AutoCaddie: The AI-Driven Smart Golf Coach

Joshua Glaspey, Jarett Artman, Rylan Simpson, and Louis Deal
Dept. of Computer Engineering and Dept. of Electrical Engineering, University of Central Florida, Orlando, Florida, 32816-245

***Abstract* — The rapid progression of Artificial Intelligence across various commercial fields has left open much potential for human benefit. The AutoCaddie project seeks to bring these benefits towards training the swinging motion within the sport of golf. AutoCaddie uses a deep neural network as an AI model, trained off video data of proper swinging motions to provide feedback on the user's technique, and incorporates hardware sensor data to provide reliable information reading to cover the weaknesses of computer vision. Overall, AutoCaddie as a golf swing trainer provides a new way for golfers to improve their abilities with a low barrier of entry.**

***Index Terms* — Artificial Intelligence, Computer Vision, Deep Neural Networks, Feedback Communications, Sensor Fusion, Sports**

I. Introduction

One of the most challenging aspects of sports training and physical fitness is the lack of personalized, real-time feedback. Whether you're a professional athlete or a casual fitness enthusiast, the absence of immediate and accurate coaching can hinder performance and lead to a plateau in improvement. To address this problem, we introduce AutoCaddie, an artificial intelligence (AI) driven smart coaching system that aims to redefine the way individuals train and improve in their respective sports or fitness regimes.

AutoCaddie employs a suite of sensors and data analytics tools to capture a wide range of performance metrics. These metrics are processed through a sophisticated machine learning algorithm, alongside other optimized computations, which then provide real-time personalized coaching feedback directly to the user. This system is designed to be integrated into existing training and can also function independently, offering a versatile solution adaptable to various sports and physical activities.

In addition to real-time feedback, AutoCaddie features an interface that provides a more in-depth analysis of the collected data. This includes the information of the current swing performed by the player and other data we collected from our camera footage. This feedback will be valuable to the player for their improvement of golf skills. Moreover, all the data captured by AutoCaddie can be securely stored and accessed remotely, serving multiple purposes. It can be used for further analysis, for more nuanced feedback, or even utilized for research and development in sports science.

For potential future and additional various iterations, and to support a customizable experience to better assist in a user's improvement of golf swinging skills, a key approach in the design of the hardware subsystem of the AutoCaddie project is modularity. Modularity allows the user to experience a tailored reviewing session while also incorporating a model that is easily expandable. The systems maintains focus on the arms, hips, and shoulders of the user, but with the introduction of a minimal amount of additional hardware, and a slight tweak of data acquisition from the software, it is possible to include more areas of measurement. For example, the current iteration of the AutoCaddie project emphasizes the use of inertial measurement units (IMUs) to accurately gauge the rotation of the user before and during their swing action, along with other potential kinematic data. However, because of the Modularity of the project, later iterations or packages of the AutoCaddie system, it can easily be modified to incorporate other hardware peripherals, such as foot pressure sensors, in order for a user to properly train their footwork.

As engineers with a passion for golf, we understand firsthand the challenges and limitations of traditional training methods in the sport. The existing systems often lack real-time, personalized feedback essential for meaningful improvement. AutoCaddie emerged from our desire to integrate our engineering skills with our hobbies, aiming to address these gaps. By leveraging AI and sensor technology, we aspire to enhance the act of golf training and also to make advanced, data-driven coaching accessible to golfers at all levels. Through AutoCaddie, we aim to elevate the standards of training, making the sport more engaging and rewarding for everyone involved. Overall, the function of the AutoCaddie project is to capture and analyze the biomechanics of a golfer's swing using a combination of cameras and sensors. The system processes this data in real-time through machine learning algorithms to identify areas for improvement. The analyzed information is then displayed on a computer screen in a user-friendly format, providing golfers with actionable tips to enhance their swing. This enables golfers to make instant corrections, fostering more effective practice sessions and accelerating skill development.

II. Project Outline

The core goal of AutoCaddie is to provide comprehensive feedback to a user regarding the form of their gold swings. To do so, three streams of information must be utilized to provide the feedback model with the sufficient data to provide effective reports to the user of AutoCaddie. As demonstrated by Figure 1, two video data streams are fed into the central computer for processing by the AI model. These cameras are the main source of data which is analyzed during the user's swing. The AI will monitor and provide feedback regarding the user's overall posture, arm straightness, and hip rotation.

