Research in Sports Medicine
An International Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/gspm20

Importance of anthropometric features to predict physical performance in elite youth soccer: a machine learning approach

Tindaro Bongiovanni, Athos Trecroci, Luca Cavaggioni, Alessio Rossi, Enrico Perri, Giulio Pasta, F. Marcello Iaia & Giampietro Alberti

RESEARCH IN SPORTS MEDICINE
2021, VOL. 29, NO. 3, 213–224
https://doi.org/10.1080/15438627.2020.1809410

Importance of anthropometric features to predict physical performance in elite youth soccer: a machine learning approach

Tindaro Bongiovanni[a], Athos Trecroci[b], Luca Cavaggioni[b,c], Alessio Rossi[d], Enrico Perri[b], Giulio Pasta[e], F. Marcello Iaia[b] and Giampietro Alberti[b]

[a]Nutrition, Hydration & Body Composition Department, Parma Calcio 1913, Parma, Italy; [b]Department of Biomedical Sciences for Health, Università degli Studi di Milano, Milano, Italy; [c]Department of Endocrine and Metabolic Diseases, IRCCS Istituto Auxologico Italiano, Obesity Unit and Laboratory of Nutrition and Obesity Research, Milan, Italy; [d]Department of Computer Science, University of Pisa, Pisa, Italy; [e]Medical Department, Parma Calcio 1913, Parma, Italy

ABSTRACT

The present study aimed to determine the contribution of soccer players' anthropometric features to predict their physical performance. Sixteen players, from a professional youth soccer academy, were recruited. Several anthropometric features such as corrected arm muscle area (AMA_{corr}), arm muscle circumference (AMC) and right and left suprapatellar girths (RSPG and LSPG) were employed in this study. Players' physical performance was assessed by the change of direction (COD), sprint (10-m and 20-m), and vertical jump (CMJ) tests, and Yo-Yo Intermittent Recovery Test level 1 (Yo-Yo IRT1). Using an extra tree regression (ETR) model, the anthropometric features permitted to accurately predict 10-m sprint, 20-m sprint and Yo-Yo IRTL 1 performance ($p < 0.05$). ETR showed that upper-body features as AMA_{corr}, and AMC affected 10-m and 20-m sprint performances, while lower-body features as RSPG and LSPG influenced the Yo-Yo IRTL 1 (Overall Gini importance ≥ 0.22). The model predicting COD and CMJ presented a poor level of prediction, suggesting that other factors, rather than anthropometric features, may concur to predict their changes in performance. These findings demonstrated that the upper- and lower-body anthropometric features are strictly related to sprint and aerobic fitness performance in elite youth soccer.

ARTICLE HISTORY
Received 17 March 2020
Accepted 8 August 2020

KEYWORDS
Body composition; aerobic fitness; anthropometry; change of direction; artificial intelligence; data mining

Introduction

In youth soccer, players' athleticism depends on different factors linked to the anthropometric features (e.g. derived body composition variables) and physical performance. Soccer players with less fat mass in favour of lean mass may perform better than their lower level peers in high-intensity intermittent running, change of direction (COD) speed, vertical jump and sprint performance (Hazir, 2010). A previous study has found that body

CONTACT Athos Trecroci athos.trecroci@unimi.it Department of Biomedical Sciences for Health, Università degli Studi di Milano, Milano 20129, Italy
Supplemental data for this article can be accessed here.

fat percentage and fat-free mass are associated to maximal incremental running, vertical jump-and-reach task and manoeuverability in competitive young soccer players (Esco et al., 2018). The authors concluded that elevated level of fat tissue together with lower level of muscle mass may negatively affect physical performance in youth soccer. It has been also demonstrated that somatotype and anthropometric variables are associated to repeated sprint ability performance (mean and best time) in young players of different level of competition (Campa, Semprini et al., 2019). Campa, Semprini, et al. (2019) observed that low body fat on upper arm, tight and calf areas was related to better mean and best time in elite and sub-elite soccer players. Moreover, the authors found that high calf muscle area was related to better mean and best time.

Overall, having an optimal body composition and favourable anthropometric variables might be advantageous for young player not only to develop high levels of muscle force and power but also to move their body more efficiently (Lloyd et al., 2015). Results from the study of Campa, Piras et al. (2019) revealed that team sport athletes (e.g. soccer players) with high body fat (body mass index, upper arm fat area, thigh fat area and calf fat area) showed poor movement patterns measured by the functional movement score. This information may have reasonable implications on understanding how soccer players could improve their ability to perform more complex soccer-related movements, such as COD (Li et al., 2020; Sattler et al., 2015). While some research has established significantly relationships between anthropometry, somatotype and body composition with aerobic fitness (Campa, Semprini et al., 2019; Esco et al., 2018; S. Gil et al., 2007, S. M. Gil et al. 2007; Teixeira et al., 2015), repeated sprint ability (Campa, Semprini et al., 2019), straight sprint (Campa, Semprini et al., 2019; S. M. Gil et al., 2007, 2007) and vertical jump (Campa, Semprini et al., 2019; S. Gil et al., 2007, S. M. Gil et al. 2007), to date, little information exist on their association with a field-based assessment including COD ability in young elite soccer players.

