A neural network model approach to athlete selection

Adam Maszczyk · Adam Zając · Igor Rygula

Published online: 1 October 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract In order to determine the usefulness of neural models in optimisation of recruitment processes, statistical analyses were carried out on measured results of javelin throwers using a full take off. A group of 140 Polish junior javelin throwers took part in the research. In order to choose the optimum combination of model parameters the Hellwig method was used. Linear and multilayer perceptron neural models were constructed and used to calculate combinations of variables. Statistical analysis of the results showed that the linear model was not able to describe precisely the relationship between the dependent variable and independent variables for the investigated group of young javelin throwers. For the investigated group, the perceptron network with a 4-3-2-1 structure gave the best predictive relationship for sports results of the javelin throwers.

Keywords The method of Hellwig · Non-linear neural models · Linear neural models · Multilayer perceptron (MLP) · Artificial neural networks · Sports-selection

1 Introduction

Instruments which would allow the determination of personal predispositions for achieving high sport performance have been sought for many years. One approach is the application of successive phases of training with different selection criteria, to determine the contestant’s chances of achieving high performance [18]. Statistical and mathematical forecasting methods [19] are becoming more and more significant in this area. These methods include multidimensional exploration techniques, which have only been sporadically used in the area of sport science. These methods are mainly applied to recognition of objects and their grouping [27]; concentration analysis; factor analysis; and discriminatory analysis [30]. In the opinion of many researchers, a system of organisational-methodical methods, having complex character uniting pedagogical, sociological, psychological and medical–biological methods of research should be used for the purpose of construction of a model in the definite discipline or in the group of disciplines of sport [14, 17, 19]. Based on this, Naclerio [18] showed the necessity of filling the gap existing between pedagogical theories, the practice of the sports-training and the competitive model approach in order to create this link. The foundation of this linked approach needs to consider development of competitors in terms of achieved results in comparison with the model criteria. This has become possible as a result of the mathematical formalisation of this problem by authors such
as Haykin [10], Tidow [24, 25], Naclerio [18] and Hatton [9].

The use of multidimensional exploration techniques for optimising the recruitment process should allow groups of objects or qualities to be grouped together. Regarding ‘qualities’, subsets representing homogeneous taxonomic units can contain similar information on contestants. This allows the possibility of identifying certain types of contestants, who have greater potential. This may lead to the optimisation of recruitment and the definition of different training loads [29].

The occurrence of linear and non-linear relationships between variables has led to the development of artificial neural networks (ANNs) for modelling and prediction [2, 7, 10, 14, 15, 20, 23, 28].

Previous research [1, 3, 4, 9, 12, 13, 16, 17] has covered many aspects of sports result for javelin throwers. Some of this work has shown the greater usefulness of non-linear neural models compared with regression models for prediction [1, 17] with these studies showing that, in most cases, the simplest networks give the best generalisation. Analyses, model construction and experimentally determined results indicate that the quality of prediction of neural models is comparable to that of regression analysis and regression models.

At this moment it is important to draw attention to the fact that in further research over the construction of neural models optimising the sports, selection and the prediction of the results of competitors in the javelin throw, more independent variables should be taken into consideration than were covered in this earlier research. Namely the characteristics of the biomechanical analysis of the javelin throw such as speed and angle of the throw out, angle of the attack and characteristics of the postural muscle [24, 25]. Some characteristics directly related to the techniques of the throw such as the degree of muscle stretching before contraction and various start positions were also omitted [4].

The objectives of this research are to cover some of the aspects noted earlier, namely.

1. Will linear neural models be able to describe precisely the relationship between given input and output data in the investigated group of young javelin throwers?
2. Will research results show the possibility of using MLP networks in predicting sports results?
3. What structure of neural model will be able to describe precisely the prediction of the sports result?

The aim of this study was to determine whether there is a possibility that the ANNs may be applied to assist in the process of recruiting athletes and prediction for javelin throwing.

2 Materials, methods and study tools

2.1 Participants

In order to verify the formulated hypothesis, statistical analysis of measured results for junior javelin throwers was carried out. The measured results from a group of 55 (preliminary modelling) and 85 (new learning cases) junior category javelin throwers formed the initial dataset. The measurements were carried out on the ground of sports facilities of Sports-Schools in Dąbrówka Górnicza, facilities of Sports-School in Mysłowice, Schoolboy Athletic Club “Tornado” of Elementary School 31 in Katowice, sports facilities of Interschool Sports Centre in Katowice, facilities of Athletic Club “Pogoń” Ruda Śląska, sports facilities of Athletic Club “Plomień” Sosnowiec and in the laboratory of Chair of System Analyses in Sport in Academy of Physical Education in Katowice.

Participants were selected at random from candidates with less than 2 years’ experience in javelin throwing. The basic criteria of the selection were the period of training (approaching 2 years of the throwing training from 2002) and their informed consent and active engagement in the training process. The intensity of training in terms of number of sessions per week that each competitor undertook was an experimental variable due to different methods of training used in the individual centres.

