

Computer Vision: Golf Pro

Parker Sullins



Abstract

In this paper, we attempt to rectify many of the pitfalls that come with learning golf. Using computer vision and machine learning, we aimed to produce full 3D tracking of key skeletal points with the intent of providing insights on how a beginner golfer compares to a pro on key metrics. For this study, we focused solely on the driver, the longest club in a player's arsenal, and the most used club. Using pre-trained machine learning models, we tracked 33 key points, each generating four features, an X, Y, Z coordinate tuple, and a visibility metric [4]. This generated 133 features to extract per frame. We also attempted to normalize the total frames of a swing so that each swing took approximately three seconds and was composed of 180 frames. Using the 133 key point features and the current frame feature, we were able to accurately predict whether a golfer was an amateur or a pro with 99% accuracy on testing data and 80% accuracy on validation sets. The models we tested were Linear Regression, Gradient Boosting Classifier, and Random Forest. Random Forest performed the best and was selected as the preferred model for classifying golfers as either Pro or Amateur. The goal of this research was to explore whether a machine learning model could successfully identify a highly skilled player's swing versus a low-skill player's swing and if so, understand why.

1 Introduction

Currently, golf is a highly prohibitive sport to newcomers. Beginners must shell out hundreds of dollars for clubs and equipment, and that is all before having ever swung a golf club. As most beginners soon find out, golf is not only expensive but requires a large amount of time and effort to obtain reasonable competency. This issue is only worsened by the nature of golf as an athletic venture. Often the only feedback a player receives is watching the flight of the ball after impact. It often takes weeks for a beginner to even begin striking a ball consistently, let alone possessing any accuracy. Below is a typical learning curve one could expect when attempting to

gain skill in an area, and to the right is the typical learning curve one might experience when learning golf.

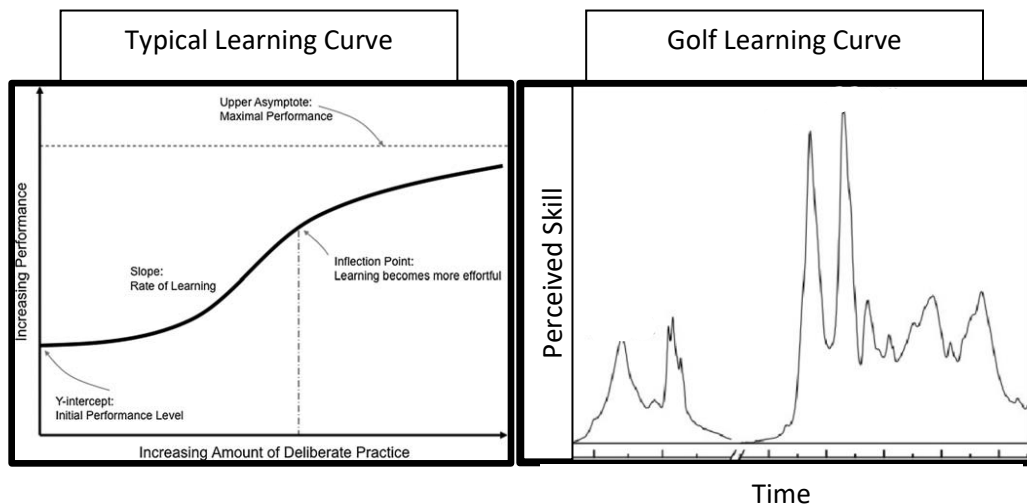


Figure 1: Typical learning curve for skill acquisition

Golf is an atypical sport, and the learning curve reflects this. The learning curve of golf is steep, unpredictable, and highly discouraging. The most common method to reduce early friction when learning golf is to hire an instructor. The cost of an instructor typically ranges from \$75 to \$150, and players can only expect results after numerous lessons. We hope to change the seemingly insurmountable challenges of learning golf for the layman. We hope to be able to provide personal insights into mistakes golfers are making so they may correct their swing without the assistance of a trainer. Ideally, players would be able to record their swings from their smartphones, upload them through a mobile app, and receive feedback on their swings after the video is processed on a cloud server. This would greatly reduce the cost of learning golf while providing exceptional feedback that could only otherwise be accessible by paying for a coach.

