

## Research Article

Jumaa A. Al-Somaydaii, Aminah T. Albadri and Faiq M. S. Al-Zwainy\*

# Hybrid approach for cost estimation of sustainable building projects using artificial neural networks

<https://doi.org/10.1515/eng-2022-0485>

received April 27, 2023; accepted July 04, 2023

**Abstract:** Inaccurate estimation in sustainable construction projects is a significant challenge for appraisers, particularly when data and knowledge about the projects are lacking. As a result, there is a need to use cutting-edge technology to solve the issue of estimation inaccuracy. Iraq's productivity estimates are now made using outdated, ineffective methodologies and procedures. In addition, it is essential to implement cutting-edge, quick, precise, and adaptable technology for productivity estimation. This study's major goal is to calculate the overall costs of sustainable buildings using the cutting-edge technique known as artificial neural networks (ANNs). For Iraq's construction industry to handle projects successfully, ANNs must be used as a new technology, a methodology developed to estimate the overall costs of sustainable construction projects. In this study, the process of cost estimation was modeled using ANNs. Investigations of a number of examples involving the creation of ANNs have also been made, including network design and internal elements and how much they impact the effectiveness of models built using ANNs. Equations were developed to determine structural productivity. These networks were shown to have extremely strong predictive power for both accounting coefficients ( $R$ ) (93.33%) and the overall costs of sustainable construction projects, with a prediction accuracy of 87.00 and 93.33%, respectively.

**Keywords:** cost estimation, sustainable buildings, neural network, correlation coefficient

## 1 Introduction

Sustainable buildings consider how to consume less energy, resources, and materials, as well as how to minimize their adverse impacts on the environment during construction and usage while increasing their harmony with the natural world. They are also designed, implemented, operated, and maintained in ways that respect the environment. With the first energy crisis in the 1970s, when engineers questioned the sense of having box structures surrounded by glass and steel that needed large-scale heating and costly cooling systems, sustainable architecture was born. For more energy-efficient structures, the volume was increased [1].

The emphasis shifted to the long-term environmental effects of building operation and maintenance, with a vision that went far beyond the project's initial costs. This viewpoint has now found support in a number of building assessment systems, including the 1990 AD implementation of the BREEAM standard in Britain. In the United States of America, the term "LEED standards" refers to the in Energy and Environmental Design Leadership Standard, which was created by the US Green Building Council and introduced in 2000. The most important recommendations to establish the concept of sustainable buildings are to rationalize the ways of construction and housing and to take into account sustainable development that requires the use of existing resources without compromising the resources needed for future generations [1].

In addition to developing an integrated construction plan with the environment in mind, sustainable buildings are also interested in the use of green energy, or renewable energy, particularly solar and wind energy, since our Arab cities have a significant supply of solar energy. By integrating the following three fundamental components, sustainable buildings establish harmony between an individual, his or her society, and the environment [1]:

- 1) Effective utilization of materials and resources.
- 2) Effective management of the local environmental, political, and social circumstances.

\* **Corresponding author: Faiq M. S. Al-Zwainy**, Forensic DNA Center for Research and Training, Al-Nahrain University, Jadriya, Baghdad, Iraq, e-mail: [faiqalzainy@gmail.com](mailto:faiqalzainy@gmail.com)

**Jumaa A. Al-Somaydaii:** College of Engineering, University of Anbar, Ramadi, Iraq

**Aminah T. Albadri:** College of Engineering, Al-Nahrain University, Jadriya, Baghdad, Iraq

- 3) Meeting basic human necessities while protecting the rights and requirements of future generations.

Cost is a topic that has gained significant relevance in the building business over many decades. Coping with the cost issue begins when the owner or developer has the first concept and continues throughout the course of the project [2,3].

Cost estimate, which defines the project's baseline cost at different phases of project development, is one of the most crucial project management procedures. During a certain stage of project development, the cost estimate is an engineer's or cost estimator's projection based on the facts at hand [4,5].

The success of a building project depends heavily on precise cost estimates throughout the early stages of the project. Yet when documents, drawings, etc., are still lacking, it is challenging to precisely and swiftly estimate the building costs at the planning stage. Several approaches are used in this manner to accurately estimate building costs at an early stage, when project information is few. While each approach has advantages and disadvantages, little work has been done to identify the technique that performs the best in terms of cost estimates [6–8]. From this point on, the researcher will concentrate on conducting building cost calculations using one estimating approach, neural network (NN).