2.2 Data collection and tools of the statistical analyses

The characteristics of the first group of 55 16- to 17-year-old javelin throwers were used to build the neural models and were measured twice (in May 2004 and in May 2005). The developed models were supplemented by 85 new teaching cases (competitors at the same age and with the same period of training) whose independent variables were obtained as a result of measurements in April 2006. This led to an improvement in these models, expressed by greater generalisation and better predictions. The verification took place in May 2007 by comparing the predicted model results (using the module Run One-Off Case) with the actual javelin throw results obtained as an average of the three test throws after a 30-min warm-up. Twenty athletes were selected (at the same age and the same period of training) to participate in javelin throwing in May 2006 based on model predictions of their potential using measures determined from the May 2004 and May 2005 studies. The results of this group of 20 throwers were not used in building the model, but their May 2007 javelin throw distances were used to assess the quality of the ANN model for recruitment.

The structure of the variables used in the model was RXn|Yn, a multi-valued dependent variable (Yn) and n
multivalent independent variables ($X^n$) taking into consideration the principle of randomisation ($R$). The distance of a javelin throw (average result of three throws after a 30-min warm-up) was the dependent variable ($Y$) in all tests. The initial a-priori approach meant that a very wide variety of independent variables were sampled initially for later reduction through modelling.

Independent variables for multivariate analysis were obtained by measuring the different characteristics of the athletes in the following groups:

### 2.3 Anthropometrical measurement

- The Rohrer's indicator (WSR) calculated by the formula
  $$\text{WSR} = \frac{(\text{WS}/\text{MS})^3 \times 100}{\text{where WS is the height}}$$
  and MS is the body mass). This measurement was determined using a BIA-101/S.C on impedance analyser by body tissue. The device sampled (for a period of 5-s) the resistance ($R_z$) and the reactance ($X_c$) values of tissues and their volume. From these measurements, the program enumerated six indicators of body composition:
  - Quantity of fat tissue (SC-FM in \%)
  - Quantity of fat tissue (SC-FM in kg)
  - Quantity of muscle tissue (SC-MM in \%)
  - Quantity of muscle tissue (SC-MM in kg)
  - Body mass index (SC-BMI in kg/m$^2$)
  - Body cell mass index (SC-BCMI in kg/m$^2$)

### 2.4 The flexibility measurement

The coefficient of flexibility (WG). From a standing position, the athlete pushes a dowel (100–110 cm) overhead with the shoulders after which the dowel is positioned across the athlete’s shoulders below the neck by twisting the shoulders joints. WG is the distance measured from the horizontal dowel position underneath the cervical pit, to the dowel length measured from the internal part of the palm. This is a modification of Cook’s test [6] and the values are quoted in cm (to two decimal places).

### 2.5 The coordinative abilities measurement

- The ability of the connection of movements (KLG): performing three rolls forward (tuck rolling), with an accuracy of 0.1 s [5].
- The ability of high-frequency movements (KBST): side step–test, obtained as a result for maximum number jumps performed within 20 s [5].
- The conformability of movements (KSM): involving jumping forward and back, from a standing position, three times; KSM is the difference between starting and finishing position [5].

### 2.6 The special efficiency measurement

- The cross-step up to the throw position (PZMPW): calculated from time, $t$, measurement $r$ and an assessment of the correctness of movement on a scale of 1–5.
- Distance of a javelin throw from the standing position (SPPM) with 10 cm accuracy.
- Distance of a javelin throw with 10 cm accuracy after the cross-step (SPPR).
- The degree of leg work during the cross step (PNPK) assessed as the validity of movement on a scale of 1–5.
- Speed and accuracy of movement in the throwing zone (VSR) measured on a time scale, $t$, with assessment of movement validity on a scale of 1–5.
- Estimation of overall special efficiency (OSS) determined on the basis of the above results of measurements of special efficiency on the basis of taxonometry analysis tools.

In order to calculate the power efficiency of selected muscle groups the following measurements were carried out:

- The efficiency force of the shoulder girdle (FOB) represented by the number of push-ups performed in 10 s.
- The efficiency force of the shoulder girdle and trunk (FOBT): the distance a medicine ball (2 kg) is thrown forward (overhead medicine ball throw with an accuracy of 10 cm).
- The efficiency force of shoulders and the trunk (FRT): the distance a medicine ball (2 kg) is thrown forward from the straddle seat (with 5 cm accuracy).
- The efficiency force of the abdominal muscles (FMB): the number ($n$) of full lifts of the torso from a lying position and return to the ground completed within 10 s.
- Grip power (FCH): the dynamic register tightening at maximum force kept in the left- and then in the right hand.

### 2.7 The stress capacity measurement

The measurement was based on Wingate test, using a cycle ergometer type “Monark” 814 E and PC using MCE-1 v.2.3 software.

The following variables were specified:

- Maximum power ($P_{\text{max}}$): measured over 15 s, in W/kg.
- The threshold of anaerobic transformations (PPB) in HR.
- The Klonowicz coefficient of restitution (WSKR) calculated by the formula $\text{WSKR} = C_2 - C_3/C_2 - C_1 \times 100$ (in percent), where $C_1$, the frequency of pulse before effort; $C_2$, the frequency of pulse in first minute after effort; $C_3$, the frequency of pulse in fifth minute after effort.
2.8 The measurement of motor ability

- The time taken to run a distance of 5 m (SZB5).
- The time taken to run a distance of 10 m (SZB10).

The sprint measurement for distances of 5 and 10 m were made on an asphalt athletic track from a standing start. The research made it possible to measure, record and archive data for both distances during one race.

