labeled data for repetitions (left) and camera angles (right)

With the data properly curated for the focus of this research and appropriate “ground truth” labels generated, the next components of the Physio AI Companion were finalized. The following techniques were used to determine the best approaches for each of the remaining components.

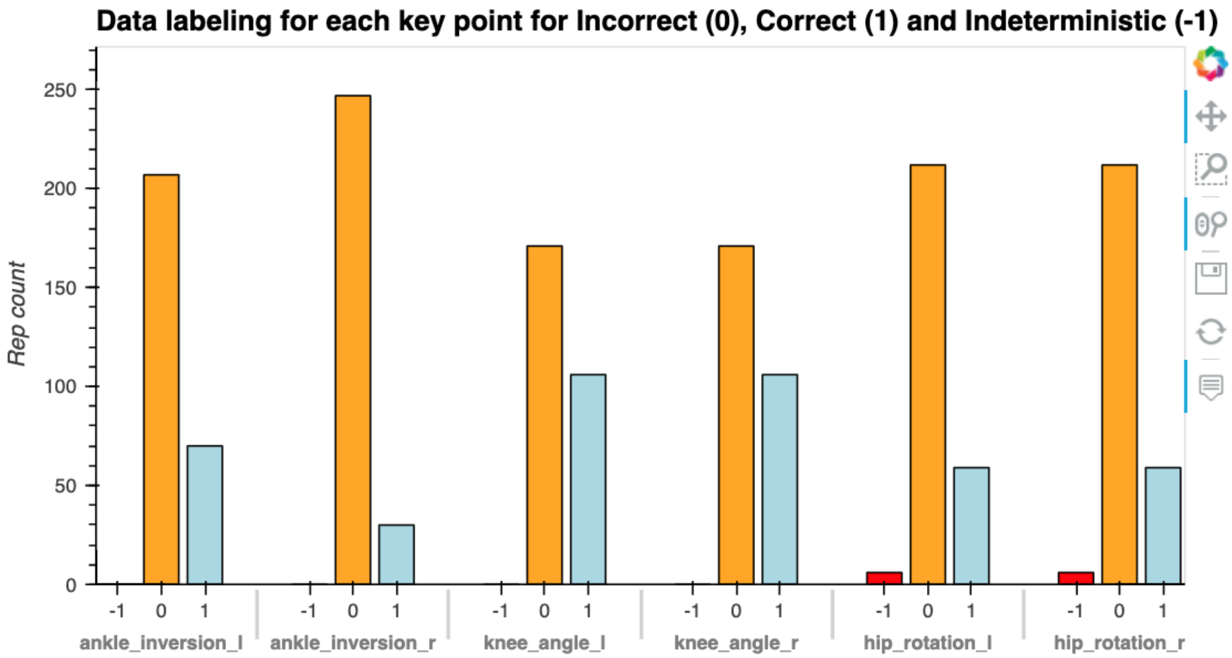

Figure 7: Manual labeling of correctness

Pose Estimation

Research was conducted to determine the best approach for pose estimation. Initially, 2D pose estimation was considered, but it proved insufficient due to issues with varying orientations, missed key points, and significant noise in the output data. 3D pose estimation was found to be crucial for several reasons:

1. **Orientation and Perspective Variability**: 3D models accurately capture a person's pose regardless of their orientation relative to the camera.
2. **Information about the third dimension**: 3D pose estimation predicts the information regarding the third dimension, allowing for precise measurements of joint positions and movements, essential for applications like physiotherapy.
3. **Reduction of Noise**: 3D models reduce noise by considering spatial relationships between body parts across frames.
4. **Holistic Movement Analysis**: 3D models enable comprehensive analysis of human movement in three dimensions.

The WHAM model was selected for 3D pose estimation after thorough market research(Figure 8) and empirical testing:

1. **Market Research**: Various 3D pose estimation models, including state-of-the-art approaches like VIBE, were evaluated.
2. **Empirical Testing**: The team conducted tests using video datasets, comparing WHAM's output to ground truth data and other models' results.

3. **Performance Evaluation**: WHAM demonstrated robust performance in capturing accurate and consistent 3D poses, achieving the lowest MPVPE (Mean Per Joint Position Error).

Initially, the variable ‘poses’ was used for pose estimation but found it lacking the information required for further calculations. Switching to the 'poses_world' variable provided the information required for the Physio AI Core Model. WHAM used 31 keypoints, combining data from Human3.6M and 3DPW datasets. The team selected the first 17 key points to display, focusing on the most relevant key points for accurate and efficient pose estimation.

