TECHNICAL PAPER

Machine learning-integrated 5D BIM informatics: building materials costs data classification and prototype development

Saeed Banihashemi[1] · Saeed Khalili[2] · Moslem Sheikhkhoshkar[3] · Abdulwahed Fazeli[3]

Abstract

Non-informatics cost estimation is a tedious process and requires substantial amount of time and manual operations. However, BIM adoption approaches have attracted significant attention with this respect. Since BIM models are object-based with built-in parametric information, it is easier to capture the quantities of building elements and deliver more accurate estimates with less errors and omissions. As most of the current cost estimation standards are designed and developed based on old-fashioned construction project delivery systems, a lack of compatibility between their classification and BIM-based informatics is observed. This study, therefore, aims to develop an informatics framework to integrate a cost estimation standard with BIM in order to expedite the 5D BIM process and enhance the digital transformation practices in construction projects. The developed framework is considered to be a new approach which can automatically estimate the cost of building elements using machine learning-integrated algorithms and MATLAB engine for its effective implementation.

Keywords Building information modelling · 5D BIM · Machine learning · Cost estimation standard

Introduction

One of the critical and vital processes in construction projects is a precise cost estimate. As the project becomes larger and more complex, the more accurate estimate of the cost of a project, over time, becomes a major factor in the project success [1]. Cost estimation refers to the process of calculating and projecting the costs of three main parts including all building materials, labors and equipment within the scope of the project [2, 3]. It involves performing quantity takeoff (QTO), which is a list of quantities of the items and materials needed for a project.

Cost items contain the quantity and unit price of particular resources, connected with one or several tasks [4]. The cost of each task is estimated by multiplying the required quantities with their unit cost items and the total cost is, then, the sum of corresponding cost items [5, 6]. Since the cost estimation is considered as one of the main parts of the projects and decision-making at the conceptual stage, its accuracy plays a key role in the success and quality of the project [7, 8], while cost overrun is the major problem, particularly in projects with tight budgets. Indeed, the cost overrun can trigger the failure of projects [9] in developing countries like Iran where its economy suffers from huge inflation and uncertainty [10–12]. On the other hand, the cost estimation process is still labor-intensive and error-prone [13]. In fact, estimators should manually take off the information from printed CAD drawings and 2D-based documents. This process is not only error prone and time consuming but also it is heavily relied on the estimators' knowledge and experience. So, it makes it difficult to process in large and complex projects [14–16]. Another challenging task is when estimators may need to identify different design conditions in their projects [17]. Especially, if the design changes, estimators should manually determine affected costs and adjust tasks and resources based on the updated project design. This is mainly error

✉ Saeed Banihashemi
    saeed.banihashemi@canberra.edu.au

    Saeed Khalili
    saeid.khalili@bimsolutions.ir

    Moslem Sheikhkhoshkar
    msheikhkhoshkar@iibimsolutions.ir

    Abdulwahed Fazeli
    ab_fazeli@alumni.iust.ac.ir

1   University of Canberra, Canberra, Australia

2   Bana Yar Mehraz Iranian Company, Tehran, Iran

3   Iran University of Science and Technology, Tehran, Iran

prone and leads to inconsistencies and inefficiencies of cost estimation process [18]. Therefore, there is still a need for practical frameworks and prototypes to accurately and precisely classify building cost informatics, estimate the total cost of the project and automate its estimation process [2].

The advent of the BIM can resolve many of the problems caused by the use of traditional estimation methods [19, 20]. BIM offers a powerful solution to consistently produce take-offs, counts and measurements through the updated model [6, 21]. As opposed to traditional methods, all of the cost estimation data can be exchanged through different software with the help of collaborative platforms and open data exchanges standard for BIM such as Industry Foundation Class (IFC) [22, 23]. This information exchange can significantly reduce the amount of workloads and human-oriented errors [24, 25]. Furthermore, BIM is a digital representation of physical and functional traits of facilities. The product centric and object-oriented information of the BIM model is enriched with parametric building objects as properties [26, 27]. All of these information can be extracted from BIM models for future analyses and so, the BIM model can have the ability to assist decision making in different areas of construction [14, 28, 29]. Ultimately, Davidson [30] indicates that BIM has the ability to takeoff the cost analysis at any stage of a project lifecycle.

Since the traditional BIM-based cost estimation requires human interference and there is no standardized cost specification for being connected with each building element, the BIM-based cost estimation is still a time-consuming, error-prone process and suffers from a low return on investment [31, 32]. Therefore, this study aims to eliminate the human intervention by developing a machine learning and 5D BIM-integrated cost estimation prototype which classifies building cost elements, links with the 5D BIM process and enhances the adaptation of the cost classification systems (Iranian Cost Estimation Standard, in this case) with BIM-based informatics. This research also extends a new approach to the automated cost estimation process through applying a machine learning-integrated method. In which, first, an informatics prototype is built, so that all construction cost items can be classified, collected, developed and transferred to the Excel software. Second, the database of BIM is connected and updated through the Iranian Cost Estimation Standard (ICES) [33], and finally, the developed application extracts QTO from each item-cost and estimates the cost of the whole project. Furthermore, in order to automate the process, a suite of machine learning; Artificial Neural Network (ANN) algorithms including Multilayer Perceptron (MLP), Radial Basis Function (RBF) and Adaptive Neuro Fuzzy Inference System (ANFIS) is developed and applied via setting the parameters of the project as the inputs and the result of costs as the outputs.

Literature review

There are many studies which emphasize on and present the BIM-based cost estimation process. Balali, Zalavadia [34] proposed a virtual reality (VR)-based cost estimation approach for improving the ability of design changes and cost updates during preliminary phase of the project. However, the main approach only represents elemental costs of the project and it fails to estimate the cost of labors or equipment in the project. Bademosi, Tayeh [35] also developed a framework that utilizes VR as a method of visualization in order to improve the efficiency of cost estimation process and concluded that using VR can significantly reduce the time of estimation process as well as streamline the process of interactive estimates for the project. proposed a knowledge-based approach for automating the process of code-compliant QTO using the combinations of standard method of measurement rules and semantic information. Despite the effectiveness of the proposed method, the main approach still needs human intervention for extracting the QTO-related information which is prone to errors and faults. Valentini, Mirarchi [36] compared two different methods of cost estimation including traditional and BIM-based methods and found that using BIM can both increase the precision and reduce the cost of the process. developed a BIM-based system for analyzing incremental costs of prefabricated projects in China. This two-dimensional incremental cost index system was validated in a real project in the Shanghai, China. developed a BIM based method for estimating the cost of steel frames in the early phase of building projects. The proposed integrated steel design approach can estimate the whole costs of steel frames including materials, fabrications and erection. However, the main framework can be only limited to steel projects while there are many other types of projects like wood and concrete projects.

Akanbi and Zhang [37] developed a semantic natural language processing-based approach which can automate the process of information extraction and matching in the main process of cost estimation. Khosakitchalert, Yabuki [38] proposed an automatic approach for improving the accuracy of QTO in the BIM model for compound elements. The proposed method lays the groundwork for more optimized cost estimation as many overlapping regions are eliminated in this process. Wang, Wang [39] developed a new framework for cost estimation which combines project schedules from work breakdown structure (WBS) and cost information from cost breakdown structure (CBS) in the BIM area. Firstly, all quantities for one cost item in the BIM are taken off; then, the pop-up menus are represented for estimating the cost of that item. This utilizes keynotes, assembly codes and family types in the BIM and the user

can manually explore the names items and gain the cost of the desired items. However, despite the practicality of the proposed method, the model has less user-friendly interface which complicates the application of the model. In addition, the user must have sufficient knowledge of BIM for efficient performance of the model. Wu, Wang [40] developed a new cost estimation framework which enables the designers to estimate the cost of structural elements in the preliminary stage of the project by linking ETABS and BIM model. The proposed method is, then, examined through a reinforced concrete building and its compliance with Taiwanese architectural laws and guidelines. Ma, Liu [13] developed an ontology-based framework that streamlines the adoption of the formulized representation of cost specifications as to improving the efficiency of cost specifications implementation in computer programs. The main advantage of this study is its capability to adapt to the other cost specifications based on the bill of quantity method. Ma and Liu [41] proposed a new approach to identify and derive the construction information in order to increase the automation of cost estimation for tendering of building projects based on the BIM-based design. Joosung and Jaejun [42] developed a 4D simulation framework which can seek out more efficient management in modular construction projects. According to this methodology, projects can benefit from the reduction in time and cost required for module yields. Zanni, Sharpe [43] examined the process of integration between the whole lifecycle cost and BIM. The method only concentrates on the emerging private rental sector since the built-to-rent market is repeatable task and similar workflow pattern. The proposed model has the potential of improving BIM lifecycle management via an augmented decision-making approach which is integral to the natural design development process. Fazeli, Dashti [44] developed a BIM-based cost estimation method which can link Iranian cost specifications with MasterFormat and UniFormat.

There are also other studies which focus on the application of machine learning in the process of cost estimation. Jiang [45] proposed the deployment of ANN in cost estimation and compared the results with RBF neural networks. The results indicate the higher performance of ANN in comparison with RBF neural networks. To optimize the accuracy of the model, the paper also examined other types of projects. Huang and Hsieh [46] proposed a BIM-based method for prediction of labor costs using the combination of random forest and simple linear regression in construction projects. The proposed method was then tested through nineteen real projects in Taiwan. Chandanshive and Kambekar [47] examined the potential of multilayer feed forward ANN along with a backpropagation learning algorithm for cost estimation in 78 building projects in India. To prevent overfitting of the model and improvement in the generalization

competency of neural networks, the paper also used early stopping and Bayesian regularization approaches. Elghaish, Abrishami [48] developed an approach for the process of cost estimation using the combinations of 5D BIM and integrated project delivery. This study also used Monte Carlo simulation, target value design and activity-based costing in the main process of cost estimation which can make the process unique. Bala, Ahmad Bustani [49] used backpropagation ANN for predicting the cost of building projects in Nigeria. Despite the effectiveness of the model, the study only takes the institutional type of projects into account and there is a void for estimating other type of projects. Cheng, Tsai [50] applied the hybrid of ANN with fuzzy logic as a tool for reducing uncertainties in the process of cost estimation. The results confirmed the effectiveness of the model combination in comparison with only ANN approach for predicting the cost of projects in early phase of the project.

With regards to the literature review, the following limitations still exist in the previous studies. Firstly, although the previous approaches aim to propose a new method for cost estimation, few of them apply standardized resource specifications such as ICES items. ICES not only considers the cost of materials, but it also takes into account the cost of labors and equipment which can make the process more accurate and precise along with the full rich of standardized items which can be updated annually. Furthermore, the holistic view on the literature indicates that there is still a room of improvement for BIM-based cost informatics classification and estimation processes, particularly through the application of machine learning tools and their proper integration with the national cost guidelines to increase the digital transformation uptake in the construction industry. Finally, few studies try to classify different building elements in the project and integrate the classified system with machine learning algorithms to make the process more automated and deliver superior results.

Methodology

This study aims to develop a machine learning-integrated 5D-BIM informatics prototype for estimating the cost of buildings and validating the process through using ICES. Figure 1 demonstrates the research design of this work in which the process starts with a quantitative analysis and the workflow design, so that the BIM-based cost estimation is produced. Then, machine learning algorithms are developed through the simulation method involving several calculations and formulas in the MATLAB engine. Through the simulation, researchers are enabled to test the influences of causes and effects on various variables and likelihood reactions of different scenarios [51]. In this phase, a workflow should be developed to integrate machine learning with BIM. The
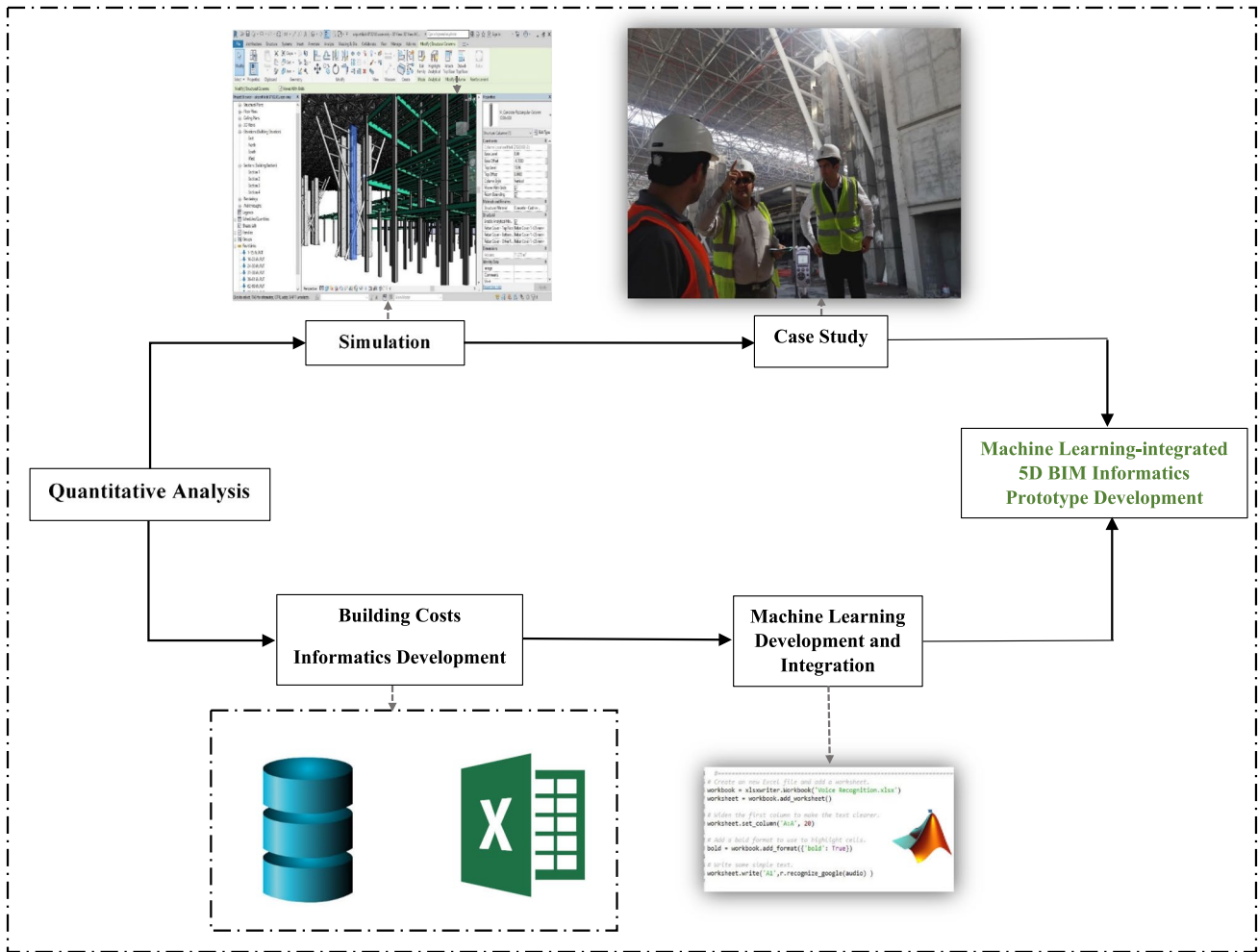
Fig. 1 Research design

algorithms are then implemented through training datasets by machine learning, and the automated cost estimation process is generated. According to Hancock and Algozzine [52], a viable case study demonstrates the values and performance of the workflow, models or algorithms implemented within a real-life setting. The algorithms are, hence, performed on a practical case study for training machine learning and developing an automated cost estimation through BIM. Finally, the control dataset is utilized to test the process and find the optimum performance for machine learning algorithms.