2 Team Organization

Team organization consisted of solo research performed by Parker Sullins. Additionally, several amateur golfers volunteered their time, and without them, this would not have been possible. Also, special thanks to Taylor Made for providing excellent footage of professional golfers.

3 Related Work

Currently, single-camera swing analysis is in the research phase, and no current consumer products can provide computer-vision-based motion tracking feedback on a player's body orientation relative to the golf ball. However, sensor-based three-dimensional tracking systems are available. Currently, sensor-based tracking consists of attaching sensors to locations throughout the body and the golf club. Using multiple high-speed cameras, they can track player movement exceptionally well using motion-tracking technology developed for filmmaking.[2] The drawbacks of this system are that a user must attach numerous tracking nodes which can be cumbersome when performing athletic movements. Additionally, they require several high-speed specialty cameras and require a studio that must be professionally set up to make the environment acceptable for the systems software to run correctly. These systems range from tens of thousands of dollars on the low side to upward of one hundred thousand dollars. This means such systems are practically impossible to obtain by the average golfer.



Figure 2: The OPTITRACK'S GEARS GOLF SYSTEM

Notice the multiple highspeed cameras in red and motion tracking bodysuit, hat, and shoes in green.

There also exist flight tracking systems that have been on the market for years and perform competently when simulating the flight path and distance of a golf ball after impact.[1] These systems range from hundreds to thousands of dollars and are more obtainable to the average golfer. However, these systems fail to address the underlying issue that beginner golfers face. These systems do not provide feedback on swing form or tell a user how to correct their swing.

Academic Research in this field is ongoing, and several studies exist which try to solve a similar problem using machine learning and computer vision. These papers use computer vision to classify the golf swing into various states and to identify which state a golfer is in throughout a swing.[7] The goal of this is to see if machine learning can correctly identify the key sequences in a golf swing. This research is not geared at helping golfers and is only concerned with motion tracking of the golf sequence, and still makes use of a motion capture suit. Additionally, several papers attempt to use inertial sensors to measure physical characteristics of a golfer's swing, such as angular velocity and gyroscope alignment of a golfer's body throughout the swing [5]. These papers were successful at motion tracking and digital signal processing of the golf swing, but they were unsuccessful at understanding what makes a good versus bad golf swing, and do not provide feedback to a player. Currently, this is the only research into golf motion tracking using only computer vision, without the use of a motion-capture suit.

4 Data Acquisition and Processing

4.1 Data Collection

In this paper, golf swing data was collected from various volunteers. Firstly, we collected swing data from amateur golfers who have some experience with golf but who have never had formal instruction. Data collection from the amateur golfers consisted of a single iPhone 10X camera, utilizing 60 fps in MOV format. This format is commonly used in smartphones as it allows for better compression and video quality when compared to MP4. Videos were taken approximately 17 feet behind the golfers and slightly off-center to the right. This was done to best match the professional footage of pro golfers produced by Taylor Made's golfing YouTube channel. Twelve amateur swings were recorded from three different golfers. Additionally,

twelve professional swings were downloaded as MP4's and used as the data for the professional swing. All videos were converted to MP4's for compatibility with the OpenCV library. Videos were slowed or sped up so that all swings took approximately three seconds to complete. The videos were then exported at 60 frames per second so that each video was segmented into roughly 180 frames each. All videos were then fit to a 9:16 ration, and golfers were placed squarely in the middle, with a similar distance from the golfer to the edges for all videos. Data was collected on each key point generated by Google's MediaPipe model, a machine learning library for pose estimation, and mapped to the appropriate frame. Data generated by MediaPipe's model was then fit with a supervised Random Forest Classifier. The classifier optimized a solution for a binary dependent variable, Professional or Amateur golfer, and reported a probability for that video being of a pro or amateur.

4.2 Data Description

Data was collected from thirty-three key point body segments and mapped to the frame in which they occurred. Each point was composed of a four-part tuple, an X, Y, and Z geometric coordinate, and a visibility estimation denoted V. The X and Y coordinates were generated relative to the image dimensions and the Z coordinate is estimated by the model. Visibility represents the confidence in which the model feels it can accurately pinpoint where a certain segment on the body is at each frame.

