Video Recognition Technology: Enhancing Golf Swing Performance for Players of All Levels

Joel Klein   
*Luddy School of Informatics, Computing, and Engineering*  
*Indiana University - Bloomington*Salt Lake City, UT  
joeklein@iu.edu

Gabriel Levy   
*Luddy School of Informatics, Computing, and Engineering*  
*Indiana University - Bloomington*Bloomington, IN  
gaalevy@iu.edu

Wes Martin   
*Luddy School of Informatics, Computing, and Engineering*  
*Indiana University - Bloomington*Houston, TX  
wemarti@iu.edu

*Abstract*—This research project focuses on the use of computer vision and deep learning techniques to simplify golf swing analysis. Golf is experiencing significant growth in the US, and its economic activity is valued at $84.1 billion. Golfers must invest a substantial amount of time and money to improve their game, with the golf swing being a complex and challenging movement to perfect. Perfecting the golf swing is a key driver of success in the sport and requires extensive training, attention to biomechanical detail, skill, physical ability, and practice. The project extends pre-existing golf swing sequencing research by McNally et al. and uses optical flow processed videos to train a hybrid CNN and LSTM deep learning algorithm effectively identifying eight key event frames within swing videos. Identifying key swing event frames simplifies swing analysis and enables extraction of key biomechanical features for comparison to previous swings or even professional golf swings. This research offers more accessible opportunities for amateur golfers to improve their game.

# Introduction

### Since the COVID-19 pandemic, golf is one of the fastest growing sports in America with 3.3 million Americans playing golf for the first time in 2022. Golf’s overall reach is an estimated 119 million people with approximately 41.1 million Americans playing golf either on or off-course in 2022 [1]. World Golf Foundation as of 2016, values the golf industry’s economic activity at $84.1 billion [2]. There is no doubt there is substantial economic and business opportunity within the sport. Whether its green fees, driving range/practice fees, equipment, travel, food/beverage, or lessons, most golfers undertake a significant time commitment and financial hit to play and improve on the game they love (and sometimes hate on a bad day).

### The objective as a golfer is hitting the golf ball in the hole in as few shots as possible. This, to many golfers, is much easier said than done. The key driver to a golfer’s success is the ability to perfect the golf swing, a rigorous, unique full body motion requiring fluid coordination from head to toe in order to achieve optimal results. Building and improving a player’s golf swing(s) requires substantial time in training, meticulous attention to biomechanical detail, high levels of skill, sheer physical ability, and substantial practice. To put this more in perspective, the average tour pro golfer spends about three hours a day practicing his/her golf swing [3]. The golf swing, due to its biomechanical complexity, is heavily analyzed, fine-tuned, adjusted, and repeated to improve. Often, it takes amateur golfers several years to make even the slightest incremental adjustments and improvements in their swing. It is very common for golfers to seek out external mechanisms to improve.

### Most amateur golfers often turn to a series of learning and practice methods to achieve better performance. Like many sports, watching and observing professional players helps many players get started on the basic motion. However, it soon grows evident to the beginning golfer that more structured coaching is required to reach slightly higher levels of play. Golfers typically analyze and adjust their golf swing via one of two common approaches: golf instructors and technology.

### This project research focuses on using computer vision and deep learning techniques to streamline amateur golfer swing analysis via golf swing sequencing extending pre-existing research by McNally, et. al. Golf sequencing is the process of identifying the eight key frames within swing videos: Address, Toe-up, Mid-backswing, Top, Mid-downswing, Impact, Mid-follow-through, Finish. Identifying these key events, simplify and streamline the ability to analyze golfer biomechanics throughout the swing. After identifying the key events/images of a golf swing, this research uses MediaPipe’s Pose human pose estimation technology to extract key golfer biomechanical features at each event. Upon upload of a swing video, these biomechanical features can be extracted and compared to professional golfers and/or previous swings facilitating more accessible opportunity for improvement.

# Background and Related Work

## Golf Instruction

## Golf instruction for beginners often starts with the basics and general rules of thumb such as club grip, stance, club head placement, backswing motion, and ball striking. Once beyond improving the basic motions, golf instruction grows increasingly difficult as the instructor’s naked eye can not completely analyze the fast-paced biomechanical motion throughout the entire swing. Slow motion swing video often mitigates this key issue and aids the instructor and golfer to more steadily analyze the motion picture and suggest improvements throughout the key points in time of the swing.

### Although slow motion cameras are highly accessible to the average American today, there are two major underlying challenges hindering this learning approach: cost and instruction consistency. The typical cost for golf lessons or instruction from a professional or coach is $50-$75 per 30 minutes. This expensive option is a massive barrier to entry for middle class, amateur golfers who are most often budgeting to simply get on the course and play the game. Perhaps the bigger challenge is the quality, value, and consistency of the instruction. An instructor doesn’t obtain the ability on their own to provide data, biomechanical measurements, or motion analysis for consistent feedback in improvement. This sometimes leads to subjective, biased, and inconsistent feedback to the golfer.

