

Can Health Status and Lifestyle Indicators Predict Amateur Soccer Players Performance Level? A Preliminary Study

Beatrice De Lazzari^{1,2}^a, Giuseppe Vannozzi¹^b, Federico Caramia¹, Filippo Lupi²,
Paolo Salvatore² and Valentina Camomilla¹^c

¹Department of Movement, Human and Health Sciences, Interuniversity Centre of Bioengineering of the Human Neuromusculoskeletal System, University of Rome "Foro Italico", Rome, Italy

²GoSport srl, Rome, Italy

Keywords: Anamnesis Sheet, Amateurs, Performance Evaluation, Football.

Abstract: Introduction: Soccer is one of the most popular sports in the world. To determine the performance potential of an athlete, various tests are typically performed in elite athletes, but not in amateur ones. Aim: To evaluate if and which health status and lifestyle indicators can be useful predictors of physical performance level in amateur soccer players. Methods: A group of 32 male subjects (age: 32±12 years, mass: 77±10 kg, stature: 1.78±0.06 m) voluntarily participated in the study. To assess their functional capacities, five in-field tests were conducted, while as an anamnesis sheet, a questionnaire was developed that investigated: body mass index (BMI), age, physical activity level, lifestyle, alcohol consumption and smoking habits, sports career, occurring injuries, and medical history. A stepwise backward regression was then conducted. Results: A significant $R^2=0.722$ was found between the questionnaire outputs and the physical tests, using only six of the nine investigated indicators. Conclusions: With a simple questionnaire, an estimate of amateur athletes' physical performance can be obtained. Prospectively, a wider dataset, including women, will allow for the definition of a synthetic biometric index.

1 INTRODUCTION

Soccer is one of the most popular sports in the world, practiced by elite athletes and amateurs, and it is an intermittent sport (Tessitore et al, 2005; Bangsbo, 1994). Players Key Performance Indicators (KPI) can be assessed through in-field tests for different soccer-related functional capacities (Figueiredo et al, 2011; Taher and Shahbazi, 2013). While these tests are usually performed by elite athletes and research is mainly focused on them, there is a lack of information about amateurs' functional capacity. This is more evident for amateurs not belonging to sport clubs or associations, that are our target, where there is a reduced accessibility to physical testing facilities and a limited knowledge of coaches about the players. To overcome this issue, a technological support is proposed to provide an accessible way to assess and

predict the level of performance of the athletes. Biometric quantities, such as BMI and physical activity level, were proved to be related to functional capacity of a player (Nikolaidis, 2012; Gil-Rey et al, 2015) and can be collected through a purposely developed questionnaire, together with players' in-field performance tests (Kaplan, 2010). Soccer research attempted to identify the link between biometric quantities and in-field test scores (Campa et al, 2019; Nikolaidis, 2012), but these do not cover all biometric quantities in relation with many different performance tests. With this work we aimed to investigate health status and lifestyle indicators as possible predictors for the player performance level, considering different functional capacity tests that contribute to the overall performance assesment of the player.

^a <https://orcid.org/0000-0002-2887-9139>

^b <https://orcid.org/0000-0002-2359-6076>

^c <https://orcid.org/0000-0002-7452-120X>

From the questionnaire results, and through a regression analysis, we aimed at identifying the main indicators with the potential of being used in the future to implement a new collective index, related to biometrics, that could allow to predict the physical status of an athlete when it is not possible or appropriate to perform in-field tests.

The great advantage related to this potential index is that it could be an accessible tool if implemented into a smartphone app specifically devised to support soccer users. Through this type of app, big data could be gathered to provide knowledge about amateurs soccer players. The app could also make available physical testing procedures for this underinvestigated population. As a fallout, personalized training plans also based on training time and player role could be promoted for amateur soccer players.

2 METHODS

To identify the main biometric indicators with the potential to predict an athlete's physical status, a questionnaire of 20 questions was developed. The questionnaire evaluates 9 different aspects: body mass index (BMI), age, physical activity level (through IPAQ), lifestyle, alcohol consumption and smoking habits, sports career, occurring injuries, and medical history. From the multiple responses available for each question, a single score is obtained as descriptive of each abovementioned aspect.

