Talent Classification of Motoric Parameters with Support Vector Machine

Hanife Kanat Usta[1], Naci Usta[2], Adil Deniz Duru[3], Hasan Birol Çotuk[4]

Abstract

Aim: In recent years, the methods of analysis of data science have started to be used frequently in talent selection in sports and the evaluation of athletes. Based on the motor and physical measurements of the future athletes, determining which sport branch they are prone to is important in terms of training and resource planning. Within the scope of this study, it was aimed to propose a classification system to determine which sports branches the participants are suitable for, based on motor and physical measurements.

Material and Methods: Measurements of height, arm span, body weight, 20-meter sprint test, vertical jump height, 1 kg medicine ball throw, back strength, hand grip strength, flexibility test and standing long jump values were recorded with the contribution of 1240 participants who are 9 years old. Afterwards, grouping procedures were carried out with classification methods based on Support Vector Machines (SVM). Radial based functions are used as kernel functions of SVM. The results of evaluations made by consulting expert opinion beforehand were accepted as actual values, compared with the classification results and the performances of the classifiers were calculated. Within the scope of this study, participants were classified into four as rapidity branch (E), strength branch (F), height branch (G) and other group (H).

Results: The accuracy values of classification of support vector machines were found ranging from 96% to 100% in each class, and 98% in average. Minimum value of sensitivity was found to be 93% while it was 99% in maximum. On the other hand, precision varied between 92% and 100%.

Conclusion: In the light of the information provided, successful classification of the test dataset using the the model that is formed by the training dataset, points out a possible high classification accuracy of big test datasets even in the use of a small dataset in the training phase.

INTRODUCTION

Given that the sport contributed much to the individual and the society, it is an undeniable fact that it has become a social need requiring persistence. It is thought that directing athletes to the branches where they can use the features they have at a young age in the most appropriate way is the most fundamental and the most important step in their achievement, in order for them to achieve success in the sports branch they are involved in. In studies of talent selection, a framework that is the golden standard in practice has not been established until ten years ago (Vaeyens, Lenoir, Williams & Philippaerts, 2008). "Sports talent" usually refer to the individuals who are thought to have a high degree of predisposition or a particular tendency for sports efficiency due to hereditary or later gained behavioral conditions (Karl, 2001). While trainers of elite athletes see themselves as experts when their knowledge and skills they gained in their branches are enriched with their experiences, sports scientists have a strong belief that only the measurements carried out can represent reality (Buekers, Borry & Rowe, 2015). These two ideas should be blended for success in sports. In recent studies, qualitative and numerical methods are used in the literature to determine the position of the players for the team game. Based on numerical methods, Bayesian networks, Decision trees and k-nearest neighbors have been shown to achieve high accuracy in making decision by eliminating personal prejudices (Razali, Mustapha, Yatim & Aziz, 2017). The mental, physical and technical skills of the players are taken into account and the players' states in different positions are classified with 98% of accuracy. The trainer was provided with a system with which he could quantitatively measure the strengths of the players, and it was thought to support him in deciding. As a matter of fact, a test scale was developed by bringing it into the use of 20 trainers in order to evaluate the system qualitatively and 80 per cent of the users found the system usable. In a recent study, relation between performance and anthropometric measurements obtained from young

wheelchair basketball players were investigated and the sitting height and functional ability were found to be correlated with performance (Cavedon, Zancanaro & Milanese 2015).

Today, advances in computer science have facilitated the processing of large-scale data and accelerated the information extraction. In a recent study carried out by Woods, Veale, Fransen, Robertson & Collier (2018), players were classified in terms of their positions according to their technical abilities. In their study, 12 game positions of 4 Australian elite football players were classified using three separate analyzes, and achieved a maximum accuracy rate of 70 per cent in the analysis results. All these developments can be considered to be a sign of the importance of the talent orientation decision support system in directing the players to the branches.