Recent research investigated the relationship between body composition (total body fat and lean mass), strength and power with speed (10-m and 20-m sprint time) and COD ability (505 test total running time and deficit) in professional female soccer players (Emmonds et al., 2019). The authors failed to observe significant association between total body fat and lean mass with COD ability, although they found it could be predicted with high accuracy ($R^2 \geq 0.999$) when computed together with 10-m and 20-m sprint performance. Moreover, in the latter study, COD ability was evaluated by total running time and deficit variables. However, total running time has been demonstrated not be the most appropriate variable identifying COD ability. As such, the COD deficit variable has been introduced to provide a more isolated measure of COD ability, without being biased by linear sprint capacity (Nimphius et al., 2016). Unfortunately, there is a dearth of information regarding potential association between field-based physical performance including COD ability (measured by COD deficit) and anthropometric features (including derived body composition variables). Given the importance of COD performance in soccer, such additional knowledge might be useful to develop training programme to improve the athleticism of elite young soccer players.

Therefore, the aim of this study was to examine the association between anthropometric characteristics with physical performance including change of direction ability in elite young soccer players. Data mining approaches such as Machine Learning permits to develop algorithms based on mathematical models able to discover multidimensional linear and non-linear patterns into the data. Thus, using a machine learning approach, this study aimed to determine the overall contribution of anthropometric features to predict physical performance by a multidimensional approach.

Materials and methods

Study design

An observational study design was adopted to evaluate the contribution of anthropometry to predict physical performance in a group of elite soccer players from a professional youth soccer academy. The testing assessments were performed in June. At the time of the study, all players had at least 3 years of experience in soccer training and performed 3–4 regular training sessions per week (about 90–120 min per session) and played one official soccer game per week. The participants were advised to abstain from using dietary supplements before and the day of the study and to refrain from all training activities except for the tests included in the experimental design. Participants were also instructed to avoid caffeinated drinks on the two days prior to the testing session. The testing day were organized in two moments: the first moment occurred in the morning (10.00 am), 2 hours after a standardized breakfast, while the second moment occurred in the afternoon (3.00 pm), 3 hours after a standardized lunch.

At the testing day, all subjects followed their dietary routine consisting of a standardized breakfast (65%, 20%, 15% of total energy intake composed of carbohydrates, fat and protein, respectively), and a standardized lunch (65%, 20%, 15% of total energy intake composed of carbohydrates, fat and protein, respectively) (Balsom et al., 1999).

Participants

Sixteen males under 15 elite soccer players (ages 14.3 ± 1.0 years, body weight 63.2 ± 6.8 Kg, height 176.0 ± 5.5 cm, BMI 20.4 ± 1.4 kg/m^2, endomorphy 2.1 ± 0.5, mesomorphy 4.2 ± 0.9 and ectomorphy 3.8 ± 0.7, fat mass 10.56 ± 1.44%), from the same Italian professional club competing in the first division (Serie A), voluntarily participated in the study. After a well description of the study and an illustration of the procedures, all participants verbally agreed to the testing conditions. Written consent was obtained before to start the study from participant's parents after being fully informed about the experimental procedures, aims, and potential risks of the study. The study procedures were approved by the Ethics Committee of the local University.

Anthropometry features

The anthropometry variables included in the study were body mass, height, mid-upper arm circumference (MUAC), waist (WC) and hip (HC) circumferences, suprapatellar girths of the right (RSPG) and left legs (LSPG) and 8 skinfold thicknesses (triceps, subscapular, biceps, iliac crest, supraspinal, abdominal, anterior thigh and medial calf). Additionally, the sum of 2 (Σ2 = triceps + subscapular), 4 (Σ4 = anterior thigh + abdominal + triceps + medial calf), 7 (Σ7 = triceps + subscapular + iliac crest + supraspinal + abdominal + anterior thigh + medial calf), 9 (Σ9 = triceps + subscapular + biceps + iliac crest +

supraspinal + abdominal + anterior thigh + medial calf + pectoral) and 10 ($\Sigma 10$ = triceps + subscapular + biceps + iliac crest + supraspinal + abdominal + anterior thigh + medial calf + pectoral + axillar) skinfolds thickness were considered for the analysis. All anthropometric measurements were profiled by an accredited anthropometrist following the International Society Advancement Kinanthropometry guidelines. Height was recorded to the nearest 0.1 cm with a standing stadiometer (Seca 217, Basel, Switzerland) and body mass was measured to the nearest 0.1 kg with a high-precision mechanical scale (Seca 877, Basel, Switzerland). Body mass index (BMI) was calculated as the ratio of body mass to height squared (kg/m^2). Girths were measured to the nearest 0.1 cm with an anthropometric tape (Lufkin executive thinline, W606ME). Skinfold thicknesses were measured to the nearest 0.1 mm with a skinfold caliper (Holtain Ltd, Crymych, UK). For each anthropometrical point considered, 3 non-consecutive measurements were performed in order to compute the average. The technical error of measurement scores was required to be within 5% agreement for skinfolds and within 1% for breadths and girths (Ackland et al., 1997). Derived data for muscle area of upper arm (AMA) and thigh (TMA), and for fat area of upper arm (AFA) and thigh (TFA) were calculated according to literature (AR Frisancho, 1981). Corrected arm muscle area (AMA_{corr}) was obtained from arm muscle area values following the Heymsfield's equation (Heymsfield et al., 1982). Additionally, Arm muscle circumference (AMC) was obtained by the formula previously adopted in literature (MUAC – π * triceps skinfold) (AR Frisancho, 1981). The new youth soccer-specific prediction equation recently developed by Munguia-Izquierdo et al. (2018) was used to obtain fat mass from two skinfold sites: triceps and supraspinal. Fat-free mass (FFM) data were obtained by subtracting fat mass from body mass to obtain to the nearest of 0.1 kg. Somatotype components were also calculated according to Heath-Carter method (Carter & Heath, 1990). After completing the anthropometric and body composition assessment, the players were tested for physical performance. All players were familiar with all physical tests as encompassing a field-based soccer-specific test battery based on sprint, COD and jump ability, and aerobic fitness capacity.