## Motion Technology/Computer Vision in Golf

### High-speed cameras, motion capture systems, force plates, and pressure mats are just some of the technologies used to capture and analyze the various aspects of the golf swing. By providing detailed data on the kinematics and kinetics of the swing, these technologies can help golfers and coaches identify areas of weakness and inefficiency and develop targeted training programs to improve swing mechanics and performance. However, these technologies are not accessible to the average golfer attempting to improve their swing on their own on a budget.

### Many consumer facing companies such as K Motion offer full body wearable devices to track human motion data through the swing. In addition to full body wearables, many apps including Golf Swing Analyzer++ pair with wearable devices with sensors such as the Apple Watch to generate such tracking data. Outside of wearable devices, there are several other sensor-based options such as Arccos Caddie Smart Sensors which are sensors placed on the club heads. Again, the barrier to entry for these products is the cost reducing the accessibility for the average middle-class golfer. The K Motion K Player wearable system costs $3,500. The Apple Watch Series 8 costs $429 plus the monthly subscription to use the swing tracking application and the Arccos system costs $200.

### Prior computer vision research suggests several approaches for swing analysis simply using an uploaded golf swing video and club motion tracking. Gehrig et. al. fit a global motion swing trajectory model to estimate club location from single frames facilitating analysis on club location, speed, and trajectory [10]. Urtasun et al. also estimates swing motions by incorporating dynamic models into human body tracking to produce 3–D reconstructions from monocular swing image sequences [11]. Although these existing works are highly effective approaches for club tracking, they lack focus on the individual golfer’s posture, the most key biomechanical elements to a successful swing.

### Several researchers target posture analysis through golf swing sequencing. McNally et. al. propose a hybrid CNN and LSTM deep learning algorithm to identify key swing events using 1,400 golf swing videos in GolfDB sourced from YouTube [9]. Kim et. al. use vision-based pose estimation to find key video frames in golf swings to streamline feedback and swing improvements [12]. Liao et. al proposes a neural network system comparing two motion swing sequences and identifies keyframes in which significant differences can be observed between amateur golfers using the system and professional players [13]. Sun also proposes a similar technique detecting key frames in the video stream using pose estimation and outputs posture change recommendations based on player posture distance [14]. This research uses the swing sequencing analysis of McNally et. al. as its basis due to the model implementation simplicity and size, promising initial results with opportunity to improve, and easy access to publicly available, labeled event GolfDB video data. In addition to improving the golf sequencing performance, this research leverages the output frames from the model for human pose estimation and biomechanical swing analysis and feedback.

## Golf Swing Sequencing

### Golf swing sequencing is a task involving identifying events in videos of trimmed golf swings. By identifying the key events throughout the complex golf swing, the swing analysis is simplified for the end user and streamlines the focal points for analysis through automatic detection. Golfers or instructors who want to view a golf swing sequence can record a video of a single golf swing on a mobile device, which eliminates the need for spatio-temporal localization to the start and end points of a golf video. Because a single video swing sample will contain a fixed number of events occurring in a specific order, deep learning algorithms can be trained efficiently with less noise to enhance key event frame detection accuracy.



Fig 1. Eight events in a golf swing sequence. The swing event names from left to right are Address, Toe-up, Mid-backswing, Top, Mid-downswing, Impact, Mid-follow-through, and Finish.

## The key golf swing events defined by McNally et. al. in GolfDB are defined as follows:

### Address (A) refers to the position just before the takeaway, which is the frame just prior to the start of the backswing.

### Toe-up (TU) is when the shaft is parallel to the ground during the backswing.

### Mid-backswing (MB) is when the golfer's arm is parallel to the ground during the backswing.

### Top (T) is when the golf club changes directions at the transition from the backswing to the downswing.

### Mid-downswing (MD) is when the golfer's arm is parallel to the ground during the downswing.

### Impact (I) is the moment when the clubhead makes contact with the golf ball.

### Mid-follow-through (MFT) is when the shaft is parallel to the ground during the follow-through.

### Finish (F) is the moment just before the golfer relaxes into their final pose.

### The model architecture described in the next section attempts to identify the exact frames in the labeled GolfDB videos where these eight events occur. It is important to note a major challenge presented by the sequencing problem is that each of the eight events are not confined to a single frame. For instance, the mid-backswing may span multiple frames within the video even though a single frame is labeled Graphical user interface Description automatically generatedas the mid-backswing event. Due to this challenge, the model evaluation and loss function during training need to be more conservatively adjusted as described in the next paragraph.

Fig 2. SwingNet, a deep hybrid convolutional and recurrent network for swing sequencing. SwingNet maps a video sequence of swing images I to a sequence of event probabilities e. First, MobileNetV2 generates a sequence of feature vectors f, which are input into a bidirectional LSTM. The LSTM output at each frame t is then fed into a fully-connected layer, and a softmax function is applied generating eight event probabilities. [9]

### Evaluation: McNally et. al. introduce a tolerance δ on the number of frames to correctly detect an event in videos. This same approach is used for evaluating model performance in this effort. They propose a sample-dependent tolerance level based on the number of frames between Address and Impact for slow-motion videos. The sample-dependent tolerance is defined as δ = max(n/f,1), where n is the number of frames from Address to Impact, f is the sampling frequency, and [x] rounds x to the nearest integer. The authors introduce the PCE evaluation metric, which is the "Percentage of Correct Events" within tolerance δ. This project splits the video data into four random samples for cross-validation. The averaged PCE over the four splits is the primary evaluation metric for model performance.