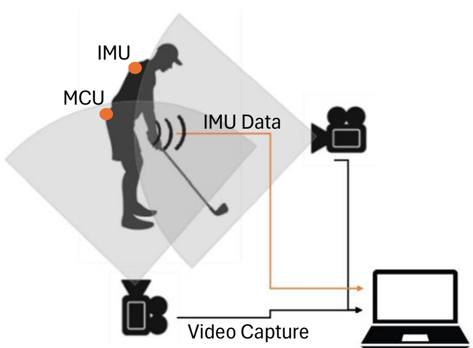


Fig. 1: High-level diagram of data sources.

Additionally, critical to the control flow of the system, the hardware sensor system connected to a wireless transceiver module. This hardware system can collect sensor data worn by the user, such as IMU data, and can wirelessly transmit this data to the computer for analysis. This hardware data seeks to supplement the analysis data generated by the AI model regarding areas that the model may be weak at. Further, as mentioned before, the hardware system is critical to the control flow of the system. The hardware system allows to user to remotely initiate the data collection sequence.

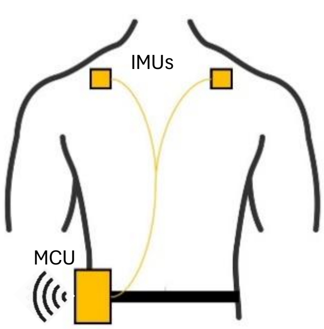


Fig. 2. High-level diagram of hardware placement.

At a high level, the control flow of AutoCaddie can be seen by Figure 3. Following initial setup by the user, the system will detect and calibrate all known peripheral devices. Once the calibration setup phase is complete, the system will enter a standby phase, waiting for the user to start the data collection process. Once the user initiates the process, the system will notify the user that it has begun, while briefly waiting for the user to enter their swing preparation posture. The system will collect sensor data during this phase, such that feedback can be given on the user's posture in preparation for their swing. After this, the system will prompt the user to swing, during which both video camera data and sensor data is collected. After the swing action, the user then references the graphical user interface (GUI) to determine if they are satisfied with their swing and want to move forward with the processing of it. Should they accept, they assist the GUI in matching the exact swing period, following which the AI then begins analyzing the video data. Once all data analysis is performed, feedback is generated, and is then displayed to the user.

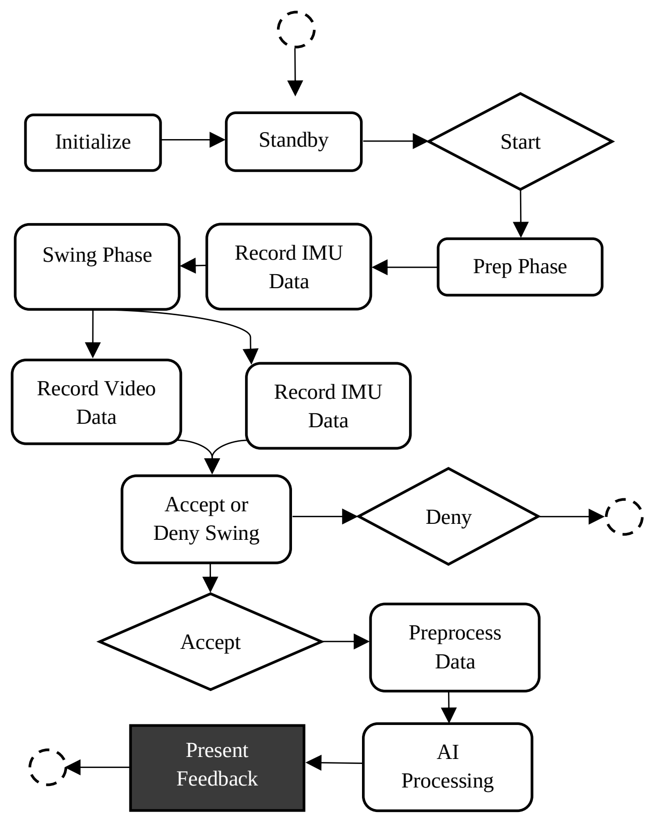


Fig. 3. Control-flow diagram of the AutoCaddie System Hardware-Software interface..

