Modeling Golf Player Skill Using Machine Learning

Rikard König, Ulf Johansson, Maria Riveiro, Peter Brattberg

Modeling Golf Player Skill using Machine Learning

Rikard König[1], Ulf Johansson[1], Maria Riveiro[2] and Peter Brattberg[1]
[1] University of Borås, Borås, Sweden,
[2] University of Skövde, Skövde, Sweden
rikard.konig@hb.se, ulf.johansson@hb.se,
maria.riveiro@his.se, peter.brattberg@hb.se

Abstract. In this study we apply machine learning techniques to Modeling Golf Player Skill using a dataset consisting of 277 golfers. The dataset includes 28 quantitative metrics, related to the club head at impact and ball flight, captured using a Doppler-radar. For modeling, cost-sensitive decision trees and random forest are used to discern between less skilled players and very good ones, i.e., Hackers and Pros. The results show that both random forest and decision trees achieve high predictive accuracy, with regards to true positive rate, accuracy and area under the ROC-curve. A detailed interpretation of the decision trees shows that they concur with modern swing theory, e.g., consistency is very important, while face angle, club path and dynamic loft are the most important evaluated swing factors, when discerning between Hackers and Pros. Most of the Hackers could be identified by a rather large deviation in one of these values compared to the Pros. Hackers, which had less variation in these aspects of the swing, could instead be identified by a steeper swing plane and a lower club speed. The importance of the swing plane is an interesting finding, since it was not expected and is not easy to explain.

Keywords: Classification, Decision Trees, Machine Learning, Golf, Swing analysis.

1 Introduction

Golf is a major sport that is today played by over 60 million people all over the [1]. The golf swing of a pro may look simple, but it is a combination of several complex biomechanical motions that need to be performed at high speeds with high accuracy. Hence, both pros and amateurs spend considerable time on perfecting their swings. However, due to the very complex chain of movements, the help of a teaching professional is most often needed to identify problems in a swing.

Golf instructors require a deep theoretical knowledge about golf swings and a lot of first hand teaching experience to be able to teach at a high level. Naturally, finding the problem with a swing is a crucial skill that must be mastered. Since the golf swing is such a complex movement and since the club head moves at great speeds, golf swing analysis has long been an art requiring very sharp eyes. When determining the

effectiveness of a swing, many teaching professionals often start by observing the ball flight and thereafter the club and body motion [2].

Lately, new technology, such as the TrackMan Launch Monitor Radar (TM) [3], has made it possible to measure numerous characteristics of the golf swing quantitatively. TM units use a Doppler-radar to register information about the club head at the point of impact with the ball and the trajectory of the ball. In total, TM delivers 28 metrics, where eight are related to the club head and twenty are related to the ball flight.

Of course, technology like TM is a great tool for teaching professionals when analyzing swings. However, in practice, due to the complexity of the swing and the many metrics, teachers often focus on only a few parameters they consider the most important ones. In essence, the TM solves the problem of characterizing a swing quantitatively, but does not help to identify good and bad aspects of a particular swing.

Naturally, many previous studies have focused on analyzing the swing quantitatively using high speed video, e.g. see [4] and [5]. However, due to the tedious manual labor related to video analysis, these and similar studies have only used a small number of players (20-45). A small number of example swings in combination with several metrics make an analysis of the interaction between different swing variables difficult. Hence, these and similar studies have been restricted to an analysis of single variables using statistical techniques.

By using a TM-unit, it is, however, feasible to collect quantitative data from a larger number of golfers, since all metrics are calculated automatically. More data enables the use of more powerful techniques, like machine learning, for modeling golf player skill. In this study, we evaluate whether it is possible to discern a good swing from a bad one by only using data from the club head at impact and from ball flight.

2 Background

Golf is a game in which the player aims to hit the ball from the tee to a hole in as few strokes as possible. A golf course normally consists of 18 holes, where each hole is designed to be played with a certain number of strokes. The number of intended strokes is called the *par* of the hole and ranges from 3-5 strokes. A golf course also has a par, calculated as the sum of the par of each hole, which is normally 72 strokes.

Golf has a handicap system which is designed to let golfers with different levels of skill compete against each other. There are several different handicap systems of which the EGA [6] and USGA are dominant. The EGA system is predominant in Europe and the USGA in the USA. In essence, a handicap system lets players deduct strokes according to their handicap (hcp). A hcp lies in the range of 36 to -4 (called +4). One simple way of calculating the final score using a hcp is to subtract the player hcp from the total number of strokes. When golfers play better than their hcp, e.g., the score minus the hcp is lower than the course par, their hcp is reduced a fraction and if they play worse it is increased. Hence, a player's hcp is supposed to be an estimation of that golfer's current skill level, where a better player has a lower hcp.
Broadie in [7], sorted golf shots on a course into four different categories:

- *Long game* - shots longer than 100 yards.
- *Short game* - shots shorter than 100 yards, not including sand shots.
- *Sand game* - shots from bunker no longer than 50 yards.
- *Putting* - shots on the green.

To become a skilled player, all of these shots must be mastered, but according to Broadie, the long game is the most important one for amateur players' hcp.

The long game consists of shots from the tee and longer shots on the course. Tee shots can be hit with the driver, which is designed to hit the longest shots, or iron clubs, which are designed for different specific distances. Shots from the fairway are most often done using iron clubs.

Swings can be described in numerous ways using different terms; however, in this study, the terminology of the TM software is used consistently for simplicity. In total, Trackman delivers eight metrics related to the club head (CLUB):

- *ClubSpeed* - Speed of the club head at the instant prior to impact.
- *AttackAngle* - Vertical movement of the club through impact.
- *ClubPath* - Horizontal movement of the club through impact. (+) = inside-out, (-) = outside-in.
- *SwingPlane* - Bottom half of the swing plane relative to the ground.
- *SwingDirection* - Bottom half of the swing plane relative to the target line.
- *DynLoft* - Orientation of the club face, relative to the plumb line, at point of impact (POI).
- *FaceAngle* - Orientation of the club face, relative to the target line, at POI. (+) = open face, i.e., for a right-handed golfer to the right of the target line (-) = closed face.
- *FaceToPath* - Orientation of the club face, relative to the club path, at POI. (+) = open path, i.e., for a right-handed golfer to the right of the club path (-) = closed path.

