# Importance of fundamental movement skills to predict technical skills in youth grassroots soccer: A machine learning approach

Duncan, M., Eyre, E., Clarke, N., Hamid, A. & Jing, Y

# Importance of fundamental movement skills to predict technical skills in youth grassroots soccer: A machine learning approach

Michael J. Duncan[1] , Emma L. J. Eyre[1], Neil Clarke[1],
Abdul Hamid[2], and Yanguo Jing[1,3]

## Abstract

This study determined the contributors to soccer technical skills in grassroots youth soccer players using a machine learning approach. One hundred and sixty-two boys aged 7 to 14 (mean ± SD = 10.5 ± 2.1) years, who were regularly engaged in grassroots soccer undertook assessments of anthropometry and maturity offset (the time from age at peak height velocity (APHV)), fundamental movement skills (FMS), perceived physical competence, and physical fitness and technical soccer skill using the University of Ghent dribbling test. Coaches rated player's overall soccer skills for their age. Statistical analysis was undertaken, using machine learning models to predict technical skills from the other variables. A stepwise recursive feature elimination with a 5-fold cross-validation method was used to eliminate the worst-performing features and both L1 and L2 regularisation were evaluated during the process. Five models (linear, ridge, lasso, random forest, and boosted trees) were then used in a heuristic approach using a small subset of suitable algorithms to achieve a reasonable level of accuracy within a reasonable time frame to make predictions and compare them to a test set to understand the predictive capabilities of the models. Results from the machine learning analysis indicated that the total FMS score (0 to 50) was the most important feature in predicting technical soccer skills followed by coach rating of child skills for their age, years of playing experience and APHV. Using a random forest, technical skills could be predicted with 99% accuracy in boys who play grassroots soccer, with FMS being the most important contributor.

## Keywords

Association football, children, maturity, motor competence, peak height velocity, talent identification

## Introduction

Soccer remains the most popular sport in the world with over 265 million players engaged in the sport from grassroots levels up.[1] The majority of this engagement comes via grassroots soccer which FIFA define as recreational soccer taking place predominantly in children from the age of 6 years on to promote mass participation in the sport.[2] This definition of grassroots soccer was employed in the current study. Grassroots soccer not only provides a means to engage in physical activity to promote health in children, but also provides the foundation and pathway to more specialised soccer performance via long-term athlete development and talent development programmes employed by professional soccer academies and national governing bodies in various countries across the world.[3,4] Within such programmes there has traditionally been a focus on measuring physical fitness as a predictor of soccer potential[5–9] and the use of physical fitness as a key driver of selection or deselection in junior academy football.[10]

Such an approach ignores the fact that the development of soccer-related motor ability is essential for success in soccer.[11] This is particularly the case for youth players where skill development is a key determinant of success in the sport.[12] It has consequently been recognised that coaches need to consider potential talent from a multidimensional standpoint.[13] Recently, there has been an emerging interest in the importance of fundamental movement skills (FMS) in the development of soccer-related talent in children,[14–17] with this aforementioned work[14–17] suggesting that the development of FMS is important in the development of technical soccer skills.[14] FMS are the basic motor skills that form the foundation of specialised skills needed to engage successfully in sports, games, dance and other contexts of physical activity and are considered the foundation for subsequent sports skills.[18] Importantly, the development of FMS is a feature of governing body coaching awards in soccer.[3,4] However, despite the theoretical basis suggesting that children with better FMS will perform better in sports, there appears to be a theory–practice gap in coaching behaviour and practice from grassroots to elite levels of youth soccer.[19]

Research in the last three years has emerged which suggests FMS are key prerequisites of soccer skills in children and youth. For example, Jukic et al.[15] reported that children classified as first-team players had better FMS, but similar physiological fitness, than those classed as second-team players in a small sample ($N = 23$) of 9- to 10-year-olds. Kokstejn et al.[16] also demonstrated that the relationship between physical fitness and soccer dribbling skills was mediated by FMS in a sample of 40 elite Czech youth players. More recently, Duncan et al.[14] reported that FMS and perceived competence mediated the effect of fitness on technical skills in soccer (comprising dribbling, passing and shooting) in a sample of 70, 7- to 12-year-old boys. Duncan et al.[14] suggested that focusing on physical fitness, without an emphasis on FMS and perceived competence in FMS, is likely to be less effective in the development of technical ability in grassroots soccer.