III. HARDWARE

As discussed previously, the hardware system serves a dual purpose. The first responsibility is to collect sensor data relevant to the user's preparation and swing action. This data is collected via a microcontroller unit (MCU), from which the data is wirelessly transferred to the central computer for analysis via a Bluetooth serial link. The current selection for the MCU is the ESP32-WROOM-32E, as its support for the Arduino environment and mature source of libraries makes programming the MCU more streamlined. For the wireless transceiver module, the HC-06 device was chosen. This device supports Bluetooth serial communication links and has a reliable connection strength. In the current iteration of the AutoCaddie project, IMUs were selected as the primary hardware sensor for this project. IMUs allow for the measurement of the rotational movement and acceleration of certain points of the body. Placing IMUs on the user's shoulder blades allows for the measurement of the user's squareness with the ground during the swing preparation phase. For this purpose, the MPU6050 IMU was selected, as its wide availability contributes to its cost effectiveness and library support.

Facilitating user interactivity and understanding for each stage of the AutoCaddie process, the hardware system has several peripherals which allow the user to control or to receive cues from the system. A tactile button allows the user to easily choose to start the recording sequence when they are ready, directly from the MCU board which will be fastened to the user's hip. An LED is used to indicate when the MCU is ready for recording, and the active buzzer allows the MCU to indicate what recording stage has been reached.

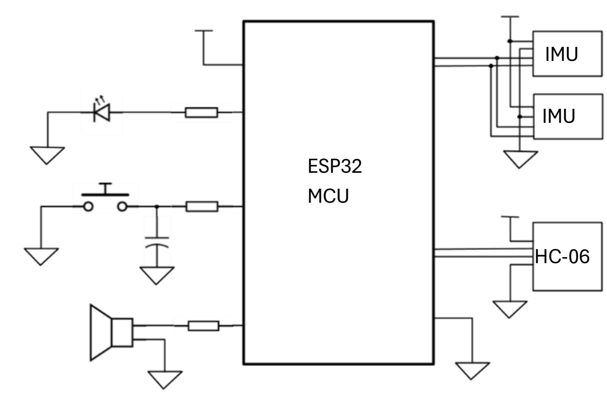


Fig. 4. Simplified Circuit Diagram of MCU Connections.

Due to the rapid rotation of the user and the swinging of the golf club during the user's swing, the impedance of the user's movement from the hardware system is a primary concern. This concern is exactly why a wireless communication protocol was chosen to send data from the MCU to the central processor computer. Any wires extending from the IMUs to the MCU board will be sheathed and guided by suspender and Velcro straps (discussed later). Further, the profile of the MCU board will be restricted to be less than 100cm^2, to not generate any discomfort for the user during the swing. The current profile of the larger PCB is 71.25cm^2

Regarding the data collected by the IMUs, quaternion data will be collected to determine the rotational vectors of the user's shoulders. Quaternions are preferable in comparison to other types of rotational data, such as Euler vectors as they do not suffer from gimbal lock [1]. Gimbal lock, when two or more rotational axes match and "lock" with each other, can cause precision errors and loss of degrees of freedom. This is particular to the measurement of the user's "squareness" with the ground during the swing preparation phase. For this, quaternion axes can be analyzed individually and compared with threshold values demarcating a straight posture. Additionally, the linearity of these quaternion vectors in a plane can be determined from the angle determined by Equation 1.

$$\theta = \cos^{-1}(2\langle q_1, q_2\rangle^2 - 1) \tag{1}$$

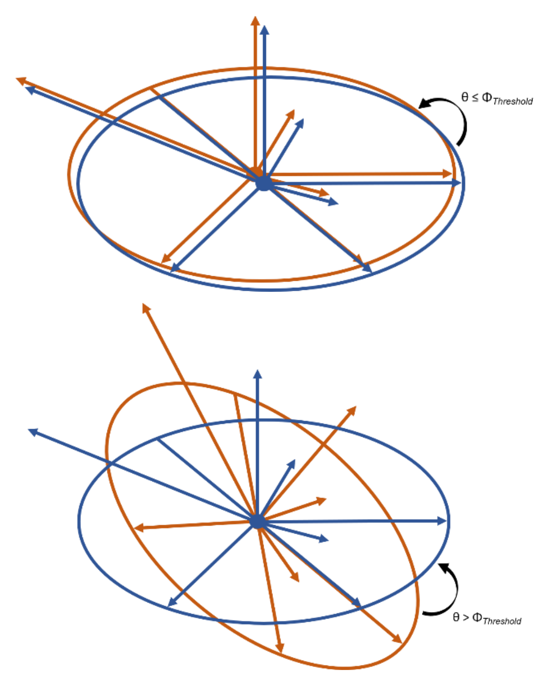


Fig. 5. Quaternion diagrams demonstrating quaternion linearity (top) and nonlinearity (bottom).