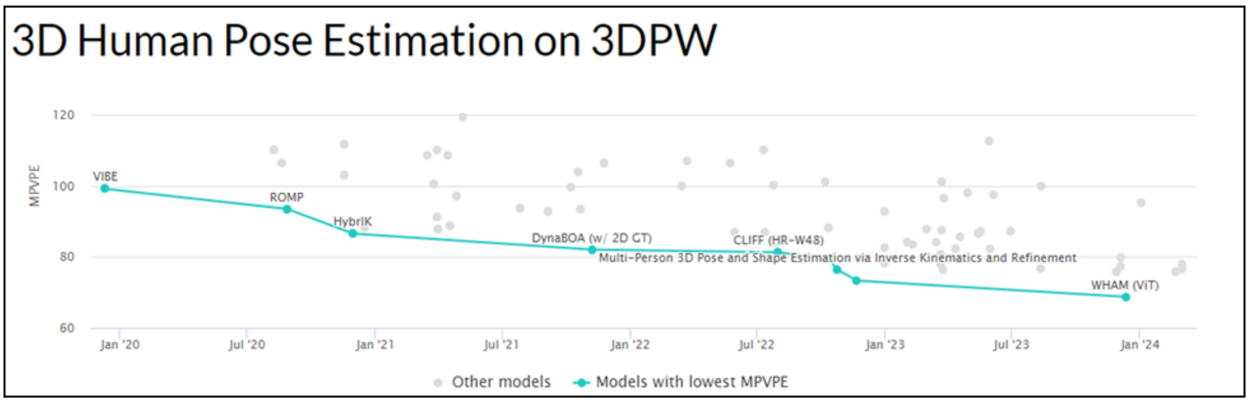

Figure 8: WHAM achieves the lowest MPVPE among the compared models

Repetition Identification

Persons performing exercises typically repeat a given exercise. The speed of repetition and number of repetitions often vary from a person to person and hence one video to another. For accurate score calculation, it was important to programmatically identify the number of repetitions as well as start and end frames for each repetition.

During the initial research into identifying repetitions of exercises, the repetition identification model employed a variety of computer vision approaches and customized algorithms to track detections. The computer vision models trained to track the movement of the head and count the repetitions as the head passed a certain point. A primary challenge with this model was lack of sufficient training data and also additional computational resources it would incur affecting the performance of the overall pipeline. Additional research was done on how to calculate the number of reps in an exercise video using only a computational approach from the data generated by the WHAM model.

For the final approach, joint angle measurements from key points such as right knee and left knee are analyzed for peaks and valleys. The computational algorithm was defined to identify two adjacent valleys that contain a single peak between them as start and end of a repetition and the corresponding frame numbers were extracted. A smoothing spline algorithm was applied to filter out noisy and rugged peaks. Several smoothing spline techniques were researched to select a compatible interpolation algorithm to work with generated data measures. This approach facilitated an accurate count of exercise repetitions each of which consisted of one valley-peak-valley measurement. We also applied a filtering technique using the mean values of these peak values to eliminate small peaks unqualified to be labeled as a repetition.

Table 3 highlights the comparison between the initial head position based approach and the joint angle based approach.

Approach	Number of Videos from Set3 Dataset	Set3 Reps Identified	Set3 Actual Reps	Accuracy
Computer Vision Based Approach (tracking the head)	34	243	405	60%
Computational algorithm Based Approach (tracking joints)	34	347	405	86%

Table 3: Repetition Identification Results

Based on the above table, the signal processing approach improved accuracy of repetition identification by 43%.

Incorrectness scoring

Specific joint angles were analyzed against predefined thresholds to assess the correctness of the movements. For the case of overhead squats, the lower body joints - hips, ankles and knees were identified as key points by the domain advisors who also provided the acceptable range of movement for each of the key points.

For knee joints, the function rounds the normalized joint angle measurements and identifies the maximum peak value within each repetition. Since knees have a single degree of freedom, which is knee bend, the algorithm evaluates whether the peak value of the knee movement falls within the correct range (75 to 90 degrees) and labels the corresponding repetitions as correct or incorrect accordingly. Similarly, for hip and ankle joints, the function rounds the respective measurements and determines correctness based on predefined angle ranges. These measurements were chosen after verification by the domain advisors.

This step involved MapReduce-like functionality. Frames were reduced to the number of repetitions making it easier to compute the incorrectness score. Score was then calculated for each rep and for each joint. This was aggregated and presented to the user as percent of repetitions incorrect.

Data Visualization

Data Visualization is an important aspect of the Physio AI Companion and many data science research activities. The data visualization needs to explain the results of the data science aspects of the Physio AI Companion to each type of user that would use this system. Keeping the use cases in mind and identifying the users generically as practitioners and clients, the following Data Visualizations were developed.

Client Portal

The Client Portal is used to address the first two use cases (increasing the effectiveness of health assessments and increasing the effectiveness of unsupervised exercise) described in the introduction of this paper. The Client Portal provides an interactive visualization of the results generated from the Physio AI Companion data pipeline as shown in Figure 9.