Despite interest in how factors other than physical fitness, such as FMS, might contribute to soccer performance, only one of these studies[14] considered psychological variables where perceived ability was identified as a key variable in the study of soccer-related skill development.[16] There are also other key variables that have been related to soccer skill development in the literature, such as quartile of birth,[20] maturation,[21] years of experience in soccer training[22] and coach perception of player skill[23] which have either only been considered in isolation in the literature or have not been considered alongside FMS, perceived ability and physical fitness in explaining technical skills in soccer. Machine learning approaches offer potential to address this multidimensional issue as machine learning is better at handling large amounts of input variables.[24] Such machine learning approaches have been used in the

context of soccer performance in predicting winning/losing,[24] and have been successful in predicting injury[17] and physical performance[9] in youth soccer. There is robust evidence demonstrating the use of machine learning to predict playing positions from in-game behaviour in professional players.[25] However, to date, the use of machine learning approaches to predict performance in youth soccer is relatively sparse. Based on a PubMed search at the time of writing, only four studies using machine learning in youth soccer have been published and none have examined prediction of soccer skill.[26,27] Machine learning methods such as neural networks and random forests are capable of discovering multidimensional and non-linear patterns in data.[9] The present study therefore sought to determine the contributors to soccer technical skills in grassroots soccer players aged 7 to 14 years using a machine learning approach.

## Methods

### Participants

One hundred and sixty-two boys aged 8 to 14 years (mean $\pm$ SD of age = $10.5 \pm 2.1$ years, height = $145.1 \pm 14.8$ cm, and body mass = $38.0 \pm 11.4$ kg) who were regularly engaged in grassroots soccer participated in the study following institutional ethics approval (Coventry University Ethics Committee: P131207), informed parental consent and child assent.

To be eligible to participate, children had to be aged between 7 and 14 years and registered (and playing) with a grassroots soccer club with at least 1 year of playing experience prior to taking part and including participation in training and organised fixtures against other grassroots teams within the County FA structure in England. Eligibility criteria also stipulated that participants had to be currently training and playing in grassroots football with a minimum of one training session and one match per week, as this is the common minimum standard for grassroots football in England. The mean $\pm$ SD of years of playing experience for the sample was $4.3 \pm 2.4$ years. Participants in the current study were engaged in two to three grassroots football sessions per week, including one organised fixture against another grassroots soccer club within the same County FA. Participants were recruited from junior grassroots clubs ($n = 4$) within Birmingham County FA via contact with club officials. Players then volunteered to participate, provided they were eligible. In regard to age band distribution, participants were from the following age bands under 8 ($n = 24$), under 9 ($n = 19$), under 10 ($n = 36$), under 11 ($n = 39$), under 12 ($n = 19$), under 13 ($n = 14$), and under 14 ($n = 9$).

## Procedures

All assessments took place over two days, separated by 24 h. On the first day of assessment, psychometric questionnaires

were completed, followed by anthropometric assessment, assessment of FMS and fitness. This was followed on the second day by a technical skills assessment. All assessments were conducted by trained researchers and the participants' soccer club coaches were not involved in any way. Prior to participation parents also provided information regarding the quartile of birth to account for potential relative age effects in subsequent analysis.[28] Assessment (with the exception of anthropometry and self-report methods) took place on a 3G synthetic astroturf pitch as is typically used for grassroots football in the UK. During the assessment period, environmental conditions were stable with no rain, temperature between 17°C and 19°C, and wind speed of 2 miles/h, which is considered 'calm' according to the MET Office.[29]

## Anthropometry and maturity offset

Stature (cm), sitting height (cm) and body mass (kg) were assessed to the nearest 0.1 cm and 0.1 kg using a SECA anthropometer and weighing scales (SECA Instruments Ltd, Hamburg, Germany), respectively. Moore et al.'s[30] prediction equation was used to determine maturity offset using measures of height and body mass as a marker of biological maturation.

