Andrej Panjan[1, 2]
Nejc Šarabon[3]
Aleš Filipčič[4]

THE PROGNOSTIC VALUE OF MACHINE LEARNING METHODS IN TENNIS

NAPOVEDNA VREDNOST METOD STROJNEGA UČENJA V TENISU

ABSTRACT

The purpose of this study was to assess the possibilities of predicting playing successfulness in competitive tennis by using machine learning methods applied to young players' motor abilities and morphological test results. The classification of players according to their competitive successfulness was performed using several methods: the naive Bayes classification method, decision tree, the C4.5 algorithm, the k-nearest neighbour, support vector machine (SVM), and logistic regression. After discretising the players' successfulness into quality classes, the possibility of automatically identifying the most promising attributes was tested using the ReliefF method and the wrapper approach. Both the naive Bayes method with ReliefF and logistic regression with the wrapper approach proved to be accurate predictors of competitive performance in the age group under 12 years and in the age group between 12 and 16 years. The most promising attribute was racquet ball handling. Predictions of the competitive performance of tennis players proved to be a highly complex issue because the accuracy of the prediction models in our study, based on morphological and motor factors, was relatively poor.

Key words: tennis, identification, selection, predictability, competitive performance, machine learning

[1]*S2P Ltd., Laboratory for Motor Control and Motor Behaviour, Ljubljana, Slovenia*

[2]*Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Koper, Slovenia*

[3]*University of Primorska, Andrej Marusic Institute, Department of Health Study, Koper, Slovenia*

[4]*University of Ljubljana, Faculty of Sport, Ljubljana, Slovenia*

Corresponding author:
Prof. Dr. Sci. Nejc Šarabon
University of Primorska, Andrej Marusic Institute,
Department of Health Study, Muzejski trg 2,
SI-6000 Koper, Slovenia, e-mail: nejc.sarabon@s2p.si"

IZVLEČEK

Namen te študije je bila ocena zmožnosti napovedovanja igralne uspešnosti v tekmovalnem tenisu z uporabo metod strojnega učenja na rezultatih motoričnih in morfoloških testov mladih tekmovalcev. Razvrstitev tekmovalcev glede na njihovo tekmovalno uspešnost je bila narejena z več metodami: naivni Bayes, odločitveno drevo, C4.5 algoritem, k-najbližjih sosedov, metoda podpornih vektorjev in logistična regresija. Po razvrstitvi igralcev v kakovostne razrede sta bili za avtomatsko iskanje najobetavnejših atributov uporabljeni metodi ReliefF in metoda ovojnice. Za napovedovanje tekmovalne uspešnosti v starostni skupni pod 12 let in starostni skupini med 12 in 16 let sta bili najuspešnejši metodi naivni Bayes z ReliefF in logistična regresija z metodo ovojnice. Obvladovanje žogice z loparjem se je izkazalo za najbolj obetaven atribut. Napovedovanje tekmovalne uspešnosti teniških igralcev se je izkazalo za zelo kompleksen problem, zato ker je bila točnost napovedovalnih modelov na podlagi morfoloških in motoričnih dejavnikov sorazmerno slaba.

Ključne besede: tenis, identifikacija, izbira, napovedovanje, tekmovalna uspešnost, strojno učenje

INTRODUCTION

Tennis is one of the most popular sports in the world, both in terms of its widespread competition system and the attention the media pay to it, as well as the large numbers of people of both sexes and all ages that participate in it. As is the case of all sports today, tennis is no exception when it comes to the fact that it is the top players who steer the development of the sport as well as of the training and diagnostic technologies used. The absolute competitive performance of a tennis player is indicated by their position on the ATP ranking list. This position is the cumulative result of numerous factors, procedures and activities. Factors with an important role in the development of any young tennis player include those involved in the initial selection procedure, in planning and in organisation, as well as how the subsequent training and its supervision are performed. In particular, supervision over the effects of training in various areas of the tennis player's bio-psycho-social status can influence the efficiency of the sports training system as a whole to a large extent, and thus of the absolute competitive performance of the player as well. In view of this, tests pertaining to main body characteristics and motor abilities are an indispensable part of the process, one of whose goals is a more in-depth understanding of the relations between the bio-functional potential and the player's actual competitive performance. With regard to analyses of this kind, classical statistical approaches and subjective estimates by experts have been the most commonly used in kinesiology. In our opinion, some state-of-the-art data analysis methods from the artificial intelligence field can assist us in the search for more reliable and objective means of evaluations of such a kind, and thereby contribute to taking more efficient decisions in the course of the development of young players.