Thirty-two male participants (age: 32 ± 12 years, mass: 77 ± 10 kg, stature: 1.78 ± 0.06 m) completed the questionnaire and performed the following five tests, as detailed in section 2.1, to assess their KPIs: 30 m sprint test, Yo-Yo Intermittent Recovery Test Level 1 (Yo-Yo IR1), countermovement jump (CMJ), standing long jump (SLJ), repeated sprint ability test (RSA).

Each test result is compared to the typical score performed by an elite soccer player (whose scores come from the literature cited in Section 2.2, Table 1), and a related percentage value is obtained. At the end of all tests and once the percentage values are obtained, a KPI is calculated for each test. An overall KPI_t value is then obtained as average of these specific values. A stepwise backward regression analysis is performed using IBM SPSS 28 by receiving the questionnaire scores as independent variables of the regression analysis. A collective index is then obtained as the result of a mathematical relation among the independent variables which minimises the distance to KPI_t as value obtained from the physical tests (see 2.1.5).

2.1 Performed Tests

In this section, the tests performed by the participants are briefly described.

2.1.1 Yo-Yo Intermittent Recovery Test Level 1 (Yo-Yo IR1)

This test is typically used for the evaluation of agility and aerobic capacity in intermittent sports (Bangsbo et al, 2008; Bangsbo et al, 2006). In untrained individuals, it evaluates the ability to perform a repeated and intense test and the capacity of recovery. It is a maximal test, and it is evaluated as follows: the participant must run 20 m back and forth across a marked track keeping time with beeps. After a 20+20 m run, the participant has 10 s of recovery before starting again. Going over the time, the time interval in which the 20 m must be performed decreases, while the recovery time remains the same. In the 10 s of recovery, the participant can run slower or walk 5 m back and forth before starting again, performing an active recovery, as shown in Figure 1.

The KPI value for this test, KPI_1 , is calculated as the total distance run by the subject in percentage of the typical score of an elite athlete (reported in Table 1).

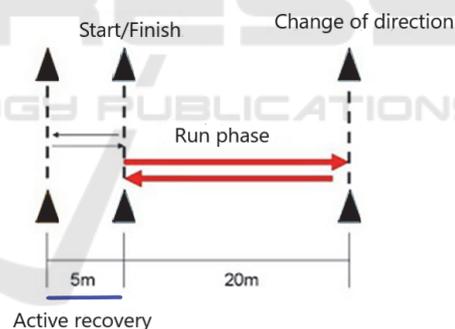


Figure 1: Yo-Yo Intermittent Recovery Test Level 1 (IR1) scheme.

2.1.2 Repeated Sprint Ability (RSA)

This test is composed of six shuttle runs of 20 m with change of direction and a time recovery of 20 s among runs, as shown in Figure 2. This test evaluates the change of direction ability. This type of test is particularly influenced by the aerobic capacity (Rampinini et al, 2009) and reproduces the metabolic phenomena that characterize the most intense phases in a match, such as pH reduction and anaerobic glycolysis activation (Rampinini et al, 2007).

The values of interest are the time intervals needed by the subject to perform each shuttle. The

minimum time spent in a trial by an elite athlete (see Table 1) is used to obtain the KPI of interest for this test, KPI_2 .

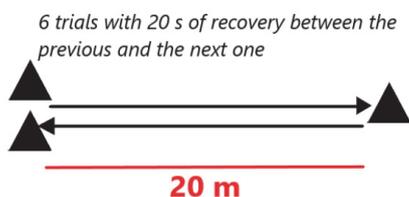
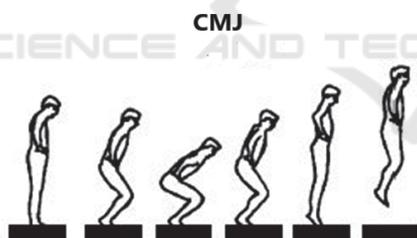


Figure 2: Repeated Sprint Ability (RSA) scheme.

