

# Morphology, body composition and maturity status of young Colombian athletes from the Urabá subregion: A k-Medoids and hierarchical clustering analysis

DIEGO A. BONILLA<sup>1</sup> , JAVIER O. PERALTA<sup>2</sup>, JHONNY A. BONILLA<sup>2</sup>, WILSON URRUTIA-MOSQUERA<sup>2</sup>, SALVADOR VARGAS-MOLINA<sup>3</sup>, ROBERTO CANNATARO<sup>4</sup>, JORGE L. PETRO<sup>5</sup>

<sup>1</sup>Research Division, DBSS International SAS, Bogotá, Colombia

<sup>2</sup>Agroindustrial, Livestock and Tourism Technology Complex, National Training Service SENA, Antioquia, Colombia

<sup>3</sup>University of Wales Trinity Saint David, Málaga, Spain.

<sup>4</sup>Department of Pharmacy, Health and Nutritional Sciences, University of Calabria, Rende, Italy

<sup>5</sup>University of Córdoba, Montería, Colombia

## ABSTRACT

The Urabá subregion is one of the most prominent cradles of Colombian elite athletes and, therefore, highly recognized within the “*Land of Athletes*” framework of the Colombian Ministry of Sports. In order to contribute to the young talent identification and selection of sports specialization, the aim of this STROBE-based cross-sectional study was to determine the morphological characteristics (MC), body composition (BC) and maturity status (MS) of U16 athletes from this subregion (7 municipalities). Eighty-one young athletes (66 weightlifters, 15 boxers) with at least one regional-competition of experience participated (33F; 48M; 14.9 ± 1.4 years; 62.28 ± 16.6 kg; 162.8 ± 9.9 cm). After parental informed consent, ISAK-standardized anthropometric data were collected during a youth sports championship. Athletes were subdivided in clusters using the PAM (*k*-Medoids clustering) and the bottom-up agglomerative (hierarchical clustering) algorithms. Both clustering methods were based on 55 variables that encompassed MC (raw variables, indices, somatotype), BC (five-compartment model, %BF-equations,  $\Sigma S$ ) and MS (maturity offset, PHV, inter alia). The number of clusters was predefined as  $k = 2$  since was the best solution according to 18 criterion-algorithms (100 bootstrap simulations). Non-parametric tests showed significant differences between sex, sports, municipalities and clusters for certain analysed variables. Internal validity of the clustering showed that sport type might explain the variation in the data; thus, it is noteworthy reasonable to recommend the implementation of unsupervised machine learning strategies along with other supervised methodologies in the identification and characterization of young talents and early sports specialization in Colombian athletes with Olympic projection but further research and support is needed.

**Keywords:** Anthropometry; Somatotype; Biological maturation; Early sport specialization; Youth sports; Sports medicine.

### Cite this article as:

Bonilla, D.A., Peralta, J.O., Bonilla, J.A., Urrutia-Mosquera, W., Vargas-Molina, S., Cannataro, R. & Petro, J.L. (2020). Morphology, body composition and maturity status of young Colombian athletes from the Urabá subregion: A k-Medoids and hierarchical clustering analysis. *Journal of Human Sport and Exercise*, 15(4proc), S1367-S1386. doi:<https://doi.org/10.14198/jhse.2020.15.Proc4.34>

 **Corresponding author.** Research Division, DBSS International SAS, Bogotá, Colombia. <https://orcid.org/0000-0002-2634-1220>

E-mail: [dabonilla@dbss.pro](mailto:dabonilla@dbss.pro)

Abstract submitted to: Spring Conferences of Sports Science. [Costa Blanca Sports Science Events](#), 19-20 June 2020. Alicante, Spain.

JOURNAL OF HUMAN SPORT & EXERCISE ISSN 1988-5202

© Faculty of Education. University of Alicante

doi:10.14198/jhse.2020.15.Proc4.34

## INTRODUCTION

Kinanthropometry (*kine-* movement, *anthropos-* human, and *metron-* measurement) is a discipline that not only encompasses basic measurements, such as body mass, stature, and arm span, but also other important variables that provide important information about morphology and body composition (skinfolds, girths, breadths and lengths) (Esparza-Ros, Vaquero-Cristóbal, & Marfell-Jones, 2019). In this sense, kinanthropometry offers to exercise and health professionals a scientific procedure of acquiring surface anatomical dimensional measurements, which in spite of its limitations is easily accessible, inexpensive, short-in-duration, mobile, and routinely used in many sports (Zemski, Keating, Broad, & Slater, 2018), and allows to analyse the influence of several factors of the training, recovery, and nutrition interventions on morphology and body composition. In fact, the International Olympic Committee research working group on body composition, health, and performance recommends the procedures established by the International Society for the Advancement of Kinanthropometry – ISAK (Larson-Meyer, Woolf, & Burke, 2018), as the entity that developed international standards for anthropometric assessment and an international anthropometry accreditation scheme (ISAK, 2020).

Considering that kinanthropometry might help to study the quantitative interface between structure and function across the lifespan, significant efforts have allowed evaluating the maturity status of the young athletes from anthropometric data (Patel, Nevill, Cloak, Smith, & Wyon, 2019; Sherar, Mirwald, Baxter-Jones, & Thomis, 2005; Söğüt et al., 2019). Maturation is a process that occurs in all tissues, organs and systems of the human body in a time- and biological-dependent manner. This gives the opportunity to identify and track different physiological (e.g., hypothalamic-pituitary-gonadal and hypothalamic-pituitary-adrenal axes) and morphological variables (e.g., pubic and axillary hair, skeletal age) as indicators of progress towards maturity. Hence, maturation is assessed in terms of status (level of maturation at the chronological age of observation, and timing), which although related are not equivalent (Malina, Rogol, Cumming, Coelho e Silva, & Figueiredo, 2015). The importance of understanding the maturity status and how it affects children and youth performance is highly valuable for coaches and sports professionals, considering that it plays an important role in neuromuscular development, physical conditioning and injury-risk among other variables (Beunen & Malina, 2007; Faigenbaum et al., 2009).

In Colombia, the Urabá subregion is considered one of the most strategic sites in terms of sports potential, including young talents identification and sport specialization. Urabá is a coastal region of Antioquia on the Colombian Caribbean Sea with 11,664 square kilometres and almost 700,000 inhabitants. This subregion is characterized by exotic landscapes, great cultural diversity (Afro-descendants, indigenous, mulattos, zambos, whites and mestizos populate this sub-region) and, as aforementioned, tremendous athletic potential (Cardona, 2013). In fact, athletes born or trained in the Urabá area contribute not only to the Antioquia delegation with the 60% of the medals in national championships but also to Colombian delegations with about 20% of the medals in athletics, weightlifting and boxing at the Olympic (IOC, 2016) and Pan American level (Panam Sports, 2019). Thus, the Urabá subregion is highly recognized within the “*Land of Athletes*” framework of the Colombian Ministry of Sports, which is a program that involves all the directorates of this national entity in order to support children and future talents (MinDeporte, 2019); however, there is a need to characterize and extract relevant information related to the talent identification and sports specialization processes since no quality research has been conducted with this population before.

### **Objective**

The aim of this observational study was to determine by the first time the morphological characteristics, body composition and maturity status of young athletes with competition experience from the Urabá subregion

through a machine learning approach. We hypothesize that the variation in the data might be explained by the overlapping between the clusters obtained from the unsupervised algorithm and sexes or sport.

## MATERIAL AND METHODS

### **Study design**

A cross-sectional study of anthropometric data in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology – STROBE guidelines (Vandenbroucke et al., 2009) was performed.

### **Setting**

The study was conducted during the month of October 2018 with the support of the Servicio Nacional de Aprendizaje – SENA. This research is part of the activities related to the "applied research projects in physical activity and sports training sciences", valid from 2016 to 2019, within the framework of the SENNOVA Line 23: Technological Updating and Modernization Program of the Training Centres, with a SIGPS verification code of submission: 981411626632015FD4F61A93C1F4A67C (confirmation number: 20155).

### **Participants**

A total of eighty-eight young boxing and weightlifting athletes (ages 12 to 17) with previous experience in youth sport competitions and participating in the Urabá regional games (*Juegos Regionales de Urabá*) were suitable for eligibility. The athletes were from seven municipalities in the Urabá subregion in Antioquia, Colombia (Apartadó, Carepa, Chigorodó, Mutatá, Necoclí, San Juan de Urabá and Turbo). Informed consent with the respective signature was requested from the coaches and parents of each participant. After reviewing the informed consents, the participants were asked verbally if they agreed with the assessment and were given clarification on their freedom to withdraw from the investigation at any time and without any repercussions. In the consent, detailed information was given about the objective of the study, the measurements to be made and their conditions (comfortable clothing and characteristics of the anthropometric assessment) and the approximate duration of the evaluation. All procedures were developed in accordance with the latest version of the Declaration of Helsinki (World Medical Association, 2002).

### **Variables**

The following continuous variables were measured: body mass (kg), stature (cm), sitting height (cm), wingspan (cm), skinfolds (mm), girths (cm) and breadths (cm). From these variables we derived several anthropometric indices, the somatotype, the body composition (five-compartment model), the maturity offset, the peak height velocity (PHV), the biological maturation classification and the height prediction. Other variables were also collected from the informed consent (age, sex, sport and place of origin).

### **Data sources / measurement**

All the measurements on the young athletes were made at the Combat Coliseum in Carepa, a municipality of the Urabá subregion (Antioquia, Colombia). Four folding screens were installed in the central training area, taking advantage of good lighting and ventilation. In addition, two additional fans were available to mitigate the heat and avoid inconveniences in the measurements (<22 °C, <60% humidity). Evaluations were conducted between 9:00 am and 5:00 pm (GMT-5).