Machine learning is an artificial intelligence field which deals with discovering knowledge in data by data analysis and by the automatic generation of knowledge databases for expert systems for the construction of numeric and qualitative models using classification and regression analyses etc. In recent years, we have witnessed a rapid increase in the volume of data in digital form. Machine learning is becoming an important tool for transforming these data into useful information since the manual processing of such a vast quantity of data has become impossible. The increased recognition of machine learning is also reflected in a rising number of commercial systems within the sectors of industry, medicine, economics, banking etc. The core principle of machine learning is the automatic modelling of data. Learned models attempt to interpret the data from which the models were constructed. They can assist in making decisions when it comes to studying the modelled process in the future (predictions, diagnosis, control, verification, simulations etc.).

Predictions of competitive performance can be made using either classification or regression methods. What both approaches have in common is that out of a multitude of data (independent variables or attributes) they can construct a model whose output is a dependent variable (class) (Hand, Mannila, & Smyth, 2001). There are two types of class: discrete and continuous. The output of the classification model is a discrete class, whereas the output of the regression model is a continuous class. Classification and regression can employ various methods based on various approaches, which is why there are differences in how well different methods perform with regard to different issues. It is possible to improve the reliability of individual methods by selecting only the most promising attributes (Kohavi & John, 1997; Kononenko & Kukar, 2007). In doing so, the attributes that do not influence the class are eliminated and, therefore, mostly exert a

negative effect on the performance of the learned model. Individual attributes may be eliminated manually or by using various automatic methods.

The aim of this study was to assess the possibilities of predicting performance in competitive tennis by employing machine learning methods on the basis of results of measurements of motor and morphological tests of young tennis players. The efficiency of predicting competitive performance was studied with regard to various age groups in male and female categories, both within the individual age categories, as well as in advanced categories. Our final aim was to identify those attributes which prove most useful in making predictions. For this purpose, we employed two methods for attribute selection.

A complementary study on the same sample of subjects was published in 2010 (Panjan, Šarabon, & Filipčič, 2010). This study provides an additional in-depth presentation of various classification methods and results of regression analysis without using methods for attribute selection.

Materials and methods

Subjects

The sample of subjects included those Slovenian tennis players who were positioned on the ranking list of the Slovenian Tennis Association in individual periods and who also underwent morphological and motor measurements in those individual periods. Measurement data were collected for 593 male tennis players and 409 female tennis players, i.e. 1,002 individual tennis players in total. The data collection procedures met international ethical standards and were consistent with the Declaration of Helsinki. The selected subjects were divided into three age groups: 12 years and under; between 12 and 16 years; above 16 years. The entire sample of measurements was then divided into age categories for the analysis of the predictability of competitive performance:

– Age category U12/U12: for subjects in the age group of 12 years and under on the basis of measurements performed in the period of 12 years of age and under, consisting of 170 male tennis players (age 12.14±1.02 years, body height 153.53±7.95 cm, body weight 42.85±7.58 kg) and 157 female tennis players (11.85±.75 years, 155.74±8.16 cm, 44.02±8.33 kg).

– Age category 12–16/12–16: for subjects in the age group between 12 and 16 years on the basis of measurements performed in the period between 12 and 16 years of age, consisting of 341 male tennis players (14.88±1.20 years, 170.35±10.08 cm, 58.43±11.49 kg) and 215 female tennis players (14.80±1.19 years, 166.65±6.18 cm, 55.93±7.46 kg).

– Age category A16/A16: for subjects in the age group above 16 years on the basis of measurements performed in the period above 16 years of age. This sample consisted of 82 male tennis players (18.87±2.53 years, 182.73±5.72 cm, 73.37±6.69 kg) and 37 female tennis players (18.07±1.78 years, 169.88±6.22 cm, 62.59±7.95 kg).

– Age category 12–16/U12: for subjects between 12 and 16 years, but on the basis of measurements performed in the period of 12 years of age and under consisting of 89 male tennis players (12.03±1.02 years, 156.34±7.82 cm, 44.82±7.76 kg) and 84 female tennis players (11.96±.71 years, 157.68±7.03 cm, 45.42±8.12 kg).