Artificial Neural Networks (ANN) as a part of machine learning strategy was implemented to perform the prediction of sports results (Bunker & Fadi, 2017). The prediction of results regarding to horse races, (Davoodi & Khanteymoori, 2010) and the results of rugby and soccer matches (McCabe & Trevathan, 2008) were computed by the use of previous data in ANN.

Within the scope of this study, classification results were achieved by comparing the values obtained with the application of the classification techniques which accept the parameters including the physical and motor characteristics as input to the trainer views. The grouping of the athlete candidates based on their abilities was carried out according to the trainer's view, and grouped by using Support Vector Machines.

In the materials and methods section, information is provided about how motor properties are gathered and the classification method being applied. The findings section contains the results of the classification method applied. Finally, the discussion section includes the comparison of the classification findings with the literature.

METHODS

Physical parameters

In order to train and test the classification methodology, the participants' measurements include height, arm span, body weight, 20-meter sprint test, vertical jump height, 1 kg medicine ball throw, back strength, hand grip strength, flexibility test and standing long jump tests.

Height was measured by using tape measure with the help of a tape sticked to the wall. The athletes, without shoes, put their heels together and lean on the tape measure, and the arms are hanging freely on their sides. The back, hips, back of the head and heels stand upright verging to the vertical scale. The subject stands by first taking a deep breath and in this position the ruler is brought to the top of the head and the hair is compressed sufficiently and the measurement is noted (Coşan, Demir & Mengütay, 2002) The subject leans on the wall, arms are stretched parallel to the floor and back of the hand touches the wall; while in this position, the arm span parameter was calculated with the help of a tape measure by measuring the distance between the middle fingers of the right and left arms (Coşan et al.) Body weight (kg) is known to be strongly related to muscle mass. In some other sports branches, the basic force is predominant (Gündüz, 1997). The 20 m sprint (secs) is a 20 meters flat track that is precisely measured and its start and finish lines are clear. The starting position is standing. The subject starts when he is ready and tries to complete the 20-meter course as soon as possible. When they start, the photocell works and when they reach the end zone, the photocell stops. They run twice and the best rating is taken into account (Kamar, 2003). An amount of chalk enough to leave a mark on the wall is applied on the middle finger of the athlete for the vertical jump height measurement (cm), and he touches vertically the highest point where the arm can reach, beneath the jumping platform. This part is taken as the zero point and its difference with the highest point is determined as the jump test. At the moment of jump, the subject stands at the tip of the toe as the knee is slightly twisted. With a sudden movement, the vertical legs bounce and the arms are lifted up to the highest point that he can touch on the marking platform. The distance between the marking part and the jump height gives the jump value. The measurement is taken twice and the best attempt is recorded (Coşan et al.). For 1 kg medicine ball throw (m), the athlete tries to throw a medicine ball weighing 1 kg, with both hands over the head, as far away as possible from where he stands. The athlete takes his position just behind the starting line and one step ahead. It is not allowed to run at the moment of throwing. By bending his body backwards, the subject can obtain the necessary acceleration for the throw. The medicine ball should be thrown using both hands. The test