In addition to these CLUB-related metrics, TM also registers twenty metrics related to the ball flight (BALL):

- *BallSpeed,BallSpeedC* - Ball speed the instant after impact, speed at landing.
- *SmashFactor* - Ball speed / club head speed at the instant after POI.
- *LaunchAngle* - Launch angle, relative horizon, immediately after impact.
- *LaunchDirection* - Starting direction, relative to the target line, of the ball immediately after impact. (+) = right, (-) = left.
- *SpinRate* - Ball rotation per minute the instant after impact.
- *SpinAxis* - Tilt of spin axis. (+) = fade / slice, (-) = draw / hook.
- *VertAngleC* - Ball landing angle, relative to the ground at zero elevation.
- *Height, DistHeight, SideHeight* - Maximum height of shot at apex, distance to apex, apex distance from the target line.
- *LengthC, LengthT* - Length of shot, C = calculated carry at zero elevation, T = calculated total including bounce and roll at zero elevation.
- *SideC, SideT* - Distance from the target line, C = at landing, T = calculated total including bounce and roll. (+) = right, (-) = left.

Access to launch monitors producing quantitative data, and a consensus about impacting variables, have allowed a revision of swing theory. Most importantly, im-

proved analysis has produced a fundamental change in the understanding of the ball flight. Traditionally, the club path was believed to determine the starting direction and the face angle (opening or closing the club head) to be responsible for the curvature of the shot. Modern golf theory, however, has established that it is in fact mainly the face angle that determines the starting direction [8]. Using Trackman terminology, the CLUB parameters, *FaceAngle* and *ClubPath* are the most important factors responsible for the starting direction of the ball, where *FaceAngle* is credited for 85% of the direction [8]. Similarly, the curvature of the shot is a direct function of the difference between the ClubPath and the FaceAngle (*FaceToPath*), where a negative value represents a right-to-left movement, i.e., a draw for a right-handed golfer. Naturally, the aim in most cases, is to produce a fairly straight shot down the target line, which is achieved with a *ClubPath* and *FaceAngle* of approximately zero. The *LaunchAngle* of the ball is dependent on the *DynamicLoft* which, in turn, is related to the *AttackAngle*; a higher *DynamicLoft* will produce higher shots.

Golf theory maintains that the swings for iron clubs and those for the driver differ slightly; first, the driver is designed for longer shots and is hence longer than iron clubs, which results in higher *ClubSpeeds*. Secondly, when hitting with the driver, the ball is placed towards the front facing leg, instead of in the middle, as done with iron clubs. This is done to hit the ball on the way up of the swing, i.e., with a positive *AttackAngle*, thus decreasing the *DynLoft* and thereby producing a longer shot due to a lower *SpinRate*. Shots with iron clubs are instead hit on the way down in the swing with a negative *AttackAngle*, in order to produce spin and thereby better control the length and direction of the shot.

3 Related work

Golf is a sport that many people are passionate about and, hence, much research regarding all aspects of the game has been carried out. However, not as much work based on quantitative data has been carried out, since the necessary technology has only become available in recent years. The following section presents a selection of some recent relevant studies where quantitative data was analyzed and related to players' skill.

In [4] Fradkin Sherman and Finch performed a quantitative study of how the *ClubSpeed* correlated with player hcp. Here, *ClubSpeed* speed was measured using high speed video and averaged over the strokes. Data was collected from forty-five male Australian golfers with hcps in the range of 2-27. The results show a very strong correlation (0.95) between hcp and club head speed.

Sweeny et al. performed another quantitative study in [5] and noted that even if many coaching and biomechanical texts describe how the kinematics of the club head at impact lead to distance and accuracy, there is limited quantitative evidence for these claims. Hence, an opto-reflective system was used to analyze the swings with the driver of 21 male golfers. Using the kinematics of the club at impact, i.e. *ClubSpeed*, orientation path and centeredness, five kinematics of early ball flight, i.e., *BallSpeed*, *LaunchAngle*, *LaunchDirection*, *SpinRate* and *SpinAxis*, were modeled. Experiments show that these club kinematics could explain a significant part of the early ball flight, i.e., R^2 values between 0.71 - 0.82 were achieved.

In [9] which is the only identified study that analyzed a larger group of players, 10 driver shots were recorded for 285 players. These shots were recorded mainly using five 1000Hz high speed cameras, but also the TM-launch monitor. The aim of this study was to evaluate the variability in club head presentation at impact and the resulting ball impact location on the club face, for a range of golfers with different hcps. The variability of *ClubSpeed*, *SmachFactor*, *AttackAngle*, *ClubPath*, *FaceAngle* and impact location was evaluated and compared between the different hcp groups. Statistical tests based on 10 shots from each player were used to show that overall players with lower hcps, i.e., players with hcp <=11.4, exhibited significantly less variation in all of the evaluated variables.

A rather different and interesting approach of evaluating golf player skill was taken in [7], where golf players registered real course shots in a database, using a computer. In total, 40.000 shots were registered from 130 different golfers. Each shot was then compared to a shot from a scratch player in the same situation. An interesting aspect of this study is that all parts of the games were analyzed, including the long game, short game, sand game and putting. The results show that for players with higher hcps, inconsistency was the main cause of a bad score, i.e. a few really poor shots often ruin the score. Another interesting result is that proficient players tend to be better at all parts of the game, but it is the long game, i.e. shots over 100 yards, that is the biggest influencing factor between low and high hcp players.

What all of these studies have in common is that they evaluate each parameter separately using statistical methods. To summarize the results, it is good to have a high club head speed and low variability in the presentation of the club head at impact. It is, however, not clear how low variability in the club head can be achieved or how different club head parameters relate to each other. Hence, these results demonstrate little about the difference in the swings of players with low and high hcps.

4 Method

The aim of this paper is to explain what distinguishes really proficient players from amateurs and novices. An early design choice was to model player skill on the basis of the long game, which is supposed to be the most important hcp factor for less skilled players [7]. The long game consists of both driver and iron shots and, hence, it was decided to record shots from each player with both the driver and one iron club, i.e. the 7 iron (i7). The next crucial design choice was how to discern good players from those less skilled. Since the EGA hcp system is designed to account for the skill of a golfer, it was selected as the basis of skill estimation in this study. Figure 1 shows the handicap distribution of the 277 players participating in the study. The average hcp was 12.8; the majority of players (73%) were in the hcp range of 4-18, indicating a higher skill than normal, i.e. the mean male hcp in Sweden is 21.5 and 48% of players fall into the 4-18 range [10].

The EGA hcp system divides players into five different categories based on their hcp. The best category has hcp numbers equal to or lower than 4.5. Hence, this was first chosen as a group of proficient players. However, to make the dataset a little less unbalanced, this group was expanded to include golfers with hcps <=5. For simplicity, players with hcps <=5 are henceforth called Pros and players with higher hcps are

called Hackers. In practice, this meant that 58 players, with hcps of 5 or lower, were classified as Pros and the rest (219 players) as Hackers.