– Age category A16/U12: for subjects in the age group above 16 years, but on the basis of measurements performed in the period of 12 years of age and under. This sample consisted of 47

male tennis players (12.09±1.25 years, 157.02±8.26 cm, 45.20±8.51 kg) and 35 female tennis players (11.92±1.18 years, 159.22±6.94 cm, 46.73±7.28 kg).

– Age category A16/12–16 for subjects in the age group above 16 years, but on the basis of measurements performed in the period between 12 and 16 years of age. This sample consisted of 125 male tennis players (14.80±1.27 years, 175.00±9.19 cm, 63.66±11.11 kg) and 79 female tennis players (14.94±1.17 years, 168.29±6.03 cm, 58.88±7.54 kg).

Data collection

Measurements were made on a selection of independent attributes whose usefulness in predicting competitive performance in tennis had already been identified. These measurements test both general and tennis-specific motor abilities of players, as well as morphological attributes. They

Table 1: Applied morphological and motor tests

Abbreviation	Measure and Test	Ability/Dimension
ATV	Body height	Morphology
ATT	Body weight	Morphology
BMI	Body mass index	Morphology
AMASPP	Fat tissue percentage	Morphology
AMISP	Muscle tissue percentage	Morphology
AKOSP	Bone tissue percentage	Morphology
MSARG	Sargent test	Explosive power – lower ext.
MM2	Medicine ball throw (2 kg)	Explosive power – upper ext.
MSKOK4	Four-jumps test	Explosive power – lower ext.
MDT60	Sit-ups	Muscular endurance – trunk
MT20	20-metre sprint	Sprint acceleration
MT9X6	9 x 6-metre sprint test	Agility
MREAK	Reaction pole	Reaction time
MTAPNO	Foot tapping	Alternative movements' frequency – lower ext.
MTAPRO	Hand tapping	Alternative movements' frequency – upper ext.
MTPK	Forward bend	Passive flexibility – lower ext.
MZVIN	Sprain with a stick	Passive flexibility – upper ext.
MIZPK	Lunge	Active flexibility – lower ext.
MPAH	Fan	Agility
MHEK	Hexagon test	Agility
MHST	Stamping test	Coordination – lower ext.
MPOL	Obstacle course backwards	Coordination – whole body
MOZL60	Racquet ball handling	Coordination – tennis-specific
MOSMI	Figure-of-eight sprint with bending	Agility – tennis specific
MOBRAT	Balance beam turnarounds	Dynamic balance
MHOJA	Balance beam walk with racquet ball handling	Dynamic balance
MPRIS	Side steps on balance beam	Dynamic balance
MT2400	2400-metre run	Endurance

were conducted annually in the laboratories of the Faculty of Sport in Ljubljana between 1993 and 2008. The tests of general and tennis-specific motor abilities examined all key areas of a player's motor and functional abilities (strength, speed, agility, flexibility, balance, coordination, endurance). Table 1 presents the composition of this test battery.

The position on the ranking list of the Slovenian Tennis Association for an individual year was used as the primary criterion for estimating competitive performance. This ranking list takes the five best results achieved in that competition year into account. The position on the Tennis Association ranking list is determined on the basis of a coefficient which represents the total number of points won by an individual player, divided by the number of tournaments played.

Data processing

In analysing the predictions of competitive performance with classification algorithms, it was necessary to discretise the sample data since classification algorithms do not work with continuous classes. The class was determined by the position on the ranking list of the Slovenian Tennis Association, and was divided into two quality groups, i.e. the top ten players, and others. The reason for this reflected our aim to separate top players from the rest because only top players can succeed on the international level. The average position across all years in a period was taken into the analysis. Similarly, the attributes were also averaged across all years in a period. Classification was performed by means of several methods: the naive Bayes classification method, decision tree, the C4.5 algorithm, the k-nearest neighbour, support vector machine (SVM), and logistic regression. These methods use considerably different approaches (with the exception of the decision tree and the C4.5 algorithm since the C4.5 algorithm is a variation of the decision tree). Each of these methods is able to predict discrete classes. The simplest method is the k-nearest neighbour method, whereas the most complex one is SVM (Kononenko & Kukar, 2007). The evaluation of the performance of classifiers was conducted with classification accuracy using the 10-fold cross-validation method.