- The standing long jump (S-SDM)
- The standing triple jump (S-TR)
- The standing quintuple jump (S-PS)

The jump measurements were carried out on an asphalt athletic track with an accuracy of 1 cm.

2.9 The measurement of movement abilities

- Movement abilities (UR)

The measurement of this feature was carried out on the basis of the Ryguła test of movement abilities with modifications [21]. The test consisted of five exercises involving a standing jump with rotation around the longitudinal axis of the body at a random angle. The measurement was carried out on a special mat with a square (59 cm side) drawn on it. This was divided into four rectangles and placed inside a circle with a radius of 41.7 cm. The circumference of the circle was divided into eight sections, each with the measurement of \(\frac{45}{8}\) (Fig. 1). The place of jump was drawn at a distance of 25 cm from the square.

The test included the following: pirouette 180°, pirouette 270°, pirouette 360°, the maximum turnover to the left and the maximum turnover to the right. The result was measured in points. For the proper execution of the first three trials \(UR_{\text{max}}(3) = 30\) points; for another two \(UR_{\text{max}}(360L_{\text{R}}) = 8\) could be obtained or more in the fourth trial and \(UR_{\text{max}}(360R) = 8\) or more in the last trial. Finally, \(UR_{\text{max}}\) of the five tests could be (a priori): \(UR_{\text{max}}(5) = 460 \pm 1–2\) points. The choice of the test was dictated by a crucial factor, namely the test of Rygula, which correlates significantly with other ability tests, but it does not show a significant correlation with the throwers’ motion abilities and with their body construction.

2.10 The measurement of personality and temperament

- Raven’s progressive matrices test (OPTR): OPTR are multiple-choice tests of abstract reasoning. In each test item, a candidate was asked to identify the missing segment required to complete a larger pattern. Many items are presented in the form of a \(3 \times 3\) or \(2 \times 2\) matrix, giving the test its name. Standard progressive matrices: the booklet comprises five sets (A to E) of 12 items each (e.g. A1 through to A12), with items within a set becoming increasingly difficult, requiring ever greater cognitive capacity to encode and analyse information. All items are presented in black ink on a white background.

- Technical intelligence-technical/mechanical skills (OITK): used Squares Test [8]. The maximum that can be obtained is \(OITK_{\text{max}} = 45\) points.

The measurement of personality differences in terms of temperament was defined by means of the Strelau questionnaire, in Angleitner authorised version [22]. The obtained variables are

- The power of processes of stimulation (SPP)
- The power of processes of braking (SPH)
- The motility of nerve processes (RPN)
- The equilibrium of nerve processes (RWPN)

2.11 The measurement of the overall physical fitness

- The International Test of the Physical Fitness (MTSF): this variable was obtained on the basis of score tables of international physical fitness test. The test battery was as follows:
  1. 50-m test-relative power, speed.
  2. Flexed-arm hang-upper body relative strength and endurance.
  3. 10-m shuttle run-relative power, speed and suppleness.
  4. Back throw-absolute power, speed and suppleness.
  5. 1,000-m run-aerobic/anaerobic capacity, stamina.

In summary, 42 independent measurements were made to compare with one dependent variable \(Y\)—distance of javelin throw (Table 1).
Measurements of the data were conducted first for the 55 cases, in May 2004 and in May 2005. Then for 85 new learning cases in April 2006 for the testing group (the May 2007 javelin throw results for this group were used to assess the actual quality of the model).

In order to carry out statistical analysis of the studied javelin throwers group, basic statistical measures, such as: arithmetic average \( \bar{x} \), standard deviation \( S \), variation coefficient \( V \), asymmetry \( A_s \), kurtosis \( K_{U-3} \) and Pearson coefficient of correlation \( r \), were calculated.