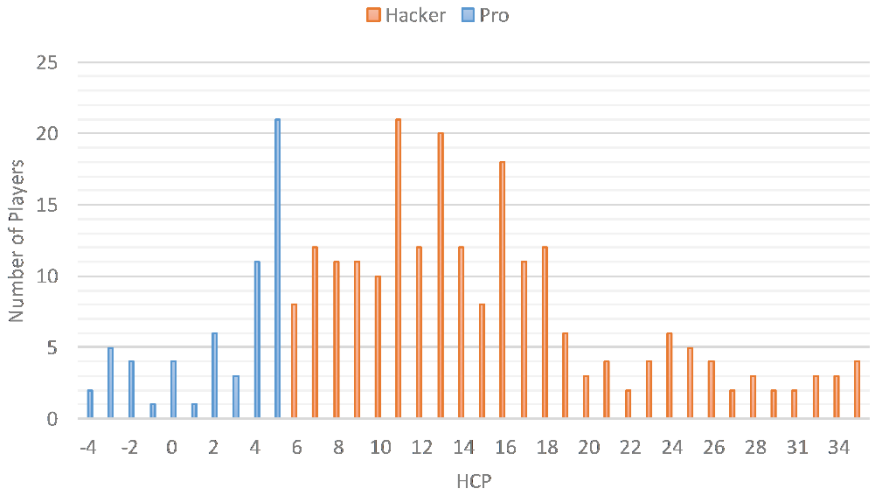

Fig. 1. Handicap distribution among players

4.1 Data Collection

Most of the data was collected during 18 days in March, April and May. Due to a rather skewed sample with only a few Pros, another collection phase with only Pros was conducted, resulting in 15 additional players. All data was collected at the same training facility [11], where a special section was reserved for each day of data collection. Since the aim of this study was to find general patterns among larger groups of players, only male golfers hitting right-handed were targeted. Female and left to right stroking players only represented a small fraction of the players at the selected training facility and could hence not be recorded in sufficient numbers.

To record a player, the radar was positioned three meters behind and slightly to the right of the hitting mat. Next, the radar was aimed (using the Trackman Performance Studio software) at a flag approximately 250 meters straight in front of the hitting mat. Before a player was recorded, he was first allowed to warm up, in order to reach a comfortable stroking state. Thereafter, five consecutive strokes were recorded using the players own seven iron followed by five strokes with the driver.

Naturally, an identical setting with regard to temperature and wind would have been preferable. However, this was impossible, since an indoor facility was not available. Instead, TrackMan's built-in normalization functionality was used. When normalizing ball data, TrackMan utilizes information from the club head at impact to correct deviation caused by wind, temperature, altitude and ball type. Hence, the players were told to hit the balls in the direction of the flag using a normal full stroke, while disregarding any wind if present.

4.2 Preprocessing

Trackman Performance Studio output a total of 28 metrics, of which eight were related to the club head and twenty to the ball flight. Most metrics are measured directly, but some are calculated on the basis of other metrics. If the radar cannot measure some aspects accurately, which rarely happens but frequently enough to be an issue, it does not output any values. This typically occurs for club head related metrics, i.e. *ClubSpeed*, *AttackAngle*, *ClubPath*, *SwingPlane* and *SwingDirection*, at lower swing speeds. Ball related metrics are most often recorded properly, with the exception of *SpinRate* which is sometimes estimated instead or not reported at all. Of the 2780 shots that were recorded, 187 had at least one missing value.

One approach of obtaining representative measurements, for a single player and club, could be to use the average value of all five recorded shots. If, however, one of the shots were to be a really poor one, the average values could be quite misleading. Instead, the more robust approach, argued by Broadie in [8], of using the median value was chosen. The median value disregards both the worst and the best shots and should hence be a better estimate of normal standards. The question of how to define the median shot of a group of shots still remains. In this study, we chose the median shot based on *LengthC* of the ball, since the length of a shot is one of the most important aspects of a good shot. The distance from the target line, i.e. *SideC*, could, of course, be another alternative, but since a straight shot is not effective if it is not long enough, we settled for length in this study. In eleven cases, the median shot, i.e. the third longest shot, had some missing value. In these cases, the second best shot was used instead. If the second best longest also happened to have a missing value, the fourth best shot was used instead. For one single player all shots with the driver had at least one missing value and, hence, this player was removed from that dataset.

Since previous research has shown strong correlations between skill and consistency, the standard deviations of each of the 28 metrics were also calculated for each club using all five shots. Shots with missing values were not excluded, except in the calculation of the metric with the missing value. The average number of shots with no missing values was 4.7 for a specific club.

Another issue was how to best represent each metric for a predictive modeling technique. Most metrics, like *LengthC*, have a straightforward representation, but metrics related to angles need some extra consideration. *FaceAngle* is one example where the representation plays an important role, since the angle can be both positive and negative, i.e. representing an angle to the right or left of the target line. If no transformation is done, a big negative angle would be considered as smaller than a small positive angle. However, in relation to the target line, which is more relevant for the quality of a swing, the opposite is true. Hence, each metric related to the target line was replaced with two new variables, where the first one was the absolute value and the second was a binary variable with the same name but preceded with a *P-*, to represent whether the original angle was positive or not. Metrics related to vertical angles, i.e. *AttackAngle*, *LaunchAngle*, were not modified.

4.3 Experiments

The aim of this paper was mainly to investigate how well golf skill can be modeled using quantitative data from a TM-Radar and to understand which factors are the most important ones for a good golf swing. Naturally, this requires that the predictive models are transparent and comprehensible and, hence, WEKAs J48, which is an implementation of the famous decision tree algorithm C4.5 [12], was used in the main experiments. However, to evaluate how much predictive power the data actually contained, a Random Forest [13] was first applied as an upper benchmark.

The separation of the players into Pros and Hackers was rather unbalanced, i.e. 58 Pros vs 219 Hackers. Unbalanced data sets often result in models favoring the majority class at the cost of very few predictions for the minority class. Initial experiments confirmed that this was also the case for both the 7i- and the driver datasets in this study, resulting in true positive rates of around 35%. Hence, since the main goal was to model the proficient players, a cost of miss-classification equal to the unbalance rate was associated with the minority class. This was done for both RF and J48 using WEKA's meta *cost sensitive classifier* by assigning a cost of 219/58=3.78 for miss-classifications of Pros. To keep the total weight of the instance to the original 277, WEKA, in practice, assigned a weight of 0.632 to Hackers and 2.389 to Pros.

When modeling using J48, the main goal is to describe good general swings in a comprehensible way. Hence, it was decided that at least 10% of the Pros, i.e. six players, should be present in any pure Pro leaf node. However, due to WEKAs re-weighting of the instances, a minimum instance per leaf of 6 would only require three Pros. To avoid this and to ensure that in practice there were always at least six instances in a pure Pro leaf, the minimum number was instead set to 6 *2.389=14.

Another setting motivated by the class imbalance was to use *Laplace* estimates in J48, since previous studies, e.g. [14], have shown that it most often improves the probability estimates and thereby the ranking ability of the produced model.