From the initiation of each phase in the data collection sequence, the MCU will provide headers within the transmitted quaternion data. These headers include "READY", "PREP", "SWING", and "END". Data is collected and transmitted during the "PREP" and "SWING" phases.

The power consumption of the hardware system is another primary concern. The hardware system has been targeted to last for more than two hours. As such, a sufficient power supply must be provided. For this, a 6.6Ah Li ion battery was selected. In combination with lower power modes during standby, with a maximum draw of 1.3A from the ESP32, the hardware may have enough power to run for over 5 hours. This allows a user to use the AutoCaddie system for an extended period, allowing them to experience a thorough and effective training session, unrestricted by time.

To mount the hardware peripherals on the user, Velcro straps in conjunction to provided enclosure materials can be provided. As the main point of attachment, a pair of suspenders will be provided to the user. These elastic suspenders will have Velcro-ready material applied to the shoulder region, such that containers designed for the IMUs with Velcro pads can be easily attached and removed from the user. With this, the user can easily relocate the IMUs, such that they are in the correct position for effective measurement. Additionally, a belt-like Velcro strap will be used to attach the MCU enclosure to the user. This will allow the user to place the hardware peripherals in a position that is comfortable for the user, and in a position that will not affect their swings.

IV. SOFTWARE

The software side of AutoCaddie aims to contain all communications, and act as an interactive tool for navigating the product. The system can be described using the flow diagram shown in Figure 7. It will serve as the main processor and feedback controller. For each of the data acquisition subsystems, the software needs to be able to receive results, manipulate data, then display it back to the user. The format of the output will be described in a later section.

In terms of programming languages, Python was selected for usage in all components of the project, including using the GUI, performing data processing, and interacting with a deep neural network model. Python is revered for its simplicity and versatility and will allow for the communication between different systems in the project to be completed more seamlessly. Also, this grants access for the project to utilize complex mathematics libraries, specifically NumPy [2], to process matrices representing image data.

In contract, Python is notorious for being slower during runtime due to it being a dynamic language. However, this fact can be overlooked as the processing is performed after the user completes a swing sequence and does not disrupt this action.

A. Subsystems

AutoCaddie's architecture is comprised of two main systems: a video system and a physical sensor system. The relationship between AutoCaddie's subsystems is shown in Figure 6.

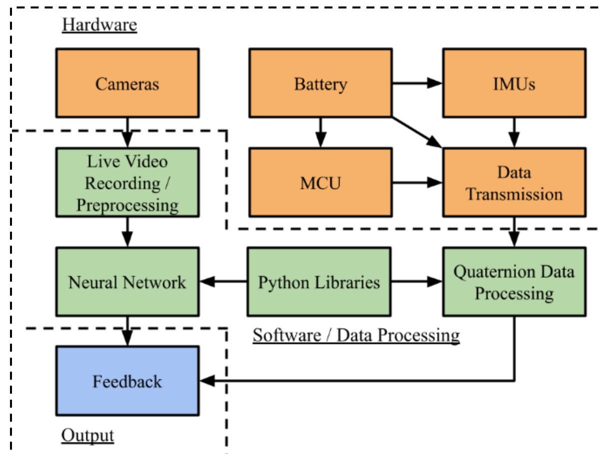


Fig. 6. Relationship diagram between AutoCaddie's subsystems.

The video system oversees the implementation of computer vision, or the genre of distinguishing patterns from pixels comprising an image. Computer vision is pivotal for the success of AutoCaddie, as it allows for the usage of modern technology to analyze the user in real time.

This video system comprises two main components of AutoCaddie: visual pattern distinction with user angle measurements over time, and AI analysis using a deep neural network. The neural network will be explained in Section V. Using OpenCV2 [3], an open-source computer vision library, it is possible to employ video editing, image processing, and computer vision tasks. The computer vision tasks are to be able to record the user during a golf swing sequence, be able to decompose video feed into images, be able to create a seamless recording process by referencing the same camera feed at different instances of time, and be able to create video input taken from the user for analysis by the AI model.