The naive Bayes classifier is a simple probabilistic classifier based on applying the Bayes theorem with assumptions of conditional independence of values of different attributes with regard to the given class. In spite of this, it performs much better than might be expected when it comes to a number of complex actual issues (Zhang, 2004). The decision tree is a tree-like structure whose leaves represent classifications, whereas its internal nodes are conjunctions of attributes which lead to classifications. Interpretations of such structures are simple, which is one reason that decision trees are quite commonly used in practice. C4.5 is an algorithm used to generate a decision tree which was developed by Quinlan (1993). The k-nearest neighbour is an algorithm for classification based on closest training examples in the attribute space. The classification of a new case is made on the basis of k-nearest neighbour votes (the neighbour being assigned to a certain class) by selecting the class which receives the majority of votes. SVM is one of the most successful classification methods. Unlike the majority of machine learning algorithms, which aim to minimise the number of attributes, the SVM method uses as many attributes as possible, out of which the method itself selects a suitable combination that leads to the needed information. Logistic regression is a method which generates a linear model on the basis of a transformed predictor variable. The transformed variable is approximated by using the linear function in the same way as with linear regression (Witten & Frank, 2005).

Regression analysis works with a continuous class and therefore no discretisation was necessary. Here, the class was also represented by the position on the ranking list of the Slovenian Tennis Association. Before running the regression analyses, the position on the ranking list and independent attributes were averaged across all years in a period. Regression analysis was conducted by linear regression and by regression trees. Linear regression is commonly used in practice and is based on modelling the relation between attributes and the class so that a linear model is obtained. The linear regression model was calculated by the approach of minimising the sum of the squares error. In principle, regression trees are the same as decision trees, except that they are able to predict a continuous class (Witten & Frank, 2005). This is why in the leaves there are functions which transform the attribute value vector into the continuous class. The evaluation of performance was conducted by means of the relative absolute error (Orange, 2010). Both regression algorithms are unable to handle cases with missing values and thus such cases must be removed before the analysis.

The possibility of automatically identifying the most promising attributes was tested using both the ReliefF method and the wrapper approach. ReliefF (Robnik-Šikonja, & Kononenko, 2003) works independently of the learning algorithm and assumes neither the *a priori* nor conditional independence of attributes. Consequently, it also works efficiently when dependent attributes are involved. The wrapper approach (Kohavi & John, 1997) conducts a search in the space with one of the search algorithms and adds or removes one or several attributes in each iteration. Each iteration also includes a test on selected attributes of the learning algorithm and calculations of the learning performance. In this study, the hill-climbing search algorithm was used, while the cross-validation method was used as a measure for evaluating the learning performance.

RESULTS

Both the naive Bayes method with ReliefF and the logistic regression with the wrapper approach proved to be the most accurate in predicting competitive performance in U12/U12 and in 12–16/12–16 among all of the classification methods. In these two cases, other methods produced somewhat less accurate results. In A16/A16, all methods for predicting competitive performance failed. In predicting competitive performance in 12–16/U12, in A16/U12 and in A16/12–16 only some tests went beyond the limit of classification accuracy of 0.60, which is evident in Table 2. Relative frequencies of the majority classes (not the top ten players) for age categories were: for U12/U12 males and females 94%; for 12–16/12–16 males 97% and females 95%; for A16/A16 males 88% and females 73%; for 12–16/U12 males 89% and females 88%; for A16/U12 males 79% and females 71%; for A16/12–14 males 92% and females 87%.

Tests for predicting competitive performance by using regression methods in A16/A16 and in A16/U12 for male and female tennis players, and in A16/12–16 for female tennis players were not conducted due to an excessive number of cases with missing values. Missing values of some of the variables resulted from the development of the testing procedures over a longer period of time (i.e. years).