The practical steps for this research design include 6 main technical phases. Firstly, a database of ICES items is created in Excel and automatically imported into BIM authoring tool; Revit to align the materials information along with their extra data. After that, the user should create elements based on the generated Iranian materials and the system will then estimate the cost of the building through ANN and show the results in both BIM and Excel in the form of reports. However, one of the main challenging tasks for the

proposed method is to adjust the elements based on ICES items. Therefore, this study also implements machine learning for facilitating the process of cost estimation of every element by using MATLAB functions. By doing so, the user can estimate cost of the project in the shorter time. So, a BIM-based extension with six major functions is required to be developed in Revit to facilitate and automate the cost estimation process. The development of the proposed approach, as represented in Fig. 2, is explained through the following six steps.

Phase 1: developing the library of building materials

Since Revit software has relatively modest number of building components, a comprehensive library should be created by designers to develop a BIM model with high details of information and components. This library of components enables designers to develop a BIM model with full details.
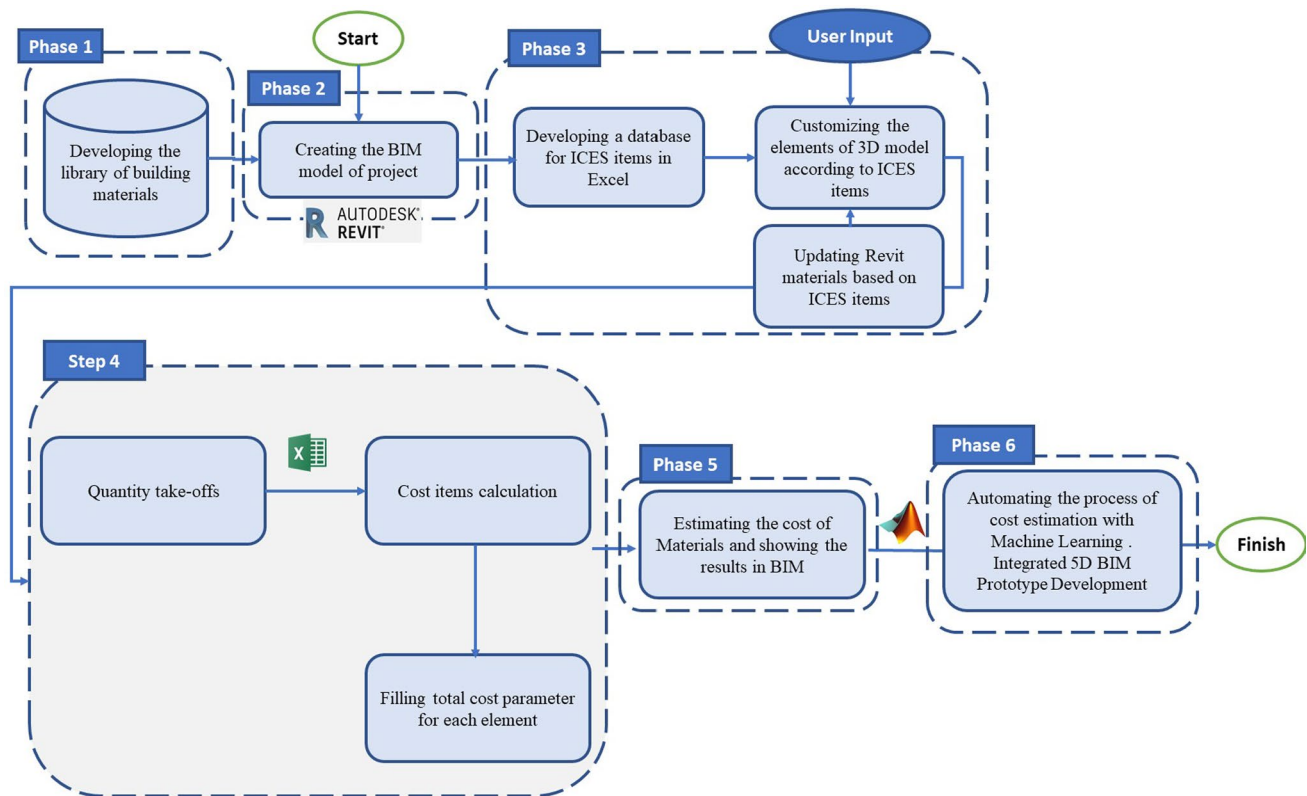
Fig. 2 5D BIM prototype development

It is also imperative for a quick and accurate cost estimation according to their price list. The designer can define different scenarios for the project by changing the different components (see "Appendix 1") and it will help stakeholders for efficient future decision in terms of cost saving. The components of a project are provided in the formats of.rvt and.rfa. So, the designer can download them from formal webpages such as SmartBIM or ArCat libraries and import them in the Revit file projects.

Phase 2: creating the BIM model of project

3D model is considered as one of the main benefits of BIM. This evident merit can greatly assist all practitioners of projects to detect clashes of building and include data for each component. This information is needed for an accurate cost estimation of the project, and therefore, a 3D model should be created for capturing all quantities of objects.

Phase 3: developing a database for ICES items in Excel

For estimating the cost of a project in Iran, it is needed to apply Iranian cost items in the BIM model. For doing so, a database of ICES items is created in the Excel environment

in this phase. This database includes all information of ICES items including Name, ID, ICES Description and unit costs.

The ICES database includes 909 items which is mainly used for the cost estimation of structural and architectural elements of projects. For better understanding about the process, the process is described in the form of a sample wall. Figure 3 shows items of ICES for this sample Wall. Table 1 also indicates all Excel information of ICES items of this sample wall which is extracted from the main database. The first column is the name of ICES items and the second, third, fourth and fifth columns are ICES items, unit cost, formulas and description, respectively.

Updating Revit materials based on ICES items

In this phase, a plugin is created to import all created ICES items in the BIM environment. By activating this plugin, Revit calls the Excel file, the information of ICES are automatically imported into Revit and the respected table is updated. So, 909 items of ICES were added to the Revit material list. Figure 4 shows the updated table of Revit after implementing the created plugin. As shown in Fig. 4, the name of materials, cost parameters and unit costs are corresponded to their ICES names and Keynotes of materials in Revit.
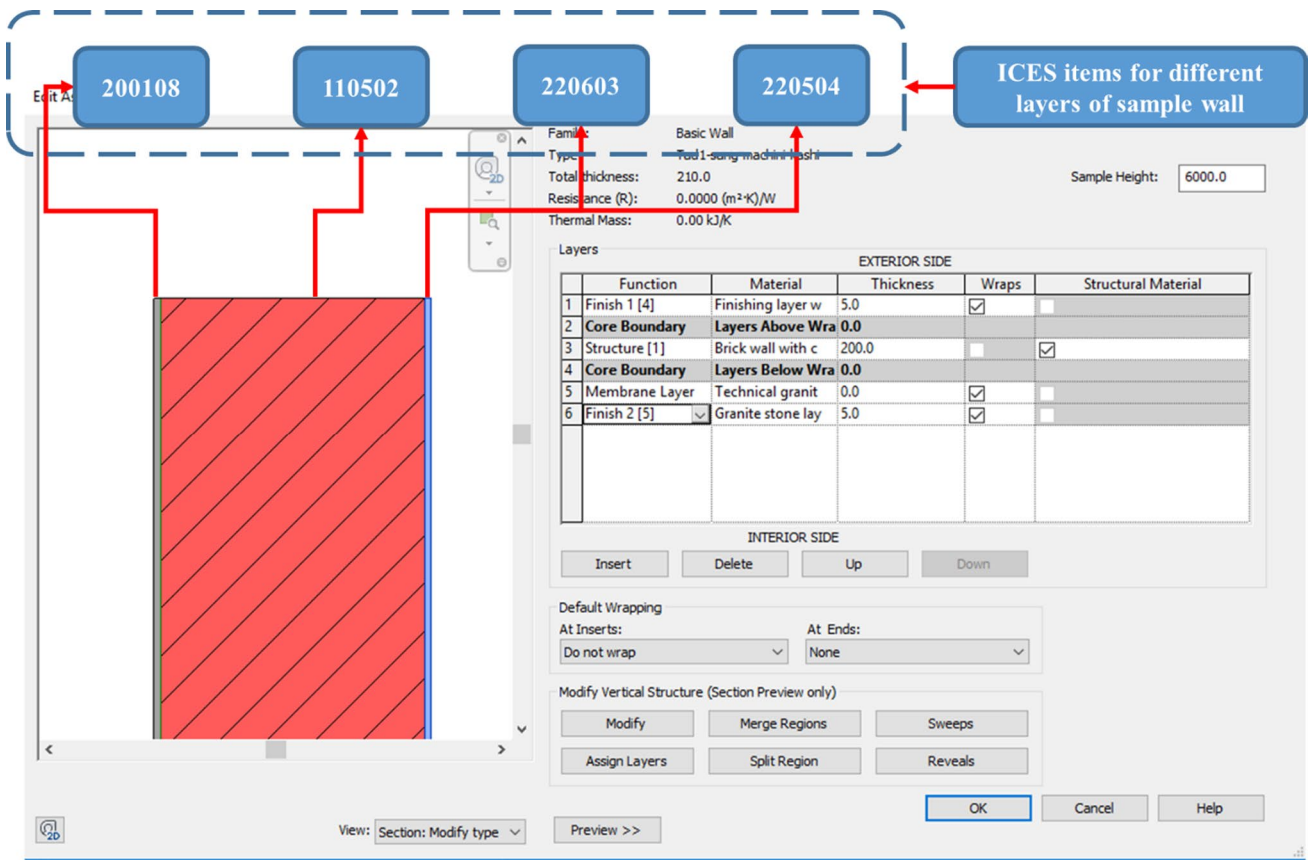
Fig. 3 Different layers for a sample wall

Table 1 ICES items for the sample wall

Name	ICES item	Unit cost	Formula
Finishing layer with glazed tile	200,108	399,500	Area
Brick wall with cement to sand ratio mortar of 1.5	110,502	425,000	Area
Technical granite stone provision (Extra cost for granite stone layer)	220,603	73,000	Area
Granite stone layer on the vertical surface	220,504	674,500	Area

Customizing the elements of 3D model according to ICES items

By compiling the database of ICES items in Revit, the user updates the elements based on ICES items. For doing so, the user must update the layers of components according to ICES items. Figure 5 shows how the user can add ICES items in the sample wall. As it can be shown in this Fig. 5, the wall has four layers and each layer is chosen from the ICES item database.

