FormFit: A Real-Time Application for Exercise Posture Analysis and Correction

Parth Manish Thapliyal
Department of Computer Science
Stony Brook University
SBUID: 116437211

Abstract

Engagement in fitness is essential for preserving comprehensive health and wellness. Nevertheless, discerning the appropriate starting point and executing exercises proficiently can prove to be difficult, particularly due to the vast and frequently contradictory information accessible on the internet. A significant number of individuals find themselves perplexed and doubtful, resulting in incorrect form leading to injuries and less than optimal outcomes. Hiring a personal trainer to resolve this can get too expensive. This study introduces a fitness assistant application, FormFit, aimed at mitigating these challenges by offering users immediate feedback regarding their exercise technique via camera-based pose detection. The application employs Mediapipe for the identification of joints and tracking of movements, OpenCV for the integration of camera functionalities, and PyQt5 to deliver a smooth and user-friendly Graphical User Interface. A heuristic methodology was employed in the design of the user interface to guarantee its simplicity and user-friendliness, accommodating individuals with varying levels of experience. Assessments of live video performance have been carried out to evaluate system efficacy, resulting in an average accuracy rate of 98.2% across two specific exercises: bicep curls and lateral raises. The application functions as a comprehensive resource for novices to acquire fundamental exercise knowledge while simultaneously providing intermediate users the opportunity to enhance their skills. The real-time feedback system facilitates ongoing improvement in users' form, thereby promoting the attainment of optimal fitness outcomes.

Index Terms

Real Time, Fitness Assistant, Mediapie, OpenCV, PyQT5(GUI), Human Computer Interaction,

I. INTRODUCTION

Exercises are the major tools for health and fitness, but without proper execution, apart from causing serious injury in extreme cases, this would result in poor performance. Proper form has remained a big concern in the spheres of fitness for both the amateur person or professional. Real-time feedback becomes crucial for effectively working and avoiding injury with the flood of information on fitness through the internet, where it is often difficult to identify which movements are correct. Since hiring personal trainers for guidance can be very expensive, an application of State of the art Machine Learning and AI techniques can offer cost effective and easy solutions to our predicament.

The work presented here, FormFit, is a fitness assistant application developed to bring in the latest advances of pose detection into an easy-to-use, user-friendly interface. A system with a conventional camera tracks the body movements against a set of standards in exercises and scores the same. The **Mediapipe** application used allows tracking of joint positions and identification of poses that deviate from correct form. The library **OpenCV** will be used for the real-time processing of video input, while the integrated interface with **PyQt5** will immediately show visual feedback to help users return to proper form and technique..

This design has further been made more accessible and convenient, since it requires only a device with a camera and no additional hardware. Implementation of functionality testing through scientific fitness guidelines of exercises like bicep curls and lateral raises to test the detection of deviation with immediate corrective feedback really constituted a base for proving that the system is sensitive and highly articulate in recognizing infringements of form and offering practical insights.

It is lightweight to run on low-cost hardware without the need for a GPU. Application porting is highly suitable for mobile smartphones, hence assured ease of access and ease of operation for the end-user without high-end computing devices. With Mediapipe tracking the joints, immediate response and accuracy are warranted without the usage of too much computational power. This system becomes much more viable for general application in the field of fitness and exercise correction.

This paper outlines the system's architecture, the development methodology, and the evaluation results. It also discusses the user experience that is possible with the Graphical User Interface and how the inclusions of real-time pose detection and feedback mechanisms in the system improve safety and efficiency during workouts. The proposed application plays the role of connecting technological integration with fitness coaching and, as a result, making personalized coaching more accessible to larger groups of people.