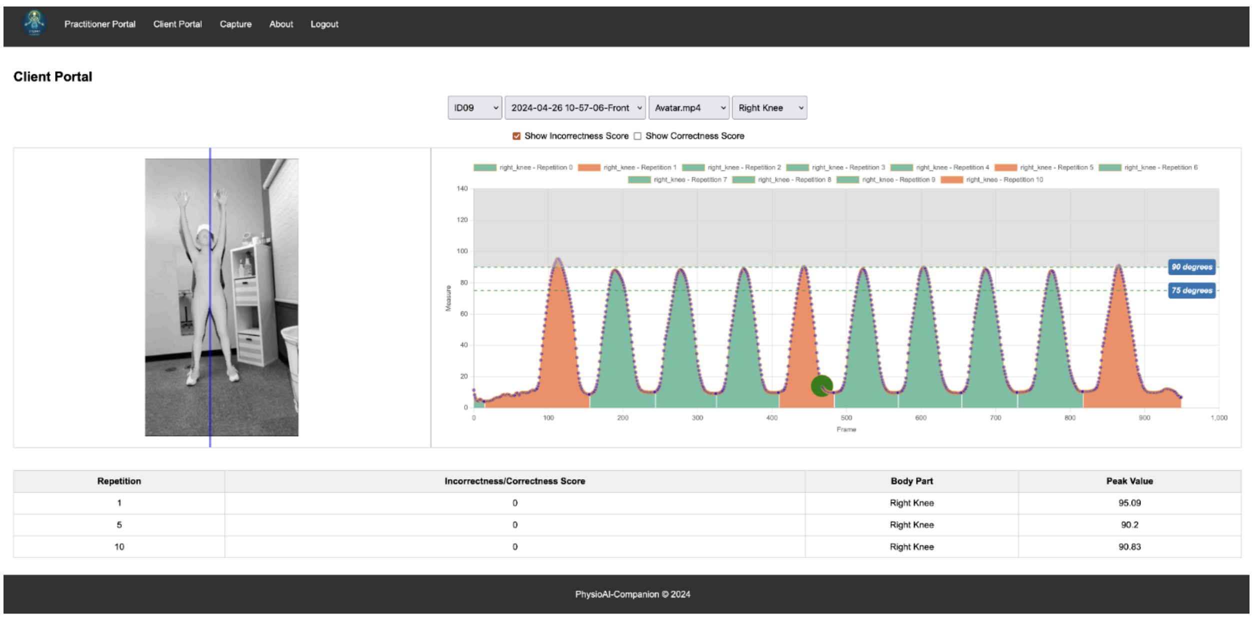

Figure 9: Client Portal UI

In Figure 9, the UI shows four sections of content. The first section on the top is a set of filters or selection criteria that alter the content of the following sections - 1: video playback, 2: graph plot, and 3: tabular view. The combination of these for sections allows the users to understand how well they are performing the overhead squat exercise.

The video playback component has three options of what type of video to play based on the user's selection of Avatar, Skeleton, and Standard video. The avatar option enables the user to replay their exercise with a 3D body model overlaid on their image. The skeleton option enables them to have a colored line drawing overlaid on the video, as shown in Figure 10.The lines represent connections between keypoints on the body. The mark of a point is used to display the key points tracked in this research (i.e. knees, hips, and ankles). The channel of color is used to indicate incorrectness in red or correctness in yellow. The standard option plays back the video without any overlays. Regardless of the video option selected there is a vertical line that can be slid across the video. The domain advisors wanted this specific feature as it helps with use case 1, assessments, by visually letting the practitioner see the asymmetrical aspects of the body across the vertical line.

Figure 10: Video Feedback to users indicating when and which joint the incorrectness was observed

The graph plot component provides detailed posture feedback on the correctness and incorrectness of biomechanical movement, highlighting specific body parts (e.g., right knee) and the degree of movement (e.g., 90 degrees, 75 degrees) and along with movements outside the optimal range (i.e., below 75 degrees or above 90 degrees). For this view the idiom of an area chart was chosen to show trends and changes with the biomechanical motion over time, making it easy to track the progression and also enabling users to monitor multiple variations in one place.The mark of points are used to show specific measured value for a key point on the body (angle, rotation, etc.) for each frame. The graphical plot with measures against frames provides a clear metric for tracking progress over time, helping users track patterns in their movements and improves the cognitive understanding of the flow of the biomechanical movement with clear key metrics boundary conditions displayed to help with the explainability of the model and why a measure is marked as incorrect or not.

The final component of this page, the table view, provides a straightforward way to view correctness and incorrectness scores by repetition, key part of the body by providing the peak value associated with the repetition and score giving a clear quantifiable result.