Construction cost prediction using computational intelligence techniques has gained a lot of attention recently. Thanks in large part to the widespread use of artificial neural networks (ANNs) in this and other completely unrelated fields of management, science, and the construction project industry [9–11].

The novelty of this research lies in the addition of a replacement technique within the field of machine intelligence, that of NN, which has been applied to the sector of sustainable building projects. Very few researchers have examined the viability of prediction models applied to construction cost prediction [12].

The major focus of this study is on utilizing association NN to estimate costs for sustainable construction projects, with the intention of helping project managers and estimators make decisions. Finding the most efficient NN activity in terms of model parameters and structure is the research's contribution to cost prediction.

The practical application of this study may likely be the use of ANN, a machine intelligence technology, as a reliable tool for predicting building costs. In order to get financing or convince decision-makers to approve the budget, project managers in Iraq are often forced to estimate the cost of sustainable building projects early on and accurately [13,14]. As a result, finding the most straightforward

method for quickly and accurately determining the cost of sustainable buildings projects is vital.

The most significant prior studies on using ANNs to calculate the costs of sustainable structures between 2018 and 2021 are included along with their sources.

A study conducted by Tatari *et al.* [15] developed an ANN model for predicting the costs of green building projects. The model was trained and tested using data collected from actual green building projects in Egypt. The study concluded that ANN is a reliable and accurate tool for cost estimation of green buildings.

In another study, Jung *et al.* [16] proposed a hybrid approach that combines ANNs with genetic algorithms for the cost estimation of sustainable building projects. The hybrid model was able to achieve better accuracy than the traditional regression-based models.

Research by Asgari *et al.* [17] utilized an ANN-based model for predicting the costs of sustainable building projects in Iran. The model was trained and validated using data collected from actual sustainable building projects. The results showed that ANN is a highly effective tool for predicting the costs of sustainable buildings.

A study conducted by Naghizadeh *et al.* [18] proposed a novel approach that combines ANNs with particle swarm optimization for the cost estimation of sustainable building projects. The hybrid model was tested using data collected from actual sustainable building projects in Iran. The study concluded that the proposed model is highly accurate and effective for cost estimation of sustainable buildings.

A study conducted by Faiq *et al.* [4] proposed a modern method to predict the residual strength of sustainable self-consolidating concrete exposed to elevated temperature using artificial intelligent technique in sustaining the buildings of Iraq.

## 2 Identification of ANN model variables

The historical data used in this study were collected from the Ministry of Housing and Construction in the Republic of Iraq/Engineering Construction Company, and it includes 75 sustainable projects implemented in Baghdad city for the period from 2015 to 2022. These historical data are shown in Table 1, and can be divided into:

- 1) **Dependent variables:** The cost of the sustainable buildings project is defined as the dependent variable.
- 2) **Independent variables:** These variables are classified into three types of variables and can be measured using the unit of measurements, such as the area of the

**Table 1:** Historical data of sustainable project in Republic of Iraq

| Parameters | $Y$ = cost of project (\$) | $F1$ = area ( $M^2$ ) | $F2$ = height | $F3$ = site conditions |
|------------|----------------------------|-----------------------|---------------|------------------------|
| Max.       | 750,000                    | 985                   | 25            | 3                      |
| Min.       | 250,000                    | 485                   | 15            | 1                      |
| Range      | 500,000                    | 500                   | 10            | 2                      |
| Average    | 695,000                    | 760                   | 20            | 2                      |

sustainable building ( $F1$ ), height of sustainable building ( $F2$ ), and site conditions ( $F3$ ).

### 3 Software ANN selection: GMDH Shell

A professional NN program called GMDH Shell uses ANNs to provide forecasts for several aspects of the construction industry, including cost, productivity, cash flow, and more, intending to make construction forecasting possible for even novice planners and estimators. As shown in Figure 1, the GMDH Shell is made up of numerous plugs connected in a chain. Figure 2 also depicts the design for the GMDH Shell.