### **Anthropometry**

All anthropometric measurements were carried out in accordance with the international standards for anthropometric assessment published by the International Society for the Advancement of Kinanthropometry – ISAK (Esparza-Ros et al., 2019). The body mass was measured with a digital scale with a 100-gram

graduation (Tanita BC-543, Amsterdam, Netherlands). For the measurement of stature and wingspan, a portable stadiometer with a 1-millimeter graduation was used (Seca 700, Medical Scales and Measuring Systems, Hamburg, Germany), which was fixed to the walls and calibrated with a cable, a plumb bob and a tape measure. A 50 cm high wooden anthropometric box was used to measure the sitting height. The skinfold thickness of the triceps, subscapular, supraspinale, abdominal, front thigh and medial calf were measured with a skinfold calliper calibrated in load cell with constant closing compression of  $10 \text{ g}\cdot\text{mm}^{-2}$  (Gaucho Pro, Rosscraft SRL, Buenos Aires, Argentina). The measurement of the girths (head, arm [relaxed], arm [flexed and tensed], forearm, thorax, chest [mesosternale], waist [minimum], gluteal [hips], thigh [1 cm gluteal], middle thigh and calf [maximum]) was carried out using a non-extensible metal tape of 0.7 mm thickness (Rosscraft SRL, Buenos Aires, Argentina). The measurement of breadths was performed with a Campbell 10 (18 cm) small sliding calliper and a Campbell 20 (54 cm) wide sliding calliper with AP branches (Rosscraft SRL, Buenos Aires, Argentina). Hand length and distal styloids were evaluated with a segmometer SEG4 (Rosscraft SRL, Buenos Aires, Argentina) according to the method established by Visnapuu and Jürimäe (2007). For data analysis, averages of two measurements of each anthropometric variable were calculated and processed. To reduce the technical error of measurement, skinfold measurements were taken in triplicate if the difference between the first and second measurement was greater than 1.0 mm. The intra-observer technical error of measurement of the anthropometrists and research assistants was less than 7.5% for skinfolds and 1.5% for the other measurements, which is considered acceptable by the ISAK recommendations (Norton & Eston, 2018).

### *Morphology*

To perform the morphological comparison and characterization, we analysed the raw data obtained for each anthropometric variable (girths, breadths and skinfolds) and calculate the skinfold-corrected muscle girths according to the expression  $\text{girth} - (\pi \times \text{skinfold})$  (Martin, Spenst, Drinkwater, & Clarys, 1990). Moreover, we derived the following anthropometric indices: i) cormic index, which provides an estimation of relative trunk length, calculated as sitting height/stature (Catikkas, Kurt, & Atalag, 2013); ii) the muscle-to-bone ratio, as the relationship between muscle mass and bone mass, both in kilograms (Holway & Garavaglia, 2009); iii) the adipose-to-muscle tissue ratio, as the ratio of adipose tissue to muscle mass, both in kilograms (Alastrue Vidal et al., 1988); iv) locomotive index, which describes a relationship between the load and musculoskeletal tissues to analyse the efficiency of the locomotor system, calculated as the ratio of the sum of adipose and residual tissues to the sum of muscle and bone mass (Fernández, Jara, & Zamudio, 2017).

The somatotype method of Heath and Carter (1967) was implemented to identify the endomorphy (relative fatness), mesomorphy (musculoskeletal component), and ectomorphy (linearity).

### *Body composition*

A multi-component model of body composition was used to predict skin, adipose, muscle, bone and residual tissue masses (Kerr, 1988). This method has been shown to be appropriate to determine muscle mass (Berral de la Rosa, Rodríguez-Bies, Berral de la Rosa, Rojano Ortega, & Lara Padilla, 2010) and is frequently used as fractionation method to study body composition (Bernal-Orozco et al., 2020; Holway & Garavaglia, 2009). In addition, we estimated the percentage of body fat according to the two skinfold-based equations developed by Slaughter et al. (1988), which have shown to be accurate among non-obese children (Freedman, Horlick, & Berenson, 2013) and are commonly used in Colombian population due to the lack of an specific equation for this population (Alzate Salazar, Ramos Bermúdez, & Melo Betancourt, 2011; Correa & Enrique, 2008; Palma et al., 2021; Petro, Rodríguez Arrieta, & Montenegro Arjona, 2017). For convention, “%FM Slaughter\_1” and “%FM Slaughter\_2” represent the equations that used skinfolds of triceps + subscapular and triceps + calf, respectively, as input. Finally, the sum of six skinfolds ( $\Sigma 6S$ ) was calculated, which is an

absolute measurement (expressed in millimetres) that not only gives information concerning local fat distribution in the body but also provides an index to determine adiposity, since subcutaneous fat reflects the amount of fat present in the adipose tissue (Ballard, Dewanti, Sayuti, & Umar, 2014; Garrido-Chamorro, Sirvent-Belando, González-Lorenzo, Blasco-Lafarga, & Roche, 2012).

#### *Basal metabolic rate*

To report the estimated basal metabolic rate of the young athletes several sex- and age-specific equations were used. On the one hand we calculated the approximate surface area (Du Bois, 1916) and subsequently we input the necessary data for the Fleisch (1951) equations accordingly. On the other hand, the basal metabolic rate was also estimated according to the Schofield (1985) equations.

#### *Maturity status*

Estimation of biological maturation was calculated according to the practical recommendations reported by Sherar et al. (2005), which is based on the valid, non-invasive and simple method developed by Mirwald, G. Baxter-Jones, Bailey, and Beunen (2002). In brief, this method estimates maturity offset (error of 1 year 95% of the time) as follows:

*In Boys* ( $R = 0.94$ ;  $R^2 = 0.891$ ;  $SEE = 0.592$ ):

Maturity Offset =  $-9.236 + 0.0002708 \times [\text{Subisqual length and Sitting Height interaction}] - 0.001663 \times [\text{Age and Subisqual length interaction}] + 0.007216 \times [\text{Age and Sitting Height interaction}] + 0.02292 \times [\text{body mass by stature ratio}]$ .

*In Girls* ( $R = 0.94$ ;  $R^2 = 0.890$ ;  $SEE = 0.569$ ):

Maturity Offset =  $-9.376 + 0.0001882 \times [\text{Subisqual length and Sitting Height interaction}] + 0.0022 \times [\text{Age and Subisqual length interaction}] + 0.005841 \times [\text{Age and Sitting Height interaction}] - 0.002658 \times [\text{Age and body mass interaction}] + 0.07693 \times [\text{body mass by stature ratio}]$ .

Subsequently, the age at PHV was calculated according to the expression: chronological age – maturity offset, which also allowed to categorize the young athlete as early, average or late maturing. Age at PHV refers to the estimated chronological age at maximum rate of growth in height during the adolescent spurt, which begins with acceleration in rate of growth in height (take-off), continues to accelerate until it reaches a peak, and then decelerates, eventually terminating in the late teens/early 20s (Malina et al., 2015). Finally, the prediction of adult height in adolescent children was performed within  $\pm 5.35$  cm 95% of the time in boys and  $\pm 6.81$  cm 95% of the time in girls (Sherar et al., 2005).

#### **Study size**

After the announcement to participate in this study, only the young boxers and weightlifters that fulfil all inclusion criteria were considered as enrolled (non-probability sampling).

#### **Statistical methods**

Descriptive statistics was expressed as mean and standard deviation (95% CI). Normality was assessed with the Shapiro-wilk test. The data were analysed with the Mann-Whitney U and Kruskal-Wallis tests to determine differences between men and women and between municipalities of origin, respectively. The young athletes were subdivided into groups (clusters) using the partitioning around medoids (PAM) algorithm, also known as k-Medoids clustering, which unlike the k-means algorithm considers the median as the centre of a cluster. This has shown to enhance robustness against outliers and reduce noise in the unsupervised machine learning process (Mannor et al., 2011). In addition, we performed a hierarchical clustering with bottom-up approach that was plotted as a dendrogram, where the length of an edge between a cluster and its split is proportional to the dissimilarity between the split clusters, which provides an easy-to-interpret view of the

clustering structure (Wittek, 2014). The number of clusters was determined using 18 criterion algorithms comparing the two-to eight-cluster solutions with the R package 'NbClust'. The internal validation of clustering results was performed with the 'cValid' package, which implements a Dunn Index, Silhouette, and Connectivity to assess the quality of the clustering (Brock, Pihur, Datta, & Datta, 2008). The packages 'Factoextra' and 'potly' were used to visualize clustering results. A significance level of  $p < .05$  was established. All tests were performed in the IBM SPSS v26 (IBM Corp., Armonk, NY, USA) and within the free software environment for statistical computing and graphics R v4.0.2 (Team, 2017).

## RESULTS

The results of this research were reported following the STROBE guidelines for observational studies (von Elm et al., 2014). The results are presented as mean (SD).

### Participants

All data measurements were obtained for 81 young athletes (33F, 48M; 66 weightlifters, 15 boxers;  $14.9 \pm 1.4$  years;  $62.28 \pm 16.6$  kg;  $162.8 \pm 9.9$  cm) from the Urabá subregion (35 Apartadó, 25 Carepa, 15 Chigorodó, 2 Mutatá, 1 Necoclí, 1 San Juan de Urabá, 2 Turbo), which accomplished the inclusion criteria of this study.

### Descriptive data

The general characteristics of the population are shown in Table 1.

Table 1. Characteristics of the young athletes (n = 81) by sex.