Phase 4: quantity takeoffs

In this phase, all quantities of materials are taken-off from Revit files and sent into the Excel. So, all required data for cost estimation such as cost items and unit costs in the ICES along with the geometric information such as area, volume and length of building elements are exported and categorized based on BIM parameters (Fig. 6).
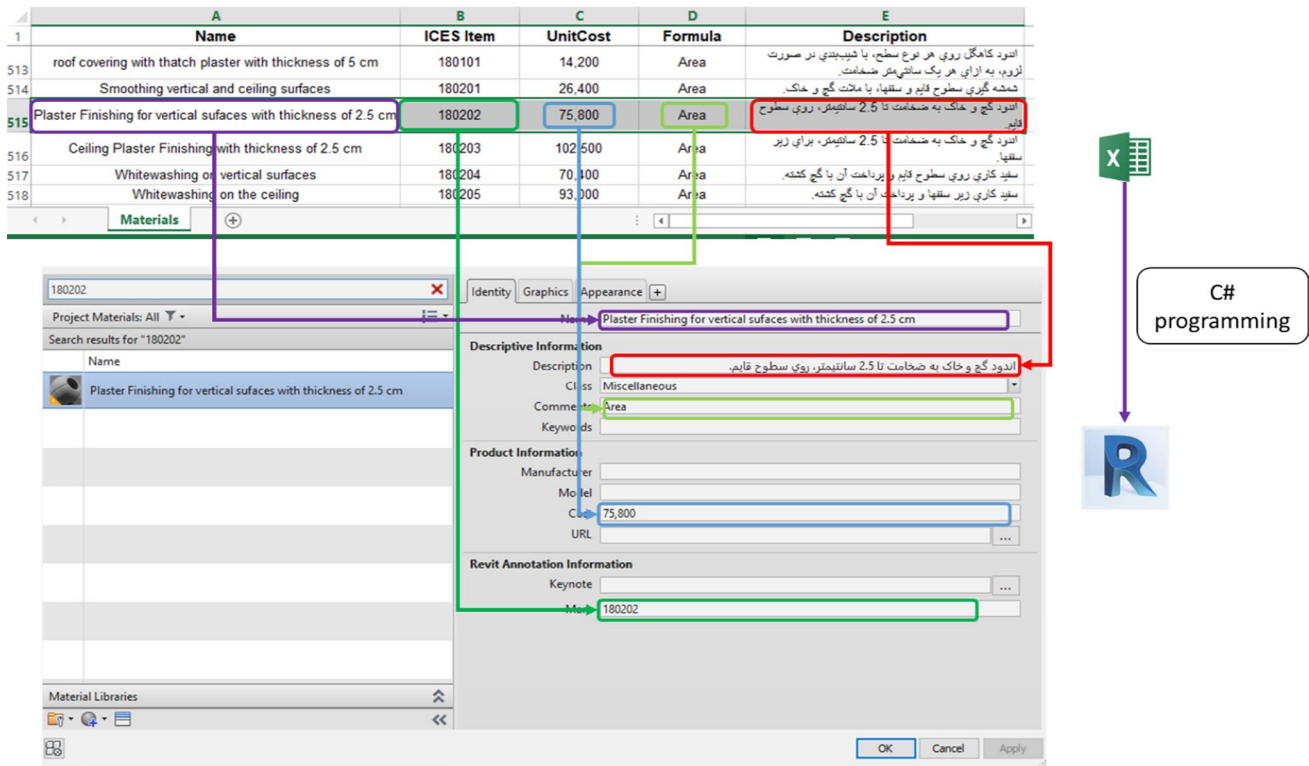
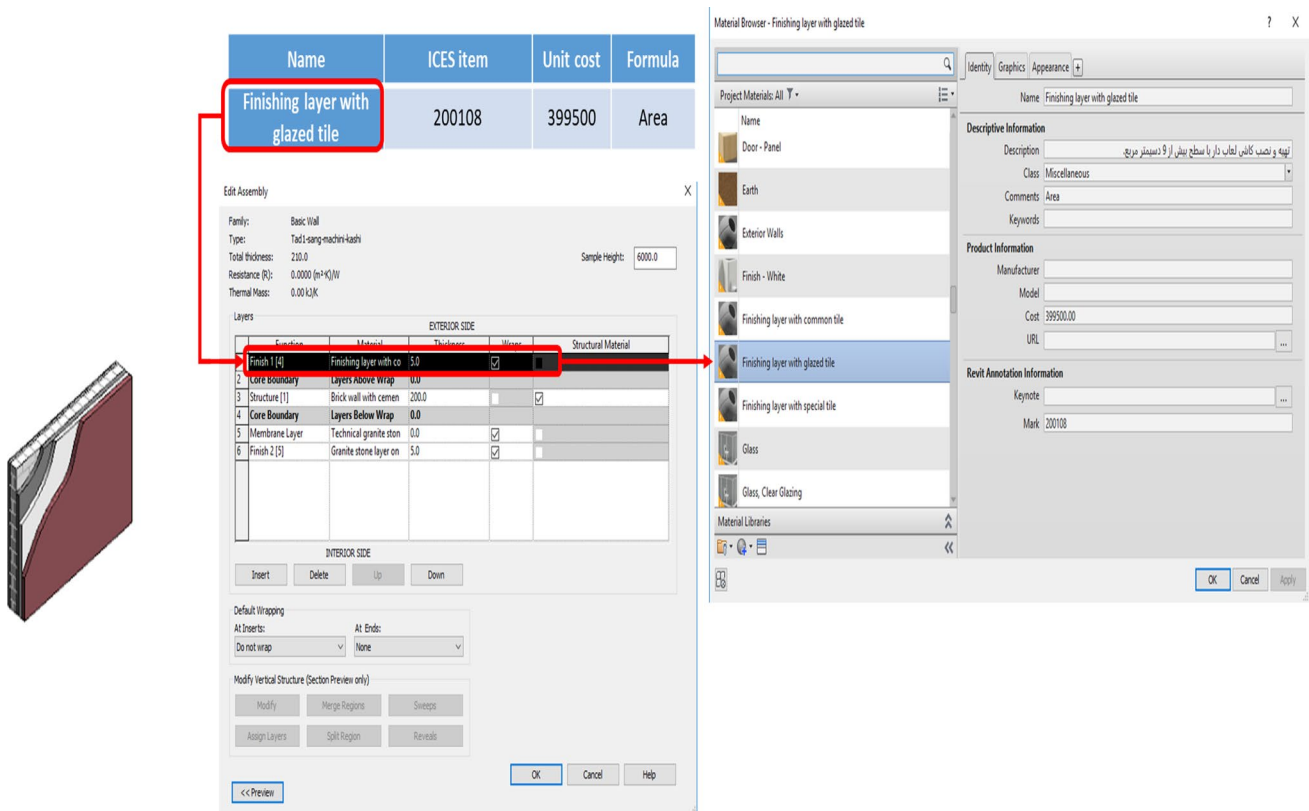**Fig. 4** Updated database of ICES items in Revit materials

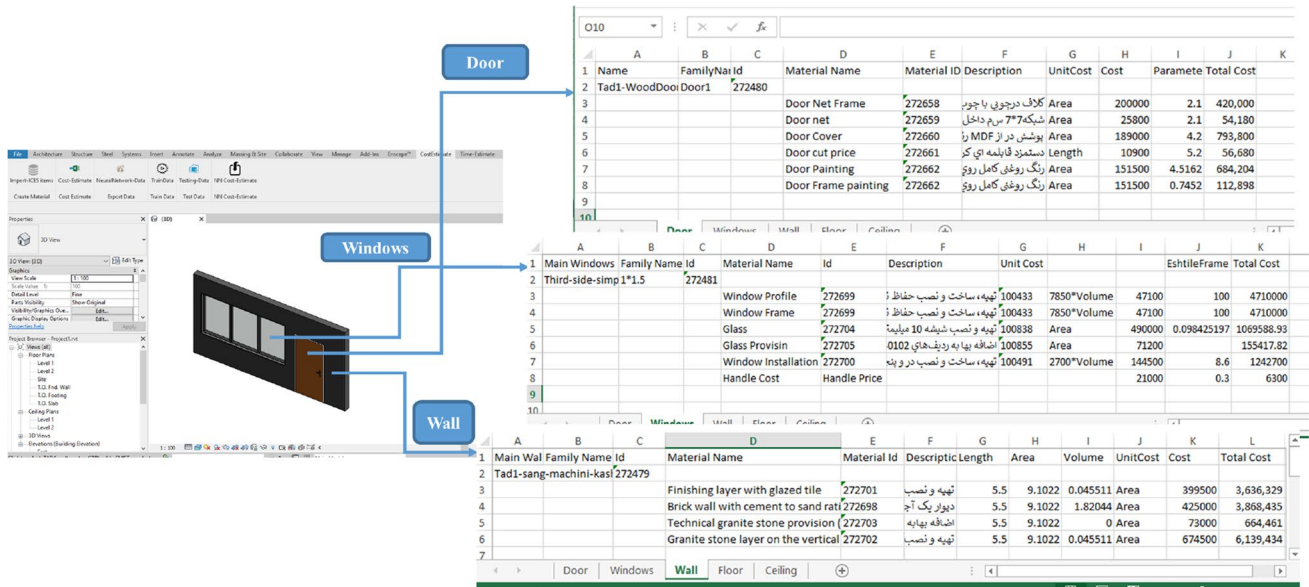**Fig. 5** Customizing the elements of 3D model according to ICES items

Fig. 6 Cost estimation process and reporting to Excel file

Cost items calculation

In this step, all required calculations are performed by multiplying all related ICES cost items, corresponding ICES unit costs and their appropriate quantities of components. There are different units in the ICES items including square meter, meter, cubic meter and kilogram. It should be considered that all of these processes were done before by using formula instead of units of ICES items (Table 1). Figure 6 demonstrates a cost estimation for the sample wall along with their doors and windows. As can be seen in Figs. 4, 6, categories are estimated including floor, wall, door and windows and for each of them, a separated sheet is created and the results of each category and costs are shown in the Excel.

Filling total cost parameter for each element

In this step, the generated total cost parameter is filled by the function of "Cost-Estimate" in the developed extension. For doing so, first, all cost information of elements is summed up in the Excel. The resulted cost data are then set in the total cost parameter for every component. This parameter is needed for any future cost prediction using ANN algorithms. These data would be considered as the output data for machine learning algorithms and would be exported to MATLAB for getting run (Fig. 7).

Phase 5: estimating the cost of materials and showing the results in BIM

In the phase 5, the total cost of different categories of building components is shown in the Revit environment using the "Cost-Estimate" function. As being explained in the former steps, all data of elements are sorted according to the Revit parametric elements and so, the cost figures are summed up to deliver the total cost estimate (Fig. 8). The framework also exports the whole reports of the BIM-based cost estimation to the excel software. This report can help cost estimators to show all stakeholders of the project the details of cost estimation. With using this report, cost estimators can easily change their desired ICES items for each element using the associated ID and if there is any error including wrong ICES items or wrong type of elements, they can easily fix that problem. In fact, the main objective of this step is to reduce errors resulting from staff interventions during selecting ICES items in the phase 3 of the study.

Phase 6: automating the process of cost estimation with machine learning-integrated 5D BIM prototype development

The main objective of this phase is to develop an algorithm for facilitating the process of cost estimation of each element in the project using the neural network for training data. Since the process of selecting ICES items for different elements in the project was time-consuming and error-prone, this step can help all stakeholders of construction projects to automatically estimate the cost of elements. In addition, for understanding the performance of different algorithms of neural network, three versions of neural networks are examined including MLP, RBF and ANFIS. Figure 9 shows
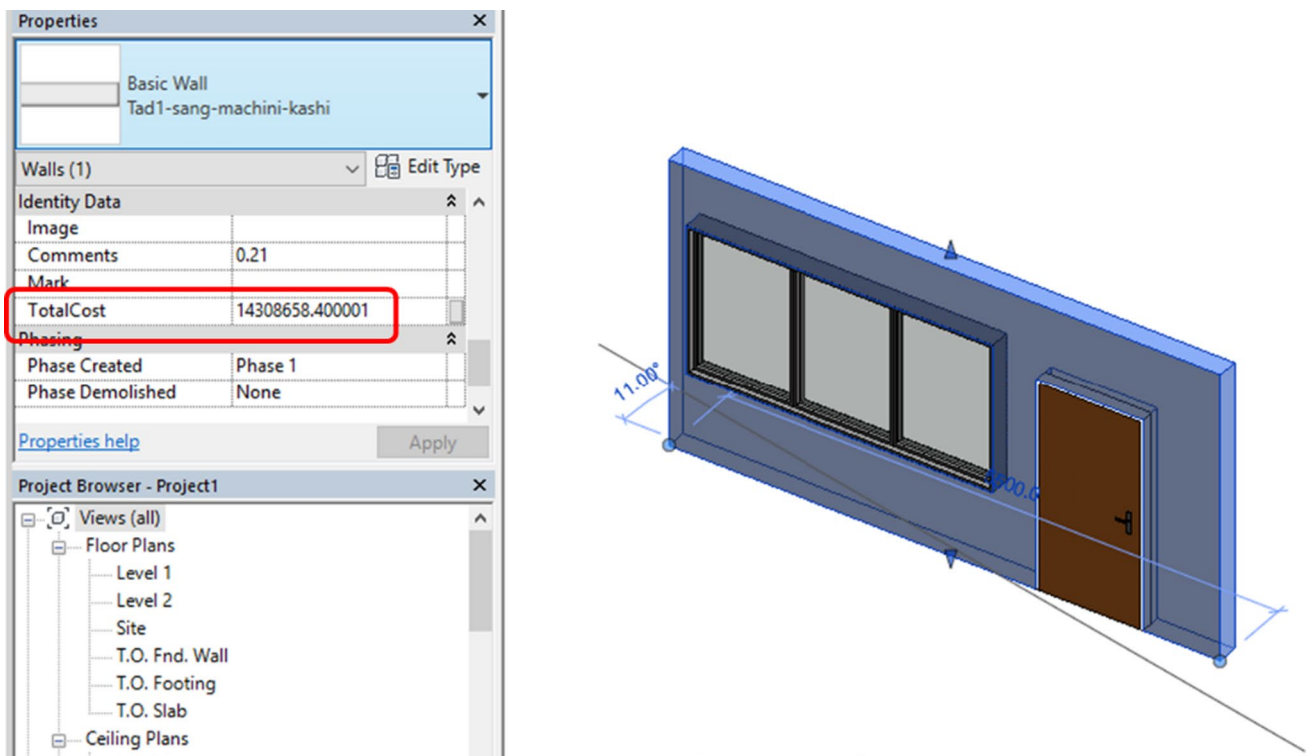
Fig. 7 The resulted total cost for each component in Revit

the overview of implementation of the neural network algorithms in MATLAB.

Step 1: exporting effective criteria and total cost in Excel (input and output data)

The main challenge for this step is to identify main effective criteria which considerably impact the cost of compound elements. For doing this, the digitization technique was applied and the components were corresponded to detect significant criteria of the total cost of elements. Figure 10 demonstrates a sample wall number 163. As it can be shown, the first number (1) is gypsum wall which is located as the below finishing layer of wall, the second number (6) represents engineering brick which is the structural layer of wall and the third number (3) demonstrates the trowel finish concrete wall which is laid in the upper layer of wall finishing. The similar method is applied for all other building components, parameters and cost elements.

The total cost parameter is also exported from Revit into Excel. This parameter is the output for neural network algorithms. All outputs are exported into the Excel using "NeuralNetwork-Data" function in the extension.

Step 2: data call by MATLAB

MATLAB is selected in this study because it benefits from the powerful and flexible engine and is linked well with BIM models [53]. By having the data in Excel, another plugin is created which can transfer them to MATLAB, implement neural network algorithms and finally show the results in the BIM. Hence, the implementation of the ANN algorithms encompassing MLP, RBF and ANFIS are explained in the following three steps.