Physical performance

All players underwent a field-based testing session including countermovement jump (CMJ), 10-m and 20-m sprint, 90° change of direction test (90 COD), and yo-yo intermittent recovery test level 1 (Yo-Yo IRT 1) with the same order. This testing order was chosen to avoid potential fatigue-related effects planning an adequate recovery period of 10 min between each test (A. Trecroci et al., 2018; A. Trecroci, Longo, et al., 2019; A. Trecroci, Porcelli et al., 2019). Running time performance for 10-m and 20-m sprint, 90 COD and S90 was obtained by an electronic timing gates system (Microgate, Bolzano, Italia) with the gates fixed at 0.7 m above the ground and placed 0.3 m back from the starting line. All tests were conducted outdoor on an artificial turf except to CMJ that was performed inside a gym.

Countermovement jump (CMJ)

The Optojump next system (Microgate, Italy) was used to obtain vertical jump height for each participant. The obtained jump height was then utilized to compare the jumping trials to each other. The participants performed three vertical CMJ trials and the best performance was used for the analysis. During the trial, the participants were asked to jump keeping their hands on the hips and without bending the legs from takeoff and landing phase. A recovery of 2 min was allowed between trials.

10-m and 20-m sprint performance. When ready, the participants were requested to accelerate maximally up to 20 m (with 10 m split time) starting from a standing position. They performed three trials with a 2 min recovery in between. The best performance time recorded at 10 m and 20 m were considered for the analysis.

90° Change of direction test (90° COD test)

Participants were instructed to perform three bouts to each right and left sides interspersed by 2 min of passive recovery with a single change of direction at 90°. The distance between the starting line to the cone and between the cone and the finish line was 5 m each. At the turning point, all subjects were instructed to change direction using the same side-step technique to avoid any influence due to different COD execution technique (Rouissi et al., 2015). In case of hitting or touching the cone at the turning point, the subject was stopped and invited to repeat the bout after an adequate recovery (2 min). The 90° COD performance was measured by the total running time to complete the 5 m + 5 m course, and the fastest trial of each direction was used for further analysis. Furthermore, COD deficit was calculated by subtracting the 10 m sprint time from the total time (Cuthbert et al., 2017; Nimphius et al., 2016). Based on the recommendations of Nimphius et al. (2016), we decided to employ COD deficit within the analysis while reporting total running time only for descriptive purposes.

Yo-Yo intermittent recovery test level 1

The Yo-Yo IRT1 test will be performed after 15 min of standardized warm-up. The test consists of 2 × 20-m shuttle runs at increasing speeds, interspersed with 10 s of active recovery, controlled by audio signals. The test terminates when the subject is no longer able to maintain the required speed and the distance achieved will records as result (Bangsbo et al., 2008).

Statistical analysis

The assumption of normality was verified by the Shapiro Wilk's test for each variable. Relative and absolute reliability were assessed for 10-m and 20-m sprint and 90° COD tests using the Intra-class Correlation Coefficient (ICC) based on average measurements (ICC 3, k) and standard error of the measurement (SEM), respectively. Statistical analysis was performed using the IBM SPSS Statistics software (v. 21, New York, U.S.A.). Decision Tree Regression (DTR), Random Forest Regression (RFR), Extra Tree Regression (ETR), AdaBoost Regression (ABR) and Gradient Boosting for Regression (GBR) models were used to predict which of the anthropometric features (independent variables) is of importance for each physical performance (dependent variables). To validate the prediction of these regression models, a leave-one-out cross-validation approach was employed. This represents a K-fold cross validation that take to its logical extreme, i.e. with K equal to N (number of players). That means that N separate times, the regression models are trained across the data except for one player and a prediction is made for that player. In order to control for possible type 1 error, due to the high number of independent features, we reduced the independent features space in each K-fold by selecting independent variables showing a correlation coefficient higher than 0.47 ($p < 0.5$ with df = 16). The root mean squared error (RMSE), the coefficient of determination (r^2) and the Pearson correlation coefficient (*r*) were computed between predicted and observed values and used to evaluate the goodness of each model prediction (for detailed information, see the supplementary data/supporting information). Correlation coefficient significance was set at 0.47 in accordance with the degree of freedom (*N* = 16) (Kazemitabar et al., 2017). Gini importance (GI) were computed for each independent variable used in the model to predict the dependent one. GI calculates each independent variable importance as the sum over the number of splits (across all decision trees) that include the independent variables itself, proportionally to the number of samples it splits. A GI less than 0.05 was not considered statistically significant. The overall machine learning analysis was programmed using the software Python 3. Based on the GI values, the corresponding independent variables were employed in a successive regression analysis to quantify the extent of its association with the dependent variable. Data are reported as mean ± SD. Ninety-five percent confidence intervals were calculated and reported (95% CI) for each data set. An alpha threshold of $p < 0.05$ was set to identify statistical significance.