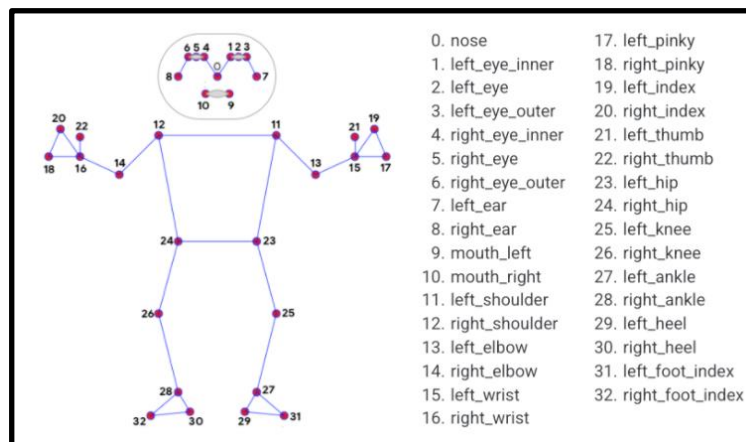
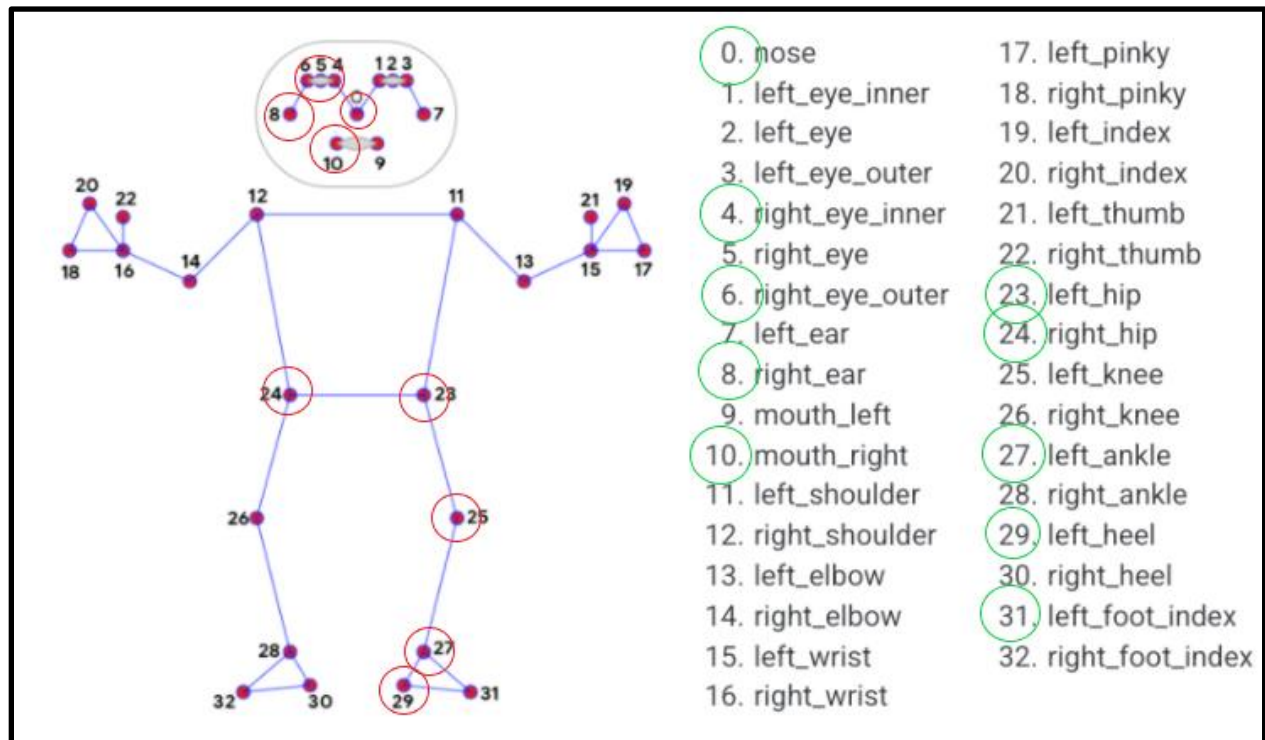


Figure 3: A diagram showing where each feature maps to in the MediaPipe Pose Model

We collected over 3,000 rows of data for each of the 134 features, generating over 1,000,000 data points.