MLP neural networks implementation in MATLAB The MLP network is known to be a prevalent and one of the most prominent networks applied in neural networks. By MLP, the bias terms and weighted sum of the inputs pass to an activation level by a transfer function in order to generate an output. The units are arranged in a feed-forward layered topology coined as the feed forward neural network. MLP has some features including any number of inputs, hidden layers with any number of units, linear combination functions in the input layers, activation functions in the hidden layers and any number of outputs with any activation function [54]. Each layer involves a set of neurons and nodes. The input and output layers are linked by a hidden layer.
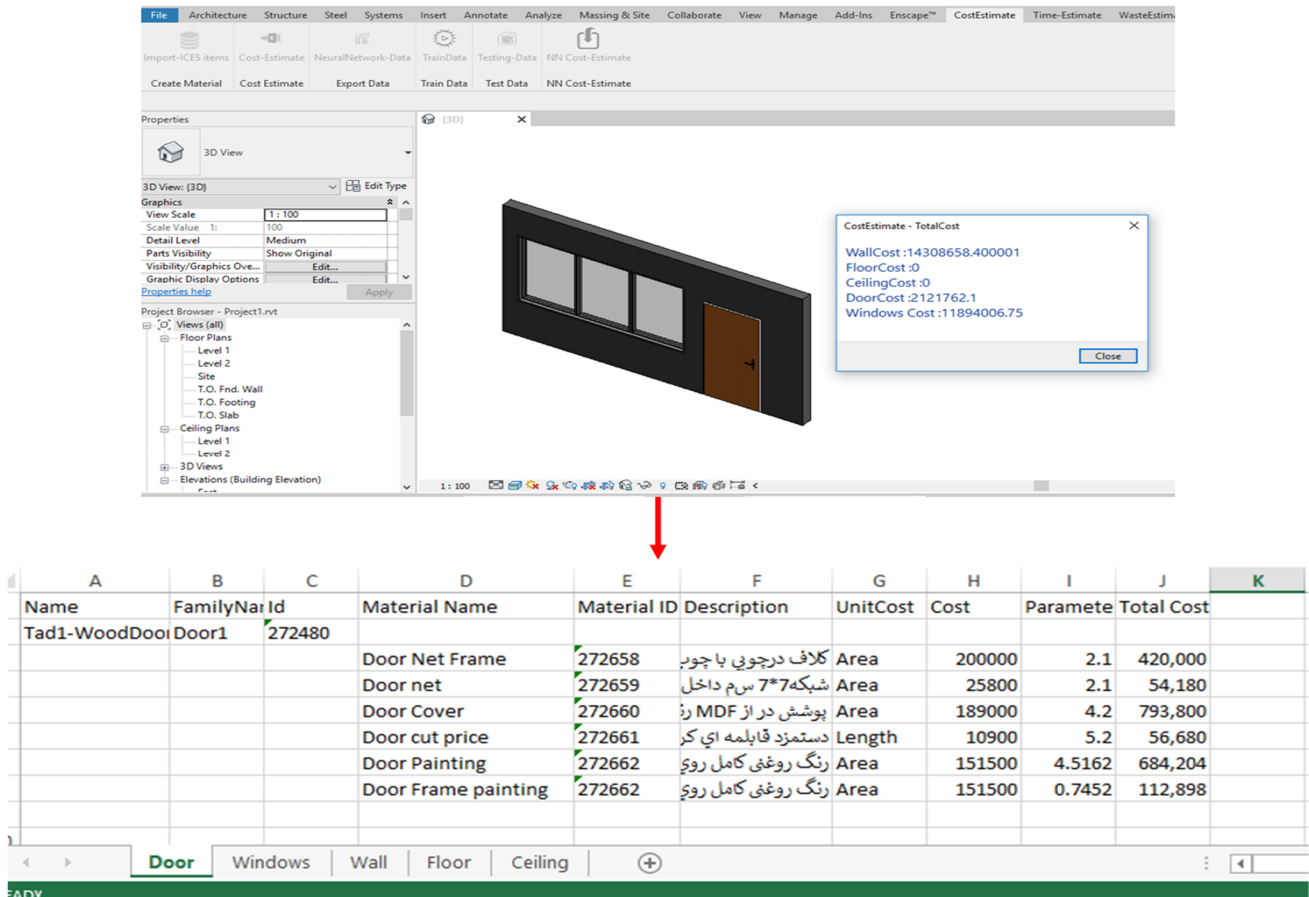
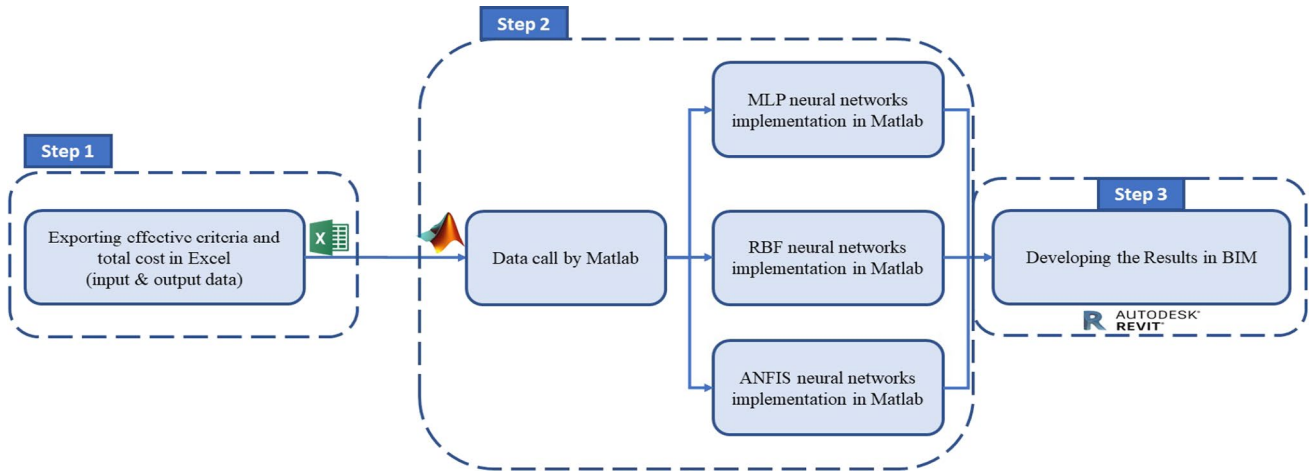**Fig. 8** Total cost representation

Fig. 9 Machine learning implementation process in MATLAB

All connections between the input layers and hidden layers contain the input weight matrix and all connections between hidden layers and output layers encompass the output weight matrix [55]. Each neuron, by computing the weighted sum of its input (x_j, for $j = 1,2 ,…,n$), utilizes a nonlinear activation function in order to achieve the output signal of y as Eq. 1:
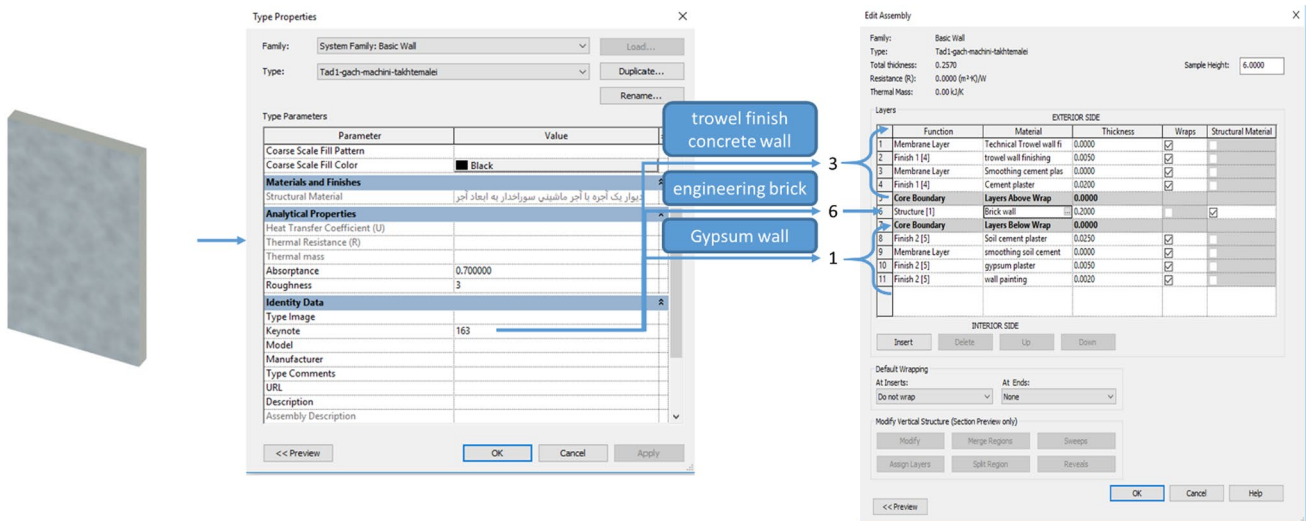
Fig. 10 Classification of compound elements in Revit

$$y = f\left(\sum_{j=1}^{n} w_i x_j\right) \tag{1}$$

where f is the sigmoid activation function which is used in most feed-forward neural network applications. The function is demonstrated as Eq. 2:

$$\text{sgm}(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

After clicking on the "TrainData" function in the developed extension, Revit asks from users to choose which algorithm should be implemented. By choosing MLP algorithm function, the data are transferred and processed in MATLAB and MATLAB sends back the processed data to Revit. Figure 11 shows how this back-and-forth application works.

RBF neural networks implementation in MATLAB RBF neural networks as the pattern recognition functions were proposed by several researchers to develop training methods and increase the precision of the algorithm in comparison with other neural networks [55]. The RBF networks is well working at modelling nonlinear function approximations. One-stage-training can achieve a good result rather than applying a repetitive process in MLP, as well as early learning of the application [56]. It is also used in the data classification, system modelling and control [55]. It includes one input layer, one hidden layer with a nonlinear RBF activation function along with one linear output layer. Each RBF has two key parameters, representing the location of the functions center and its deviation or width. The hidden unit calculates the gap between an input data and the center of its RBF. There also exists a single hidden layer and two sets of weights, connecting hidden layers with input and output layers. The weights which connect hidden layer with input layer include the parameters of basis functions. The weights are applied to constitute linear combinations of the activations of the basis functions (hidden units) to produce the network output [54]. The output of the network is as Eq. 3:

$$y_k(x) = \sum_{J=1}^{M} W_{KJ} \phi_j(x) + W_{KO} \tag{3}$$

where M is the number of basic functions, x is the input data vector, w_{kj} represents a weighted connection between the basic function and output layer, and ϕ_j is the nonlinear function of the jth unit, which is typically a Gaussian function (Eq. 4):

$$\phi_j(x) = \exp\left(-\frac{\left|x - \mu_j\right|^2}{2\sigma_j^2}\right) \tag{4}$$

where x and μ are the input and the center of RBF unit, respectively, and σ_j is the spread of the Gaussian basis function.

By clicking on the "TrainData" function in the developed extension, the same process as MLP gets activated.

ANFIS neural networks implementation in MATLAB ANFIS is a combination of neural networks and fuzzy systems that contain the benefits of both developed by [57, 58]. Its structure is based on the feed forward neural network system where each layer is formed from the neuro-fuzzy system components [55]. For facilitating the learning and adaptation process, the ANFIS performs
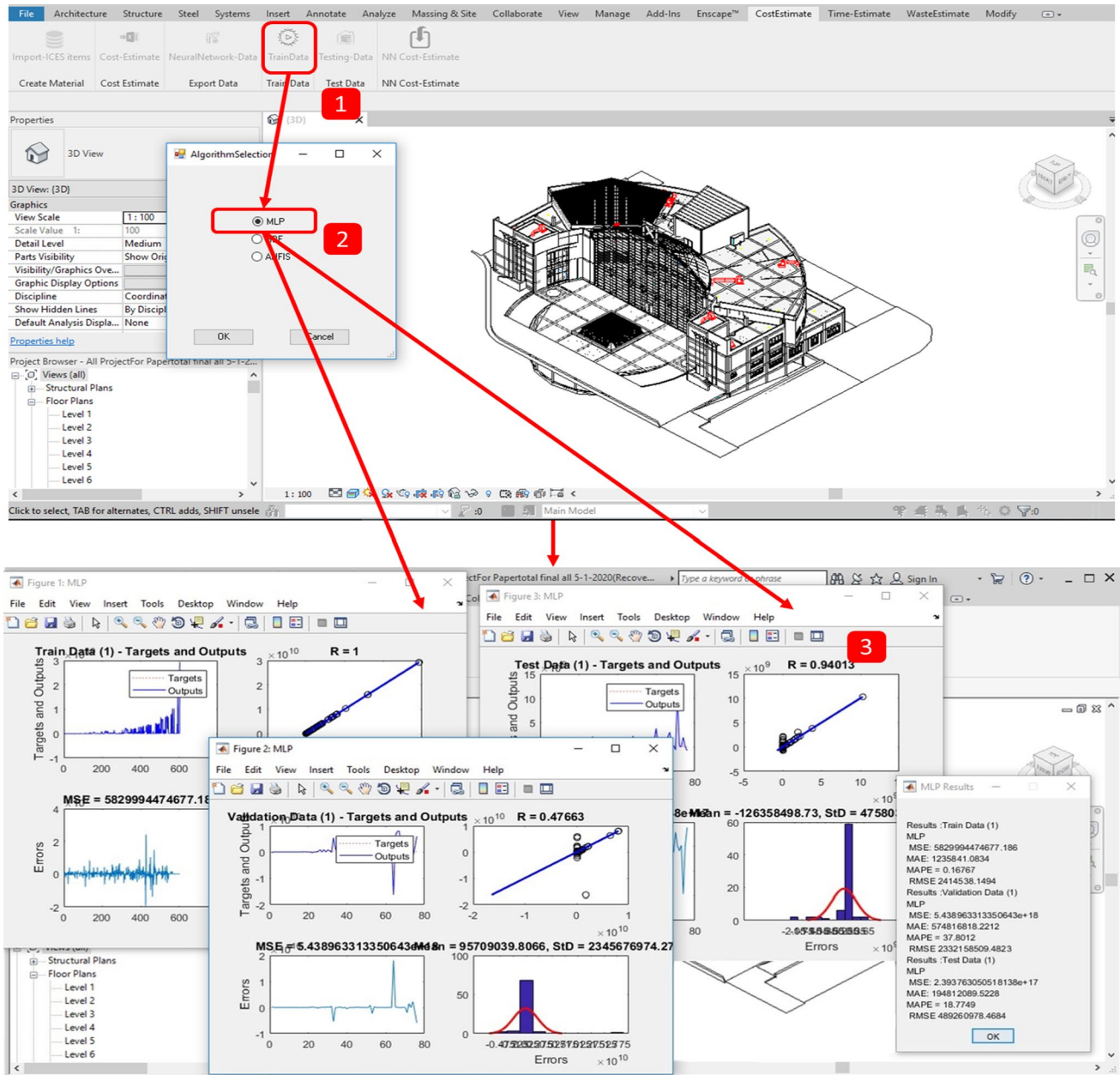
Fig. 11 Back and forth MLP implementation

according to the fuzzy surgeon model in an application of adaptive system [59]. It makes ANFIS modeling more systematic and without necessity for expert knowledge. Figure 12 represents a surgeon fuzzy system that includes two input parameters, one output parameter and two rules [55, 60]. The corresponding ANFIS Architecture is also indicated in Fig. 12 [54, 55]. Its rules are as follows:

Rule 1   If x is A_1 and y is B_1 Then $f = p_1x + q_1y + r_1$
Rule 2   If x is A_2 and y is B_2 Then $f = p_2x + q_2y + r_2$

where x and y are the inputs, A_i and B_i are the fuzzy sets, f_i is the output, and p_i, q_i, and r_i are the design parameters determined during training [57]. It has five layers in which each output signal is entered into the nodes of the next layer.