## Fundamental movement skills

FMS were assessed using the test of gross motor development - 3rd edition [TGMD-3].[31] The following skills were selected: run, jump, hop, overhand throw, underhand throw, and catch to reflect a balance of locomotor and object control skills, without the inclusion of kick to avoid confounding the assessment of FMS and technical soccer skills. This is congruent with recent research examining the utility of FMS in soccer.[14] Each skill is comprised of 3 to 4 behavioural components and skill mastery on the TGMD-3 requires each component to be present. For example, for the run skill, the behavioural components are: (a) arms move in opposition to elbows with elbows bent, (b) brief period where both feet are off the ground, (c) narrow foot placement landing on heel or toe, and (d) non-support leg bent to approximately 90°.

Trials of each skill were video recorded (Sony Handicam CX405b, Sony, UK) and subsequently edited into single film clips of individual skills with Quintic Biomechanics analysis software v21 (Quintic Consultancy Ltd, UK). Scores from two trials were summed to create a total FMS score (scored 0 to 50) following the recommended TGMD-3 test administration guidelines.[31] Two experienced FMS researchers analysed the video clips after training in two separate 2 to 3 h sessions by watching videoed skills of children's skill performances and rating these against a previously rated 'gold standard' rating. Congruent with prior research,[32] training was considered complete when

each observer's scores for the two trials differed by no more than one component per trial from the instructor score for each skill (>80% agreement).[28] We performed inter- and intra-rater reliability analysis for all skills between the two raters on 10% of all the videos. Intraclass correlation coefficients for inter- and intra-rater reliability were .92 and .98, respectively, implying satisfactory inter- and intra-rater reliability.

## Perceived competence

The Perceived Physical Ability Scale for Children (PPASC) was used to assess perceived competence.[33] The PPASC is a valid and reliable tool for children of the ages taking part in this study which assesses physical self-efficacy.[33] It is a 6-item measure comprising questions reflecting strength and coordinative abilities. Items are structured in response scales with a 1 to 4 format. Labels are attached to each point of the response scales to assist in giving meaning to the items for the children. For example, scores for the first item range from 1 (I run very slowly) to 4 (I run very fast). The children were asked to think of themselves when playing/training in soccer and were asked to choose one of the four sentences that best represented their perceived ability. Administration followed recommended guidelines[29] with a potential score of 6 to 24, where higher scores represented a high self-perception of physical competence.

## Physical fitness

Two measures of physical fitness were computed: 15 m sprint time and standing long jump. Each participant's 15 m sprint time was assessed using infra-red timing gates (Fusion Sport, Coopers Plains, Australia) with sprint time converted to speed in m/s. Standing long jump was determined as the distance from take-off to the back of the closest heel on landing and was assessed using a tape measure. For sprint speed and long jump, the best of two trials (fastest speed in m/s; longest jump in centimetres) was selected for analysis. Intraclass correlation coefficients for the two measures of fitness were .9 for the 15 m sprint, .94 and for the standing long jump, indicating good reliability. Testing was completed individually, and we calculated a Z-score for each of our three measures of fitness and summed these Z-scores to create a composite product measure of physical fitness. The use of a composite Z-score was employed as a means to bring together three aspects of physical fitness as a theoretical concept as per guidelines for the use of composite variables.[34] Such a process has also been used previously in the context of youth football performance.[14]

## Technical skills

The University of Ghent (UGent) dribbling test was employed as a measure of technical soccer skills in this

study. This test was chosen given its documented reliability in children,[35] where the reliability of other available soccer skill tests has not been demonstrated.[36] All testing took place on a grass surface with participants wearing soccer boots and was completed with the official ball size for the age band (Size 3 for U8 and U9, Size 4 for U10 to U14) as recommended by the Football Association. Testing was completed individually by the participants to minimise any peer pressure to perform as previously described by Vandendriessche et al.[37] Participants completed a set circuit with four left and four right turns at different angles with a distance between cones ranging between 1 and 2.2 m.[37] Following familiarisation and a practice trial each participant undertook two attempts at the test. Each test was performed as quickly as possible in two steps per test: the first step was made without the ball and the second step was with the ball. The time of each attempt was measured to the nearest 0.01 s with a handheld stopwatch. The time taken to complete the dribbling course without the ball was deducted from the time with the ball to give a skill differential reflecting the dribbling skill (labelled as Z_UGentBall in the machine learning analysis and Supplemental material). This is the outcome variable of interest from this test reflecting dribbling ability. This test has good reliability for the dribbling with the ball component in children.[35]