| L.p | Variables                                      | \( x \) | \( S \) | \( V \) | \( A_s \) | \( K_{U-3} \) |
|-----|------------------------------------------------|--------|--------|--------|----------|-------------|
| 1   | WS The height cm                                | 173.31 | 7.48   | 4.32   | -1.54    | 2.29        |
| 2   | MS The body mass kg                             | 66.12  | 8.89   | 13.45  | -1.26    | 1.86        |
| 3   | SZB5 The time to run the distance of 5 m s      | 1.09   | 0.09   | 8.58   | 0.25     | -1.20       |
| 4   | SZB10 The time to run the distance of 10 m s    | 1.53   | 0.21   | 13.59  | 0.14     | -1.13       |
| 5   | UR Movement abilities Points                    | 33.18  | 3.18   | 9.58   | -0.71    | -0.24       |
| 6   | OITK Technical intelligence Points              | 30.84  | 2.79   | 9.04   | -0.39    | -0.67       |
| 7   | \( P_{Max} \) Maximum power W/kg               | 9.69   | 0.30   | 3.11   | -0.18    | -0.89       |
| 8   | PPB The threshold of anaerobic transformations HR| 169.22 | 3.57   | 2.11   | -0.10    | -0.73       |
| 9   | WSR The Rohrer’s indicator g/cm^3              | 68.23  | 6.96   | 10.20  | -0.92    | -0.01       |
| 10  | PZMPW The cross step with acceptance of the throw position s | 1.15   | 0.09   | 7.87   | 0.81     | -0.02       |
| 11  | RWPN The equilibrium of nerve processes Points  | 0.71   | 0.11   | 16.01  | -1.01    | 1.09        |
| 12  | S-PS The standing quintuple jump m              | 8.26   | 0.78   | 9.40   | 0.12     | -1.14       |
| 13  | OSS The estimation of the special efficiency Points | 3.84   | 1.01   | 26.43  | -0.54    | -0.26       |
| 14  | SC-MM Quantity of muscle tissue %               | 62.24  | 8.55   | 13.74  | -0.39    | -0.10       |
| 15  | SC-MM Quantity of muscle tissue kg              | 42.93  | 7.38   | 17.20  | -0.24    | -0.45       |
| 16  | SC-BMI Body mass index kg/m^2                   | 21.00  | 1.40   | 6.69   | 0.45     | 1.14        |
| 17  | SC-BCMI Body cell mass index kg/m^2             | 10.68  | 1.92   | 18.01  | 0.02     | 0.16        |
| 18  | FCH Power of the grip kg/kg                     | 41.38  | 5.79   | 13.99  | -0.54    | -0.58       |
| 19  | WG The coefficient of flexibility cm            | 0.44   | 0.06   | 12.78  | 1.73     | 5.80        |
| 20  | PZMPW The cross step with acceptance of the throw position s | 2.64   | 1.02   | 38.88  | -0.17    | -1.07       |
| 21  | PNPK The legs work in the phase of the cross step Points | 3.09   | 0.80   | 25.88  | -0.17    | -1.41       |
| 22  | SPPM Distance of javelin throw from standing position m | 58.83  | 5.31   | 9.02   | 0.04     | -1.31       |
| 23  | SPPR Distance of javelin throw after cross step m | 64.03  | 6.07   | 9.48   | 0.00     | -1.27       |
| 24  | VSR The speed of moving in the throwing zone s  | 4.33   | 0.12   | 2.87   | 0.81     | -0.47       |
| 25  | VSR The speed of moving in the throwing zone Points | 3.18   | 0.72   | 22.70  | -0.29    | -1.01       |
| 26  | S-SDM The standing long jump m                  | 2.71   | 0.31   | 11.57  | -0.18    | 0.28        |
| 27  | S-TR The standing triple jump m                 | 4.74   | 0.28   | 5.92   | 0.28     | -0.56       |
| 28  | SC-FM Quantity of fat tissue %                  | 10.74  | 3.81   | 35.45  | 0.08     | -0.86       |
| 29  | SC-FM Quantity of fat tissue kg                 | 7.49   | 2.93   | 39.17  | 0.27     | -0.72       |
| 30  | MTSF The International test of the physical fitness Points | 478.11 | 60.97  | 12.75  | 0.45     | -0.54       |
| 31  | FRT The efficiency forced of the shoulders and trunk cm | 7.07   | 0.52   | 7.34   | 0.33     | -0.96       |
| 32  | FOBT The efficiency force of the shoulder girdle and trunk cm | 18.70  | 2.57   | 13.72  | -0.20    | -1.34       |
| 33  | FOB The efficiency force of the shoulder girdle n | 15.20  | 4.45   | 29.30  | 0.12     | -1.26       |
| 34  | FMB The efficiency force of the abdominal muscles n | 18.56  | 4.60   | 24.79  | 0.04     | -1.10       |
| 35  | WSKR The Klonowicz coefficient of restitution % | 1.27   | 0.11   | 8.47   | 0.18     | -0.64       |
| 36  | KLG The ability of the connection of movements s | 12.40  | 1.51   | 12.18  | 0.01     | -1.31       |
| 37  | KBST The ability of the high frequency of movements n | 27.24  | 3.63   | 13.31  | 0.03     | -0.86       |
| 38  | KSM The conformity of movements cm              | 33.57  | 9.20   | 27.39  | -0.16    | -0.16       |
| 39  | OPTR Raven’s progressive matrices test Points   | 38.05  | 2.90   | 7.62   | -0.46    | -0.64       |
| 40  | SPP The power of processes of stimulation Points | 59.09  | 2.81   | 4.76   | 0.00     | -1.09       |
| 41  | SPH The power of processes of braking Points    | 63.80  | 2.74   | 4.30   | 0.07     | -1.03       |
| 42  | RPN The motility of nerve processes Points      | 58.06  | 2.81   | 4.84   | 0.00     | -0.95       |
| 43  | Y (2005) Distance of javelin throw m            | 51.69  | 3.73   | 7.21   | -0.70    | 0.30        |
Table 2  Values of integral capacities of carriers of the information $H$, for each variable combinations in the investigated group of junior javelin throwers

| $X_1$ the height (cm), $X_2$ the body mass (kg), $X_3$ the time to run the distance of 5 m (s), $X_{10}$ the cross step with acceptance of the throw position (s), $X_{13}$ the estimation of the special efficiency (points), $X_{15}$ the standing quintuple jump (m), $X_{18}$ power of the grip (kg) | The choice of the optimum data of explanatory variables |
|---|---|
| The best combination of VII independent variables | $X_1, X_2, X_3, X_{10}, X_{12}, X_{13}, X_{18}$ |
| No. explanatory variables | Integral capacity of information $H$ |
| The best combination of VIII independent variables | $X_1, X_3, X_{10}, X_{11}, X_{12}, X_{13}, X_{17}, X_{18}$ |
| No. explanatory variables | Integral capacity of information $H$ |
| The best combination of IX independent variables | $X_1, X_2, X_3, X_{10}, X_{11}, X_{12}, X_{13}, X_{17}, X_{18}$ |
| No. explanatory variables | Integral capacity of information $H$ |
| The best combination of X independent variables | $X_1, X_2, X_3, X_7, X_{10}, X_{11}, X_{12}, X_{13}, X_{17}, X_{18}$ |
| No. explanatory variables | Integral capacity of information $H$ |

During the course of further mathematical and statistical analyses, multiple regression and discriminatory analysis were used. In order to choose the optimum combination of model parameters Hellwig’s algorithm was used.