The layers from the input to output are as follows [55]:

Layer 1, Fuzzy layer, includes the membership functions. Triangular and bell-shaped functions are the most popular membership functions in this layer. Every node in this layer is an adaptive node.

$$f_1 = p_1 + q_1 + r_1$$

$$f_2 = p_2 + q_2 + r_2$$

$$f = \frac{w_1 f_1 + w_2 f_2}{w_1 + w_2}$$

$$= \bar{w}_1 f_1 + \bar{w}_2 f_2$$
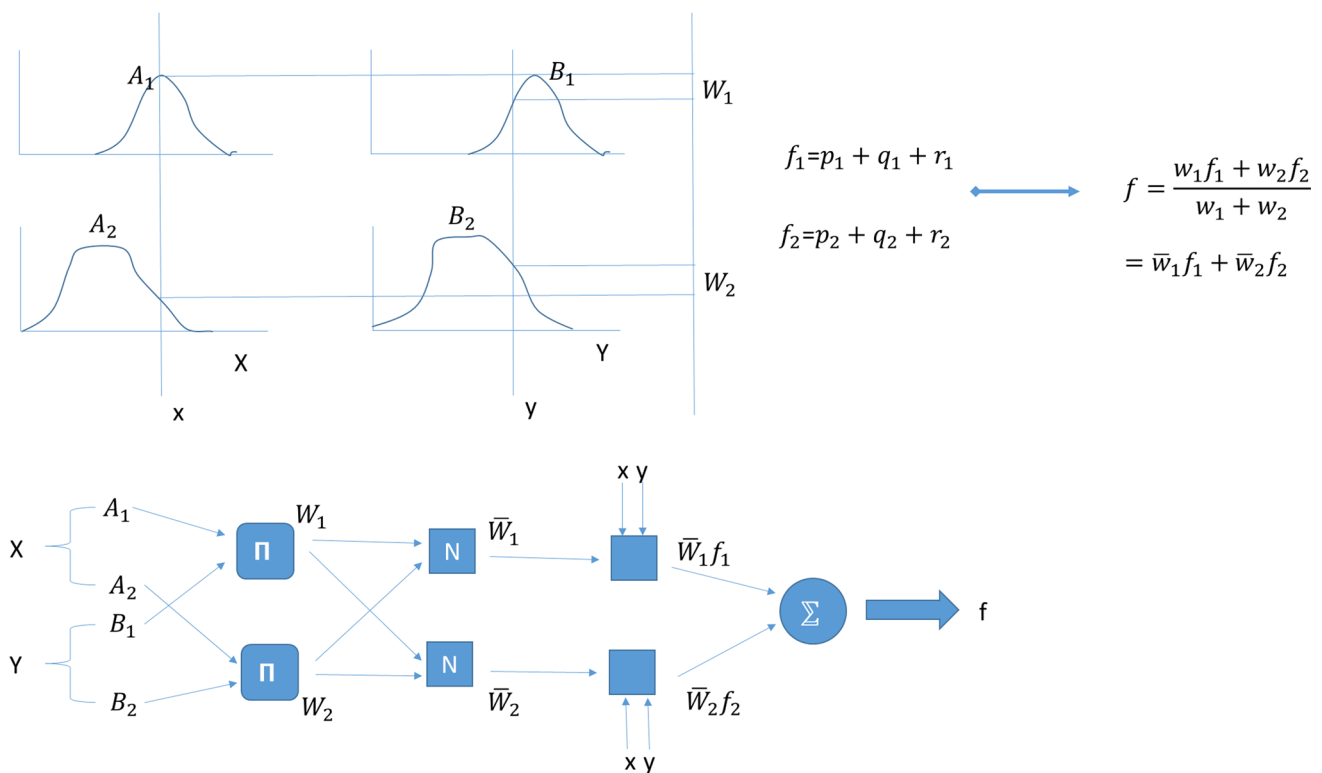
Fig. 12 ANFIS process in MATLAB [54, 55]

Layer 2, product layer, represents the strength of each rule. In fact, it evaluates the grade of activation of incorporated rules. Every node in this layer is a fixed node.

Layer 3, normalized layer, represents the normalized strength of each rule. The ith node computes the ratio of the ith rules firing strength. Every node in this layer is a fixed node.

Layer 4, defuzzied layer, contains a linear combination of multiplied inputs by the normalized strength w. It produces the output of layers.

Layer 5, total output layer, calculates the sum of the outputs in layer 4 (Fig. 12).

The same back-and-forth method is run for the ANFIS as well.

Step 3: developing the results in BIM

In this step, the results of the desired algorithm trained by MATLAB are shown for the user in Revit. Therefore, 4 values are shown including train and test charts, all data results from the algorithms and the measure of errors in the algorithms by using different methods including mean square error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and root-mean-square error (RMSE) for all charts. Figure 11 shows the sample charts of MLP algorithm after implementing the framework. As

it can be shown, for each chart, there are number of tables. The first table is the data of training or test in MATLAB, the second one is to show the linear regression for results, the third one is to demonstrate errors and the final one shows the variants of errors. All of these processes are automatically implemented on the "TrainData" function in the developed extension. It should be considered that MATLAB can save the trained data. So, for testing the trained data for different algorithms, another plugin has been created. This plugin enables BIM engineers to figure out the best performing neural network, the one which performs as close as that of BIM-based cost estimation (Fig. 13).

An application called "Testing-Data" has been produced in order to estimate and compare the cost of elements using the trained data with BIM-based cost estimation. The plugin extracts data from the new 3D model and the user can finally estimate the cost of elements using trained data. This application can give an estimation cost by different methods in the form of charts along with the reports in MATLAB (Fig. 13).

By finding out the best ANN algorithm, there is no need to repeat the whole previous process of examining different types of neural networks and BIM engineers can simply estimate the cost of building elements in the future projects using "NN Cost Estimate" plugin. In fact, this tool estimates the cost of elements using the trained algorithm and associates the results to each building element. Figure 14 shows
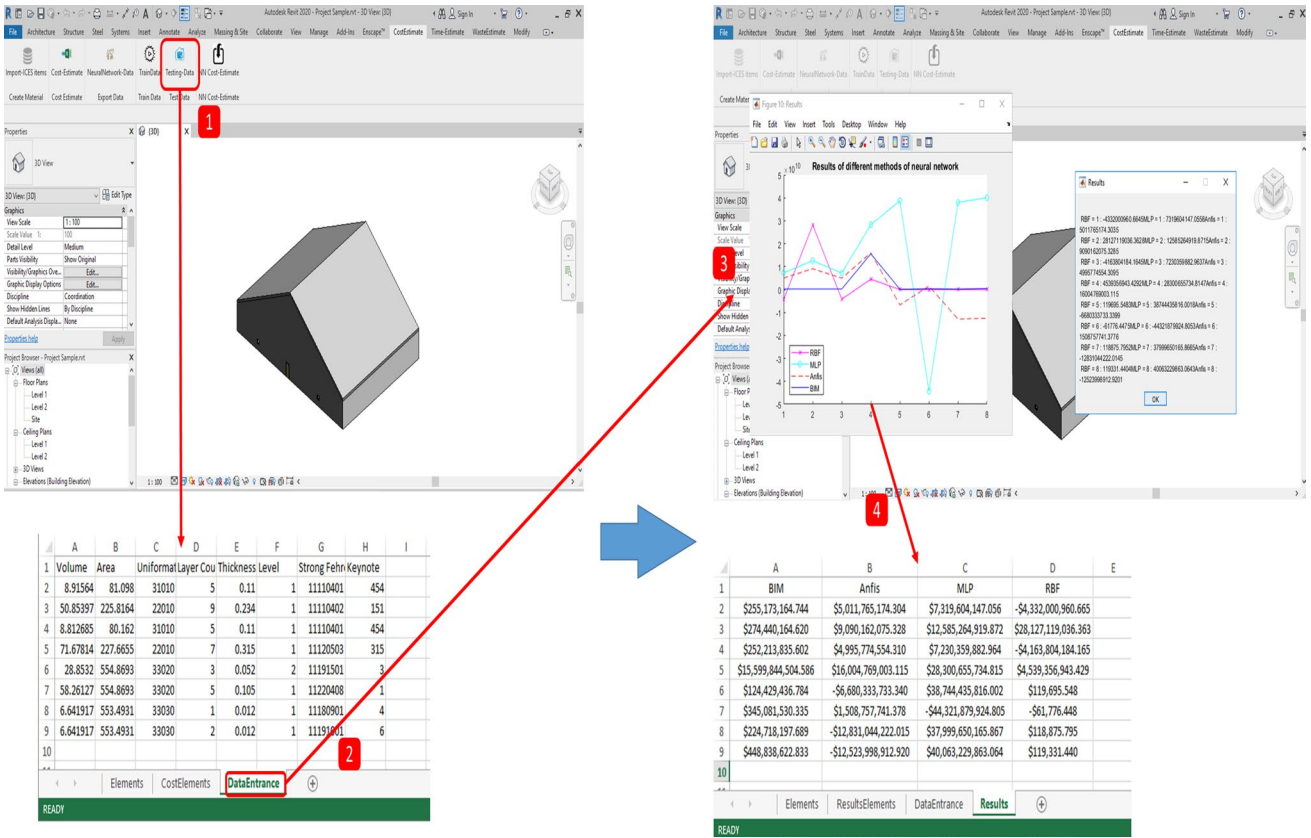
Fig. 13 Testing data framework process

how this plugin works. As it can be seen, firstly, BIM engineers need to choose the best neural network algorithm for the current project and then a new parameter called "NN-estimation" is generated and the estimated cost is drawn for each element. The generated plugin also delivers the total cost of each element.

The generated prototype

In order to implement the proposed method, an extension and compilation of 6 plugins is developed. Figure 14 demonstrates an overview of the generated prototype. As it can be shown, the first plugin called "ImportICES" sends ICES items into Revit using the ICES database created in Excel. The second plugin CostEstimate" is to estimate the cost of each element and produce a report to the user. This plugin also creates a total cost parameter for each element in the project and fills in the blank parameters with estimated costs. The next part of the plugin, "TrainData" application, is for sending created data and implementing ANN algorithms in MATLAB. The next plugin is to test new data and enable BIM engineers to test different algorithms in different projects and select the most suitable one for that project. The

final plugin is for the cost estimation using the algorithm and importing results to the Revit (Fig. 15).

Prototype validation

To validate the developed prototype, a real Construction Engineering Organization Office located in Qom, a city of Iran, was used. This building can accommodate around 900 people. The project includes four main floors with one basement for parking and a facility room. The construction site of the project has a total of 6800 square meters. Figure 16 depicts an overview of the project.

For estimating the cost of the project, all ICES items in the project are imported using "ImportICES item" plugin. Then, the user must determine appropriate ICES items for each element of the project. After doing this, the user can apply the "cost estimate" plugin for the project and the plugin, then, estimates the cost and shows a popup window for summarized cost estimation report. The plugin also reports the full ICES items along with the details in the Excel file. It also produces a parameter which is named the total cost for each element and the user can see the cost by clicking on each. Figure 17 shows how the user should
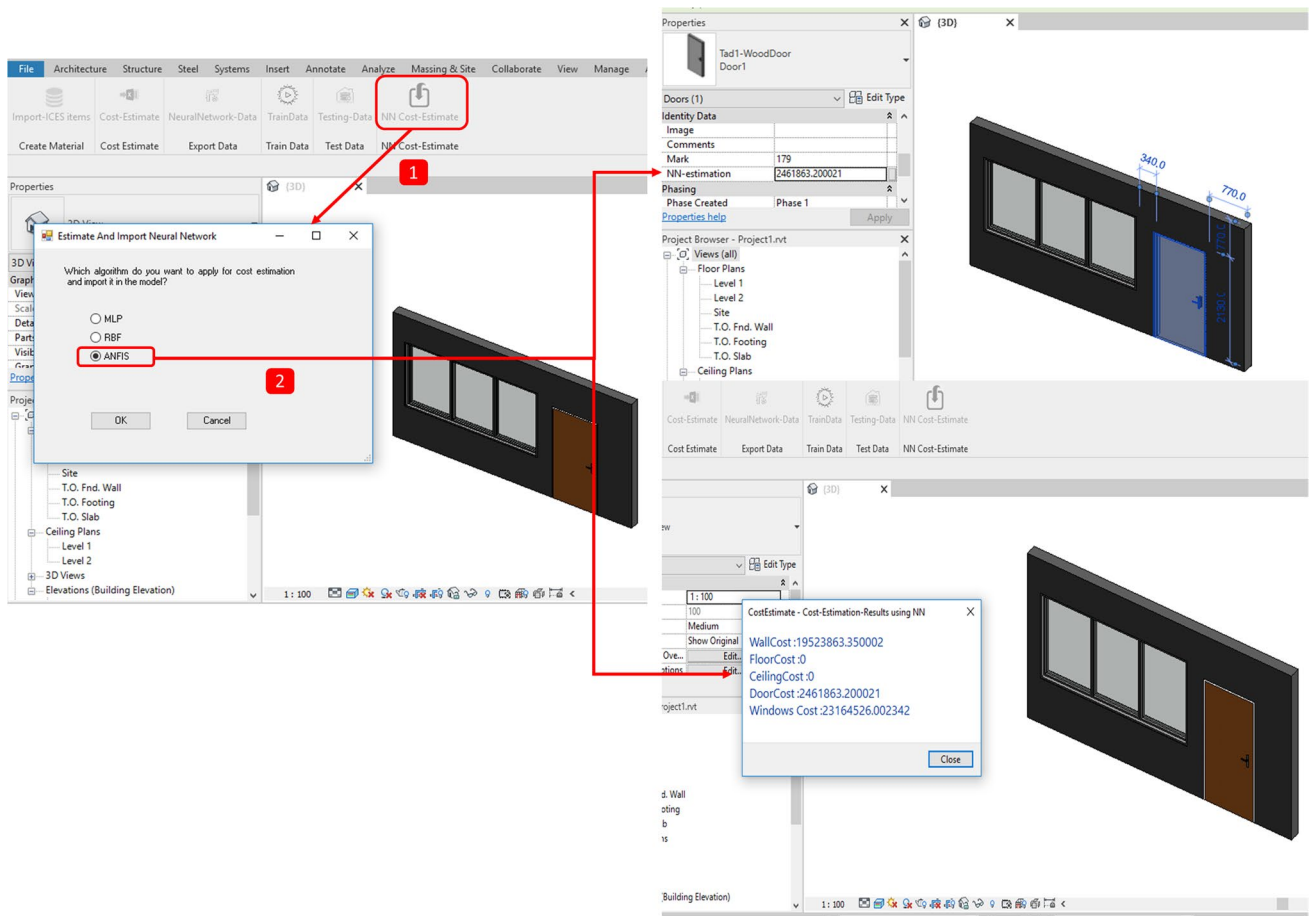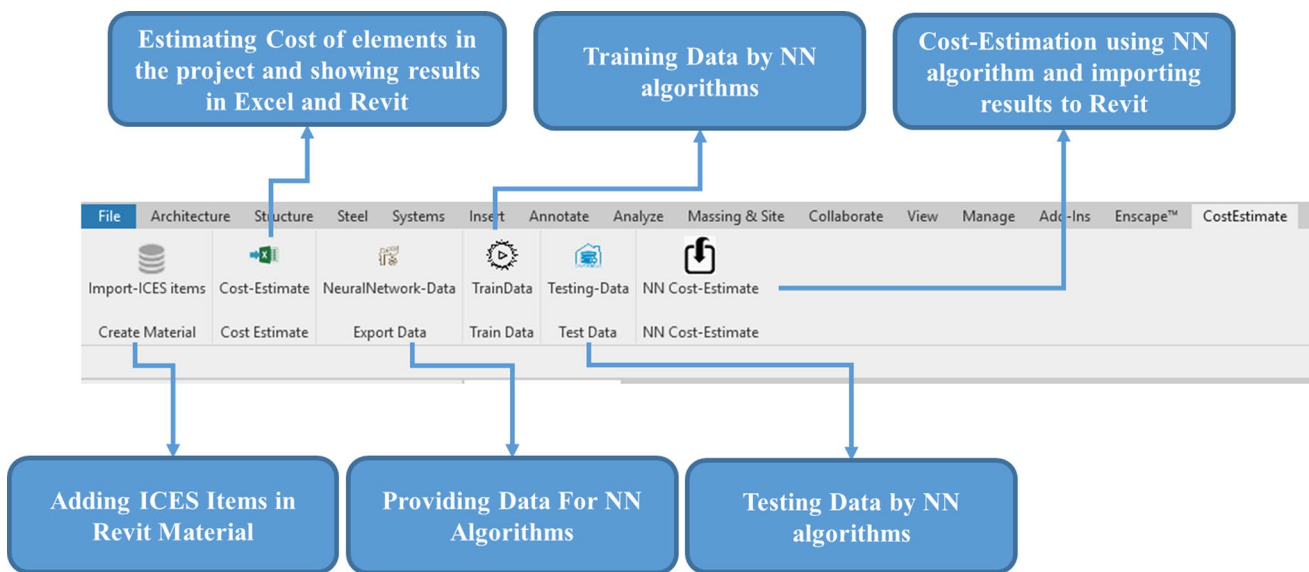
Fig. 14 Importing the algorithms estimation results

Fig. 15 The overview of generated prototype in Revit

Fig. 16 The overview of the case study

choose ICES items for the elements, and Fig. 18 demonstrates the summarized cost estimation reports. It should be noticed that the most recent ICES version of 2020 was used for this case study [33].