## SwingNet: A Swing Sequencing Network

### McNally et. al. proposes the first network architecture SwingNet (figure 2), a hybrid CNN plus LSTM model to identify the eight golf swing events within swing videos. The CNN architecture is based on MobileNetV2, fitting for mobile vision applications because of its inverted residual structure and lightweight depth wise convolutions [9]. SwingNet takes a sequence of images from the swing video and maps a corresponding sequence of nine event class probabilities, including eight golf swing events and one no-event class.

### Identifying golf swing events requires information from prior and after frames within the sequence, requiring temporal information within the model structure. SwingNet captures temporal information using a sequence of feature vectors produced by the MobileNetV2 CNN, which are input to an N-layer bidirectional LSTM with H hidden units in each layer. The LSTM is trained bidirectionally as event predictions benefit from frames surrounding each side of the event as contextual information. At each frame, the output of the LSTM is fed through a final fully connected layer, and a softmax is applied to obtain class probabilities. The weights of the fully connected layer are shared across frames. McNally et. al. performs an ablation study to identify the appropriate suite of configurations for the full architecture. They report the model parameters generating the best PCE result of 76.1% as follows:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Input**  **Size** | **Seq. Length** | **Batch Size** | **ImageNet Weights** | **LSTM Layers** | **LSTM Hidden** | **Bidirectional** |
| **160** | **64** | **24** | **Yes** | **1** | **256** | **Yes** |

Table 1: SwingNet model configuration parameters from McNally et. al. best performing model

## This network configuration serves as the basis for experiments to run throughout this study. Rather than experimenting further with differing model architectures or configurations, this study focuses on testing several preprocessing techniques to the input video sequence to improve PCE results. After improving swing event detection, this study proceeds to feed the event frames detected from the model into a human pose based biomechanical swing analyzer.

## MediaPipe Pose: Human Pose Estimation

### MediaPipe Pose is an open-source library developed by Google that enables real-time human pose estimation on image and video files. It uses a deep neural network to estimate the 2D key points of a person's pose in real-time, which can be used for a variety of applications such as fitness tracking, motion analysis, and, in this case, golf swing biomechanical analysis. The API identifies 33 3D landmarks (figure 3) which could then be used to calculate biomechanical features such as hip angle, arm bend angle, and footing.

Chart, radar chart

Description automatically generated

Fig 3. MediaPipe’s Pose human pose joint landmark features

### In addition to pose estimation, MediaPipe Pose includes a module to perform human segmentation, effectively removing background from human videos. Such background removal could add significant value to reduce noise from swing videos taken throughout different sceneries on a golf course.

### MediaPipe Pose provides pre-trained models for upper body, full body, and hand pose estimation, and allows users to adjust the model complexity and required confidence in the estimation in order to return a valid landmark value. Users are also allowed to train their own custom human estimation models for specific tasks. Although this project doesn’t address building a custom human pose estimation model for golf swing analysis, it is a potential future option to improve swing sequencing and biomechanical features.

### It is important to note that the library supports popular platforms such as Android and iOS, making a great fit as a module in the golf swing analyzer intended for mobile device deployment. MediaPipe Pose is a powerful and flexible tool for developers looking to incorporate real-time human pose estimation into their mobile applications.

## Golf Swing Biomechanics

### The biomechanics of the golf swing have been studied extensively in recent years, with the aim of understanding the mechanics of the movement and identifying ways to optimize performance and reduce the risk of injury. Golf swing biomechanics involves the analysis of the body movements, joint kinematics, and muscle activation patterns involved in the swing, as well as the forces and torques generated by the club and ball.

### Human pose estimation can provide valuable information about the golfer's body posture, joint angles, and movement patterns, which can be used to analyze and improve the swing biomechanics. By estimating the positions of key body joints such as the hips, shoulders, elbows, and wrists throughout the golf swing, we can calculate joint angles that are indicative of the golfer's performance and technique. One common approach to calculating joint angles is to use the law of cosines. This involves calculating the angles between the vectors defined by three adjacent joints. For example, the backswing and downswing planes of the club can be determined by the angles between the hands and the golfer's shoulders and hips. The sequencing and timing of the different phases of the swing, such as the backswing, downswing, and follow-through, in addition to the amount of joint rotation or flexion/extension can also be analyzed using joint angle data. This information can be used to identify areas of improvement in the golfer's technique and develop personalized training programs that target specific weaknesses.

### One of several factors that influence the biomechanics of the golf swing is the individual differences in golfers' physical characteristics and swing styles. For example, the length and flexibility of a golfer's muscles, bones, and joints can affect their range of motion and ability to generate clubhead speed. Similarly, the golfer's swing style, such as their grip, stance, and swing plane, can influence the trajectory and spin of the ball, as well as the efficiency and consistency of their swing. Normalizing for differences in golfer physical attributes is a key challenge to overcome when generating swing comparisons of users to professional golfers. This research attempts to address some, but not all, of these normalization challenges by focusing on biomechanical features unaffected by physique.