Results

ICC values showed excellent reliability in 10-m sprint (ICC = 0.93, 95% CI = 0.81 to 0.97; SEM = 0.02 s), 20-m sprint (ICC = 0.96, 95% CI = 0.89 to 0.98; SEM = 0.03 s), 90° COD test (ICC = 0.94, 95% CI = 0.86 to 0.96; SEM = 0.02 s) and CMJ (ICC = 0.97, 95% CI = 0.92 to 0.99; SEM = 0.99 cm). Data collection of physical performance, anthropometry, skinfolds and body composition is shown in Table 1. Table 2 shows the performance of the best regression model for each dependent variable and the mean ± SD of the GI for the independent variables to predict the dependent variable in the model. ETR showed the highest accuracy and the lowest error to predict 10-m sprint (AMA_{corr}, FFM, AMC and MUAC), 20-m sprint (AMC, AMA_{corr}, FFM and MUAC), and Yo-Yo IRTL 1 (RSPG, LSPG, anterior thigh, WC, HC, Σ4 SKF, TFA and medial calf) with a relative low explained variance (Table 2). In COD deficit and CMJ, ETR showed a poor accuracy to predict the dependent variable from anthropometric characteristics expressing a low and non-significant correlation ($r < 0.33$). The importance of the independent features in the regression models show a low variability (<6%) meaning that for each fold in cross-validation the importance of them are very consistent (i.e. are always the same showing similar importance in each model). This suggest that the DTs build from the accurate regression models are valid to assess the physical performance by anthropometric features (please see supplemental materials for a better DT interpretation).

Discussion

In the present study, the derived ETR models were able to detect the importance of a group of selected anthropometry features to predict 10-m and 20-m sprint, and Yo-Yo IRTL1 performance in elite young soccer players. However, for COD deficit and CMJ, the low and non-significant *r* values indicated that the derived model ETR, including several anthropometric features, had a poor level of prediction.

Table 1. Overall physical and anthropometric features.

Variables	Mean ± SD	95% CI
Physical performance		
90 COD test – total running time (s)	2.47 ± 0.06	2.43 to 2.50
90 COD test – COD deficit (s)	0.58 ± 0.09	0.52 to 0.63
10-m sprint – running time (s)	1.89 ± 0.09	1.83 to 1.94
20-m sprint – running time (s)	3.23 ± 0.13	3.16 to 3.31
CMJ – jump height (cm)	34.60 ± 5.67	31.57 to 37.62
Yo-Yo IRTL 1 – distance (m)	1528.75 ± 341.52	1346.76 to 1710.74
Anthropometry features		
MUAC (cm)	25.96 ± 1.73	25.04 to 26.89
RSPG (cm)	36.68 ± 1.86	35.68 to 37.68
LSPG (cm)	36.67 ± 1.95	35.63 to 37.71
WC (cm)	73.02 ± 3.21	71.31 to 74.74
HC (cm)	92.00 ± 2.33	90.76 to 93.25
Supraspinal (mm)	6.87 ± 1.91	5.85 to 7.88
Subscapular (mm)	8.81 ± 1.78	7.86 to 9.76
Pectoral (mm)	5.51 ± 1.26	4.84 to 6.19
Axillar (mm)	5.72 ± 0.96	5.21 to 6.23
Biceps (mm)	4.11 ± 0.59	3.79 to 4.42
Triceps (mm)	6.84 ± 1.40	6.09 to 7.59
Iliac crest (mm)	11.55 ± 2.97	9.97 to 13.14
Abdominal (mm)	9.96 ± 3.18	8.26 to 11.65
Anterior thigh (mm)	9.79 ± 1.67	8.90 to 10.68
Medial calf (mm)	6.54 ± 1.02	6.00 to 7.08
2 SKF (mm)	15.65 ± 2.59	14.27 to 17.03
4 SKF (mm)	33.03 ± 5.08	30.33 to 35.74
7 SKF (mm)	61.02 ± 11.11	55.10 to 66.94
9 SKF (mm)	70.64 ± 12.28	64.09 to 77.18
10 SKF (mm)	76.35 ± 13.11	69.37 to 83.34
AFA (cm^2)	8.48 ± 1.64	7.60 to 9.35
AMA_{corr} (cm^2)	35.42 ± 6.99	31.70 to 39.14
AMC (cm)	23.82 ± 1.87	22.82 to 24.82
TFA (cm^2)	25.79 ± 5.17	23.04 to 28.55
TMA (cm^2)	201.96 ± 21.62	190.43 to 213.48
TMC (cm)	50.31 ± 2.73	48.86 to 51.77
Fat mass (%)	10.56 ± 1.44	9.80 to 11.33
FFM (kg)	56.57 ± 6.16	53.29 to 59.86

COD = change of direction speed, CMJ = countermovement jump, Yo-Yo IRTL 1 = yo-yo intermittent recovery test level 1, MUAC = mid-upper arm circumference, RSPG = right suprapatellar girth, LSPG = left suprapatellar girth, WC = waist circumference, HC = hip circumference, Σ2 SKF = sum of 2 skinfolds, Σ4 SKF = sum of 4 skinfolds, Σ7 SKF = sum of 7 skinfolds, Σ9 SKF = sum of 9 skinfolds, Σ10 SKF = sum of 10 skinfolds, AFA = arm fat area, AMA_{corr} = corrected arm muscle area, AMC = arm muscle circumference, TFA = thigh fat area, TMA = thigh muscle area, TMC = thigh muscle circumference, FFM = fat-free mass, SD = standard deviation, CI = confidence interval.

