

## Research Article

# ANN Based Approach for Estimation of Construction Costs of Sports Fields

Michał Juszczyk , Agnieszka Leśniak, and Krzysztof Zima 

Faculty of Civil Engineering, Cracow University of Technology, 24 Warszawska St., 31-155 Cracow, Poland

Correspondence should be addressed to Michał Juszczyk; [mjusczyk@izwbit.pk.edu.pl](mailto:mjusczyk@izwbit.pk.edu.pl)

Received 29 September 2017; Accepted 13 February 2018; Published 14 March 2018

Academic Editor: Andrej Soltesz

Copyright © 2018 Michał Juszczyk et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cost estimates are essential for the success of construction projects. Neural networks, as the tools of artificial intelligence, offer a significant potential in this field. Applying neural networks, however, requires respective studies due to the specifics of different kinds of facilities. This paper presents the proposal of an approach to the estimation of construction costs of sports fields which is based on neural networks. The general applicability of artificial neural networks in the formulated problem with cost estimation is investigated. An applicability of multilayer perceptron networks is confirmed by the results of the initial training of a set of various artificial neural networks. Moreover, one network was tailored for mapping a relationship between the total cost of construction works and the selected cost predictors which are characteristic of sports fields. Its prediction quality and accuracy were assessed positively. The research results legitimize the proposed approach.

## 1. Introduction

The results presented in this paper are part of a broad research, in which the authors participate, aiming to develop tools for fast cost estimates, dedicated to the construction industry. The main aim of this paper is to present the results of the investigations on the applicability of artificial neural networks (ANNs) in the problem of estimating the total cost of construction works in the case of sports fields as specific facilities. The authors propose herein a new approach based on ANNs for estimating construction costs of sports fields.

*1.1. Cost Estimation in Construction Projects.* Cost estimation is a key issue in construction projects. Both underestimation and overestimation of costs may lead to a failure of a construction project. The use of different tools and techniques in the whole project life cycle should provide information about costs to the participants of the project and support a complex decision-making process. In general, cost estimating methods can be classified as follows [1, 2]:

(i) Qualitative cost estimating:

- (1) Cost estimating based on heuristic methods
- (2) Cost estimating based on expert judgments

(ii) Quantitative cost estimating:

- (1) Cost estimating based on statistical methods
- (2) Cost estimating based on parametric methods
- (3) Cost estimating based on nonparametric methods
- (4) Cost estimating based on analogous/comparative methods
- (5) Cost estimating based on analytical methods.

The expectations of the construction industry are to shorten the time necessary to predict costs, whilst on the other hand, the estimates must be reliable and accurate enough. There are worldwide publications in which the authors report the research results which respond to these expectations. The examples of the use of a regression analysis (based on both parametric and nonparametric methods) are as follows: application of multivariate regression to predict accuracy of cost estimates on the early stage of construction projects [3], implementation of linear regression analysis methods to predict the cost of raising buildings in the UK [4], proposal and discussion of the construction cost estimation method which combines bootstrap and regression techniques [5],

and application of boosting regression trees in preliminary cost estimates for school building projects in Korea [6]. Another mathematical tool for which some examples can be given is fuzzy logic, for example, implementation of fuzzy logic for parametric cost estimation in construction building projects in Gaza Strip [7] or proposal and presentation of a fuzzy risk assessment model for estimating a cost overrun risk rating [8]. Case based reasoning (CBR) is also an approach which can be found in the publications dealing with the construction cost issue, for example, implementation of the CBR method improved by analytical hierarchy process (AHP) for the purposes of cost estimation of residential buildings in Korea [9] or the use of the case based reasoning in cost estimation of adapting military barracks also in Korea [10]. The examples of the publications which report and discuss the applications of artificial neural networks in the field of cost estimation and cost analyses in the construction process are presented in the next subsection.

*1.2. Artificial Neural Networks Cost Estimation in Construction Projects.* Artificial neural networks (ANNs) can be defined as mathematical structures and their implementations (both hardware and software), whose mode of action is based on and inspired by nervous systems observed in nature. In other words, ANNs are tools of artificial intelligence which have the ability to model data relationships with no need to assume a priori the equations or formulas which bind the variables. The networks come in wide variety depending on their structures, way of processing signals, and applications. The theory in this subject is widely presented in the literature (e.g., [11–15]). Main applications of ANN can be mentioned as follows (cf., e.g., [11, 12, 15]): prediction, approximation, control, association, classification and pattern recognition, associating data, data analysis, signal filtering, and optimization. ANNs features which make them beneficial in cost estimating problems (in particular for cost estimating in construction) are as follows:

- (i) Applicability in regression problems where the relationships between the dependent and many independent variables are difficult to investigate
- (ii) Ability to gain knowledge in the automated training process
- (iii) Ability to build and store the knowledge on the basis of the collected training patterns (real-life examples)
- (iv) Ability of knowledge generalization; predictions can be made for the data which have not been presented to the ANNs during a training process.

Some examples of ANN applications reported for a range of cost estimating and cost analyses in construction are replication of past cost trends in highway construction and estimation of future costs trends in this field in the state of Louisiana, USA [16], computation of the whole life cost of construction with the use of the concept of cost significant items in Australia [17], prediction of the total structural cost of construction projects in the Philippines [18], estimation of site overhead costs in the dam project in Egypt [19],

prediction of the cost of a road project completion on the basis of bidding data in New Jersey, USA [20], and cost estimation of building structural systems in Turkey [21]. The authors of this paper also have their contribution in studies on the use of ANN in cost estimation problems in construction. In some previous works, the authors presented the ANN applications for conceptual cost estimation of residential buildings in Poland [22–24] and estimation of overhead cost in construction projects in Poland [25, 26].