Table 2 shows the cost of each category of the project elements. As it can be seen, the cost of the wall category is 1.564 E+12 Rials which constitutes the largest cost of the project. Floor and windows are in the second to the third ranks by representing 2.852 E+9 Rials, and 2.344 E+9 Rials, respectively. Door components make up the least amount of the cost by 1.970 E+8 Rials.

The main challenge for the developed prototype is to manually select the ICES items which is a very time consuming and error-prone task. The main solution for this issue

is to estimate the cost of elements using machine learning approach. Therefore, all data of the elements in the project are applied as the training data for ANN. So, all data were extracted from Revit and imported into Excel using "Neural-Network Data" plugin. By its activation, all required inputs and outputs of the project were imported in the "Elements" and "CostElements" of the Excel and as a result, a record of 756 items of elements along with their costs were sent to Excel. Figure 19 demonstrates the overview of the created Excel.

After compiling the dataset for neural network, the machine learning algorithms using MATLAB engine are utilized. In fact, by "TrainingData" plugin, the user must choose the desired algorithm and MATLAB finds the Excel generated data, implements the algorithm and trains the data. In this study, all algorithms were conducted for the case study. Figure 20 shows the results of MLP algorithm. As it can be shown on the Figure, 80% of the data were used for training and 10% was dedicated for the validation and test, respectively. To assess the performance of the algorithm, four major indicators of MSE, MAE, MAPE and RMSE were applied. The MSE, MAE, MAPE and RMSE for the test data were found as 1.36 E+16, 40.13 E+6, 1.9983 and 11.664 E+7, respectively.

Figure 21 demonstrates the application of RBF algorithm on the data. The data were performed by the maximum neurons of 525 neurons. 80% of data were utilized for training and the rest was used for testing. As it is shown in Fig. 20, the MSE, MAE, MAPE and RMSE performance indexes
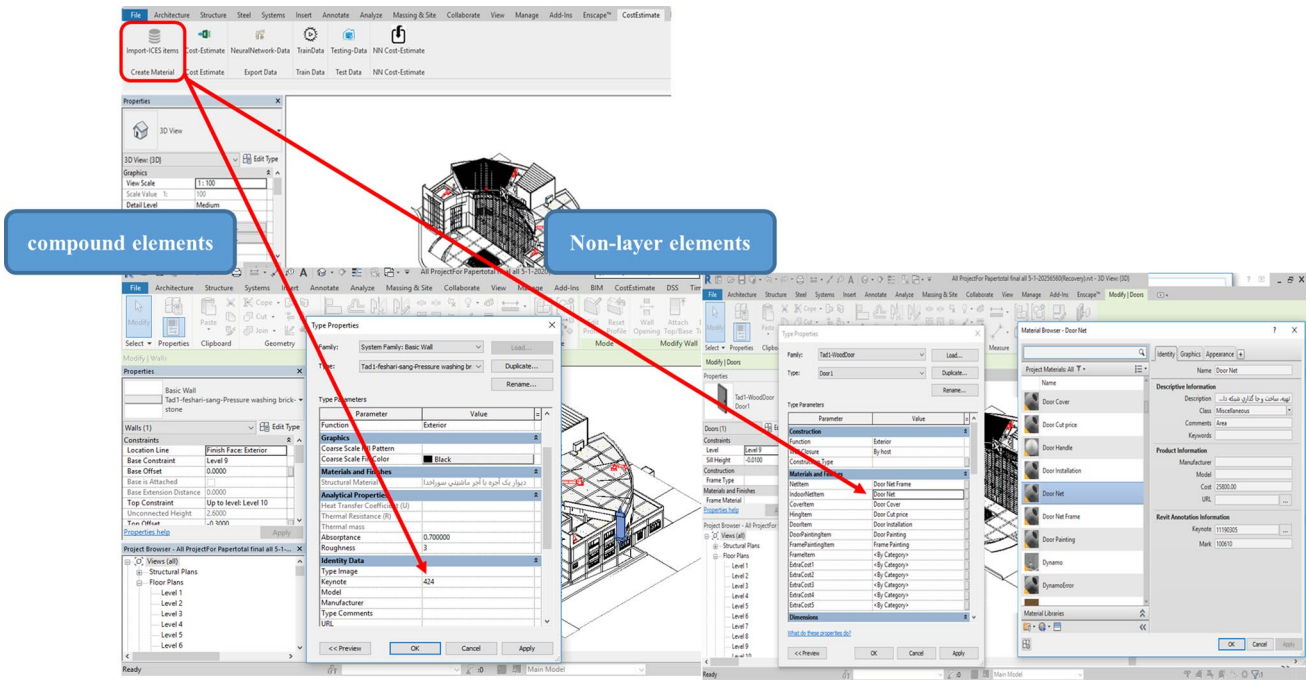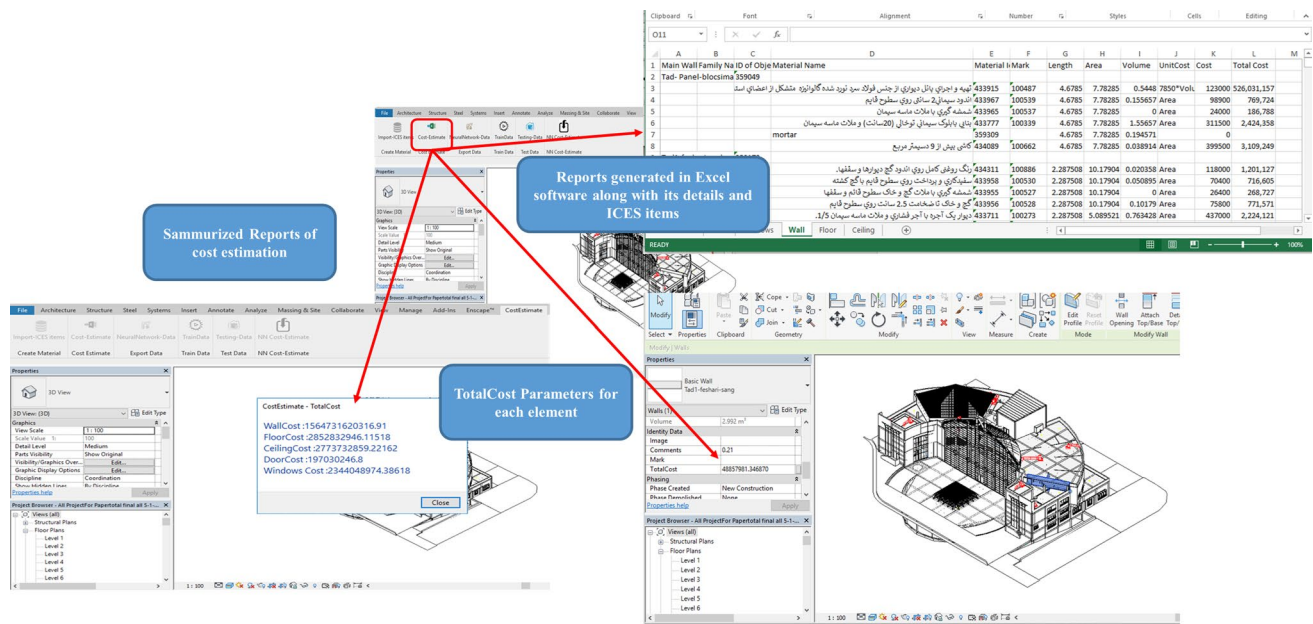

Fig. 17 Selecting ICES items by user

Fig. 18 Results application by the prototype

Table 2 Cost estimation results for the case study

Category	Cost (Rials)
Wall	1,564,731,620,316.910
Floor	2,852,832,946.115
Door	197,030,246.800
Windows	2,344,048,974.386
Total Cost	1,572,899,265,343.430

of the algorithm for testing data are 1.58 E+17, 99.07 E+6, 1.57 and 39.79 E+7, respectively.

Figure 22 indicates the results of the ANFIS. The initial fuzzy inference system for data was performed by the maximum iteration of 450 through the surgeon function. 70% of data were utilized for training and the rest was used for testing. As it is seen in this Figure, the MSE, MAE, MAPE and RMSE performance indexes of algorithm for testing data are 5.91 E+17, 414.51 E+6, 37.59 and 76.90 E+7, respectively.

Once the data were trained, the algorithms are prepared by adding new inputs and comparing the outputs for the validation stage. In fact, the main objective of this step is to compare different types of neural networks and choose the optimized algorithm for cost estimation. For doing so, new inputs were added in the" Entrance Data" sheet of the Excel.

By generating new inputs, the "TestingData" plugin was applied for estimating the cost of elements using trained algorithms. The plugin, first, estimated the cost of elements using BIM and, then, predicted the costs using trained Data. Figure 23 shows the generated results produced by different algorithms. As it is inferred from this Figure, the best

algorithm which generated the most optimum results is ANFIS. It has perfectly generated the cost estimates matched with the 5D BIM system estimation. Then, RBF is in the next place and MLP shows the worst results. It can be shown from this figure that despite the good results of training, the solo application of neural networks cannot, accurately, estimate the cost of building elements in this prototype. This can be due to the insufficient number of data which should be used for training of this type of algorithm. However, ANFIS has performed very superior against RBF and MLP and also delivered the results well-aligned with 5D BIM. It is, therefore, a recommended machine learning tool for 5D BIM integration.

By choosing ANFIS as the best algorithm, there is no need to repeat the process of testing different algorithms and BIM engineers can choose ANFIS, hereafter, as the most optimized algorithm for cost estimation of the future projects. Therefore, a one-floor residential house, located in Tehran, is chosen for cost estimation applying this algorithm. Figure 24 shows the result of cost estimation using ANFIS on this case study.