score is the recording of the throwing distance in meters and centimetres. The best score obtained at the end of two trials is recorded (Kamar). For back strength measurement, the subject stands on the dynamometer platform with his back straight, his head upright and his knees tight. While grasping the bar, the right-hand palm faces the body and the back of the left hand faces the body and the chain is adjusted to form the intended uprightness. The subject pulls the bar strongly upwards using the back muscles, without bending backwards. The shoulders move backwards during pulling. Prior to the movement, the subject should bend his body forward very mildly, with his head held upright. The needle of the dynamometer stops at the point where the maximum is reached. Two trials are carried out with a minute interval. For the hand grip strength test, the hand grip measure is usually adjusted in a way that the subject can use easily. The hand length can be measured with a caliper and the value found can be used for the optimal grip size. The subject stands upright, his arms are on his side. The dynamometer is held on the side, in a parallel position to the body. The dynamometer is squeezed as strongly as possible without moving the arm. Generally, 3 trials are carried out with 1-minute intervals, for both hands (Özer, 2006). For the flexibility test, the subject is seated on the floor and the sole of the foot is rested against the test stand in an upright position. Leaning forward from the body, the knees being straight, the hands being held ahead, the subject reaches forward as far as he can. He waits for two seconds before stretching forward or backward. The test is repeated twice and the higher value is accepted (Tamer, 1991). For standing long jump, tape measure is sticked onto a flat ground. The athlete is positioned behind the starting line. Tips of the athlete's toe take a relaxed standing position behind the starting line. Tips of the subject's toe should be behind the starting line. With the command heard, the subject pulls the arms backwards and tries to jump as far as possible from the starting line. The distance between the starting line and the nearest heel of the athlete to this line is recorded as the score. Two trials are performed and the best rating is recorded in cm (Kamar)

Participants

This study involves the measurement values of 9-year-old students who participated in the study voluntarily within the scope of the project called "Talent Identification in Sports and Sports Orientation", conducted jointly by Gaziosmanpaşa District Governorate, Gaziosmanpaşa Municipality, Gaziosmanpaşa District Directorate of National Education and Gaziosmanpaşa Youth Services and District Directorate of Sports in the province of Gaziosmanpaşa, Istanbul. This study received ethical approval from Ethics Committee of Institute of Health Sciences of Marmara University dated 08.01.2018 and by decision no 2018-13. In the scope of the study, data from 4183 participants were collected between 2015 and 2017 and all participants were divided into four different groups according to the expert trainers' opinions. These groups were characterized as height, strength, rapidity, and other. The number of members of the groups was 310, 314, 444 and 3115, respectively. Since the number of participants included in the last group is very high compared to other groups and it can bring about bias in classification performance, it is necessary to equalize the number of participants in the groups. This can be done in two different ways. Increasing the number of participants in the group with the smallest number of members can be considered as a solution, but it does not seem practical in the context of this study. On the other hand, another method is to select participants from the other groups by random sampling, equal to the number of members in the group with the smallest number of members. Within the scope of this study, 310 participants from each group were randomly selected, and a total of 1240 participants' data were selected for use. 60% of these participants were used to train classifiers, and the remaining 40% were used for testing purposes. Thus, the samples of 186 participants belonging to each group were used for the training of classifiers, while the samples of 124 participants were used for testing purposes.

Classification with Support Vector Machines: Support vector machines, as a supervised learning method, perform multiclass classification process and all possible binary classifications, and complete the classification process by using the knowledge of which classes each training data belongs to the most. Support vector machines use hyperplanes to reduce the classification error by dividing training data belonging to two classes so that data is farthest from each other. In this study, classification process was applied in Phyton environment. SVC from the Scikit library was used to implement classifier. Radial basis kernel was used in SVM.

Classifier Performance Metrics: The performance of the classifier is evaluated with precision, sensitivity, accuracy, f-score and support parameters by looking at the number of correct and incorrect measurements. Table 1 represents the confusion matrix for four classes.

Table 1. Definitions of confusion matrix belonging to multi-class classifier

	True: Class E	True: Class F	True: Class G	True: Class H
Predicted: Class E	True	False 2-> 1	False 3-> 1	False 4-> 1
Predicted: Class F	False 1-> 2	True	False	False
Predicted: Class G	False 1-> 3	False	True	False
Predicted: Class H	False 1-> 4	False	False	True

When Table 1 is examined, the correct classification is observed as diagonal elements. In addition, while the columns show the correct classes, the lines represent the predicted classes. Accuracy is the ratio of the sum of the numbers getting to the diagonal elements to the sum of the numbers of the elements appearing in all the cells. The ratio of the diagonal element on a column to the sum of the elements on that column is called the accuracy parameter of that class. Precision is the ratio of the number of correctly predicted data, which are called true positives, to the data determined incorrectly to be in that class, which are called false positives. False positives are also referred to as type 1 error in statistics. The ratio of true positives to the sum of true positives and false negatives is defined as recall/sensitivity. False negative is the state of mistakenly identifying a faulty condition.