Variable	Female	95% IC	Male	95% IC	p-value
	(n = 33)	(min. max)	(n = 48)	(min. max)	
Age (years)	14.67 (1.31)	14.21. 15.14	15.09 (1.48)	14.66. 15.53	.175
Body mass (kg)	58.06 (14.68)	52.85. 63.26	65.18 (17.43)	60.11. 70.24	.016
Stature (cm)	155.86 (6.53)	153.55. 158.18	167.68 (9.04)	165.05. 170.31	<.001
BMI (kg·m <sup>-2</sup> )	23.73 (4.78)	22.03. 25.42	22.96 (4.70)	21.60. 24.33	.322
<b>Morphology</b>					
Sitting height (cm)	82.39 (3.42)	81.17. 83.60	86.25 (5.08)	84.77. 87.72	<.001
B. biacromial	31.78 (2.88)	30.76. 32.81	34.74 (2.87)	33.91. 35.57	<.001
B. biilocrystal	25.64 (3.91)	24.25. 27.03	22.65 (4.48)	24.35. 26.95	.965
B. transverse chest	25.61 (2.14)	24.85. 26.37	26.92 (2.64)	26.15. 27.69	.005
B. A-P chest depth	16.28 (4.18)	14.80. 17.77	17.79 (3.62)	16.74. 18.84	<.001
B. humerus	5.97 (0.60)	5.76. 6.18	6.97 (0.58)	6.80. 7.14	<.001
B. femur	8.60 (0.68)	8.36. 8.84	9.29 (0.79)	9.06. 9.52	<.001
G. head	54.38 (1.84)	53.72. 55.03	55.22 (2.10)	54.61. 55.83	.16
G. arm (relaxed)	27.98 (4.05)	26.54. 29.42	28.42 (4.21)	27.20. 29.64	.513
G. arm (flexed & tensed)	28.83 (4.22)	27.33. 30.33	31.13 (3.88)	30.01. 32.26	.002
G. corrected arm	22.46 (3.43)	21.24. 23.68	23.33 (5.51)	21.73. 24.93	.1
G. forearm (maximum)	24.48 (2.61)	23.56. 25.41	27.20 (2.76)	26.40. 28.00	<.001
G. chest (mesosternale)	84.22 (8.07)	81.35. 87.08	88.89 (9.39)	86.16. 91.62	.002
G. corrected chest	80.34 (7.98)	77.51. 83.17	86.03 (8.53)	83.55. 88.51	<.001
G. waist (minimum)	70.65 (11.47)	66.58. 74.72	74.31 (9.91)	71.43. 77.19	.009
G. corrected waist	65.62 (10.14)	62.02. 69.21	70.97 (8.36)	68.54. 73.40	<.001
G. Gluteal (hips)	92.46 (9.28)	89.16. 95.75	90.32 (9.85)	87.46. 93.18	.226
G. Thigh (1 cm gluteal)	58.80 (8.14)	55.91. 61.69	55.80 (8.51)	53.32. 58.27	.045
G. Middle thigh	51.06 (5.72)	49.03. 53.09	51.53 (8.73)	49.00. 54.07	.641

G. corrected thigh	45.04 (4.61)	43.40. 46.68	48.25 (8.05)	45.91. 50.58	.088
G. Calf (maximum)	32.29 (6.08)	30.13. 34.45	30.68 (8.80)	28.12. 33.23	.882
G. corrected calf	27.46 (6.06)	25.31. 29.61	27.55 (8.58)	25.06. 30.04	.06
S. Triceps	17.56 (7.53)	14.89. 20.23	16.21 (11.91)	12.75. 19.67	.056
S. Subscapular	12.35 (4.56)	10.73.13.97	9.10 (4.31)	7.85. 10.35	<.001
S. Supraespinale	12.78 (5.43)	10.86. 14.71	8.71 (5.56)	7.10. 10.33	<.001
S. Abdominal	16.03 (5.98)	13.90. 18.15	10.61 (6.72)	8.66. 12.56	<.001
S. Front thigh	19.15 (6.77)	16.74. 21.55	10.46 (4.68)	9.10. 11.82	<.001
S. Medial calf	15.39 (5.74)	13.35. 17.43	9.95 (4.31)	8.70. 11.21	<.001
Cormic index	0.52 (0.02)	0.52. 0.54	0.51 (0.02)	0.51. 0.52	.001
Muscle-to-bone ratio	4.38 (0.88)	4.07. 4.70	4.40 (1.30)	4.02. 4.78	.624
Adipose-to-muscle ratio	0.84 (0.28)	0.74. 0.94	0.72 (0.43)	0.60. .085	.002
Locomotive index	0.89 (0.22)	0.81. 0.97	0.79 (0.32)	0.70. 0.88	.008
Endomorphy	4.60 (1.33)	4.013. 5.07)	3.38 (1.65)	2.90. 3.86	.001
Mesomorphy	4.42 (1.70)	3.82. 5.02	4.42 (2.31)	3.75. 5.10	.686
Ectomorphy	1.50 (0.97)	1.15. 1.84	2.40 (1.48)	1.91. 2.83	.005
<b>Body composition</b>					
Adipose tissue mass (kg)	19.05 (4.93)	17.30. 20.80	17.97 (5.89)	16.26.19.68	.217
Muscle tissue mass (kg)	23.61 (7.09)	21.09. 26.12	28.55 (10.44)	25.51. 31.58	.016
Residual mass (kg)	6.21 (2.37)	5.36. 7.05	7.78 (2.21)	7.14. 8.42	<.001
Bone tissue mass (kg)	5.48 (1.63)	4.90. 6.06	6.62 (2.10)	6.01. 7.23	<.001
Skin mass (kg)	3.49 (0.48)	3.32. 3.66	3.88 (0.49)	3.73. 4.02	.001
Structured mass 5C	54.57 (14.44)	49.45. 5.69	60.60 (18.86)	55.12. 66.07	.1
%FM Slaughter_1 (%)	25.20 (7.19)	22.65. 27.75	20.24 (10.13)	17.29. 23.18	.013
%FM Slaughter_2 (%)	22.98 (6.53)	20.66. 25.29	22.06 (11.25)	18.79. 25.33	.701
Sum of six skinfolds	93.28 (28.25)	83.26. 103.30	65.08 (28.02)	56.94. 73.22	<.001
<b>Basal metabolic rate (kcal)</b>					
Schofield's equation	1411.58(142.81)	1360.94. 1462.22	1804.93(291.25)	1720.35. 1889.50	<.001
Fleisch's Equation	1484.3 (164.52)	1425.89. 1542.57	1706.08(215.75)	1643.43. 1768.73	<.001
<b>Maturity status</b>					
Maturity offset	1.77 (0.88)	1.46. 2.09	0.93 (1.31)	0.55. 1.31	.002
Age at PHV	12.90 (0.66)	12.66. 13.13	14.16 (0.75)	13.94. 14.38	<.001
Predicted adult height (cm)	160.02 (5.62)	158.03. 162.02	177.35 (6.66)	175.42. 179.29	<.001
Distance left to grow (cm)	4.15 (3.43)	2.94. 5.37	9.67 (7.88)	7.38. 11.96	<.001

Data are expressed as media (SD). B: breadths (in centimetres); G: girths (in centimetres); S: skinfolds (in millimetres); 5C: five-compartment; %GC Slaughter\_1: percentage of body fat calculated with the Slaughter equation using triceps and subscapularis skinfolds as input variables; %GC Slaughter\_2: percentage of body fat calculated with the Slaughter equation using triceps and calf skinfolds as input variables; PHV: peak high velocity. p-value for the Mann-Whitney U test.

Several morphological variables, body composition, basal metabolic rate and maturity status were significantly different between female and male (Table 1). Young female athletes had significantly higher thigh girth, skinfolds (excepting for triceps), cormic index, adipose-to-muscle ratio, locomotive index, endomorphy, %FM Slaughter\_1,  $\Sigma$ 6S and maturity offset than boys. Male participants had higher stature, body mass, various breadths, girths, ectomorphy, muscle tissue, bone tissue, basal metabolic rate, and age at PHV than women. In regards to maturity categorization, 54.54% of the girls were in average maturing and

45.45% were late matures; on the other hand, 75% of the boys were average matures, 18.75% late and 6.25% in early maturing.

### Outcome data

Characteristics of the sample of young athletes were also analysed by municipality of origin by the Kruskal-Wallis test. As reported in Table 2, Turbo and Apartadó participants had the highest stature, breadth humerus, and corrected girth thigh, while Chigorodó and Mutatá participants had the lowest ones (with the exception of corrected girth thigh that was slightly higher in Chigorodó participants than others). Carepa participants had the highest breadths (excepting for antero-posterior chest depth and apedicular breadths [humerus and femur]), cormic index, mesomorphy and muscle tissue mass, while Chigorodó participants had the lowest ectomorphy and biacromial breadth. We found significant differences ( $p < .05$ ) between municipalities in the breadth humerus, the triceps skinfold, the %GC Slaughter\_1 and the cormic index; notwithstanding, the post hoc analysis (Bonferroni correction) revealed significant differences solely for breadth humerus between Apartadó and Chigorodó ( $p = .014$ ); and cormic index between Apartadó and Chigorodó ( $p = .028$ ) and also between Apartadó and Carepa ( $p = .027$ ). The young female athletes were born in Apartadó ( $n = 10$ ), Chigorodó ( $n = 10$ ), Carepa ( $n = 9$ ), Mutatá ( $n = 1$ ), Necoclí ( $n = 1$ ) and San Juan de Urabá ( $n = 1$ ). The male athletes were distributed by municipality as follows: Apartadó ( $n = 24$ ), Carepa ( $n = 16$ ), Chigorodó ( $n = 15$ ), Turbo ( $n = 2$ ) and Mutatá ( $n = 1$ ). Regarding the calculation for early, average, and late matures across municipalities we found the following, respectively: Apartadó, 0% - 68.57% - 31.42%; Carepa, 12% - 60% - 28%; and Chigorodó, 0% - 60% - 40% (all other municipalities had average matures considering the low sample size from each location). It should be noted that all boxers ( $n = 15$ ) were born in Apartadó, while the weightlifters were from Apartadó ( $n = 20$ ), Carepa ( $n = 25$ ), Chigorodó ( $n = 15$ ), Mutatá ( $n = 2$ ), Necoclí ( $n = 1$ ), San Juan de Urabá ( $n = 1$ ) and Turbo ( $n = 2$ ).