Table 2. Classification accuracies of all models for the studied samples of tennis players. (NB – naive Bayes, DT – decision tree, C4.5 – C4.5 algorithm, kNN – k-nearest neighbour, SVM – support vector machine, LR – logistic regression, WA – wrapper approach, R – ReliefF)

Method	U12/U12 Male	U12/U12 Female	12-16/12-16 Male	12-16/12-16 Female	A16/A16 Male	A16/A16 Female	12-16/U12 Male	12-16/U12 Female	A16/U12 Male	A16/U12 Female	A16/12-16 Male	A16/12-16 Female
NB	0.67	0.66	0.67	0.62	0.48	0.57	0.55	0.54	0.59	0.43	0.61	0.61
NB WA	0.62	0.70	0.68	0.61	0.56	0.50	0.53	0.52	0.55	0.45	0.58	0.50
NB R	0.67	0.73	0.66	0.67	0.52	0.63	0.59	0.63	0.57	0.54	0.58	0.58
DT	0.59	0.62	0.65	0.56	0.45	0.63	0.46	0.47	0.49	0.54	0.54	0.52
DT WA	0.63	0.65	0.51	0.62	0.49	0.52	0.58	0.59	0.60	0.56	0.56	0.50
DT R	0.66	0.63	0.53	0.66	0.45	0.57	0.63	0.62	0.62	0.48	0.58	0.46
C4.5	0.58	0.62	0.60	0.58	0.54	0.45	0.48	0.55	0.54	0.38	0.59	0.46
C4.5 WA	0.66	0.63	0.59	0.64	0.58	0.53	0.57	0.56	0.55	0.51	0.56	0.48
C4.5 R	0.65	0.61	0.60	0.66	0.57	0.57	0.58	0.52	0.46	0.53	0.48	0.48
kNN	0.64	0.58	0.55	0.61	0.54	0.50	0.56	0.57	0.43	0.44	0.53	0.49
kNN WA	0.53	0.63	0.62	0.55	0.45	0.54	0.61	0.44	0.45	0.50	0.55	0.52
kNN R	0.62	0.56	0.55	0.63	0.52	0.63	0.60	0.60	0.54	0.53	0.54	0.48
SVM	0.57	0.64	0.61	0.55	0.54	0.54	0.41	0.52	0.42	0.38	0.38	0.53
SVM WA	0.65	0.61	0.55	0.54	0.57	0.50	0.48	0.53	0.60	0.52	0.58	0.43
SVM R	0.60	0.59	0.60	0.58	0.54	0.48	0.51	0.52	0.49	0.43	0.48	0.50
LR	0.60	0.56	0.58	0.60	0.47	0.50	0.48	0.53	0.47	0.48	0.51	0.52
LR WA	0.69	0.67	0.67	0.62	0.43	0.61	0.51	0.61	0.49	0.55	0.58	0.57
LR R	0.67	0.65	0.66	0.63	0.48	0.60	0.57	0.66	0.52	0.57	0.54	0.57

Table 3. The relative absolute error of the regression models for the studied samples of tennis players

Method	U12/U12 Male	U12/U12 Female	12-16/12-16 Male	12-16/12-16 Female	12-16/U12 Male	12-16/U12 Female	A16/U12 Male
Regression tree	1.02	0.98	1.16	1.06	0.78	1.16	1.08
Linear regression	0.59	0.63	0.76	0.77	0.14	0.03	0.42

In five out of the seven tests, the relative absolute error of predictions of the regression tree is above 1.0 (Table 3), indicating that predicting the competitive performance of tennis players by using this method in practice serves no useful purpose.

Concerning the effect of considering only the most promising attributes, the performance of the logistic regression method was influenced to the largest extent. In U12/U12 and in 12–16/12–16, the wrapper approach on average improved in accuracy by 0.08, whereas the ReliefF method improved by 0.07 (Figure 1). All other methods improved in accuracy by ≤ 0.04. The difference in improving the accuracy between the wrapper approach and the ReliefF methods was 0.02 on average for individual classification methods in U12/U12 and in 12–16/12–16.
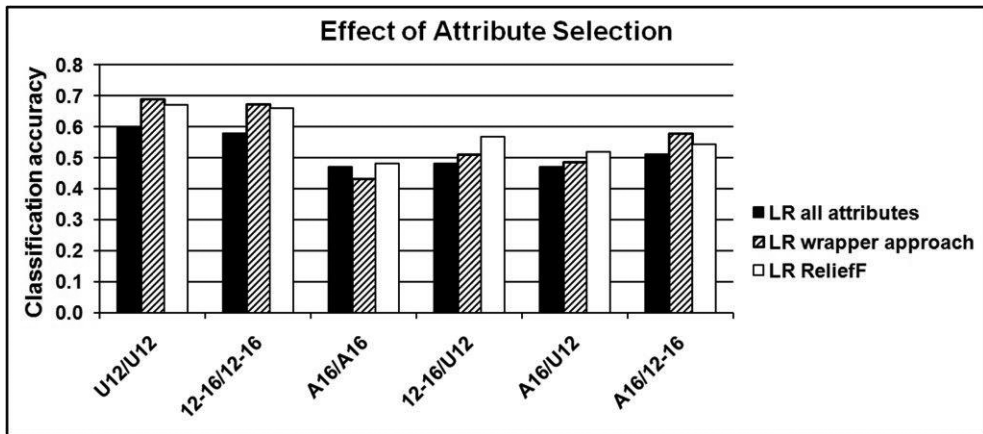
Figure 1. Effect of attribute selection on the logistic regression method