In Hellwig’s method the number of combinations is provided by the formula: $L = 2^m - 1$ where $m$ is the number of independent variables. The individual contribution of each independent variable in the combination is given by the formula: $h_{ij} = (r_{0j})^2 / (1 + \sum_{i=1}^{n} |r_{ij}|)$, where the individual capacity of information for the j-th variable in the l-th combination ($h_{ij}$) is equal to the ratio between the coefficient for correlation between j-th variable (independent) and the dependent variable ($r_{0j}$) and the coefficient correlation between i-th and j-th variable [11]. Summing the individual capacities gives the overall capacity for a given combination, $H_l$. This parameter represents the correlation between a group of individual parameters and the final results so that values closer to unity represent more appropriate combinations for prediction of the final result. Iterating for groups of parameters (from the 42 listed in Table 1) for the May 2004 measurements verified using a regression analysis made on the grounds of a stepwise progressive regression algorithm identified a smaller sub-set of seven parameters that gave the highest $H$ value. The four highest combinations are summarised in Table 2.

This analysis allowed the number of independent variables to be reduced from 42 to seven, namely: height (cm), body mass (kg), the time to run the distance 5 m (s), the cross step from the place into the front with acceptance of the throw position (points), the estimation of the special efficiency (points), the standing quintuple jump (m) and power of the grip (kg/kg) (Table 2).

Results of calculations using Hellwig’s method were verified by multiple regression analysis against the dependent variable $Y$ (the distance of javelin throw) and gave the weighting factors for each of the seven chosen parameters, which are summarised in Table 3. Statistical analysis of the weighting factors derived from this regression equation for dependent variable $Y$ indicated non-coincidence in the case of variables—the power of the grip, the estimation of the special efficiency (within the meaning of the equation: sign $(r(x_i,y)) = sign (a_j)$) and non-statistical significance of variable—the body mass.

The use of Hellwig’s method and regression analysis has identified [based on calculated regression equation coefficients (st. error B)] four parameters having the most significant effect of javelin throw distance. In order of importance these are (Table 4):
- time to run 5 m;
- the cross-step to the throwing position;
- standing quintuple jump; and height.

Table 3 Preliminary structural parameters of regression equation for dependent variable $Y$—distance of javelin throw, in the investigated group of junior javelin throwers (measurement period—2004)

| Parameter | $B$ | St. error $B$ |
|---|---|---|
| Intercept | 33.923 | 13.83 |
| $X_1$-The height | 0.107 | 0.069 |
| $X_3$-The time to run the distance of 5 metres | $-6.547$ | 4.850 |
| $X_{10}$-The cross step with acceptance of the throw position | $-7.224$ | 3.813 |
| $X_{12}$-The standing quintuple jump | 0.461 | 0.574 |
| $X_5$-The body mass | 0.041 | 0.054 |
| $X_{13}$-The estimation of the special efficiency | $-1.944$ | 0.700 |
| $X_{18}$-The power of the grip | $-0.004$ | 0.098 |

Summing up of dependent variable regression $Y$—distance of javelin throw: $r = 0.859$, $r^2 = 0.727$

Corrected $r^2 = 0.687$, $F = 16,833$, $p < 0.00000$; standard estimation error $= 2.043$
These four parameters can be used as a basis to predict the dependent variable \( Y \) — distance of javelin throw, e.g. if the time to run the distance of 5 m is decreased by 1-s then the value of our explained variable distance of javelin throw will increase by 8.53 m, assuming that the remaining variables remain unchanged (Table 4).

Based on variables obtained using Hellwig’s method and verified by regression analysis, linear and MLP ANN models were constructed based on the four essential statistical (the height, the time to run the distance of 5 m, the cross step with acceptance of the throw position and the standing quintuple jump). The ANNs were trained by the error back propagation learning algorithm, Quasi Newton’s and Levenberg–Marquardt’s algorithms. All algorithms have been tested, but in this paper the network results were obtained by using the Levenberg–Marquardt’s algorithm. This algorithm was also used in the final learning of the perceptron network. Initially, a model was constructed which describes the correlations between the four independent variables and the dependent variable \( Y \)—distance of javelin throw, in order to determine the usefulness of the linear neural model. If the investigated dependence has a linear character, it is possible to use a linear network for its description, but if the linear dependence does not appear, the constructed linear network will not be able to describe precisely the relationship and a non-linear model is needed.

The results for the linear ANN are shown in Table 5. Analysing Normalised root mean square error (NRMSE) values (a frequently-used measure of the differences between values predicted by a model and a good measure of accuracy) of 0.59, 0.49 and 0.37 for the learning, validation and test series, respectively, indicated that this linear network cannot precisely describe the observed trend suggesting that the relationship has some non-linear character. Hence, ANNs using MLP were constructed in order to account for the non-linearity.

