

Article

Forecasting Construction Cost Index through Artificial Intelligence

Bilal Aslam ¹, Ahsen Maqsoom ², Hina Inam ³, Mubeen ul Basharat ⁴ and Fahim Ullah ^{5,*}

¹ School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ 86011, USA; ba924@nau.edu

² Department of Civil Engineering, COMSATS University Islamabad, Wah Cantt 47040, Pakistan; ahsen.maqsoom@ciitwah.edu.pk

³ College of Electrical and Mechanical Engineering, National University of Sciences and Technology, Rawalpindi 44000, Pakistan; hinam.ee20ceme@student.nust.edu.pk

⁴ Department of Computer Science and Engineering, HITEC University, Taxila 47080, Pakistan; mubeen.ul.basharat@student.hitecuni.edu.pk

⁵ School of Surveying and Built Environment, University of Southern Queensland, Springfield, QLD 4300, Australia

* Correspondence: fahim.ullah@usq.edu.au

Abstract: This study presents a novel approach for forecasting the construction cost index (CCI) of building materials in developing countries. Such estimations are challenging due to the need for a longer time, the influence of inflation, and fluctuating project prices in developing countries. This study used three techniques—a modified Artificial Neural Network (ANN), time series, and linear regression—to predict and forecast the local building material CCI in Pakistan. The predicted CCI is based on materials, including bricks, steel, cement, sand, and gravel. In addition, the swish activation function was introduced to increase the accuracy of the associated algorithms. The results suggest that the ANN model has superior prediction results, with the lowest Mean Error (ME), Mean Absolute Error (MAE), and Theil's U statistic (U-Stat) values of 0.04, 28.3, and 0.62, respectively. The time series and regression models have ME values of 0.22 and 0.3, MAE values of 30.07 and 28.3, and U-Stat values of 0.65 and 0.64, respectively. The proposed models can assist contractors, project managers, and owners through an accurately estimated cost index. Such accurate CCIs help correctly estimate project budgets based on building material prices to mitigate project risks, delays, and failures.



Citation: Aslam, B.; Maqsoom, A.; Inam, H.; Basharat, M.u.; Ullah, F. Forecasting Construction Cost Index through Artificial Intelligence. *Societies* **2023**, *13*, 219. <https://doi.org/10.3390/soc13100219>

Academic Editor: Manfred Max Bergman

Received: 22 May 2023

Revised: 26 September 2023

Accepted: 9 October 2023

Published: 11 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: building materials; construction cost index (CCI); developing countries; cost estimation; artificial neural network (ANN)

1. Introduction and Background

The construction industry is vital for a country's economy as it creates jobs for skilled workers and supports economic stability. However, accurate cost estimation for construction projects in developing countries is challenging due to uncertainties associated with project durations, material supply, poor documentation, and the overall economy. The associated outdated methods of project cost and duration estimation in developing countries can lead to cost overruns, delays, and financial losses [1]. To mitigate these challenges, sophisticated cost estimation methods that account for material costs and project timelines should be adopted. Accurate cost estimation is essential for justifying a project's economic feasibility during the approval and financial planning stages [2]. Cost overruns can negatively impact the financial viability of projects.

To address the challenge of accurate cost estimation in construction projects, many are calling for localized Construction Cost Indices (CCIs) to forecast the cost of local construction materials, especially in developing countries [3,4]. CCIs are weighted aggregate indices that detail the price changes of construction materials over time, enabling contractors to propose accurate bids during the tender phase and project owners to estimate

project costs and allocate realistic budgets. Contractors rely on the CCI to determine the most competitive pricing strategy for winning the bid while ensuring profitability. Project owners use it to estimate the probable cost of the project.