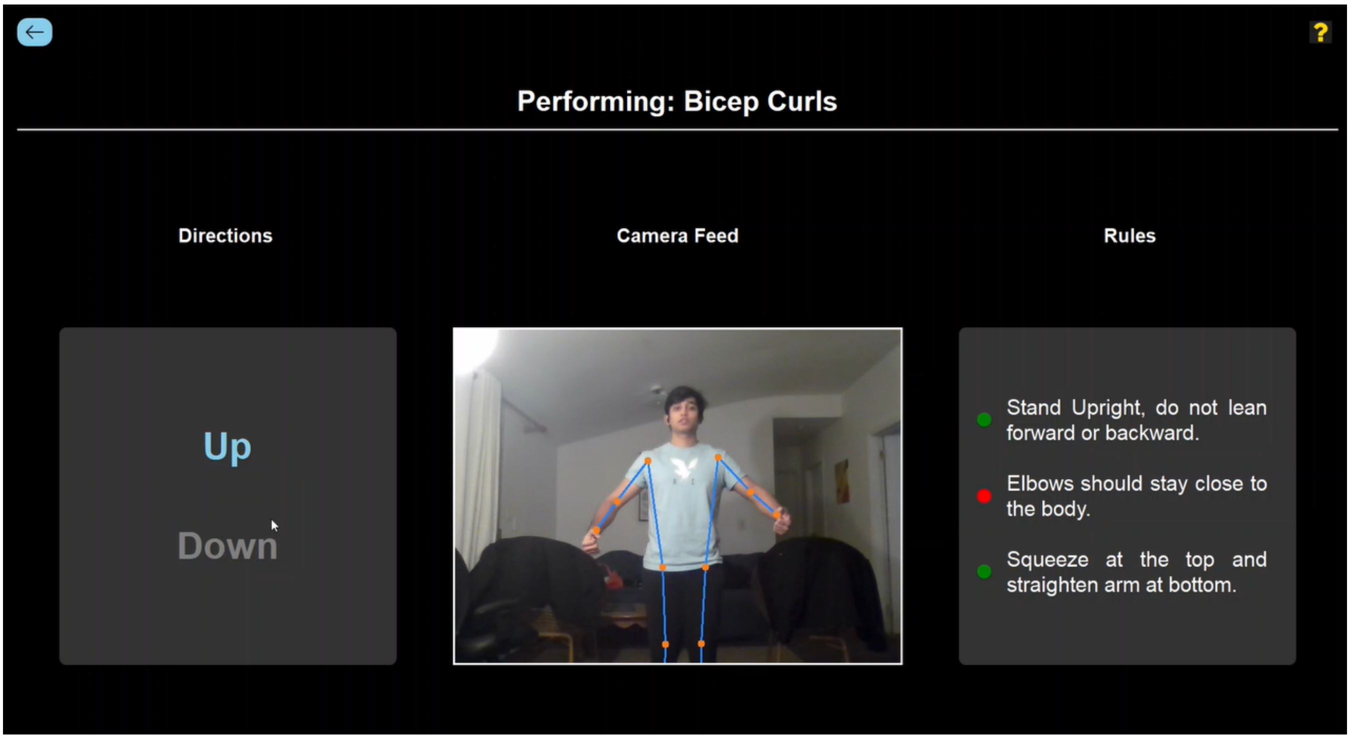


Fig. 1: Screenshot of Application User Interface with real-time feedback for Bicep Curls

II. LITERATURE SURVEY

The relevant scientific discussion of human pose estimation and correction of exercise techniques underlines several developments in the application of computer vision combined with machine learning techniques for monitoring and improving physical activities.

Study [1] focuses on a neural network approach for classifying static exercises, such as planks and isometric squats, and providing real-time feedback to correct posture using a mobile application. While it has achieved very accurate classification for postures such as correct, hips too low, or hips too high, the fact that the exercises are static means that it cannot be applied on a more general basis.

[2] proposed an SVM-based approach for classifying and correcting dynamic fitness exercises like squats and pull-ups in real time with the best accuracy of 96.87%. While the system is effective in preventing faulty forms across most of the four activities included, it is not versatile enough in absorbing other forms of exercises within its scope.

[3] applies physiotherapy by way of novelty dimensionality reduction using the Logistic Regression Recursive Feature Elimination technique in conjunction with random forests to estimate body pose, yielding very good accuracy but at great computational cost due to reliance on high-dimensional datasets.

Most of the literature nowadays focuses on either general exercises of fitness or physiotherapy protocols without mentioning the specific exercises, such as bicep curls and lateral raises, that need to be performed in order to address certain muscle groups. Furthermore, whereas several research efforts rely on training neural networks with large image datasets for pose estimation, I use Mediapipe, one of the most sophisticated joint localization tools which economizes the computational burden with no loss of accuracy. Human pose estimation has advanced significantly with computer vision and machine learning, enabling real-time tracking and analysis of movements for physiotherapy and fitness training.

The study [4] focuses on addressing high feature dimensionality in pose datasets, employing methods like Logistic Regression Recursive Feature Elimination (LogRF) to improve machine learning performance. These advancements support accurate posture correction, injury prevention, and personalized exercise regimens through efficient feature selection and model optimization.

Building upon the work done in prior literature, I propose an integrated system of Mediapipe to track joints in real time and rule-based evaluation customized for particular exercises like bicep curls and lateral raises. This paper will show how movement direction rules and visual feedback, through a user-friendly GUI, ensure robust posture monitoring in my system. My solution tends to scale and efficiently solves the problem of exercise correction by overcoming the limitation of prior methods, which focused on less explored exercises, using state-of-the-art(SOTA) pose estimation technology to improve user experience and performance.

III. System Architecture and Implementation

The system architecture for our exercise assistant system is composed of three core modules: the Pose Detection Module, the Form Analysis Module, and the Feedback Module. It is designed with a specified pipeline wherein each module executes a particular subsection of the application, integrated together for the successful smooth monitoring and analysis of an exercises. In the following diagram is an overview of a flowchart for the sequence of events in this system:.

