Power System Technology
ISSN:1000-3673

Received: 06-10-2024 Revised: 15-11-2024 Accepted: 10-12-2024

# Perspectives of Machine Learning for Anthropometry Data Processing

**Victoria Zaborova [1,2], Ekaterina Chebotareva [1], Oxana Zolnikova [1], Natiya Dzhakhaya [1], Elena Bueverova [1], Alla Sedova [1], Anastasia Kurbatova [1], Dmitry Zyuzin [1], Hassan Shafaei* [3], Marina Kinkulkina [1].**

1 Institute of Clinical Medicine, Sechenov First Moscow State Medical University, 119991 Moscow, Russia

2 Moscow Institute of Physics and Technology, Institutsky lane 9, Dolgoprudny, Moscow Region, 141700, Russia

3 Department of Physical Rehabilitation, massage and health-improving physical culture named after I. M. Sarkizov-Serazini RSUFKSMiT, 105122, Moscow, Russia

Corresponding Author: Hassan Shafaei, hassanshafai37@yahoo.com, ORCID: 0000-0001-5802-3385

**Abstract**

The most popular machine learning methods with a teacher are linear regression, the method of near neighbors and decision trees. In the course of this work, the most important anthropometric parameters for male skiers were determined through the use of artificial intelligence. The data was evaluated in 3 stages using various machine learning methods. As a result, for male skiers, the most important indicators were shoulder length, elbow diameter, wrist circumference, skin-fat fold of the wrist in front, wrist diameter, shoulder width and frontal diameter of the chest. Thus, we found out that machine learning methods can be used to determine the most important anthropometric parameters for a particular sport.

**Keywords:** machine learning, physical condition, athletes

**Introduction**

The assessment of the parameters of a person's physical condition is of great importance. These parameters can be estimated by anthropometry, bioimpedance analysis (BIA).

It should be noted that the physical development of athletes is associated with professional selection. During professional selection, hereditary features are realized, as well as qualities acquired as a result of systematic training sessions. Anthropometric parameters such as skeletal lengths and diameters are inheritable, while girths and fat folds undergo changes during exercise. Anthropometry measures various characteristics (height, weight, girth, width of bones and skin folds), BIA evaluates fat and muscle mass using a weak alternating electric current that flows at different speeds depending on the physique [1]. Tissues rich in water and electrolytes (muscles, blood) easily conduct current, while spaces containing bone, air, and fat, on the contrary, delay current [2].

This assessment is carried out for specific sports and is aimed at determining physical fitness and competition results in a particular sport [3], as well as for screening or assessing the physical fitness of athletes of various skill levels during pre-season training [4].

Anthropometric indicators and body composition may vary in different sports depending on the physical requirements of the sport and the level of athletic training. This suggests that different sports, such as swimming and cross-country skiing, will have different effects on the physical development of athletes.

Power System Technology
ISSN:1000-3673
Received: 06-10-2024 Revised: 15-11-2024 Accepted: 10-12-2024

Cross—country skiing is a cyclical winter sport in which competitors compete in time over a specially prepared snow track using cross-country skis and ski poles. The main styles of movement on skis are "classic (flat) style" and "free (skate) style". The "classic" ski moves are divided by the method of pushing off with sticks into alternating and simultaneous ones. "Free style" implies that the skier is free to choose the method of movement over the distance, but since the "classic" stroke is inferior in speed to the "skate", "free style" is, in fact, synonymous with the "skate stroke", similar to the skating technique [5].

In this machine learning study, we analyzed which anthropometric parameters are most important for monitoring and predicting success in cross-country skiing.

**Materials and methods**

To begin with, it was decided to test the general hypothesis about the possibility of distinguishing the level of training of an athlete based on his anthropometric data. The verification was performed using anthropometric data using a regression model. A default two-class classifier based on a regression model was used for verification. When training and testing a model, objects of only 2 classes are submitted to the input of such a classifier, and the expert system learns to distinguish them from each other. It should be noted that the accuracy of the model's distinction between the two classes should not be lower than 0.5, since a 50% probability gives a banal guess of the answers. If the accuracy of the classifier is below 50%, the interpretation of the answers should be reversed, which will increase the accuracy above 50%.

The most popular machine learning methods with a teacher are linear regression, the method of near neighbors and decision trees.

In practice, we used the C++ and Python programming languages to build expert systems. There are libraries for both languages that significantly simplify the work of constructing and interpreting mathematical models of machine learning.

To begin with, we wrote a script in the Python programming language that collects anthropometric data of athletes from separate tables dedicated to one person into one summary table. The output was a table, each row of which corresponded to the anthropometric measurement of one person.