The construction of the neural models was started with the division of the dataset into the three series: learning (44 cases), validation (6 cases) and test (5 cases), in May 2005. The obtained models were complemented with 85 new cases in April 2006 (now, total 140 cases), which led to improvement of the quality of these models, expressed by better generalisation and prediction. Finally, in May 2007, measurements of javelin throw results were carried out, which were used to verify the quality of model predictions obtained during recruitment.

All statistical analyses were carried out on a PC using the statistical package STATISTICA 6.0 (2001 Polish edition) and the module of STATISTICA Neural Networks (Release 4.0 E).

### Table 4

| Parameter | Beta  | St. Error beta | B     | St. Error B | t     | p level |
|-----------|-------|----------------|-------|-------------|-------|---------|
| Intercept | 0.320 | 0.084          | 30.782| 11.457      | 2.687 | 0.010   |
| The height| −0.222| 0.121          | 0.154 | 0.040       | 3.815 | 0.000   |
| The time to run the distance 5 metres | −0.191 | 0.093 | −8.528 | 4.665 | −1.828 | 0.074 |
| The cross step with acceptance of the throw position | 0.355  | 0.118          | 1.652 | 0.549       | 3.009 | 0.004   |
| The standing quintuple jump | 0.320  | 0.084          | 30.782| 11.457      | 2.687 | 0.010   |

Summing up of dependent variable regression \( Y \) distance of javelin throw: \( r = 0.901, r^2 = 0.827 \)
Corrected \( r^2 = 0.822, F = 17.934, p < 0.00000 \), standard estimation error = 2.023

### Table 5

| Linear network for \( Y \)—distance of javelin throw |
|-----------------------------------------------------|
| Data standard deviation ratio | Learning series | Validation series | Test series |
| Normalised root mean squared error | 0.51 | 0.49 | 0.37 |
| \( r \) | 0.85 | 0.92 | 0.93 |

\( r \) correlation coefficient in series cases in April 2006 (now, total 140 cases), which led to improvement of the quality of these models, expressed by better generalisation and prediction. Finally, in May 2007, measurements of javelin throw results were carried out, which were used to verify the quality of model predictions obtained during recruitment.

All statistical analyses were carried out on a PC using the statistical package STATISTICA 6.0 (2001 Polish edition) and the module of STATISTICA Neural Networks (Release 4.0 E).

### 3 Results and model development

ANNs were constructed with three MLP networks, for four input neurons (representing the four independent variable used in linear network, identified above) and one output neuron (the dependent variable \( Y \), using the architectures: 4-2-1, 4-3-1 and 4-3-2-1, which were trained by the Levenberg–Marquardt algorithm (Table 6).

The construction of these neural models commenced with the division of the dataset into three series as for the linear case above.

The ANN model having a 4-2-1 network has values of standard deviation ratio for the validation series which are not satisfactory, Table 6. NRMSE values for the learning series (0.37), validation series (0.40) and test data (0.46) were too high and not satisfactory to claim that this model...
correlated well with the learning data. However, the values for the learning and test series (0.37 and 0.38) and the correlation coefficient in those groups (0.95 and 0.97) suggested that models with a greater number of neurons in the hidden layer could fit the learning data in the first set more accurately.

The neural model with a 4-3-1 network showed better results than that with a 4-2-1 network (Table 6). This is seen in the reduced values for the standard deviation ratio for learning, validation and testing data (0.32, 0.35 and 0.13, respectively). Correlation coefficients for these data also had better values generally than for the previous model.

The value of NRMSE for learning data using the neural model with a 4-3-2-1 network was 0.26, which indicates a satisfactory fit to the learning data: better than previous models.

Analysis of model and experimental results showed that, for the model with a 4-3-2-1 network, the NRSME value for learning data is 0.19, indicating a good fit to the learning data; better than previous models.

This was supported by a value of 0.97 for the correlation coefficient. Analysis of the results for the validation (0.21) and test (0.15) data further supported the ability of the model in terms of generalisation and prediction of results. The practical value of the model was confirmed by the high correlation coefficients (0.97, 0.96 and 0.99) recorded.

In accordance with statistical theory, the selected model was complemented with new learning cases, which should improve the quality of this model in terms of better generalisation and prediction. Hence, the trained neural model with 4-3-2-1 network was complemented with 85 new cases (new group of competitors in the same 16- to 17-year-old age group with the same period of training) which were added to the previous 55 cases and divided into learning (80 cases), validation (35 cases) and testing (25 cases) datasets. The neural network was trained using Levenberg–Marquardt’s algorithm. The comparisons of model and experimental data were analysed and the results summarised in Table 7.

