Research Article

Neural Network-Based Model for Predicting Preliminary Construction Cost as Part of Cost Predicting System

Diana Car-Pusic ⓘ,[1] **Silvana Petruseva** ⓘ,[2] **Valentina Zileska Pancovska** ⓘ,[2] **and Zlatko Zafirovski** ⓘ[2]

[1]*Faculty of Civil Engineering, University of Rijeka, Rijeka 51000, Croatia*
[2]*Faculty of Civil Engineering, University "Ss. Cyril and Methodius", Skopje 1000, North Macedonia*

Correspondence should be addressed to Diana Car-Pusic; diana.car.pusic@uniri.hr

A model for early construction cost prediction is useful for all construction project participants. This paper presents a combination of process-based and data-driven model for construction cost prediction in early project phases. Bromilow's "time-cost" model is used as process-based model and general regression neural network (GRNN) as data-driven model. GRNN gave the most accurate prediction among three prediction models using neural networks which were applied, with the mean absolute percentage error (MAPE) of about 0.73% and the coefficient of determination R^2 of 99.55%. The correlation coefficient between the predicted and the actual values is 0.998. The model is designed as an integral part of the cost predicting system (CPS), whose role is to estimate project costs in the early stages. The obtained results are used as Cost Model (CM) input being both part of the Decision Support System (DSS) and part of the wider Building Management Information System (BMIS). The model can be useful for all project participants to predict construction cost in early project stage, especially in the phases of bidding and contracting when many factors, which can determine the construction project implementation, are yet unknown.

1. Introduction

The complex cost estimation problem in the field of building construction is the problem which is traditionally burdened by lack of data, uncertainties, and risks, but at the same time very important for the success of a construction project. Due to all of these, numerous construction projects are faced with significant cost overruns, which are elaborated extensively in the paper. The causes of this condition are complex and are the subject of research presented in this paper and supported by data. One of the causes, which is to be particularly emphasized, is the focal point of this paper. This important cause is an early initial cost prediction, which is often of unsatisfactory accuracy. The reason is the lack of information in the initial stages and the desire to get results in a short time, not going too far into its accuracy and the extent of the consequences such data could have on the project. Such a superficial and inaccurate assessment results in a number of further steps in the project, resulting in multiple negative consequences that could jeopardize the implementation of project goals. The desire of the parties to come up with information about the costs as soon as possible is understandable and will always be present, regardless of the type of project or of its size. Therefore, there is a need to create a reliable cost prediction system.

The unsatisfactory and uncertain cost prediction [1] and their overrun in construction projects are a very frequent [2–4] and not easily solvable problem. Due to the uniqueness, diversity, complexity of projects, and the ever-present risks, establishing the model for enough accurate assessment of the project costs is doubtless a challenging task. That is why for many researchers this problem is often the subject of their research, whereby they use different approaches and methods often for a certain type of buildings and structures [1, 5–16]. The aim is to establish as accurate a cost estimation model as possible that would be applied in the initial project

phases. The fact that the contracted costs are often exceeded is also evidence of the claim that the cost prediction is in a lot of projects inadequate. The cause may be "... a heavily experience-based process" according to Alex et al. (2010) as it is cited in [1], which means that the estimation is not based on scientifically proven methods, then the application of low-accuracy models or inadequate models for the case under consideration, or even intentional miscalculations [3]. Data on cost overruns of completed projects that are the subject of numerous scientific studies are the evidence of a previous claim of the frequency of cost overruns [2–4]. As stated by Žujo et al. [2], one of the reasons is "...the absence of a thorough expert analysis of conditions, circumstances, and possible risks when concluding a contract."

There are numerous reasons why research often focuses on construction cost. Cost is factor that can be expressed quantitatively and unambiguously. When conducting research regarding construction costs in different countries, numerous researchers indicate frequent significant cost overruns of many construction projects [3, 4, 17, 18]. Hence, for example, Baloi and Andrew [19] have presented the results of the Morris and Hough [20] research resulting in significantly exceeded costs in 63% of the 1778 projects financed by the World Bank, constructed between 1974 and 1988.

The authors in [19] state that cost overrun is more a rule than an exception. Moreover, according to the reports from the World Bank in 2007, road construction in India suffers about 25% of contracted price overrun [21]. According to the research conducted in China [22], where various types of reconstructed structures were considered, the construction contracted price overrun of more than 10% was recorded at 26.39% of the structures and 5–10% at 55.56% of the structures.

In Slovenia, a research was conducted on a sample of 92 traffic structures built in the period from 1993 to 1998. The average contracted price overrun was 51% [23]. A similar study was conducted in Australia in the period from 1992 to 1999. 93 structures were analyzed and the cost overrun was recorded in 21 or 22.58% of structures [24].

Within the scientific research project conducted in Croatia [25], 333 structures were investigated in the period from 1996 to 1998. Price overrun at 81% of structures was recorded.

A similar research was conducted in Bosnia and Herzegovina on 177 structures built from 1995 till 2006. The results indicated that the contracted date was not met in 51.40% of structures and the contracted price was not met in 41.23% of structures [26].

It can be concluded that construction cost overrun is present not only in underdeveloped countries and developing countries but also in developed countries. This was also confirmed by Baloi and Andrew [19], stressing that "... in most developing countries ... the problem is more acute." Reasons are surely multifaceted and multilayer and deserve a deeper analysis of the issue.

Therefore, estimating construction costs already in the initial stages of the project is the subject of special attention of the researchers, which does not lose on the actuality. In doing so, the special attention of the researchers is focused on modeling the interdependence of costs and other variables, primarily on the duration of the construction.

Considering the complexity and the significance of the problem, other opportunities should be explored, which have a greater potential for solving such complex tasks, which are undoubtedly integrated management information systems whose prediction cost system should be an integral part.

2. The Main Objectives and the Research Framework

One of the main objectives of the research is to evaluate the results of applying the proposed combined process-based and data-driven cost estimation model, that is, hybrid model, and compare its accuracy with the results of simple models. The second objective is to propose a basic concept of cost prediction system (CPS) as a part of a Building Management Information System (BMIS), with a more detailed elaboration of NNs module which includes also hybrid models.

The recommendation about applying the results of the considered case of the proposed hybrid model in CPS will also be presented.

Steps in researching, implementing, and displaying the results are as follows:

(1) Review of the existing references on cost prediction in construction projects.

(2) Review of the existing references on CPS ontology basics.

(3) Creating a proposal for cost prediction system ontology.

(4) Predicting construction costs by using a hybrid process-based and data-driven model.

(5) Recommendations for the results' integration into the CPS.

3. Literature Review: Construction Cost Prediction

The Australian Bromilow was the first to investigate financial execution in relation to construction time for a total of 329 structures in the building construction area (built in Australia between 1963 and 1967). The research resulted in establishing the so-called "time-cost" model (hereinafter BTC or TC model) [27, 28]. The simple linear regression analysis method was applied whose suitability was also proven in numerous later researches [18, 29]. Despite being originally a "time-cost" model, it also served as a template for examining the interdependence between construction costs and construction time. It was noted that construction cost prediction and also cost interdependence with time (as quantitative factors) can be mathematically modeled according to Bromilow "time-cost" model by using simple linear regression [2, 29]. Furthermore, scientific studies indicate that there is a dependency between the contracted

construction price/cost and time at various construction markets [3, 4, 17, 18, 30].