To begin with, we applied the linear regression method. In addition to the basic algorithm, I used the Gridsearch tool, which allows me to analyze network training parameters in order to get the most accurate data.

Such hyperparameters of the network as the type of regularization (l1 and l2) and the regularization coefficient C varied.

Then we tried using other learning models, such as the k nearest neighbors (kNN) method. Gridsearch was also launched for this method. The number of neighbors and the presence/absence of weight varied depending on the distance.

For the model based on random trees, the criteria for building a tree (criterion), the maximum depth of the tree (max_depth) and the number of trees (n_estimators) in the decisive forest were varied.

The table contains records of 78 athletes, of whom 10 had 3 years of athletic experience, 33 had 5 years, 25 had 7 years, and 10 had more than 10 years (Figure 1).

Power System Technology
ISSN:1000-3673
Received: 06-10-2024 Revised: 15-11-2024 Accepted: 10-12-2024

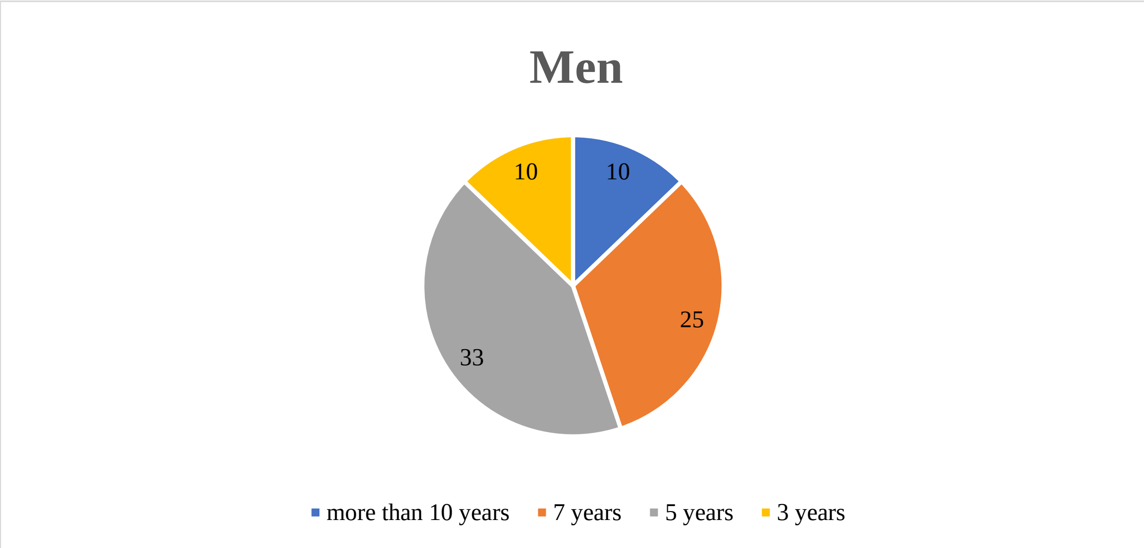



Figure 1. The ratio of male skiers by sports categories.

**Results**

***Primary data processing***

Figure 2 shows the obtained accuracy of the pairwise separation. For clarity, a point at the 0.5 level has been added to each graph, indicating that the class corresponds to itself. To improve the accuracy of the grid performance assessment, testing was performed 100 times. The graph shows the average data from 100 predictions.

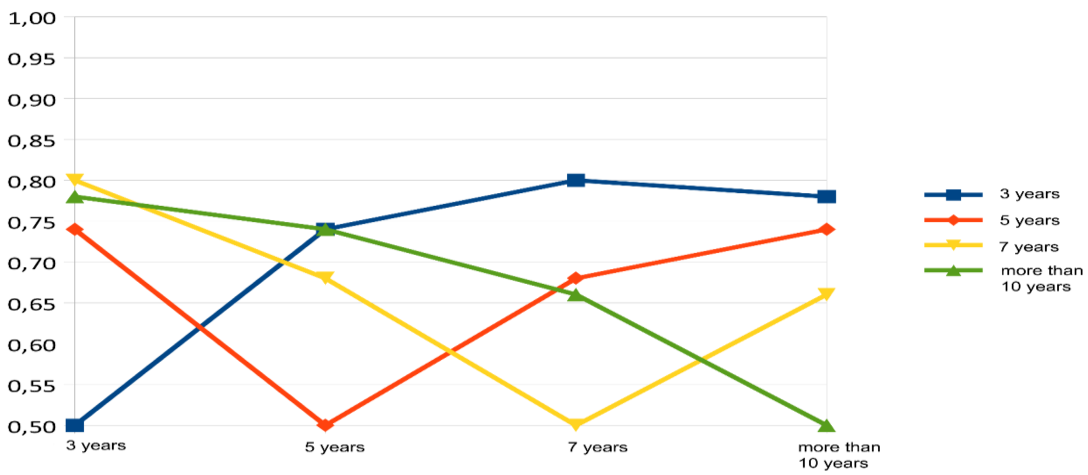



Figure 2. The accuracy of the mutual pairwise classification of the processed data, men, skiing.