## Coach rating of player skill

Prior to assessment, the coaches of the participants were asked to rate the football skills of each child. The stem question 'Please rate player [name of player] in terms of their ability relative to their current age group' on a scale of 1 to 10 with 1 being poor and 10 being excellent, was posed to the coaches. Each coach was asked questions with the same stem question but specifically asking them to rate the players' technical football ability (labelled as 'coach rating of technical skill'), social ability (labelled as 'coach rating of social skill'), physical ability (labelled as 'coach rating of physical skill'), the effort made by players in training and games (labelled as 'coach rating of effort') and their overall football ability (labelled as 'coach rating of player skill') and overall coach rating for their age (labelled as 'coach overall rating for age'). Similar assessment methods have been used by coaches to rate the skill ability of youth players in prior work and have demonstrated reliability in rating and validity in predicting soccer skill.[23]

## Birth quartile

Recognising that the relative age effect has been demonstrated in both FMS[33] and soccer skills,[20] each participant provided dates of birth which were subsequently grouped into quartiles of birth (starting at the school year cut-off

date of 1st September) and subsequently labelled: Q1, Q2, Q3, and Q4 with Q1 corresponding to the period of 1st September to 31st November, Q2 from 1st December to 30th February, Q3 from 1st March to 31st May, and Q4 from 1st June to 31st August, as per prior studies in this topic.[38]

## Statistical analysis

The statistical analysis approach undertaken sought to train machine learning models to predict the output variable of interest (Technical soccer skill via the UGent dribbling test). The feature correlation matrix and the UGent dribbling test (Z_UGentBall) distribution analysis that were used to examine the influence of multiple input methods and for eliminating collinear features are presented in Supplemental Figures S1 to S3. An 80%/10%/10% training/validation/test split per age group was employed. The data set and its feature transformation were normally distributed (see Supplemental Figures S2 and S3). The Gaussian distribution test data for the UGent dribbling test (Z_UgentBall) is presented in Supplemental Figures S3 and S4 at the end of the manuscript. As a consequence, a parametric modelling approach was appropriate where UGent dribbling test performance was predicted from 11 other variables; chronological age, years of playing experience, quartile of birth, height, body mass, leg length, age at peak height velocity (APHV), 15 m sprint time, perceived competence, total FMS and coach rating of player skill. To help determine the best-performing predictor variables a recursive feature elimination method was used to eliminate the worst-performing features and any collinear features using linear regression (see Supplemental Figure S4). Five models (linear, ridge, lasso, random forest, and boosted trees) were then used in a heuristic approach to analyse the predicted values against the test set to better understand their potential predictive capabilities (please see Supplemental Figures S4 to S7 for penalty parameters that were used). The initial experiment produced an accuracy range of baseline scores from 34% to 66% (See Table 1). For the ridge regression model, the penalty value of 100 was used (see Supplemental Figure S5). For the lasso regression model, a penalty value of 0.1 was used (see Supplemental

**Table 1.** Prediction accuracy using different machine learning techniques when applying a randomised training and testing approach without accounting for age band within the sample.

| Model | Prediction accuracy( %) |
| --- | --- |
| Linear regression | 56.9 |
| Ridge regression | 47.8 |
| Lasso regression | 34.1 |
| Random forest | 67.7 |
| Boosted trees | 66.1 |

Figure S7). The objective of the heuristic approach was to analyse the data set using a small subset of suitable machine learning algorithms using a stepwise recursive feature elimination with a 5-fold cross-validation method to achieve a reasonable level of accuracy within a reasonable time frame. All analyses were performed in Python (Python Software Foundation. Delaware, USA). The approach used provides a reference point from which to compare various machine learning algorithms and a means to measure performance changes and has been proven effective in other domains related to human movement.[39]