Table 3 compares the results of the ANFIS case study (Fig. 24) with the BIM-based cost estimation which is the same as the planned cost estimation of the building elements. The comparison indicates the acceptable performance for the Wall category where there is around 13% of variation between the ANFIS and BIM whereas the variations for the other elements including floor, doors and windows are higher. This finding can be in view of benefiting from the larger number of data for the wall category and having smaller dataset of floor, windows and doors in the
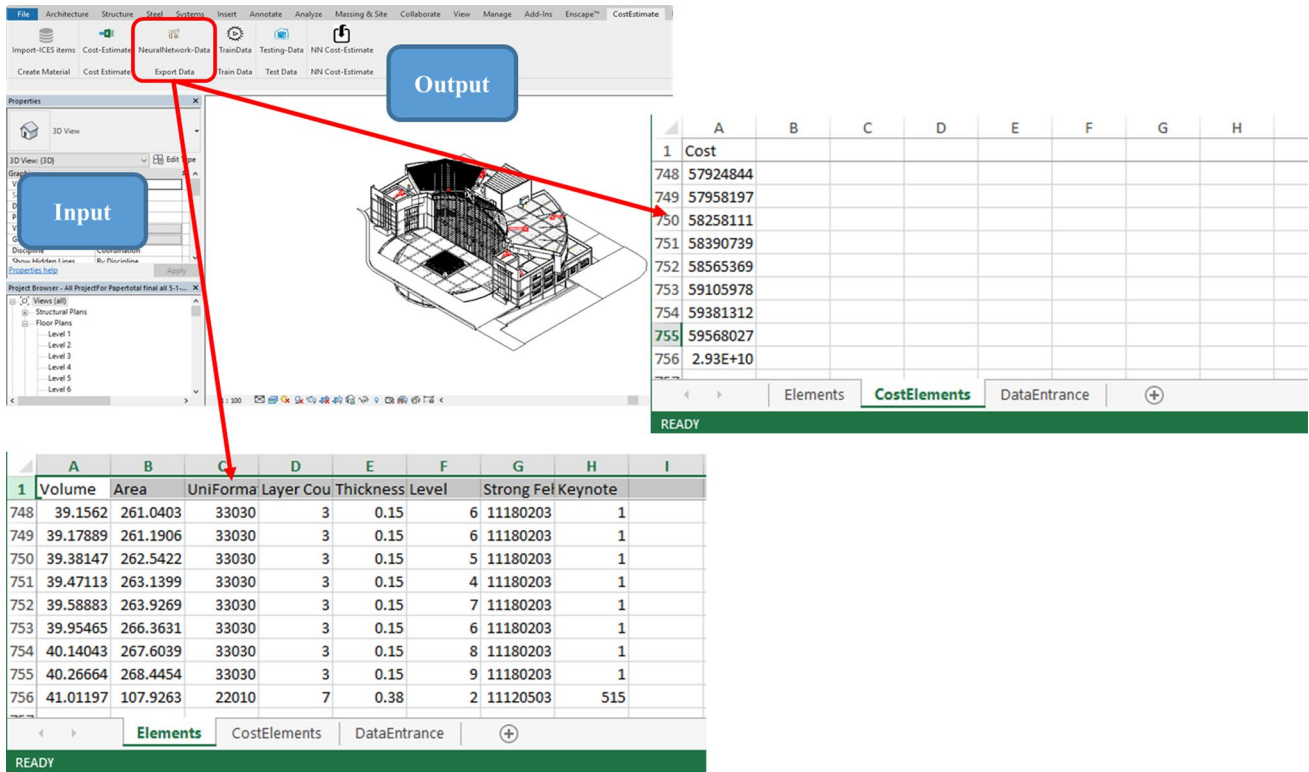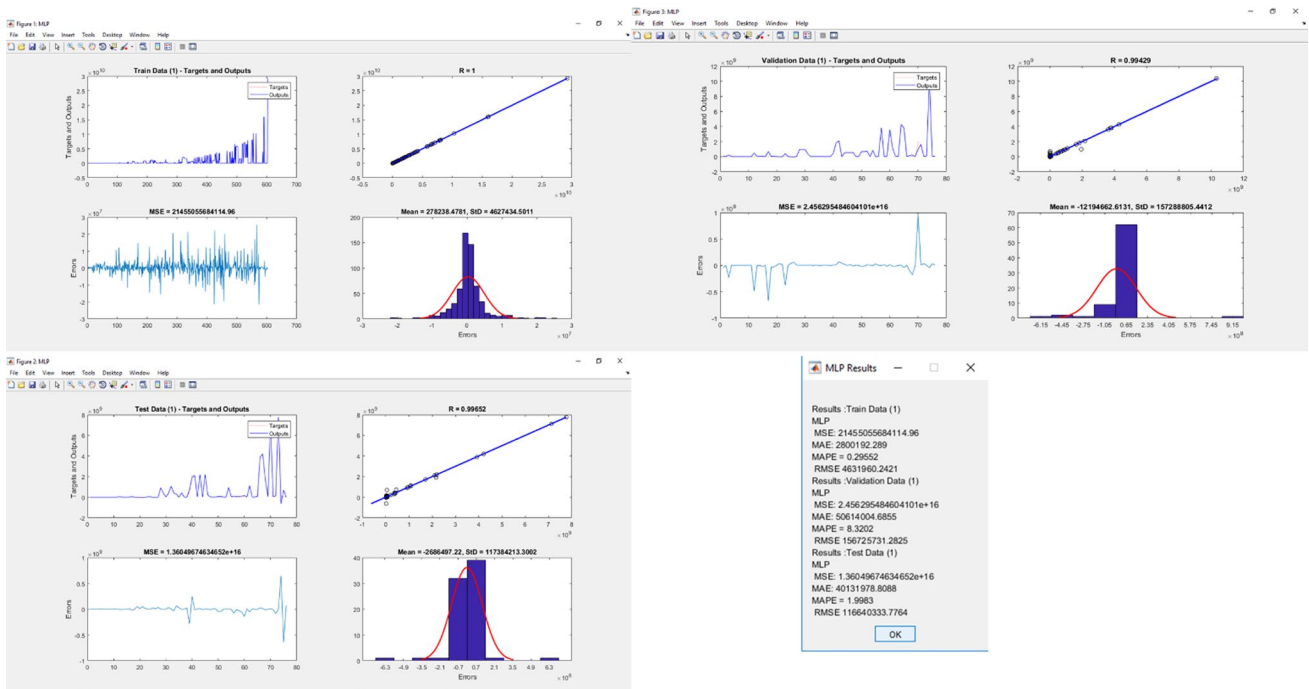
Fig. 19 Extracting data for machine learning algorithms
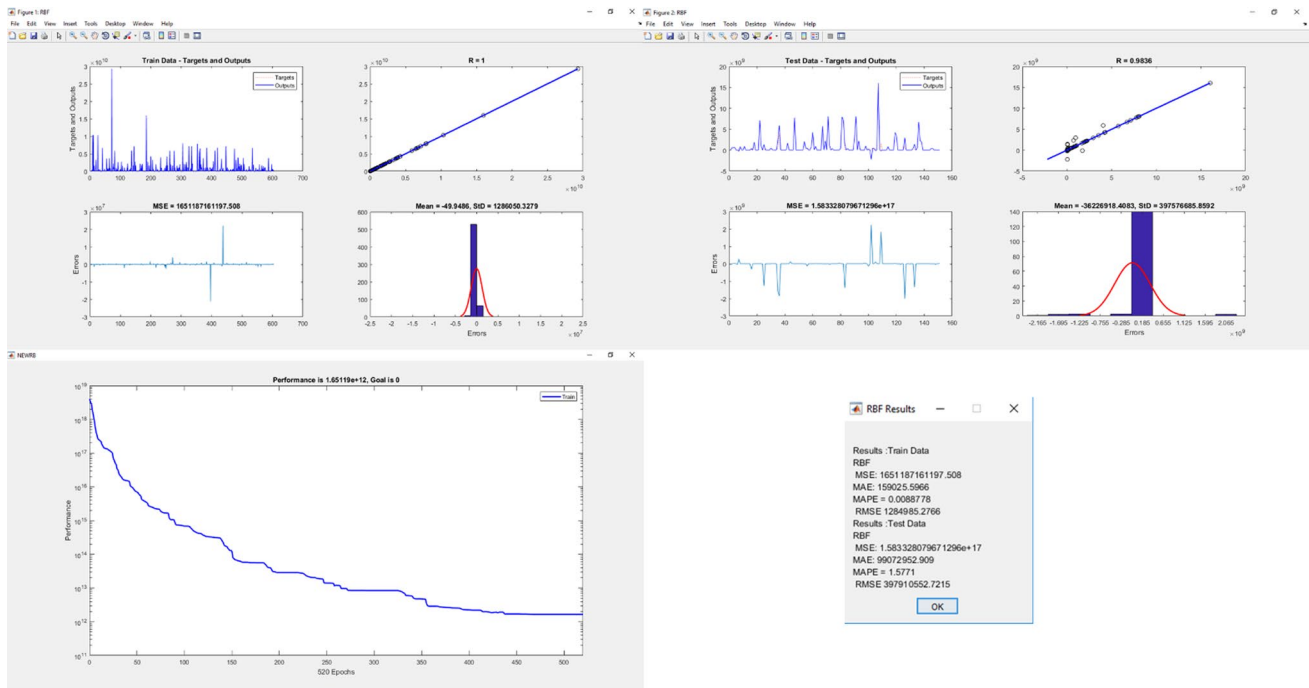

Fig. 20 MLP algorithm results

Fig. 21 RBF algorithm results

Discussion and contribution

The lack of connection between BIM and cost specifications is one of the main barriers for adoption of BIM in the construction industry of developing countries such as Iran. Therefore, this paper aimed to bridge the gap of connecting building cost specifications and BIM-based system. ICES items are used as the representative of specification where BIM engineers can apply the similar system for other BIM-integrated applications. Previous studies made attempt to separately estimate the cost of materials, equipment and labors. However, this study develops an integrated system in which all of the building cost items are considered. Furthermore, non-informatics project cost estimation process suffers from a great deal of time and errors. Thus, there is a research priority to develop a classified building cost component and integrate with 5D BIM in order to facilitate the procedure of cost estimation. This necessity is particularly significant where BIM-driven parametric capabilities can alleviate this error-prone process. Therefore, this study developed a new informatics prototype which can connect ICES items with BIM authoring

applied case study. All in all, ANFIS has indicated a good performance in estimating the cost of elements, while there is still a room for better performance by expanding the dataset for more enhanced training.

tools. However, it should be noted that this process was still semi-automatic and involved handiworks as users, cost estimators and BIM managers should designate and choose appropriate ICES items for each element in projects. So, cost estimation could be still time-consuming and error-prone. In order to solve this problem, this study employed machine learning and ANN algorithms in order to fully automate the process. Thus, three main types of ANN including MLP, RBF and ANFIS were implemented and their performance was tested and compared via the real-life construction case study. The results indicated the most optimum performance for the ANFIS algorithm in comparison with the other ANN algorithms. The reason is in light of the integration of fuzzy and neural networks in the ANFIS which declines the negative effects of the black box mechanism of neural networks. However, the validation comparison between ANFIS and BIM-based cost estimation implies that the algorithm should be trained with larger and more expanded datasets in order to deliver more accurate and closer results with that of BIM system.

In addition to the above advantages, cost estimators can be benefited from the following items:

1. Cost estimators can automatically find out the quantities of all construction elements along with the cost estimation details in the form of reports and schedules.
2. A prototype application is developed which can use a formulated process map and algorithms in the process of
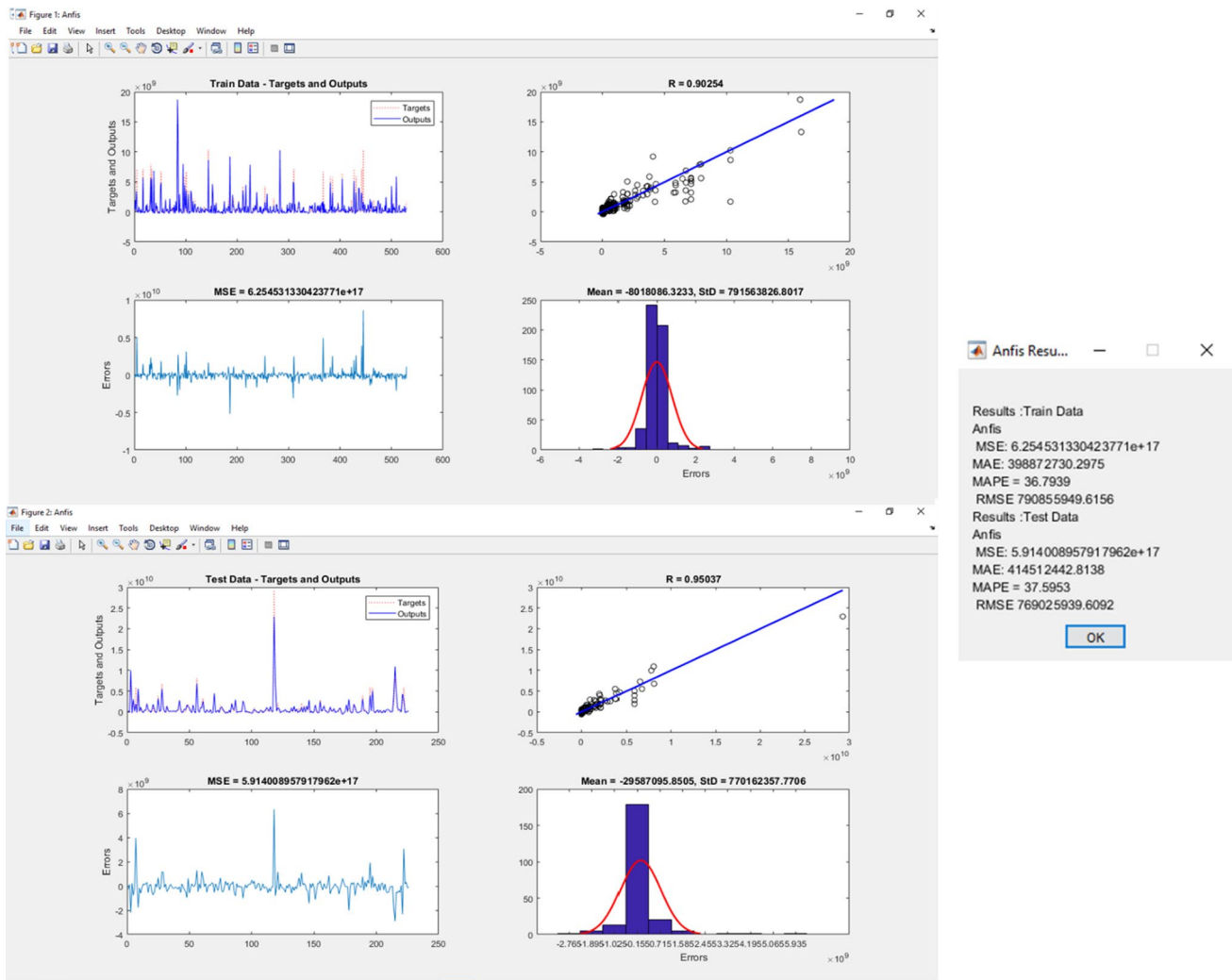
Fig. 22 ANFIS algorithm results

cost estimation. The process can significantly decrease estimators' workloads and errors.

3. The total project cost can be identified in the preliminary steps of the project development.

Conclusion

Traditional cost estimation process involves selecting individual elements, using a tool to determine the dimensions for the takeoff, and inputting the quantities into the QTO list. These shortcomings require estimators, engineers and construction managers to spend a substantial amount of time, errors and omissions. However, using BIM systems can facilitate the process of accurate QTO, as they are enriched by object-based and built-in parametric information. Although many studies have been done in the area of implementing BIM approach in the cost estimation, very little is known regarding an integrated and standardized BIM-based system for cost estimation using neural networks algorithms in the design phase of construction projects. This is particularly prominent for those developing countries such as Iran, where the BIM implementation level and construction digital transformation are in the embryonic steps.