As we can see, there is an observed correlation between the mutual distance of the discharges (and hence the physical fitness of athletes) and the accuracy of network prediction. However, there is a theoretically poorly explained failure in the accuracy of separating 7 years of athletic experience from 10 years. Let's now try to make a multiclass classification, that is, class definitions from all, rather than a choice of two.

Power System Technology
ISSN:1000-3673
Received: 06-10-2024 Revised: 15-11-2024 Accepted: 10-12-2024

***Data processing.***

We will also use the regression model. The result is shown in Figure 3. Here and further, graphs of this kind along the abscissa axis show the results of network predictions for each person, compared with the control response.

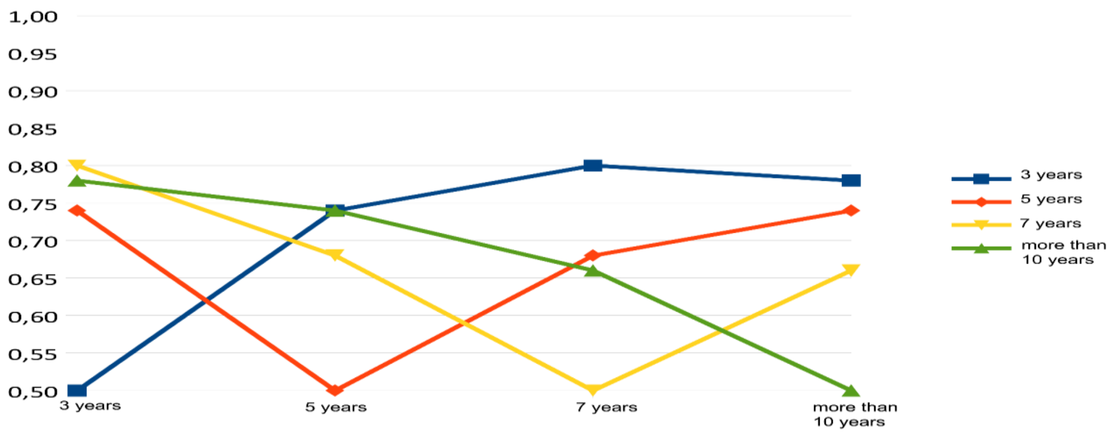



Figure 3. The accuracy of the mutual pairwise classification of male skiers.

From this Figure, it can be seen that the accuracy of the classifier was about 70%. That is, even with multiclass classification, the expert system's ability to determine an athlete's level of training is observed. From these graphs, it can be seen that the general trends observed in the sample among women can also be seen in the database for men. The main difference is greater basic accuracy in the male sample for the two-grade classification, which can be explained by a larger training sample, which means better model training. The accuracy is worse for a multiclass classifier.

***Testing various machine learning methods (men, skiing).***

As a result of the selection of hyperparameters, models of the following type and accuracy were obtained.

Regression model (l2, C = 0.09), accuracy on the deferred sample was 0.25 (Figure 4).

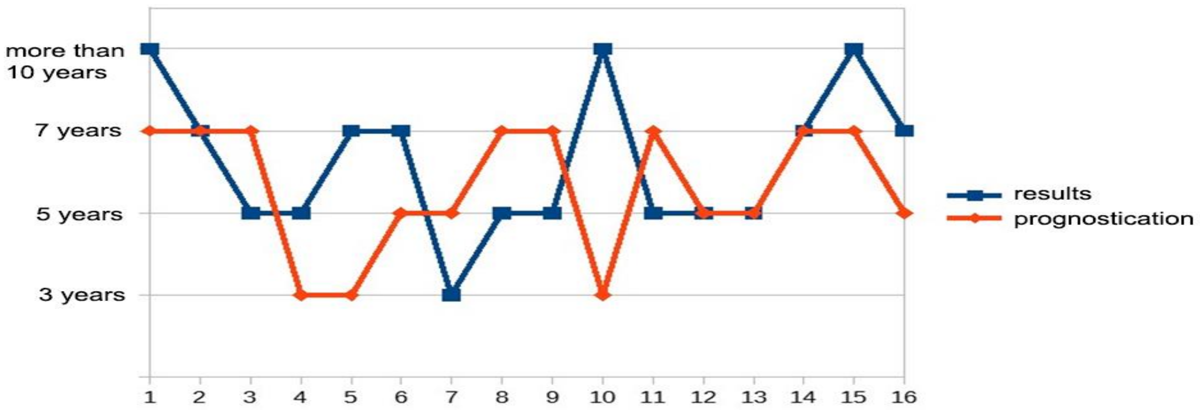



Figure 4. The best regression model for deferred sampling, men, skiing.