To facilitate a comparison of the predictive power of variables related to club delivery and the ball flight, the original datasets were used to create four new ones based on four subsets of the original variables. The attribute subsets used to create the new datasets were:

- CLUB Variables related to the club delivery.
- C-STD Player standard deviations for CLUB variables.
- BALL Variables describing ball Launch and flight, Carry flat and Est. Total.
- B-STD Player standard deviations for BALL variables.

Based on these subsets, five datasets were then created; only CLUB, CLUB and S-STD, only BALL, Ball and B-STD and one final dataset containing all variables.

4.4 Model interpretation

Since all decision trees techniques optimize some kind of information gain criteria, starting at the root, splits closer to the root are normally considered more important. Information gain is the difference in purity, e.g. *gini diversity index* (GDI) [15] or *entropy* [16], between the original dataset and the resulting subsets. The equation below defines a general way to calculate the information gain given a specific purity measure, where $P(D_i)$ is the proportion of the dataset D that is placed in the subset

D_i. The split resulting in the highest purity gain is selected and the procedure is then repeated recursively for each subset in this split.

$$gain(D,S) = purity(D) - \sum_{i=1}^{s} P(D_i) * purity(D_i) \quad (1)$$

Consequently, the importance of the final splits can be, in the same way, evaluated by considering the number of training instances affected by the split and the resulting purity of the subsets.

5 Results

The following sections present the results of the experiments. First, basic statistics are presented followed by the predictive performance of the models. Finally, a selection of models are presented and interpreted, to ensure that they are practical and based on rational relationships.

5.1 Basic statistics

Table 1 shows the basic statistics of the players labeled as Hackers and Pros. For each attribute and club, the mean value and the standard deviation are presented. Note that the standard deviations are within each group and not for each player which is another variable used in the experiments discussed above.

The most obvious differences between Hackers and Pros are in the values for *Ball-Flight*, *CarryFlat* and *LengthT* variables, i.e. Pros hit the ball longer and straighter. More specifically, Pros on average hit the ball 29 meters longer with the 7i and 49 meters longer with the driver. However, they also hit the ball straighter than the Hackers, (4 meters smaller deviation from the target line for both 7i and the driver), in spite of the longer shots. Note that since the Pros hit the ball substantially further, the difference in accuracy is essentially much greater than the 4 meter deviation implies. Furthermore, the standard deviations are much smaller (for all variables) for Pros, showing that they are a more homogeneous group, in terms of how they hit the ball.

Regarding the variables related to club delivery, the arguably biggest differences (always in favor of the Pros) concern *ClubSpeed*, *AttackAngle ClubPath* and *FaceTo-Path*, i.e. the Pros hit harder, more down on the ball (i7) and deliver the club and club face better in relation to the target line, resulting in longer and straighter shots. One variable that appear to be rather similar for the Hackers and Pros is the *SwingPlane*, which only differs by 1.6 degrees for the 7i and 2.1 for the driver.

The longer shots of the Pros can be further explained by the Ball Launch variables. Pros have a higher *BallSpeed* which is a result from the higher *ClubSpeed* and a better *SmashFactor*, which means that more energy is transferred from the club to the ball. Another interesting variable is the *SpinRate*, where a lower spin rate promotes longer shots while higher rates give more control over the ball flight. As could be expected, Pros have a lower *SpinRate* for the driver, thus maximizing distance, but a higher *SpinRate* for the 7 iron, thus enabling more control compared to the Hackers. Another

basic assumption that can be verified is that it is common among Hackers to hit the ball with a slice or fade. Slice or fade is signified by a positive *SpinAxis*, which 69% of the Hackers have. The distribution of positive and negative *SpinAxis* among the Pros is even and *SpinAxis* itself is much smaller, resulting in just a small fade or draw.

Table 1. Basic statistics for Hackers and Pros

	i7				Driver			
	Mean		STD		Mean		STD	
	Hacker	Pro	Hacker	Pro	Hacker	Pro	Hacker	Pro
Club Delivery								
ClubSpeed	80.2	87.2	8.6	5.9	94.5	104.7	10.7	7.7
AttackAngle	-3.2	-4.8	2.4	2	-1.7	-1.1	3.5	2.9
ClubPath	4	2.9	3	2.4	3.8	3.6	3.2	2.7
P-ClubPath	0.45	0.6	0.5	0.49	0.49	0.69	0.5	0.46
SwingPlane	61.9	59.3	4.4	3.6	49.9	47.8	4.5	3.4
SwingDir.	4.8	3.5	3.4	2.6	5.6	4.8	4.1	3.6
P-SwingDir.	0.35	0.26	0.48	0.44	0.42	0.53	0.49	0.5
DynLoft	25.3	23.3	4.8	2.9	14.3	12.6	4.9	2.7
FaceAngle	3.3	2.1	2.6	1.6	4.2	2.6	3.1	1.8
P-FaceAngle	0.38	0.52	0.49	0.5	0.33	0.45	0.47	0.5
FaceToPath	4	2.5	3.4	1.7	4.6	3.4	3.7	2.4
P-FaceToPath	0.42	0.38	0.49	0.49	0.33	0.24	0.47	0.43
Ball Launch								
BallSpeed	104	116.1	11	6.7	133.1	151.8	16.5	10.4
SmashFactor	1.3	1.33	0.06	0.04	1.41	1.45	0.07	0.04
LaunchAngle	20	18.4	3.8	2.8	12.3	11.2	4.2	2.7
LaunchDir.	3	2	2.3	1.7	3.8	2.6	2.8	1.7
P-LaunchDir.	0.4	0.55	0.49	0.5	0.34	0.47	0.48	0.5
SpinRate	5470	5686	1532	847	2837	2207	1318	613
SpinAxis	6.1	3.3	5	2.3	10.7	8.2	7.8	5.4
P-SpinAxis	0.69	0.5	0.46	0.5	0.6	0.47	0.49	0.5
Ball Flight								
DistHeight	86.8	102.7	14.8	8.5	118	148.4	26.2	18.5
Height	24.7	30.1	7.2	5.7	19.6	22.8	9	6.6
SideHeight	4.8	3.2	4.1	2.9	8.1	6.4	6.3	5.1
P-SideHeight	0.45	0.53	0.5	0.5	0.39	0.47	0.49	0.5
Carry Flat								
LengthC	142.1	163.3	21.1	11.4	192.2	238.1	36.8	25.8
SideC	9	5.7	7.9	5.1	15.6	12	12	10.1
P-SideC	0.46	0.5	0.5	0.5	0.44	0.47	0.5	0.5
VertAngleC	42.3	45.9	6.7	4.6	29	29.2	9.9	6.5
BallSpeedC	54.7	56.6	3.1	2.4	65.6	68.6	7.7	4.6
Est. Total								
LengthT	153.2	172.2	21.2	11.3	223.3	271.8	36.4	24.8
SideT	9.1	5.5	8.1	4.9	17.5	13.2	13.3	11.1
P-SideT	0.46	0.5	0.5	0.5	0.46	0.5	0.5	0.5

Fig. 2 presents the players in a scatter plot where the y-axis is the carry distance and the x-axis is the deviation in meters from the target line. As Table 1 also shows, Pros, in general, hit longer and straighter than Hackers. However, there are many Hackers who hit as far as the Pros and the longest shot (207m) was actually made by a Hacker (who, it transpired, was competing in the longest drive competition). The point is that some Hackers can easily be discerned on the basis of distance and side deviation, but far from all.