*1.3. Justification for Research.* It needs to be emphasized that, despite a number of publications reporting research projects on the use of artificial neural networks in cost analyses and cost estimation in construction, each of the problems is specific and unique. Each of such problems requires an individual approach and investigation due to distinct conditions, determinants, and factors that influence the costs of construction projects. An individual approach to cost estimation in construction is primarily due to specificity of the facilities, including sports fields. The costs of a sport field are significant not only for the construction stage but also later in terms of its maintenance. The decisions made about the size, functionality, and quality are crucial for the future use and operational management of sport fields. The success in investigation of ANNs applicability in the problem will allow proposing a new approach for estimation of the construction cost of sport fields. The new approach, based on the advantages offered by neural networks, will allow predicting the total construction cost of sport fields much faster than with traditional methods; moreover, it will give the possibility of checking many variants and their influence on the cost in a very short time.

## 2. Formulation of the Problem and Research Framework

*2.1. General Assumptions.* The general aim of the research was to develop a model that supports the process of estimating construction costs of sports fields. The authors decided to investigate implementation of ANNs for the purpose of mapping multidimensional space of cost predictors into a one-dimensional space of construction costs. In a formal notation, the problem can be defined generally as follows:

$$f : X \longrightarrow Y, \quad (1)$$

where  $f$  is sought-for function of several variables,  $X$  is input of the function  $f$ , which consists of vectors  $x = [x_1, x_2, \dots, x_n]$ , where variables  $x_1, x_2, \dots, x_n$  represent cost predictors characteristic of sports fields as construction objects, and  $Y$  is a set of values which represent construction costs of sports fields.

In the statistical sense, the problem comes down to solving a regression problem and estimating of a relationship  $f$  between the cost predictors being independent variables belonging to the set  $X$  and constructions cost of a sports field being dependent variable belonging to the set  $Y$ . According to the methodology in cost estimating based on statistical methods, one can distinguish between two

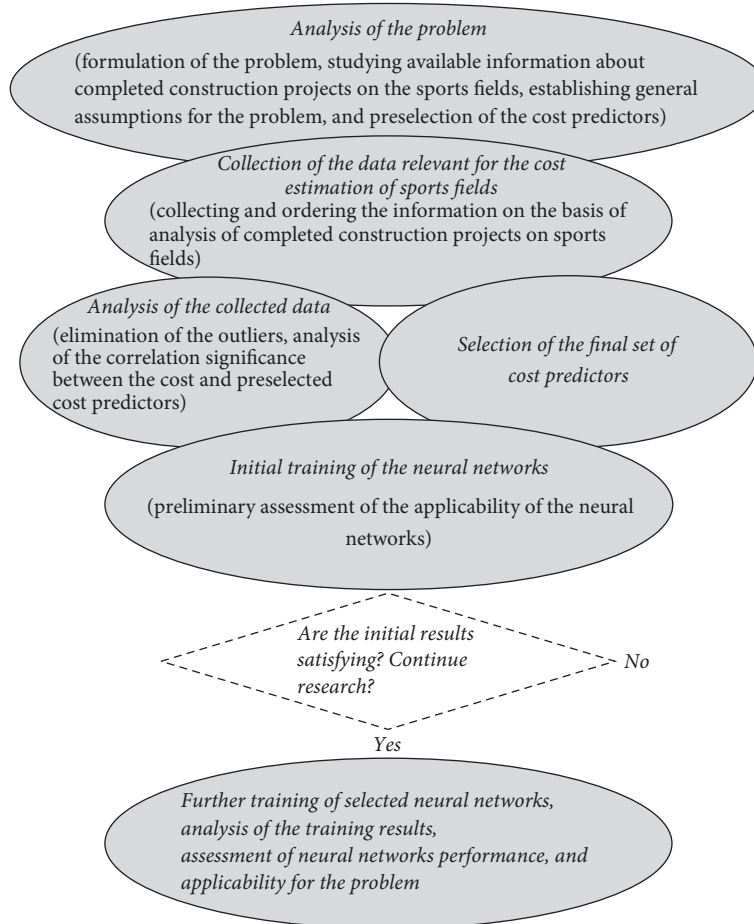


FIGURE 1: Scheme of the research framework (source: own study).

main approaches: estimating based on parametric methods and estimating based on nonparametric methods (cf. [1, 2, 25]). Both methods rely on the real-life data, that is, representative samples of cost predictors values and related construction costs values. In the case of the use of parametric methods function  $f$  is assumed a priori and the structural parameters of the model are estimated. On the other hand, nonparametric methods are based on fitting the function  $f$  to the data. According to the assumptions made for the research presented in this paper, the sought-for function was supposed to be implemented implicitly by ANN.

A general framework of the adopted research strategy is depicted in Figure 1.

**2.2. Characteristics of Sports Fields Covered by the Research.** Sports fields are facilities for which some types of works are usually repeated during the construction stage. The main types of works that can be listed are

- (i) geodetic surveying,
- (ii) earthworks (topsoil stripping, trenching, compacting of the natural subgrade, etc.),
- (iii) works on subgrade preparation for the sports field surface,

- (iv) works on sports fields surface (usually surfaces are either natural or synthetic grass),
- (v) assembly of fixtures and in-ground furnishings (e.g., football/handball gates, basketball goal systems, volleyball, or tennis poles and nets),
- (vi) works on fencing and ball-nets installation,
- (vii) minor road works and works on sidewalks,
- (viii) landscape works and arranging green areas around the sports fields.

In the course of the research, a number of completed projects on sports fields in Poland were investigated. Both the fields dedicated to one discipline and multifunctional fields were taken into account. The facilities subject to the analysis differed in size of the playing area, arranged area for communication, arranged green area, and fencing. The surfaces were of two types: either natural or synthetic grass. It must be stressed here that the quality expectations for surfaces varied significantly and played an important role in the construction costs. The completed facilities are located all over Poland both in the urban areas (in cities of different sizes) and outside the urban areas (in the villages).

TABLE 1: Characteristics of dependent variables and independent variables considered initially to be used in the course of a regression analysis (source: own study).