Power System Technology
ISSN:1000-3673
Received: 06-10-2024 Revised: 15-11-2024 Accepted: 10-12-2024

The nearest neighbors method is the number of neighbors 2, weighted voting. The accuracy on the deferred sample was 38% (Figure 5).

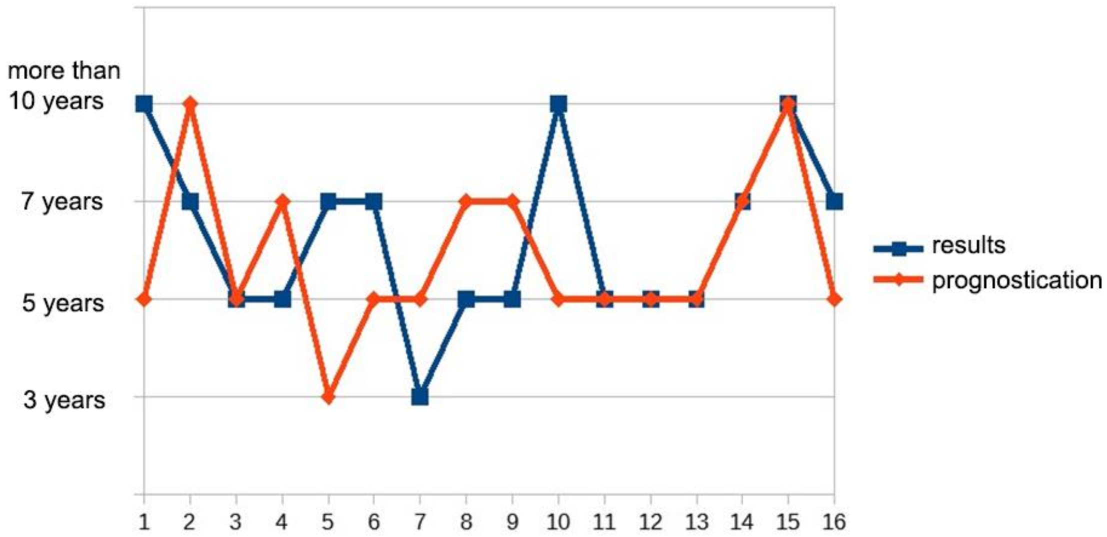



Figure 5. The best knn model in the deferred sample, men, skiing.

For the decision tree, the optimal hyperparameters were as follows: ncriterion — 'entropy', max_depth — 12, n_estimators — 9. The accuracy was also 38% (Figure 6).

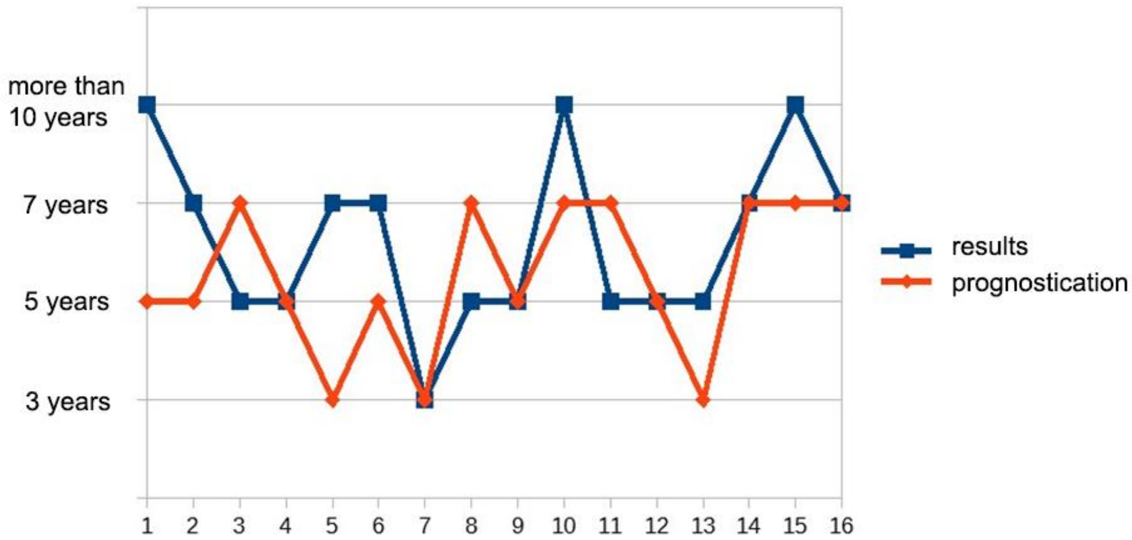



Figure 6. The best model based on decision trees in a deferred sample, men, skiing.