### Table 6: Regression statistics of assessment of non-linear neural models for dependent variable \(Y\)—distance of javelin throw, in a group of junior throwers (\(N = 55\) cases, four input neurons, May 2005)

|            | MLP 4-2-1 | MLP 4-3-1 |
|------------|-----------|-----------|
| Learning series |          |           |
| Normalised root mean squared error | 0.37      | 0.38      |
| Validation series | 0.40      | 0.32      |
| Test series |           |           |
| Normalised root mean squared error | 0.93      | 0.95      |
| Validation series | 0.97      | 0.95      |
| Test series |           |           |
| Normalised root mean squared error | 0.13      | 0.94      |

### Table 7: Regression statistics of assessment of 4-3-2-1 structure MLP model for dependent variable \(Y\) in a group of junior throwers (four input neurons data collection, April 2006)

|            | MLP 4-3-2-1 |
|------------|------------|
| Learning series |          |
| Normalised root mean squared error | 0.19      |
| Validation series | 0.21      |
| Test series |            |
| Normalised root mean squared error | 0.15      |
| Validation series | 0.97      |
| Test series |            |
| Normalised root mean squared error | 0.99      |

\(r\) correlation coefficient in series
4 Discussion

Using Hellwig’s method (verified by multiple regression analysis against the dependent variable Y), seven variables for the future regression model were chosen: height, body mass, the time to run the distance 5 m, the cross-step approaching the throw position, the estimation of the special efficiency, the standing quintuple jump and power of the grip (Table 3). Application of sports theory would indicate that increasing values for the estimation of special efficiency and grip power should increase javelin throw distance, whereas the weighting factors from statistical analysis (Table 3), are negative. These two parameters should therefore be omitted from the models due to non-coincidence when applying fuzzy logic. This and the non-significance of body weight reduced the essential inputs to four that agree broadly with javelin thrower and coach experience [24, 25]; [9, 19, 26], namely height, the time to run the distance of 5 m, the cross step with acceptance of the throw position and the standing quintuple jump.

The NRMSE values (0.59, 0.49 and 0.37 for learning, validation and test series respectively), showed that a linear ANN model is not capable of describing precisely the trend’s dependences between the four chosen independent variables and dependent variable Y. ANNs using MLPs gave better results with a 4-3-2-1 structure giving the best fit to the learning, validation and test data (0.19, 0.21 and 0.15 respectively) for the three structures considered.

The usefulness of the model in selection of athletes was assessed by selecting 20 new javelin throwers on the basis of the 4-3-2-1 structure MLP ANN. Comparing the calculated model values with module Run One-Off Case used with real results of throws confirmed the excellent work of this model. It also showed that after 1 year of training for the 20 test throwers the total network error was only 2.50 m and in the case of the eight throwers calculated predictions differed from actual results in the range of ±0.01 to ±0.19 cm (Table 8). In the case of five throwers (nr 1, nr 4, nr 5, nr 11 and nr 13), the calculated prediction differed from actual results by only ±0.01 ± 0.05 cm, and for another three (nr 12, nr 14 and nr 16), respectively, by

| L.P. | Real distance of javelin throw (m) | MLP 4-3-2-1 | Network error (m) |
|------|-----------------------------------|-------------|------------------|
|      | Values of models prediction (m)   |             |                  |
| 1    | 61.39                             | 61.38       | 0.01             |
| 2    | 57.60                             | 57.06       | 0.54             |
| 3    | 52.75                             | 53.93       | -1.18            |
| 4    | 56.24                             | 56.20       | 0.04             |
| 5    | 51.70                             | 52.73       | -1.03            |
| 6    | 56.40                             | 55.85       | 0.55             |
| 7    | 52.80                             | 53.56       | -0.76            |
| 8    | 56.61                             | 56.37       | 0.24             |
| 9    | 56.49                             | 56.42       | 0.07             |
| 10   | 55.75                             | 55.23       | 0.52             |
| 11   | 55.48                             | 55.46       | 0.02             |
| 12   | 55.45                             | 55.26       | 0.19             |
| 13   | 60.60                             | 60.55       | 0.05             |
| 14   | 58.51                             | 58.54       | 0.06             |
| 15   | 53.56                             | 53.12       | 0.44             |
| 16   | 51.45                             | 51.64       | -0.19            |
| 17   | 56.95                             | 56.44       | 0.51             |
| 18   | 56.45                             | 56.74       | -0.29            |
| 19   | 54.26                             | 53.24       | 1.02             |
| 20   | 50.80                             | 49.11       | 1.69             |
| Sum  |                                   |             | 2.50             |
\[ \pm 0.06 \text{ and } \pm 0.19 \text{ cm. Thus, the great possibilities of algorithms of neural models were confirmed in the determining of prediction (Fig. 1). Positive values of the total network error indicate that the model is less accurate for weaker throwers (difference 1.62–0.19 m, between values of models prediction and the real distance of javelin throw, average 0.88 m). At the same time, in the case of the analysed group, it can be established that the MLP model adapted better to throwers achieving average and very good results. In the very good throwers’ group (throws above 56 m) an average value of differences between real distances of javelin throw and values of models prediction, was about 0.26 m (Table 8).

The results of this study indicate that MLP ANN can usefully be used in athlete selection.

5 Conclusions

The models developed in this research, based on the architecture of perceptron networks showed the capability for generalisation and prediction, which supported the use of neural networks in optimisation of the recruitment and selection process for javelin throwing. The results of versatile statistical analysis showed that the developed linear model, for the investigated group of young javelin throwers, was not able to describe precisely in a linear way the relationship between the dependent variable and each of the independent variables. Simultaneously, the results confirmed experimental possibility of using the perceptron models for prediction of javelin throw distances (the sports-result) in the investigated group for which the perceptron network using the 4-3-2-1 structure has the best predictive proprieties of sports-result for junior javelin throwers.