Several factors should be considered to develop an accurate CCI, including the objective of the index, the choice of cost elements, the weight of each component, and the base year. Different countries have developed and used their own CCIs. An example is the Engineering News Record (ENR) in North America which provides a weighted aggregate index of consistent amounts of Portland cement, structural steel, lumber, and labor [4]. Other examples include the construction labor cost index in Hong Kong, the CCI of Taiwan, and Malaysia [5–7]. However, such indices are rare in developing countries.

Construction costs in developing countries can be volatile due to inflation, which can cause problems for construction projects, especially mega projects. To address this issue, CCIs can be leveraged to forecast future material prices and provide a more accurate estimate of project costs. However, predicting future prices is complex and involves many variables that can be challenging to address in developing countries. In recent years, various techniques have been developed to improve the accuracy of future price predictions, including machine learning algorithms such as Artificial Neural Network (ANN), multiple linear regression, time series analysis, the Delphi technique, simple linear regression, and managerial estimates [8]. These techniques aim to reduce the risk of cost overruns in construction projects and can be classified into three categories: traditional, modern, and quantitative methods. Traditional and modern methods are econometric, while quantitative methods are non-econometric [4].

The traditional method for forecasting construction costs is based on the relationship between dependent and independent variables [9]. Linear regression is an example of this method, which examines the linear relationship between the variables. Although various researchers have used it, the drawback is that all the independent variables must be identified before the method can be applied [10], which is not preferred in construction as it can significantly impact the project cost. Evaluating multiple variables can also make the assessment tricky and susceptible to errors. Lundberg and Lee [11] suggested that linear regression is not accurate enough for forecasting future values.

The modern method for predicting future costs is based on time series analysis, which involves analyzing historical data and internal statistical relationships to forecast future values [12]. This method arranges data points in time order and equal intervals, with time as the independent variable and other variables predicted based on it. Lee et al. [13] proposed a multivariate time series analysis to forecast raw material prices, primarily focusing on the impact of iron core costs on steel items. The authors found their method was 2.3 times more accurate than previous studies in predicting project profitability. Time series analysis was used to predict the CCI of Taiwan, stock returns, probabilistic forecasting, and financial prediction [14,15]. Ashuri and Lu [16] used various time series models to predict CCI and found that seasonal autoregressive integrated moving average (ARIMA) models are the most accurate. Multiple studies have reported that modern time series methods effectively predict construction costs [4]. However, these methods require historical data and involve complicated operations compared to other techniques. Forecasting errors increase exponentially for projects with longer durations, leading to estimating mistakes. Multivariate time series models coupled with deep learning methods such as LSTM can resolve some of these issues for accurate CCI forecasting.

Quantitative methods use computational techniques such as ANN, which simulates human brain learning to yield reliable results with manageable inputs. Various researchers in construction have used ANN for cost estimation, such as predicting highway construction cost increments, tender prices, and cash flow forecasting [17,18]. ANN is suitable for long-term index prediction as it is not solely reliant on past data. Shiha et al. [19] predicted construction material prices in a developing society (Egypt) using ANN. Other studies have also used ANN to achieve higher prediction accuracy [4].

ANN models have been compared to other methods for developing CCIs. Williams [20] predicted changes in CCI value using ANN, multiple regressions, and exponential smoothing, and found that ANNs provide better results than the other techniques. Similarly, Kim et al. [21] developed construction cost models using ANN, case-based reasoning, and regression methods and found that ANN outperformed the other methods.

The case-based reasoning model outperformed ANN regarding long-term use and accuracy vs. time tradeoffs, while the combined model of Kim and Kwak [22] showed improved results with higher accuracy and fewer errors in predicting cost overruns. Elfahham [4] used ANN, regression, and time-series models to predict the CCI of Egypt. The results showed that the time-series model outperformed the other models. Cao, Cheng, and Wu [5] predicted CCI for Taiwan using a self-adaptive structural radial basis neural network intelligence machine (SSRIM). Fan and Sharma [23] used Support Vector Machine (SVM) and Least Squares SVM (LSSVM) for CCI prediction. Wang and Ashuri [24] utilized the K-Nearest Neighbors algorithm (K-NN) and Program Evaluation and Review Technique (PERT), and Moon [25] used the Vector Error Correction Model (VECM) with search query frequencies to forecast CCI.