In soccer, performing COD efficiently is vital as provides players more chances to evade/mark an opponent, to create space for her/his teammates and to score a goal (A. Trecroci et al., 2018). According to the model proposed by Sheppard and Young (2006), anthropometry is a sub-component, which could play a role in influencing COD performance in team sports athletes. Regardless specific kinematic elements (e.g. height of the athlete's centre of gravity), body fat level is one of the main anthropometric factors influencing COD performance (Sheppard & Young, 2006). Of note, if we look at the ETR model for COD deficit, fat mass appears one of the most predictor capable to affect its performance at a greater extent compared with the others features (with a coefficient < 0.045). On the other hand, the poor value of *r* does not allow data to be interpreted with a good prediction. A possible explanation of such a controversial outcome may be attributable to the nature of COD deficit itself. Recently, COD deficit has been proposed

Table 2. The importance of each independent variable and the related best regression model.

Dependent variables	Model	RMSE	r^2	r	Importance		Linear regression	
					Predictors	GI ± SD	Coefficient	SSE
90 COD test COD Deficit (s)	ETR	0.10	0.11	0.33	AMA_{corr} (cm^2)	0.19 ± 0.04	0.008	0.004
					WC (cm)	0.18 ± 0.06	0.018	0.008
					Fat mass (kg)	0.18 ± 0.05	0.045	0.021
					MUAC (cm)	0.15 ± 0.05	0.038	0.015
					AMC (cm)	0.15 ± 0.06	0.032	0.013
					FFM (kg)	0.09 ± 0.03	0.009	0.004
					HC (cm)	0.07 ± 0.03	0.023	0.010
Sprint 10 m (s)	ETR	0.09	0.25	**0.50***	AMA_{corr} (cm^2)	0.36 ± 0.05	−0.009	0.005
					FFM (kg)	0.25 ± 0.04	−0.009	0.006
					AMC (cm)	0.21 ± 0.05	−0.033	0.020
					MUAC (cm)	0.17 ± 0.03	−0.037	0.022
Sprint 20 m (s)	ETR	0.12	0.32	**0.56****	AMC (cm)	0.37 ± 0.06	−0.048	0.035
					AMA_{corr} (cm^2)	0.27 ± 0.03	−0.013	0.009
					FFM (kg)	0.21 ± 0.04	−0.013	0.010
					MUAC (cm)	0.15 ± 0.04	−0.049	0.037
CMJ (cm)	ETR	6.36	0.05	0.23	AMC (cm)	0.30 ± 0.04	1.898	1.256
					AMA_{corr} (cm^2)	0.29 ± 0.04	0.507	0.337
					FFM (kg)	0.24 ± 0.05	0.507	0.375
					MUAC (cm)	0.17 ± 0.04	1.982	1.348
Yo-Yo IRTL 1 (m)	ETR	250.58	0.43	**0.66****	RSPG (cm)	0.22 ± 0.08	−116.289	55.466
					LSPG (cm)	0.20 ± 0.05	−128.936	55.839
					anterior thigh (mm)	0.16 ± 0.07	−116.894	60.260
					WC (cm)	0.15 ± 0.05	−64.189	31.742
					HC (cm)	0.10 ± 0.05	−94.933	44.679
					4 SKF (mm)	0.07 ± 0.02	−34.006	19.210
					TFA (cm^2)	0.06 ± 0.03	−35.270	19.097
					medial calf (mm)	0.04 ± 0.03	−187.611	98.230

COD = change of direction speed, CMJ = countermovement jump, Yo-Yo IRTL 1 = yo-yo intermittent recovery test level 1, ETR = extra tree regression model, RMSE = root mean squared error, GI = gini importance, SD = standard deviation, SSE = standard error of estimate, MUAC = mid-upper arm circumference, RSPG = right suprapatellar girth, LSPG = left suprapatellar girth, WC = waist circumference, HC = hip circumference, Σ4 SKF = sum of 4 skinfolds, AMA_{corr} = corrected arm muscle area, AMC = arm muscle circumference, FFM = fat-free mass, TFA = thigh fat area.
*$p < 0.05$ and **$p < 0.01$ Correlation coefficient significance set at 0.47 in accordance with the degree of freedom.

as a more suitable measure of COD ability than total running time, because of its peculiarity of not being affected by sprint capacity (Nimphius et al., 2016). It might be that other factors rather than fat mass are of importance COD ability prediction. Overall, the fact that ETR had a poor level of prediction may inform about the fact that the selected anthropometric features are unlikely to explain COD performance measured by COD deficit. In support of this notion, the study of Emmonds et al. (2019) did not found significant association between total body fat and lean mass with COD ability measured by COD deficit. It can be assumed that changing direction efficiently imply not only greater (lateral) force production, but also an optimal decrease of an individual's body momentum through dynamic balance and braking strategy (adjusting stride length and frequency) (Dos'Santos et al., 2017). Thus, other factors (e.g. inter- and intra-limb coordination, whole-body balance, rate of force development), rather than anthropometric features, may possibly concur to its prediction (Emmonds et al., 2019; Young et al., 2002). On the other hand, further studies are warranted to establish potential association between anthropometry and COD deficit in a sample of elite youth soccer players.