## Results

Features in the data set were selected to avoid correlated features. Correlated features were removed to improve the model's generalisation ability.[40] A correlation matrix depicting the pairwise significance of the correlated features is presented in Supplemental Figure S1 (Leg Length, Mass, UGent with ball, Run, Hop, Catch, Object Control Score and Hoff Passing Test were correlated). A subset of the correlated features with the lowest average correlation was included with the uncorrelated features resulting in 'Age on test' and 'UGent Fastest with ball' being included in the machine learning model. Initially, a first test was run using a linear regression model using accuracy as a measure of prediction success with a random split of the initial dataset between the training data set and test data set. The linear regression model was the first model used for analysing the correlation, identifying both positive and negative correlations, and for evaluating the validity and usefulness of the simple predictive model. We used a random 70%/15%/15% training/validation/test split, respectively. The model produced an initial accuracy level of just over 50%. However, when trying to optimise the model it was overfitting and predicting poorly. Hence, we subsequently employed a random 80%/10%/10% training/validation/test split from each age group for testing, which resulted in a significant improvement in the predictive performance. To improve the machine learning model instead of randomising the choice of the test-set data throughout the entire data set, the machine learning models were set to extract a certain percentage of data points from the test set (e.g. 10% or 20%) from each chronological age band (e.g. under 11s, under 12s, etc.). This process was undertaken as when randomly selecting training samples from the entire data set some age groups are under-represented or not represented in the training at all. By making sure that each age group is equally represented in the training it ensures that the modelling includes the whole sample and, as a consequence, independently represents the whole sample. Randomisation where there may be a possibility that some age groups are under-represented or not represented may be a valid reason for obtaining a model with low prediction accuracy, hence

why we chose to ensure all age bands were represented in the data. After implementing the described split of the original data set into training and testing according to all the age band values there was a significant improvement in prediction accuracy to above 90% irrespective of the machine learning technique that was employed. A 5-fold cross-validation approach was employed based on the sample size and the distribution of samples across the age bands to further help verify the validity of the results (see Supplemental Figure S8).

Following the determination of prediction accuracy, a variable importance analysis was conducted to determine the most important features included in the dataset. The results are represented in Figure 1 and suggest that 'Total FMS' was the most important feature closely followed by 'Coaches Overall Rating for Age', 'Playing Experience', and 'APHV'. The participant's 'Chronological Age on Test' and 'Birth Quartile' were the least important features to account for the variability in participants' soccer performance.

## Results from machine learning training

Table 1 presents the prediction accuracy using different machine learning techniques when applying a randomised training and testing approach without accounting for age bands within the sample. Table 2 presents the prediction accuracy when including each age band in the machine learning training model to predict soccer skill.

## Discussion

This study extends understanding related to technical development in grassroots soccer. This is the first study to evaluate how different machine learning models predict the participant's soccer skill performance (UGent dribbling test) in grassroots youth soccer. The present study identifies important impact factors including FMS, coaches' overall rating for age, years of playing experience, and APHV in predicting technical soccer skill in youth soccer. FMS was the most influential variable in predicting technical skills in youth grassroots soccer. A Random Forest achieved 98.6% prediction accuracy with FMS, coaches' overall rating for age, years of playing experience, and APHV as predictors.