Description of the variable	Variable type	Values
Total cost of construction works	Quantitative	Cost given in thousands of PLN
Playing area of the sports field	Quantitative	Surface area measured in m <sup>2</sup>
Location of the facility	Quantitative	Urban area (big cities, medium cities, small cities) or outside the urban area (villages)
Number of sport functions	Quantitative	Number of sports that can be played in the field
Type of the playing field surface	Categorical	Natural or artificial grass
Quality standard of the playing field's surface	Categorical	Quality standard assessed according to the available information in the tender documentation
Ball stop net's surface	Quantitative	Surface area measured in m <sup>2</sup>
Arranged area for communication	Quantitative	Surface area measured in m <sup>2</sup>
Fencing length	Quantitative	Length measured in m
Arranged green area	Quantitative	Surface area measured in m <sup>2</sup>

### 3. Variables Analysis

*3.1. Preselection of Variables.* As the problem was formally expressed and the assumption about the use of ANNs was made, the authors focused on the analyses which allowed them to preselect cost predictors. The preselection was preceded by studying both technical and cost aspects of the construction projects on sports fields. This stage of the research allowed collecting the necessary background knowledge about the nature of sports fields as specific construction objects with their characteristic elements, range and sequence of construction works which must be completed, and the clients' quality expectations.

In the next step, 129 construction projects on sports fields that were completed in Poland in recent years were investigated. For the purposes of fast cost analysis, the authors preselected the following data to be the variables of the sought-for relationship:

- (i) Total cost of construction works as a dependent variable
- (ii) Playing area of a sports field, location of a facility, number of sport functions, the type of the playing field's surface (natural or artificial), quality standard of the playing field's surface, ball stop net's surface, arranged area for communication, fencing's length, and arranged greenery area as independent variables.

The criteria for such preselection were the availability of the data in the investigated tender documents and ensuring enough simplicity of the developed model due to which the potential client would be able to formulate the expectations about the sport field to be ordered by specifying values for potential cost predictors in the early stage of the project.

Most of the mentioned variables were of a quantitative type. In the case of the location of the facility, type of the playing field's surface, and quality standard of the playing field's surface, only descriptive information was available; the

three variables were of the categorical type. Table 1 presents synthetically the characteristics of all of the variables that were preselected in the course of the analysis of the problem.

The next step included data collection and scaling categorical variables. In the case of three of the variables (namely, location of the facility, type of the playing field surface, and quality standard of the playing field's surface), categorical values were replaced by numerical values. The studies of the problem, analyses of the number of completed construction projects on sports fields, and, especially, the analyses of construction works costs brought the conclusions of how the categorical values of the three variables are associated with the costs of construction works. Table 2 explains how the change of the categorical values stimulates the costs of construction works in general. The observation made it possible to order the values and transform them into numbers in the range from 0.1 to 0.9.

Categorical values for location of the facility have taken numerical values as follows: urban area: 0.9 for big cities (population over 100,000), 0.66 for medium cities (population between 20,000 and 100,000), and 0.33 for small cities (population below 20,000); outside the urban area: 0.1 for villages. In the case of the type of the playing field, surface artificial grass has taken the value of 0.9 and natural grass has taken the value of 0.1. Finally, depending on the client's expectations and specifications available in the tender documents, the descriptions of the demands for quality standard of the playing field's surface took values from the range between 0.1 and 0.9.

The studies of tender documents for public construction projects where completion of sports fields was the subject matter of the contract allowed for collecting the data for 129 projects. The data were collected for projects completed in the last four years all over Poland. The collected information was ordered in the database. After the analysis of outliers, the authors decided to reject some extreme cases for which the total construction cost was unusually high or unusually

TABLE 2: General relationship between the three categorical variables and cost (source: own study).

Location of a facility	Type of the playing field surface	Quality standard of the playing field's surface	Cost
Urban area, big cities	Artificial grass	High demands	Higher
Urban area, medium cities		Moderate demands	
Urban area, small cities	Natural grass	Low demands	Lower
Outside the urban area, villages			

TABLE 3: Significance of correlations between the variables (source: own study).

Preselected cost predictors	Is the correlation between the dependent variable, total cost of construction works, and independent variable significant for $pvalue < 0.05$ ?	Variable's symbol (for accepted variables only)
Playing area of the sports field	Yes	$x_1$
Location of the sports ground	No	-
Number of sport functions	No	-
Type of the playing field surface	Yes	$x_2$
Quality standard of the playing field's surface	Yes	$x_3$
Ball stop net's surface	Yes	$x_4$
Arranged area for communication	Yes	$x_5$
Fencing length	Yes	$x_6$
Arranged greenery area	Yes	$x_7$

low. After the elimination of outliers, the data for 115 projects remained.

**3.2. Selection of the Final Set of Variables.** Further analysis included the investigation of the significance of correlations between the dependent variable and all of the initially considered independent variables, preselected cost predictors. The significance of correlations for  $p$ -value  $< 0.05$  was assessed. The results of this step are synthetically presented in Table 3.

As the correlations for the two of preselected cost predictors (namely, location of the sports ground and the number of sport functions) appeared to be insignificant, they were rejected and no longer taken into account as the cost predictors.

Table 4 presents ten exemplary records with the specific numerical values of the dependent variable  $y$  (total cost of construction works) and independent variables  $x_j$  (cost predictors) accepted for the model.

Table 5 presents descriptive statistics for the variables accepted for the model. Average, minimum, and maximum values are presented for each of the variables as well as the standard deviation.

It is noteworthy that minimum value, namely, 0.00, for the variables  $x_4$ ,  $x_5$ ,  $x_6$ , and  $x_7$  corresponds with the fact that in case of certain sports fields elements such as ball stop nets, arranged area for communication, fencing, and arranged greenery have not been included in a project's scope (cf. Table 4). Moreover, there is some regularity in Table 4 which manifests in the distribution of average values closer to minimum values than to maximum values. This is due to the fact that the number of small-sized and medium-sized sports fields in the database was relatively greater than the number of large-sized ones. This can be explained by a general rule

valid for all kinds of construction. The number of small-sized and medium-sized facilities of all types, either newly built or existing, is always greater than those large-sized.