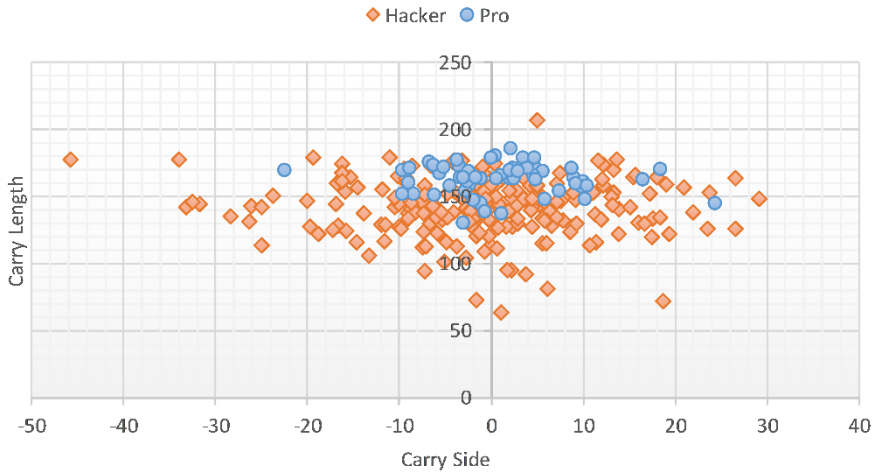

Fig. 2. i7 - Carry Length vs. Carry side

Next, Fig. 3 instead shows the standard deviation of the five recorded 7i-shots for each player. Obviously, the Pros are more consistent, i.e. deviate less, than the Hackers, in spite of hitting the ball further. Again there is no clear distinguishing boundary separating the two classes, except for the really substandard Hackers.

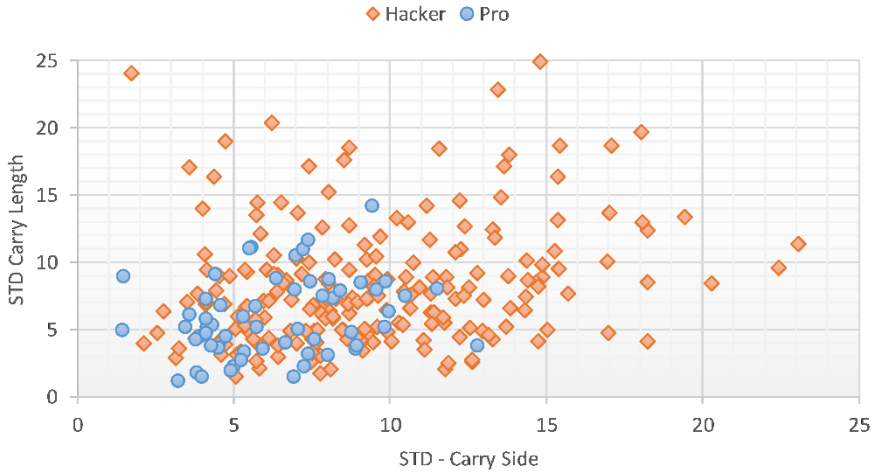

Fig. 3. i7 - STD of Carry Length vs. Carry side

The same pattern is apparent in Fig. 4, i.e. the group of Hackers has a greater variation in their *SwingPlane* and *FaceAngle*, however, only some can be easily discerned from the Pros using only these variables.

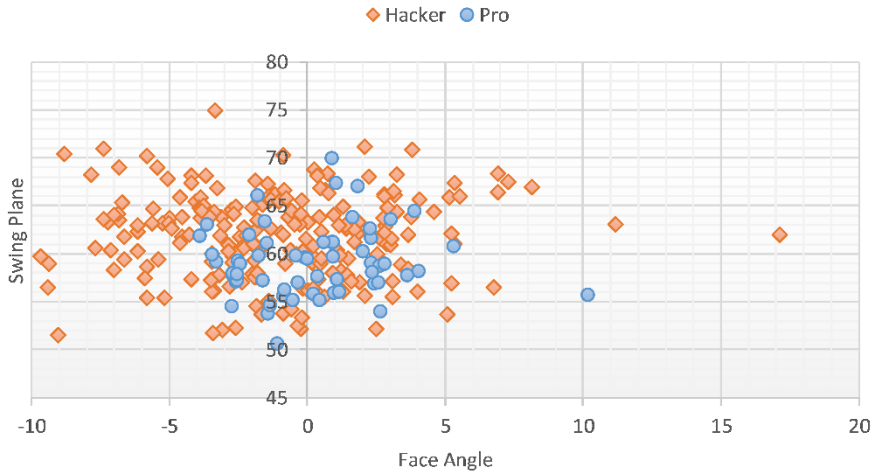

Fig. 4. i7 - SwingPlane vs. Face Angle

5.2 Predictive Performance

Table 2 below shows the results for RF, using n-fold cross validation for the different datasets. The first columns show the club that was used and the following four define the data that was present in each dataset.

Table 2. Random forest n-fold results

RF	CLUB	C-STD	BALL	B-STD	ACC	AUC	TP
i7	X				0.773	0.792	0.534
i7	X	X			0.848	0.886	0.655
i7			X		0.798	0.833	0.69
i7			X	X	0.874	0.901	0.776
i7	X	X	X	X	0.866	0.907	0.724
Driver	X				0.801	0.824	0.569
Driver	X	X			0.797	0.872	0.638
Driver			X		0.826	0.839	0.655
Driver			X	X	0.855	0.902	0.776
Driver	X	X	X	X	0.866	0.908	0.724

The first observation that can be made is that the results for datasets based on i7 and driver data are rather similar overall, even if they differ slightly for a specific dataset. Furthermore, BALL and B-STD are, in general, slightly better than the corresponding combination of CLUB and C-STD.

Using all the data, RF achieves a rather high accuracy of 86.6% for both clubs, while the naive approach of only predicting Hacker would achieve an accuracy of 79%. Due to the class imbalance, it is arguably more interesting to look at the area under the ROC curve (AUC), since it is a measure of how well the model can rank the players. Again, the RF performs well and reaches a high AUC of 0.901 for both clubs. The true positive rate (TP) for the Pro class is 72.4%, which is a reasonable level with the class imbalance in mind.