# Data

### The video data for the project will be sourced from GolfDB, a video database for golf swing sequencing that consists of 1400 high-quality trimmed golf swing videos. The creation of GolfDB by McNally et. al. marks the first significant dataset that is exclusively intended for computer vision applications in the realm of golf.

### Four annotators were employed to label each sample's frames, identifying full golf swings and eight events. However, the accuracy of event labeling was occasionally affected by the sample's native framerate. Besides event labeling, the annotators drew bounding boxes, entered club and view type, and noted whether the sample was in real-time or slow-motion. The annotators also identified the player's sex and name, and domain-specific knowledge was given before annotation, and the dataset's quality was verified by an experienced golfer.

### The database consists of both slow-motion and real-time swings shot from all over the standard golf course (driving range, tee boxes, fairways, and sand traps) from 248 PGA, LPGA, and Champions Tour golfers via YouTube. The videos are sampled at 30 fps and 720p resolution (both a lower sample rate and resolution than that of the modern-day smartphone camera offers). In future experiments, adjusting these parameters may lead to significantly better golf sequencing results.

# Methods

## Golf Swing Sequencing

### The project seeks to improve the existing SwingNet baseline hybrid CNN/RNN deep learning algorithm to sequence and identify eight key frames within swings.

### We explore new methods for achieving prediction of correctly identified golf swing events. The primary method we explore is bias correction of videos using either optical flow or background removal video transformation prior to passing the videos as the input to the existing SwingNet model. More specifically, we would like to know if the model overfits based on the location of the video as opposed to focusing on the amateur golfer’s motion as well as if an excessive amount of irrelevant information is present within the swing videos.

### In addition to video bias correction, we explore extending the LSTM portion of the SwingNet architecture adding state of the art human pose estimation features as input to determine its effects on improving golf sequencing accuracy. There are 99 features (x, y, z coordinates for 33 joint landmarks) that we append to the output of the MobileNetV2 CNN before feeding to the bidirectional LSTM.

### To measure if the model qualifies as “correct” algorithmic approach, we will evaluate the updated model to detect eight golf swing events at an average PCE rate better than baseline model of 76.1% and six out of eight events at a rate of 91.8%. The model accuracy is measured by the percent of correct events within a tolerance—the percentage of correct events identified within a frame tolerance of the actual event occurrence within a video. For instance, if the predicted action is within three frames of the actual event and the tolerance is plus or minus ten frames, the predicted action will be classified as correct. To estimate the model PCE, the data is randomly split into 4 training sets and 4 validation sets. The average PCE across the 4 splits is reported for the best model.

### The Google Collaboratory cloud environment was chosen as it is a fully managed Jupyter notebook python environment with the ability to train Pytorch models on graphical processing units (GPUs) operated and made available by Google. This compute environment provided the fastest way to experiment versus alternative options where the team would need to set up a custom environment.

### Optical Flow: Optical flow can be used on golf swings to capture and analyze the motion of the golfer and the club throughout the swing. Optical flow is a computer vision technique that involves tracking the movement of pixels between consecutive frames of a video. In the context of golf swings, optical flow can be used to estimate the velocity and acceleration of different parts of the body and the club, as well as to track the trajectory of the clubhead.

### Optical flow is the pattern of apparent motion between two consecutive frames as the result of objects or camera movement. More specifically a 2D vector where each vector is a displacement vector detailing the movement of points from frame to frame [4]. OpenCV provides two optical flow algorithms. Given the frames are cropped to a bounding box container only the golfer, we used the Dense Optical Flow algorithm, which computes the optical flow for all the points in the frame.

### To investigate the effectiveness of optical flow for reducing bias, videos are transformed to only the video’s optical flow capturing the appearance of invariant motion features only, inherently removing the effects of the background before inputting into the model architecture.

### The optical flow frames are preprocessed prior to model training and are input to the existing SwingNet model architecture during training. After processing the training videos to optical flow using OpenCV, the resulting videos appeared to have noisy frame data in the beginning and ending parts of the videos. Initially unsure if the noise would impact the training results, we processed an additional set of optical flow videos minimizing the computed magnitude, as opposed to the OpenCV example of normalizing, which resulted in output noise reduction (fig 4). Both sets of optical flow training videos were fed into the SwingNet CNN/RNN deep learning algorithm for training and performance results compared to the baseline SwingNet model.

# A picture containing crowd Description automatically generated No image Description automatically generated

Fig 4. Normalized magnitude calculations (left) and minimized magnitude calculations (right).

### Background Removal: Other preprocessing methods explored include explicitly removing background from the video before feeding to model. MediaPipe Pose performs human segmentation, effectively removing background from the swing videos by using the human silhouette as a mask. Such background removal could add significant value to reduce noise from swing videos taken throughout different sceneries on a golf course.