A key strength of this study is the use of machine learning. Machine learning, a branch of artificial intelligence, is becoming more popular as a modelling technique to understand various aspects of soccer performance. This has included predicting injury in youth players[17] and predicting physical performance based on anthropometric variables in youth soccer.[9] The benefit of machine learning is in using mathematical models able to discover multidimensional linear and non-linear patterns in data in an unbiased manner. We recognise that the present study is exploratory
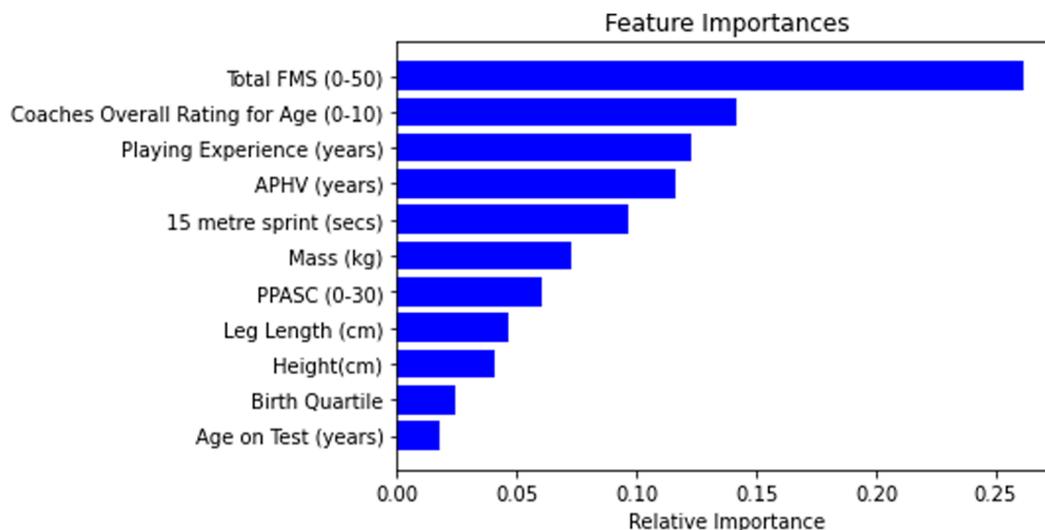
**Figure 1.** Feature importance analysis of the data set (higher relative importance indicates a more important feature in the data set in predicting technical skills).

**Table 2.** Prediction accuracy when including each age band in the machine learning training model to predict soccer skill.

| Model | Prediction accuracy (%) |
| --- | --- |
| Linear regression | 91.9 |
| Ridge regression | 91.2 |
| Lasso regression | 94.1 |
| Random forest | 98.6 |
| Boosted trees | 96.1 |

in examining how technical skills might be explained by other variables that are often cited as important in youth soccer development. There exists robust research using machine learning techniques alongside performance analysis data from videos of games in professional leagues.[25] One outcome of this aforementioned work is a suggestion that machine learning techniques might be applied to youth teams to detect and identify players with particular qualities.[25] The current study aligns with this assertion, albeit different in nature to Garcia-Aliaga et al.,[25] in that we use machine learning techniques to identify some of the predictors that explain technical skills in youth grassroots soccer players.

Total FMS score was the most important predictor of technical skills in the present study. Such a finding agrees with recent research.[14–16] The results of the current study would also support the assertions of the Athletic Skills Model in regard to the importance of FMS and might suggest soccer coaches at grassroots levels may benefit from refocusing away from solely concentrating on soccer-specific practices during training and focusing more broadly on the movement skills that form the foundation of technical skill.[18] That is not to say soccer-specific practices

cannot also develop some FMS but the Athletic Skills Model[18] highlights that a sole focus on sport-specific practice only, will not develop the broad base of movement skills that underpin sport-specific skills. Once the player is proficient in those FMS then a sole sport-specific focus may be beneficial.[18] It is also perhaps not surprising that playing experience and the coach's rating of player skill were the second and third most important predictors of technical skills as prior work has demonstrated that coaches can successfully evaluate player performance[23] and the role of experience in developing skill is well known.[41] It is however perhaps surprising that perceived competence was not as important a contributor to technical skills compared to some of the other variables in our model. Prior work has posited a perception of competence as a key facilitator of movement skill development[42] and soccer skill development in youth.[14] The results of the current study in this respect suggest that perceived competence may not be as important as actual motor competence and experience in the development of technical skills.

The current study has limitations. The participants were all boys and, therefore, the conclusions drawn here should not be inferred for girls. Future work on this topic should focus on girls in particular, especially given the increasing numbers of girls participating in grassroots soccer. Longitudinal research examining the trajectories of development of FMS, technical skills and related factors would also be useful alongside the use of machine learning models to predict soccer talent in children and youth. While we used a measure of technical skills as our outcome variable in the present study, a useful next step would be to establish how the variables used as predictors in the current study might relate to in-game soccer performance.