## 4 Development of ANNs model

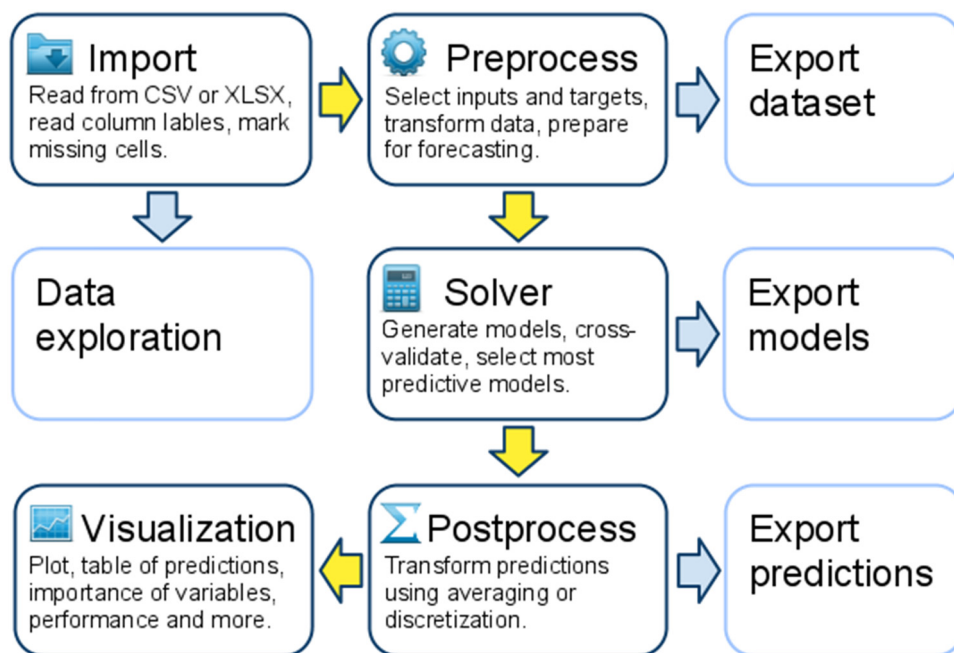
### 4.1 Development of model inputs

ANN is a machine learning model inspired by the structure and functioning of the human brain. An ANN consists of layers of interconnected nodes (neurons) that process and transmit information, each neuron takes in inputs, performs a mathematical computation on those inputs, and then outputs a result. The connections between neurons have weights that determine the strength of the signal being transmitted. The weights are adjusted through a process called training, in which the network learns to recognize patterns in the data.

A single output – the total cost of the sustainable building projects – and just three inputs – the area of the sustainable building ( $F1$ ), height of the sustainable building ( $F2$ ), and site circumstances ( $F3$ ) – were determined by the researcher using the technique (method of previous knowledge).

### 4.2 Data division

The act of separating a dataset into two or more pieces in order to train and test an ANN is known as data division, sometimes known as data splitting. The most popular