### The team used MediaPipe Pose human segmentation to remove background from the swing videos by using the human silhouette generated from the python API as a mask. Each video is preprocessed through a human segmentation pipeline and stored prior to model training. The human segmentation pipeline uses a Media Pose segmentation workflow with complexity mode set to 2, min detection confidence and min tracking confidence both set to 0.5, and a human segmentation smoother. The background is then blurred by converting the background intensity values to 10% of its original intensity value. The background removed videos were fed into the SwingNet CNN/RNN deep learning algorithm for training.

### 

Fig 5. Example golf swing video with background removed from sample of frames.

### A known potential downside of this approach is that the club shaft, club head, and ball will likely all be masked out of the video prior to model fitting. These are essential objects likely needing to be retained in the video for improving golf sequencing performance. The major limitation of the out-of-the-box MediaPipe Pose human segmentation algorithm is that it is not fine tuned to maintain these non-human objects.

### Adding Human Pose Estimation Features: Estimating the positions of key body joints such as the hips, shoulders, elbows, and wrists throughout the golf swing and passing these landmark coordinates into the existing SwingNet model architecture could improve event detection accuracy. The LSTM layer in the existing SwingNet model architecture is adjusted to take as input the output features from the MobileNetV2 layer concatenated with the 99 features (x,y,z coordinates for 33 joint landmarks).

### This slight adjustment in the architecture benefits from both the raw image and human pose landmark feature sequences. Intuitively, as the joint positions, such as the hands, move within the sequences of images, the LSTM can more effectively adjust the output probabilities according to identify the appropriate frames.

## Biomechanical Analysis

Fig 6. MediaPipe 33 Joint Feature Landmarks

### Realtime 2D pose estimation will be used to track the location of an amateur golfer and produce biomechanical insight from the golf swing sequences. Table 2 contains the biomechanical features captured for each golf swing event of interest. Analyzing and improving these features will lead to more power, accuracy, and consistency.

|  |  |
| --- | --- |
| Event | Biomechanical Features |
| Address | Feet placement and weight balance between them, ball position relative to the golfer's stance, angle and position of the spine [[8](https://docs.google.com/document/d/1tPX0IUMb-I3s5RZKEuG789wlfmtvT0XD/edit#bookmark=id.5aja9791apkg)] and hips [[7](https://docs.google.com/document/d/1tPX0IUMb-I3s5RZKEuG789wlfmtvT0XD/edit#bookmark=id.f70mmbs3ldxi)], grip and hand position on club |
| Toe-up | Degree of wrist hinge, club position relative to body |
| Mid-backswing | Degree of wrist hinge, club position relative to body, hip, shoulder, and spine angle [[8](https://docs.google.com/document/d/1tPX0IUMb-I3s5RZKEuG789wlfmtvT0XD/edit#bookmark=id.5aja9791apkg)] and position, hips and torso rotation, tempo of backswing (start to middle to end of backswing) [[5](https://docs.google.com/document/d/1tPX0IUMb-I3s5RZKEuG789wlfmtvT0XD/edit#bookmark=id.9vgw80hfids1)] |
| Top | Club position relative to body, club angle, hip translation [[7](https://docs.google.com/document/d/1tPX0IUMb-I3s5RZKEuG789wlfmtvT0XD/edit#bookmark=id.f70mmbs3ldxi)], shoulder tilt [[6](https://docs.google.com/document/d/1tPX0IUMb-I3s5RZKEuG789wlfmtvT0XD/edit#bookmark=id.jix7nwcnyb5v)], spine angle [[8](https://docs.google.com/document/d/1tPX0IUMb-I3s5RZKEuG789wlfmtvT0XD/edit#bookmark=id.5aja9791apkg)], hips and torso rotation |
| Mid-downswing | Degree of wrist hinge, club position relative to body, hip, shoulder, and spine angle [[8](https://docs.google.com/document/d/1tPX0IUMb-I3s5RZKEuG789wlfmtvT0XD/edit#bookmark=id.5aja9791apkg)] and position, hips and torso rotation |
| Impact | Club position relative to body, degree of shaft lean, hip, shoulder, and spine angle [[8](https://docs.google.com/document/d/1tPX0IUMb-I3s5RZKEuG789wlfmtvT0XD/edit#bookmark=id.5aja9791apkg)] and position, hips and torso rotation, compare to address [[7](https://docs.google.com/document/d/1tPX0IUMb-I3s5RZKEuG789wlfmtvT0XD/edit#bookmark=id.f70mmbs3ldxi)] |
| Mid-follow-through | Club position relative to body, hip, shoulder, and spine angle [[8](https://docs.google.com/document/d/1tPX0IUMb-I3s5RZKEuG789wlfmtvT0XD/edit#bookmark=id.5aja9791apkg)] and position, hips and torso rotation |
| Finish | Club position relative to body, hip, shoulder, and spine angle [[8](https://docs.google.com/document/d/1tPX0IUMb-I3s5RZKEuG789wlfmtvT0XD/edit#bookmark=id.5aja9791apkg)] and position, hips and torso rotation |

Table 2. Biomechanical features captured for each golf swing event of interest.

### After capturing the frames of interest and biomechanical features, we compare the swing biomechanics to previous user golf shots and/or to professional golfers using the GolfDB dataset with side-by-side visuals and displayed biomechanical features.