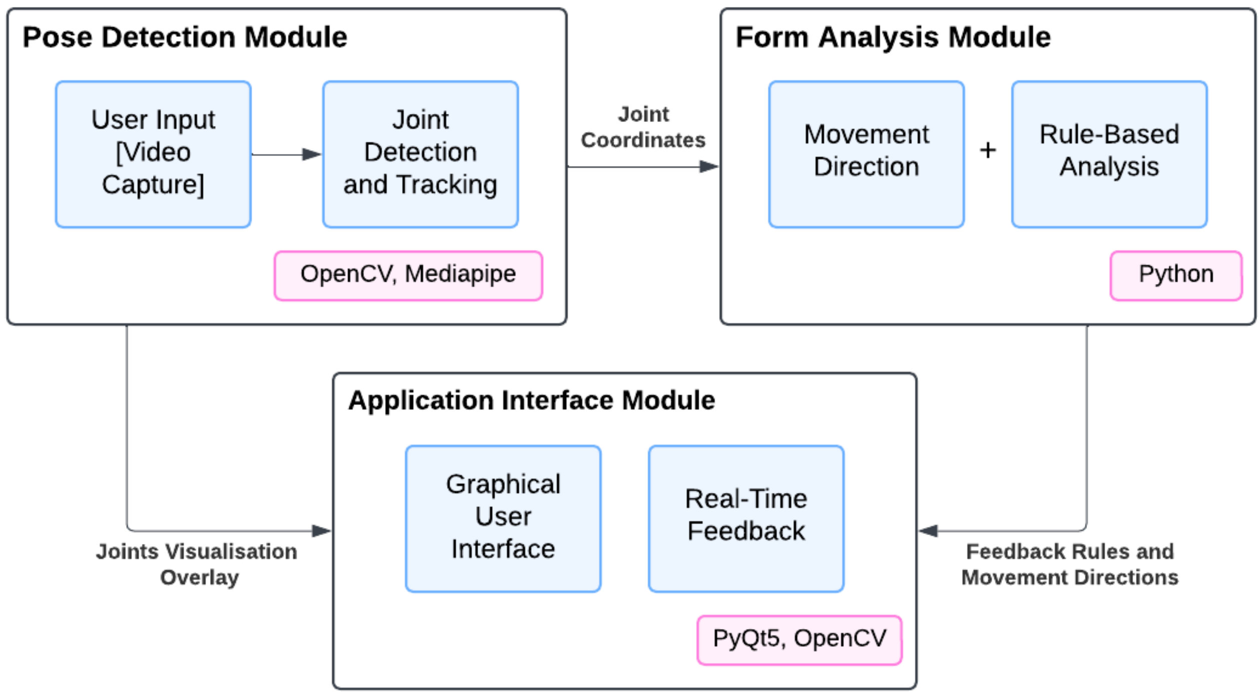


Fig. 2: Modular Representation of System Architecture and Technologies used

In this case, preprocessing starts by capturing input video from the client's webcam, which it uses to detect relevant joints, form, and posture, providing insightful feedback to the user through the Camera Overlay module and visual feedback 1. The system ensures that data flows efficiently from the video capture stage, through real-time pose detection and form analysis, and is ultimately presented to the user in a meaningful way through visual overlays and alerts. Each module interacts with the others to maintain accurate exercise tracking, prevent injuries, and optimize workout form for better efficiency and muscle engagement.

A. Pose Detection Module

The Pose Detection module is responsible for tracking and identifying the user's key body joints in real-time. This module utilizes the combined power of OpenCV and Mediapipe, two robust computer vision libraries that are well suited for video processing and pose estimation.

OpenCV connects the user's camera and the computation system. It captures the video frames coming from the user's inputted camera and sends frames continuously into the system. OpenCV is one of the multi-functional libraries for real-time computer vision. It provides a wide variety of functions regarding the processing of images and videos, directly connecting to the user's camera device for a smooth flow of uninterrupted video, as well as providing a camera overlay of visuals captured as well as other texts or alerts. On capturing the video frames, OpenCV sends those images to MediaPipe for pose estimation.