Thus, it is noticeable that for male skiers, the most successful models were the decision trees and the method of the nearest neighbors.

Power System Technology
ISSN:1000-3673
Received: 06-10-2024 Revised: 15-11-2024 Accepted: 10-12-2024

***Evaluation of the importance of parameters.***

Theoretically, it is classifiers based on decision trees that can most clearly assess the contribution of each parameter. This is due to their architecture and algorithms for building a decision tree — the most defining parameters are often located closer to the root in order to divide the data into classes as efficiently as possible. The tree-based classifier also showed good accuracy in testing.

It is for the above reasons that it is proposed to use the proprietary feature_importances_ method of the ExtraTreesClassifier class of the sklearn library to assess the importance of each parameter of anthropometric measurements in predicting athletes' results.

Let's build a table and graph for male skiers (Table 1, Figure 7).

Table 1. The importance of anthropometric parameters, men, skiing.

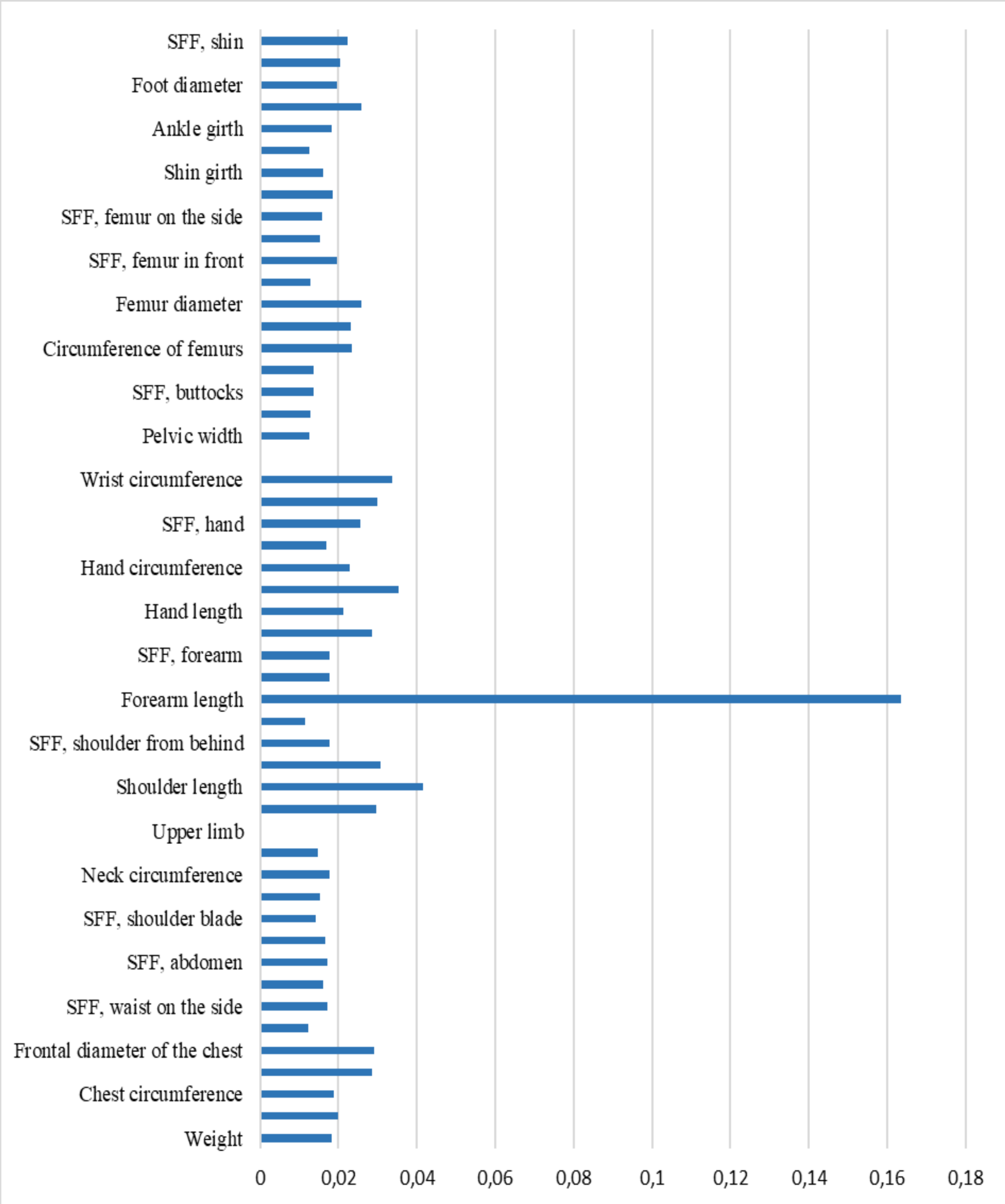
| **Parameter** | **Value** |
|---|---|
| Weight | 0,018375 |
| Height | 0,020016 |
| Chest circumference | 0,018958 |
| Sagittal chest size | 0,02852 |
| Frontal diameter of the chest | 0,02914 |
| Waist circumference | 0,012206 |
| SFF, waist on the side | 0,017319 |
| SFF, chest | 0,016066 |
| SFF, abdomen | 0,017274 |
| SFF, back | 0,016595 |
| SFF, shoulder blade | 0,014343 |
| SFF, xiphoid process | 0,015409 |
| Neck circumference | 0,017681 |
| Head circumference | 0,014663 |
| **Upper limb** | |
| Shoulder width | 0,029563 |
| Shoulder length | 0,041584 |
| SFF, shoulder in front | 0,030755 |
| SFF, shoulder from behind | 0,017743 |
| Shoulder girth | 0,01147 |
| Forearm length | 0,16359 |