In the present study, it was found that AMA_{corr} and AMC were the most important predictors of 10-m and 20-m, respectively. Considering the coefficients in Table 2, for a unit increase of 1 cm^2 of AMA_{corr} and 1 cm of AMC, there is a decrease of 0.009 s and

0.048 s in 10-m and 20-m sprint time, respectively, suggesting that the size of upper limb would be associated with sprint performance. This seems to be in line with the study of Barbieri et al. (2017). The authors reported top sprinters had significantly greater relaxed and contracted upper arm circumference than their slower peers. The arms drive helps to counterbalance the rotary momentum of the legs leading the body to a more stable position (Macadam et al., 2018). Moreover, according to Macadam et al. (2018), the arms has been observed contributing up to 10% of the total vertical propulsive forces during sprinting. Intuitively, the contribution of upper muscle arms in terms of an increased arm size appear to influence sprint running time possibly via an improved balance and body posture of the upper body (lean position) that would influence step length with a consequent speed increase (Ralph, 1981). Furthermore, it appears that sprint performance would be better in individuals with less fat and more lean body mass (FFM). Although FFM did not represent the most important predictor, it was selected by the ETR model within the features with the higher GI. This would suggest that changes in FFM are associated to changes in sprinting over 10-m and 20-m distance. This result is also in accordance with the study of Barbieri et al. (2017). The authors found that less ectomorphic athletes, having a greater FFM, presented a different sprint performance compared with their peers with less FFM. Take all together, these findings indicate that anthropometric features, especially those focusing on the size of the upper limb, would useful to explain sprint performance in a sample of young elite soccer players. Therefore, practitioners may be encouraged integrating AMA_{corr}, AMC and MUAC measures to their field-based assessment to better interpret any potential change in players' sprint performance.

As regards Yo-Yo IRTL, the most important predictors outlined by the ETR model were RSPG, LSPG, anterior thigh, WC, HC, Σ4 SKF, TFA and medial calf. Taken all together, these results indicate that lean lower limbs (e.g. low anterior thigh SKF) would be associated to a better aerobic fitness capacity. Moreover, it would also appear that focusing on a few number of measures (Σ4 SKF, including thigh and calf SKFs), rather than assessing >4 SKFs, may provide informative and practical data linked to the Yo-Yo IRTL 1 outcomes. On the other hand, as outlined by the ETR model, their GI was inferior than RSPG and LSPG, which were similar. According to the regression output, our data would indicate that increases in bilateral suprapatellar girths were associated to a decrease in Yo-Yo IRTL 1 distance. Unfortunately, given no comparable data has been published on the association between RSPG or LSPG and aerobic fitness, any interpretation is arduous, and it is likely to mislead. Changes in suprapatellar girth have been observed to give information on the suprapatellar bulk (muscle mass) of the vastus muscles, and it was previously used in patients recovering from anterior cruciate ligament reconstructive surgery (Soderberg et al., 1996). However, suprapatellar girths are nothing but a measure of circumference, which unlikely transfer information on muscle mass, especially without a measure of SKF in situ. Additional studies are warranted to add comparative data in literature on the use of RSPG and LSPG in combination with other measures (e.g. suprapatellar SKFs) to predict aerobic fitness and assessing their association with physical performance in youth elite soccer.

Regarding CMJ, the ETR model showed a low prediction accuracy suggesting that the anthropometric features did not related to vertical jump height. This result should be taken with caution as a dearth of information exist on the use of specific upper-body anthropometric measures (i.e. AMC, AMA_{corr}, MUAC) to predict vertical jump height. In the study of Reeves et al. (2008), results showed that both upper arm length and size were not strong predictors of vertical jump performance in 17 healthy active subjects. The authors concluded that other parameters (e.g. strength and power levels, flexibility, balance and motor coordination), rather than the arm-related features, could represent more capable predictors of vertical jump height. However, with little data in support of this consideration, conclusive interpretations are not allowed. Of note, within the selected features, the ETR included FFM within the important predictors of CMJ as previously outlined in 25 young elite soccer players. It might be that our small sample size could represent a further explanation for such an outcome even though it appeared appropriate for detecting the best predicting models by machine learning approach.

This study presents two main limitations that should be acknowledged. First, body composition (e.g. fat mass) was assessed using derived parameters from direct anthropometric measures (sum of several SKFs) rather than with a laboratory technique (i.e. total body dual-energy X-ray absorptiometry, DXA). In fact, it should be noted that, considering body mass in a three-compartment view (FM, bone mineral content and lean soft tissue), DXA has been recognized as an optional and accurate technique for determining FM) and FFM in sports field (Esco et al., 2018). Additional studies incorporating DXA are needed to estimate body composition with more accuracy and investigating its potential association with COD ability expressed by COD deficit. Second, even though the present sample size was in line with a previous similar study (Emmonds et al., 2019), its small statistical power limits the interpretation of the results and future studies should adopt larger sample sizes. Anyhow, involving elite highly trained athletes may reduce the within-group variability allowing coaches and practitioners to infer field-based information of the highest standard.