2.1.3 Lower Limb Strength

The Counter-Movement Jump (CMJ) is a test repeated three times and the variable of interest is the jump height, measured by an inertial sensor, Gyko (Migrogate, Italy, sampling frequency = 1000 samples/s; full scale: $\pm 16g$, $\pm 2000deg/s$). The execution of the jump is described in Figure 3. In the CMJ, the participant starts with their hands on their hips and, following the experimenter's start, performs a sudden bend in their legs and, without stopping the movement, jumps upwards. To calculate a jump related KPI_{CMJ} , the maximum jump height among these trials is taken as the reference value and normalised by the typical jump height of an elite soccer athlete found in the literature (see Table 1).



3 trials, the best one is taken as reference

Figure 3: Counter-Movement Jump (CMJ) vertical execution (the frames refer to the same position in space).

The Standing-Long Jump (SLJ) is repeated three times and performed as shown in Figure 4. In the SLJ, the participant starts with their arms at their sides, and following the go-ahead given by the experimenter, performs a forward jump using their arm swing and by landing on the ground with their feet close together. The variable of interest is the jump length, taken with a tape as the heel-to-heel distance. The longest jump is taken as the value to compare with the result performed by an elite athlete (see Table 1) to obtain the related KPI_{SLJ} .



3 trials, the longest one is taken as reference

Figure 4: Standing-Long Jump (SLJ) execution.

Both SLJ and CMJ describe the lower limb power. In particular, CMJ gives an estimate of the capacity to storage and use the elastic energy of extensor muscles (Tessitore et al, 2005). For this reason, KPI_3 is calculated as the mean of the two results coming from KPI_{CMJ} and KPI_{SLJ} .

2.1.4 The 30 m Sprint Test

The 30 m test is a sprint test in which the participants must run 30 m at their maximum velocity. The test is performed three times and the variable of interest is the time required to run 30 m. The value of the best attempt is compared to the score of elite athletes to calculate KPI_4 (see Table 1).

This was considered as a sport-specific test because high-intensity sprints are frequent in soccer matches: in fact, 96% of sprints are shorter than 30m and their frequency is one every 90 s (Izzo et al, 2018).



Figure 5: 30 m sprint scheme.

2.1.5 KPI Total (KPI_t)

The KPI_t is the final KPI index that considers the KPI values coming from the performed tests as follows:

$$KPI_t = (KPI_1 + KPI_2 + KPI_3 + KPI_4) / 4 \quad (1)$$

2.2 Statistics of the Tests

In Table 1, the results of the tests are reported. For Yo-Yo IR1, 30 m, RSA, CMJ and SLJ tests, the mean and standard deviation are reported. These values are obtained from the best results performed by each participant. In the last column, the scores of elite athletes found in literature are reported (Schmitz et al, 2018; Rampinini et al, 2007; Sporis et al, 2009; Zapartidis et al, 2009; Chaouachi et al, 2010).

Table 1: In the second and third columns, there are the mean and standard deviations of the results of the tests, in the fourth column, the value of reference of elite athletes as found in the above-cited literature.

Test	Mean	Standard deviation	Elite score
Yo-Yo IRI [m]	785	473	2800
RSA [s]	7.4	0.5	6.5
CMJ [m]	0.35	0.06	0.46
SLJ [m]	2.1	0.4	>2.8
30 m [s]	4.8	0.4	4.0

2.3 Questionnaire Implementation

The proposed questionnaire is composed of 20 questions and investigates nine different aspects about health status and lifestyle as possible predictors of the performance of the athlete. In particular, the acquired information is about:

- anthropometric data, as weight and height of the subject to calculate the BMI, age and sex;
- health information: smoking habits, alcohol consumption, medical history and occurring injuries in the last 12 months;
- habits information, such as sports career and work;
- physical activity data, evaluated through the IPAQ-FS questionnaire (Lee, 2011).