Power System Technology
ISSN:1000-3673
Received: 06-10-2024 Revised: 15-11-2024 Accepted: 10-12-2024

| | |
|---|---|
| Forearm girth | 0,017613 |
| SFF, forearm | 0,017664 |
| Arm length | 0,028584 |
| Hand length | 0,021285 |
| Elbow diameter | 0,035304 |
| Hand circumference | 0,02284 |
| Hand diameter | 0,016884 |
| SFF, hand | 0,025542 |
| Wrist diameter | 0,029829 |
| Wrist circumference | 0,033792 |
| **Lower limb** | |
| Pelvic width | 0,012632 |
| SFF, anterior iliac spine | 0,012935 |
| SFF, buttocks | 0,013768 |
| Leg length | 0,013545 |
| Circumference of femurs | 0,023379 |
| One femur`s circumference | 0,023088 |
| Femur diameter | 0,025929 |
| Femur length | 0,01299 |
| SFF, femur in front | 0,019686 |
| SFF, femur from behind | 0,015431 |
| SFF, femur on the side | 0,01597 |
| SFF, thigh inside | 0,018547 |
| Shin girth | 0,016165 |
| Shin length | 0,012664 |
| Ankle girth | 0,018272 |
| Ankle diameter | 0,02577 |
| Foot diameter | 0,019591 |
| Foot circumference | 0,020553 |
| SFF, shin | 0,02248 |

Power System Technology
ISSN:1000-3673
Received: 06-10-2024 Revised: 15-11-2024 Accepted: 10-12-2024

*SFF – skin-fat fold




*SFF – skin-fat fold

Figure 7. The importance of anthropometric parameters, men, skiing.

Power System Technology
ISSN:1000-3673
Received: 06-10-2024 Revised: 15-11-2024 Accepted: 10-12-2024

It can be noted that in skiing, the upper body is more important for determining the level of training of athletes. Skiers also seem to be interested in lung volume, which is achieved, among other things, due to the sagittal and frontal diameters of the chest. At the same time, the length of the leg and its sections have an unexpectedly small effect on the result (below average).

**Discussion**

In this study, we examined the possibility of using expert systems to assess the level of training of athletes, taking into account changes in specific anthropometric parameters characteristic of a particular sport. There is a correlation between anthropometric parameters and the level of athletes and it is discernible on the available data sets.

Athletic performance was often associated with anthropometric parameters. This applies to team sports such as volleyball, in which higher results were associated with the position used, a low percentage of fat, greater muscle mass and height, to individual sports such as swimming, cycling, judo, running [6,7,8,9,10]. The Prieske study drew the following conclusions:

- young male judoka showed a greater increase in body muscle mass, hand grip strength, and CMJ during the 10-month training period compared to female athletes;

- a significant increase in height and body weight, as well as dynamic balance (Y-balance test) was observed before and after testing, regardless of gender;

- Men's height/body weight and performances were higher than those of female athletes. while the body fat mass was lower in young men compared to judoka, regardless of the time [11].