## Conclusions

This study demonstrates that machine learning models can predict technical soccer skills in boys who play grassroots soccer with up to 99% accuracy, with FMS being the most important contributor in addition to coach rating of skill, playing experience and APHV. We are not suggesting coaches should depart from soccer-specific practices. However, coaches should not solely focus on sport-specific practice if their child athletes have not yet become proficient in their FMS, as FMS provide the foundation for subsequent sport-specific skill development. The machine learning analysis presented in the present study confirms this assertion.

### ORCID iDs

Michael J. Duncan (iD) https://orcid.org/0000-0002-2016-6580
Yanguo Jing (iD) https://orcid.org/0000-0001-9581-4215

### Supplemental material

Supplemental material for this article is available online.

### References

1. FIFA. FIFA big count. FIFA Communications Division. FIFA, Zurich, Switzerland, 2011.
2. FIFA. Grassroots football. Zurich: FIFA, https://www.fifa.com/who-we-are/news/grassroots-goes-live-fifa-com-1462069 (2011, accessed 17 March 2023).
3. The FA. Fundamental movements in the foundation stage. Staffordshire: The FA, http://www.thefa.com/learning/coaching/fundamental-movements-in-the-foundation-phase (2016, accessed 17 March 2023).
4. North J, Lara-Bercial S, Morgan G, et al. *The identification of good practice principles to inform player development and coaching in European youth football.* Zurich: UEFA, 2014.
5. Meylan C, Cronin J, Oliver J, et al. Talent identification in soccer: the role of maturity status on physical, physiological and technical characteristics. *Int J Sports Sci Coaching* 2010; 5: 571–592.
6. Unnithan V, White J, Georgiou A, et al. Talent identification in youth soccer. *J Sports Sci* 2012; 30: 1719–1726.
7. Höner O, Leyhr D and Kelava A. The influence of speed abilities and technical skills in early adolescence on adult success in soccer: a long term prospective analysis using ANOVA and SEM approaches. *PLoS One* 2017; 12: e0182211.
8. Serrano JS, Pizarro AP, García-González L, et al. Evolution of tactical behavior of soccer players across their development. *Int J Perf Anal Sport* 2017; 17: 885–901.
9. Bongiovanni T, Trecroci A, Cavaggioni L, et al. Importance of anthropometric features to predict physical performance in elite youth soccer: a machine learning approach. *Res Sports Med* 2012; 29: 213–224.
10. Noon MR, Eyre ELJ, Ellis M, et al. The influence of recruitment age and anthropometric and physical characteristics on the development pathway of English academy football players. *Int J Sports Physiol Perf* 2020; 16: 199–207.
11. Rampinini E, Impellizzeri FM, Castagna C, et al. Technical performance during soccer matches of the Italian Serie A League: Effect of fatigue and competitive level. *J Sci Med Sport* 2009; 12: 227–233.
12. Huijgen BC, Elferink-Gemser MT, Ali A, et al. Soccer skill development in talented players. *Int J Sports Med* 2013; 34: 720–726.
13. Sieghartsleitner R, Zuber C, Zibung M, et al. Science or Coaches' eye? – Both! Beneficial collaboration of multidimensional measurements and coach assessments for efficient talent selection in elite youth football. *J Sports Sci Med* 2019; 18: 32–43.
14. Duncan MJ, Eyre E, Noon MR, et al. Actual and perceived motor competence mediate the relationship between physical fitness and technical skill performance in young soccer players. *Eur J Sport Sci* 2022; 22: 1196–1203.
15. Jukic I, Prnjak K, Zoellner A, et al. The importance of fundamental motor skills in identifying differences in performance levels of U10 soccer players. *Sports* 2019; 7: 178.
16. Kokstejn J, Musalek M, Wolanski P, et al. Fundamental motor skills mediate the relationship between physical fitness and soccer-specific motor skills in young soccer players. *Front Physiol* 2019; 10: 596.
17. Rommers N, Rössler R, Verhagen E, et al. A machine learning approach to assess injury risk in elite youth football players. *Med Sci Sports Exerc* 2020; 52: 1745–1751.
18. Wormhoudt R, Savelsbergh G, Teunissen J, et al. *The athletic skills model.* London: Routledge, 2017.
19. Cushion C, Ford PR and Williams AM. Coach behaviours and practice structures in youth soccer: implications for talent development. *J Sports Sci* 2012; 30: 1631–1641.
20. Figueiredo P, Seabra A, Brito M, et al. Are soccer and futsal affected by the relative age effect? The Portuguese Football Association case. *Front Psychol* 2021; 12: 679476.
21. Towlson C, Salter J, Ade JD, et al. Maturity-associated considerations for training load, injury risk, and physical performance in youth soccer: one size does not fit all. *J Sport Health Sci* 2021; 10: 403–412.
22. Malina RM, Ribeiro B, Aroso J, et al. Characteristics of youth soccer players aged 13–15 years classified by skill level. *Br J Sports Med* 2007; 41: 290–295.
23. Hendry DT, Williams AM and Hodges NJ. Coach ratings of skills and their relations to practice, play and successful transitions from youth-elite to adult-professional status in soccer. *J Sports Sci* 2018; 36: 2009–2017.
24. Geurkink Y, Boone J, Verstockt S, et al. Machine learning-based identification of the strongest predictive variables of winning and losing in Belgian professional soccer. *Appl Sci* 2021; 11: 2378.
25. Garcia-Aliaga A, Marquina M, Coterón J, et al. In-game behaviour analysis of football players using machine learning techniques based on player statistics. *Int J Sports Sci Coaching* 2020; 16: 148–157.