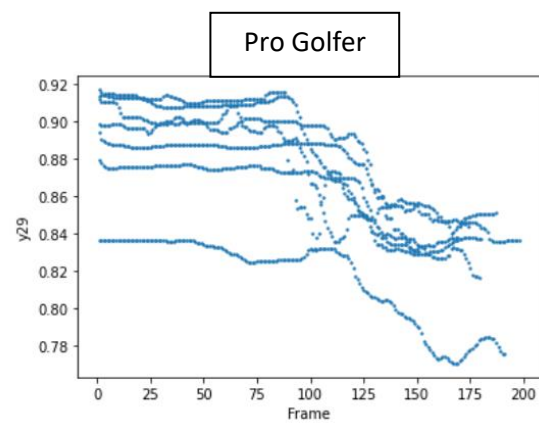
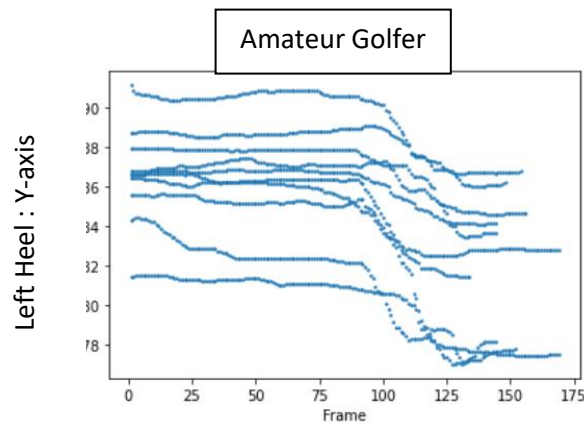
4.3 Visualization

Data visualization was used to explore relationships between various features and how they were classified by the Random Forest model. Feature Importance was calculated and used to determine which features to explore based on their relative weight in the model. Below is an illustration of the 10 key points with the highest contribution to the classification of a swing.



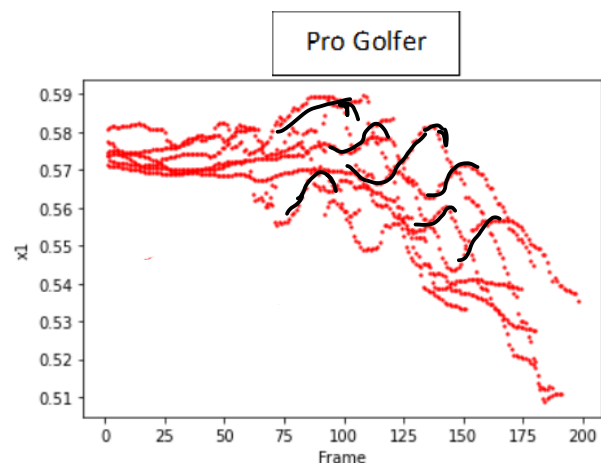
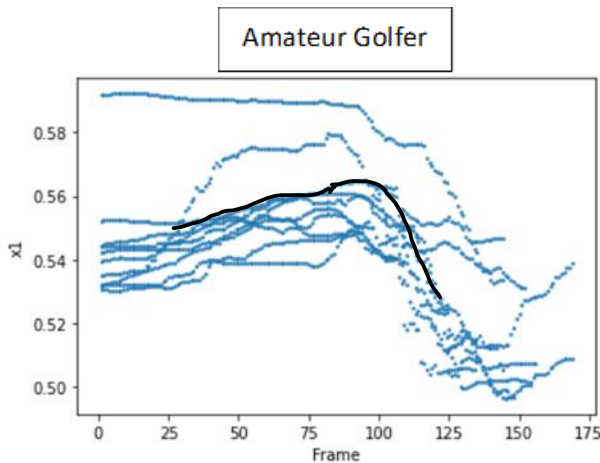
The results of the feature importance analysis are somewhat unexpected. For example, no feature related to the upper body is recorded until the 40th most important feature, and no upper body feature contributed more to the model than its relative size in the feature set, i.e., 1:134. The most important feature was the left heel, which is unexpected as the left heel is typically stationary throughout the golf swing with little to no movement.