Since judo is a sport of the weight category, therefore, body composition and, in particular, body fat and muscle mass are two important aspects that require systematic monitoring during the long–term development of an athlete [12]. In general, a lower percentage of body fat and a higher percentage of body muscle increase the relative strength level of athletes. In addition, it has been suggested that these two body composition indicators are related to success in competitions [13]. The analysis revealed a significantly higher body fat mass in young women compared to male judoka. This is in good agreement with other studies, for example, Little revealed a 10% higher percentage of body fat in elite judoka compared to their male peers [14,15]. Also, in a sample of Spanish elite teenage judoka, Torres-Lucke et al. observed a higher isometric hand grip strength in men compared to women, while the height of CMJ did not differ significantly [16].

Differences in anthropometric parameters are also observed in mountaineering and bouldering. Different forms of sport climbing require different sets of skills and physiological conditions [17]. Leading rock climbing is characterized by more static, slow, and controlled movements than bouldering [18]. Moreover, the leading climbers are more capable of overcoming longer routes than boulderers, with a shorter climbing time (30 seconds for bouldering versus 2-7 minutes for climbing ahead). On the other hand, bouldering routes are shorter, but the movements are carried out with maximum effort [19]. This suggests that bouldering can be classified as a strength discipline, while lead climbing is characterized by an endurance effort [20]. Several studies have shown that boulderers are characterized by a higher explosive strength of the forearm muscles compared to leading climbers [21].

Comparable links have also been found between muscle mass and performance in both female and male skiers, and especially in ski sprints [22,23]. The reason for the present results may be related to the advantages of better upper and lower body strength for forward movement in skiers with the highest muscle mass, as evidenced by the same links found between arm muscle mass and results on flat terrain and downhill. This is also confirmed

Power System Technology
ISSN:1000-3673
Received: 06-10-2024 Revised: 15-11-2024 Accepted: 10-12-2024

by Larsson and Henriksson–Larsen, who suggest the importance of body muscle mass in the arms for XC skiing results in junior skiers [24,25].

**Conclusion**

Thus, we have identified the importance of machine learning in sports, namely, determining the most important anthropometric parameters for a particular sports discipline. The system is based on machine learning algorithms that process large amounts of data. This makes it possible to develop personalized training programs that take into account the unique characteristics of the user and his current condition, which surpasses traditional standards of recommendations. This system is versatile and can be customized for various sports and physical fitness levels. This makes it suitable for both professional athletes and amateurs, which is its unique advantage over existing specialized solutions.

**References**

1. J.J. Ramos-Álvarez; J.J. Montoya.; C. Solís-Mencia; F. Miguel-Tobal,; P. López-Tapia; A.J. Sánchez-Oliver; R. Domíngue; J.M. Martínez-Sanz. Anthropometric Profile Assessed by Bioimpedance and Anthropometry Measures of Male and Female Rugby Players Competing in the Spanish National League. Appl. Sci., 11, 11759. 2021.
2. S.Y. Lee; D. Gallagher. Assessment methods in human body composition. Curr. Opin. Clin. Nutr. Metab. Care, 11, 566–572. 2008] [U. Mulasi; A.J. Kuchnia; A.J. Cole; C.P. Earthman,. Bioimpedance at the bedside: Current applications, limitations, and opportunities. Nutr. Clin. Pract., 30, 180–193. 2015.
3. Schwesig R., Hartmann M., Leuchte S., Fischer D., and Kuß O., Validity of a complex soccer-specific field test and a non-specific sprint test-assessments for test and match performance, Sportverletzung Sportschaden. 2013.
4. Barfod, K.W.; Feller, J.A.; Clark, R.; Hartwig, T.; Devitt, B.M.; Webster, K.E. Strength Testing After Anterior Cruciate Ligament Reconstruction: A Prospective Cohort Study Investigating Overlap of Tests. J. Strength Cond. Res. 2019.
5. FIS | International Ski and Snowboard Federation https://www.fis-ski.com/.
6. Mielgo-Ayuso, J.; Urdampilleta, A.; Martínez Sanz, J.M.; Calleja González, J.; Seco, J. Relationship of Sport Competitive Level on Dietary Intake, Body Composition and Somatotype between Elite and Amateur Female Volleyballers. Ann. Womens Health, 2017.
7. dos Santos, M.A.M.; Henrique, R.S.; Salvina, M.; Silva, A.H.O.; Junior, M.A.d.V.C.; Queiroz, D.R.; Duncan, M.J.; Maia, J.A.R.; Nevill, A.M. The influence of anthropometric variables, body composition, propulsive force and maturation on 50m freestyle swimming performance in junior swimmers: An allometric approach. J. Sports Sci. 2021.
8. Basset, F.; Billaut, F.; Joanisse, D. Anthropometric Characteristics Account for Time to Exhaustion in Cycling. Int. J. Sports Med. 2014.
9. Giudicelli, B.B.; Luz, L.G.O.; Sogut, M.; Sarmento, H.; Massart, A.G.; Júnior, A.C.; Field, A.; Figueiredo, A.J. Chronological Age, Somatic Maturation and Anthropometric Measures: Association with Physical Performance of Young Male Judo Athletes. Int. J. Environ. Res. Public Health 2021.
10. Dessalew, G.W.; Woldeyes, D.H.; Abegaz, B.A. The Relationship Between Anthropometric Variables and Race Performance. Open Access J. Sports Med. 2019.
11. Olaf Prieske, Helmi Chaabene, Martijn Gäbler, Michael Herz, Norman Helm, Adrian Markov and Urs Granacher. Seasonal Changes in Anthropometry, Body Composition, and Physical Fitness and the Relationships with Sporting Success in Young Sub-Elite Judo Athletes: An Exploratory Study. Int. J. Environ. Res. Public Health, 2020.