The database records (whose number equalled 115) were used as training patterns  $p$  for ANNs in the course of the research.

The values of the variables were scaled automatically before and after each of the ANN's training. This was done due to the functionalities of the ANN's software simulator used in the course of the research. The variables were scaled linearly to the range of values appropriate for activation functions employed for certain investigated ANN. The results, especially ANNs' training errors, presented further in the paper, are given as original, not scaled values.

## 4. Initial Training of Neural Networks

After the selection of independent variables, a formal notation of the relationship in the statistical sense can be given as follows:

$$y = f(x_1, x_2, x_3, x_4, x_5, x_6, x_7) + \varepsilon. \quad (2)$$

Consequently a prediction can be formally made:

$$\hat{y} = f(x_1, x_2, x_3, x_4, x_5, x_6, x_7), \quad (3)$$

where  $y$  is dependent variable, total cost of construction works, as observed in real life,  $\hat{y}$  is predicted total cost of construction works,  $f$  is sought-for function, implicit relationship implemented by ANN,  $x_1, x_2, x_3, x_4, x_5, x_6, x_7$  are independent variables, selected cost predictors as presented in Tables 3 and 4, and  $\varepsilon$  are random deviations (errors) for which  $E(\varepsilon) = 0$ ,

TABLE 4: Exemplary records of the database including training patterns (source: own study).

$p$	$y$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$
5	565.8	968	0.9	0.6	0.0	196.8	602.5	0.0
13	1359	3292	0.9	0.5	600.0	2142.0	0.0	0.0
23	427.6	1860	0.9	0.3	1116.0	78.0	37.5	1000.0
37	489.5	1860	0.1	0.3	240.0	100.0	0.0	0.0
46	323.0	800	0.9	0.5	192.0	181.8	139.7	0.0
59	181.3	800	0.1	0.3	72.0	0.0	0.0	400.0
67	1972.3	4131	0.9	0.5	396.0	1096.0	207.1	2586.4
82	161.1	1650	0.1	0.1	300.0	0.0	0.0	0.0
94	250.0	1470	0.1	0.2	344.5	93.9	38.5	0.0
101	800.5	1104	0.9	0.8	295.9	1469.1	93.3	374.6

TABLE 5: Descriptive statistics for the models' variables (source: own study).

	Variable's symbol	Average value	Minimum value	Maximum value	Standard deviation
Dependent variable	$y$	457.56	33.30	2592.50	373.14
	$x_1$	1333.79	275.00	5600.00	788.49
	$x_2$	0.59	0.10	0.90	0.27
	$x_3$	0.49	0.10	0.90	0.16
Independent variables	$x_4$	300.26	0.00	2212.00	345.27
	$x_5$	193.16	0.00	2142.00	325.84
	$x_6$	105.24	0.00	602.50	121.13
	$x_7$	324.29	0.00	3000.00	603.26

The aim of this stage of the research, namely, the initial training of ANNs, was to assess their applicability to the problem in general and to take a decision whether to continue the research or not. A variety of feed forward ANNs were trained in the automatic mode. The overall number of networks equalled 200; the authors took into account 100 multilayer perceptron (MLP) networks and 100 radial basis function (RBF) networks as the types appropriate for the regression analysis and suitable for the formulated problem.

The main criteria to assess the applicability of the ANNs were the quality of predictions made by trained networks and the errors. The measure for quality of predictions was Pearson's correlation coefficient  $R(y, \hat{y})$  between that in real life and that predicted by networks values of the independent variable, the total construction cost, whereas the measures of error was the root mean squared error (RMSE):

$$R(y, \hat{y}) = \frac{\text{cov}(y, \hat{y})}{\sigma_y \sigma_{\hat{y}}}, \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_p (y^p - \hat{y}^p)^2},$$

where  $\text{cov}(y, \hat{y})$  is covariance between  $y$  and  $\hat{y}$ ,  $\sigma_y$  is standard deviation for  $y$ , and  $\sigma_{\hat{y}}$  is standard deviation for  $\hat{y}$ .

In the case of RBF networks, both the quality of predictions and errors were so dissatisfying that the authors decided to focus on the MLP networks only. The results for the MLP networks were satisfying; they are presented synthetically below in Figure 2 and in Table 5 and discussed briefly. The main assumptions for this stage were as follows:

- (i) From the database of training patterns, learning ( $L$ ), validating ( $V$ ), and testing ( $T$ ) subsets were randomly drawn 10 times.
- (ii) The overall number of available training patterns was divided into the three subsets in relation:  $L/V/T = 60\%/20\%/20\%$ .
- (iii) For each drawing, 10 different networks were trained.
- (iv) The networks varied in the number of neurons in the hidden layer,  $H$ .
- (v) Distinct activation functions, such as linear, sigmoid, hyperbolic tangent, and exponential, were applied in the neurons of a hidden and output layer.

Learning and validating, that is,  $L$  and  $V$ , subsets were used in the course of training process. The third subset,  $T$ , was used for testing purposes after completing the training process as an additional check of the generalization capabilities (cf., e.g., [11]). Number of neurons in the hidden layer,  $H$ , was assessed according to the following equation [27] and inequality [28]:

$$H \approx \sqrt{NM}, \quad (5)$$

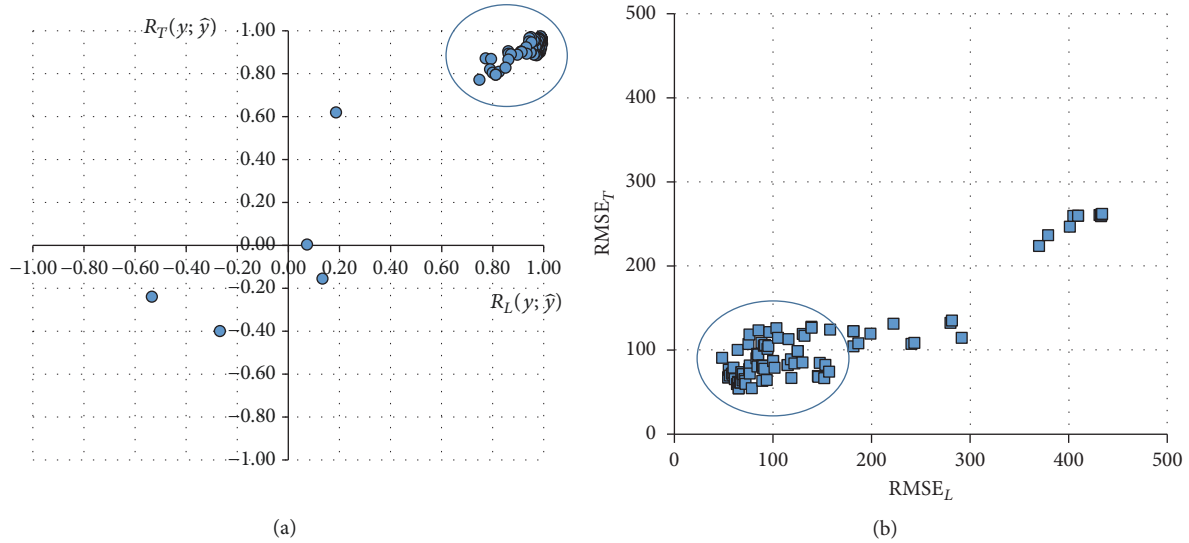
where  $N$  is number of neurons in the input layer and  $M$  is number of neurons in the output layer.

$$NNP = NNW + NNB < L, \quad (6)$$

where  $NNP$  is number of ANN's parameters,  $NNW$  is number of ANN's weights,  $NNB$  is number of ANN's biases, and  $L$  is cardinality of a learning subset.

TABLE 6: Summary of the initial training of ANNs for MLP networks (source: own study).

Subset	$R(y, \hat{y})$			RMSE	
	Average	Standard deviation	$R > 0.9$	Average	Standard deviation
$L$	0.901	0.240	82.1%	145.5	111.7
$V$	0.879	0.228	81.0%	108.8	57.8
$T$	0.881	0.233	80.5%	126.9	83.3

FIGURE 2: Quality end errors of ANNs after the initial training phase: (a) scatter diagram of Pearson's correlation coefficients and (b) scatter diagram of errors (RMSE);  $L$  and  $T$  stand for learning and testing subsets accordingly (source: own study).

From (5),  $H \approx 2,646$ . According to the assumptions about the cardinality of  $L$  subset and from inequality (6),  $NNP < 69$ . Compromising these two conditions, the number of neurons in the hidden layer,  $H$ , varied between 2 and 6.

The overall results, in terms of the quality of predictions and errors, are presented in Figures 2 and 3.

Part (a) of Figures 2 and 3 depicts scatter plots of Pearson's correlation coefficients calculated for each network after the training process. Figure 2 shows coefficients for learning and testing subsets, whereas Figure 3 shows the coefficients for learning and validating subsets. In part (b) of both figures, one can see scatter plots of errors (namely, RMSE). Figure 2 presents the scatter plot for learning and testing subsets, and Figure 3 presents the scatter plot for learning and validating subsets.

Table 6 presents the summary of the initial training of 100 MLP networks. The average and standard deviation of  $R(y, \hat{y})$  as well as percentage of the cases for which  $R(y, \hat{y})$  is greater than 0.9 are presented in the table. Additionally, the average and standard deviation for RMSE errors are given. All values were calculated for learning, validating, and testing subsets separately.

As can be seen both in Figures 2 and 3 and in Table 6, the correlations for most of the cases are very high. For more than 80% of networks,  $R(y, \hat{y})$  was greater than 0.9 in case of learning, validating, and testing. There are evident clusters of points in Figures 2(a) and 3(a) which represent networks with the potential of the acceptable or good quality of prediction.

There are only few cases outside the clusters for which the correlations coefficients are very low, due to the failure of the training process. RMSE errors are in the acceptable range at this stage.

This stage of the research confirmed the general applicability of ANNs to the investigated problem. The decision was to continue the research. Moreover, it allowed choosing a group of MLP networks to be trained in the next stage.

## 5. Results of Neural Networks Training in the Closing Phase of the Research

With respect to the initial training results, a group of 5 networks was chosen for the closing phase of the research. The details including networks' structures and activation functions are given in Table 7. (All of the networks consisted of 7 neurons in the input, the number of neurons ranging from 2 to 5 in one hidden layer and one neuron in the output layer. All of the networks were trained with the use of Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm.)

Assumptions for the networks training and testing were different than in the initial stage. From the set of 115 training patterns, the testing subset,  $T$ , was chosen randomly in the beginning. This subset remained unchanged for all of the networks and it was used to assess the generalization capabilities after all of the training processes. Cardinality of  $T$  subset equalled 10% of the total number of training patterns.