AutoCaddie is consistent of two USB cameras, tethered to the main computing device. These cameras record at their native framerates, but for the purpose of this project, a minimum of 30 frames per second (fps) is required. This allows for the cameras to not omit important information during the swing if the user's motion is faster than the updating rate. It is important to utilize two cameras to capture every characteristic of the user's swing. By nature, constraining the video feed to one perspective limits the

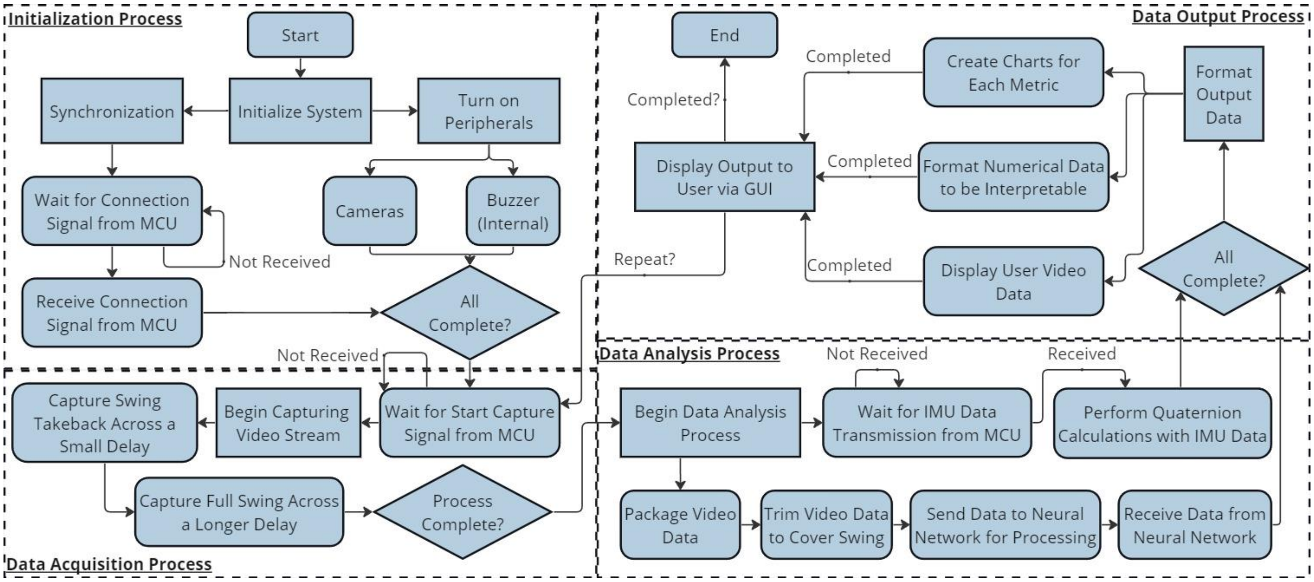


Fig. 7. Software Flow Diagram of AutoCaddie.

perspective of motion to two dimensions. This is solvable by incorporating two perpendicular viewports, allowing for three-dimensional capture. The webcam selected for this project is the HP 320 FHD due to its cost affordability for this project while also surpassing the minimum fps limit.

The other side of AutoCaddie consists of the hardware system, as described by Section III. Each IMU device is mounted on one of the user's shoulders, where it transmits quaternion data to the MCU via a cable. The MCU then sends the data to computer on request, which is stored as a comma-separated-value (CSV) file. This allows for easy reference at any point in the project.

Moreover, recording data as quaternions and using matrix operations allows for more dynamic processing. It is possible to represent each quaternion axis as a row, and each individual point as a column in a matrix. This way, the number of columns can be unlimited, and the algorithmic implementation remains the same.

$$Q_{ALL} = [Q_1, Q_2, \dots, Q_n] = \begin{bmatrix} q_{1\theta} & q_{2\theta} & \cdots & q_{n\theta} \\ q_{1x} & q_{2x} & \cdots & q_{nx} \\ q_{1y} & q_{2y} & \cdots & q_{ny} \\ q_{1z} & q_{2z} & \cdots & q_{nz} \end{bmatrix} \quad (2)$$

$$n = Number\ of\ measurements$$

Storing the data in this format allows for simple processing. In terms of rotational measurements, each quaternion measures the rotation around a specified axis. This can be demonstrated by Figure 8. Thus, for the relevancy of this project, AutoCaddie focuses on the rotation of the z-axis, represented as the bottom row of the matrix formed in Equation I. This will be justified in the Feedback subsection of this section.