MediaPipe is a SOTA cross-platform framework developed by Google using some of the most advanced machine learning models with the purpose of performing real-time pose estimation. Detecting key landmarks of the user's body, such as elbows, shoulders, wrists, and knees, it estimates the spatial coordinates in 2D or 3D data points, that is, (((x, y, z))). Instead of displaying all detected landmarks as shown in 3, the system focuses only on the joints that are crucial for tracking specific exercises, such as lateral raises and bicep curls. For instance, it only tracks the shoulder and elbow positions for bicep curls. These joint locations and their connection lines are then passed on to the Camera Overlay

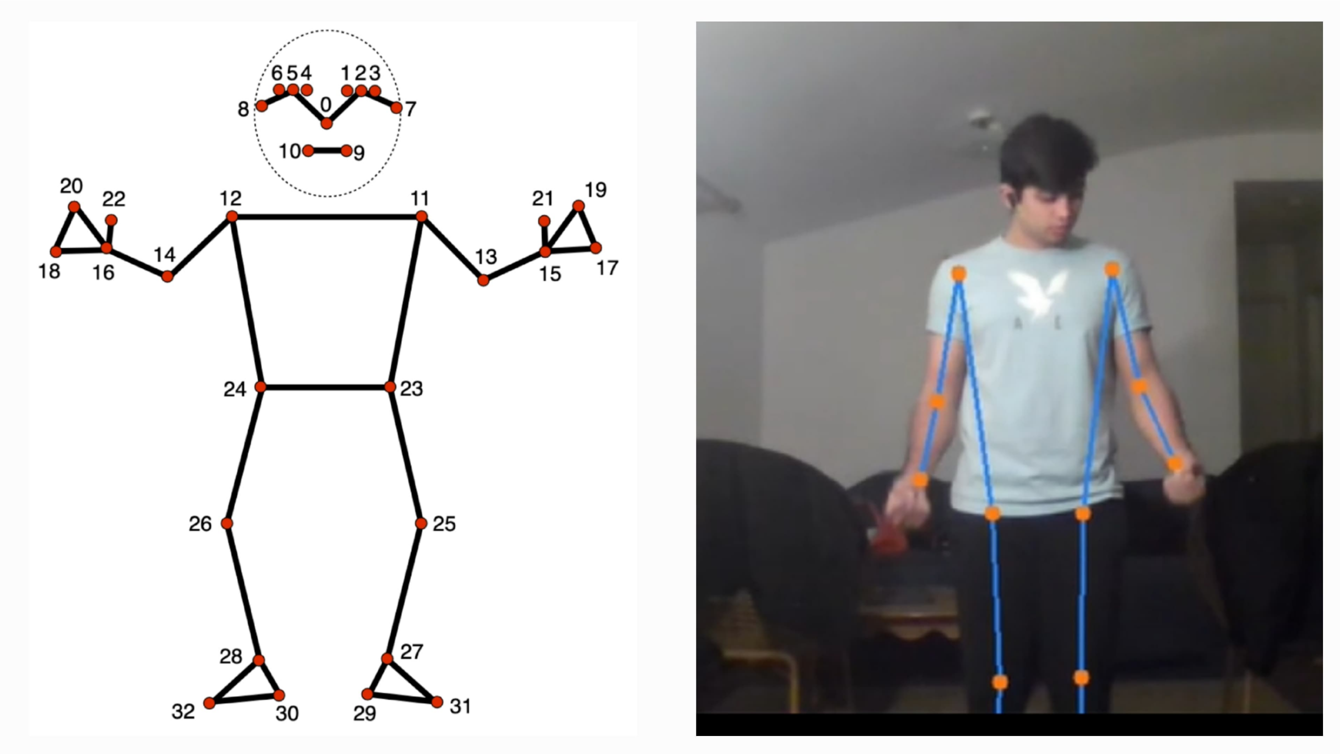


Fig. 3: All Mediapipe Landmarks (left) and Filtered Landmarks(ressential) for selected exercises (Right)

module for visualization and analysis. This ensures that the feedback provided to the user remains relevant and tailored to their specific exercise of choice.

B. Form Analysis Module

The Form Analysis module builds on the Pose Detection module's data to analyze the user's form and posture during an exercise session. It stresses two major components: **Posture Detection** and **Joint Angles**. Both components are vital in giving assurance over the correctness of the user's form, which is vital for diminishing the risk of injury and enhancing effectiveness during exercises.

1) Posture Detection: The posture detection relies on the analysis of the user's position on the z-axis, which is indicative of the distance between the user and the camera. It is an important measure as it gives useful information about the general orientation of the user while standing-that is, if the user keeps upright or leans forward or backward. By comparing the distance of lower torso. or hip joints, with respect to the shoulder joints distance, we can analyze to what extent the user is leaning forwards or backwards.