MOZL60 and MPAH were the most commonly selected attributes by means of the logistic regression method in combination with the wrapper approach for predicting competitive performance in U12/U12 and in 12–16/12–16; whereas the ReliefF method selected MT9X6 and MOZL60.

DISCUSSION

According to the classification analysis, a classification accuracy value above 0.60 was considered a satisfactorily accurate result. In view of this criterion, Table 2 can be divided into two parts: U12/U12 and 12–16/12–16, where the majority of classification accuracies are higher than 0.60; and A16/16, 12–16/U12, A16/U12 and A16/12–16, where only some of the classification accuracies were higher than 0.60. A conclusion that may be drawn from this is that the machine learning methods tested are suitable for predicting competitive performance in U12/U12 and 12–16/12–16. With a well-selected sample, a classification accuracy of around 0.50 can be achieved with a random classifier (Kononenko & Kukar, 2007) alone, making models with a classification accuracy below 0.60 unsuitable for the issues in question. The most accurate classification models are the naive Bayes method with ReliefF and the logistic regression with the wrapper approach, the average classification accuracy of which varies by 0.02, which is a negligibly small difference. They are therefore equally suitable for predicting performance regarding the issues in question. While the SVM method is usually considered one of the most reliable when it comes to complex actual issues (Caruana & Niculescu-Mizil, 2006; Kononenko & Kukar, 2007), in our case, contrary to expectations, the SVM models produced the least accurate results (in common with the k-nearest neighbour models). In any case, the classification accuracy did not exceed the accuracy of the default classifier (which classifies all cases in the majority class), although a comparison of classifiers is still possible.

Linear regression is a considerably more accurate method for predicting competitive performance than regression trees, which proved to serve absolutely no purpose in solving the issue in question. The relative absolute error of the linear regression models for 12–16/U12 for male and female tennis players (Table 2) is close to 0, which is considered a near-ideal regression function. However, in examining both samples it turned out that they contain a large number of cases

with missing values, influencing the linear regression learning, which then fails to use subjects with missing values and thus, by excluding these cases, the learning set becomes considerably reduced. In other linear regression models the relative absolute errors are between 0.42 and 0.76. As a result, only a very rough evaluation of tennis players' competitive performance can be made, which is not particularly useful in practice.

In comparison with predicting competitive performance using all of the morphological and motor attributes, little (0.02 on average) was gained by selecting the most promising attributes when using the ReliefF method and the wrapper approach for U12/U12 and 12–16/12–16. It can be concluded that this is a consequence of the high interdependence among the attributes, as was also confirmed by correlation analysis. It might be expected that the ReliefF (which has been shown to work well on dependent attributes (Kononenko & Kukar, 2007; Robnik-Šikonja & Kononenko, 2003)) will, on average, considerably improve the reliability of methods, although this was only in the case relating to the logistic regression method. Similarly, the wrapper approach only improved reliability to a considerable extent when it was used with the logistic regression method.

The results of predicting competitive performance by means of classification methods for individual categories were expected since the competitive performance in U16 is considerably influenced by morphological and motor factors. In A16, mental abilities, practical and technical competencies and competitive experience play an increasingly more important role. Since the ReliefF method and the wrapper approach both selected MOZL60, it can be concluded that the results obtained by each of these two methods are comparable, whereas MPAH and MT9X6 are based on a very similar functional mechanism and are also highly correlated. Both tests involve agility. Agility can be defined as the motor ability to carry out acceleration/deceleration types of locomotor movements effectively, including changes in direction. All of these are based on neuromuscular power, quickness of reaction/response and feet coordination. Studies (Filipčič, 1996; Filipčič & Filipčič, 2005; Šerjak, 2000; Unierzyski, 1994) have established that agility tests elucidate competitive performance at a statistically significant level.