Pakistan's construction industry is crucial to its economy but faces various challenges, leading to project delays or failures [26]. These challenges include inaccurate project scheduling, inadequate planning, inefficient management, inflation, poor documentation, and lack of proper project cash flows due to inaccurate cost estimates, features common in most developing countries [26,27]. Moreover, despite the cost being a key factor in awarding a construction project, accurate estimation is hindered by multiple variables due to the volatile local economy. These variables include fluctuating currency rates, unstable political situations, high and fluctuating fuel costs, improper project durations and unrealistic delivery milestones, poor documentation, and others [28–30].

Construction material prices in Pakistan are subject to unsystematic revisions due to inflation, resulting in inaccurate cost estimates. Such poor estimates lead to project issues such as delays, poor quality work, customer dissatisfaction, and cost and time overruns [31]. Improper planning and cost estimation have led to the failures of several local mega projects, including the new Islamabad airport, M-5 motorway, Islamabad-Rawalpindi metro bus project, Sher Shah bridge Karachi, Neelum-Jhelum hydropower plant, Kalabagh dam, and Lahore ring road [32].

A localized CCI can help prevent future project failures or delays by accurately predicting project costs. However, the absence of a dedicated agency for CCI development in the country contributes to the deteriorating conditions of the local construction industry. The Pakistan Bureau of Statistics (PBS) currently reports on the costs of various materials but does not explicitly focus on the construction industry. PBS collects the primary data through various surveys, including the Labor Force Survey (LFS), Household Integrated Economic Survey (HIES), Pakistan Integrated Household Survey (PIHS), Pakistan Demographic Survey (PDS), and more. Meanwhile, secondary data, such as foreign trade statistics, industrial statistics, transport and communication statistics, social statistics, agriculture statistics, environment statistics, etc., were obtained from the official records of relevant Ministries/Departments by PBS. Further, it does not consider future fluctuations in construction material rates.

This study proposes a method to develop CCI for developing countries using cost data from 2000–2020 for prime construction elements in Pakistan. Three techniques—Artificial Neural Network (ANN), time series, and regression methods—were utilized to predict future CCI values. The swish activation function was used to enhance the accuracy of the time series model's predictions. The study focuses on the costs of elements used in reinforced concrete buildings, as most local structures are based on reinforced concrete. Developing a local CCI using advanced prediction techniques will enable contractors and owners to manage their budgets better, reduce cost overruns, and assist in developing a more efficient and competitive construction industry, driving economic growth and

development. The research can also serve as a basis for further research worldwide in similar developing countries.

The proposed mechanism for developing CCI is valuable to the literature and pertinent to developing countries. The associated novel approaches, such as utilizing the swish activation function, enhance the accuracy of the forecasting tool. The models can help make accurate project cost estimates, prevent financial losses, and contribute to stabilizing Pakistan's and developing countries' fragile economies.

2. Research Methodology

This study is conducted in three phases, as shown in Figure 1. In the first phase, the literature on building material CCI and its calculation techniques are reviewed and analyzed using scientific literature repositories, including Web of Science, Scopus, and Google Scholar, following Ullah and Al-Turjman [32]. In the second phase, a multi-step approach is taken to extract construction element prices from historical data. Accordingly, a formula is developed to calculate and predict CCI values for the next five years using three techniques: ANN, regression, and time series. In the final phase, the results are compared with each other and with previous studies, limitations are identified, and future research directions are suggested.