However, the researchers did not only rely on modeling the interdependence of building time costs but have also introduced new predictors, for example, number of floors, gross floor area, type of facility, and type of client. In their research, some researchers emphasized the risk factors that cause cost overruns. Thus, Le-Hoai et al. [4] apply the factor analysis technique to categorize the causes. Ranking of causes in terms of occurrence and severity was conducted. "Poor site management and supervision, poor project management assistance, financial difficulties of owner, and financial difficulties of contractor are ranked as the first problems." Spearman's rank correlation tests do not point out differences in ranking the main causes among three groups of respondents (owners, contractors, and consultants).

Multiregression analysis is also applied as a mathematical method. Hence, Alshamrani [5] developed a multiregression model for conceptual initial cost estimation of conventional and sustainable college buildings in North America. The obtained model can predict the initial cost in USD/ft^2 in dependence on the following predictors: height of one floor, building space, number of floors, sustainability index (1 for conventional and 2 for sustainable), and structure type.

Multiple regression analysis is also used by authors [6] to develop an early parametric model, that is, a model for early cost estimation. The research was based on data for thirty-three real-constructed road tunnel projects. It was concluded that the employed approach using multiple regression analysis is valid for heavy construction projects.

In addition to researching the application of regression analysis to estimate the cost of construction projects, another direction of research has been focused on the application of neural networks to obtain expected project costs. Thus, Ahiaga-Dagbui and Smith [1] in their research on 98 water-related construction projects built in Scotland in the period 2007–2011 applied ANN to determine models for cost estimation. Impacts, such as construction site conditions, price changes, purchases, various possible risks, and contractual changes, were taken into account.

Separate cost models for normalized target cost and log of target costs were developed. Variable transformation and weight decay regularization were then explored to improve the final model's performance. As a prototype of a wider research, the final model's performance was very satisfactory, demonstrating ANN's ability to capture the interactions between the predictor variables and final cost. Ten input variables, all readily available or measurable at the planning stages for the project, were used within a Multilayer Perceptron Architecture and a quasi-Newton training algorithm [1].

El Sawy et al. [31] pointed out that cost prediction is one of the tasks of successful management of construction projects, that is, cost management. Cost prediction is a demanding task. Instead of the usual methods, one should turn to the more sophisticated ways of predicting. In the mentioned research [31], the researchers used the ANN

approach to develop a parametric cost-estimating model for site overhead costs. The research was conducted on 52 real-life cases of building projects constructed in Egypt during the seven-year period from 2002 to 2009. N-Connection Professional Software version 2.0 was used for the development of neural network models. The neural network architecture is presented for the estimation of site overhead costs as a percentage of the total project price.

When it comes to the problem of construction site overhead costs, it is worth noting the quite new research from Poland from 2019 [16] for a few reasons. The authors claim that the "Construction site overhead costs are key components of cost estimation in construction projects. The estimates are expected to be accurate, but there is a growing demand to shorten the time necessary to deliver cost estimates." After considering and then combining several types of neural networks, in order to select the members of the ensemble, the authors developed three models intending to predict a construction site overhead cost index.

It was proved that proposed models offer better cost prediction than those based on single neural networks [16].

Neural networks are also applied by Petroutsatou, Georgopoulos, Lambropoulos, and Pantouvakis [7] for early cost estimation for 33 twin tunnels with a total length of 46 km in Greece. As first, the authors determined the parameters that affect the temporary/final support and the final cost of tunnel construction, such as geometrical, geological, and data related to quantities of works. After that, the data were analyzed using two neural network types: the first was multilayer feed-forward network (MLFN), and the second was a general regression neural network (GRNN). In the next step, model results have been compared with costs and quantities from the real projects. It was concluded that the usage of developed models leads to fairly accurate cost estimation and quantities of works for road tunnels. It was also concluded that the NNs usage for cost estimation is beneficial, due to NNs capability for modeling nonlinear data relationships.

A very interesting artificial neural network (ANN) approach to predicting index of indirect cost of construction projects in Poland was applied in research presented in [32]. Based on the quantitative study of 72 cases of building projects constructed in Poland, "the factors conditioning indirect costs and the actual costs incurred by enterprises during project implementation" have been determined [15].

Another relevant research was carried out by Juszczyk et al. [8] on a sample of 129 sports field construction projects that have been implemented in Poland in recent years. The possibility and justification of the application of the NN for the assessment of total construction costs for sports' fields were explored. As one of the research tasks was to establish a set of cost predictors, 7 predictors regarding the technical and functional characteristics were established. After that, the data were analyzed using two neural network types: multilayer perceptron networks (MLP) and radial base function networks (RBF). By applying Pearson's correlation coefficient between real and predicted values of construction costs and by using the root mean square error (RMSE) as the measure of prediction errors, satisfactory results were

established for MLP networks. This proved the applicability of the cost estimation network. In the next step, the analysis for a group of 5 MLP networks was performed and the results were compared. As a comparison measure, Pearson's correlation coefficient was used between the actual and predicted construction cost and the root mean square error (RMSE) as the measure of the prediction errors. The accuracy of the estimation was tested using mean absolute percentage error (MAPE). The best results for all assessors were established for one network. In conclusion, this type of network can be recommended for estimating the sports field construction costs.

It was to be expected that the course of modeling development of these interdependencies would be redirected towards the comparison of the accuracy and applicability of the models obtained using various techniques. In this respect, comparative models obtained by applying different regression techniques without neural networks, as well as using neural networks, supporting vectors, case-based reasoning techniques, and others, have been developed.

Kim et al. [9] have been exploring the performance of three cost estimation models. A database of 530 implemented project costs of Korean residential buildings has been used. Three-type techniques have been applied for estimating construction costs and their results have been compared: multiple regression analysis (MRA), neural networks (NNs), and case-based reasoning (CBR). Model performance was measured by the Mean Absolute Error Rate (MAER) as the measure of the difference between estimated and actual construction costs. Comparing results from 40 test data, the best MAER of 2.97% with the 48% of the estimates within 0–2.5% of the actual error rate and 98% within 10% has been established. The CBR model gave MAER of 4.81% with 43% of the estimates within 2.5% and 83% within 10%. In spite of these results, the authors point to slowness in establishing a NN model because of the trial and error process. They point to the need to take into account the compromise between accuracy, speed, and clarity when explaining the cost and choosing an estimation model. In this sense, CBR is considered a better model. Future research is expected to create a hybrid model that would combine different techniques.

On the other hand, the research, which was carried by the authors [33], compared the accuracy of cost estimation using two types of models-linear based regression models and vector support vector machines (SVMs) models. The models were applied on a database of 75 buildings built on the territory of the Federation of Bosnia and Herzegovina. The usual estimators, the coefficient of determination R^2, and the mean absolute percentage error (MAPE) were used. MAPE is a measure of accuracy, so a better result was established for SVM which also has a better R^2. The weakness of the SVM model is the speed of convergence in relation to the LR model.

From all of the aforementioned, it can be concluded that neither one of the techniques nor one of the estimation models can be considered absolutely the best for all the conditions and circumstances of the construction of this type of structure. Olawe and Sun [34] and Ahsan and

Gunawan [35] stated that, despite the availability of various control techniques and project control software, many construction projects still do not achieve their cost goals.

4. Cost Predicting System (CPS) as the Part of Building Management Information System (BMIS)

4.1. Basic Framework. Timely cost estimation of satisfying accuracy is one of the crucial factors that affect project performance and thus represent essential management information for the highest level of management in business systems. In this regard, a cost predicting system (CPS) is proposed as a possible integrative component of the system responsible for improving the effectiveness and effectivity of the construction through cost planning and predicting different levels of detail, phases, and project as a whole. All of these systems are integrated into the Building Management Information System (BMIS) as shown in Figure 1.