In U16 it was established that by using classification methods top players can be separated from others on the basis of morphological and motor factors with an accuracy of approximately 0.66 (Table 1), whereas in the age group above 16 years attempts at classification on the basis of morphological and motor factors failed. Classification based on measurements conducted previously also turned out to be unreliable. The most accurate methods for predicting competitive performance were the naive Bayes method and logistic regression with the wrapper approach. Among the regression methods, linear regression proved to yield the most satisfactory results.

Similar findings to these were recorded by Filipčič (1996) who compared the uniformity of estimates made by means of regression analyses in the fields of motor, morphological and functional dimensions of tennis players aged between 12 and 14 with estimates of potential performance made using expert modelling. The correlation coefficient between both estimates is 0.72. According to the author, the relatively low correlation between the estimates stemming from both procedures can be attributed to the fact that the estimates by the expert system do not reflect current relations between criterion and prediction variables, but aim to predict relations that will arise in the future.

The relationship between the motor, morphological and functional dimensions and competitive successfulness of young male and female tennis players was studied in Filipčič, Filipčič and

Leskošek (2004), Filipčič and Filipčič (2005), Filipčič, Pisk and Filipčič (2010). The aim of these studies was to find out how the selected motor variables (explosive and elastic power, repetitive strength, speed and acceleration, speed of alternative movements, agility, static and dynamic balance, flexibility), functional variables (running endurance) and anthropometrical measures (longitudinal and transversal dimensions, skin folds and body mass) can explain the variance of the criterion variable (competitive successfulness).

Filipčič and Završki (2002) reported a medium-high (R = .52 and .43) and statistically significant correlation between two functional variables (VO2max, running test on 2,400 m) and competitive successfulness. In the study by Filipčič, Filipčič and Leskošek (2004), the results of a regression analysis showed that the system of tennis motor, functional and anthropometric variables explains 49% (R = .70) of the variance of the criterion variable in female tennis players, and 54% (R = .74) in male tennis players. For female tennis players three variables (elastic leg power test, balance and running endurance) and two variables for male players (agility test and body height) were found to be statistically significant for explaining the variance of the criterion variable. In a similar study by Filipčič and Filipčič (2005), the results revealed a statistically significant correlation between the group of selected tennis-specific motor variables and the criterion variable (R = .83) and the system of predictor variables explained 69% of the competitive successfulness of young female tennis players. The variables that measure the muscular power of the arms and shoulders, speed, flexibility, hand-eye coordination and dynamic balance significantly explain competitive successfulness.

In their study, Filipčič, Pisk and Filipčič (2010) examined the relationship between selected motor tests and competitive successfulness in tennis for different age categories of young tennis players. The competitive successfulness of players of both genders was defined by their position on the national ranking list. Several motor abilities were investigated: the neuromuscular power of the arms, the elastic power of the legs, the dynamic muscular strength endurance of the trunk, acceleration, agility, hand-eye coordination, dynamic balance and running endurance. The results of a regression analysis showed that in all categories there was a moderate, statistically significant correlation between the system of predictor variables and the criterion variable. A group of eight motor variables described 34% (R = .58) of the criterion variable variance in the category of 12- to 14-year-old girls and 54% (R= .73) in the same age category of boys, whereas for the 15- to 18-year-old players the predictor variables described 52% (R= .73) of the criterion variable variance in girls and 34% (R= .58) in boys. The running endurance test in girls and the hand-eye coordination test in boys partially described competitive successfulness in the category of 12- to 14-year-olds. In the category of 15- to 18-year-olds, the criterion variance was partially described by the dynamic muscular strength endurance of the trunk in girls and hand-eye coordination and acceleration in boys. The results of the study underline the importance of several motor abilities for competitive successfulness in particular age categories of young tennis players.

The above findings suggest it is possible to predict present or future performance in competitive tennis on the basis of results of measurements of motor, morphological and functional tests of young tennis players. In conclusion, the results showed that the correlation between motor, morphological and functional dimensions and competitive successfulness is higher in younger age categories (12- to 14-year-olds) than in older ones. This leads to the conclusion that the tactical, technical and mental dimensions of tennis players should be included in the test battery.