Sensitivity criterion and precision criterion, when evaluated separately, are not sufficient for us to get meaningful comparison outcome. In order to obtain the correct result, the parameter defined as the weighted harmonic mean of precision and sensitivity is called F-score. It is obtained when twice the product of sensitivity and precision parameters is divided by the sum of the sensitivity and precision parameters. The number of elements per class is indicated by the support parameter.

RESULTS

The accuracy values of the SVM reached to 98% in average over the classes. The classification evaluation parameters achieved the values close to %100 which is an indicator of the high performance classification. Results of the SVM was summarized in Figure 1 and Table 2. In Figure 1. the relationship between the true positive and false positive values was drawn. The dotted black line represent the random classification performance. As seen from Figure 1., for each class, accuracy values were 100%, 99%, 98% and 96%, respectively.

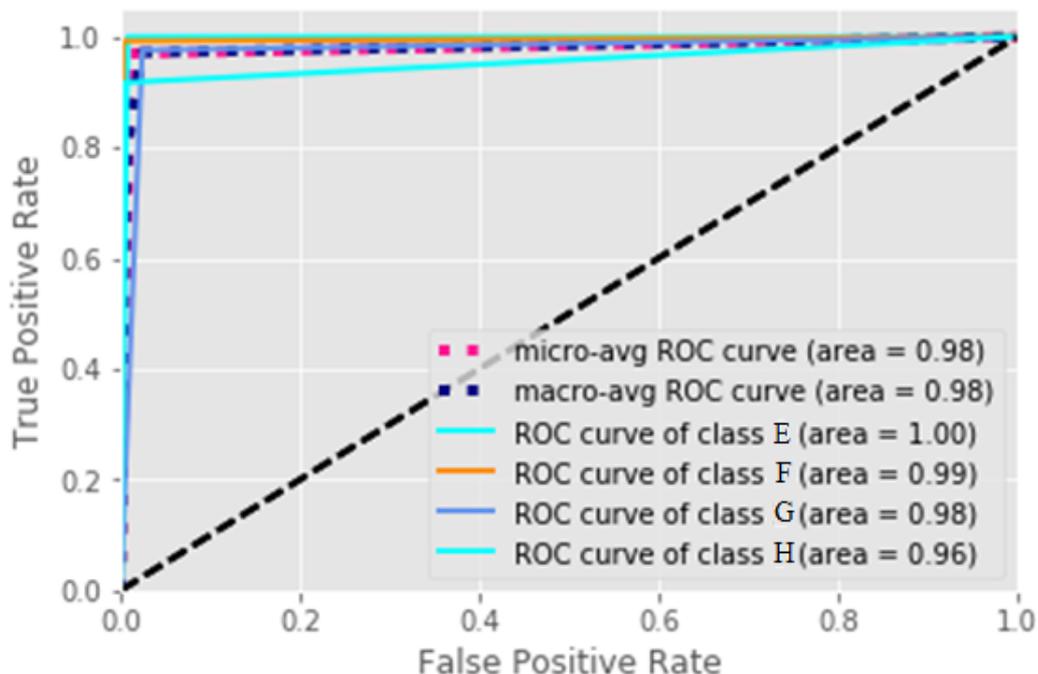**Figure 1.** Support vector machines classification scores

Table 2 Decision support tree classification technique test data confusion values

Variables	E	F	G	H
Precision	1	0.99	0.98	0.92
Recall	0.97	0.99	0.93	0.98
F1-score	0.99	0.99	0.95	0.95
Support	116	126	130	124