The software of AutoCaddie is divided into modules, each storing different components of functionality. The abstraction of these modules makes the software modular, which makes the addition of new features or expansions easier in the future. Specifically, the three modules are data processing, GUI, and neural network.

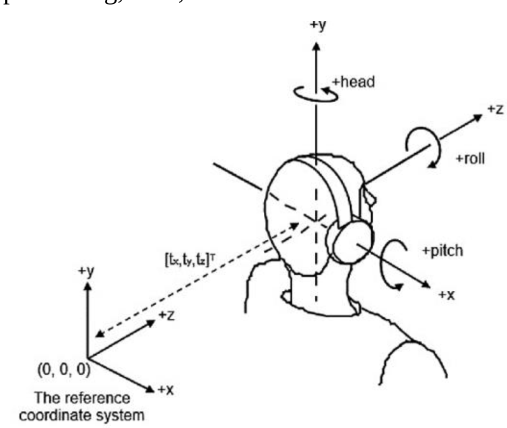


Fig. 8. Representing the quaternion axis of measurement along a user. AutoCaddie focuses on roll (rotation along the z-axis). Referenced via [4].

B. Feedback

There are several modes of feedback given to the user. The user will be evaluated in terms of their initial posture, inner arm straightness, hip rotation, and shoulder rotation. A description of each mode of feedback follows this paragraph.

The initial posture of the user is measured as to how square the shoulders are before beginning the taking back of the club during a swing. As displayed by Figure VI, the roll of the user represents the angle around the z-axis pointing through the user. By measuring the roll of each shoulder with an IMU mounted on each, it is possible to determine the relative squareness of the shoulders before beginning the sequence. This is important for contributing to proper preparation for a swing.

The inner arm straightness refers to the non-dominant arm's measure of bend during the swing sequence. This is important through the downward acceleration into a golf ball. The inner arm acts as a guide for the swing, and maintaining its straightness ensures a consistent motion.

The hip rotation is critical for generating power in a golf swing. The rotational movement of the hips allows for the transfer of energy from the lower body to the upper body, allowing for greater club speed.

Shoulder squareness during the swing is important for two primary factors: balance and injury prevention. Ensuring that the shoulders are properly aligns keeps the upper body straight during the swing, which helps center the user's center of mass. Additionally, maintaining balance helps prevent strain on specific muscles that may control more weight than the rest of the body.

As mentioned prior, the user's initial upper body posture is recorded by the hardware system. The user's inner arm straightness, hip rotation, and shoulder rotation will be extracted using computer vision, and analyzed by the neural network model, which will be explained in Section V.

C. Graphical User Interface

The GUI is the user interface for AutoCaddie. It serves as the means of controlling the system and being able to be given visual feedback. It is implemented in a cyclic design manner, where the entire process is repeatable without dependence on earlier executions. The GUI can be divided into several stages: calibration, recording, user-processing, and feedback.

The calibration stage is the process where the user initializes two camera feeds and a Bluetooth serial connection with the MCU. The user cannot proceed until both are established. AutoCaddie's performance is boosted through the usage of threading, or the idea of processing events in parallel. This allows for the GUI to be reflective of the status of a Bluetooth connection thread and camera connection thread, neither of which will introduce blocking code into the main thread. The user may also have multiple cameras and Bluetooth connections active than required by the system and still execute. The system can allow the user to select the correct cameras to use, as well as only establishing a connection with a serial port under the specification of "HC-06."

The recording stage of the GUI happens as a sequence of events. This can be described as a "Ready, Set, Go!" process. The physical buzzer sends an initial beep, signifying for the user to prepare for the swing. A second beep will sound, signifying to be in the setup position. During this stage, the IMUs will collect data while mounted on the user's shoulders. A third beep, sounded as two short beeps, will sound, signifying that it is time to complete a swing. The user's swing will here be recorded through the two separate camera feeds. Lastly, a final beep will sound, signifying recording has finished.

This moves the user to the processing stage of the GUI. To assist in creating accurate feedback, the video must conform to the same starting and ending points as the video data used by the neural network model. The user selects a starting and ending frame, where the starting frame will be as close to the beginning of the swing takeback as possible, and the ending frame will be as close to the end of the follow through as possible.