**Figure 1:** Component of GMDH Shell.

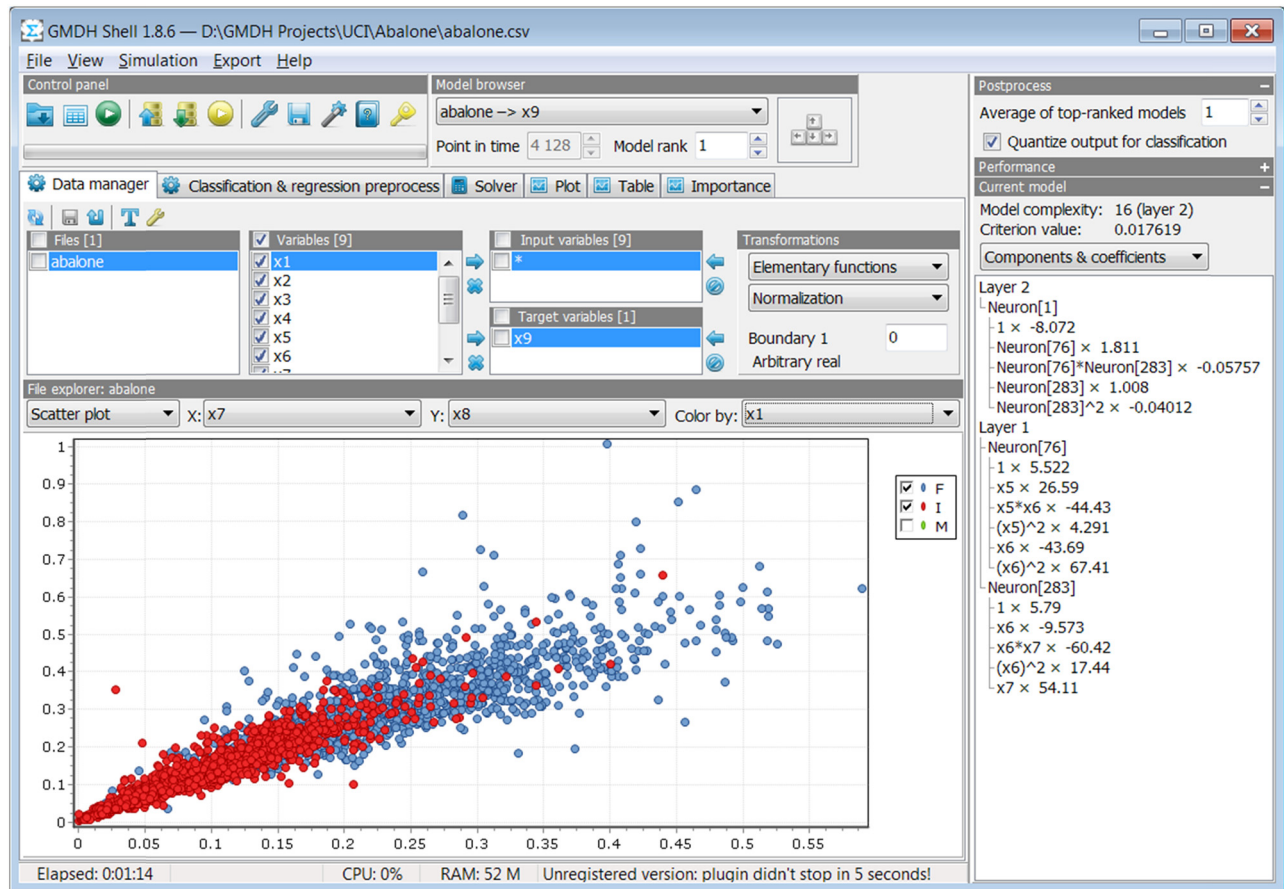


Figure 2: Graphing of GMDH Shell.

strategy is to separate the data into three groups: the training set, validation set, and test set. The test set is used to assess the model's final performance after the validation set has been used to fine-tune the model's hyperparameters. Depending on the size of the dataset and the particular assignment, each set's size will vary. Generally, the training set is the biggest and comprises up to 70–80% of the data, while the validation set and test set each have 10–15% of the data. Data division is crucial in ANN as it aids in preventing overfitting, which happens when the model grows too complicated and matches the training data too closely, leading to worse performance on new, unforeseen data. We may get a more precise assessment of the model's performance on fresh data by measuring it against a different test set.

The training set, testing set, and validation set are subsets of the accessible data that the researcher is using in this study. With the help of the GMDH Shell program, a trial-and-error procedure was used to choose the best division. The findings are described in Table 2, where the best division is calculated as 75% for the training set, 10% for the testing set, and 15% for the validation set ( $r$ ) 93.33%.

### 4.3 Method of division

There are several methods for dividing a dataset into training, validation, and test sets in ANN. The following are three common methods:

- 1) A training set and a test set are created from the dataset using the straightforward hold-out approach. The model is trained using the training set, and its ultimate performance is assessed using the test set. This method's drawback is that it could not accurately reflect the complete dataset, which might lead to inaccurate estimations of model performance.
- 2) k-fold cross validation: This technique divides the dataset into  $k$  equal "folds," or portions. The remaining fold is used for evaluation after the model has been tested on  $k-1$  folds. Each fold serves as the test set once over the  $k$ -time course of this operation. The model's performance is then averaged across the  $k$  folds. Nevertheless, this technique may be computationally costly. It offers a more accurate evaluation of model performance.
- 3) When a dataset is unbalanced, meaning that the classes or labels are not equally represented, stratified

**Table 2:** Effect of data division on NN model for sustainable building project

| Data division (%) |           |           | Training error (%) | Testing error (%) | Coefficient of correlation [r] (%) |
|-------------------|-----------|-----------|--------------------|-------------------|------------------------------------|
| Training          | Testing   | Querying  |                    |                   |                                    |
| 80                | 10        | 10        | 8.89               | 7.40              | 90.22                              |
| 75                | 5         | 20        | 8.99               | 6.76              | 91.32                              |
| 75                | 10        | 15        | 7.98               | 6.84              | 91.34                              |
| 70                | 5         | 25        | 7.99               | 6.92              | 92.56                              |
| 70                | 20        | 10        | 6.73               | 5.81              | 92.57                              |
| <b>75</b>         | <b>10</b> | <b>15</b> | <b>5.62</b>        | <b>4.53</b>       | <b>93.33</b>                       |
| 80                | 10        | 10        | 6.34               | 6.64              | 91.00                              |

Bold values refer to selected data in ANN model as shown in section 4.2 Data Division.

sampling is utilized. This approach ensures that the percentage of each class or label in each set is the same as that in the training and test sets while dividing the dataset into these three categories: training, validation, and testing. By using this technique, it is possible to prevent the model from being biased toward any one class or label and to train it on a representative sample of the data.