As the assessment procedures themselves are demanding in terms of required knowledge as well as time-consuming, it is necessary to integrate them into a single information management system that possesses the necessary historical and other data used in these models and forms part of the Decision Support System (DSS) in business construction and project systems. Ma et al. [36] point out a large amount of information that is collected on a daily basis, thanks to information systems in construction companies. Authors call them "reusable legacy information" and discuss two approaches to their possible use, using general or specialized software.

Reflecting on the future development of construction through the prism of past experience and knowledge as well as of new development trends, the integration of separate segments is a development challenge and therefore probably an imperative. The solutions it brings have a synergistic potential with the ability to improve significantly the operational, functional, economic, management, and quality dimensions of the construction. Watson [37] classifies the "fragmented structure" into one of the underlying, inherent construction industry problems. Egan [38], in his famous Rethinking Construction, advocates "... the use of computer modeling to predict the performance for the customer." The same author considers one of the goals to be "annual reductions of 10% in construction cost and construction time."

It can be argued that the strength and potential of computer modeling, as a technical platform, are unprecedented at the obtained development level. What needs to be reexamined is the utilization of such potential and of new possibilities. Again, utilization should be linked to a human factor, that is, lack of readiness, engagement, organizational, and managerial competencies, that is, attitude and commitment to integration. Egan [38] points out that "... the way forward for achieving the ambition of a modern construction industry lies in commitment." Recognizing the benefits that integration can bring and the commitment to integration is a longer-term process that will underpin the future development of construction industry.
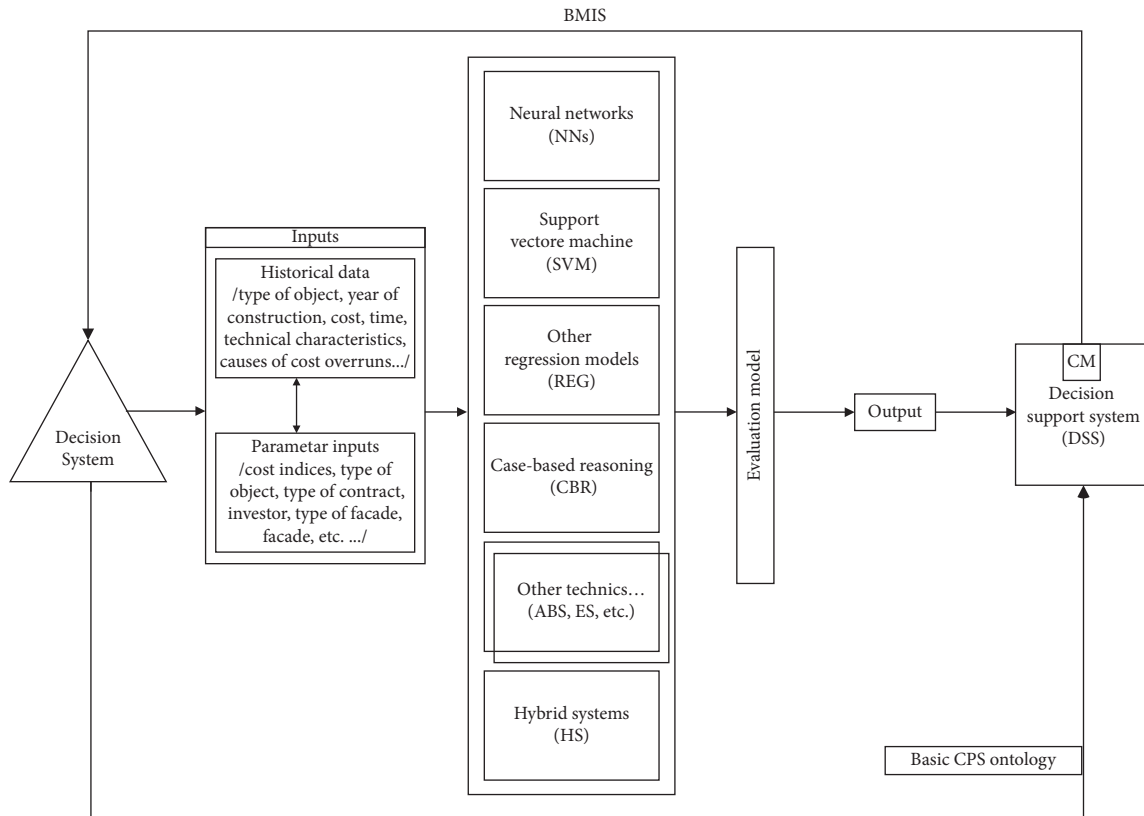
FIGURE 1: CPS ontology.

The development of a unique information management system is inevitable technical support for the operation of an integrated construction system. The authors of this paper advocate an open modular, upgradeable, flexible, and adaptable system that would in its generic form be widely applicable, with the possibility of incremental adaptation and upgrade depending on the local needs. In this regard, it is worth highlighting the developed multimodel-based Management Information System concept as one of the results of the "Mefisto" research project presented by Scherer and Schapke in their paper [39]. The conceptual multilevel model of the information system is presented in the paper, in which the third level is foreseen for construction economic (cost and time) and specification models.

4.2. CPS Ontology. When it comes to the starting points for creating the proposed CPS ontology, it is the result of previous and subsequent research of other authors and own research results [1, 8, 10, 18, 29, 30, 33, 40].

Its essential determinants are as follows:

(i) Integration of different models and cost prediction techniques.

(ii) Use of historical data for implemented projects.

(iii) Valorization of results obtained by applying two or more estimation models.

(iv) Foreseeing the application of hybrid models depending on the degree of their development.

(v) Integration of output into the Decision Support System (DSS).

Although the proposed CPS ontology integrates different models of cost predicting, this paper focuses on NNs due to their specific characteristics and capabilities identified by previous research [7, 8, 10–12, 33, 40, 41].

Based on the above, the following benefits of NNs should be highlighted:

(i) Self-learning ability in the training-process.

(ii) Knowledge-generalization ability.

(iii) Possible prediction on other data sets.

(iv) Processing rate.

(v) Rate of estimation of a large number of variants.

(vi) Applicability for problems in which it is difficult to determine the functional dependence between dependent and independent variables.

(vii) Good predictive ability in conditions of insecurity and incomplete data.

(viii) Prediction based on previous cases and so on.

Figure 1 shows the basic structure of CPS ontology as part of the comprehensive Building Management Information System (BMIS).

Although NN also has deficiencies ("black box" decisions), it can generally be said that they are more pronounced when it comes to other intelligent techniques applied to cost estimation. Significant contribution in terms

of comparing "intelligent techniques in construction project cost estimation" was made by Elfaki et al. [10] The authors compared five categories of intelligent cost estimation techniques: Machine Learning Systems (ML) techniques-neural networks and the support vector machine (SVM), Knowledge-Based Systems (KBS) techniques, expert systems and case-based reasoning (CBR), evolutionary systems (ES) used as optimization tools, Agent-Based System (ABS) simulating actions, and interactions and evaluating the effects on the system. Hybrid Systems (HS) is the fifth and perhaps the most challenging category because it represents a set of different techniques. This enables overcoming the limitations of each individual technique.

Thus, for example, the authors list deficiencies of the KBS systems to be "difficulty of self-learning and time-consuming during the rule acquisition process," while for ES somewhat difficult generalization is listed.

Based on the above, the proposed CPS ontology was structured, consisting of the following components:

(i) Input-data component.

(ii) Central-processing component.

(iii) Output with the evaluation module.