Lastly, the user will be given feedback on each metric described earlier. For each metric, there are two main forms of feedback: visual and numerical. The visual feedback consists of two videos. The first video represents the user's swing in real time, with a skeleton mapped over their joints. The angle history of the swing will be displayed in the top of the video. Additionally, a video representing all points of the user's skeleton in a 3D space will also be provided, which gives a visualization of the important features of the swing for the user to reference. There will also be a timeline provided for the user to observe their swing. It will be mapped against AutoCaddie's AI predicted swing, which takes the first couple frames of motion from the user, and generates a proper swing. By comparing the actual angles against the predict ones, the user can compare what part of their swing they need to adjust.

Moreover, for inner arm straightness, there are expected angles the user can anticipate must be met. Therefore, this metric features pre-selected angles, which will be provided to the user. These are to be referenced at key points in the swing, to ensure that the right motion is being followed.

Fig. 9. Feedback for front arm angle.

V. NERUAL NETWORK MODEL

The heart of AutoCaddie's capability to provide nuanced feedback on golf swings lies within its deep learning architecture, a neural network model meticulously designed to analyze the biomechanics of golf swings. This section delineates the processes involved in capturing data, extracting meaningful features through advanced computer vision techniques, and preparing this data for the neural network training, which collectively empower AutoCaddie to offer personalized coaching insights.

A. Data Capture and Preprocessing

The initiation of our neural network model's training process begins with an elaborate data capture setup involving two high-definition cameras, meticulously positioned to capture the golfer's swing from distinct vantage points. This dual-camera setup not only enhances the depth of the data collected but also mitigates the limitations posed by single-view analysis, thus allowing for a comprehensive capture of the swing dynamics. Following the capture, the videos undergo a preprocessing stage where each frame is analyzed using the MediaPipe [5] framework. This advanced computer vision technique facilitates the extraction of critical postural landmarks from the golfer, such as the positioning of arms, hips, and shoulders during the swing. This stage is crucial for identifying key biomechanical angles and rotations that are indicative of a proficient swing.

B. Feature Extraction

With the preprocessing phase providing a foundational layer of raw data, the subsequent step involves the intricate extraction of features pivotal for assessing the quality of the golf swing. Through the calculation of angles and rotations between specific landmarks identified in the preprocessing phase, we derive quantitative measures that embody the swing's biomechanics. These measures include the front arm angle, hip rotation, and shoulder rotation. Each feature serves as a critical indicator of swing effectiveness and is instrumental in detecting common swing faults.

A visualization of this is displayed in Figure 10. The swing metrics are displayed to the user for every frame within the swing, and if the measured angle is beyond an error threshold, it is signified as problematic with red text. Through testing, a consistent threshold value is 25 degrees.

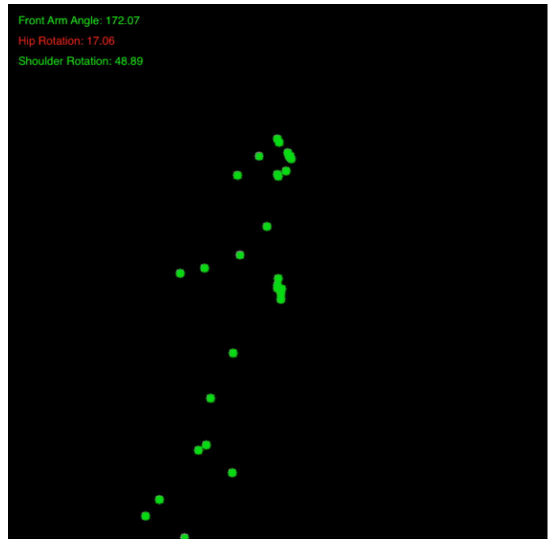


Fig. 10. 3D point cloud representing the user's swing at an exact frame during the driving motion.

C. Sequence Data Preparation and Model Training

The culmination of the data preparation process involves the transformation of the extracted features into a format conducive for neural network training. Recognizing the dynamic nature of a golf swing, the model is trained on sequences of data points that represent the temporal progression of a swing. This approachhip rot allows the neural network to not only assess individual postures but also understand their progression and coordination throughout the swing. The input to the model consists of sequences of angles and rotations, capturing the essence of the swing dynamics, while the training process fine-tunes the model to identify patterns associated with both effective and faulty swings.