Table 2. Characteristics of the young athletes ( $n = 81$ ) by municipality.

Variable	Apartadó ( $n = 35$ )	Carepa ( $n = 25$ )	Chigorodó ( $n = 15$ )	Mutatá ( $n = 2$ )	Necoclí ( $n = 1$ )	San Juan de Urabá ( $n = 1$ )	Turbo ( $n = 2$ )
Age (years)	14.96 (1.69)	14.84 (1.26)	15.16 (1.17)	14.95 (1.01)	13.08	14.56	14.55 (1.37)
Body mass (kg)	62.50 (15.39)	62.92 (18.44)	62.57 (19.47)	49.05 (2.47)	64.2	68.5	57.30 (13.57)
Stature (cm)	164.55 (9.59)	161.92 (9.62)	160.46 (12.13)	157.50 (6.36)	165.5	161	168.25 (12.37)
BMI ( $\text{kg}\cdot\text{m}^{-2}$ )	22.98 (4.82)	23.68 (4.84)	23.96 (5.02)	19.78 (0.60)	23.43	26.42	20.05 (1.84)
<b>Morphology</b>							
Sitting height (cm)	84.22 (5.24)	85.18 (4.43)	84.63 (5.32)	83.75 (5.30)	86.5	82.5	87.75 (2.47)
B. biacromial	33.54 (2.97)	33.73 (3.67)	33.38 (3.30)	32.95 (4.31)	35.3	29	34.20 (1.41)
B. biilocrystal	25.42 (3.41)	26.22 (5.79)	25.88 (3.65)	24.35 (0.63)	25.5	26	21.80 (0.98)
B. transverse chest	26.37 (2.61)	26.74 (2.48)	26.26 (2.75)	25.75 (2.19)	25	26.5	24.40 (2.26)
B. A-P chest depth	17.82 (4.17)	16.50 (2.24)	16.01 (1.83)	16.05 (0.35)	17	16.7	14.40 (1.97)
B. humerus	6.83 (0.56)*	6.58 (0.81)	6.00 (0.85)	5.55 (0.49)	6.3	7	6.85 (.021)
B. femur	9.16 (0.77)	8.99 (0.83)	8.84 (0.93)	7.90 (0.28)	8.7	9.5	8.90 (0.42)

G. head	55.00 (2.14)	54.73 (1.80)	54.79 (2.50)	55.45 (2.19)	54	55.5	54.85 (0.07)
G. arm (relaxed)	28.28 (4.39)	28.33 (3.70)	28.42 (4.92)	25.00 (0.00)	28.9	30.6	27.00 (2.82)
G. arm (flexed and tensed)	30.59 (4.44)	30.2 (3.80)	29.90 (4.80)	27.00 (1.41)	29.8	32	30.20 (2.82)
G. corrected arm	21.34 (5.90)	24.48 (3.25)	23.78 (3.73)	22.25 (1.44)	24.97	25.26	25.42 (2.38)
G. forearm (maximum)	26.38 (3.27)	26.48 (2.76)	25.30 (3.00)	22.65 (0.63)	25	27.5	25.55 (2.19)
G. chest (mesosternale)	87.34 (8.90)	87.78 (9.29)	85.47 (11.11)	83.10 (2.96)	87.5	91	83.90 (5.79)
G. corrected chest	84.36 (8.85)	84.33 (8.65)	81.63 (10.26)	80.74 (2.96)	83.73	86.13	81.93 (5.24)
G. waist (minimum)	73.14 (11.16)	73.84 (10.95)	71.46 (11.17)	65.55 (2.05)	73	77.7	69.30 (3.11)
G. corrected waist	69.63 (9.83)	69.47 (9.74)	66.57 (9.87)	62.80 (3.04)	67.81	71.57	67.33 (2.77)
G. Gluteal (hips)	89.96 (8.61)	91.26 (10.5)	94.75 (11.53)	84.10 (2.96)	97	101	84.50 (4.94)
G. Thigh (1 cm gluteal)	56.28 (8.89)	57.67 (8.48)	58.74 (8.60)	52.05 (4.31)	57	62	51.60 (5.09)
G. Middle thigh	51.27 (8.95)	51.34 (6.69)	52.73 (7.01)	44.70 (0.84)	50.5	51.5	49.15 (4.31)
G. corrected thigh	47.15 (8.33)	46.91 (6.29)	47.43 (6.02)	41.08 (2.84)	45.78	45.53	47.18 (4.64)
G. Calf (maximum)	27.35 (9.98)	34.42 (3.60)	34.94 (3.28)	30.50 (0.70)	32	36.5	33.40 (0.42)
G. corrected calf	23.60 (9.79)	30.49 (3.36)	30.72 (2.85)	27.59 (1.70)	27.76	33.35	31.51 (0.42)
S. Triceps	22.09 (12.11)	12.26 (6.06)	14.76 (6.77)	8.75 (4.59)	12.5	17	5.00 (1.41)
S. Subscapular	9.47 (3.94)	11.00 (4.54)	12.21 (6.28)	7.50 (0.00)	12	15.5	6.25 (1.76)
S. Supraespinale	10.10 (5.47)	10.32 (6.21)	12.20 (6.67)	7.25 (2.47)	10	12.5	4.5 (0.70)
S. Abdominal	11.17 (6.74)	13.92 (6.81)	15.56 (7.35)	8.75 (3.18)	16.5	19.5	6.25 (1.06)
S. Front thigh	13.12 (7.71)	14.10 (5.86)	16.86 (7.46)	11.50 (6.36)	15	19	6.25 (1.06)
S. Medial calf	11.92 (5.76)	12.52 (6.05)	13.43 (5.07)	9.25 (3.18)	13.5	10	6.00 (0.00)
Cormic index	0.51 (0.02)*	0.52 (0.02)†	0.52 (0.01)	0.53 (0.01)	0.52	0.51	0.52 (0.02)
Muscle-to-bone ratio	4.07 (1.42)	4.56 (0.83)	4.66 (0.62)	4.33 (0.21)	4.29	5.38	5.84 (2.06)
Adipose-to-muscle ratio	0.94 (0.50)	0.63 (0.18)	0.70 (0.17)	0.61 (0.17)	0.73	0.67	0.43 (0.01)
Locomotive index	0.96 (0.37)	0.72 (0.16)	0.77 (0.13)	0.71 (0.14)	0.9	0.77	0.54 (0.03)
Endomorphy	4.26 (1.48)	3.50 (1.63)	4.07 (1.86)	2.54 (0.98)	3.63	4.79	1.41 (0.34)
Mesomorphy	3.86 (2.66)	5.09 (1.30)	4.69 (1.41)	3.05 (0.06)	3.75	6.53	4.55 (1.48)
Ectomorphy	2.28 (1.46)	1.77 (1.28)	1.72 (1.31)	2.91 (0.74)	1.67	0.59	3.47 (0.19)

<b>Body composition</b>							
Adipose tissue mass (kg)	19.58 (5.27)	17.24 (5.52)	18.87 (6.13)	13.16 (2.03)	18.92	20.65	12.76 (3.71)
Muscle tissue mass (kg)	24.84 (10.31)	28.43 (9.13)	27.39 (9.43)	21.69 (2.81)	25.67	30.59	29.19 (7.67)
Residual mass (kg)	7.53 (2.38)	7.06 (2.57)	6.49 (2.39)	5.72 (0.80)	9.8	7.37	6.17 (1.59)
Bone tissue mass (kg)	6.22 (1.62)	6.38 (2.52)	5.95 (2.20)	5.03 (0.89)	5.97	5.68	5.08 (0.47)
Skin mass (kg)	3.84 (0.51)	3.62 (0.53)	3.65 (0.54)	3.26 (0.07)	3.72	3.88	3.75 (0.78)
Structured mass 5C	57.19 (16.80)	60.00 (18.18)	59.10 (20.58)	45.98 (3.25)	61.69	63.58	51.97 (9.52)
%FM Slaughter_1 (%)	25.83 (9.84)	19.29 (8.35)	21.77 (7.65)	14.81 (6.53)	20.96	21.57	9.08 (1.03)
%FM Slaughter_2 (%)	26.06 (9.87)	19.47 (8.52)	21.67 (8.77)	17.27 (8.51)	17.65	22.82	9.37 (3.39)
Sum of six skinfolds	77.89 (28.46)	74.12 (32.51)	85.04 (36.26)	53.00 (19.79)	79.5	93.5	34.25 (3.8)
<b>Basal metabolic rate (kcal)</b>							
Schofield's equation	1666.44 (250.17)	1680.50 (370.97)	1574.89 (368.45)	1435.80 (180.44)	1507.79	1522.8	1677.61 (237.62)
Fleisch's equation	1634.45 (180.27)	1623.06 (260.59)	1575.72 (281.67)	1427.13 (92.43)	1678.4	1643.57	1638.61 (218.57)
<b>Maturity status</b>							
Maturity offset	1.09 (1.43)	1.21 (0.95)	1.80 (1.21)	1.22 (0.03)	1.54	1.99	0.72 (1.12)
Age at PHV	13.86 (1.02)	13.63 (0.85)	13.35 (0.80)	13.73 (1.04)	11.53	12.56	13.83 (0.24)
Predicted adult height (cm)	173.22 (9.48)	169.42 (11.46)	165.10 (11.06)	163.98 (6.51)	169.92	163.99	178.64 (4.21)
Distance left to grow (cm)	8.67 (8.40)	7.49 (5.43)	4.64 (5.94)	6.48 (0.15)	4.42	2.99	10.39 (8.16)