Overall, the client portal's visualizations demonstrate expressiveness, effectiveness, and interactivity creating a positive user experience, helping users perform exercises correctly, reducing the risk of injury, and shortening rehab time by improving use case 1 and 2.

Practitioner Portal

Practitioner Portal UI is a data visualization designed to address use case 3, Increasing the practitioner's visibility on exercise programs' effects on their client base. The UI is where practitioners can log in to view a summary of their facilities overall patient activities and biomechanical motion results. This functionality enables practitioners to review overall trends with their facility to help determine where clients are faltering on their exercises. This enables them to look how their overall exercise plans are affecting their clients overall to help determine how changes in those plans affect their clients overall as shown in Figure 11.

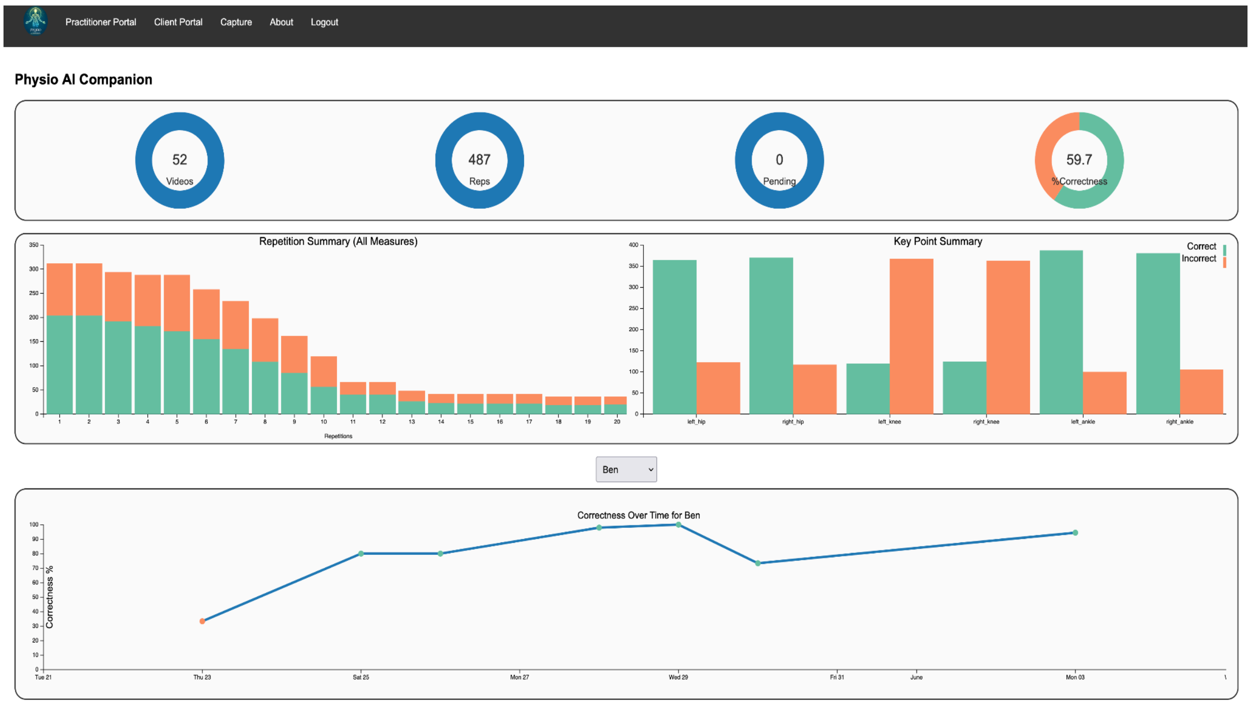

Figure 11: Practitioner Portal UI

Figure 11 illustrates the Practitioner portal UI, divided into three sections - 1: Top Section Facility Summary; 2:Middle section, Incorrectness by repetition and by key body points; 3: Bottom Section, Client History. The top section contains four components: First a count of the number of videos collected. Secondly, the count of total repetitions captured in those videos. The third component informs the practitioner if there are any videos that have not completed processing. The fourth component is the key reason for the idiom of a doughnut chart being used. The donut chart was used to show a comparison between the total number of incorrectness scores and incorrectness scores both visually with the green and orange colors in the color channel as well as a quantitative percentage value displayed in the center of the doughnut chart.

The middle section has two stacked bar charts to provide insights into the summary of incorrectness/correctness of repetitions on the left and key points on the right. The idiom of a stacked bar chart was chosen to enable quantitative values (quantity of videos) to be displayed for two categorical values (incorrectness and correctness). The channel color was chosen to represent the two categorical values - green correctness and orange incorrectness. The color choice of a more orange color vs a more definite red color to enable those with colorblindness to be able to still visually see the difference between the two categories. This information enables practitioners to understand overall which particular body part (key point) is being performed incorrectly the most. Additionally, the charts enable the practitioner to see at which repetition the clients start to show the most drop off in performance (i.e., incorrectness scores increase the most).