CONCLUSION

The popularity of sports throughout the world has increased, ensuring more focus on the performance of the athletes. Unless the physical characteristics of the athletes are not suitable for the sport branch they are in, their performance will not reach adequate levels (Aydos, 1991). In individual and team sports, athletes must have body structures that are in accordance with the characteristics of the branch in order to reveal their physiological, physical and motoric properties. In these conditions, it is necessary to select athletes considering the needs of the branches. The properties required by various sports branches differ. Badminton sports, for instance, require sudden moves in every direction, bounces and sudden changes of direction (Sharif, George & Ramlan, 2009). Since Badminton is a sport involving speed, talent, mobility, reaction and aesthetics, badminton athletes should have good reaction, jumping strength and speed, and they should have a slim body structure. It is believed that if the badminton player has a good level of speed in terms of performance, this will provide a significant advantage for the elbowroom in the court to reach from the midcourt to the corners and from the corners to the center point (Omosegoard, 1996). Athletics, which is another branch, embodies distance, time and height. The runners represent the fight against time, the jumps against the distance, and the throws against the height (Çalışkan, 2013). The height criteria of the athletes have a direct effect on performance in some sports branches, but they also affect performance indirectly in some other branches (Gündüz, 2005). The anthropometric feature and jumping capacity of the volleyball player is an important factor directly affecting the performance of the team (Clarke, 1975). In branches such as volleyball and handball, it is a visible feature that athletes are tall, as well as having low body fat percentage. It is thought that it provides an important advantage in defensive and offensive positions in basic technical and tactical games (Pehlivan, 1997). According to the International Amateur Wrestling Federation (FILA), wrestling is defined as a struggle for supremacy between two individuals, without the use of any tool, on a mattress in a designated size, with the use of techniques, skills, powers and mind and within a framework of rules (Öcal, 2007). In addition, wrestling is a sport that makes progress depending on the strength of the body to a large extent (Cicioğlu, Kürkçü, Eroğlu & Yüksel, 2007). Boxing is one of the fighting sports in the world that requires a high level of power and has a complex structure due to its dynamic and static characteristics (Mitchell, Willams & Reter, 1999). Boxing workouts provide the athlete with a major improvement in physical and physiological characteristics, including aerobic power, muscle strength and endurance, hand eye coordination and flexibility, rapidity and reflexes. In order to be able to exhibit high performance in the football field, it is considered that if the athletes have a high level of explosive speed and durability in terms of physical properties, and if their motor skills are blended with anthropometric properties, it is an inevitable success (Figueirdo, Gonçalves & Coelho, 2009; Vaeyens, Malina, Janssens, Van Renterghem, Bourgois, Vrijens et al., 2006). In a high-level soccer match, it was measured that the professional footballers covered approximately 10 km of the total distance they had during the game with an intensity close to anaerobic threshold, at a maximal heart rate (80-90%) (Stolen, Chamarı, Castagna, & Wisloff, 2005) Football is a sports branch where aerobics and anaerobic loads are used intensively, intermittent rests are present, and performance is created by speed, agility, flexibility, mobility, coordination and muscular endurance (Akgün, 1994). In terms of its technical properties, judo is a defensive art that does not show resistance to the reaction of the opponent, and even uses opponent's strength to defeat the opponent himself (Manfred, 1979). One of the most important factors in the basketball branch is the height. By measuring the height of the athlete, a number of predictions can be made for his height in the future. Evaluating the athlete's parents' height also contributes positively to the estimates (Magill, 1989). Looking at the studies above, speed, strength, height, coordination and endurance characteristics are needed for each sport branch. However, when the needs of the branches are examined, the priority order of these properties is constantly changing. In line with this need, we can classify sport branches as rapidity branches, strength branches and height branches.

Metodologically, there is a limited usage of classical statistical tests in the talent grouping studies since the datasets being used have high inter parameter correlation as well as high number of parameters (Till, Ben, Jones, Cobley, Morley, O'Hara et al., 2016).