The input part consists of a database of historical project data and an input parameter database. These bases are complex and structured according to certain predetermined criteria (e.g., by category and type of structures or by other defaults), so that data selection can be made according to a variety of criteria. This allows the creation of homogeneous databases that provide accurate time estimation data while processing. The historical database includes data on constructed structures, planned and incurred costs, and time of construction, as well as reasons for cost and deadline overruns (risks). It also includes categories and types of structures, their purpose, and their technical characteristics, for example, the number of floors, size, surface, type of facade, year of construction or reconstruction, type of client, and type of contract.

The input parameter database contains the appropriate parameters that are the inputs in the estimation models and defines the individual features from the historical database. There are, for example, price indices by months and years, currency rates, parameters that determine the technical characteristics of the structure (e.g., various types of facades can be encoded with certain numbers), parameters of the purpose of the structure, parameters related to the type of risk, and type of client and contract.

The process part integrates appropriate prognostic software systems that use these data, so that, through processing, the estimated costs for a particular structure based on its characteristics are obtained, and by data processing, a more similar and homogeneous group of previously constructed structures is determined. In the specific case, the processing can be done using one (which is not recommended), two, or more models within the system, and in the evaluation part, the accuracy of the results is compared using statistical indicators (most often it will be MAPE and R^2) as the usual measures of accuracy and suitability of the

model. As can be seen from Figure 2, the NNs module itself integrates different types of networks (GRNN, MLP, Multilayer Perceptron, RBF NNs, Polynomial NNs, Cascade Correlation NN, Probabilistic NNs, etc.), which are suitable for different data types so that, by processing, the optimal type and network architecture for the structure in question are determined on the homogeneous database as possible. The homogeneity of the base is achieved by a series of parameters, not only by the type of structures but also by the financial value of the investment, similar technical characteristics, the type of client and contract, and so on. The homogeneity of the database positively influences the reliability of the estimation. If necessary, the normalization of the input data is performed. As the result of the processing of a particular database, the optimal network type with all the indicators that define it is obtained. The results are stored in the DSS system and used in cost estimation and future business decisions. In this paper, the GRNN network is presented as part of the NNs of the CPS module. Optimal data processing results would be integrated into the DSS system together with the parameters of the selected network and data processing architecture (the number of neurons per layer, the number of hidden layers, the activation function, the sigma parameter value, the number of iterations, the conjugate algorithm gradient, the validation method, and other).

One of the future development trends should certainly be sought in the development and application of hybrid models that carry significant synergistic potential in solving cost prediction problems, but also other complex problems.

The chosen most accurate result, together with all the relevant features of the processed model, becomes part of the Decision Support System (DSM), which has a complex structure, Cost System (CS) being an integral part of it, while both are part of the wider Building Management Information System (BMIS). Regardless of the choice of modules and techniques, it is clear that such an integrated cost prediction system provides a powerful tool for fast multivariate data processing and evaluation of results and saves time compared to conventional unintegrated partial and time-consuming processes. This is a strong argument for applying such a system as systematic support in making business decisions.

5. Predicting Construction Costs by Using Process-Based and Data-Driven Model

5.1. Methods. As the first phase of the investigation, a survey was conducted to collect data for estimated and real construction cost of the structures, construction time (predicted and real), year of the construction, structure type (purpose), construction site region, technical characteristics of the structure, and other data (e.g., about risk factors), but not relevant for this research. Data were collected by the questionnaires and, due to the sensitivity of some data, during face-to-face interviews with project participants (investors, contractors, designers, and construction surveyors). The survey covered one hundred and sixteen structures constructed in the Republic of North Macedonia
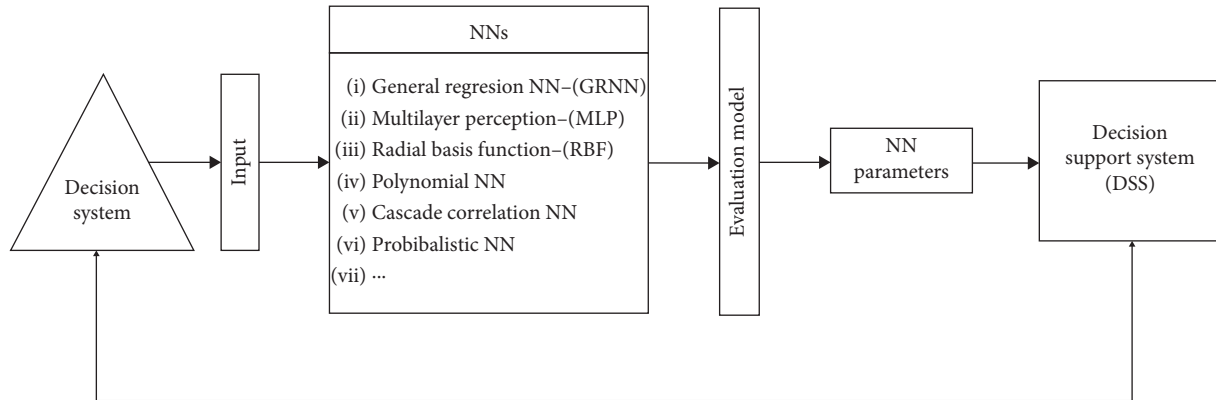
FIGURE 2: NNs as part of CPS.

and in the Republic of Croatia during the last two decades. The database will be described in more detail below. In the next phase of the investigation, the historical data for constructed structures were used in the process for developing the construction cost prediction model.

In order to predict the price of the construction accurately, the combination of two types of methods (models) is used: process-based method and data-driven method. The main difference between process-based models and data-driven (statistical) models is that the process-based models are based on the assumed knowledge of the actual process. Process-based models use the laws of the considered physical process, so that their results have broad applicability. To develop a process-based model, a very good understanding of the process is required, along with accurate and extensive data in order to obtain that analytical law (mathematical formulae) for the process [32].

The data-driven (statistical) models are based only on the observed relationships in the data and do not assume knowledge about the laws between the input and output variables in the actual process; they use only the actual values of the input and output variables and need only good selection of relevant independent variables and an appropriate output (dependent) variable which will describe the process well.

When the estimations of the parameters for the process-based models are difficult to be obtained, when they are not precise, or when the data for the development of the process-based models are not available, then the data-driven models can be used [32]. In civil engineering, data-driven models became popular because of the increasing availability of the data in the construction industry. They make maximal use of the available data, extracting useful relationships and conclusions from the existing data sets.

5.2. Process-Based Model.
The process-based model used in this paper for predicting the construction cost is Bromilow's well-known "time-cost" model [28], which gives the relation between the construction time and construction price (equation (1)).

$$A = P \cdot B^Q, \tag{1}$$

where A is the contracted time, B is the contracted price, P is the model parameter showing the average time needed for construction of a monetary value, and Q is the parameter that shows time dependence of cost change [28].

Equation (1) is used in this paper for the relation of contracted time and contracted price and also for real time and real price, because these data are available in the input data:

$$A_1 = P_1 \cdot B_1^{Q_1}, \tag{2}$$

$$A_2 = P_2 \cdot B_2^{Q_2}, \tag{3}$$

where A_1 and B_1 are contracted time and contracted construction cost, respectively, and A_2 and B_2 are real time and real construction cost, respectively.

In order to obtain simpler equations for modeling, equations (2) and (3) will be logarithmized:

$$\ln(A_1) = \ln(P_1) + Q_1 \ln(B_1), \tag{4}$$

$$\ln(A_2) = \ln(P_2) + Q_2 \ln(B_2). \tag{5}$$

By summing up equations (4) and (5), (6) is obtained:

$$\ln A_1 + \ln A_2 = \ln(P_1) + \ln(P_2) + Q_1 \ln(B_1) + Q_2 \ln(B_2). \tag{6}$$

From equation (6), the dependence of B_2 (real cost) from A_1, A_2, and B_1 can be obtained:

$$\ln(B_2) = \frac{1}{Q_2} \left(\ln A_1 + \ln A_2 - \ln P_1 - Q_1 \ln B_1 - \ln P_2 \right). \tag{7}$$

Because of equation (7), as input data for the artificial neural network used in this paper, the actual values for real price, real time, contracted price, and contracted time are not used, but logarithm of their values.