As may be expected, an analysis of the results for the attribute subsets shows that including the standard deviations for either the club or the ball flight provides extra predictive power. The second best results are achieved using datasets with BALL and B-STD, closely followed by CLUB and C-STD.

The data sets without any standard deviations, i.e. BALL and CLUB for i7 and driver, all clearly result in models with lower predictive performance. Using just BALL data, compared to BALL and B-STD, results in a drop of 8.6% in TP for 7i and 12% for the driver. Similarly, using only CLUB data results in a loss in TP of 12% for i7 and 26% for the driver, compared to using CLUB and C-STD.

Next, the results of J48 are presented in **Table 3**. Compared to the results of RF in **Table 2**, it becomes apparent that RF is superior for all datasets and metrics, except for a few cases where J48 achieves a higher TP.

Table 3. J48 n-fold results

J48	CLUB	C-STD	BALL	B-STD	ACC	AUC	TP	SIZE
i7	X				0.773	0.755	0.741	13
i7	X	X			0.762	0.771	0.672	13
i7			X		0.765	0.797	0.69	13
i7			X	X	0.783	0.83	0.69	5
i7	X	X	X	X	0.823	0.835	0.69	9
Driver	X				0.699	0.729	0.638	11
Driver	X	X			0.83	0.778	0.759	13
Driver			X		0.71	0.732	0.724	11
Driver			X	X	0.764	0.782	0.672	13
Driver	X	X	X	X	0.815	0.835	0.621	11

The general results for J48 are also similar to those of RF, where the datasets containing the standard deviation of the ball (STD-B) again give the best results, regardless of the club that was used. This also correlates with previous research showing that consistency is a very important aspect of the swing. For both clubs, the best result is achieved using all variables, closely followed by only BALL and B-STD. The da-

tasets containing only CLUB and C-STD variables were slightly worse, but still surprisingly accurate. An interesting detail is that adding STDs to the CLUB data only had a minimal effect for the 7i but a rather large one for the driver.

The most interesting result is, however, that the J48 achieves a rather high predictive performance with relative small and simple models. Hence, it can be argued that the trees capture a substantial part of what distinguishes good swings from bad ones while still being small and comprehensible. The predictive power is, of course, higher for RF, but these models are opaque and hence of limited value for human inspection and analysis. Next, we present a few sample J48 models to evaluate how well they fit with existing golf theory and how usable they would be in a teaching situation.

5.3 Interpretation of models

The decision trees presented below were created using data from all the players, but do otherwise use an identical setup to the main experiment. Hence, the overall accuracy of these models for new players should be judged on the results in Table 3.

The result for each leaf is, however, still presented, to show the accuracy of each rule for the existing golfers. The trees are presented in textual form where each leaf is given a number preceded with # to facilitate referencing. Leaves defining Pros and splits leading to Pro leaves are marked in green, while Hacker leaves and splits are marked in red. Splits that do not directly lead to a leaf are marked in black, since they do not define a property exclusively for Pros or Hackers. Each leaf also reports the number of golfers of the leaf class, the true positive rate (TP) of the leaf and the average hcp of all players classified by the leaf, i.e. (#players of leaf class|TP|hcp). Four trees, one based on all the data and one based on only CLUB data for each club, are discussed below. The reason for this choice is to present the most accurate trees (using all the data) and the more usable trees based on only club data.

The general result of the interpretation of the models below is that they also concur with general golf theory. Another general result is that Hackers can be identified with less splits than Pros. In practice, this means that you must master several skills to become a Pro, while just a few deficiencies in the swing are often enough to categorize a hacker. Another interesting observation is that all the trees tend to group players with similar hcps in a leaf, thus creating an implicit ranking of the players, in spite of the problem being setup as a classification task. This signifies that each split has predictive power in itself and thus represents a skill that needs to be mastered by a golfer. The contrary would apply if all the leaves relating to a particular class (Hacker or Pro) would have a similar average hcp. This would mean that all the splits leading to a rule would be needed to discern Hackers from Pros. Looking at this explains the higher AUC which exactly measures the ability to rank the players.

Model based on all data for i7

Next, Fig. 5 below, depicts a tree created from all the recorded data for the i7, which resulted in the most accurate model. Looking at the particular tree presented in Fig. 5, the first group of Hackers, i.e. rule #5, is signified by having a high STD in their launch direction. In practice, this means that they have problems controlling the starting direction of their shots. The average hcp for this large group of 115 Hackers is 17.

The next group of Hackers classified by rule #1 (62 Hackers with an average hcp of 14.7) is better at controlling the *LaunchDirection* but has a lower *DistHeight* than the better golfers in leaf #3-#4. A lower *DistHeight* means that they do not hit the ball far enough, i.e. the apex and the ball trajectory are not as far away from the player, which naturally results in a shorter shot.

```
S-LaunchDirection <= 2.14
|       DistHeight <= 89.5 #1: Hacker (62|.95|14.7)
|       DistHeight > 89.5
|       |       S-SmashFactor <= 0.029 #2: Pro (39|.76|4.2)
|       |       S-SmashFactor > 0.029
|       |       |       S-DynLoft <= 1.42 #3: Pro (13|.52|7.4)
|       |       |       S-DynLoft > 1.42 #4: Hacker (18|.90|11.0)
S-LaunchDirection > 2.14 #5: Hacker (115|.99|17.0)
```

Fig. 5. Tree based on i7: CLUB, STD-C, BALL and STD-B

The first group of Pros, the largest group of 39 Pros with an average hcp of 4.2, is discerned by leaf #2. These Pros hit the ball further than the Hackers in group #1 and have a lower STD of their *SmashFactor* than players in groups #3 and #4. In essence, this means that they are more consistent in how well they hit the ball, which of course is important for controlling the length and accuracy of a shot.

The last two groups (#3 and #4) have a higher STD in their *SmashFactor* and a higher average hcp (7.4 for #3 and 11 for #4). Rule #3 defines the better golfers of these two groups which are characterized by their ability to better control the STD of their *DynLoft*. The *DynLoft* has a very strong correlation to the launch angle, which is an important factor for *DistHeight* and thereby the length of the shot. Hence, the difference between the Pros in rule #3 and the Hackers in rule #4 is that the Pros are better at controlling the length of their shots. The difference in hcp is, however, not large, which explains the low TP of .52 for rule #3.

To summarize, this tree tells us that the first skill a golfer must learn is to start the ball consistently on the target line and thereafter to hit long enough. Once mastered, the next skill to conquer is the ability to consistently hit the ball the same distance.