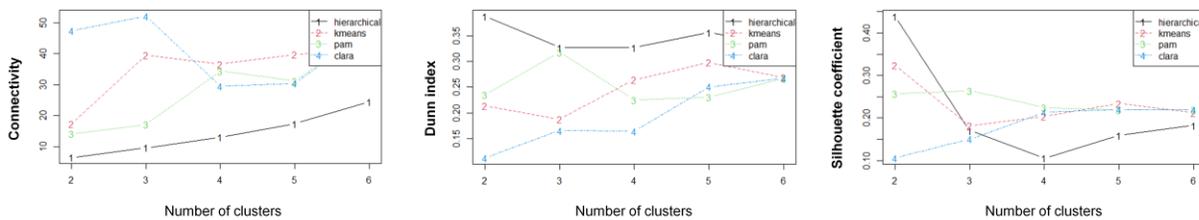
Data are expressed as media (SD), excepting for Necoclí and San Juan de Urabá where  $n = 1$ . B: breadths (in centimetres); G: girths (in centimetres); S: skinfolds (in millimetres); 5C: five-compartment; %GC Slaughter\_1: percentage of body fat calculated with the Slaughter equation using triceps and subscapularis skinfolds as input variables; %GC Slaughter\_2: percentage of body fat calculated with the Slaughter equation using triceps and calf skinfolds as input variables; PHV: peak high velocity. \*  $p$ -value < .05 for the Kruskal-Wallis test (Bonferroni post hoc) between Apartadó and Chigorodó. †  $p$ -value < .05 for the Kruskal-Wallis test (Bonferroni post hoc) between Apartadó and Carepa.

## Main results

In order to evaluate potential profiles among the young athletes and to identify the relative overlap/matching with categorical variables such as sex, sports or maturity categorization, a cluster analysis was performed using the full set of variables. The criterion algorithms and the internal validation allowed the identification and confirmation of two clusters ( $k = 2$ ) to analyse the data, respectively. Additionally, the `cValid()` function revealed that PAM clustering had better performance and resulted more appropriated to cluster our data in comparison to the hierarchical clustering (Figure 1).

Figure 1. Internal validation and clusters identification. A. After executing the `fviz_nbclust()` function, performing the enrichment with an extra loop for a total of 18 criterion algorithms and validating internally with `cValid()`, we identified two clusters and PAM as the best option to analyse our data. We highlight that connectivity should be minimized, while the Dunn index and Silhouette width should be maximized. B. Cluster plot of the k-Medoids analysis (left) and dendrogram of the bottom-up agglomerative clustering (right).

**A. Internal Validation**



**B. Cluster plot & cluster dendrogram**

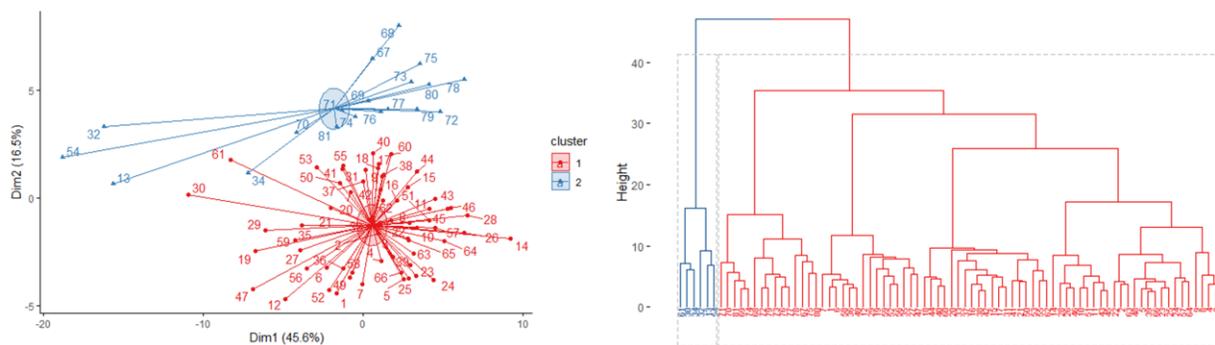


Figure 1. Internal validation and clusters identification.

It is interesting to note that the majority of the weightlifters were grouped in the cluster one, while all boxers were grouped in the cluster two after the PAM analysis. Thus, the two specific clusters highly matched with the sport type. As the internal validation predicted, the results of the hierarchical clustering were not efficient to group the data and overlap with any of the initial categorizations (Table 3).

Table 3. Matching of the clusters with review groups.

	n = 81 55 input variables	Sex		Sport *		Maturity categorization		
		F	M	Boxing	Weightlifting	Early	Average	Late
k-Medoids clustering	Cluster 1 n = 62	29	33	0	62	2	41	19
	Cluster 2 n = 19	4	15	15	4	1	13	5
Hierarchical clustering	Cluster 1 n = 75	31	44	15	60	3	50	23
	Cluster 2 n = 6	2	4	0	6	1	4	1

\* High precision matching between clusters obtained from the k-Medoids algorithm and sports.

Table 4 shows the characteristics of the young athletes by sports (boxing vs. weightlifting) and PAM clusters (cluster 1 vs. cluster 2) in order to analyse the distributions between the matching groups. All morphological variables (breadths, girths, cormic index, muscle-to-bone ratio, adipose-to-muscle ratio, locomotive index, endomorphy and mesomorphy), body composition (all masses of the 5C-model, %FM,  $\sum 6S$ ) and basal metabolic rate estimations that were significantly different between sports were also different between clusters (the high matching between significantly different variables is impressive). Notwithstanding, the maturity categorization by sports revealed that boxing had 73.3% and 26.6% for average and late matures,

respectively, while weightlifters showed 4.54%, 65.15% and 30.30% of early, average and late maturing. Conversely, the clusters were so similar between them in regards to maturity categorization (cluster 1: 3.22% - 66.12% - 30.64% vs cluster 2: 5.26% - 68.42% - 26.31% for early, average and late matures, respectively).

Table 4. Characteristics of the young athletes (n = 81) by sport and cluster.

Variable	Weightlifting (n = 66)	Boxing (n = 15)	p- value	Cluster 1 (n = 62)	Cluster 2 (n = 19)	p- value
Age (years)	14.93 (1.42)	14.91 (1.48)	.808	14.86 (1.39)	15.13 (1.56)	.361
Body mass (kg)	63.18 (17.87)	58.28 (8.91)	.597	60.00 (12.45)	69.70 (25.10)	.227
Stature (cm)	161.56 (9.74)	168.60 (9.11)	.01	160.60 (9.08)	170.26 (9.28)	<.001
BMI (kg·m <sup>-2</sup> )	23.92 (4.92)	20.42 (2.01)	.002	23.13 (3.63)	23.75 (7.34)	.251
<b>Morphology</b>						
Sitting height (cm)	84.68 (4.84)	84.63 (5.02)	.966	84.25 (4.59)	86.07 (5.50)	.198
B. biacromial	33.34 (3.30)	34.40 (2.70)	.121	32.95 (2.91)	35.45 (3.46)	.005
B. biilocrystal	25.69 (4.63)	25.45 (1.67)	.395	24.86 (2.93)	28.19 (6.45)	.014
B. transverse chest	26.46 (2.64)	26.07 (1.92)	.903	26.06 (2.16)	27.45 (3.29)	.118
B. A-P chest depth	16.95 (3.39)	18.16 (5.71)	.426	16.64 (3.24)	18.92 (5.29)	.021
B. humerus	6.51 (0.79)	6.81 (0.60)	.165	6.43 (0.74)	7.01 (0.70)	.006
B. femur	8.92 (0.84)	9.41 (0.55)	.026	8.83 (0.75)	9.60 (0.76)	.001
G. head	54.76 (2.07)	55.40 (1.78)	.118	54.61 (1.95)	55.75 (2.07)	.021
G. arm (relaxed)	28.61 (4.29)	26.61 (2.84)	.104	27.97 (3.34)	29.14 (6.06)	.933
G. arm (flexed and tensed)	30.42 (4.30)	29.20 (3.36)	.398	29.82 (3.48)	31.43 (5.79)	.455
G. corrected arm	24.62 (3.45)	15.73 (2.39)	<.001	24.20 (2.96)	18.96 (7.02)	<.001
G. forearm (maximum)	26.14 (3.15)	25.89 (2.31)	.927	25.76 (2.84)	27.16 (3.35)	.125
G. chest (mesosternale)	87.24 (9.68)	85.86 (6.22)	.724	85.61 (7.29)	91.48 (12.71)	.163
G. corrected chest	83.81 (9.19)	83.28 (6.52)	.976	82.28 (6.94)	88.38 (12.04)	.091
G. waist (minimum)	73.36 (11.62)	70.43 (3.85)	.657	71.26 (7.66)	77.90 (16.45)	.2
G. corrected waist	68.94 (10.30)	68.16 (4.02)	.627	67.12 (6.95)	74.25 (13.85)	.035
G. Gluteal (hips)	92.03 (10.17)	87.52 (5.56)	.144	90.40 (7.86)	93.77 (13.88)	.721
G. Thigh (1 cm gluteal)	58.35 (8.65)	51.16 (3.80)	.001	57.06 (7.15)	56.91 (11.98)	.147
G. Middle thigh	52.42 (7.90)	46.58 (3.31)	.001	51.45 (6.99)	50.98 (9.57)	.169
G. corrected thigh	47.68 (7.41)	43.69 (3.42)	.023	46.96 (6.95)	46.88 (7.37)	.589
G. Calf (maximum)	34.72 (3.45)	16.44 (1.16)	<.001	34.32 (3.12)	21.61 (10.36)	<.001
G. corrected calf	30.82 (3.18)	12.94 (2.15)	<.001	30.64 (3.10)	17.31 (8.99)	<.001
S. Triceps	12.70 (6.30)	34.64 (1.99)	<.001	11.97 (5.65)	32.40 (5.26)	<.001
S. Subscapular	10.93 (4.83)	8.20 (3.14)	.027	10.60 (4.56)	9.86 (5.12)	.246
S. Supraespinale	10.75 (6.28)	8.73 (2.71)	.507	9.71 (4.69)	12.52(8.37)	.296
S. Abdominal	14.09 (6.89)	7.23 (3.58)	<.001	13.18 (5.96)	11.63 (9.54)	.035
S. Front thigh	15.09 (7.10)	9.20 (4.37)	<.001	14.29 (6.03)	13.05 (9.80)	.096
S. Medial calf	12.40 (5.83)	11.13 (4.43)	.465	11.70 (4.91)	13.68 (7.38)	.468
Cormic index	0.52 (0.01)	0.50 (0.01)	<.001	0.52 (0.01)	0.50 (0.01)	<.001
Muscle-to-bone ratio	4.79 (0.83)	2.68 (0.65)	<.001	4.82 (0.83)	2.99 (0.88)	<.001
Adipose-to-muscle ratio	0.63 (0.17)	1.38 (0.46)	<.001	0.63 (0.18)	1.24 (0.50)	<.001
Locomotive index	0.72 (0.15)	1.30 (0.28)	<.001	0.72 (0.15)	1.19 (0.33)	<.001
Endomorphy	3.58 (1.65)	5.18 (0.62)	<.001	3.41 (1.55)	5.39 (0.72)	<.001
Mesomorphy	5.15 (1.50)	1.22 (0.84)	<.001	5.00 (1.37)	2.53 (2.81)	<.001
Ectomorphy	1.73 (1.26)	3.35 (1.05)	<.001	1.83 (1.23)	2.69 (1.60)	.028