5.3. Data-Driven Model.
The data-driven model used in this paper is artificial neural network (ANN), more specifically, general regression neural network (GRNN), which will be described below.

Over the last two decades, artificial neural networks (ANNs) were of great interest in civil engineering, because they have demonstrated very good and often very accurate solutions to the wide range of complex nonlinear computation problems from many branches of civil engineering [40, 42]. ANNs are empirically derived modeling methods and versatile predictors that are being trained using a comprehensive set of examples of the problem, which is being solved, and their target solutions. Inspired by biological neural systems, they learn from experience, that is, from many input patterns and their appropriate outputs. The success of ANN applications depends mostly on selecting appropriate type and structure of the NN for solving the problem and the quality of the data used for training of the ANN.

For different type of data, different type of ANN or modeling method will be suitable. Several types of modeling methods should be always tested in order to choose the one which will give the most accurate results. In this research, multilayer perceptron (MLP), radial basis function (RBF), and general regression neural network (GRNN) were tested and the most accurate predicting was obtained using GRNN.

5.4. General Regression Neural Network (GRNN). GRNN is a neural network with a highly parallel structure that provides estimation of numerical variables and converges to a linear or nonlinear regression surface. This NN can be used for any nonlinear regression problem, for prediction, mapping, and modeling, or as a controller [43].

GRNN needs only a few training samples in order to converge to the basic function of the data, which makes this NN be very useful tool for application in practice, particularly for sparse data.

GRNN is very similar to RBF (radial basis function NN) with many nodes and, in comparison with well-known MLP NN (multilayer perceptron NN), it is faster to train and in many cases more accurate, but it is slower than MLP at classification of new cases and needs more memory space for storing the model.

The basic regression equation, from the statistical theory, is

$$y(X) = E\left[\frac{y}{X}\right] = \frac{\int_{-\infty}^{\infty} y f(X, y) \mathrm{d}y}{\int_{-\infty}^{\infty} f(X, y) \mathrm{d}y}. \tag{8}$$

E[y/X] is the conditional expectation of y for given X and $f(X, y)$ is the joint probability density function (jpdf) of the vector X and scalar y. When the function $f(X, y)$ is not known, it is being estimated from any of the Parzen estimators [44] using a finite set of observations of X and y and Gaussian Kernel [43]:

$$f\left(\frac{X}{y}\right) = \frac{1}{(2\pi)^{((p+1)/2)} \sigma^{p+1}} \frac{1}{n} \sum_{i=1}^{n} \exp\left(-\frac{\left(X - X^i\right)^T \left(X - X^i\right)}{2\sigma^2}\right) \exp\left(-\frac{\left(y - Y^i\right)^2}{2\sigma^2}\right), \tag{9}$$

where p is the dimension of the input vector X, n is the number of training samples, σ is the smoothing parameter, X is the input vector for which y should be estimated, X^i is i-th training sample, and Y^i is the appropriate measured value of y.

The integration over y in equation (8) can be computed by substitution equation (9) in equation (8), and the obtained estimation for Y is given in equation (10) [45].

$$\hat{Y}(X) = \frac{\sum_{i=1}^{n} Y^i exp\left(-\left(\left(X - X^i\right)^T \left(X - X^i\right)/2\sigma^2\right)\right)}{\sum_{i=1}^{n} exp\left(-\left(\left(X - X^i\right)^T \left(X - X^i\right)/2\sigma^2\right)\right)}. \tag{10}$$

The architecture of GRNN is shown in Figure 3 [46]. There is the same number of neurons in the input layer as predictor variables and input neurons feed the values of input variables to the neurons in the hidden layer. Each neuron from the hidden layer contains the data for each row (case) from the training set, that is, the values of all predictors and target value for one case. The hidden layer computes the Euclidean distance of the test case from the neuron's center and applies kernel RBF function. The resulting value is fed to the next pattern layer. Pattern layer has only two neurons: numerator summation unit which for each hidden neuron adds up the weight values multiplied by the actual value of the target variable and denominator summation unit which adds up the weight values from the hidden neurons. The value from the numerator summation unit is divided by the value from the denominator summation unit in the decision layer.

In the next section, the results for the prediction is going to be presented.

5.5. Database. Database consists of 116 structures data, built on the territory of the Republic of North Macedonia, 75 in total, and 41 built on the territory of the Republic of Croatia during the last two decades. The database consists of 51 buildings data, 53 construction structures, and 12 others (e.g., gas stations, multilevel car parking, electrical substations, and storage buildings). For future research, homogenization of bases is recommended to obtain more accurate results. In this research, the focus was on the number of cases in the database and the analysis and comparison of multiple models with an emphasis on the evaluation of the hybrid model.

6. Results

For modeling the data and predicting the real construction price, general regression neural network (GRNN) from the
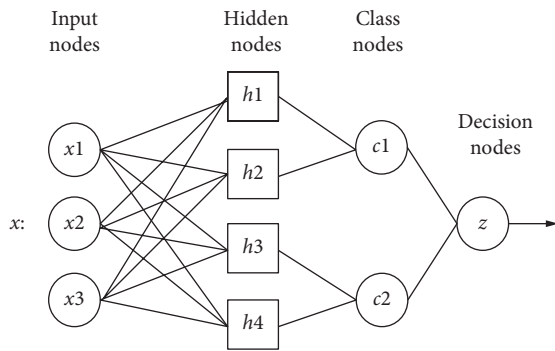
Figure 3: GRNN architecture [46].

predicting modeling software DTREG [46, 47] was used. The standard estimators of the model, the mean absolute percentage error (MAPE), and the coefficient of determination R^2 which reflects the overall fit of the model are MAPE = 0.73% and R^2 = 99.55%. The coefficient of correlation between actual and predicted values of the target variable is 0.998 (Table 1, Validation data).

The available data used for modeling were purpose of the facility, planned (contracted) price, real price which was achieved, and contracted and real construction time.

Bromilow time-cost model is used for choosing the input values for the target and predictors. According to the discussion in the previous section, ln (real price) is used as a target variable, and ln (real time), ln (contracted time), ln (contracted price), and purpose of the facility are used as predictors. Initial input knowledge that is available is the values of the target variable and predictors for 116 built structures.

For all numerical variables (predictors and target), DTREG obtains their minimal, maximal, mean value, and their standard deviation (Table 2).

For validation of the model, DTREG offers 4 choices:

(1) Random percent of the rows are held out when the model is being made and after the building of the model, that number of rows is run through the model and the error is evaluated.

(2) Control variable is used to select which rows will be selected to be held out for testing.

(3) Cross-validation with the chosen number of folds.

(4) Cross-validation with one row left out of each built model.

In Table 1, the results for the training and validation data are given using cross-validation method with 10-fold.

DTREG computes the relative importance of each predictor to the quality of the model, using sensitive analysis. Table 3 shows this importance with an accuracy of 3 decimal places. The displayed values are percentage values of the importance of every predictor in the model for predicting the target variable (real cost).

It can be seen that the most important predictor for predicting the real price is the planned (contracted) construction price.

Table 1: Results for the training and validation data (DTREG software).