Models based on CLUB data for i7

Even if this tree, presented in Fig. 5, is one of the more accurate trees, it does not say much about how the swing itself should be performed, since this is mainly based on BALL-data. Hence, a model for 7i based only on CLUB-data.

The first split in this tree is based on the *DynLoft* with a split point of 29.6. This value is far from the Pros' mean value of 23.5 and the average Hacker value of 25.3 presented in Fig. 6. Hence, the 24 players that are all correctly classified as Hackers by leaf #7 have a way to high dynamic loft. The next group of players discerned by leaf #1 is again Hackers with a more reasonable *DynLoft* but a rather low *ClubSpeed* of 77.1, resulting in shorter shots. It is not a big difference from the average *ClubSpeed* for Hackers, which is 80.2, but far from the Pros at 87.2. Leaf #5 shows that the next group of Hackers has high *FaceAngle*, resulting in shots in the wrong direction, if there is no corresponding curvature of the ball (hook or slice) to counteract the ini-

tial direction. Nonetheless, swings with higher *FaceAngles* are harder to control and repeate consistently.

Players in group #5 are again classified as Hackers who are here discerned from the remaining golfers by a rather large *ClubPath* angle (4.8) in relation to the target line. A high *ClubPath* value will also make it harder to control the ball flight, for more or less the same reasons as the *FaceAngle*. It is, however, interesting to note that *FaceAngle* is used higher up in the tree, which signifies that it is a more important factor than the *ClubPath*. This concurs with modern golf theory which has only been applied a few years.

Leaf #4 is the first leaf to classify players as Pros. Players classified by this rule (including 37 Pros and 24 Hackers) have avoided the rules leading to the leaves #1,#7,#6,#5 and thus have a reasonable swing, but also a high *ClubSpeed*, i.e. higher than 86.2. Players who have avoided the leaves #1, #7,#6,#5, but have a lower *ClubSpeed* than the Pros in #4, are classified as Pros or Hackers by leaves #2 and #3, depending on the SwingPlane. Here, a lower swing plane is favorable, which by itself is interesting, especially since Table 1 does not show any major difference in the swing plane between Hacker and Pros.

It should be noted that even if the TP rate for leaf #2 is only 0.41, it still classifies the players as Hackers due to the cost-sensitive learning applied. Nonetheless, the average hcp of this leaf is lower than all the Hacker leaves and 5 points lower than the neighboring Hacker leaf (#3). Hence, the split is still important for discerning Hackers from Pros.

```
DynLoft <= 29.6
|    ClubSpeed <= 74.7 #1: Hacker (42|1.0|17.7)
|    ClubSpeed > 74.7
|   |   FaceAngle <= 4
|   |   |   ClubPath <= 4.8
|   |   |   |   ClubSpeed <= 86.2
|   |   |   |   |   SwingPlane <= 60.4 #2: Pro (13|.41|10.1)
|   |   |   |   |   SwingPlane > 60.4 #3: Hacker (32|.91|15.1)
|   |   |   |   ClubSpeed > 86.2 #4: Pro (37|.61|5.7)
|   |   |   ClubPath > 4.8 #5: Hacker (30|.91|11.6)
|   |   FaceAngle > 4 #6: Hacker (34|.94|15.3)
DynLoft > 29.6 #7: Hacker (24|1.0|17.4)
```

Fig. 6. - Tree based on i7: CLUB

Models based on all data from the driver

In the following section, models based on data from shots with the driver are presented. Fig. 7 depicts a decision tree created using all the available data. The first group of Hackers, i.e. leaf #6 in Fig. 7, is selected on the basis of the STD of the *LaunchAngle* (2.89), which lies close to the Pros value of (2.7). Consequently, a large number of golfers, (93) are selected in this leaf but, surprisingly, all are correctly classified as Hackers. Next, leaf #5 selects 27 Hackers on the basis of a relatively large STD of the *FaceAngle*. More Hackers are classified in leaf #1 on the basis of a *BallSpeed* less than 128.5, which is rather low compared to the average *BallSpeed* of Pros at 151.8. Obviously, these players cannot hit the ball far enough.

Leaf #4 is characterized by a higher STD of the *SmashFactor*, which relates to how consistently the player hits the ball. The split point of 0.04 is identical to the Pros mean value and, consequently, a few Pros (3) are miss-classified as Hackers in this leaf. The final split divides Pros from Hackers on the basis of their *SwingPlane*, where it is again favorable to have a flatter plane. The chosen split point of 51.89 is higher than the average driver *SwingPlane* of both Pros and Hackers, thus discerning rather good Hackers with atypical swings, i.e. the average hcp of rule #3 is 9.8.

```
S-LaunchAngle <= 2.89
|   S-FaceAngle <= 3.66
|   |   BallSpeed <= 128.52 #1: Hacker (26|1.0|15.1)
|   |   BallSpeed > 128.52
|   |   |   S-SmashFactor <= 0.04
|   |   |   |   SwingPlane <= 51.89 #2: Pro (52|.61|6.4)
|   |   |   |   SwingPlane > 51.89 #3: Hacker (19|.83|9.8)
|   |   |   S-SmashFactor > 0.04 #4: Hacker (19|.86|10.4)
|   S-FaceAngle > 3.66 #5: Hacker (27|1.0|17.2)
S-LaunchAngle > 2.89 #6: Hacker (93|1.0|18.2)
```

Fig. 7. Tree based on Driver: CLUB, STD-C, BALL and STD-B

Models based on CLUB data from driver

The final tree presented in Fig. 8 is based on only CLUB variables for the driver. This tree selects exactly the same subset of variables as the i7 tree presented in , but in a slightly different order. The split point values are, however, quite different; which is natural because the driver is a longer club and designed to launch the ball at a lower angle.

The first group of Hackers in leaf #6 has a *DynLoft* at more than 18.26 degrees, which is much too high, compared to the mean values of the Pros which are at 12.6 degrees. A high *DynamicLoft* results in a higher *LaunchAngle* and thereby a higher and shorter shot. Next, leaf #5 discerns players with a high *FaceAngle*, i.e. 5 degrees higher than that of the Pros, which results in shots in the wrong direction or curved shots that are harder to control.

The third group, defined by leaf #4, is the first to classify golfers as Pros on the basis of a *ClubSpeed* of 104.14, which is very close to the average value for Pros. Since the leaves #5 and #6 are based on rather extreme values, the main factor discerning these really proficient players, with a mean hcp at 5.5, from the rest of the golfers is high swing speed and reasonable values for *FaceAngle* and *DynLoft*. Of the remaining golfers, 44 players are classified as Hackers in leaf #3, on the basis of a rather steep swing plane. Finally, the remaining players are classified by leaves #1 and #2, again on the basis of their *ClubSpeed*. Players with club speeds lower than 89.54, which is very low compared to Pros at 104.7, are classified as Hackers and the rest as Pros. The TP rate for this last set of Pros is low (0.38), but due to the cost-sensitive training the leaf still classifies golfers as Pros. It could, however, be argued that the predictive power of this leaf is higher than indicated by this value, since the average hcp of this rule is 4.5 points lower than any leaf classifying Hackers. Obviously, players misclassified by this rule are still rather good players.