<b>Body composition</b>						
Adipose tissue mass (kg)	17.58 (5.67)	22.03 (2.71)	<.001	16.56 (4.01)	24.42 (5.54)	<.001
Muscle tissue mass (kg)	28.65 (8.95)	17.23 (5.41)	<.001	27.37 (7.42)	23.79 (14.25)	.008
Residual mass (kg)	7.01 (2.54)	7.69 (1.52)	<.001	6.57 (1.78)	8.99 (3.16)	<.001
Bone tissue mass (kg)	6.09 (2.16)	6.42 (0.93)	.065	5.71 (1.38)	7.60 (2.87)	.001
Skin mass (kg)	3.65 (0.52)	4.02 (0.43)	<.001	3.58 (0.44)	4.18 (0.51)	<.001
Structured mass 5C	60.00 (18.35)	49.98 (8.44)	.029	56.71 (12.76)	62.80 (27.52)	.763
%FM Slaughter_1 (%)	19.54 (8.09)	34.20 (2.49)	<.001	18.57 (7.09)	34.30 (4.05)	<.001
%FM Slaughter_2 (%)	19.94 (8.22)	33.40 (7.19)	<.001	19.12 (7.70)	33.26 (6.70)	<.001
Sum of six skinfolds	75.98 (33.71)	79.14 (16.95)	.355	71.48 (28.84)	93.16(33.70)	.012
<b>Basal metabolic rate (kcal)</b>						
Schofield's equation	1643.55 (330.81)	1649.61 (198.42)	.451	1593.73 (257.36)	1810.92 (404.05)	.019
Fleisch's Equation	1611.90 (237.88)	1632.42 (154.26)	.368	1575.20 (190.21)	1747.86 (276.46)	.012
<b>Maturity status</b>						
Maturity offset	1.36 (1.24)	0.87 (1.08)	.066	1.27 (1.18)	1.29 (1.36)	.616
Age at PHV	13.56 (0.88)	14.03 (1.14)	.03	13.59 (0.90)	13.83 (1.11)	.196
Predicted adult height (cm)	168.57 (10.26)	177.85 (8.75)	.003	167.89 (10.01)	178.14 (8.56)	<.001
Distance left to grow (cm)	7.01 (7.12)	9.25 (6.18)	.072	7.28 (7.18)	7.88 (6.39)	.532

Data are expressed as media (SD). B: breadths (in centimetres); G: girths (in centimetres); S: skinfolds (in millimetres); 5C: five-compartment; %GC Slaughter\_1: percentage of body fat calculated with the Slaughter equation using triceps and subscapularis skinfolds as input variables; %GC Slaughter\_2: percentage of body fat calculated with the Slaughter equation using triceps and calf skinfolds as input variables; PHV: peak high velocity. p-values for the Mann–Whitney U test.

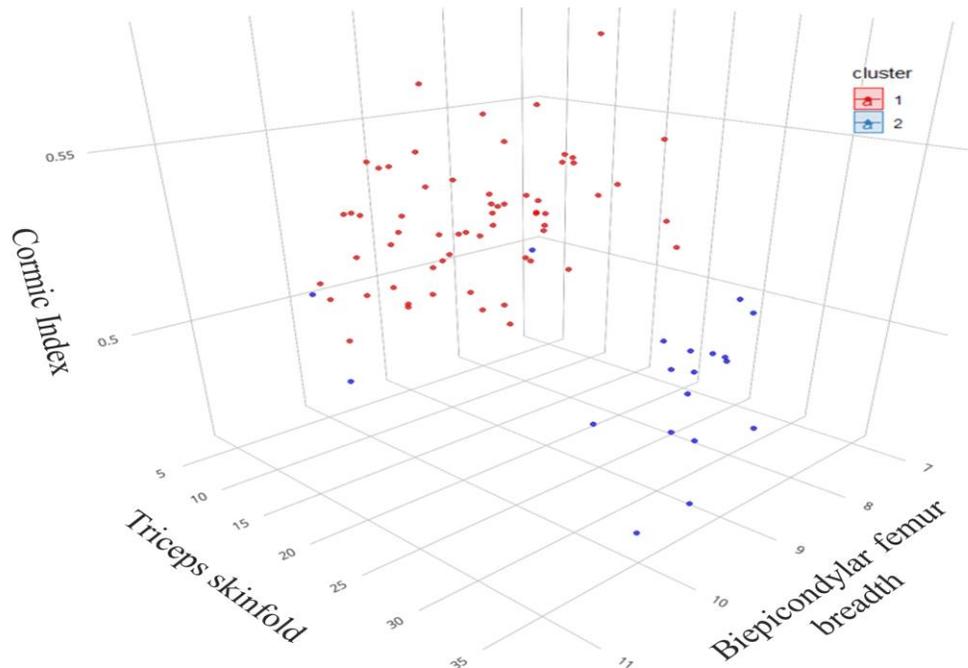


Figure 2. 3D point clustering of three relevant variables.

After looking for conserved variables that were different between municipalities and between sports, in order to compare against the clusters, we identified that breadths, cormic index and triceps skinfold where among the most relevant. Actually, biepicondylar femur breadth was also significantly different between clusters, and triceps skinfold influences the body fat percentage calculated with the Slaughter's equations, which were also different in several comparisons; hence, we developed a 3D scatter plot with these variables grouping by cluster number with impressive and clear separation of the data (Figure 2).

## DISCUSSION

### **Key results**

The aim of this study was to characterize using unsupervised machine learning analysis the morphology, body composition, and maturity status of young Colombian athletes after collecting anthropometric data during a youth sports championship in the Urabá subregion. Although recent studies have used machine learning regressions to develop equations based on anthropometric data (Ms. Shiela David, 2020) or to design a neural network architecture able to learn the body shapes using anthropometric measurements (Piccirilli, 2018), to our knowledge (based on literature searching) this is the first study that implemented machine learning strategies to analyse anthropometric cross-sectional data to characterize a young athletic population in order to extract valuable information regarding talent identification and sports specialization.

It should be noted that being selected to represent each region at the national level should indicate high-level competitive skills in the athlete; however, this evaluation frequently comes from local “*expert selectors*” without standardized criteria, lack of academic background and unfamiliarity with data analysis methodologies. In this sense, one of the important components to track is the morphology of the athletes, which refers to the measurable dimensions of the human body, such as stature, length of legs and arms, breadths, circumferences/girths and thickness of skinfolds. Several studies have been performed in different countries to have an optimal monitoring process of the young talents (Kristiansen & Stensrud, 2020; Till & Baker, 2020) and to identify the ideal morphology (e.g., somatotype) for an athlete according to the requirements of the particular sport (Cárdenas-Fernández, Chinchilla-Minguet, & Castillo-Rodríguez, 2019; Fry, Ryan, Schwab, Powell, & Kraemer, 1991; Gryko, Kopiczko, Mikołajec, Stasny, & Musalek, 2018). In this study we report relevant morphology variables derived from anthropometric data that might help to the monitoring process of young athletes. In fact, we found significant differences between sexes, municipalities and sports, which might be explained by the degree of development and adaptation among the participants. For example, the somatotype of the young athletes that participated in this study showed a dominance of mesomorphism in the weightlifters but endomorphism in boxers. From a kinanthropometry point of view, Urabá young athletes are well constituted and highly differentiated according to the type of sport, although there is a need for evaluating the physical and technical conditioning to establish more accurate statements in regards to the sports potential they might achieve. Hence, the specific functions required for different weight category sports (e.g., boxing and weightlifting) are expected to be associated with body composition adaptations (i.e., increased muscle, decreased fat), which was partially supported by our study (given that we did not measure association directly) but found interesting differences between sports (Table 4). Finally, the maturity categorization (early, average, and late) in regards to the sex of the participants was similar to previous reports (Sherar et al., 2005), given that most of the young female athletes were average matures and men had a higher rate of early maturation (6.25% in men vs. 0% in women). Interestingly, all early matures were born in Carepa, but due to the heterogeneity of the sample size among municipalities is difficult to extract relevant conclusions in this regard; however, more research is needed with complementary assessments of the maturity status in a larger sample size per municipality.