The bottom section displays the progress of individual clients over time, showing how effectively exercises are performed and monitoring both correctness and incorrectness over the period of time.The idiom of a line graph is used to choose as it is an ideal approach for showing a trend of a quantitative value overtime (i.e., correctness). The use of an interactive dropdown filter allows for the selection of specific clients.This visualization allows the practitioner to look further into each client’s history to understand how clients are progressing overtime.

The Practitioner Portal offers a user-friendly interface that provides a quick and comprehensive overview of patient activities in summary form. Its interactive graphs and visual features enable practitioners to swiftly visualize data and focus their attention on patients requiring the most care.

Data Visualization Conclusions

The research team developed a data visualization that provides user interfaces that address each of the three use cases identified by the domain advisors. Research was conducted to ensure the data visualizations were compelling, reduced the cognitive load on the users, and provided the content needed for the three use cases. The data visualizations were presented to the domain advisors to gather qualitative feedback to determine if the resulting data visualization did meet the three use cases as intended. The domain advisors unanimously thought the user experience was of value to their practice in addressing all three use cases as well as looking “awesome” and/or “cool”.

Findings and Reporting

The team tested the solution via daily exercise video uploads as well as batch processing of curated videos from Gym (Set4) totaling 50+ videos. Analysis of incorrectness score indicates a large percentage of repetitions for knee joints to be incorrect as shown in Figure 12 below.

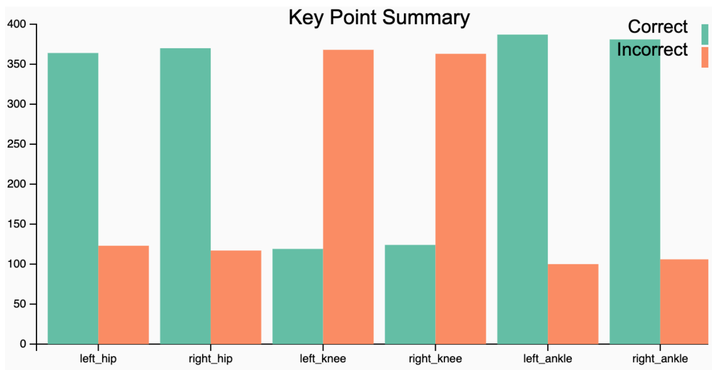

Figure 12: Score summary for key points

Further, the team worked with domain advisors to label Set3 videos independently. Comparing the output of the Physio AI Companion and the domain advisor's expert assessment labels was conducted. The data consisted of 24 videos. Figure 13 shows a confusion matrix for the domain advisor's labels vs solution output for each of the key points.