Conclusions

This study showed the anthropometric features are important predictors of sprint performance and aerobic fitness in a sample of youth elite soccer players. According to the present study, the model predicting COD deficit and CMJ presented a poor level of prediction, suggesting that other factors, rather than anthropometric features, may concur to predict their changes in performance. The use of machine learning may be encouraged to determine which anthropometric features are of importance to predict players' performance. This would allow practitioners to monitor selectively a few numbers of variables in an attempt to maximize their sprint and aerobic performance.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

F. Marcello Iaia http://orcid.org/0000-0003-0981-6471

References

Ackland, T. R., Schreiner, A. B., & Kerr, D. A. (1997). Absolute size and proportionality characteristics of World Championship female basketball players. *Journal of Sports Sciences*, *15*(5), 485–490. https://doi.org/10.1080/026404197367128

Balsom, P., Wood, K., Olsson, P., & Ekblom, B. (1999). Carbohydrate intake and multiple sprint sports: With special reference to football (soccer). *International Journal of Sports Medicine*, *20*(1), 48–52. https://doi.org/10.1055/s-2007-971091

Bangsbo, J., Iaia, F. M., & Krustrup, P. (2008). The Yo-Yo intermittent recovery test: A useful tool for evaluation of physical performance in intermittent sports. *Sports Medicine*, *38*(1), 37–51. https://doi.org/10.2165/00007256-200838010-00004

Barbieri, D., Zaccagni, L., Babić, V., Rakovac, M., Mišigoj-Duraković, M., & Gualdi-Russo, E. (2017). Body composition and size in sprint athletes. *The Journal of Sports Medicine and Physical Fitness*, *57*(9), 1142–1146. https://doi.org/10.23736/S0022-4707.17.06925-0

Campa, F., Piras, A., Raffi, M., & Toselli, S. (2019). Functional movement patterns and body composition of high-level volleyball, soccer, and rugby players. *Journal of Sport Rehabilitation*, *28*(7), 740–745. https://doi.org/10.1123/jsr.2018-0087

Campa, F., Semprini, G., Júdice, P., Messina, G., & Toselli, S. (2019). Anthropometry, physical and movement features, and repeated-sprint ability in soccer players. *International Journal of Sports Medicine*, *40*(2), 100–109. https://doi.org/10.1055/a-0781-2473

Carter, J., & Heath, B. (1990). *Somatotyping – Development and applications*. Cambridge University Press.

Cuthbert, M., Thomas, C., Dos'Santos, T., & Jones, P. A. (2017). The application of change of direction deficit to evaluate cutting ability. *Journal of Strength and Conditioning Research*, *33*(8), 2138–2144. https://doi.org/10.1519/JSC.0000000000002346

Dos'Santos, T., Thomas, C., Jones, P. A., & Comfort, P. (2017). Mechanical determinants of faster change of direction speed performance in male athletes. *Journal of Strength and Conditioning Research*, *31*(3), 696–705. https://doi.org/10.1519/JSC.0000000000001535

Emmonds, S., Nicholson, G., Begg, C., Jones, B., & Bissas, A. (2019). Importance of physical qualities for speed and change of direction ability in elite female soccer players. *Journal of Strength and Conditioning Research*, *33*(6), 1669–1677. https://doi.org/10.1519/JSC.0000000000002114

Esco, M., Fedewa, M., Cicone, Z., Sinelnikov, O., Sekulic, D., & Holmes, C. (2018). Field-based performance tests are related to body fat percentage and fat-free mass, but not body mass index, in youth soccer players. *Sports*, *6*(4), 105. https://doi.org/10.3390/sports6040105

Frisancho, A. R. (1981). New norms of upper limb fat and muscle areas for assessment of nutritional status. *The American Journal of Clinical Nutrition*, *34*(11), 2540–2545. https://doi.org/10.1093/ajcn/34.11.2540

Gil, S., Ruiz, F., Irazusta, A., Gill, J., & Irazusta, J. (2007). Selection of young soccer players in terms of anthropometric and physiological factors. *The Journal of Sports Medicine and Physical Fitness*, *47*(1), 25–32. https://www.minervamedica.it/en/journals/sports-med-physical-fitness/article.php?cod=R40Y2007N01A0025

Gil, S. M., Gil, J., Ruiz, F., Irazusta, A., & Irazusta, J. (2007). Physiological and anthropometric characteristics of young soccer players according to their playing position: Relevance for the selection process. *The Journal of Strength and Conditioning Research*, *21*(2), 438–445. https://doi.org/10.1519/R-19995.1

Hazir, T. (2010). Physical characteristics and somatotype of soccer players according to playing level and position. *Journal of Human Kinetics*, *26*(1), 83–95. https://doi.org/10.2478/v10078-010-0052-z

Heymsfield, S. B., McManus, C., Smith, J., Stevens, V., & Nixon, D. W. (1982). Anthropometric measurement of muscle mass: Revised equations for calculating bone-free arm muscle area. *The American Journal of Clinical Nutrition*, *36*(4), 680–690. https://doi.org/10.1093/ajcn/36.4.680

Kazemitabar, J., Amini, A., Bloniarz, A., & Talwalkar, A. S. (2017). Variable importance using decision trees. In *Proceeding of the 31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA.