We partially confirmed our initial hypothesis, given that the clustering analysis revealed that sport type might explain the variation in the data but not sex. The differences between clusters were more sensitive to variations in morphological variables and basal metabolic rate, while body composition behaves similarly within both sports and clusters (except for the sum of skinfolds and bone tissue mass, which were different between clusters but not between sports). Interestingly, the clusters did not differ in the maturity status variables (except for predicted adult height). Thus, most of the morphological and body composition variables that showed significant difference between sports and between clusters were the same, which highlights the importance of results in this study and open a huge panorama for future sport-related studies that aim to take advantage of the unsupervised machine learning algorithms to strengthen the cross-sectional data analysis when the aim is to characterize populations. In fact, this clustering methodology can be optimized as we previously demonstrated after applying pre-processing with principal component analysis and a Boosting algorithm, which allowed increasing the resolving power of a K-means clustering algorithm while reducing noise in data during gene expression analysis (Molina-Mora, Mata, & Bonilla, 2017).

According to (Malina et al., 2015) “*talent development from novice to elite is superimposed on a constantly changing base—physical growth, biological maturation and behavioural development. These processes occur simultaneously and interact with each and with the demands of sport*”. Thus, we believe that complementary skills in programming and data science are mandatory for the enrichment analysis that might take place in sports studies for talent identification, especially because clustering is not the only valuable type of machine learning that can be explored. For instance, there have been recent proposals to measure anthropometrical features by means of image processing and machine learning algorithms, such as sub-pixel processing and finding convexity hull defects (Nguyen, Nguyen, & Zhukov, 2015), machine learning regressions (i.e., support vector regression, Gaussian processes regression, and neural network regression) instead of conventional linear regressions to estimate stature and body mass (Rativa, Fernandes, & Roque, 2018), or imputation methods based on naïve Bayes classifier, artificial neural networks, and support vector machine to estimate stature in humans (Son & Kim, 2020).

### **Limitations**

This study has several flaws. First, there was certain heterogeneity in the sample across each municipality and also regarding the amount of female and male athletes. Second, we did not record sports performance variables which makes difficult to interpret the data in those terms. More research is needed to have a characterization that search for associations between morphology, body composition, maturity status and sports performance. Third, although there is no universally applicable criterion or ‘gold standard’ methodology for body composition assessment, more robust body composition methods are needed, highlighting the desire for an economical laboratory or field approach to body composition assessment that is both accurate and objective (Ackland et al., 2012). In addition, the limited number of dissected cadavers cannot be representative of the range of body types and compositions among athletes in our analysis. Finally, we are aware of the limitations of the maturity status assessment based on Mirwald et al. (2002); however, besides providing some insights and distribution with this method the aim was to optimize the derived information through machine learning analysis and promote this methodologies in future studies. In fact, the Research Division of DBSS International will start soon a project with the support of IMDER Apartadó, the Colombian Ministry of Sport, and SENA Urabá to continue collecting quality data in this mission to potentiate talent identification and derivate/develop a Colombian method for the evaluation of maturation in young athletes.

### **Interpretation**

This study not only provides the relevant information for the first characterization of young Colombian athletes from Urabá subregion in order to have a better comprehension of the physical differences and maturity status

of athletes with Olympic projection, but also contributes with methodological procedures to analyse/interpret anthropometric datasets under the machine learning paradigm.

### **Generalizability**

The relative absence of athletes' homogeneity per municipality, per sex and per sports does not allow to extend generalizability to each specific geographical location or to the whole Urabá subregion.

## **CONCLUSIONS**

PAM clustering showed that sport type might explain the variation in the morphology, body composition and maturity status data in U16 athletes from the Urabá subregion. Our study revealed significant differences between sex, sports, municipalities and clusters for certain analysed variables. Considering the matching between sports and clusters, it is noteworthy reasonable to recommend the implementation of unsupervised machine learning strategies (e.g., clustering) along with other supervised methodologies in the identification and characterization of young talents, given the powerful features (such as precision, accuracy, speed, automation, pattern identification, inter alia) over traditional methods of data analysis. This study not only contributes with the first integration of anthropometric data, machine learning and talent identification but also to the early sports specialization in Colombian athletes with Olympic projection; however, further research and socio-economic support to this region is needed, not to mention the alphabetization of exercise professionals in terms of data science to optimize the use of this methods.

## **FUNDING**

This research was supported by the Complejo Tecnológico, Agroindustrial, Pecuario y Turístico del Servicio Nacional de Aprendizaje – SENA in collaboration with the Research Division of DBSS International. The conference/article processing charges were partially funded by this last.

## **ACKNOWLEDGMENTS**

The authors thank the young athletes who participated in this study, the coaches and parents for their support, and the Servicio Nacional de Aprendizaje - SENA for facilitating the development of this research. The preliminary results of this study were presented during the 2nd Ibero-American Symposium on Sport and Physical Activity; Nutrition & Training-SIDANE II in Medellin, Colombia.

## **REFERENCES**

- Ackland, T. R., Lohman, T. G., Sundgot-Borgen, J., Maughan, R. J., Meyer, N. L., Stewart, A. D., & Müller, W. (2012). Current Status of Body Composition Assessment in Sport. *Sports Medicine*, 42(3), 227-249. <https://doi.org/10.2165/11597140-000000000-00000>
- Alastrue Vidal, A., Rull Lluch, M., Camps Ausas, I., Ginesta Nus, C., Melus Moreno, M. R., & Salva Lacombe, J. A. (1988). [New norms and advices in the evaluation of anthropometric parameters in our population: adipose tissue-muscle index, weight indices and percentile tables of anthropometric data useful in nutritional assessment]. *Medicina Clínica*, 91(6), 223-236.
- Alzate Salazar, D. A., Ramos Bermúdez, S., & Melo Betancourt, L. G. (2011). Tejido adiposo en escolares entre 7 y 18 años de edad. *Hacia la Promoción de la Salud*, 16, 85+.

- Ballard, R. J., Dewanti, R. A., Sayuti, S., & Umar, N. (2014). Correlation between Sum of 8 Skinfolds to Predicted % Body Fat Range as a Reliable Measure of Body Composition Assessment for Well-Trained Athletes. *Asian Social Science*, 10(5). <https://doi.org/10.5539/ass.v10n5p12>
- Bernal-Orozco, M. F., Posada-Falomir, M., Quiñónez-Gastélum, C. M., Plascencia-Aguilera, L. P., Arana-Nuño, J. R., Badillo-Camacho, N., . . . Vizmanos-Lamotte, B. (2020). Anthropometric and Body Composition Profile of Young Professional Soccer Players. *Journal of Strength and Conditioning Research*, 34(7), 1911-1923. <https://doi.org/10.1519/JSC.0000000000003416>
- Berral de la Rosa, F. J., Rodríguez-Bies, E. C., Berral de la Rosa, C. J., Rojano Ortega, D., & Lara Padilla, E. (2010). Comparison of Anthropometric Equations for Estimation Muscle Mass in Badminton Player. *International Journal of Morphology*, 28(3). <https://doi.org/10.4067/S0717-95022010000300022>
- Beunen, G., & Malina, R. M. (2007). Growth and Biologic Maturation: Relevance to Athletic Performance. In *The Young Athlete* (pp. 3-17). <https://doi.org/10.1002/9780470696255.ch1>
- Brock, G., Pihur, V., Datta, S., & Datta, S. (2008). cIValid: AnRPackage for Cluster Validation. *Journal of Statistical Software*, 25(4). <https://doi.org/10.18637/jss.v025.i04>
- Cárdenas-Fernández, V., Chinchilla-Minguet, J. L., & Castillo-Rodríguez, A. (2019). Somatotype and Body Composition in Young Soccer Players According to the Playing Position and Sport Success. *Journal of Strength and Conditioning Research*, 33(7), 1904-1911. <https://doi.org/10.1519/JSC.0000000000002125>
- Cardona, V. P. (2013). El deporte en el departamento de Antioquia y su materialización en política pública. *Analecta Política*, 4(5), 413-433.
- Catikkas, F., Kurt, C., & Atalag, O. (2013). Kinanthropometric attributes of young male combat sports athletes. *Coll Antropol*, 37(4), 1365-1368.
- Correa, B., & Enrique, J. (2008). Determinación del perfil antropométrico y cualidades físicas de niños futbolistas de Bogotá. *Revista ciencias de la salud*, 6(2).
- Du Bois, D. (1916). Clinical Calorimetry. *Archives of Internal Medicine*, XVII(6\_2). <https://doi.org/10.1001/archinte.1916.00080130067006>
- Esparza-Ros, F., Vaquero-Cristóbal, R., & Marfell-Jones, M. (2019). International Standards for Anthropometric Assessment. In Murcia, Spain: The International Society for the Advancement of Kinanthropometry.
- Faigenbaum, A. D., Kraemer, W. J., Blimkie, C. J. R., Jeffreys, I., Micheli, L. J., Nitka, M., & Rowland, T. W. (2009). Youth Resistance Training: Updated Position Statement Paper From the National Strength and Conditioning Association. *Journal of Strength and Conditioning Research*, 23, S60-S79. <https://doi.org/10.1519/JSC.0b013e31819df407>
- Fernández, P. C. G., Jara, M. d. C. B., & Zamudio, R. A. U. J. R. R. d. e. d. J. o. S. T. (2017). El índice locomotivo, una perspectiva antropométrica de la eficiencia biomecánica. 31(3), 3-11.
- Fleisch, A. (1951). [Basal metabolism standard and its determination with the "metabocalculator"]. *Helvetica Medica Acta*, 18(1), 23-44.
- Freedman, D. S., Horlick, M., & Berenson, G. S. (2013). A comparison of the Slaughter skinfold-thickness equations and BMI in predicting body fatness and cardiovascular disease risk factor levels in children. *Am J Clin Nutr*, 98(6), 1417-1424. <https://doi.org/10.3945/ajcn.113.065961>
- Fry, A. C., Ryan, A. J., Schwab, R. J., Powell, D. R., & Kraemer, W. J. (1991). Anthropometric characteristics as discriminators of body-building success. *Journal of Sports Sciences*, 9(1), 23-32. <https://doi.org/10.1080/02640419108729852>
- Garrido-Chamorro, R., Sirvent-Belando, J. E., González-Lorenzo, M., Blasco-Lafarga, C., & Roche, E. (2012). Skinfold Sum: Reference Values for Top Athletes. *International Journal of Morphology*, 30(3), 803-809. <https://doi.org/10.4067/S0717-95022012000300005>