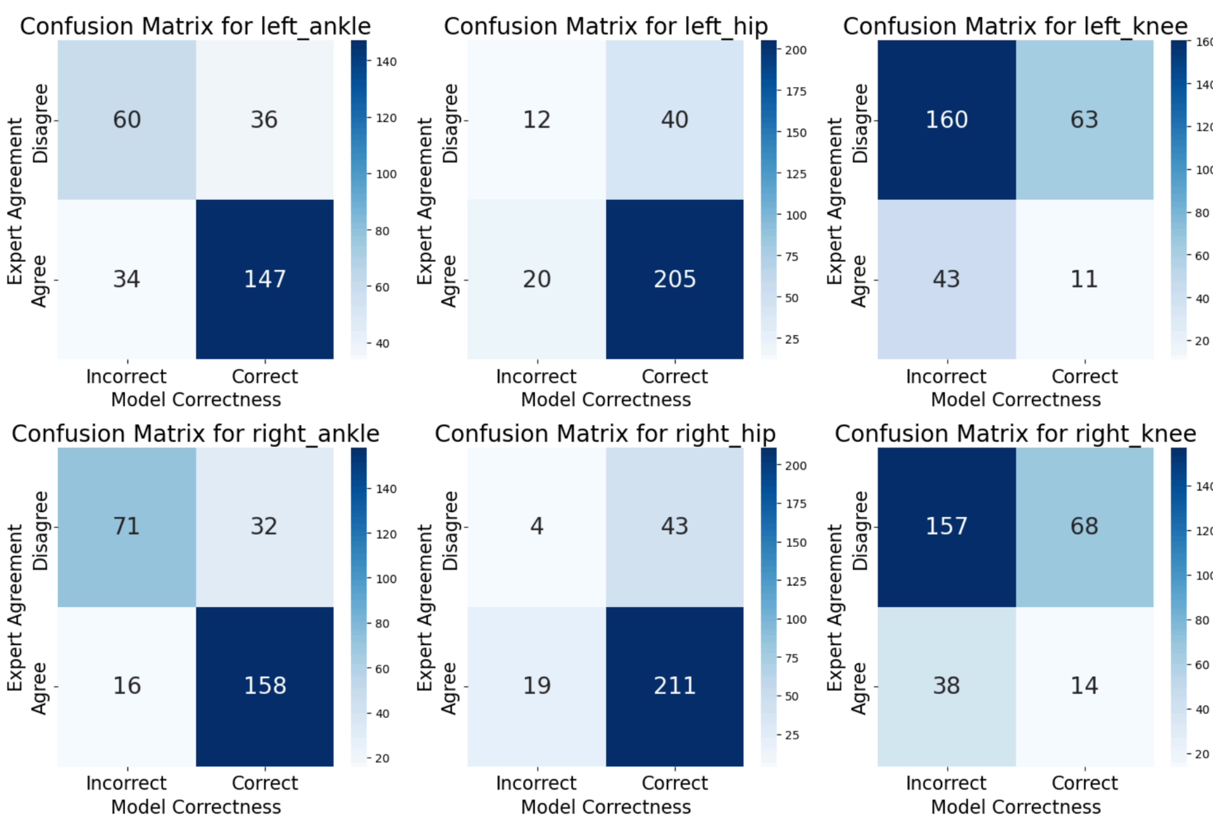

Figure 13: Model output vs Domain advisor's labels

There is a significant percentage of disagreement between the model output and the domain advisor for both knees. There are few possibilities for this:

1. The domain advisor labeling is based on visual estimation by observing the videos and the correctness may be subjective
2. One or more steps in the solution introduced significant error resulting in incorrect measurements and calculations.
3. The input capture device (camera) may need to be calibrated to remove variance and additionally the model may need to be calibrated against physical measurements using kinematic or other devices.
4. A visual observation was made by the testers that the model may produce incorrect results depending on the speed of exercises.
5. Model values for preprocessed videos were observed to be more than 22% different from the model values for original videos when a person was not facing the camera.

Solution Architecture, Performance and Evaluation

Figure 14 shows the solution architecture, realization of the data pipeline shown in Figure 4.

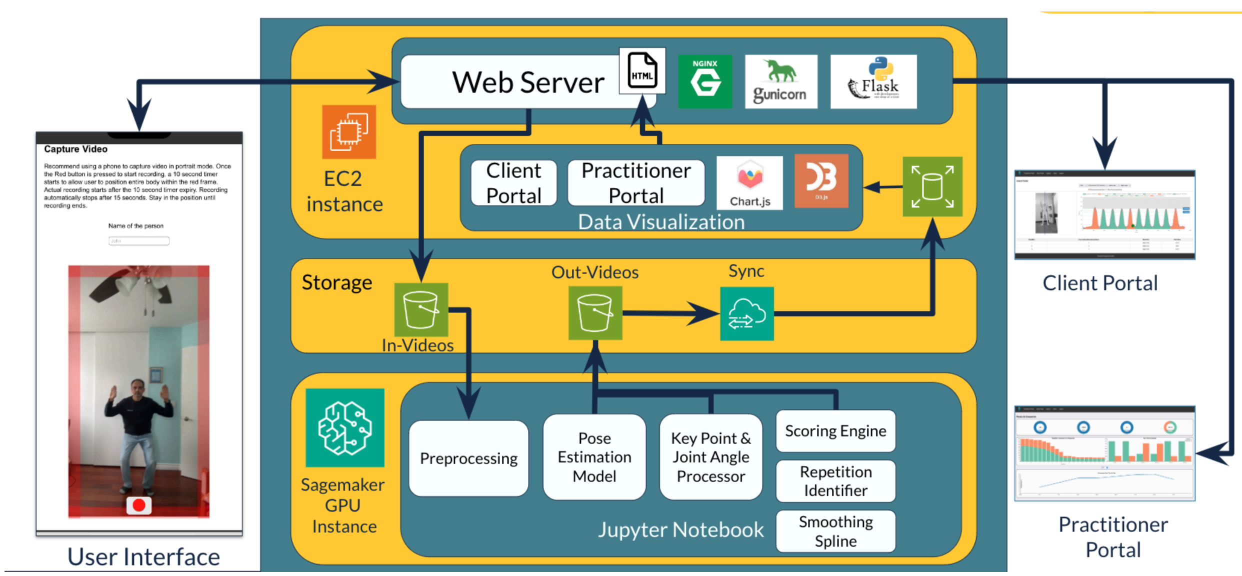

Figure 14: Solution architecture

The solution architecture uses a number of AWS platform components as follows:

1. User interface/Front End - Elastic Compute Cloud (EC2) Instance running a web server and JavaScript
2. Physio AI Companion core - SageMaker GPU Notebook instance
3. Storage - S3 and EBS