To summarize, there are two groups of Pros who all have reasonable *DynLoft* and *FaceAngle*. The best groups of Pros are, in addition, characterized by a high *ClubSpeed*. The other group of Pros has a lower *ClubSpeed* but they distinguish themselves from the Hackers by a flatter *SwingPlane*.

```
DynLoft <= 18.26
|   FaceAngle <= 7.26
|   |   ClubSpeed <= 104.14
|   |   |   SwingPlane <= 51.89
|   |   |   |   ClubSpeed <= 89.54 #1: Hacker (22|.96|15.5)
|   |   |   |   ClubSpeed > 89.54 #2: Pro (36|.38|10.6)
|   |   |   SwingPlane > 51.89 #3: Hacker (44|1.0|15.0)
|   |   ClubSpeed > 104.14 #4: Pro (35|.59|5.6)
|   FaceAngle > 7.26 #5: Hacker (29|1.0|19.6)
DynLoft > 18.26 #6: Hacker (41|1.0|19.2)
```

Fig. 8. Tree based on Driver: CLUB

6 Discussion and conclusions

From the results presented above, it is clear that it is possible to model golf player skill on the basis of quantitative data from the club head at impact or the ball flight. Predictive models with high ranking ability, i.e. AUC, can be created using both RF and J48 with an advantage to RF. For the 7i datasets containing only CLUB data, there was no difference in the predictive performance of the models. Obviously, including STDs and ball data facilitates more complex relationships, which the random forest can model but not the decision tree. The models do, nonetheless, concur in the variables they consider important, which adds credibility to the found models.

It is also interesting to note that the ranking ability is high, in spite of a setup as a binary classification task. When interpreting the models, this can be seen in the fact that players in leaves longer from the root, in general, have a lower average hcp. Obviously, each split has predictive power and leaves further from the root require that more skills need to be fulfilled by a player.

Even if the random forests are more accurate, they do not give insights as actionable as the decision trees. Future work could however address this deficiency by applying a sensitivity analysis showing the relative importance of the variables. The relative importance would possibly give a teaching professional sufficient knowledge on the particular aspect of the swing that they should improve to lower a player's hcp. This is an important point, since teaching professionals are normally mainly concerned with improving a player's swing more ideal, which does not always result in a lower hcp, the goal of many Hackers.

All J48 models are, however, small, comprehensible and concur with modern swing theory. More specifically, consistency is more important than any specific value of the club head or ball flight. Looking at only the CLUB variables, it is clear that the *ClubSpeed*, *FaceAngle*, *ClubPath* and *DynLoft* are the most important variables

for discerning good players from those less skilled. This also concurs with golf theory and is hence a strong result for the correctness of the models.

A more surprising and interesting result is, however, that the *SwingPlane* is also a strong indicator of skill, where proficient players tend to have flatter swings. This is surprising, since there is only a small deviation in the average *SwingPlane* of Hackers and Pros. The reason for this result is not obvious from modern golf theory and we can only speculate about the reasons; e.g., a flatter swing could possibly lead to higher *ClubSpeed* at the bottom of the swing, due to a more circular movement of the club. However, both decision tree models created from only CLUB data contradict this theory, since they split the players on the basis of *ClubSpeed* before *SwingPlane*. Hence, the *SwingPlane* is used to discern among players with rather similar *ClubSpeed*. Naturally, the flatter swings found at better players in this study is an important finding that needs to be studied in more detail. Obviously, a flatter plane gives rise to a different chain of movements in the swing, thus affecting the release pattern and the impact position. Specifically, a flatter swing plane is associated with swinging from the inside and a release leading to a lower dynamic loft and higher smash factor, which was found to be favorable in the induced models.

Finally, the J48 models show a clear distinction in the swing of Hackers and Pros. When modeled using only CLUB data, the swings of golfers classified as Pros were characterized by reasonable values for *FaceAngle*, *ClubPath* and *DynLoft* while having a high *ClubSpeed*. The majority of Hackers could be identified by a large deviation (from the average Pro value), for one of these variables. Better Hackers were discerned by having either a slightly lower *ClubSpeed* or a steeper *SwingPlane* than the Pros.

References

1. Wheeler, K. and Nauright, J., "A Global Perspective on the Environmental Impact of Golf," *Sport Soc.*, vol. 9, no. 3, pp. 427–443, 2006.
2. Smith, A., Roberts, J., Wallace, E. and Forrester, S., "Professional golf coaches' perceptions of the key technical parameters in the golf swing," *Procedia Eng.*, vol. 34, pp. 224–229, 2012.
3. "TrackMan A/S." Denmark, 2015.
4. Fradkin, A., Sherman, C., and Finch, C., "How well does club head speed correlate with golf handicaps?," *J. Sci. Med. Sport*, vol. 7, no. 4, pp. 465–472, Dec. 2004.
5. Sweeney, M., Mills. P. M., Alderson, J., and B. C. Elliott, "The influence of club-head kinematics on early ball flight characteristics in the golf drive," *Sport. Biomech.*, vol. 12, no. 3, pp. 247–258, 2013.
6. E. G. A. EGA, "EGA Handicap System," 2012.
7. Broadie, M., "Assessing Golfer Performance Using Golfmetrics," *Sci. Golf V Proc. 2008 World Sci. Congr. Golf*, no. 1968, pp. 253–262, 2008.
8. Tuxe n, F., "The Secret of the Straight Shot II," 2009.
9. Betzler, N. F., Monk, S. A., Wallace, E. S. and Otto, S. R., "Variability in clubhead presentation characteristics and ball impact location for golfers' drives," *J. Sports Sci.*, vol. 30, no. 5, pp. 439–448, 2012.
10. SGA, "Swedish Golf Association," 2015. [Online]. Available: www.golf.se. [Accessed: 04-Mar-2015].
11. "World of Golf." Västra Frölunda, Sweden, 2015.

12. Quinlan, J. R, *C4. 5: programs for machine learning*, vol. 240. Morgan Kaufmann, 1993.
13. Breiman, L. “Random forests,” *Mach. Learn.*, pp. 5–32, 2001.
14. Provost, F. and Domingos, P., “Tree induction for probability-based ranking,” *Mach. Learn.*, vol. 52, no. 3, pp. 199–215, 2003.
15. Breiman, L. *Classification and regression trees*. Chapman & Hall/CRC, 1984.
16. Quinlan, J. R. “Induction of decision trees,” *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986.