Each explored domain pertains to the sphere of personal health and is widely considered as risk factor in the literature. Specifically, BMI is one of the screening factors for subjects' health (Erickson, 1998), and it is correlated to their quality of life (Bottcher et al, 2020). Age is one of the main risk factors for the development of chronic degenerative diseases, according to the guidelines of WHO (2004), and for what concerns soccer, it's one of the factors for the risk of injury in both elite and amateur players (Ekstrand et al, 2011; Gebert et al, 2018; Dallinga et al, 2012; Arnason et al, 2004).

Smoking habits can affect the health of the subject following the 2002 guidelines of the WHO. For the aim of this work, the questionnaire evaluates simply if the participant smokes or not, on the basis of the study of Jeon (2021) which compared the performance of smokers and not-smoker populations. Following this study, it is found that smokers have a reduced performance about 21% in the test used for that scope.

Alcohol consumption is an indicator obtained as the result of two questions of the questionnaire which evaluates alcohol consumption in a day and with which frequency. Alcohol consumption is included as variable in this study since it is one of the main factors

for the health state of the subject, according to the guidelines of WHO (2004).

Medical history is investigated through a question in which the subject indicates the presence or not of chronic pathologies, divided into controlled and severe ones.

Information on occurring injuries is another indicator associated to one question in which the subject has to identify the time spent without doing physical activity due to the presence of the injury. The time reference is related to the last 12 months. This parameter is important in amateurs, who have high relapse risks, which can affect their health status.

The sports career indicator associates the subjects to three main groups, depending on their experience as athlete: not professionals, former elite athletes, former amateur athletes.

Work indicator is obtained as the result coming from the answers of two questions which evaluate the type of work that the subject does, and how they arrive to their place of work.

The level of physical activity is investigated through the IPAQ-FS questionnaire, which evaluates the time spent (in terms of hours and days) in doing different physical activities.

Each indicator derived from the questionnaire is associated with a score between 0% and 100%. Once completed the questionnaire, the nine indicators will be represented by a percentage value.

2.4 Stepwise Regression Analysis

A stepwise regression analysis is performed, considering the results coming from the questionnaire as independent variables.

The regression allows estimating the contribution of each indicator coming from the questionnaire to a potential synthetic index related to individual biometric characteristics. Using this technique, it is possible to evaluate different models that, step by step, don't consider the less significant predictor. A total of 6 models are thus obtained, with the first one characterized by the presence of all the 9 variables above mentioned and the last one which considers just the four main indicators. For each model, a R^2 value is obtained.

The choice of the best model is identified by the evaluation of the R^2 : as the number of input variables decreases, the R^2 decreases, too. As it is possible to see in Figure 6, an R^2 value is reported for each computed model. A model is selected as good if its R^2 changes less than 3% with respect to the model including all the 9 variables (MO9), which includes all the input variables.

3 RESULTS

The model including 6 variables (MO6) is the last to have a good R^2 according to the set criterion (Figure 6). In the following, MO9 and MO6 are compared to see the impact of removing variables on a collective biometric index.

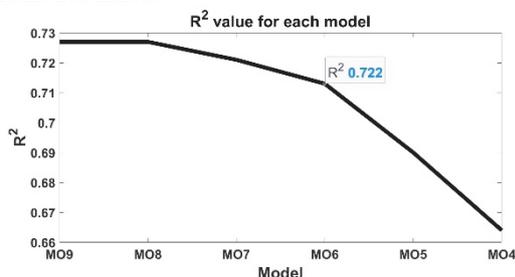


Figure 6: R^2 values of the regressive models are here reported. MO9 considers all the nine variables, MO4 considers only four main variables.

In MO9, all the variables contribute to the estimate of the global KPI, with $R^2=0.734$, $F(9,22) = 6.744$, $p<0.05$. Considering MO6, the situation changes only slightly: $R^2=0.722$, $F(6,25) = 10.842$, $p<0.05$.