Computer Vision: Golf Pro



Keep in mind that movement is calculated relative to the joint's position in the 1020px by 720px frame, and thus the physical coordinates are not as important as the pattern generated by the movement. Here we have the left heel's change in the y-axis throughout the swing. Interestingly, I do not see anything that seems to be a significant difference between the pro and amateur. Yet, this feature contributes over 6% to the model's decision-making.

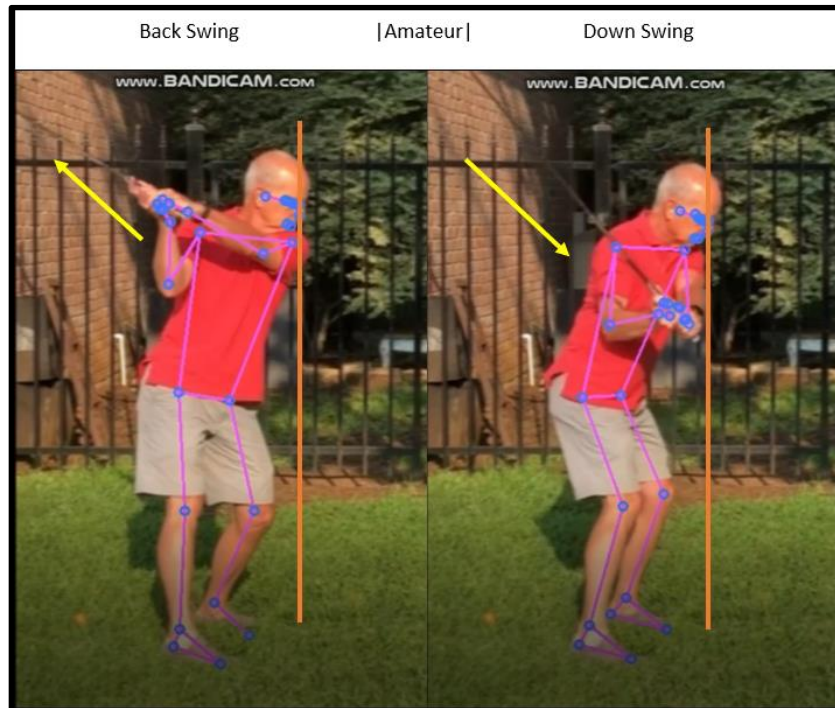
Nose Movement relative to the X-axis



In the above illustration, we have movement of the nose relative to the x-axis. This is the second most important feature, and while the patterns are subtle, their differences become obvious when comparing the data to the videos of the golfers themselves. In this data, we can see the markings of what is often called the golfers squat. It is typically only present in highly skilled players as it is difficult to do well and will lead to inaccurate shots if the golfer does not possess excellent coordination. At the top of the backswing, a pro golfer will begin his downswing with a subtle squatting motion, forcing the nose in the positive x-direction for a moment. This hump is obvious in the Pro golfers but is much less apparent in the Amateur golfers' swing.



Notice the Pro's subtle head movement in the positive X direction during the golfer's squat.



The amateur's nose never moves positively in the X-direction.

Often an important distinction between amateurs and pros is the golfer's squat [1]. Beginner golfers are taught to keep their heads perfectly still to minimize room for errors. However, pros utilize the golfers' squat to generate power, and after years of practice, their coordination allows them to be accurate and powerful at the same time.

Interestingly, the third most heavily weighted feature is the left knee. In Pro players, the left knee moves precipitously on the negative X-axis as the hips explode into the downswing, before ever impacting the ball. Amateur players have a much slower hip rotation, and thus the movement of the left knee is more gradual. Notice how the pro player has not yet contacted the ball when his lower body is in the same phase as the amateur. This is a prime example of the adage "swing with the hips, not the arms", that you often hear from professional coaching [1].