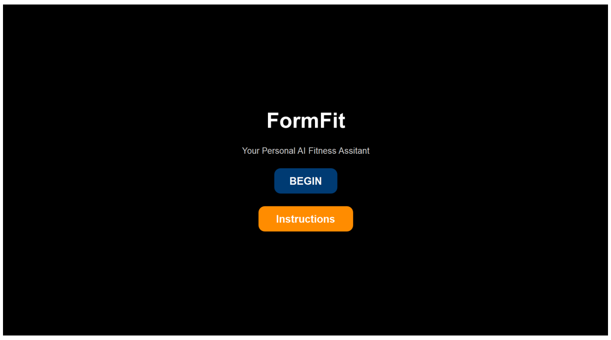
The threshold mechanism in this regard is dynamic for gaining accurate posture recognition. That means that the threshold will change by how close the user is from the camera. For example, when a user is standing 100 meters away from the camera, a fluctuation of 10 meters is acquired-a 10% difference. When the user is only 10 meters away, the change is 1 meter-makes it the same percent change. Therefore, an adaptive threshold system ensures that posture detection remains accurate across different distances, allowing real-time alignment adjustments with high precision.

2) Joint Angles: Joint angles help in evaluating the best exercise techniques, particularly for the execution of bicep curls and lateral raises. This system does an evaluation of the angulation at crucial joints like elbows and shoulders to establish whether the form of the exercise is proper or not.

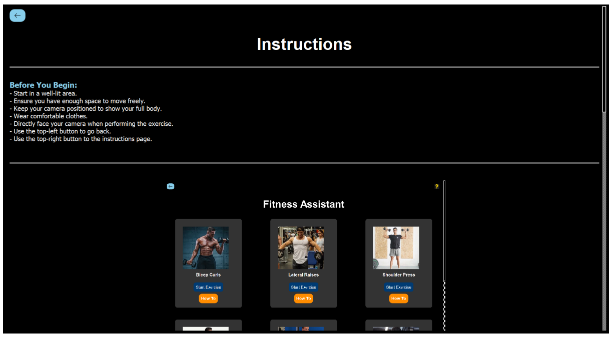
Joint angles can be used to set thresholds of rule breakage, for example if the angle between the torso and arm, signified by angle from hip-shoulder-elbow joints becomes more than 45° degrees at the shoulder, it implies distance of arm from body has increased more than it should, which shows an error. This is visible in Figure 1.

It also includes angle-based guidelines concerning the orientation a user is supposed to move in-for example, upward or downward weight lifts. Consider bicep curls, when the angle falls outside a certain threshold, i.e. 50° or exceeding 140°, a change in movement direction is required. Below 50° implies you have reached the top-most point of movement and more than 140° implies bottom most.

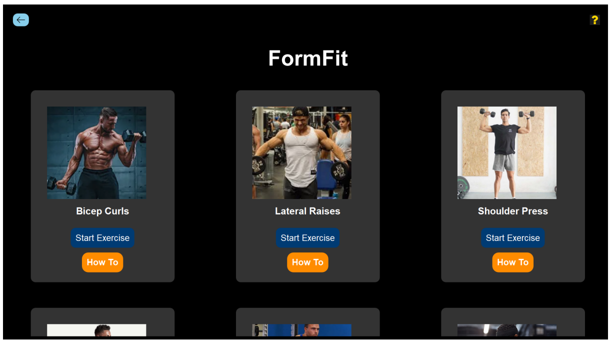
These guidelines are linked to specific thresholds that optimize muscle activation with safety considerations. Whenever the user's form exceeds the set threshold, the system will indicate to the user a change in the exercise direction and highlight rule breakage so as to enable proper stimulation of muscles without injury.



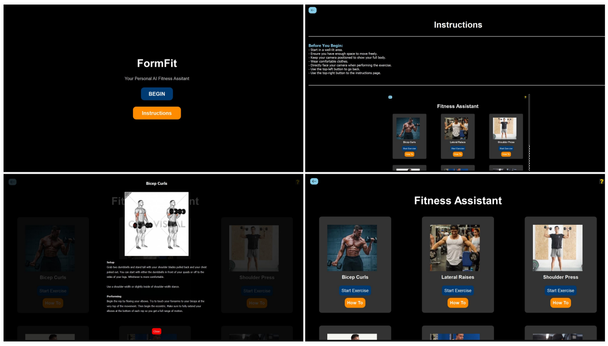
(a) Application Landing Page



(b) Application Instructions Page



(c) Application Home Page - Exercise Selection



(d) Application Pages

Fig. 4: Application interface screenshots

C. Application Interface Module

The Application Interface Module includes those elements that are shown to the user inside the application, working independently and in unison with the main integrations with Mediapipe and OpenCV. It is designed to ensure maximum user interaction throughout the flow of the workout, guiding the users to perform exercises with the right form and movement. It integrates visual elements, instructions, and real-time feedback to create an intuitive and educational workout experience.

1) User Interface (GUI): The GUI includes various screens and interface elements that guide the user through different stages of their workout journey. It is designed to be user-friendly and provides necessary information to help users perform exercises with proper form and technique.