26. Oliver JL, Ayala F, De Ste Croix MBA, et al. Using machine learning to improve our understanding of injury risk and prediction in elite male youth football players. *J Sci Med Sport* 2020; 23: 1044–1048.

27. Kelly AL, Williams CA, Cook R, et al. A multidisciplinary investigation into the talent development processes at an English football academy: a machine learning approach. *Sports* 2022; 10: 159.

28. Götze M and Hoppe MW. Relative age effect in elite German soccer: influence of gender and competition level. *Front Psychol* 2021; 11: 587023.

29. The MET Office. *Weather and climate data*. London: The MET office, 2022.

30. Moore SA, McKay HA, Macdonald H, et al. Enhancing a somatic maturity prediction model. *Med Sci Sports Exerc* 2015; 47: 1755–1764.

31. Ulrich DA. *Test of Gross Motor Development*. 2nd ed. Austin, TX: PRO-ED, 2001.

32. Barnett LM, Minto C, Lander N, et al. Interrater reliability assessment using the test of gross motor development-2. *J Sci Med Sport* 2014; 17: 667–670.

33. Colella D, Morano M, Bortoli A, et al. A physical self efficacy scale for children. *Soc Behav Person* 2008; 36: 841–848.

34. Song M-K, Lin F-C, Ward SE, et al. Composite variables: when and how. *Nurs Res* 2013; 62: 45–49.

35. Duncan M, Richardson D, Morris R, et al. Test-retest reliability of soccer dribbling tests in children. *J Mot Learn Dev* 2021; 9: 526–532.

36. Ali A. Measuring soccer skill performance: a review. *Scand J Med Sci Sports* 2011; 21: 170–183.

37. Vandendriessche JB, Vaeyens R, Vandorpe B, et al. Biological maturation, morphology, fitness, and motor coordination as part of a selection strategy in the search for international youth soccer players (age 15–16 years). *J Sports Sci* 2012; 30: 1695–1703.

38. Birch S , Cummings L, Oxford SW, et al. Examining relative age effects in fundamental skill proficiency in British children aged 6–11 years. *J Strength Cond Res* 2016; 30: 2809–2815.

39. Gjoreski H, Kaluza B, Gams M, et al. Ensembles of multiple sensors for human energy expenditure estimation. *UbiComp* 2013: 359–362.

40. Chan JY-L, Leow SMH, Bea KT, et al. Mitigating the multicollinearity problem and its machine learning approach. A review. *Mathematics* 2022; 10: 1283.

41. Adolph KE and Franchak JM. The development of motor behavior. Wiley interdisciplinary reviews. *Cogn Sci* 2017; 8.

42. Barnett LM, Morgan PJ, Van Beurden E, et al. A reverse pathway? Actual and perceived skill proficiency and physical activity. *Med Sci Sports Exerc* 2011; 43: 898–904.