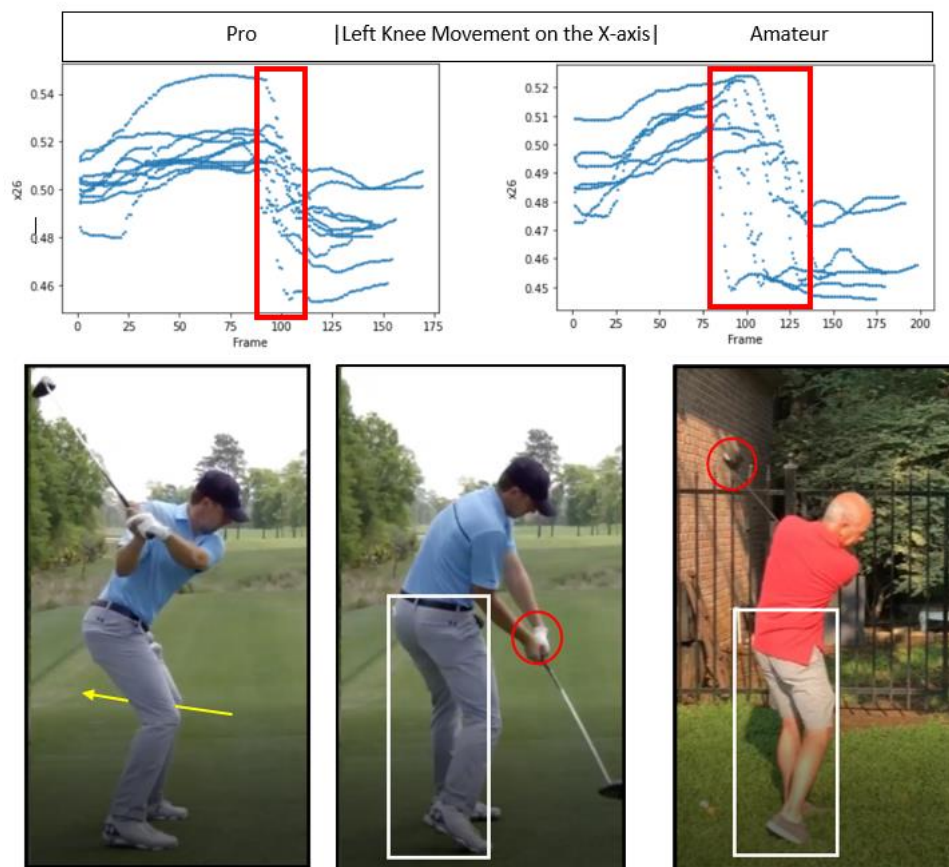


Fig 8: Notice how both golfers' hips are in the exact same position, but the pro has not yet made impact with the ball, as his hips lead his hand during the downswing.

In theory, every feature could be rigorously deconstructed, illuminating the differences in the pro vs amateur golf swing, at each key point throughout the swing. More work is needed to truly understand the complex relationships between key points so we can better provide helpful feedback to the user.

5 Software and Developing Tools

5.1 Software

This application was built using the PyCharm IDE. The most crucial software in the application is Google's MediaPipe machine learning library. This library provides great tracking of pose coordinates, which were used to build the classification model. The libraries used in this application include Pipenv, Pandas, Numpy, Scikit-learn, Matplotlib, Pickle, and Open CV [8]. Pipenv is a dependency management software, similar to a virtual environment. Open CV was used to process the video footage frame by frame. Pandas and Numpy were used for data engineering. Matplotlib was used for the exploration and visualization of the data. ScikitLearn was used to train the models and build the machine learning pipeline. Pickle was used to extract the model into a separate, reusable file. Additionally, free online software, cloudconvert.com, was used to convert MOV files into MP4s, and mp4s.org was used to convert YouTube clips into MP4s. HD Movie Maker was used to perform video editing, such as fps adjustments, video trimming, and cropping into a 9:16 window ratio.

5.2 Computer Setup

This research was conducted on a Sager NP8358F2, utilizing an Intel i7-10875H CPU, an RTX 2070 Super 8GB GPU, 32GB RAM, and a Windows 10 operating system.

5.3 Sensor Hardware

An iPhone 10X was used to capture video footage. This was the only sensor used to capture footage. The footage was shot at 60fps.