### The biomechanical features consisted of calculating angles, translations, and one feature that involved time (measured by number of frames). The translations were very straightforward, simply calculating the difference in the y-coordinate, x-coordinate, or a Euclidean distance involving both. Calculating angles proved more difficult. Mediapipe’s human pose software finds three dimensional coordinates. To calculate the angle between two vectors, one needs to fund the dot product and divide by the product of the norms. Then find the arc cosine of this number to get the angle in radians. We then convert it from radians to degrees. The result of this on the right elbow is shown below, where the vectors would be from the elbow to the shoulder and the elbow to the wrist.



Fig 7. Right elbow angle at address using 3-dimensional vectors.

### The other method we used was taken from a Github repository [14]. This method only used the y-coordinate and x-coordinate and the Numpy arctan2() function. The equation was:

np.arctan2(c[1]- b[1], c[0] - b[0]) - np.arctan2(a[1] - b[1], a[0] - b[0])

### where a represents the first landmark point (e.g. the shoulder), b represents the middle (elbow), and c is the last point (wrist). The indexes 0 and 1 represent the x-coordinate and y-coordinate, respectively. We then converted this to degrees, as this equation outputs radians. We also took an extra step, also taken from the repository, and if the angle were over 180 degrees, we subtracted it from 360. The result of this on the right elbow angle from the same frame as above is shown below.



Fig 8. Right elbow angle at address only 2-dimensional coordinates.

We can see that the second method is much more accurate, even with such a simple case where the whole joint is in view of the camera.

The other method to be discussed is how we had to calculate the shoulder and hip rotation. Since we could not rely on the *z*-coordinate, we had to find another way to simulate the rotation. We calculated the Euclidean distance between the shoulders (hips) for each frame, and either divided it by the distance at address if the camera angle were behind or divided the address distance by the frame distance if the camera angle were open. If the ratio was 1, there was no rotation. If the ratio were greater, there was more rotation.

We formatted the code to be able to either compare two videos or just analyze one video, given the user input. The user also inputs a biomechanical feature they want to display.

# Results

## Golf Swing Sequencing

### In this study, we explored three methods to improve the existing SwingNet baseline hybrid CNN/RNN deep learning algorithm to sequence and identify eight key frames within golf swings. The first two methods preprocessed the swing videos applying optical flow and background removal. We aimed to evaluate if these methods could reduce video bias and improve the accuracy of golf sequencing. The third method involved appending human pose landmark features to the output of the CNN prior to feeding the LSTM. We aimed to improve accuracy from this approach by adding explicit human joint features to the model.

### The model accuracy was measured by the percent of correct events within a tolerance. We evaluated the updated model to detect eight golf swing events at an average PCE rate better than the baseline model of 76.1% and six out of eight events at a rate of 91.8%. To estimate the model PCE, the data were randomly split into four training sets and four validation sets. The average PCE across the first split only is reported for the best model.

Table

Description automatically generated

Fig 7. Validation Split 1 Out-of-sample Percent of Correct Events (PCE) by Model. Optical Flow is the best performing model.

### Only the optical flow modeling methods improve the model's accuracy for golf sequencing. Optical flow reduced bias in the videos by capturing the appearance of invariant motion features only, inherently removing the effects of the background before inputting into the model architecture. The background removal method reduced noise from swing videos effectively removing the background, however, it also removed the club and ball throughout the video leading to its inability to outperform the baseline. Additionally, the human pose feature additions to the third proposed model failed to add any value in addition to the raw video input. This is likely due to the low confidence of the x, y, and z coordinate locations identified from the human pose estimation across videos. More details on this deficiency is outlined in the next section.

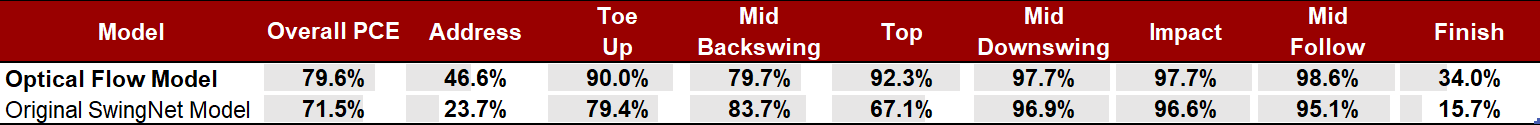
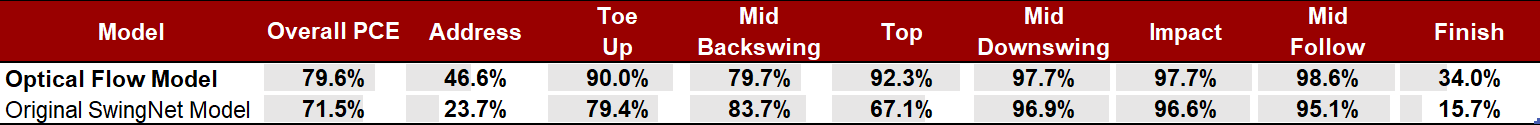
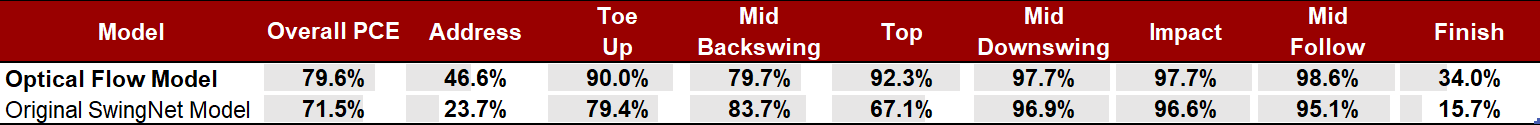
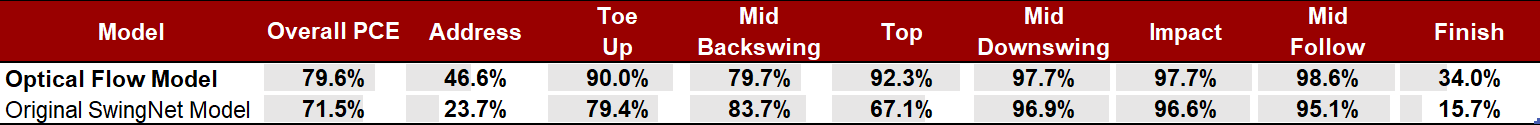