Since the impact of employing several divisional options (striped, blocked, and random) was examined in this study, the researcher employed stratified sampling (3). When random division was applied, the performance improved (Table 3).

#### 4.4 NN model architecture

The architecture of an ANN refers to the overall structure of the network, including the number and size of layers, the number of neurons in each layer, and the connections between neurons. The architecture plays a crucial role in the performance of the ANN. Following are some reasons why the ANN model architecture is important:

1) Learning ability: The architecture of an ANN determines its ability to learn complex patterns in the data. A well-designed architecture can improve the network's learning ability and help it to generalize to new, unseen data.

- 2) Computational efficiency: The architecture of an ANN can also affect its computational efficiency. A simpler architecture with fewer layers and neurons may require less computation and training time, but may also result in lower performance. On the other hand, a more complex architecture may require more computation and training time, but may achieve better performance.
- 3) Overfitting: The architecture of an ANN can help to prevent overfitting, which occurs when the model becomes too complex and fits the training data too closely, resulting in poor performance on new, unseen data. A well-designed architecture can help to balance the complexity of the model with its ability to generalize to new data.
- 4) Interpretability: The architecture of an ANN can also affect its interpretability, or the ability to understand how the model makes its predictions. A simpler architecture with fewer layers and neurons may be more interpretable, while a more complex architecture may be more difficult to interpret.

In conclusion, an ANN's design matters because it has an impact on the network's interpretability, computational efficiency, and capacity for learning. The choice of the model architecture is also among the most crucial aspects in the creation of NN models. A variety of networks with various numbers of hidden layer nodes are developed using the software's default parameters (learning rate = 0.2, momentum term = 0.8, and the transfer functions in

**Table 3:** Effect of method division on NN model for sustainable building project

| Data division (%) |         |          | Choices of division | Training error (%) | Testing error (%) | Coefficient of correlation [r] (%) |
|-------------------|---------|----------|---------------------|--------------------|-------------------|------------------------------------|
| Training          | Testing | Querying |                     |                    |                   |                                    |
| 75                | 10      | 15       | Random              | 5.62               | 4.53              | 93.33                              |
| 75                | 10      | 15       | Blocked             | 6.42               | 5.66              | 92.53                              |
| 75                | 10      | 15       | Striped             | 752                | 5.88              | 91.46                              |



hidden and output layer nodes are sigmoid), and the results are summarized in Table 4. The maximum number of nodes equals  $2I + 1$ , where  $I$  is the number of input nodes. In this model, one hidden node was selected (i.e., maximum nodes = 4). The network with a single hidden node is said to be the ideal one.

#### 4.5 NN model equation

The equation for an NN model depends on its architecture and the type of problem it is designed to solve. However, the basic equation for a neural network model can be written as equation (1) [13], this equation used in development of the NN model depended on transfer function, and it is sigmoid function, and for this the following equation was adopted and as indicated by most of the researchers specialized in this field (NN):

$$y = f(w_2 f(w_1 x + b_1) + b_2), \quad (1)$$

where,  $x$  is the input to the network,  $w_1$  is the weight matrix for the connections between the input layer and the hidden layer,  $b_1$  is the bias vector for the hidden layer,  $f$  is the activation function applied to each neuron in the hidden layer,  $w_2$  is the weight matrix for the connections between the hidden layer and the output layer,  $b_2$  is the bias vector for the output layer, and  $y$  is the output of the network.

The input  $x$  is multiplied by the weight matrix  $w_1$ , and the bias  $b_1$  is added to the result. The activation function  $f$  is applied to this sum, and the result is then multiplied by the weight matrix  $w_2$  and the bias  $b_2$  is added. Finally, the output of the network  $y$  is obtained. The choice of activation function  $f$  depends on the problem being solved and the architecture of the network. Common activation functions include sigmoid, rectified linear unit, and hyperbolic tangent (tanh). It is worth noting that this equation only applies to feedforward NNs, which are the most common type of NN. Other types of NNs, such as recurrent NNs and

convolutional NNs, have different equations and architectures [14].