The **User Interface** consists of four primary screens:

- **Landing Page**
  The landing page serves as the entry point of the application as shown in Figure 4a. It offers two options: **Begin** to start the exercise session and **Instructions** to learn how to use the application. This ensures that users are well-informed before they start interacting with the main functionalities of the app.
- **Home Page**
  Users can choose on the home page either to **start an exercise session** or to choose **"How To"** that provides a small guide for the form and technique for certain exercises. This is going to help users get what is intended from the workout guidelines to then be checked during the form analysis.
- **Instructions Page**
  This page gives detailed instructions on the application, describes the features, and provides guidelines on the workouts. It is supposed to ensure that the users know what to focus on while exercising and navigate the application accordingly.
- **Exercise Page**
  The exercise page is a central part of the interface and combines instructional and feedback elements. The page is divided into two sections:
  - **Right Panel**: Contains the form rules, such as **posture guidelines, instructions**, and other essential information to maintain correct exercise form.
  - **Left Panel**: Provides visual guidance for the **direction of movement**, showing whether the user should move the weight **up or down**.

The Exercise Page integrates the live camera feed with an overlay of joint positions. This allows users to see their form in real-time while receiving corrective visual feedback.

2) Feedback: The **Feedback module** analyzes user performance during workout sessions, ensuring that exercise execution adheres to the predefined form and movement guidelines. This module combines data from the **Form Analysis Module**, Mediapipe, and OpenCV to provide actionable feedback through real-time visual indicators.
During the exercise session, the Feedback module evaluates two crucial aspects of form correctness:

- **Posture Evaluation**
  The application uses the **z-axis value** to assess posture, which compares the distance between the camera and the user. This evaluation is important because it ensures that the user maintains a proper orientation relative to the camera. The z-coordinate depends on the user's proximity to the camera, and a **variable threshold** is implemented to account for these changes. As the distance increases, the threshold decreases, ensuring a consistent evaluation across varying distances.
- **Joint Angles and Movement Analysis**
  The module also evaluates the angles of critical joints, such as the **elbow and shoulder**, to determine proper form. The angles have **fixed acceptable ranges** that do not change with the user's distance from the camera but depend solely on movement correctness. Additionally, the module implements **direction rules**, guiding users on whether to move the weight up or down. Visual indicators with **blue text** (showing UP or DOWN), **green lights** (correct form adherence), and **red lights** (incorrect form) provide immediate corrective feedback.

By integrating these aspects, the Feedback module helps users prevent injuries, maintain proper form, and optimize workout movements. It continuously analyzes and displays joint overlay data on the central camera feed while presenting form and movement correctness indicators on the side panels.
Figure 1 illustrates an example of bicep curls in progress, where the movement direction is indicated as **UP** since the arms are in a lowered position. Additionally, a red signal is displayed for Rule 2, indicating that it is being violated, while the other rules show green signals, signifying compliance.

IV. Evaluation

A. Detection and Form Analysis Assessment

The evaluation of the system's performance for joint detection and exercise correctness was based on preset rules implemented in Python. Since these rules were defined explicitly, the only reason for errors in the model were caused due to Mediapipe joint identification, which is very rare given the robustness of the model, which will be highlighted in this section.

Experimental Setup

- For **Bicep Curls**, 100 repetitions were performed and 2 Rules were evaluated with 50 repetitions for each.
- For **Lateral Raises**, 150 repetitions were performed and 3 Rules were evaluated with 50 repetitions performed for each.

During evaluation, the exercises were performed incorrectly exactly 50% of the times to make a balanced observation between correct and incorrect repetitions. This gave the ability to calculate very accurate and effective classification metrics in the form of Accuracy, Precision, Recall and F1 Score.

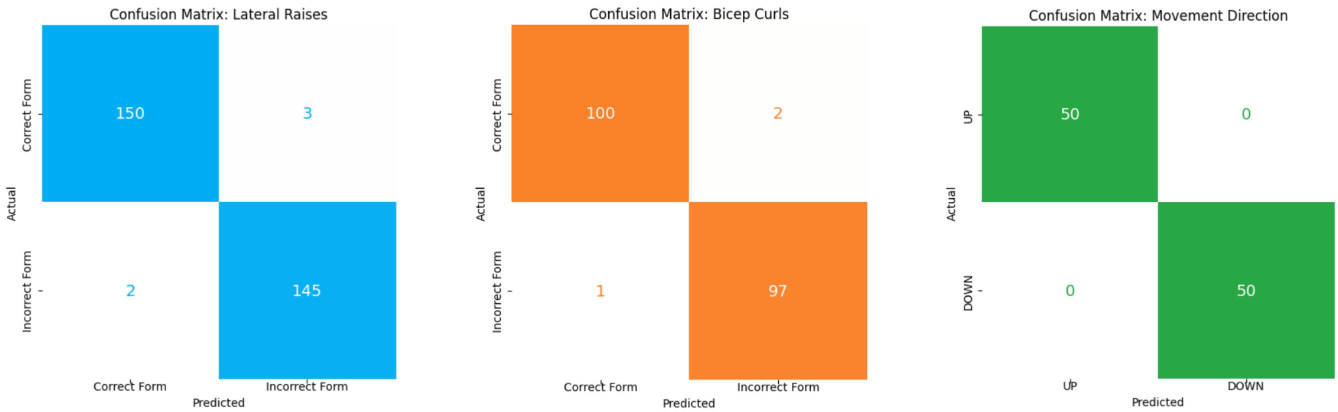
TABLE I: Performance Metrics for Bicep Curls and Lateral Raises