In a recent study, a binary classification study was performed for a group soccer group. The groups were designed as selected players and non-selected players. Based on their perceptual and cognitive parameters, high classification level was reported (93.7%) (O'Connor, Larkin & Mark 2016). However, in our study, more groups were used and better evaluation scores were achieved. Within the scope of this study, participants were classified into four as rapidity branch (E), strength branch (F), height branch (G) and other group (H). The classification accuracy of support vector machines varied from 96% to 100% in each class, and 98% in average. In the light of the information provided, it has been observed that the support vector machines have achieved the high performance. The model formed using the training set, classified the dataset which has not been used in the training with a high accuracy points out a possible high classification accuracy of big test datasets even in the use of a small dataset in the training phase.

CONCLUSION and RECOMMENDATIONS

The classification to an increased number of class is going to be enhanced by the use of physiological and psychological parameters in the training dataset. The fact that classifier techniques use a knowledge gained from existing data to create a decision support system will make it easier to introduce a model that will contribute to the talent selection. In future studies, it is aimed to classify data belonging to different age groups and to test systems with larger number of samples.

REFERENCES

Akgün, N. (1994) Egzersiz Fizyolojisi. İzmir: Ege Üniversitesi Basımevi.

Aydos, L. (1991) Fiziksel Uygunluk, Gazi Eğitim Fakültesi Dergisi, 69-79.

Bunker, R.P. & Fadi, T. (2017) A machine learning framework for sports result prediction, Applied Computing and Informatics, https://doi.org/10.1016/j.aci.2017.09.005

Buekers, M., Borry, P., Rowe, P. (2015) Talent in sports. some reflections about the search for future champions. Movement & Sport Sciences – Science & Motricit´e; 88, 3–12

Cavedon, V., Zancanaro, C., Milanese, C. (2015) Physique and Performance of Young Wheelchair Basketball Players in Relation with Classification. PLoS ONE 10(11): e0143621. https://doi.org/10.1371/journal.pone.0143621

Cicioğlu, H., Kürkçü, R., Eroğlu, H., & Yüksel, S. (2007) 15-17 Yaş güreşçilerin fiziksel ve fizyolojik özelliklerinin sezonsal değişimi: Spormetre Beden Eğitimi ve Spor Bilimleri Dergisi 5: 1 51-6.

Clarke, O.H. (1975) Exercise Physiology, Prentice Hall. New Jersey, USA.

Coşan, F., Demir, A., Mengütay, S. (Ed). (2002) Türk Çocuklarının Fiziki Uygunluk Normları. İstanbul Olimpiyat Oyunları Hazırlık ve Düzenleme Kurulu Eğitim Yayınları Yayın No 1., İstanbul.

Çalışkan, O. (2013) Özel düzenlenmiş Plyometrik antrenmanların atletizm yapan 11-13 yaş çocukların aerobik ve anaerobik güçlerine etkisi. Yüksek Lisans Tezi, Aksaray Üniversitesi Sosyal Bilimler Enstitüsü. Aksaray.

Davoodi, E. & Khanteymoori, A. (2010) Horse racing prediction using artificial neural networks, Recent Adv. Neural Networks, Fuzzy Syst. Evol. Comput., pp. 155-160

Figueiredo, A.J., Gonçalves, C.E., Coelho, E., et al. (2009): Youth soccer players, 11-14 years: Maturity, size, function, skill and goal orientation. Annals of Human Biology (36), 60-73.

Gündüz, N. (2005) Antrenman Bilgisi, Birinci baskı, İzmir, Saray Medical Yayıncılık, sf. 21 22.

Gündüz, N. (1997) Antrenman Bilgisi. Saray medikal yayıncılık 2. Yayın İzmir; s.23

Kamar, A. (2003) Sporda Yetenek Beceri ve Performans Testleri. Nobel Yayın Dağıtım. Ankara; s.40-41

Karl, K. (2001) Sporda Yetenek Arama Seçme ve Yönlendirme Antrenör Eğitim Dizisi. Bağırgan Yayınevi., Ankara.