- Gryko, K., Kopiczko, A., Mikołajec, K., Stasny, P., & Musalek, M. (2018). Anthropometric Variables and Somatotype of Young and Professional Male Basketball Players. *Sports (Basel)*, 6(1). <https://doi.org/10.3390/sports6010009>
- Heath, B. H., & Carter, J. E. L. (1967). A modified somatotype method. *American Journal of Physical Anthropology*, 27(1), 57-74. <https://doi.org/10.1002/ajpa.1330270108>
- Holway, F. E., & Garavaglia, R. (2009). Kinanthropometry of Group I rugby players in Buenos Aires, Argentina. *Journal of Sports Sciences*, 27(11), 1211-1220. <https://doi.org/10.1080/02640410903207408>
- IOC. (2016). International Olympic Committee. Retrieved from <https://www.olympic.org/olympic-results>
- ISAK. (2020). The International Society for the Advancement of Kinanthropometry. Retrieved from <https://www.isak.global/>
- Kerr, D. (1988). An anthropometric method for fractionation of skin, adipose, bone, muscle and residual tissue masses in males and females age 6 to 77 years. (Doctoral), Simon Fraser University, Retrieved from <https://core.ac.uk/download/pdf/56369359.pdf>
- Kristiansen, E., & Stensrud, T. (2020). Talent development in a longitudinal perspective: Elite female handball players within a sport school system. *Translational Sports Medicine*, 3(4), 364-373. <https://doi.org/10.1002/tsm2.143>
- Larson-Meyer, D. E., Woolf, K., & Burke, L. (2018). Assessment of Nutrient Status in Athletes and the Need for Supplementation. *International Journal of Sport Nutrition and Exercise Metabolism*, 28(2), 139-158. <https://doi.org/10.1123/ijsem.2017-0338>
- Malina, R. M., Rogol, A. D., Cumming, S. P., Coelho e Silva, M. J., & Figueiredo, A. J. (2015). Biological maturation of youth athletes: assessment and implications. *British Journal of Sports Medicine*, 49(13), 852-859. <https://doi.org/10.1136/bjsports-2015-094623>
- Mannor, S., Jin, X., Han, J., Jin, X., Han, J., Jin, X., . . . Zhang, X. (2011). K-Medoids Clustering. In *Encyclopedia of Machine Learning* (pp. 564-565). [https://doi.org/10.1007/978-0-387-30164-8\\_426](https://doi.org/10.1007/978-0-387-30164-8_426)
- Martin, A. D., Spenst, L. F., Drinkwater, D. T., & Clarys, J. P. (1990). Anthropometric estimation of muscle mass in men. *Med Sci Sports Exerc*, 22(5), 729-733. <https://doi.org/10.1249/00005768-199010000-00027>
- MinDeporte. (2019). Colombia Tierra de Atletas: la estrategia para consolidar al país como un semillero de deportistas. Retrieved from <https://www.mindeporte.gov.co/94410>
- Mirwald, R. L., G. Baxter-Jones, A. D., Bailey, D. A., & Beunen, G. P. (2002). An assessment of maturity from anthropometric measurements. *Medicine and Science in Sports and Exercise*, 34(4), 689-694. <https://doi.org/10.1097/00005768-200204000-00020>
- Molina-Mora, J., Mata, F., & Bonilla, D. (2017). Improvement of K-Means Clustering Algorithm Performance in Gene Expression Data Analysis through Pre-Processing with Principal Component Analysis and Boosting. *Research journal of life sciences bioinformatics pharmaceutical and chemical sciences*, 3(2), 53-62. <https://doi.org/10.26479/2017.0302.05>
- Ms. Shiela David, C. N. S. K. W. A. (2020). Projecting Height and Weight with Machine Learning Using Anthropometric Measurements. *International Journal of Advanced Science and Technology*, 29(10s), 7091 - 7095.
- Nguyen, T., Nguyen, T. H., & Zhukov, A. (2015). Studies of anthropometrical features using machine learning approach. Paper presented at the CEUR Workshop Proceedings.
- Norton, K., & Eston, R. (2018). *Kinanthropometry and exercise physiology*: Routledge. <https://doi.org/10.4324/9781315385662>
- Palma, L. H., Méndez, C. H., Manrique, A., Castro, J. A., Viveros, A., Garzón, K. A., . . . Restrepo, Á. J. (2021). Asociación entre la composición corporal y la condición física en estudiantes de grado sexto, pertenecientes a la institución educativa moderna de Tuluá, Colombia año 2019 (Association

- between body composition and the physical condition in sixth grade st. Retos(39), 539-546. <https://doi.org/10.47197/retos.v0i39.77988>
- Panam Sports. (2019). Pan American Sports Organization. Retrieved from <https://www.lima2019.pe/medallero-panamericanos>
- Patel, R., Nevill, A., Cloak, R., Smith, T., & Wyon, M. (2019). Relative age, maturation, anthropometry and physical performance characteristics of players within an Elite Youth Football Academy. *International Journal of Sports Science & Coaching*, 14(6), 714-725. <https://doi.org/10.1177/1747954119879348>
- Petro, J. L., Rodríguez Arrieta, A. N., & Montenegro Arjona, O. A. (2017). Perfil dermatoglífico y condición física de jugadores adolescentes de fútbol. *Educación Física y Ciencia*, 19(2). <https://doi.org/10.24215/23142561e038>
- Piccirilli, M. (2018). Machine Learning Approaches to Human Body Shape Analysis. (Doctoral), West Virginia University, West Virginia University Libraries. Graduate Theses, Dissertations, and Problem Reports database. (6417).
- Rativa, D., Fernandes, B. J. T., & Roque, A. (2018). Height and Weight Estimation From Anthropometric Measurements Using Machine Learning Regressions. *IEEE Journal of Translational Engineering in Health and Medicine*, 6, 1-9. <https://doi.org/10.1109/JTEHM.2018.2797983>
- Schofield, W. N. (1985). Predicting basal metabolic rate, new standards and review of previous work. *Human Nutrition: Clinical Nutrition*, 39 Suppl 1, 5-41.
- Sherar, L. B., Mirwald, R. L., Baxter-Jones, A. D. G., & Thomis, M. (2005). Prediction of adult height using maturity-based cumulative height velocity curves. *J Pediatr*, 147(4), 508-514. <https://doi.org/10.1016/j.jpeds.2005.04.041>
- Slaughter, M. H., Lohman, T. G., Boileau, R. A., Horswill, C. A., Stillman, R. J., Van Loan, M. D., & Bembien, D. A. (1988). Skinfold equations for estimation of body fatness in children and youth. *Hum Biol*, 60(5), 709-723.
- Söğüt, M., Luz, L. G. O., Kaya, Ö. B., Altunsoy, K., Doğan, A. A., Kirazci, S., . . . Knechtle, B. (2019). Age- and Maturity-Related Variations in Morphology, Body Composition, and Motor Fitness among Young Female Tennis Players. *International Journal of Environmental Research and Public Health*, 16(13). <https://doi.org/10.3390/ijerph16132412>
- Son, Y., & Kim, W. (2020). Missing Value Imputation in Stature Estimation by Learning Algorithms Using Anthropometric Data: A Comparative Study. *Applied Sciences*, 10(14). <https://doi.org/10.3390/app10145020>
- Team, R. C. (2017). R: A language and environment for statistical computing [Computer software]. Retrieved from <https://www.R-project.org/>
- Till, K., & Baker, J. (2020). Challenges and [Possible] Solutions to Optimizing Talent Identification and Development in Sport. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.00664>
- Vandenbroucke, J. P., Von Elm, E., Altman, D. G., Gøtzsche, P. C., Mulrow, C. D., Pocock, S. J., . . . Egger, M. (2009). Mejorar la comunicación de estudios observacionales en epidemiología (STROBE): explicación y elaboración. *Gaceta Sanitaria*, 23(2), 158.e151-158.e128. <https://doi.org/10.1016/j.gaceta.2008.12.001>
- Visnapuu, M., & Jürimäe, T. (2007). Handgrip Strength and Hand Dimensions in Young Handball and Basketball Players. *The Journal of Strength and Conditioning Research*, 21(3).
- von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., Vandenbroucke, J. P., & Initiative, S. (2014). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Int J Surg*, 12(12), 1495-1499. <https://doi.org/10.1016/j.ijsu.2014.07.013>

- Wittek, P. (2014). Unsupervised Learning. In Quantum Machine Learning (pp. 57-62). <https://doi.org/10.1016/B978-0-12-800953-6.00005-0>
- World Medical Association. (2002). World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. *Nurs Ethics*, 9(1), 105-109. <https://doi.org/10.1191/0969733002ne486xx>
- Zemski, A. J., Keating, S. E., Broad, E. M., & Slater, G. J. (2018). Longitudinal Changes in Body Composition Assessed Using DXA and Surface Anthropometry Show Good Agreement in Elite Rugby Union Athletes. *Int J Sport Nutr Exerc Metab*, 1-8. <https://doi.org/10.1123/ijsnem.2018-0019>



This work is licensed under a [Attribution-NonCommercial-NoDerivatives 4.0 International](https://creativecommons.org/licenses/by-nc-nd/4.0/) (CC BY-NC-ND 4.0).