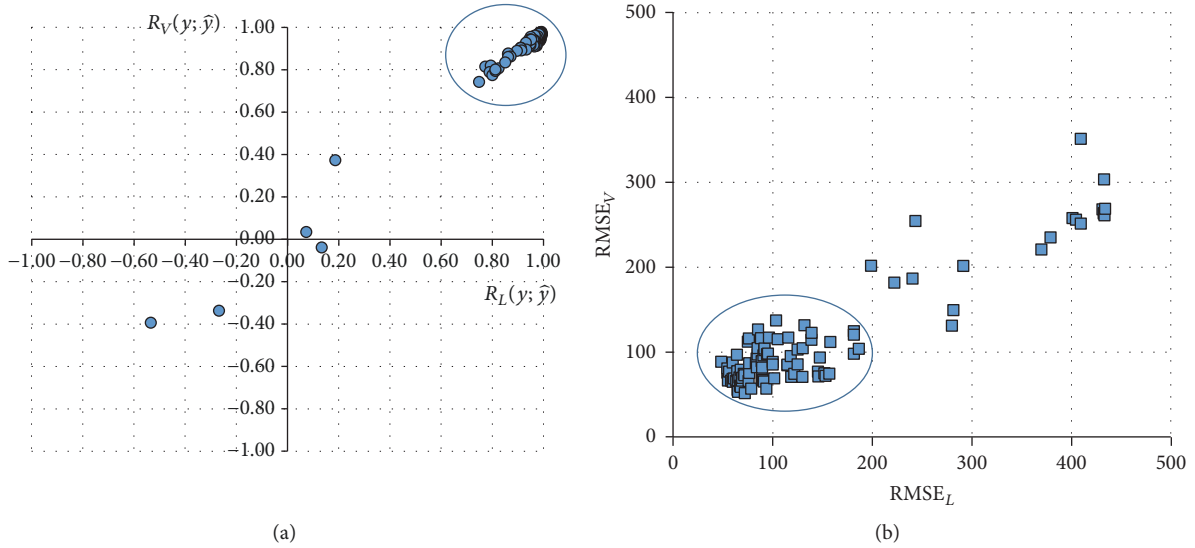


FIGURE 3: Quality end errors of ANNs after the initial training phase: (a) scatter diagram of Pearson's correlation coefficients and (b) scatter diagram of errors (RMSE);  $L$  and  $V$  stand for learning and validating subsets accordingly (source: own study).

TABLE 7: Details of the selected ANNs for further training (source: own study).

ANN	Number of neurons in the hidden layer	Activation function hidden layer	Activation function output layer	Training algorithm
$MLP_{e-1} 7-2-1$	2	Exponential	Linear	BFGS
$MLP_{e-ht} 7-2-1$	2	Exponential	Hyperbolic tangent	BFGS
$MLP_{e-1} 7-3-1$	3	Exponential	Linear	BFGS
$MLP_{s-1} 7-5-1$	5	Sigmoid	Linear	BFGS
$MLP_{ht-e} 7-5-1$	5	Hyperbolic tangent	Exponential	BFGS

The remaining patterns have been involved in the 10-fold cross-validation of the networks (cf. [11]). The patterns available for the training process after the sampling of the  $T$  subset have been divided into the 10-fold complementary learning and validating subsets. The relation of the subsets was  $L/V = 90\%/10\%$  accordingly. The sum of squared errors of prediction (SSE) was used as the error function in the course of training:

$$SSE = \sum_{p \in L} (y^p - \hat{y}^p)^2. \quad (7)$$

The performance of ANNs was assessed in general in the light of correlation between real-life and predicted values, RMSE errors of the prediction (as in the stage of initial training). Table 8 presents a summary of the training results for the five chosen networks after the training process based on 10-fold cross-validation approach. The results are given in terms of the networks' performance and errors. To synthesize and assess the performance and stability of training of each network, maximum, average, and minimum as well as the dispersion of the correlation coefficients  $R(y, \hat{y})$  between real-life and predicted values were calculated. Errors, namely, RMSE values, are presented in the same manner. Both

correlations and errors are given separately for  $L$ ,  $V$ , and  $T$  subsets. The analysis of the results made it possible to choose finally one of the five networks. The most stable performance was observed for the  $MLP_{s-1} 7-5-1$ .

Figure 4 depicts a scatter plot of the points that represent pairs  $(y^p, \hat{y}^p)$  for the finally chosen network  $MLP_{s-1} 7-5-1$ . Real-life values  $y$  are set together with predicted values  $\hat{y}$ . The points represent training results for learning and validating subsets, as well as testing results for testing subset. In the legend of the chart, apart from the letters  $L$ ,  $V$ , and  $T$  that explain membership of the certain pattern to the learning, validating, or testing subset accordingly, the numbers from 1 to 10 are given next to each letter. These numbers reveal the  $k$ th fold of the cross-validation process. In Figure 4, one can see that the points in the scatter plot are decomposed along a perfect fit line. The deviations are in the acceptable range. In respect of the analysis of  $R(y, \hat{y})$ , RMSE errors, and distribution of points in the scatter plot, the results are satisfactory.

Two more criteria, relating to the accuracy of cost estimation, were also specified for assessment of the selected network  $MLP_{s-1} 7-5-1$ . The accuracy of estimates was assessed with the use of three error measures: mean absolute percentage error (MAPE), absolute percentage error calculated for



TABLE 8: Results of the selected ANNs training (source: own study).

ANN	$R(y, \hat{y})$			RMSE		
	$L$	$V$	$T$	$L$	$V$	$T$
<b>MLP<sub>e-l</sub> 7-2-1</b>						
Max	0.992	0.997	0.997	92.043	111.521	68.006
Average	0.985	0.982	0.993	66.416	63.681	41.286
Min	0.973	0.948	0.985	49.572	20.138	26.656
Standard deviation	0.005	0.015	0.004	11.620	24.243	12.338
<b>MLP<sub>e-ht</sub> 7-2-1</b>						
Max	0.994	0.994	0.991	150.544	376.102	79.538
Average	0.978	0.979	0.983	73.351	90.999	57.609
Min	0.933	0.933	0.979	43.473	40.899	35.816
Standard deviation	0.022	0.018	0.003	31.521	95.430	13.082
<b>MLP<sub>e-l</sub> 7-3-1</b>						
Max	0.994	0.994	0.991	150.544	376.102	79.538
Average	0.978	0.979	0.983	73.351	90.999	57.609
Min	0.933	0.933	0.979	43.473	40.899	35.816
Standard deviation	0.022	0.018	0.003	31.521	95.430	13.082
<b>MLP<sub>s-l</sub> 7-5-1</b>						
Max	0.997	0.994	0.996	65.226	96.700	71.010
Average	0.992	0.983	0.991	46.222	59.770	39.252
Min	0.986	0.949	0.980	30.333	29.581	19.582
Standard deviation	0.004	0.015	0.008	12.048	17.682	18.116
<b>MLP<sub>ht-e</sub> 7-5-1</b>						
Max	0.996	0.995	0.996	129.738	211.452	83.939
Average	0.979	0.980	0.980	77.525	72.796	50.383
Min	0.946	0.957	0.952	32.191	41.891	27.691
Standard deviation	0.013	0.013	0.014	27.375	49.569	18.072

TABLE 9: MAPE and PE<sub>max</sub> errors calculated for the MLP<sub>s-l</sub> 7-5-1 (source: own study).