Exercise	Rule	Precision	Recall	F1-Score	Accuracy
Bicep Curls	Rule 1	0.98	0.98	0.98	0.98
	Rule 2	0.98	1.0	0.99	0.99
Lateral Raises	Rule 1	0.98	0.98	0.98	0.98
	Rule 2	1.0	1.0	1.0	1.0
	Rule 3	0.96	0.98	0.97	0.97

Analyzed Rules

- **Rule 1 (Posture-Related Evaluation)**
  For both exercises (Bicep Curls and Lateral Raises), **Rule 1** focused on maintaining good posture by standing upright. Since this evaluation criteria is based on a similar upright positioning of the torso, it resulted in consistent performance across both exercises, giving a high Precision, Recall, and F1-Score values.

- **Rule 2**
  For **Lateral Raises** , rule 2 has a set specific height till which the arm should be lifted. This well-defined rule resulted in **100% accuracy**, as the system could accurately determine this threshold with ease. Similarly, for **Bicep Curls**, **Rule 2** relied on a straightforward calculation of the elbow's distance from the user's body using the shoulder angle, leading to high accuracy as well.
- **Rule 3**
  This rule for Lateral Raises was more challenging to define a highly accurate logic for. It focused on determining the elbow angle accurately, but the system faced difficulties in finding a reliable mathematical formula for this evaluation. Although the results were still highly accurate, this difficulty introduced some errors compared to the previous rules. Rule 3 comprised of another rule requiring the user to keep their arms wide, this was defined by setting a minimum threshold o the distance between the elbows, this subpart of rule 3 also worked without any errors.



(a) Rules for Lateral Raises (b) Rules for Bicep Curls (c) Rules for Movement Direction

Fig. 5: Confusion Matrices for Performance Evaluation of Rules in Lateral Raises (left), Bicep Curls (middle), and Movement Direction (right)

The confusion matrices in Figure 5 highlight the robustness of the application. The lack of False Positives and Negatives in both matrices for rules. A 100% accuracy for movement direction has been shown in the third matrix.

B. User Experience Assessment

The user experience evaluation focuses on ensuring that the application is intuitive, accessible, and user-friendly. A seamless experience is achieved by aiming to follow all HCI Heuristics via the following principles:

Accessible Help and Instructions

- A dedicated help button, represented by a question mark icon at the top right, is present on every page. This provides the user with accurate **documentation** to troubleshoot whenever needed. This is visible in Figure 4.
- Clear explanation of all requirements for the application to work properly from the users has been mentioned on the Instructions page, and the option for instructions is on the landing page for first-time users. This page also has directions and explanation of every page accessed by the user, providing clear directions for use of application.

Effortless Interaction with an Intuitive Interface

- Buttons and navigation elements are clearly labeled and accurately tagged.
- Each button's functionality is **self-explanatory**, minimizing ambiguity and making interactions straightforward.
- The interface provides step-by-step instructions with direct actions.
- Features are designed to require minimal effort from the user.
- Interactions are smooth, with quick response times and intuitive feedback mechanisms.

Aesthetic and Minimalist Design

- The interface follows a **minimalist design** approach.
- Only essential elements and instructions are displayed, avoiding unnecessary distractions.
- A **consistent** color scheme and layout across all sections of the application ensure a visually appealing and professional experience.
- This fosters better engagement and interaction with the system.

Real-Time Feedback

- Visual feedback mechanisms are shown in real time and are evidently clear and logical to understand.
- Green stands for correct and red stands for incorrect, implying correct relation to real-world understanding and interpretation.
- Joints visualization on the camera feed and change in colors of the visual feedback from gray (inactive) to blue for movement direction shows **System Status** implicitly, and any errors with the camera can be directly visualized by on the screen itself.

While the evaluation here has been performed for one subject only, it must be pointed out that this must be extended to other subjects in numbers and in a more controlled manner if the robustness and efficiency of the system are to be tested rigorously. The results of an expanded test pool would be indeed enlightening about the potential system variability for different body types, exercising styles, and motion profiles. Thus, the required broad testing of the system would validate its performance and point to potential limitations that may arise, furthering refinements and optimizations for functionality towards wide usage.