Estimators of the model accuracy (DTREG)	Value
Training data	
Mean target value for input data	13.358369
Mean target value for predicted values	13.356284
Variance in input data	4.4677631
Residual variance after model fit	0.0024144
Proportion of variance explained by model R^2	0.99946 (99.946%)
Coefficient of variation (CV)	0.003678
Normalized mean square error (NMSE)	0.000540
Correlation between actual and predicted R	0.999731
Maximum error	0.3219897
RMSE (root mean squared error)	0.0491365
MSE (mean squared error)	0.0024144
MAE (mean absolute error)	0.0288461
MAPE (mean absolute percentage error)	0.2199448
Validation data	
Mean target value for input data	13.358369
Mean target value for predicted values	13.35876
Variance in input data	4.4677631
Residual variance after model fit	0.0199458
Proportion of variance explained by model R^2	0.99554 (99.554%)
Coefficient of variation (CV)	0.010572
Normalized mean square error (NMSE)	0.004464
Correlation between actual and predicted R	0.997882
Maximum error	0,5402981
RMSE (root mean square error)	0.1412296
MSE (mean square error)	0.0199458
MAE (mean absolute error)	0.0984472
MAPE (mean absolute percentage error)	0.7326534

Figure 4 shows the chart for the dependence of the predicted target values (ln (real cost, euro)) and the most important predictor (ln (planned costs, euro)).

Figure 5 shows the chart for the dependence of the actual and predicted values of the target variable.

Discussion with the Proposal of the Results Integration into Decision Support System.

Before choosing the GRNN model for predicting, other two predictive models were tested: multilayer perceptron (MLP) and radial basis function (RBF) neural network.

Because the relationship between target variable and predictors is not known in advance, several models must be tested in order to choose the best one for the actual data to provide the highest accuracy.

Table 4 presents the comparison of the accuracy among these three predictive models that were tested, using Bromilow "time-cost" model, the results for validation data for all 3 predictive models (GRNN, MLP NN, and RBF NN).

It is necessary to point out that using the Bromilow "time-cost" model drastically improved the accuracy of the prediction of these three models.

Without using the Bromilow model and by using only the actual values of numerical variables, the contracted time, and cost and the real time, as well as the target (real cost), the MAPE of the GRNN model was over 100% because of large

TABLE 2: Minimal, maximal, mean value and standard deviation for the numerical variables (predictors and target).

Continuous variables					
Variable	Rows	Minimum	Maximum	Mean	Std. dev.
ln (planned costs, €)	116	9.05358	18.36511	13.27770	2.06661
ln (real costs, €)	116	9.09459	18.51599	13.35837	2.11371
ln (planned time, days)	116	2.70805	6.81619	5.07076	1.04090
ln (real time, days)	116	2.70805	7.50934	5.14528	1.063

differences in values of the target variable. Figure 6 shows part of the input data, used for training of GRNN.

In practice, the hybrid models have demonstrated in many cases better results than when applying only one of them. Lee et al. [48] proposed a hybrid ANN (artificial neural network) called GRNNFA, which is a combination of fuzzy adaptive resonance theory model (FA) and the general regression neural network model (GRNN), developed for classification of noise data. The model removes the noise that is embedded in the training data and retains the best features of the two single models, fast training, good learning, and a network with an incremental growing structure. The performance of this hybrid model, when compared to the other published results, presented better results. The accuracy of predicting was around 96.11%.

To solve the issue of large-scale data, Wang et al. [49] proposed the TSE-GRNNs (tree-structure ensemble general regression neural networks) model. First, small-scale sample subsets are constructed using the regression tree algorithm. After that, GRNN submodels are constructed on these sample subsets, followed by the application of TSE-GRNNs method to establish the predictive model. Experiments show excellent predictive results.

Other authors also used ANN for predicting construction costs.

The authors in [50] used ANN to predict construction cost for apartment projects in Vietnam and obtained accuracy of the model with MAPE about 10%. They compared the ANN model with multiple linear regression (MLR) model and genetic algorithm model (GA), and the best accuracy was obtained with the ANN.

The author Juszczyk [13] uses several types of MLP to model the cost estimation of the construction works (residential buildings). The mean average percentage error (MAPE) for the validation data for the 5 MLP NN was from about 7% to 13%.

It is very important to mention that the accuracy of the model depends mostly on the selection of appropriate predictors for the chosen target variable and the selection of the appropriate ANN or some other regression models.

The authors in [14] have proposed data-driven methods for cost estimation of spherical storage tanks projects, based on the application of ANNs and hybridized regression models with genetic algorithm (GA), without using ANNs. The variables used in these models were thickness, tank diameter, and length of the weld. They have used two types of NNs (multilayer perceptrons): with Levenberg-Marquardt algorithm (LMNN) and Bayesian regulated (BRNN). The results have shown that both ANNs have performed

TABLE 3: Importance of the variables.

Overall importance of variables	
Variable	Importance
ln (planned costs, €)	100.000
Purpose of the facility	0.377
ln (planned time, days)	0.069
ln (real time, days)	0.042

TABLE 4: Comparison of the accuracy among GRNN, MLP NN, and RBF NN (results for validation data).

Predictive model	MAPE%	R^2	R
GRNN	0, 73	99, 55	0, 998
MLP NN	0, 97	98, 99	0, 995
RBF NN	0, 96	98, 97	0, 995

better than hybridized regression models without using ANNs. LMNN has shown better estimation than BRNN. The correlation between real data and predicted values was more than 90%, and the mean square error was around 0.4. Author's future work is focused on the comparison between this proposed model and another ANN hybridized with a metaheuristic such as GA, Bees algorithm, Ant Colony algorithm, or Artificial Bee Colony algorithm.

The author Badawy [51] has proposed hybrid model for estimation of the cost of residential buildings in Egypt. Real data were used from 174 real residential projects. The proposed model was composed of ANN model and multiple linear regression models. The MAPE of the hybrid model was 10.64% which was less than other hybrid models developed in the research. The analysis has shown that the most important factors in the cost prediction were the number of floors and the area of the floors.

In relation to DSS system, the parameters of the GRNN model can be stored in CPS: minimal and maximal sigma values, validation method (cross-validation with 10-fold), the type of kernel function, and then the parameters of the optimization algorithm of GRNN (number of iterations and absolute and relative convergence tolerance). Also, another recommendation before developing the model with some of the predictive models from the CPS system is to verify the data if the values of the input data have significant differences among them. If this is the case, then normalization of the data can be made before developing the predictive model. Also, the authors believe that in near future software can be developed which can select and try every predictive model from the CPS system and choose the most appropriate for the actual data.
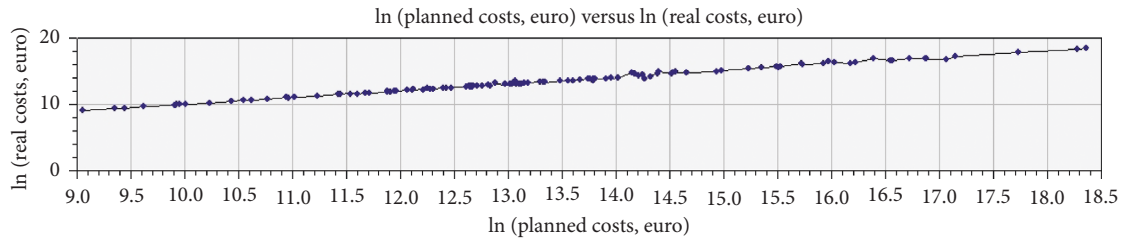
Figure 4: Dependence between the predicted target values (ln (real cost, euro)) and the values of the most important predictor (ln (planned costs, euro)).
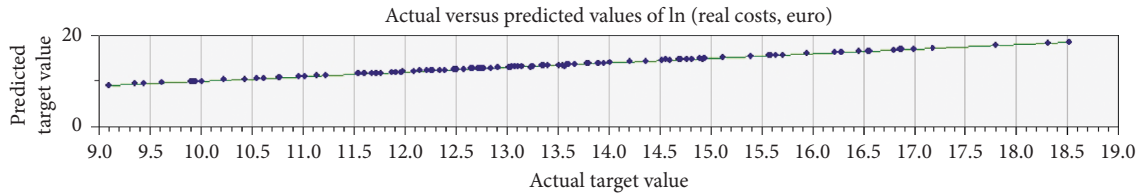

Figure 5: Dependence of the actual and predicted values of the target variable.