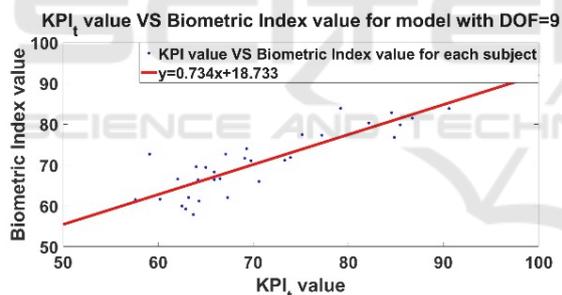


Figure 7: Relation between overall biometric index based on MO9 and the KPI_t value obtained for each subject.

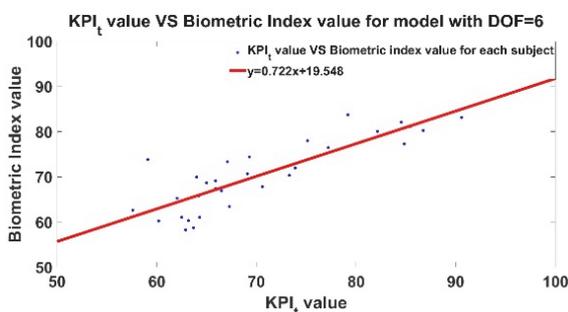


Figure 8: Relation between overall biometric index based on MO6 and the KPI_t value obtained for each subject.

In Table 2 and 3, b and β values are reported for MO9 and MO6 respectively:

Table 2: MO9 b and β values.

MO9	b	β
Constant	40.305	-
Work	0.115	0.193
Physical activity Level	0.113	0.528
Sports career	-0.028	-0.061
Alcohol consumption	-0.094	-0.132
Occurring injuries	-0.059	-0.181
Medical history	0.11	0.033
Age	0.102	0.235
BMI	0.221	0.442
Smoking habits	0.097	0.109

Table 3: MO6 b and β values.

MO6	b	β
Constant	48.026	-
Work	0.115	0.193
Physical Activity Level	0.116	0.544
Alcohol consumption	-0.096	-0.135
Occurring injuries	-0.051	-0.157
Age	0.095	0.219
BMI	0.227	0.455

4 DISCUSSIONS

These preliminary results support the possibility to determine the performance level of amateur athletes based on health and daily life indicators. Thus, a prediction of the KPI_t could be obtained without performing specific in-field tests.

Looking at β values in the previous section, BMI and physical activity level seem to be the main predictors. This is in agreement with the literature that reports a correlation between those two factors and the performance of the athlete in different in-field tests (Nikolaidis, 2012; Gil-Rey et al, 2015). In both MO9 and MO6 models, the polynomial regression explains more than the 70% of the variability of the dependent variable, obtaining an adequate predictive model, with a potential of improvement through further data sampling and updates in the regression technique.

As shown in the results, not all the aspects seem to be meaningful based on the population of interest: comparing MO9 and MO6, the R^2 decreases less than 3%, thus three out of nine indicators seem to

contribute minimally to the prediction of amateur performance level. Specifically, looking at MO6, the less important aspects are sports career, smoking habits, and medical history. For instance, a previous brilliant sporting career does not guarantee the preservation of soccer performance: in fact, following Mujika and Padilla (2000), a body that is not constantly exposed to training stimuli, can easily regress. The smoking habit may be a performance-altering factor, but it is possibly necessary to investigate the amount of tobacco consumed by the athlete (Tetelepta et al, 2019). In fact, the proposed questionnaire analyses simply if the individual is a smoker or not, so it seems to not be sufficient as discriminatory factor. Medical history possibly needs to be more specific to be predictive. For example, teenagers with diabetes can nowadays find information on how to prepare themselves to participate in different forms of physical activity and sports, both amateur and professional, without affecting their performance (Krzykała et al, 2021). Conversely, not all the diseases are well known, or it is difficult to assess if they can alter sport performance. One of these cases is COVID-19, which is still not known and its effects on physical status is not yet identified (Sarto et al, 2020).

From Table 2 and 3, in both models, BMI, IPAQ and age are the three main factors that may contribute to the identification of the performance level of the athlete, if used as predictors.