The entire process includes data collection, preprocessing, pose estimation, repetition identification, and a scoring engine to quantify incorrectness:

- **Data Acquisition Steps:** Data import from YouTube videos (publically available & team members upload) to a centralized Google Drive location for easier access, acquiring videos using our own capture webpage. Identify and exclude bad data from raw videos.
- **Model Integration and Pipeline Optimization:** Design an end to end pipeline that integrates preprocessing, pose estimation, repetition identification, and scoring into a cohesive system. This includes optimizing data flow between components and ensuring scalability to handle large video datasets efficiently.
- **Validation and Testing Framework:** Establish a validation framework using a combination of synthetic and real-world data to test the accuracy. This includes continuous integration/continuous deployment (CI/CD) practices for model updates and improvements.
- **User Interface:** Develop a user-friendly interface for both practitioners and clients, incorporating visualization tools for progress tracking and feedback. Integrate a feedback loop where user inputs can fine-tune model predictions and improve accuracy over time.

Scalability approach

Scalability of the solution is achieved via design for scale and selection of platform and open source components that provide autoscaling.

1. User Interface is via web pages hosted on a web server. The web server is implemented using Nginx which is widely regarded in industry as a small footprint and highly scalable solution.
2. Behind Nginx, Gunicorn Web Service Gateway interface is deployed to handle concurrent requests from multiple users. Gunicorn automatically scales the number of threads based on the number of incoming requests.
3. Offload - Chart drawing and data presentation on the portals is offloaded to client browsers via javascript thereby reducing the server side load allowing the server to handle a large number of requests.
4. Inline and batch processing - The Physio AI companion core component is implemented to handle two types of requests:
    a. Inline processing - when a user uploads a video via webportal, the notebook is triggered via AWS S3 upload notification and the notebook immediately starts processing the video. This notebook lifecycle scripts can be configured to invoke via a AWS Lambda function to autoscale.
    b. Batch processing - process multiple videos in a batch mode by crawling through videos in an AWS bucket and produce a set of results without user intervention.
5. Functional scaling - Initial phases of Physio AI Companion started with the one exercise, overhead squat, tested on four videos using measurements of the knees, hips, and ankles. As of today, the solution has been increased to 53 videos with different orientations of a person facing the camera and measurement of 17 key points and 24

joint angles. Additionally, the solution has been proven to process videos of different time lengths and different number of repetitions. The model now calculates the incorrectness score for more keypoints/joint angles, resulting in a growth from 126 scores to almost 3000 scores.

6. Other scalability considerations
    a. Added another dataset into the pipeline that consisted of different lighting conditions, environments, increased quantity of different body types, added an additional gender, and introduced a new obstacle to video processing (a mirror) which increased the algorithm robustness
    b. Increased robustness and scalability by reducing the computational resources and storage management required to generate the rep identifier output using trained computer vision model, by migrating to a computational approach based on WHAM joint angles
    c. Experimented with a dataset that had three different views (front, 45° angle, and side) synced to the same exercise to increase the accuracy of our scoring algorithm

Future scalability requirements are as follows:

1. Performance: Ability to perform analysis and scoring of videos in real-time
2. Datasets: Increase with additional variations of environment, lighting, people, etc.
3. Model scalability and robustness
   There are multiple ways the research could be expanded. The general approach used to date by the Physio AI Companion data pipeline can be used as a template to address broader requirements, such as:
    a. Generalize the solution to extend to support other types of exercises
    b. Generalize to support other domains such as posture correction in sports for better performance
    c. General assessment of mobility instead of just exercise incorrectness (e.g. imbalance weight between knees, range of motion of a joint is less than expected)

Scalability Evaluation plan

Scalability testing for this project primarily focused on following aspects:

1. Concurrent access of User Interface to perform actions such as video capture, client portal and practitioner portal
2. Measure and quantify the design efforts to reduce resource utilization primarily storage and compute resources
3. Batch processing capability to process large number of videos without manual intervention

Throughout the project, iterative testing was performed to identify performance issues and optimize the performance by fixing design and code issues.

Scalability evaluation report

A series of manual tests were run to prove the scalability either directly or indirectly:

1. Web server scaling: Tested by initially deploying the web server on the smallest EC2 instance available (T2 Micro) and being able to upgrade/switch the EC2 instance to T2 medium thus theoretically demonstrating the ability of the web server to scale with respect to request processing.
2. Improved individual and batch processing performance by implementing a preprocessing step in the design. The model running on a local computer took 6 hrs for some full resolution videos. With the preprocessing step, batch processing of 50 videos completed in approximately 2.5 hours. Batch mode was used extensively throughout the project once it was implemented since any changes to the core solution required a rerun of all the videos through the solution.
3. From the initial discovery stages to the final product; the number of videos, repetitions (reps) and score calculations have scaled from 4 videos, 21 reps to 53 videos and 496 reps. Initially, the score from a single key point was manually computed and now the score from 6 key points for each repetition is automatically computed with the score dataset increasing to 2976 rows.
4. During testing, the research team found that the AWS GPU notebook instance requires significant warmup time. For example, the first video typically takes a few hours to complete processing while each of the subsequent videos only take 2-3 minutes during batch processing. Further research of various technical forums showed that this is a well known issue among the developer community. Team implemented few workarounds to overcome this limitation by creating A warm-up script that performs matrix additions using GPU or a few minutes into the processing of the first video, manually stop and perform a kernel restart of the notebook