V. DISCUSSIONS AND FUTURE WORK

The performance of the similar rules stayed consistent across exercises as they are defined using the same -rule based methodology. For both bicep curls and lateral raises, the application maintained high accuracy and F1 Scores, highlghting the robustness of the approach used. Mediapipe joint tracking is highly accurate with only rare setbacks due to overlapping joints and z-axis values. The limitations of Mediapipe led to slight accuracy discrepancies in Rules 1 and 3 mainly. Rules applied in the 2D plane and with non-overlapping joints worked with virtually zero errors, as seen in Rule 2 and Movement Direction. Rule 1 utilizes z-axis values which are more difficult to fine-tune as they vary based on distance of user from camera. Rule 3 in lateral raises suffered ever so slightly from overlapping joints, which led to a slightly less stable feedback, but that too only in rare cases. An important point here is, these errors were only momentary. The errors showed in Table I were only transient and were visible on screen only for a few frames at most. For example, due to joint overlap Mediapipe, albeit rarely, reported incorrect joint locations, but given the real time response of Mediappe, it was immediately rectified and returned. Therefore, these errors had a negligible effect on the user experience. Regarding the GUI, it was designed to be very responsive and to provide immediate and clear visual feedback to the user. It was built keeping in mind all Heuristics for evaluation of a "Human-Computer Interaction" appication to provide seamless, intuitive, and engaging interface to the user.

In the future, I aim to make improvement concerning increasing the functionality of the system and improving the user experience with several key improvements. For the GUI, the next step would be make it more interactive, including more features that provide a better feel and more intuitive experience for the user. Such as adding a counter for good and bad repetitions, as well as a timer overlooking duration of every repetition which is also an important factor when exercising. Addition of live audio feedback can be a great addition to the application, which could not only point out which form is wrong, but also tell you how to correct it. To enhance the accuracy and robustness of the system, I will explore training with CNNs or other deep neural network architectures using videos of exercises. This will enable the system to learn complex patterns in joint movements and achieve 100% accuracy in detecting exercise form. Such an approach would reduce reliance on traditional rule-based methods and improve the overall adaptability of the system. Also, with neural networks trained on video data, the need for users to face the camera directly could be eliminated, therefore offering greater flexibility in the performance of exercises. The result would be a more robust, accurate, and adaptable system where users can train freely without constraints, while the quality of the form detection and exercise feedback will not be compromised.

VI. CONCLUSIONS

This project successfully demonstrated a rule-based approach to exercise performance evaluation, emphasizing accuracy and real-time feedback. The system integrated various rules, such as movement direction, joint angles, and spatial positioning, in assessing exercises across different contexts. The easiest rules to define were those for movement direction, which had an accuracy of 100% with no errors to ensure that users maintain proper form and technique.

This system would provide feedback in real time, highlighting deviations from the expected exercise form and suggesting corrections where necessary. Although Mediapipe does tend to have a few problems with joint tracking in the 3D space, especially regarding overlapping joints on screen, this usually is transient and rectified within frames, thus having minor disruption to the user experience.

Also, the GUI was very user-friendly, though there is always room for improvement by adding more interactive features and capabilities. Further developments could be done by training neural networks on videos of exercises to capture more complex patterns of joints, which may eliminate the need to directly face the camera by the user, making the system more flexible, robust, and adaptable. This would increase the effectiveness and efficiency of the evaluation process, providing users with a smooth and complete exercise assessment experience.

References

[1] Kolla, Morarjee, Phani Varma Gadiraju, and Dhruvraj Tondanoorthy. "Form Check: Exercise Posture Correction Application." Journal of Electrical Systems 20.4s (2024): 25-33..

[2] Sterbini, Andrea. "Automatic correction of complex technical exercises." 32nd Annual Frontiers in Education. Vol. 3. IEEE, 2002.

[3] Appiah, Kofi Essuming. "A Mobile-Phone Pose Estimation for Gym-Exercise Form Correction." Proceedings, Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. York, 2024..

[4] Raza, Ali, et al. "LogRF: An approach to human pose estimation using skeleton landmarks for physiotherapy fitness exercise correction." IEEE Access (2023).

[5] Rahmadani, Atharian, Bima Sena Bayu Dewantara, and Dewi Mutiara Sari. "Human pose estimation for fitness exercise movement correction." 2022 International Electronics Symposium (IES). IEEE, 2022.

[6] Mediapipe `https://ai.google.dev/edge/mediapipe/solutions/guide`

[7] PyQt5 `https://www.pythonguis.com/pyqt5-tutorial/`

[8] Pycharm (Development Enviornment) `https://www.jetbrains.com/pycharm/`

[9] OpenCv `https://opencv.org/`