This research aimed to develop a new cost estimation informatics prototype which integrates building cost components of ICES items with BIM. The automation process was also sought through the ANN algorithms including MLP, RBF and ANFIS. The proposed prototype not only introduces the standardized and integrated BIM-based system in the preliminary phase of a construction project, but it also makes a greater stride for encouraging cost specifications to be integrated with BIM. Progressively, a BIM-based application in Revit environment was also developed to facilitate the process of cost estimation and in continue, the MATLAB was used to automate
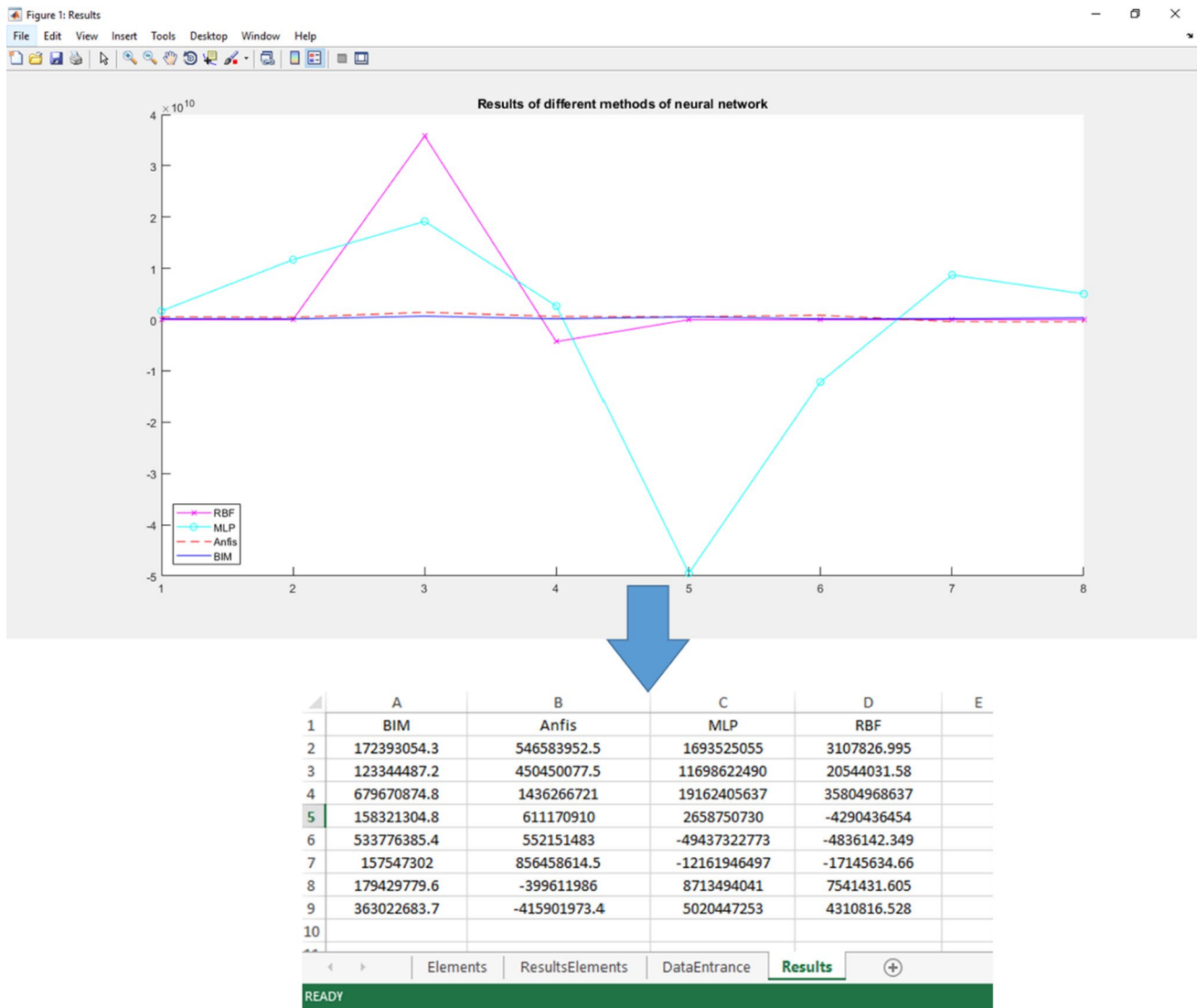
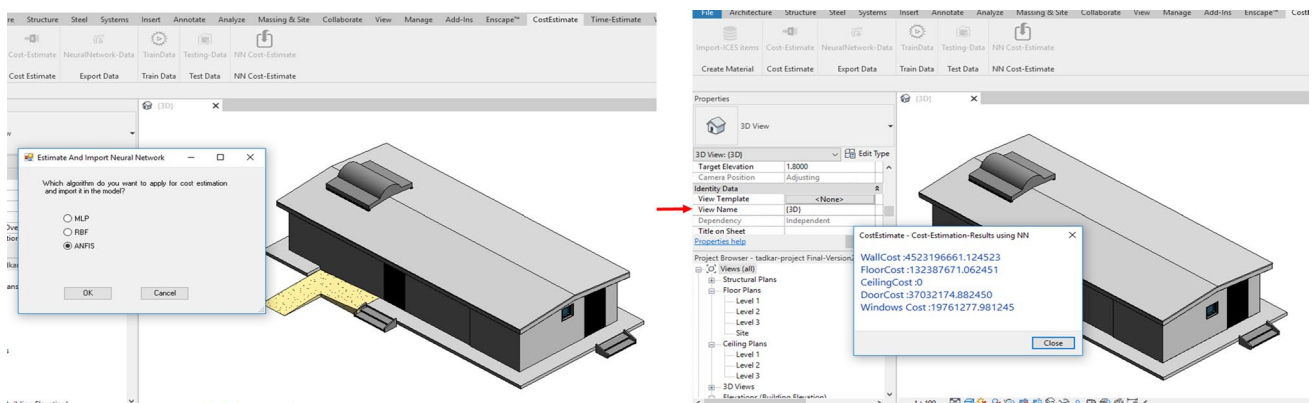**Fig. 23** The performance results of algorithms

Fig. 24 Cost Estimation using ANFIS

Table 3 Comparing the cost estimation using ANFIS and BIM-based system

	Cost estimation using ANFIS	BIM-based cost estimation	Variation%
Walls	4,523,196,661	3,986,712,706	13.4568
Floor	132,387,671.1	100,199,942.5	32.1235
Doors	37,032,174.88	46,541,225.91	− 20.431458
Windows	19,761,277.98	15,999,419.57	23.512468

the process by implementing neural network algorithms on the data.

The developed prototype was further examined through a real-life construction case study. For automating the cost estimation process, the results of the case study were utilized for training ANN algorithms. The research also compared and examined different neural networks. The results show that ANFIS delivers the best results and is the most aligned algorithm with the BIM-based cost estimation system, as compared to the other algorithms. Despite the relatively good performance of the other algorithms, the sole neural network system failed to accurately estimate the cost of components vis-à-vis that of BIM and ANFIS-integrated system. For boosting the performance of those algorithms, it is recommended to add the number of training data, separate different components and train each separately for getting better results.

In addition, only four categories of building elements including wall, floor, door and windows were included in this study, while there are other building elements which can be considered too. The current study only takes into account the architectural and structural parts of ICES items whereas there are other parts in the ICES such as electrical and mechanical elements which can be included in the future works. Moreover, there are some coefficients in the ICES to apply the rate of inflation in building elements during the construction phase of a project. It is recommended to incorporate these coefficients in the future works of cost estimation.

Appendix 1: Applying different scenarios in the cost estimation using the current methodology

Declarations

Conflict of interest All authors acknowledge that there is no conflict of interest for this publication.

References

1. Lee G, Park HK, Won JS (2012) 3D City project-Economic impact of BIM-assisted design validation. Autom Constr 22:577–586
2. Elfaki AO, Alatawi S, Abushandi E (2014) Using intelligent techniques in construction project cost estimation: 10-year survey. Adv Civ Eng
3. Ma Z, Wei Z, Zhang X (2013) Semi-automatic and specification-compliant cost estimation for tendering of building projects based on IFC data of design model. Autom Constr 30:126–135
4. Simpson A et al. (2018) Asset information requirements guide: information required for the operation and maintenance of an asset
5. Xu S et al (2016) Computers in Industry A framework for integrating syntax, semantics and pragmatics for computer-aided professional practice: with application of costing in construction industry. Comput Ind 83:28–45
6. Liu X, Li Z, Jiang S (2016) Ontology-based representation and reasoning in building construction cost estimation in China. Fut Internet 8(3):39
7. Yarnold J et al (2021) Building and construction quality: systematic literature review, thematic and gap analysis. Int J Build Pathol Adapt
8. Akbari S et al (2020) Dynamic sustainable success prediction model for infrastructure projects: a rough set based fuzzy inference system. Constr Innov
9. Elmousalami HH (2020) Artificial intelligence and parametric construction cost estimate modeling: state-of-the-art review. J Constr Eng Manag 146(1):03119008
10. Khanzadi M, Sheikhkhoshkar M, Banihashemi S (2020) BIM applications toward key performance indicators of construction projects in Iran. Int J Constr Manag 20(4):305–320
11. Hosseini MR et al. (2016) Building information modeling (BIM) in Iran: an exploratory study. J Eng Project Prod Manag
12. Banihashemi S et al (2017) Critical success factors (CSFs) for integration of sustainability into construction project management practices in developing countries. Int J Project Manage 35(6):1103–1119
13. Ma Z, Liu Z, Wei Z (2016) Formalized representation of specifications for construction cost estimation by using ontology. Comput-Aided Civ Infrastruct Eng 31(1):4–17
14. Liu H, Lu M, Al-Hussein M (2016) Ontology-based semantic approach for construction-oriented quantity take-off from BIM models in the light-frame building industry. Adv Eng Inform 30(2):190–207
15. Monteiro A, Martins JP (2013) A survey on modeling guidelines for quantity takeoff-oriented BIM-based design. Autom Constr 35:238–253
16. Banihashemi S, Zarepour Sohi S (2022) Data-centric regenerative built environment: big data for sustainable regeneration
17. Shourangiz E et al. (2011) Flexibility of BIM towards design change. In: 2nd international conference on construction and project management. IPEDR ACSIT Press, Singapore
18. Staub-French S et al (2003) A generic feature-driven activity-based cost estimation process. Adv Eng Inform 17(1):23–39
19. National Institute of Building Sciences C. (2007) National building information modeling standards: overview principles, and methodologies. NBIMS, United States, vol 1
20. Banihashemi S et al. (2019) Multi-sided platforms: a business model for BIM adoption in built environment SMEs. In: HCI in business, government and organizations. Information systems and analytics. HCII 2019. Lecture Notes in Computer Science. Springer, pp 22–32
21. Ferdosi H et al. (2022) BIM applications in sustainable construction: scientometric and state-of-the-art review. International Journal of Construction Management
22. Building Smart International, L.T.D. (2016) IFC overview
23. Oraee M et al (2017) Where the gaps lie: ten years of research into collaboration on BIM-enabled construction projects. Constr Econ Build 17(1):121–139
24. Estimator (2016) *Graphisoft*.
25. ToCoMan (2016) *Tocosoft*.
26. Banihashemi S, Tabadkani A, Hosseini MR (2018) Integration of parametric design into modular coordination: a construction waste reduction workflow. Autom Constr 88:1–12
27. Hosseini MR et al (2016) BIM adoption within Australian small and medium-sized enterprises (SMEs): an innovation diffusion model. Constr Econ Build 16(3):71–86
28. Banihashemi S (2017) Active BIM with artifical intelligence for energy optimisation in buildings. University of Technology Sydney
29. Banihashemi S, Ding G, Wang J (2015) Developing a framework of artificial intelligence application for delivering energy efficient buildings through active BIM. In: RICS COBRA 2015. RICS
30. Davidson AR (2009) A study of the deployment and impact of Building Information Modelling software in the construction industry. e-Engineering--Journal of Undergraduate Research (University of Leeds). London: Available at: http://www.engineering.leeds.ac.uk/eengineering. e-Engineering–Journal of Undergraduate Research (University of Leeds). London
31. Sompolgrunk A, Banihashemi S, Mohandes SR (2021) Building information modelling (BIM) and the return on investment: a systematic analysis. Construction Innovation
32. Sompolgrunk A et al. (2022) An integrated model of BIM return on investment for Australian small- and medium-sized enterprises (SMEs). Engineering, Construction and Architectural Management
33. ICES. Iranian Cost Estimation Standard (2022) http://sama.mporg.ir/sites/publish/SitePages/ZabetehView.aspx?mdid=5577
34. Balali V, Zalavadia A, Heydarian A (2020) Real-time interaction and cost estimating within immersive virtual environments. J Constr Eng Manag 146(2):10
35. Bademosi FM, Tayeh R, Issa RRA (2019) An immersive approach to construction cost estimating, in computing in civil engineering 2019, pp 48–54
36. Valentini V, Mirarchi C, Pavan A (2017) Comparison between traditional and digital preliminary cost-estimating approaches. Innov Infrast Solut 2(1):1–8

37. Akanbi T, Zhang JS (2021) Design information extraction from construction specifications to support cost estimation. Autom Constr 131:14

38. Khosakitchalert C, Yabuki N, Fukuda T (2020) Automated modification of compound elements for accurate BIM-based quantity takeoff. Autom Constr 113:23

39. Wang K-c et al. (2016) Applying building information modeling to integrate schedule and cost for establishing construction progress curves. Autom Constr

40. Wu C-L, Wang KC, Wang WC (2018) A BIM-based conceptual cost estimation model considering structural analysis and design. In: ISARC. proceedings of the international symposium on automation and robotics in construction, vol 35, pp 1–8

41. Ma Z, Liu Z (2014) BIM-based intelligent acquisition of construction information for cost estimation of building projects. Procedia Eng 85:358–367

42. Joosung L, Jaejun K (2017) BIM-based 4D simulation to improve module manufacturing productivity for sustainable building projects

43. Zanni M et al (2019) Developing a methodology for integration of whole life costs into bim processes to assist design decision making. Buildings 9(5):114–114

44. Fazeli A et al. (2020) An integrated BIM-based approach for cost estimation in construction projects. Engineering, Construction and Architectural Management

45. Jiang Q (2019) Estimation of construction project building cost by back-propagation neural network. J Eng Des Technol 18(3):601–609

46. Huang CH, Hsieh SH (2020) Predicting BIM labor cost with random forest and simple linear regression. Autom Constr 118:16

47. Chandanshive V, Kambekar AR (2019) Estimation of building construction cost using artificial neural networks. J Soft Comput Civ Eng 3(1):91–107

48. Elghaish F et al (2021) Revolutionising cost structure for integrated project delivery: a BIM-based solution. Eng Constr Archit Manag 28(4):1214–1240

49. Bala K, Ahmad Bustani S, Shehu Waziri B (2014) A computer-based cost prediction model for institutional building projects in Nigeria: an artificial neural network approach. J Eng Des Technol 12(4):519–530

50. Cheng M-Y, Tsai H-C, Sudjono E (2010) Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry. Expert Syst Appl 37(6):4224–4231

51. Prive NC, Errico RM (2016) Observing system simulation experiments: an overview. Maryland University, College Park

52. Hancock D, Algozzine B (2016) Doing case study research: a practical guide for beginning researchers. Teachers College Press

53. Fazeli A et al. (2019) BIM-integrated TOPSIS-Fuzzy framework to optimize selection of sustainable building components. Int J Constr Manag:1–20

54. Kashi H, Emamgholizadeh S, Ghorbani H (2014) Estimation of soil infiltration and cation exchange capacity based on multiple regression, ANN (RBF, MLP), and ANFIS models. Commun Soil Sci Plant Anal 45(9):1195–1213

55. Amid S, Mesri Gundoshmian T (2017) Prediction of output energies for broiler production using linear regression, ANN (MLP, RBF), and ANFIS models. Environ Prog Sustain Energy 36(2):577–585

56. Venkatesan P, Anitha S (2006) Application of a radial basis function neural network for diagnosis of diabetes mellitus. Curr Sci 91(9):1195–1199

57. Jang JR (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern 23(3):665–685

58. Wali WA et al (2012) Online genetic-ANFIS temperature control for advanced microwave biodiesel reactor. J Process Control 22(7):1256–1272

59. Mousavi SM et al (2021) An intelligent hybrid classification algorithm integrating fuzzy rule-based extraction and harmony search optimization: medical diagnosis applications. Knowl-Based Syst 220:106943

60. Riahi-Madvar H et al (2019) Comparative analysis of soft computing techniques RBF, MLP, and ANFIS with MLR and MNLR for predicting grade-control scour hole geometry. Eng Appl Comput Fluid Mech 13(1):529–550