No.	Purpose of the facility	ln (planned cost, euro)	Planned costs (euro)	ln (real costs, euro)	Real cost (euro)	ln (planned time, days)	Planned time (days)	ln (real time, days)	Real time (days)
1	Sport	16.72373993	18324324.32	16.85098316	20810810.81	5.899897354	365	5.899897354	365
2	Sport	16.87218537	21256756.76	16.9907451	23932432.43	6.305362462	547.5	7.286191715	1460
3	Education	16.56156799	15581081.08	16.56156799	15581081.08	6.816188085	912.5	7.063048163	1168
4	Education	16.8772583	21364864.86	16.8772583	21364864.86	6.593044534	730	7.152660322	1277.5
5	Education	16.54318672	15297297.3	16.54318672	15297297.3	6.816188085	912.5	7.509335266	1825
6	Sport	17.14771507	28000000	17.19023446	29216216.22	6.08221891	438	6.08221891	438
7	Sport	18.27861886	86756756.76	18.3138176	89864864.86	6.23636959	511	6.23636959	511
8	Sport	17.72753356	50000000	17.8054951	54054054.05	6.162261618	474.5	6.162261618	474.5
9	Education	18.36511089	94594594.59	18.51599092	110000000	5.899897354	365	5.899897354	365
10	Education	14.52208076	2027027.027	14.56130147	2108108.108	5.499419787	244.55	5.499419787	244.55
11	Water tank LISEC	12.40181722	243243.2432	12.50717774	270270.2703	4.791234729	120.45	5.499419787	244.55
12	Water tank lokvarka	13.75994071	945945.9459	13.8934721	1081081.081	5.899897354	365	6.593044534	730
13	Water tank petehovac	13.60579003	810810.8108	13.60579003	810810.8108	5.899897354	365	5.899897354	365

Figure 6: Part of the input data (13 rows, from 116) used for training.

7. Conclusions

Due to the complex cost estimation problems in the field of building construction, lack of data, uncertainty, and risks, especially in the initial phases of the project, the model of the cost prediction system (CPS) as a part of the comprehensive Building Management Information System is proposed. On the one hand, the CPS uses historical data on implemented projects and a database of appropriate parameters, and on the other hand, several models of cost prediction are based on intelligent prediction techniques. These techniques have already been tested in solving various problems of the construction industry. The paper presents CPS ontology with the indicated basic components. The NNs are singled out as especially suitable. The reasons are explained in detail.

The paper analyzes the cost estimation with a concrete database using a hybrid model which is a combination of process-based Bromilow model and data-driven GRNN network using the DTREG software. Accuracy with MAPE of 0.73% was obtained, with coefficient of determination R^2 of 99.5% and correlation coefficient of 0.998. The results were compared with the results obtained using other prognostic models with ANNs, by applying the same software. The presented processing in the proposed model would be enabled through the CPS system components and the stored data would be used. Processing results are stored in the system and used in future processing. Processing results are stored in the Decision Support System and used in future cost estimates and decision-making. The analysis and comparison of partial use of software with those included in

the cost prediction system indicate a significant time saving and an increase in the quality of the assessment in the latter case.

Therefore, the authors find the proposed model as a useful tool for all participants in the construction project for early cost prediction, when numerous factors, which determine cost, are unknown.

Finally, the authors believe that research results, particularly the experience of process-based and data-driven models combination, as well as the proposed CPS model as support for decision-making contribute to the body of knowledge in the field of cost prediction for construction projects.

The development of the proposed cost prediction system should be the subject of future research. A special emphasis in future research should be put on the development of hybrid models. This concept can be applied more widely and can also cover the problem of predicting the duration of construction projects in the early stages.

Data Availability

The authors declare that data supporting the results reported in this paper can be found in the authors' databases. The data are available upon request (contact person: Silvana Petruseva, e-mail: silvana@gf.ukim.edu.mk).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] D. Ahiaga-Dagbui and S. Smith, "Neural networks for modelling the final target cost of water projects," in *Proceedings 28th Annual ARCOM Conference*, pp. 307–316, Association of Researchers in Construction Management, Edinburgh, UK, September 2012.

[2] V. Žujo, D. Car-Pušić, and A. Brkan-Vejzović, "Contracted price overrun as contracted construction time overrun function," *Tehnički Vjesnik : Znanstveno-Stručni Časopis Tehničkih Fakulteta Sveučilišta U Osijeku*, vol. 17, no. 1, pp. 23–29, 2010.

[3] B. Flyvbjerg, M. S. Holm, and S. Buhl, "Underestimating costs in public works projects:error or lie?" *Journal of the American Planning Association*, vol. 68, no. 3, pp. 279–295, 2002.

[4] L. Le-Hoai, Y. D. Lee, and J. Y. Lee, "Delay and cost overruns in Vietnam large construction projects: a comparison with other selected countries," *KSCE Journal of Civil Engineering*, vol. 12, no. 6, pp. 367–377, 2008.

[5] O. S. Alshamrani, "Construction cost prediction model for conventional and sustainable college buildings in North America," *Journal of Taibah University for Science*, vol. 11, no. 2, pp. 315–323, 2017.

[6] C. Petroutsatou, S. Lambropoulos, and J.-P. Pantouvakis, "Road tunnel early cost estimates using multiple regression analysis," *Operational Research*, vol. 6, no. 3, pp. 311–322, 2006.

[7] K. Petroutsatou, E. Georgopoulos, S. Lambropoulos, and J. P. Pantovakis, "Early cost estimating of road tunnel construction using neural networks," *Journal of Construction Engineering and Management*, vol. 138, no. 6, 2012.

[8] M. Juszczyk, A. Leśniak, and K. Zima, "ANN based approach for estimation of construction costs of sports fields," *Complexity*, vol. 2018, Article ID 7952434, 11 pages, 2018.

[9] G.-H. Kim, S.-H. An, and K.-I. Kang, "Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning," *Building and Environment*, vol. 39, no. 10, pp. 1235–1242, 2004.

[10] A. O. Elfaki, S. Alatawi, and E. Abushandi, "Using intelligent techniques in construction project cost estimation: 10-year survey," *Hindawi Publishing Corporation Advances in Civil Engineering*, vol. 2014, Article ID 107926, 11 pages, 2014.

[11] T. P. Williams, "Predicting completed project cost using bidding data," *Construction Management and Economics*, vol. 20, no. 3, pp. 225–235, 2002.

[12] 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.

[13] M. Juszczyk, "The Challenges of nonparametric cost estimation of construction works with the use of artificial intelligence tools," *Procedia Engineering*, vol. 196pp. 415–422, Primosten, Croatia, 2017, http://www.sciencedirect.com.

[14] V. Arabzadeh, S. T. A. Niaki, and V. Arabzadeh, "Construction cost estimation of spherical storage tanks: artificial neural networks and hybrid regression-GA algorithms," *Journal of Industrial Engineering International*, vol. 14, no. 4, pp. 747–756, 2018.