Power System Technology
ISSN:1000-3673
Received: 06-10-2024 Revised: 15-11-2024 Accepted: 10-12-2024

12. Torres-Luque, G.; Hernández-García, R.; Escobar-Molina, R.; Garatachea, N.; Nikolaidis, P.T. Physical and Physiological Characteristics of Judo Athletes: An Update. Sports, 2016.
13. Casals, C.; Huertas, J.R.; Franchini, E.; Sterkowicz-Przybycień, K.; Sterkowicz, S.; Gutiérrez-García, C.; Escobar-Molina, R. Special Judo Fitness Test Level and Anthropometric Profile of Elite Spanish Judo Athletes. J. Strength Cond. Res. 2017.
14. Little, N. Physical performance attributes of junior and senior women, juvenile, junior, and senior men judokas. J. Sports Med. Phys. Fit. 1991.
15. Franchini, E.; Rodríguez Huertas, J.; Sterkowicz, S.; Carratalá, V.; Gutiérrez-García, C.; Escobar-Molina, R. Anthropometrical profile of elite Spanish Judoka: Comparative analysis among ages. Arch. Budo, 2011.
16. Torres-Luque, G.; Hernandez-Garcia, R.; Garatachea, N.; Nikolaidis, P. Anthropometric characteristics and neuromuscular function in young judo athletes by sex, age and weight category. Sport Sci. Health, 2015.
17. Lutter, C.; El-Sheikh, Y.; Schöffl, I.; Schöffl, V. Sport Climbing: Medical Considerations for This New Olympic Discipline. Br. J. Sports Med. 2017.
18. White, D.J.; Olsen, P.D. A Time Motion Analysis of Bouldering Style Competitive Rock Climbing. J. Strength Cond. Res. 2010.
19. Zieliński, G.; Pająk-Zielińska, B. Bouldering & Chronic Traumatic Encephalopathy—That Is Repeated Mild Traumatic Brain Injury from Climbing. Med. Hypotheses, 2024.
20. Stien, N.; Saeterbakken, A.H.; Hermans, E.; Vereide, V.A.; Olsen, E.; Andersen, V. Comparison of Climbing-Specific Strength and Endurance between Lead and Boulder Climbers. PLoS ONE, 2019.
21. Levernier, G.; Laffaye, G. Four Weeks of Finger Grip Training Increases the Rate of Force Development and the Maximal Force in Elite and Top World-Ranking Climbers. J. Strength Cond. Res. 2019.
22. Jones, T.W.; Lindblom, H.P.; Karlsson, Ø.; Andersson, E.P.; McGawley, K. Anthropometric, Physiological, and Performance Developments in Cross-country Skiers. Med. Sci. Sports Exerc. 2021.
23. Carlsson, M.; Carlsson, T.; Wedholm, L.; Nilsson, M.; Malm, C.; Tonkonogi, M. Physiological Demands of Competitive Sprint and Distance Performance in Elite Female Cross-Country Skiing. J. Strength Cond. Res. 2016.
24. Larsson, P.; Henriksson-Larsén, K. Body composition and performance in cross-country skiing. Int. J. Sports Med. 2008.
25. Østerås, S.; Welde, B.; Danielsen, J.; van den Tillaar, R.; Ettema, G.; Sandbakk, Ø. Contribution of Upper-Body Strength, Body Composition, and Maximal Oxygen Uptake to Predict Double Poling Power and Overall Performance in Female Cross-Country Skiers. J. Strength Cond. Res. 2016.