The NN may be converted into a rather straightforward equation using four of the connection weights produced by GMDH Shell's optimum NNs model. The prediction expression for the overall cost for a sustainable construction project is derived by using the four connection weights and the two threshold levels presented in Figure 3.

$$f(x) = T.C = [250,000 \div 1 + e^{4.1544+5.1973 \tanh x}] + 500,000, \quad (2)$$

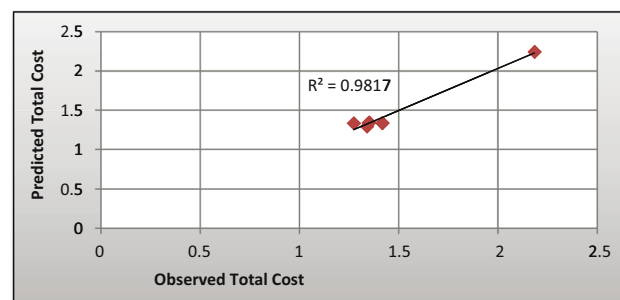
where

$$X = \{2.7891 + (1.2434 \times F1) + (-1.00246 \times F2) + (-2.4434 \times F3)\}. \quad (3)$$

Validation and verification are important steps in the development and deployment of any machine learning model, including ANN. Validation involves evaluating the performance of the ANN on a separate dataset that is not used during training. This helps to assess the generalization ability of the ANN, which is its ability to make accurate predictions on new, unseen data. Typically, the dataset used for validation is split from the original dataset into a validation set and a training set. The ANN is trained on the training set, and its performance is evaluated on the validation set. This process is repeated multiple times using different subsets of the data in a technique called cross validation. Verification involves checking that the ANN is implemented correctly, and that its outputs are consistent with the expected results. This is typically done by comparing the predicted outputs of the ANN to the actual outputs for a set of input data. Verification can also involve checking that the ANN has been trained correctly, by comparing its outputs to a set of known results. Both validation and verification are important steps in the development and deployment of an ANN. Validation helps to ensure that the

**Table 4:** Effective number of nodes on NN model for sustainable building project

| No. of nodes | Training error (%) | Testing error (%) | Coefficient of correlation [r] (%) |
|--------------|--------------------|-------------------|------------------------------------|
| 1            | 5.62               | 4.53              | 93.33                              |
| 2            | 5.82               | 4.77              | 93.11                              |
| 3            | 6.94               | 5.67              | 92.98                              |
| 4            | 6.98               | 5.88              | 91.99                              |



**Figure 3:** Comparison of predicted and observed total cost for validation data for sustainable buildings projects.

**Table 5:** Results of the ANN model for sustainable building project

| Description | NN for model |
|-------------|--------------|
| MAPE        | 13.00%       |
| AA          | 87.00%       |
| $R$         | 93.33%       |
| $R^2$       | 87.104%      |

ANN can make accurate predictions on new, unseen data, while verification helps to ensure that the ANN has been implemented correctly and is behaving as expected. By performing both validation and verification, it can have greater confidence in the performance and reliability of the ANN. Coefficient of determined, coefficient of correlation, average accuracy, and mean absolute percentage measure how well the model outputs match the target value and are given in Table 5. The outcome and advantages of this research encompass the derivation of a precise equation for forecasting the cost associated with sustainable building projects. The reliability of this equation has been demonstrated and validated, as evidenced in Table 5.

## 5 Conclusion

The goal of this research was to estimate the overall costs of constructing sustainable building projects using ANNs since sustainable buildings are now among the most significant projects. ANNs are used to predict the total costs of implementing sustainable building projects. The study emphasizes the importance of sustainable buildings in today's world and the need to find ways to accurately predict the costs associated with such projects. The study used only one ANN model, which took three variables as inputs: the area of sustainable buildings, their height, and the site conditions. The model was trained on a dataset of sustainable building projects and their associated costs. By adjusting the weights and biases of the model during training, the ANN was able to learn the relationships between the input variables and the output variable (i.e., the cost of the project). The result of the study was an equation that predicted the cost of sustainable building projects with a mathematical accuracy of 87.00%. This means that the predicted costs were within 87% of the actual costs. This level of accuracy is quite good, and it suggests that the ANN model was able to capture the underlying patterns and relationships in the data. Overall, the study demonstrates the potential of using ANN models to predict the costs of sustainable building projects. By

accurately predicting these costs, it may be possible to better plan the budget for these projects and encourage their implementation on a larger scale.