The neural network model at the core of AutoCaddie is a testament to the synergy between advanced computer vision techniques and deep learning. Through meticulous data capture, feature extraction, and sequence data preparation, the model is trained to offer personalized feedback with the goal of enhancing the golfer's technique. This AI-driven approach not only sets a new precedent for sports training technologies but also embodies our commitment to leveraging cutting-edge technology to foster athletic excellence.

D. Hybrid CNN-LSTM Architecture

Central to AutoCaddie's neural network model is the implementation of a hybrid Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) architecture, a design choice motivated by the unique requirements of processing and analyzing golf swing data. The CNN component excels at extracting high-level features from the input data by identifying patterns in the spatial arrangement of the body's landmarks across individual frames. This capability is crucial for recognizing key postures and alignments that signify the technical aspects of a proficient swing.

Building upon the feature-rich representations generated by the CNN, the LSTM layers are adept at capturing temporal dependencies and dynamics across the sequence of a golf swing. By analyzing the temporal progression of body postures and movements, the LSTM component allows for a deeper understanding of the swing's fluidity, coordination, and timing—elements that are fundamental to evaluating swing quality and identifying areas for improvement.

The fusion of CNN and LSTM architectures enables AutoCaddie's neural network to leverage the strengths of both spatial feature extraction and temporal sequence modeling. This hybrid model not only facilitates a comprehensive analysis of each golf swing but also enhances the model's predictive capability by learning from the nuanced patterns that characterize effective and ineffective swings.

VII. Conclusion

Overall, the AutoCaddie system incorporates both hardware technologies as well as a deep neural network trained AI to analyze a user’s swing form. From the feedback generated by the AutoCaddie system and presented by the graphical interface, users can engage with the system to improve their golf abilities.

ACKNOWLEDGEMENT

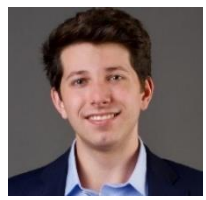
The authors wish to acknowledge the University of Central Florida and its faculty for helping in making this project possible. Specifically, the team extends its gratitude to Dr. Zakhia Abichar, Joon-Hyuk Park, and John Aedo for agreeing to act as the committee and provide assistance with the project.

BIOGRAPY

Joshua Glaspey is a 21-year-old graduating Computer Engineering student. Graduating in 2024, he will be pursuing an internship in Redmond, Washington this summer with Microsoft as a software engineer, and will further develop his education this coming fall with a master's in computer engineering at the University of Central Florida.

Rylan Simpson is a 21-year-old graduating Computer Engineering student. Graduating in 2024, he will be pursuing a career with Integrated Systems For Solutions (IS4S), focusing on hardware to software integrations, and system communications. He plans to develop a robust engineering background in industry while advancing his academic background.

Louie Deal is a 21-year-old graduating Electrical Engineering student. Graduating in 2024, he will be pursuing a full-time position with Lockheed Martin beginning in August of 2024. He will pursue a master's in electrical engineering after 5 years spent working in the defense industry.



Jarett Artman is a 22-year-old graduating Computer Engineering student. Graduating in 2024, he will be pursuing a full-time career in private equity focused on integrated systems and machine learning (ISML). And will be continuing the BS-MS track in Computer engineering.

CITATIONS

[1] J. H. Challis, “Quaternions as a solution to determining the angular kinematics of human movement,” *BMC Biomedical Engineering*, vol. 2, no. 1, Mar. 2020, doi: https://doi.org/10.1186/s42490-020-00039-z.

[2] A. Saxena, “Introduction to NumPy,” Medium, Jun. 01, 2020. https://medium.com/analytics-vidhya/introduction-to-numpy-279bbc88c615

[3] OpenCV, “OpenCV library,” Opencv.org, 2019. https://opencv.org/

[4] “A Low-Cost Infrared-Optical Head Tracking Solution for Virtual 3D Audio Environment Using the Nintendo Wii-Remote - Scientific Figure on ResearchGate.” https://www.researchgate.net/figure/Axis-rotations-pitch-yaw-and-roll-and-translation-specified-by-a-translation-vector-t_fig2_284119646

[5] “MediaPipe,” Google Developers. https://developers.google.com/mediapipe