Fig 8. Validation Split 1 Event Out-of-sample Percent of Correct Events (PCE) by Model.

### The evaluation of the updated optical flow model showed the average PCE rate for detecting eight golf swing events at 79.5% on the first validation split, superior to the reported

### A picture containing text, grass, outdoor, tree Description automatically generatedbaseline model's rate of 71.5%. For detecting the middle six out of eight events, the rate was 92.7%, better than the baseline model's rate of 86.5%. The largest performance gains training on the optical flow videos are represented in substantial PCE increases identifying the Address and Finish frames, a notable deficiency of the original SwingNet model. Our model PCE (on validation split 1) for Address and Finish frames are 46.6% and 34.0% respectively, compared to 23.7% and 15.7% respectively from the original SwingNet model. These results demonstrate that optical flow video preprocessing significantly improves golf swing sequencing accuracy.

Fig 9. Golf Swing Sequence Event Results with Confidence Levels via Baseline SwingNet (top) and Optical Flow SwingNet (bottom) on 191.mp4

Fig 10. Golf Swing Sequence Event Results via Optical Flow SwingNet on 1080p/30fps mobile video (author Joel Klein)

### These results are verified when comparing prediction event confidence levels for out-of-sample images across models (fig 9). Although each model effectively identifies the important event frames, there is a significantly higher confidence in the Address prediction for the optical flow model.

### In addition to testing the model on other out-of-sample videos within GolfDB, the model was tested on our swing videos recorded directly from an iPhone 12 at 1080p and 30 frames per second. The optical flow model correctly predicted eight swing events within tolerance.

### Table Description automatically generatedAlso noteworthy, is the limited bias in out-of-sample swing sequencing performance across golfer sex, club type, camera angle, and video speed. The PCE across different levels within each covariate are within a couple percentage points.

Fig 11. Validation Split 1 Optical Flow Model Out-of-sample Percent of Correct Events (PCE) by sex, view, club, video speed

### Our study shows that optical flow video preprocessing effectively reduces video bias and improves the accuracy of golf sequencing over the baseline SwingNet model proposed by McNally et. al. These findings could have significant implications for accessible biomechanical analysis in the realm of golf, enabling improved coaching through player swing analysis with a simple video recording.

## Biomechanical Analysis

### The resulting output of the biomechanic analysis is shown below.

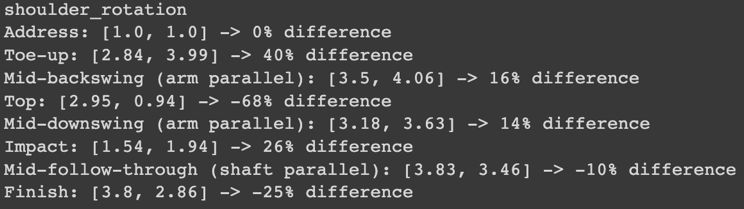


Fig 12. Output of biomechanic analysis.

We can see the results from the first video, the second video, and the percent difference from the first to the second video. The above example is for the shoulder rotation feature, and we can see how the scale works for the quasi-rotation metric.

# Discussion

## Golf Swing Sequencing

### The results of this research show that the use of computer vision and deep learning techniques can improve golf swing analysis via golf swing sequencing. By identifying the key events and extracting biomechanical features, amateur golfers can improve their swings by comparing their swings to those of professional golfers and their previous swings.

### The primary new method of bias correction of videos using optical flow video transformation prior to fitting our model resulted in significant increases in swing sequencing performance. The resulting performance gains support the initial hypothesis that removing noise within the videos would reduce error in predicting the correct frames. Due to this simple video processing transformation, our model is less constrained by the external scenery surrounding the golfer within the video. This altercation increases the model’s ability to perform swing sequencing on a wide range of input swing videos.