As a limitation, while the indices examined by the questionnaire relate to the population of interest only, genetic factors are here neglected, which might be main contributors to sports performance (MacArthur and North, 2005). Nonetheless, remaining at the in-field assessment level, this preliminary study was able to highlight how a subset of health status and lifestyle indicators can approximate player's performance level, even if further investigations and a wider dataset are needed to confirm the results.

As a future perspective, the implementation of an app able to give accessibility to these tests and to acquire data coming from amateurs can be a powerful tool to collect a larger dataset, potentially including both male and female players. The inclusion of female amateurs will bring to adjustments due to existing gender-related differences (e.g. BMI). Such larger dataset is expected to lead to more reliable regressions; thus, the definition of an overall biometric index could be provided, based on a minimal set of indicators to obtain a prediction of functional capacity in amateurs. Prospectively, the availability of a larger dataset could also open to the application of other data mining techniques to find

possible interesting factors as the evaluation of a common trend or the identification of the factors that can be used to reliably classify players from a physical performance point of view.

REFERENCES

- Arnason, A., Sigurdsson, S. B., Gudmundsson, A., Holme, I., Engebretsen, L., & Bahr, R. (2004). Risk factors for injuries in football. *The American journal of sports medicine*, 32(1_suppl), 5-16.
- Bangsbo, J. (1994). Energy demands in competitive soccer. *Journal of sports sciences*, 12(sup1), S5-S12.
- Bangsbo, J., Iaia, F. M., & Krstrup, P. (2008). The Yo-Yo intermittent recovery test. *Sports medicine*, 38(1), 37-51.
- Bangsbo, J., Mohr, M., & Krstrup, P. (2006). Physical and metabolic demands of training and match-play in the elite football player. *Journal of sports sciences*, 24(07), 665-674.
- Bottcher, L. B., Bandeira, P. F. R., Vieira, N. B., Zaia, V., & Lopes de Almeida, R. (2020). Quality of life, BMI, and physical activity in bariatric surgery patients: a structural equation model. *Obesity surgery*, 30(8), 2927-2934.
- Campa, F., Semprini, G., Júdice, P. B., Messina, G., & Toselli, S. (2019). Anthropometry, physical and movement features, and repeated-sprint ability in soccer players. *International journal of sports medicine*, 40(02), 100-109.
- Chaouachi, A., Manzi, V., Wong, D. P., Chaalali, A., Laurencelle, L., Chamari, K., & Castagna, C. (2010). Intermittent endurance and repeated sprint ability in soccer players. *The Journal of Strength & Conditioning Research*, 24(10), 2663-2669.
- Dallinga, J. M., Benjaminse, A., & Lemmink, K. A. (2012). Which screening tools can predict injury to the lower extremities in team sports?. *Sports medicine*, 42(9), 791-815.
- Ekstrand, J., Häggglund, M., & Waldén, M. (2011). Epidemiology of muscle injuries in professional football (soccer). *The American journal of sports medicine*, 39(6), 1226-1232.
- Erickson, P. (1998). Evaluation of a population-based measure of quality of life: the Health and Activity Limitation Index (HALex). *Quality of Life Research*, 7(2), 101-114.
- Figueiredo, A. J., Coelho e Silva, M. J., & Malina, R. M. (2011). Predictors of functional capacity and skill in youth soccer players. *Scandinavian journal of medicine & science in sports*, 21(3), 446-454.
- Gebert, A., Gerber, M., Pühse, U., Faude, O., Stamm, H., & Lamprecht, M. (2019). Changes in injury incidences and causes in Swiss amateur soccer between the years 2004 and 2015. *Swiss medical weekly*, (49).
- Gil-Rey, E., Lezaun, A., & Los Arcos, A. (2015). Quantification of the perceived training load and its relationship with changes in physical fitness