On the ground of developed model for measuring values of the investigated group and with the utilisation of module Run One-Off Case javelin throwers were recruited, whose javelin throw distances showed good agreement with the model prediction after 1 year of training. Hence, a MLP ANN could be a useful tool for young coaches for optimization of the selection process.

Summing up, it may be said that on the basis of experience, collected and precisely described data, the artificial neural networks may be applied to assist in the process of recruiting and predicting in javelin throwing.

Acknowledgments This project received the agreement for the research from the Bioethics Committee of Scientific Research in Academy of Physical Education in Katowice No. 16/06. The research under a title: “A neural network model approach to athlete selection”, was executed within the framework of the project: “The Multicriteria Optimization of Investigative Problems of the Sports Training”, for the Ministry of Science and Higher Education in Poland.

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Bartlett R, Müller E, Lindinger S, Brunner F, Morriss C (1996) Three-dimensional evaluation of the kinematic release parameters for javelin throwers of different skill levels. J Appl Biomech 12:7–14
2. Bartlett RM (2006) Artificial intelligence in sports biomechanics: new dawn or false hope? J Sports Sci Med 5:23–42
3. Best R, Bartlett R, Morriss C (1993) A three-dimensional analysis of javelin throwing techniques. J Sports Sci 11(4):315–328
4. Best R, Bartlett RM, Sawyer RA (1995) Optimal javelin release. J Appl Biomech 11:371–394
5. Bompa TO (2000) Training guidelines for young athletes. Champaign, IL, Human Kinetics. p 1–20
6. Cook G (2001) High-performance sports conditioning: baseline sports-fitness testing. Human Kinetics. p 19–49
7. Gregor RJ, Pink M (1985) Biomechanical analysis of a world record javelin throw: a case study. Int J Sport Biomech 1:73–77
8. Gunnell D, Magnusson P, Rasmussen F (2004) Low intelligence test scores in 18 year old men and risk of suicide: cohort study. http://www.bmj.com/content/330/7484/167.full.pdf. Accessed 22 Dec 2004
9. Hatton L (2005) Optimizing the javelin throw in the presence of prevailing winds. University of Kingston, U K, http://www.lesshatton.org/Documents/jav2007a_paper.pdf. Accessed 28 Jan 2007
10. Haykin S (1994) Neural networks, a comprehensive foundation. Macmillan College Publishing Company, New York
11. Hellwig Z (1969) The problem of the variables’ optimum choice. Stat Rev 3:33–45
12. Hubbard M, Bergman C (1989) Effect of vibrations on javelin lift and drag. Int J Sport Biomech, vol 5, pp 40–59
13. Hubbard M, Laporte S (1997) Damping of javelin vibrations in flight. J Appl Biomech 13:269–286
14. Lees A (2002) Technique analysis in sports: a critical review. J Sports Sci 20:813–828
15. Maier KD, Wank V, Bartonietz K, Blickhan R (2000) Neural network based models of javelin flight: prediction of flight distances and optimal release parameters. Sports Eng 3:57–63
16. Morriss C, Bartlett R (1996) Biomechanical factors critical for performance in the men’s javelin throw. Sports Med 21:22–37
17. Murakami M, Tanabe S, Ishikawa M, Isolehto J, Koml V, Itô A (2005) Biomechanical analysis of the javelin throwing at 11th World Championships in Athletics in Helsinki. New Stud Athl 12:7–14
18. Naclerio T (2005) Biomechanical analysis of the javelin throwing at 11th World Championships in Athletics in Helsinki. New Stud Athl 4:11–21
19. Naclerio T (2005) Teaching progressions book, reasons for throws technique breakdown by Coach Donahue. Massachusetts Northeastern University Press, Boston
20. Perl J, Lames M, Glitsch U (2002) Modeling in sport science. Hogmann, Berlin
21. Rutkowska D, Nowicki R (2000) Implication-based neuro-fuzzy architectures. Int J Appl Math Comput Sci 10(4):675–701
22. Rygul I (2000) Tools of the system analysis in the sport-training. AWF Katowice Press, Katowice
22. Strelau J (1998) Temperament: a psychological perspective. Kluwer, Norwell
23. Tadeusiewicz R (1993) Neural networks. Academic Publishing RM, Warszawa
24. Tidow G (1995) Models for teaching techniques and assessing movements in athletics. Int Associ Athl Fed: New Stud Athl 10:29–49
25. Tidow G (2000) Challenge decathlon–barriers on the way to becoming the “King of Athletes”. Part I. Int Associ Athl Fed: New Stud Athl 15:243–256
26. Velebil V, Vinduskova J (2007) Methodology training of athletic disciplines. Olympia, Praha
27. Wit A, Barton G (1996) Artificial neural networks. Compet Sport Press 9:14–22
28. Zadeh L (2002) From computing with numbers to computing with words. Int J Applied Math Comput Sci 12(2):4–12
29. Zakynthinaki MS, Sterling JR (2003) The mathematical modeling in the sport biomechanics. Competitive Sport Press 11–12:467–468
30. Zhuk E, Serikova E (2005) Effectiveness of the step-by-step discriminatory analysis at choosing the informative attributes. Automat Remote Control 66(11):1768–1781