Li, X., Li, C., Cui, Y., & Wong, D. P. (2020). Acute kinematics and kinetics changes to wearable resistance during change of direction among soccer players. *Research in Sports Medicine*. Advance online publication. https://doi.org/10.1080/15438627.2020.1770761

Lloyd, R. S., Oliver, J. L., Radnor, J. M., Rhodes, B. C., Faigenbaum, A. D., & Myer, G. D. (2015). Relationships between functional movement screen scores, maturation and physical performance in young soccer players. *Journal of Sports Sciences*, *33*(1), 11–19. https://doi.org/10.1080/02640414.2014.918642

Macadam, P., Cronin, J. B., Uthoff, A. M., Johnston, M., & Knicker, A. J. (2018). Role of arm mechanics during sprint running: A review of the literature and practical applications. *Strength and Conditioning Journal*, *40*(5), 14–23. https://doi.org/10.1519/SSC.0000000000000391

Munguia-Izquierdo, D., Suarez-Arrones, L., Di Salvo, V., Paredes-Hernandez, V., Alcazar, J., Ara, I., Kreider, R., & Mendez-Villanueva, A. (2018). Validation of field methods to assess body fat percentage in elite youth soccer players. *International Journal of Sports Medicine*, *39*(5), 349–354. https://doi.org/10.1055/s-0044-101145

Nimphius, S., Callaghan, S. J., Spiteri, T., & Lockie, R. G. (2016). Change of direction deficit: A more isolated measure of change of direction performance than total 505 time. *Journal of Strength and Conditioning Research*, *30*(11), 3024–3032. https://doi.org/10.1519/JSC.0000000000001421

Ralph, M. (1981). A kinetic analysis of sprinting. *Medicine & Science in Sports & Exercise*, *13*(5), 325–328.

Reeves, R. A., Hicks, O. D., & Navalta, J. W. (2008). The Relationship between upper arm anthropometrical measures and vertical jump displacement. *International Journal of Exercise Science*, *1*(1), 22–29. https://pdfs.semanticscholar.org/fabc/133278a1ccdb1d7cc407181586612c36d220.pdf

Rouissi, M., Chtara, M., Owen, A., Chaalali, A., Chaouachi, A., Gabbett, T., & Chamri, K. (2015). "Side-stepping maneuver": Not the more efficient technique to change direction amongst young elite soccer players. *International Journal of Performance Analysis in Sport*, *15*(2), 749–763. https://doi.org/10.1080/24748668.2015.11868827

Sattler, T., Sekulić, D., Spasić, M., Perić, M., Krolo, A., Uljević, O., & Kondrič, M. (2015). Analysis of the association between motor and anthropometric variables with change of direction speed and reactive agility performance. *Journal of Human Kinetics*, *47*(1), 137–145. https://doi.org/10.1515/hukin-2015-0069

Sheppard, J. M., & Young, W. B. (2006). Agility literature review: Classifications, training and testing. *Journal of Sports Sciences*, *24*(9), 919–932. https://doi.org/10.1080/02640410500457109

Soderberg, G. L., Ballantyne, B. T., & Kestel, L. L. (1996). Reliability of lower extremity girth measurements after anterior cruciate ligament reconstruction. *Physiotherapy Research International*, *1*(1), 7–16. https://doi.org/10.1002/pri.43

Teixeira, A., Valente-dos-Santos, J., Coelho-e-Silva, M., Malina, R., Fernandes-da-Silva, J., Cesar Do Nascimento Salvador, P., De Lucas, R., Wayhs, M., & Guglielmo, L. (2015). Skeletal maturation and aerobic performance in young soccer players from professional academies. *International Journal of Sports Medicine*, *36*(13), 1069–1075. https://doi.org/10.1055/s-0035-1549922

Trecroci, A., Longo, S., Perri, E., Iaia, F. M., & Alberti, G. (2019). Field-based physical performance of elite and sub-elite middle-adolescent soccer players. *Research in Sports Medicine*, *27*(1), 60–71. https://doi.org/10.1080/15438627.2018.1504217

Trecroci, A., Milanović, Z., Frontini, M., Iaia, F. M., & Alberti, G. (2018). Physical performance comparison between under 15 elite and sub-elite soccer players. *Journal of Human Kinetics*, *61*(1), 209–216. https://doi.org/10.1515/hukin-2017-0126

Trecroci, A., Porcelli, S., Perri, E., Pedrali, M., Rasica, L., Alberti, G., Longo, S., & Iaia, F. M. (2019). Effects of different training interventions on the recovery of physical and neuromuscular performance after a soccer match. *Journal of Strength and Conditioning Research*, *27*(1), 60–71. https://doi.org/19.1519/JSC.0000000000003269

Young, W. B., James, R., & Montgomery, I. (2002). Is muscle power related to running speed with changes of direction? *The Journal of Sports Medicine and Physical Fitness*, *42*(3), 282–288. https://www.minervamedica.it/it/riviste/sports-med-physical-fitness/articolo.php?cod=R40Y2002N03A0282