[15] A. Lesniak, "The application of artificial neural networks in indirect cost estimation," *AIP Conference Proceedings, American Institute of Physics*, vol. 1558, no. 1, pp. 1312–1315, 2013.

[16] M. Juszczyk and A. Leśniak, "Modelling construction site cost index based on neural network ensembles," *Symmetry*, vol. 11, no. 3, p. 411, 2019.

[17] L. Le-Hoai and Y. D. Lee, "Time-cost relationships of building construction project in Korea," *Facilities*, vol. 27, no. 13-14, pp. 549–559, 2009.

[18] D. Car-Pušić and M. Radujković, "Construction time-cost model in Croatia," *International Journal for Engineering Modelling*, vol. 22, no. 1–4, pp. 63–70, 2009.

[19] D. Baloi and D. F. Andrew, "Price modelling global risk factors affecting construction cost performance," *International Journal of Project Management*, vol. 21, no. 4, pp. 261–267, 2002.

[20] P. Morris and G. Hough, *The Anatomy of Major Projects*, John Wiley, Hoboken, NJ, USA, 1987.

[21] The World Bank Sustainable Development Department Finance, Economics and Urban Division., 2010 Available from Internet: http://elibrary.worldbank.org/doi/pdf/10.1596/1813-9450-5247.

[22] C. Sun and J. Xu, "Estimation of time for Wenchuan earthquake reconstruction in China," *Journal of Construction Engineering and Management*, vol. 137, no. 3, pp. 179–187, 2011.

[23] R. Nikić, ""Upravljanje rizicima kod građevinskih projekata zemlje u tranziciji", [construction project risk management in a transition country]," Master's thesis, Građevinski fakultet, Sveučilište u Zagrebu, Faculty of Civil Engineering, University of Zagreb, Zagreb, Croatia, 1998.

[24] R. M. Skitmor and S. T. Ng, "Australian project time-cost analysis: statistical analysis of intertemporal trends," *Construction Management and Economics*, vol. 9, no. 5, pp. 455–458, 2001.

[25] M. Radujković, "Upravljanje rizikom i resursima kod građevinskih projekata, znanstveno istraživački project," *MZITRH (Ministarstvo Znanosti I Tehnologije Republike Hrvatske)*, Građevinski Fakultet, Sveučilište U Zagrebu, Faculty of Civil Engineering, University of Zagreb, Zagreb, Croatia, in Croatian, 1999.

[26] V. Žujo, "Doprinos upravljanju građevinskim projektima kroz planiranje vremena" [Contribution of construction time planning to construction project management], PhD thesis, Građevinski fakultet, Univerzitet "Džemal Bijedić", Mostar, Bosnia, 2019.

[27] D. Mačková and R. Bašková, "Applicability of Bromilow´s time-cost model for residential projects in Slovakia," *Selected Scientific Papers-Journal of Civil Engineering*, vol. 9, no. 2, pp. 5–12, 2014.

[28] F. J. Bromilow, "Contract time performance expectations and the reality," *Building Forum*, vol. 1, no. 3, pp. 70–80, 1969.

[29] V. Žujo, D. Car-Pušić, V. Žileska- Pancovska, and M. Ćećez, "Time and cost interdependence in water supply system construction projects," *Technological and Economic Development of Economy*, vol. 21, pp. 1–20, 2015.

[30] W. Hu and X. He, "An innovative time-cost-quality tradeoff modeling of building construction project based on resource allocation," *The Scientific World Journal*, vol. 2014, Article ID 673248, 10 pages, 2014.

[31] I. ElSawy, H. Hosny, and M. Abdel Razek, "A neural network model for construction project site overhead cost estimating in Egypt," *IJCSI International Journal of Computer Science Issues*, vol. 8, no. 3, 2011.

[32] M. S. Gibbs, N. M. Morgan, H. R. Maier et al., "Investigation into the relationship between chlorine decay and water distribution parameters using data driven methods," *Mathematical and Computer Modelling*, vol. 44, no. 5-6, pp. 485–498, 2006.

[33] S. Petruševa, V. Zileska-Pancovska, V. Žujo, and A. Brkan-Vejzović, "Construction costs forecasting: comparison of the accuracy of linear regression and support vectore machine models," *Tehnički Vjesnik*, vol. 24, no. 5, pp. 1431–1438, 2017.

[34] Y. A. Olawale and M. Sun, "Cost and time control of construction projects: inhibiting factors and mitigating measures in practice," *Construction Management and Economics*, vol. 28, no. 5, pp. 509–526, 2010.

[35] K. Ahsan and I. Gunawan, "Analysis of cost and schedule performance of international development projects," *International Journal of Project Management*, vol. 28, no. 1, pp. 68–78, 2010.

[36] Z. Ma, N. Lu, and S. Wu, "Identification and representation of information resources for construction firms," *Advanced Engineering Informatics*, vol. 25, no. 4, pp. 612–624, 2011.

[37] A. Watson, "Digital buildings-challenges and opportunities," *Advanced Engineering Informatics*, vol. 25, no. 4, pp. 573–581, 2011.

[38] J. Egan, "Rethinking Construction," *The Report of the Construction Task Force*, HMSO, Richmond, VA, UK, 1998.

[39] R. J. Scherer and S. E. Schapke, "A distributed multi-model-based Management Information System for simulation and decision-making on construction projects," *Advanced Engineering Informatics*, vol. 25, no. 4, pp. 582–599, 2001.

[40] S. Petruseva, "Neural networks and their application in civil engineering. Isothreshold adaptive network (IAN)," *Scientific Journal of Civil Engineering, (SJCE)*, vol. 2, no. I1, 2013.

[41] A. Alqahtani and A. Whyte, "Artificial neural networks incorporating cost significant Items towards enhancing estimation for (life-cycle) costing of construction projects," *Construction Economics and Building*, vol. 13, no. 3, pp. 51–64, 2013.

[42] I. Flood, "Towards the next generation of artificial neural networks for civil engineering," *Advanced Engineering Informatics*, vol. 22, no. 1, pp. 4–14, 2008.

[43] D. F. Specht, "A general regression neural network," *IEEE Transactions on Neural Networks*, vol. 2, no. 6, pp. 568–576, 1991.

[44] E. Parzen, "On estimation of a probability density function and mode," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.

[45] D. F. Specht, "GRNN with double clustering," in *Proceedings of the International Joint Conference on Neural Networks*, Vancouver, Canada, July 2006.

[46] P. Sherrod, (2013a). DTREG Predictive Modeling Software – Tutorial. Retrieved from http://www.dtreg.com, accessed june 7, 2016..

[47] P. Sherrod, (2013b). Predictive Modelling Software. Retrieved from http://www.dtreg.com, accessed june 7, 2016.

[48] E. W. M. Lee, Y. Y. Lee, C. P. Lim, and C. Y. Tang, "Application of a noisy data classification technique to determine the occurrence of flashover in compartment fires," *Advanced Engineering Informatics*, vol. 20, no. 2, pp. 213–222, 2006.

[49] X. Wang, M. You, Z. Mao, and P. Yuan, "Tree-structure ensemble general regression neural networks applied to predict the molten steel temperature in ladle furnace," *Advanced Engineering Informatics*, vol. 30, no. 3, pp. 368–375, 2016.

[50] V. T. Luu and S. Y. Kim, "Neural network model for construction cost prediction of apartment projects in Vietnam," *Korean Journal of Construction Engineering and Management*, vol. 10, no. 10, pp. 139–147, 2009.

[51] M. Badawy, "A hybrid approach for a cost estimate of residential buildings in Egypt at the early stage," *Asian Journal of Civil Engineering*, vol. 21, no. 5, pp. 763–774, 2020.