Dual Lens	<ul style="list-style-type: none"> 12 MP, f/1.8, 28mm (wide), 1/3", 1.22µm, dual pixel PDAF, OIS 12 MP, f/2.4, 52mm (telephoto), 1/3.4", 1.0µm, PDAF, OIS, 2x optical zoom
Features	Quad-LED dual-tone flash, HDR (photo/panorama), panorama, HDR
Video	4K@24/30/60fps, 1080p@30/60/120/240fps

6 Milestones

List of Milestones

Week 7-19 to 7-26:	<ul style="list-style-type: none"> • Conduct sample footage – approx. 50 swings • Create a working model using test and training data. • Data Exploration
Week 7-26-7-31:	<ul style="list-style-type: none"> • Integrate Model into hardware • Optimize model to run efficiently.
Final Week:	<ul style="list-style-type: none"> • Final optimizations and finish documentation.

7 Results and Discussion

After extracting data from the MediaPipe pose model, and doing data cleaning and exploration, we set out on constructing a working model. Four models were chosen to be trained on the training data. We split the data frame into 70% training data and the remaining 30% was testing data. We also retained two pro swings and three amateur swings which would not be present in the training or test data and would be used for validation. The four models we selected for training were logistic regression, ridge classifier, random forest classifier, and a gradient boosting classifier. All models performed exceptionally well on both training and testing data. With accuracy scores close to 100%.

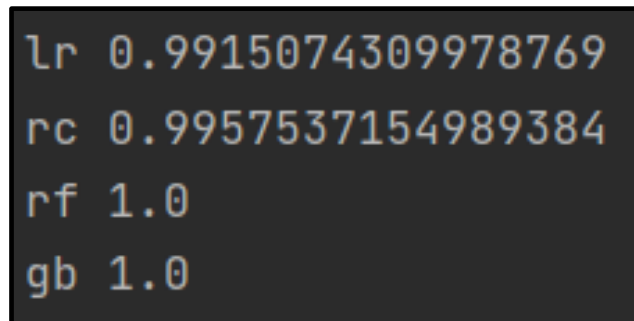
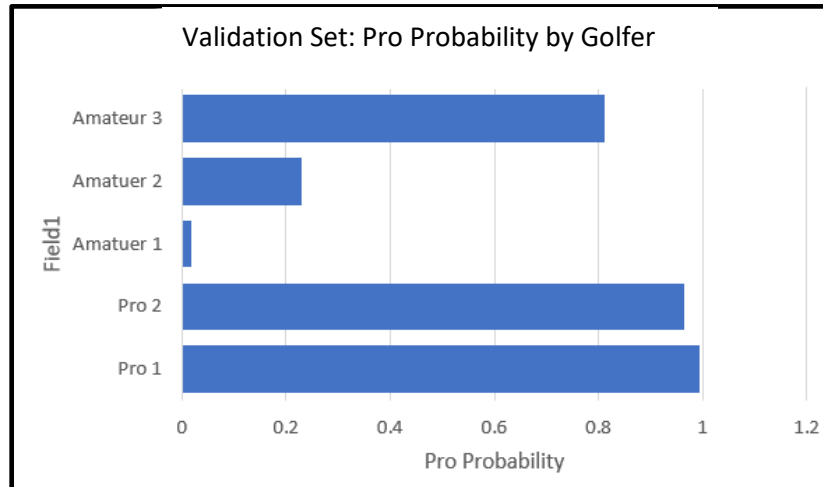


Figure 5: Test accuracy for Linear regression, ridge classifier, random forest, and gradient boosting algorithms

They performed so well it was a concern that we may be overfitting. However, the validation set exposed that only the random forest classifier model performed exceptionally well on completely novel data. Other models had a lower than 60% accuracy rating on the validation set, while the random forest model had an 80% accuracy. The most inexperienced amateur

received a 1% probability of being a pro. The second most experienced amateur golfer received a 23% probability of being a pro. The most experienced amateur, who often shoots below 90 on an 18-hole course, received an 81% probability of being a pro. All pro golfers received a greater than 97% probability of being a pro.



The final amateur exposed some weaknesses in the model. The most skilled amateur demonstrated two hallmarks of a pro that the model heavily weighs in its decision. Firstly, he demonstrated the golfers squat, and his nose moved positively on the x-axis during the downswing. Secondly, he had explosive hip movements, leading to the rapid movement of the left knee joint in the negative x-direction. However, because the model does not heavily weigh in upper body key points, it misclassified this golfer. The weakly weighted upper body key points are most likely due to the inconsistent tracking of the upper body, as the upper body joints are smaller, and move the fastest and largest distance throughout the swing.