### Although optical flow produces better swing sequence predictions, there are some limitations to this approach. The model still only identifies the Address and Finish frames with less than 50% PCE on unseen videos. These events are close to the start and end frames within the videos respectively. The study only tests one deep learning algorithm, SwingNet, which may not be the most effective algorithm for golf swing sequencing. For example, for training image sequences including the end frame, the current SwingNet architecture simply appends the start of the video to the end completing the sequence length of 64 image frames before passing into the RNN. This decision likely significantly impacts the model quality because these start and end frames significantly deviate. Future research should explore other deep learning algorithms. Additionally, the data set only contains 1400 videos. Use of larger datasets for training and evaluation ensures the generalizability of the findings, especially in predicting the Address and Finish frames.

### Several future opportunities exist to improve model performance and make its functionality available to end users. We first propose updating the model on videos shot from most modern smartphones with higher resolution and a higher frame rate. We hypothesize that model performance trained on higher quality videos would significantly increase. In addition, this updated video quality would likely improve the proposed biomechanical feature analysis via human pose estimation lacking from the low-resolution videos in GolfDB. After updating the model with new data, there is potential, based on performance, to serve the model as an endpoint within a mobile application and provide an initial public offering to amateur golfers.

## Biomechanical Analysis

The biomechanics analysis proved difficult since the *z*-coordinate was unreliable. Our calculations were then subsequently unreliable since we could not utilize all of the data available. This is especially true for the rotation measurements. This section could be greatly and easily improved by having more reliable 3-dimensional coordinates. This would be left for future work, and possibly finding new software other than Mediapipe.

# Conclusion

In this project, we were able to improve the golf swing sequencing done by McNally [9]. We found the frames of important swing checkpoints with better accuracy than McNally. Using these frames, we applied Mediapipe’s human pose software to gather body part landmark coordinates. We then used these coordinates to calculate various biomechanical features with which a user can dissect their own swing and even compare with another video’s swing.

Future work would include using software that better calculates 3-dimensional coordinates so that our calculations of the biomechanical features become more accurate. Our hope would be to eventually provide a user-friendly application with which a user can upload their videos and the application would output the biomechanical analysis. This could be an incredibly useful application for golfers and golf coaches. It can save immense amounts of time going through video and even detect features that the golfers would have no hope to find themselves. The idea behind this program was to create an easy and accessible application that anybody with a camera can use to improve their golf swing. While it still has some improvements to make, this application is already able to perform its duties.

# References

[1] National Golf Foundation. n.d. “Golf Industry Facts.” National Golf Foundation. Accessed April 2, 2023. <https://www.ngf.org/golf-industry-research/>.

[2] Matuszewski, Erik. 2018. “The Economic Impact Of Golf: $84 Billion In The U.S.” Forbes. <https://www.forbes.com/sites/erikmatuszewski/2018/04/24/the-economic-impact-of-golf-84-billion-in-the-u-s/?sh=42f1d1846eec>.

[3] Schiffman, Roger. 2011. “How to practice like a pro | This is the Loop.” Golf Digest. <https://www.golfdigest.com/story/how-to-practice-like-a-pro>.

[4] Open Source Computer Vision Library. n.d. “Optical Flow.” OpenCV. Accessed April 4, 2023. <https://docs.opencv.org/3.4/d4/dee/tutorial_optical_flow.html>.

[5] 8AM Golf. 2020. “Jack Nicklaus explains why tempo is the hidden key to the golf swing.” golf.com. <https://golf.com/instruction/flashback-jack-tempo-hidden-key-golf-swing/>.

[6] GOLFTEC Digital. 2019. “3 keys to create a good golf swing.” The GOLFTEC Scramble. <https://scramble.golftec.com/blog/2019/03/3-keys-to-create-a-good-golf-swing/>.

[7] Robinson, Ted. n.d. “Top 5 secrets to a good golf swing – Chimera Golf Club.” Chimera Golf Club. Accessed April 2, 2023. <https://chimeragolfclub.com/top-5-secrets-to-a-good-golf-swing/>.

[8] Miller, Les. 2009. “Learn golf swing basics by breaking it up into three parts.” Golf Instruction. <http://www.golfinstruction.com/golf-instruction/quick-tips/golf-swing-drill-three-parts-backswing-impact-position-follow-through-10046.htm>.

[9] McNally, William, et al. "Golfdb: A video database for golf swing sequencing." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019.

[10] Gehrig, Nicolas, Vincent Lepetit, and Pascal Fua. "Visual golf club tracking for enhanced swing analysis." Proceedings of the British Machine Conference, 2003.

[11] Urtasun, R., Fleet, D. J., & Fua, P. (2005, June). Monocular 3D tracking of the golf swing. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) (Vol. 2, pp. 932-938). IEEE.

[12] Kim, T. T., Zohdy, M. A., & Barker, M. P. (2020). Applying pose estimation to predict amateur golf swing performance using edge processing. IEEE Access, 8, 143769-143776.

[13] Liao, Chen-Chieh, Dong-Hyun Hwang, and Hideki Koike. "AI Golf: Golf Swing Analysis Tool for Self-Training." IEEE Access 10 (2022): 106286-10629