Budget Management

Being a team of 5, managing the budget was a bit challenging but we achieved it using strategic planning and resource management. Some of the steps in our planning were as follows:

1. Creating a budget plan to allocate different budgets for data storage, computing power, software tools, and any unforeseen expenses
2. Using a spreadsheet to track expenses in real-time to identify any areas where we might be overspending
3. Maximized the use of free resources like Google Colab and Google Drive for effective processing and storage until a stable data pipeline was created
4. Using batch processing and compression mechanisms for efficient storage and model processing utilizing low cost EC2 instances to process the results and model outputs.
5. Using efficient data management by compression, cleaning and preprocessing of raw video files and storing only the necessary data

Conclusions

The Physio AI Companion research effort successfully created a system for monitoring biomechanical motion during overhead squats, providing feedback to users and practitioners on motion correctness. This work focused on three business use case

1. Increasing the effectiveness of health assessments
2. Enhancing the effectiveness of unsupervised exercise
3. Increasing the practitioner's visibility on exercise programs' effects

The research team made technical decisions to balance accuracy, speed, cost, and usability to develop the Physio AI Companion solution. The solution allows practitioners to diagnose and address issues with greater accuracy, leading to more effective treatment plans. Additionally, the solution offers practitioners enhanced visibility into the effects of exercise programs across their client base. By leveraging detailed analytics and visualizations, practitioners can monitor progress, identify trends, and make data-driven adjustments to exercise regimens.

The domain advisors unanimously found the Physio AI Companion to be an useful and beneficial tool. In particular, the Physio AI Companion provided valuable insights that significantly enhance health assessments by enabling biomechanical motion analysis. Overall, the Physio AI Companion addresses the three targeted use cases effectively, delivering insights that support more accurate health assessments, safer and more effective unsupervised exercise, and improved monitoring of exercise program impacts.

References

1. Shin, Soyong, Juyong Kim, Eni Halilaj, and Michael J. Black. "WHAM: Reconstructing World-grounded Humans with Accurate 3D Motion." In *Computer Vision and Pattern Recognition (CVPR)*, 2024.
2. Jocher, Glenn, Ayush Chaurasia, and Jing Qiu. "Ultralytics YOLOv8." Version 8.0.0, 2023. [Online]. Available: https://github.com/ultralytics/ultralytics
3. Xu, Yufei, Jing Zhang, Qiming Zhang, and Dacheng Tao. "Vi{TP}ose: Simple Vision Transformer Baselines for Human Pose Estimation." In *Advances in Neural Information Processing Systems*, 2022.
4. Clark, R. A., Pua, Y.-H., Fortin, K., Ritchie, C., Webster, K. E., Denehy, L., & Bryant, A. L. (2012). Validity of the Microsoft Kinect for assessment of postural control. *Gait & Posture*, 36(3), 372-377. https://doi.org/10.1016/j.gaitpost.2012.03.033
5. Zheng, C., Wu, W., Chen, C., Yang, T., Zhu, S., Shen, J., Kehtarnavaz, N., & Shah, M. (2023). Deep Learning-Based Human Pose Estimation: A Survey. [Submitted on 24 Dec 2020 (v1), last revised 3 Jul 2023 (this version, v5)]
6. Stenum, J., Cherry-Allen, K. M., Pyles, C. O., Reetzke, R. D., Vignos, M. F., & Roemmich, R. T. (2021, November 3). Applications of Pose Estimation in Human Health and Performance across the Lifespan. *Sensors (Basel)*, 21(21), 7315. https://doi.org/10.3390/s21217315
7. Webster, D., & Celik, O. (2014, July 3). Systematic review of Kinect applications in elderly care and stroke rehabilitation. *Journal of NeuroEngineering and Rehabilitation*, 11(1), 108. https://doi.org/10.1186/1743-0003-11-108

Appendices

A. DSE MAS Knowledge Applied to the Project

This project included application of knowledge acquired via following courses over the last two years:

1. Python for Data Analysis
2. Data Integration and ETL
3. Machine Learning and Statistics
4. Scalable Data Analysis
5. Data Visualization

B. Link to the Library Archive for Reproducibility

https://doi.org/10.6075/J0HM58PG