Important aspects of potential and competitive successfulness among young tennis players are: (1) investigating the temporal stage of observed players' potential dimensions; (2) comparing the level of potential dimensions with the performance in tournaments; and (3) using the results for talent identification and selection. All three aspects have a positive impact on the quality of planning training activities, and contribute to a more ethical training process of young tennis players (Filipčič, 1996).

In future, it would make sense to carry out measurements and data collection in a more organised manner since in certain cases it was observed that a large number of values were missing. Understandably, this depends on the availability of financial resources, although, along with further developments, the situation is expected to improve. Regarding predictions of competitive performance, a future study including a larger number of factors that influence the competitive performance of tennis players would most likely produce even better results. It would also be interesting to observe improvements with regard to predictions of players' competitive performance for several years in advance. Modern technologies will allow predictions in sport to become more accurate, bringing several benefits such as: efficient long-term planning, effective goal-setting and rationalisation of the training process.

CONCLUSIONS

In this study, predictions of the competitive performance of tennis players turned out to be a highly complex issue because the accuracy of the models for prediction, based on morphological and motor factors, was relatively poor. Reasons for this lie in the fact that competitive performance was only predicted on the basis of estimates of potential performance in the fields of morphological, motor and functional dimensions and, in so doing, the players' personality traits, mental and competitive abilities, technical and tactical competencies, and experience were not taken into account. Therefore, our future goal is to use measurement procedures to cover all fundamental dimensions of athletes' bio-psycho-social status to the greatest extent possible, as well as to take into account the dynamic correlation procedures which are present among them.

ACKNOWLEDGEMENTS

REFERENCES

Caruana, R. & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In W. Cohen & A. Moore (Eds.), *Proceedings of the 23rd international conference on machine learning* (pp. 161–168). New York, NY, USA: ACM.

Filipčič, A. (1996). Evalvacija tekmovalne in potencialne uspešnosti mladih teniških igralcev. [Evaluation of the competitive and potential performance of young tennis players]. Unpublished doctoral dissertation, University of Ljubljana.

Filipčič, A., & Završki, S. (2002). Relation between two aerobic capacity tests and competitive successfulness of junior tennis players. *Kinesiologia Slovenica 8*(1), 5–9.

Filipčič, A., Filipčič, T., & Leskošek, B. (2004). The influence of tennis motor abilities and basic anthropometric characteristics on the competition successfulness of young tennis players. *Kinesiologia Slovenica 10*(1), 16–26.

Filipčič, A. & Filipčič, T. (2005). The relationship of tennis-specific motor abilities and the competition efficiency of young female tennis players. *Kinesiology, 37*(2), 164–170.

Filipčič, A., Pisk, L., & Filipčič, T. (2010). Relationship between the results of selected motor tests and competitive successfulness in tennis for different age categories. *Kinesiology 42*(2), 175–183.

Hand, D. J., Mannila, H. & Smyth, P. (2001). *Principles of Data Mining.* Cambridge, MA (USA): The MIT Press.

Kohavi, R. & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence, 97*(1–2), 273–324.

Kononenko, I. & Kukar, M. (2007). *Machine Learning and Data Mining.* Chichester (UK): Horwood Publishing Ltd.

Orange. (2010). Orange Statistics for Predictors. Retrieved 16 January 2010 from http://www.ailab.si/orange/doc/modules/orngStat.htm.

Panjan, A., Šarabon, N. & Filipčič, A. (2010). Prediction of the successfulness of tennis players with machine learning methods. *Kinesiology, 42*(1), 98–106.

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning.* San Francisco (USA): Morgan Kaufmann Publishers Inc.

Robnik-Šikonja, M. & Kononenko, I. (2003). Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning, 53*(1), 23–69.

Šerjak, M. (2000). *Povezanost izbranih motoričnih sposobnosti in tekmovalne uspešnosti mladih teniških igralk.* [Connection of selected motor variables with the competitive successfulness of young female tennis players]. Unpublished bachelor's thesis, University of Ljubljana.

Unierzyski, P. (1994). Motor abilities and performance level among young tennis players. In W. Osinski & W. Starosta (Eds.), Sport Kinetics '93 (pp. 309–313). Poznan: Institute of Sport in Warsaw.

Witten, I. H. & Frank, E. (2005). *Data Mining.* San Francisco (USA): Morgan Kaufmann.

Zhang, H. (2004). The optimality of naive Bayes. In V. Barr & Z. Markov (Eds.) *FLAIRS Conference,* AAAI Press.