Fig: 6: Skilled amateur demonstrating hip rotation leading the hands



on the left. A pro demonstrates the same positioning on the right.

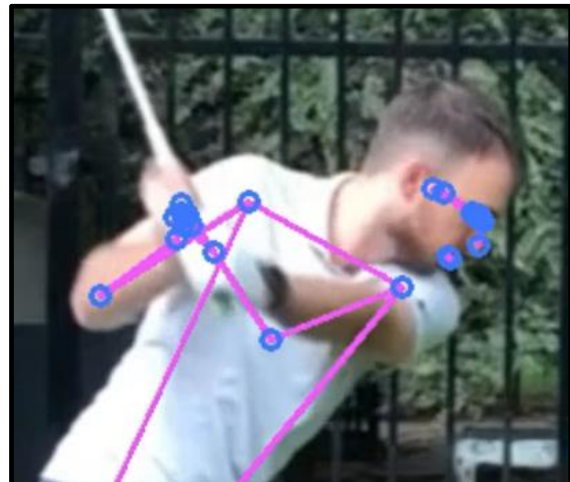


Fig 7: Example of poor upper body tracking.

8 Conclusion

In this paper, we performed a 3D golf swing analysis based on a video sequence of pro and amateur golfers. Using only a single camera, and no tracking aids aiding the computer vision model, we attempted to classify golfers based on their relative skill level. The experiment used over 3000 frames of image sequences and extracted 134 features from each frame. Using the extracted data, we trained multiple models and selected the Random Forest Model. The model performed exceptionally well on the validation set. Having an out-of-bag error rate of only 20%. The successful result of this research has opened the door for future work on single-camera, computer vision-based approaches to a self-coaching golf system. With further research, data collection, and model tuning, a complete self-coaching system could be deployed to mobile devices at an affordable price for beginner golfers.

9 Future Work

The future of this technology is exciting, and the advancements of computer vision and hardware will inevitably lead to a sophisticated and highly effective self-coaching golf system using only a single camera. In this experiment alone, we were unable to explore everything we set out to. We created the code to calculate angles between various joints but did not have the necessary time to optimize and deploy a model with the additional features. Extracting data from the angles formed by the joints could provide new insights into what separates a pro from an amateur. Additionally, building a custom pose model, focused solely on tracking golf movements, could lead to a more robust and accurate tracking of the key points, specifically the upper body segments. All of this requires further research to explore and will eventually culminate into a mature version of what was demonstrated in this paper.

10 References

- 1) “Golf Info GUIDE.” *Golf Info Guide*, 6 Mar. 2015, golf-info-guide.com/golf-tips/senior-golf-tips/reasons-and-benefits-golf-downswing-squat-senior-golf-tip/.
- 2) “Golf swing-analysis software.” *Gears Sports*, www.gearssports.com/golf/.
- 3) *Mediapipe*, google.github.io/mediapipe/.
- 4) “How to Build Your First Image Recognition Classifier with IBM Watson Visual Recognition.” *Nicholas Renotte*, 29 Jan. 2019, www.nicholasrenotte.com/how-to-build-your-first-image-recognition-classifier-with-ibm-watson-visual-recognition/.
- 5) Hsu, Yu-Liang, et al. “Golf Swing Motion Detection Using an Inertial-Sensor-Based Portable Instrument.” *2016 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, 2016, doi:10.1109/icce-tw.2016.7521016.
- 6) Huang SY, Kuo KP, Lin YH (2015) A golf swing analysis system using wii balance board and Kinect sensors for novice players. *Multimed Tools Appl* 74(23):10679–10696
- 7) Ko, Kyeong-Ri, and Sung Bum Pan. “CNN and Bi-LSTM Based 3D Golf Swing Analysis by Frontal Swing Sequence Images.” *Multimedia Tools and Applications*, vol. 80, no. 6, 2020, pp. 8957–8972., doi:10.1007/s11042-020-10096-0.
- 8) *Scikit*, scikit-learn.org/stable/index.html.
- 9) Yıldırım, Soner. “A Beginner's Guide to Supervised Machine Learning Algorithms.” *Medium*, Towards Data Science, 26 July 2020, towardsdatascience.com/a-beginners-guide-to-supervised-machine-learning-algorithms-6e7cd9f177d5.