MLP <sub>s-l</sub> 7-5-1	$L$	$V$	$T$
<b>MAPE</b>			
Max	18.76%	25.13%	21.20%
Average	12.68%	14.43%	9.97%
Min	7.49%	5.54%	5.60%
Standard deviation	3.49%	5.66%	4.44%
<b>PE<sub>max</sub></b>			
Max	115.83%	104.72%	93.42%
Average	65.47%	61.55%	46.67%
Min	34.34%	9.76%	11.87%
Standard deviation	21.35%	27.59%	17.65%

each pattern ( $PE^P$ ), and maximum absolute percentage error ( $PE_{\max}$ ):

$$\begin{aligned}
 \text{MAPE} &= \frac{100\%}{n} \sum_P \left| \frac{y^P - \hat{y}^P}{y^P} \right|, \\
 PE^P &= \left| \frac{y^P - \hat{y}^P}{y^P} \right| 100\%, \\
 PE_{\max} &= \max(PE^P).
 \end{aligned} \tag{8}$$

Table 9 shows the maximum, average, and minimum MAPE and PE<sub>max</sub> errors, as well as the standard deviations of these errors, after the 10-fold cross-validation for the selected network MLP<sub>s-l</sub> 7-5-1. The errors are given for the learning, validating, and testing, that is,  $L$ ,  $V$ , and  $T$ , subsets, respectively.

MAPE and PE<sub>max</sub> errors have been carefully investigated for the selected network in all of 10 cases of cross-validation training and testing. In respect of average MAPE errors, the results were satisfying. Average MAPE errors were expected to be smaller than 15% for learning, validating, and testing

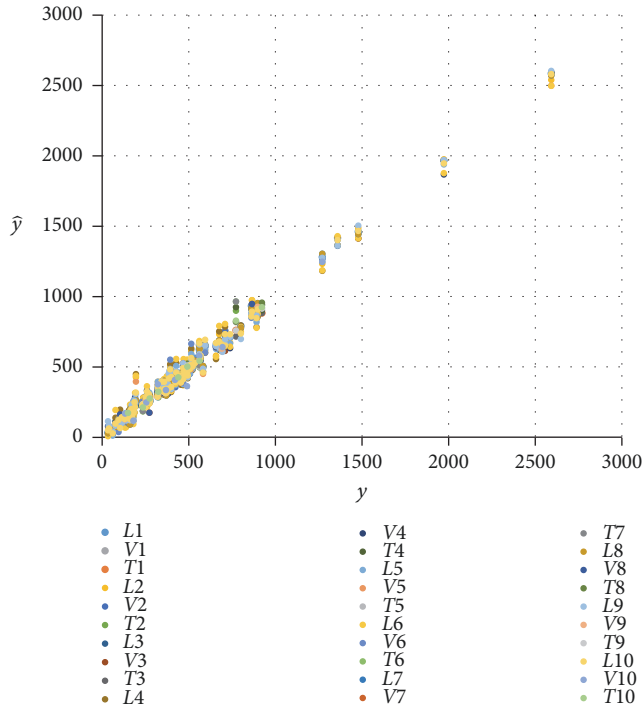


FIGURE 4: Scatter plot of  $\hat{y}$  and  $y$  values for the selected  $MLP_{s-1}$  7-5-1 network (source: own study).

subsets. This objective was achieved. In case of  $PE_{\max}$  errors, the authors expected the values lower than 30%. Table 9 reveals that the  $PE_{\max}$  errors are greater than 30%. This fact has prompted the authors to examine  $PE^P$  errors. Analysis of the distribution of  $PE^P$  errors revealed that most of these errors were smaller than 30% and only few values in each of the cross-validation folds values exceeded 30%.

A thorough analysis of the selected network, namely,  $MLP_{s-1}$  7-5-1, in terms of the training results, performance, and errors allowed selecting this network to implement implicitly the sought-for relationship  $f$ . In conclusion, the selected  $MLP_{s-1}$  7-5-1 may be proposed as the tool which supports cost estimation of construction works in the projects related to sports fields.

## 6. Summary and Conclusion

In this paper, the authors presented their investigations on applicability of ANNs in the problem of estimating the total cost of construction works for sports fields. The research allowed the authors to confirm assumptions about the general applicability of the MLP type networks as the tool which has the potential of mapping the relationship between the total cost of construction works and selected cost predictors which are characteristic of sports fields. On the other hand, the RBF type networks appeared to not be suitable for this particular cost estimation problem.

Apart from the general conclusions about the applicability of ANNs, one type of network tailored for this problem was selected from a broad set of various MLP networks. The

analysis of the results indicates a satisfactory performance of the selected network in terms of correlations between the real cost and cost predictions. The level of the MAPE and RMSE errors is acceptable. For now, the proposed approach allows the following:

- (i) Estimating cost of construction works for a couple of variants of sports fields in a very short time
- (ii) Supporting the decisions made by the client being aware of the range of the cost estimation accuracy.

The obtained results encourage continuation of the investigations which will aim to improve the model, especially to lower the  $PE_{\max}$  errors. The authors intend to collect additional data which will enable them to exceed the number of training patterns. Moreover, the authors intend to investigate in the near future more complex tools based on ANNs, such as committees of neural networks.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

The authors use the STATISTICA software in their research; therefore some of the presented assumptions for neural networks training are made due to the functionality of the tool.