Magill, A.R. (1989) Motor learning Concepts and Applications, Third Ed. Iowa, Wch, Publishers, 17-34

Manfred, B. (1979) Training, Technik, Taktik: Reinbek. International Congress on judo. 1979

Mccabe, A., & Trevathan, J. (2008). Artificial Intelligence in Sports Prediction. 1194-1197. 10.1109/ITNG.2008.203.

Mengütay. S., Demir, A., Coşan, F. (2002) Olimpiyatlar İçin Sporcu Kaynağı Projesi, Olimpiyat Hazırlık ve Düzenleme Kurulu Eğitim Yayınları. İstanbul No 2, 102.

Mitchell, H., Willams, L., & Reter, B.R. (1999) Classification of sports medicine and science in sports and exercise. American college of sports. Medicine and the American College of Cardiology, 56-85.

O'Connor, D., Larkin, P., Mark, W.A. (2016) Talent identification and selection in elite youth football: An Australian context. Eur J Sport Sci. Oct;16(7):837-44

Omosegoard, B. (1996) Physical Training for Badminton. İnternational Badminton Federation (JJBF)Denmark.

Öcal, D, (2007) Elit Güreşçilerin Somototip Özellikleri ile Anropometrik Oransal İlişkilerinin Stiller ve Sikletler Arası Karşılaştırılması. Yüksek Lisans Tezi, Gazi Üniversitesi Sağlık Bilimleri Enstitüsü ve Spor Anabillim Dalı, Ankara.

Özer, K. (2006) Fiziksel Uygunluk. 2. Baskı, Nobel Yayın Dağıtım. İstanbul s. 11, 118-120, 160.

Pehlivan, Z. (1997) 1995-1996 sezonunda Türkiye 1. Deplasmanlı Bayanlar Basketbol, Hentbol ve Voleybol Liglerinde Şampiyon olan sporcuların fiziksel ve fizyolojik özelliklerinin değerlendirilmesi; Gazi Üniversitesi Sağlık Bilimleri Enstitüsü Yüksek Lisans Tezi 1997.

Razali, N., Mustapha, A., Yatim, A.F., Aziz, A.R. (2017) Predicting player position for talent identification in association football. International Research and Innovation Summit (IRIS2017). IOP Conf. series: materials science and engineering.; 226 012087

Sharif, A.H., George, J., Ramlan, A.A. (2009) Musculoskeletal Injuries Among Malaysian Badminton Players. Singapore Med J. Nov; 50(11); 1095-7.

Stolen, T., Chamari, K., Castagna, C., Wisloff, U. (2005). Physiology of soccer. Sports Medicine. 35(6):501–536.

Tamer, K. (1991) Fiziksel Performansın Ölçülmesi ve Değerlendirilmesi. (Egzersiz Fizyolojisi Laboratuvar Rehberi) Gökçe offset matbaacılık Ankara. S.32

Till, K., Ben, L., Jones, S., Cobley, D., Morley, J., O'Hara, C., Chapman, C., Cooke, and Clive B. Beggs, (2016) Identifying Talent in Youth Sport: A Novel Methodology Using Higher-Dimensional Analysis, PLoS One. 2016; 11(5): e0155047.

Vaeyens, R., Lenoir, M., Williams, M.A., Philippaerts, M.P. (2008) Talent identification anddevelopment programmes in sport current models and future directions. Sports Med. 38 (9): 703-714

Vaeyens, R., Malina, R.M., Janssens, M., Van Renterghem, B., Bourgois, J., Vrijens, J. et al. (2006) A multidisciplinary selection model for youth soccer. The Ghent Youth Soccer Project. British Journal of Sports Medicine (40), 928-934.

Woods, T.C., Veale, J., Fransen, J., Robertson, S., Collier, N.F. (2018) Classification of playing position in elite junior Australian football using technical skill indicators, Journal of Sports Sciences, 36:1, 97-103.

CITATION OF THIS ARTICLE