**Conflict of interest:** Authors state no conflict of interest.

**Data availability statement:** The data used in this study will be available to others upon request from the authors.

## References

- [1] Al-Zwainy FMS, Awad AL-Somaydai J. Statistical evaluation of value engineering criteria's in sustainable building projects. *Iraqi J Civ Eng.* 2017;11(3):1–14.
- [2] Jaber FKh, Al-Zwainy FMS, Hachem SW. Optimizing of predictive performance for construction projects utilizing support vector machine technique. *Cogent Eng.* 2019;6(1):1–12.
- [3] Al-Zwainy FMS, Neran TH. Application artificial forecasting techniques in cost management (review). *J Eng.* 2016;22(8):1–15.
- [4] Al-Zwainy FMS, Salih SA, Aldikheili MR. Prediction of residual strength of sustainable self-consolidating concrete exposed to elevated temperature using artificial intelligent technique. *Int J Appl Sci Eng.* 2021;18(2):1–15.
- [5] Mijwel AHS, Al-Zwainy FMS, Chiad ASS, Hussein AR, Hayder G. Adopting building information modeling in claims management in construction industry. *J Appl Eng Sci.* 2022;20(4):1152–64.
- [6] Al-Zwainy FMS, Neran TH. Building a mathematical model for predicting the cost of the communication towers projects using multifactor linear regression technique. *Int J Constr Eng Manag.* 2016;5(1):25–9.
- [7] Chua DKH, Kog YC, Loh PK, Jaselskis EJ. Model for construction budget performance-neural network approach. *J Constr Eng Manag.* 1997;123(3):214–22.
- [8] Al-Zwainy FMS, Neran TH. Investigation and evaluation of the cost estimation methods of Iraqi communication projects. *Int J Eng Manag Res.* 2015;5(6):41–8.
- [9] Boussabaine A, Thomas R, Elhag T. Modelling cost-flow forecasting for water pipeline projects using neural networks. *Eng Constr Archit Manag.* 1999;6(3):213–24.
- [10] Lippmann RP. An introduction to computing with neural nets. In: Vemuri V, editor. *Artificial neural networks; Theoretical concepts.* The Computer Society. Washington, USA; 1988. p. 36–54.
- [11] Hammoody O, Al-Somaydai J, Al-Zwainy F, Hayder G. Forecasting and determining of cost performance index of tunnels projects using artificial neural networks. *Int J Comput Civ Struct Eng.* 2022;18(1):51–60. doi: 10.22337/2587-9618-2022-18-1-51-60.
- [12] Zamim SKh, Faraj NS, Aidan IA, Al-Zwainy FMS, AbdulQader MA, Mohammed IA. Prediction of dust storms in construction projects using intelligent artificial neural network technology. *Period Eng Nat Sci.* 2019;7(4):1659–66.
- [13] Al-Zwainy FMS, Aidan IA. Forecasting the cost of structure of infrastructure projects utilizing artificial neural network model (highway projects as case study). *Indian J Sci Technol.* 2017;10(20):1–12.
- [14] Al-Zwainy FMS. The use of artificial neural networks for productivity estimation of finish works for building projects. *J Eng Dev.* 2012;16(2):42–60.

- [15] Tatari O, Kucukvar M. Cost premium prediction of certified green buildings: A neural network approach. *Build Environ.* 2011;46(5):1081–6.
- [16] Jung S, Pyeon JH, Lee HS, Park M, Yoon I, Rho J. Construction cost estimation using a case-based reasoning hybrid genetic algorithm based on local search method. *Sustainability.* 2020;12(19):7920.
- [17] Asgari M, Kheyroddin A, Naderpour H. A proposal model for estimation of project success in terms of radial based neural networks: a case study in Iran. *Civ Eng J.* 2017;3(10):904–19.
- [18] Naghizadeh Vardin A, Ansari R, Khalilzadeh M, Antucheviciene J, Bausys R. An integrated decision support model based on BWM and fuzzy-VIKOR techniques for contractor selection in construction projects. *Sustainability.* 2021;13(12):6933.