## References

- [1] P. M. M. Foussier, *From Product Description to Cost: A Practical Approach, : The Parametric Approach*, vol. 1, Springer Science Business Media, 2006.
- [2] A. Layer, E. T. Brinke, F. Van Houten, H. Kals, and S. Haasis, "Recent and future trends in cost estimation," *International Journal of Computer Integrated Manufacturing*, vol. 15, no. 6, pp. 499–510, 2002.
- [3] S. M. Trost and G. D. Oberlender, "Predicting accuracy of early cost estimates using factor analysis and multivariate regression," *Journal of Construction Engineering and Management*, vol. 129, no. 2, pp. 198–204, 2003.
- [4] D. J. Lowe, M. W. Emsley, and A. Harding, "Predicting construction cost using multiple regression techniques," *Journal of Construction Engineering and Management*, vol. 132, no. 7, Article ID 002607QCO, pp. 750–758, 2006.
- [5] S. Varghese and S. Kanchana, "Parametric Estimation of Construction Cost Using Combined Bootstrap And Regression Technique," in *Proceedings of the International Journal of Civil Engineering and Technology*, vol. 5, pp. 201–205, 2014.
- [6] Y. Shin, "Application of boosting regression trees to preliminary cost estimation in building construction projects," *Computational Intelligence and Neuroscience*, vol. 2015, Article ID 149702, 2015.
- [7] N. I. El Sawalhi, "Modelling the Parametric Construction Project Cost Estimate using Fuzzy Logic," in *Proceedings of the International Journal of Emerging Technology and Advanced Engineering*, vol. 2, pp. 63–636, 2012.

- [8] I. Dikmen, M. T. Birgonul, and S. Han, "Using fuzzy risk assessment to rate cost overrun risk in international construction projects," *International Journal of Project Management*, vol. 25, no. 5, pp. 494–505, 2007.
- [9] S.-H. An, G.-H. Kim, and K.-I. Kang, "A case-based reasoning cost estimating model using experience by analytic hierarchy process," *Building and Environment*, vol. 42, no. 7, pp. 2573–2579, 2007.
- [10] S. H. Ji, M. Park, and H. S. Lee, "Case adaptation method of case-based reasoning for construction cost estimation in Korea," *Journal of Construction Engineering and Management*, vol. 138, no. 1, pp. 43–52, 2011.
- [11] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, New York, NY, USA, 1995.
- [12] S. Haykin, *Neural networks: A Comprehensive Foundation*, Prentice Hall PTR, 1994.
- [13] S. Osowski, "Sieci neuronowe w ujęciu algorytmicznym," *Wydawnictwa Naukowo-Techniczne*, 1996.
- [14] S. Osowski, "Sieci neuronowe do przetwarzania informacji," *Oficyna Wydawnicza Politechniki Warszawskiej*, 2000.
- [15] R. Tadeusiewicz, *Sieci neuronowe*, vol. 180, Akademicka Oficyna Wydawnicza, Warszawa, 1993.
- [16] C. G. Wilmot and B. Mei, "Neural network modeling of highway construction costs," *Journal of Construction Engineering and Management*, vol. 131, no. 7, pp. 765–771, 2005.
- [17] A. Alqahtani and A. Whyte, "Artificial neural networks incorporating cost significant items towards enhancing estimation for (life-cycle) costing of construction projects," *Australasian Journal of Construction Economics and Building*, vol. 13, no. 3, pp. 51–64, 2013.
- [18] C. L. C. Roxas and J. M. C. Ongpeng, "An Artificial Neural Network Approach to Structural Cost Estimation of Building Projects in the Philippines," in *Proceedings of the DLSU Research Congress*, p. 1, Manila, 2014.
- [19] I. Elsayy, H. Hosny, and M. A. Razeq, "A neural network model for construction projects site overhead cost estimating in Egypt," *International Journal of Computer Science Issues*, vol. 8, no. 3, pp. 273–283.
- [20] T. P. Williams, "Predicting completed project cost using bidding data," *Construction Management and Economics*, vol. 20, no. 3, pp. 225–235, 2002.
- [21] H. M. Günaydin, S. Z. Dogan, H. M. Günaydin, and S. Z. Doğan, "A neural network approach for early cost estimation of structural systems of buildings," in *Proceedings of the International Journal of Project Management*, vol. 22, pp. 595–602, 2004.
- [22] M. Juszczyk, "The Use of Artificial Neural Networks for Residential Buildings Conceptual Cost Estimation," in *Proceedings of the 11th International Conference of Numerical Analysis and Applied Mathematics 2013, ICNAAM 2013*, pp. 1302–1306, Greece, September 2013.
- [23] M. Juszczyk, "Application of committees of neural networks for conceptual cost estimation of residential buildings," in *Proceedings of the International Conference on Numerical Analysis and Applied Mathematics 2014, ICNAAM 2014*, Greece, September 2014.
- [24] M. Juszczyk, "The Challenges of Nonparametric Cost Estimation of Construction Works with the use of Artificial Intelligence Tools," in *Proceedings of the Creative Construction Conference, CCC 2017*, pp. 415–422, Croatia, June 2017.
- [25] M. Juszczyk, A. Lesniak, and A. Leśniak, "Site overhead cost index prediction using RBF Neural Networks. Transactions on Economics and Management," *ICEM*, pp. 381–386, 2016.
- [26] A. Leśniak, "The application of artificial neural networks in indirect cost estimation," in *Proceedings of the 11th International Conference of Numerical Analysis and Applied Mathematics 2013, ICNAAM 2013*, pp. 1312–1315, Greece, September 2013.
- [27] E. Pabisek, *Systemy hybrydowe integrujące MES i SSN w analizie wybranych problemów mechaniki konstrukcji i materiałów*, Wydawnictwo Politechniki Krakowskiej, Kraków, 2008.
- [28] Z. Waszczyszyn, "Zastosowanie sztucznych sieci neuronowych w inżynierii lądowej, XLI Konferencja Naukowa KILiW PAN i Komitetu Nauki PZITB," *Kraków-Krynica*, vol. tom 9, pp. 251–288, 1995.



Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