- performance in junior soccer players. *Journal of sports sciences*, 33(20), 2125-2132.
- Impellizzeri, F. M., Rampinini, E., Castagna, C., Bishop, D., Bravo, D. F., Tibaudi, A., & Wisloff, U. (2008). Validity of a repeated-sprint test for football. *International journal of sports medicine*, 29(11), 899-905.
- Izzo, R., De Vanna, A., & Varde'i, C. H. (2018). Data Comparison between Elite and Amateur Soccer Players by 20 Hz GPS Data Collection. *Journal of Sports Science*, 6, 31-35.
- Kaplan, T. (2010). Examination of repeated sprinting ability and fatigue index of soccer players according to their positions. *The Journal of Strength & Conditioning Research*, 24(6), 1495-1501.
- Krzykała, M., Domaszewska, K., Woźniewicz-Dobrzyńska, M., Kryściak, J., Konarska, A., Araszkiwicz, A., ... & Konarski, J. M. (2021). Characteristics of Selected Somatic and Motor Abilities of Youth Soccer Players with Diabetes Type 1 Treated with Insulin Pump Therapy. *International Journal of Environmental Research and Public Health*, 18(7), 3493.
- Lee, P. H., Macfarlane, D. J., Lam, T. H., & Stewart, S. M. (2011). Validity of the international physical activity questionnaire short form (IPAQ-SF): A systematic review. *International journal of behavioral nutrition and physical activity*, 8(1), 1-11.
- MacArthur, D. G., & North, K. N. (2007). Genes and human elite athletic performance. *East African Running*, 241-257.
- McMurray, R. G., Hicks, L. L., & Thompson, D. L. (1985). The effects of passive inhalation of cigarette smoke on exercise performance. *European journal of applied physiology and occupational physiology*, 54(2), 196-200.
- Mujika, I., & Padilla, S. (2000). Detraining: loss of training-induced physiological and performance adaptations. Part I. *Sports Medicine*, 30(2), 79-87.
- Nikolaidis, P. (2012). Association between body mass index, body fat per cent and muscle power output in soccer players. *Open Medicine*, 7(6), 783-789.
- Rampinini, E., Bishop, D., Marcora, S. M., Bravo, D. F., Sassi, R., & Impellizzeri, F. M. (2007). Validity of simple field tests as indicators of match-related physical performance in top-level professional soccer players. *International journal of sports medicine*, 28(03), 228-235.
- Rampinini, E., Sassi, A., Morelli, A., Mazzoni, S., Fanchini, M., & Coutts, A. J. (2009). Repeated-sprint ability in professional and amateur soccer players. *Applied Physiology, Nutrition, and Metabolism*, 34(6), 1048-1054.
- Sarto, F., Impellizzeri, F. M., Spörri, J., Porcelli, S., Olmo, J., Requena, B., ... & Franchi, M. V. (2020). Impact of potential physiological changes due to COVID-19 home confinement on athlete health protection in elite sports: a call for awareness in sports programming. *Sports medicine*, 50(8), 1417-1419.
- Schmitz, B., Pfeifer, C., Kreitz, K., Borowski, M., Faldum, A., & Brand, S. M. (2018). The Yo-Yo intermittent tests: a systematic review and structured compendium of test results. *Frontiers in physiology*, 9, 870.
- Sporis, G., Jukic, I., Ostojic, S. M., & Milanovic, D. (2009). Fitness profiling in soccer: physical and physiologic characteristics of elite players. *The Journal of Strength & Conditioning Research*, 23(7), 1947-1953.
- Taher, A. V., & Shahbazi, M. (2013). Functional capacities and mental skills necessary for expertise in soccer. *Iranian Journal of Health and Physical Activity*, 4(1), 44-50.
- Tessitore, A., Meeusen, R., Tiberi, M., Cortis, C., Pagano, R., & Capranica, L. (2005). Aerobic and anaerobic profiles, heart rate and match analysis in older soccer players. *Ergonomics*, 48(11-14), 1365-1377.
- Tetelepta, Y. W., Muslim, E., & Moch, B. N. (2019, May). Physical ergonomics analysis of the effect of smoking on sport performance. In *IOP Conference Series: Materials Science and Engineering* (Vol. 505, No. 1, p. 012037). IOP Publishing.
- Zapartidis, I., Varelzis, I., Gouvali, M., & Kororos, P. (2009). Physical fitness and anthropometric characteristics in different levels of young team handball players. *The Open Sports Sciences Journal*, 2(1), 22-28.