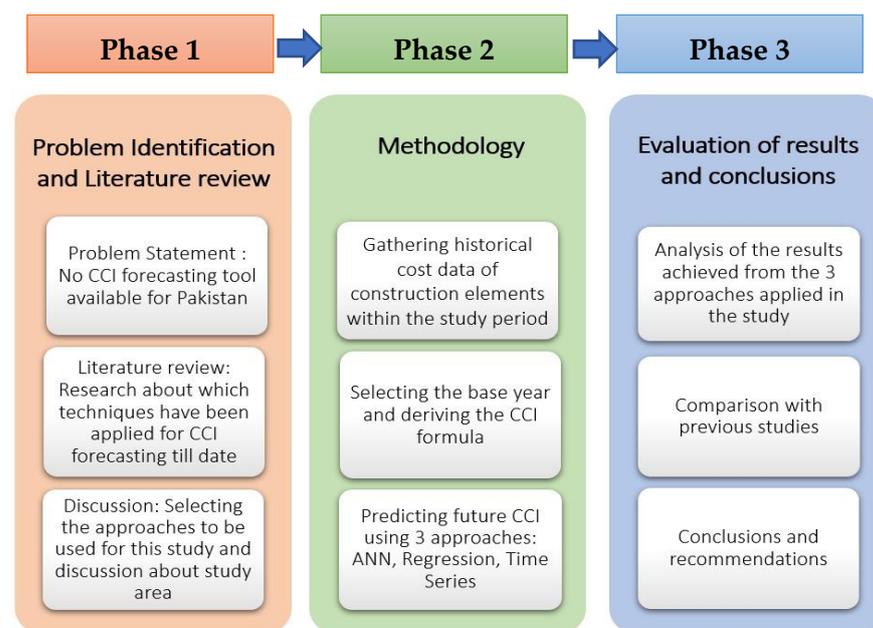


Figure 1. Research methodology for the current study.

The formula for CCI in this study is derived using the prices of key elements of reinforced concrete buildings, such as Portland cement, sand, gravel, brick, and structural steel. A multi-stepped method is adopted to achieve this, as illustrated in Figure 1. A crucial step in this process is to derive the CCI prediction formula and carefully select the base year from the study period, significantly impacting the formula's accuracy. Once the CCI formula is derived, the subsequent step is to predict the CCI for the next five years using the three techniques employed in this study. The following sections provide detailed information about the data sources, CCI calculations, and the three techniques used to forecast the CCI.

2.1. Data Sources

ENR has a well-established method of predicting CCI using "spot prices" for all tracked materials from a single source where the same suppliers are surveyed each month to obtain updated prices [30]. Accordingly, this research used their formula for predicting

building material CCI as the base formula. In addition, modifications were made to the formula to incorporate the common construction materials of Pakistan and similar developing countries. ENR has been used by Kim et al. [33] to develop time series models for the average pipe and labor costs of 20 cities from 1995 to 2016. Further, Kim et al. [34] used ENR data to improve accuracy in predicting city-level CCIs through a combination of ANNs and ARIMA. Similarly, Shrestha [35] used ENR to adjust the operational costs of highway rest area facilities.

The most commonly used construction materials in America are structural steel, lumber, and Portland cement [4]; therefore, the ENR CCI is only related to these elements [36]. However, in Pakistan, lumber is not used as a construction material, and the local construction is mainly based on reinforced concrete and bricks [37]. Accordingly, for calculating CCI, the current study used the cost data of the commonly used construction materials in Pakistan, including Portland cement, sand, gravel, brick, and structural steel [38].

The study collected data on construction material prices from reports published by PBS. These reports contained information on the total quantity of locally produced materials and their corresponding prices. The units used in the study were standard in Pakistan, with prices measured in Pakistani Rupees (PKR) and material quantity in kilograms (kg). To calculate the average unit prices of the materials, the total production quantity per year was divided by the respective cost per year, as shown in Equation (1).

$$\text{Avg Unit Price of Material X} = \frac{\text{Total production quantity of Material X per year (in kg)}}{\text{Cost of Material X per year (in PKR)}} \quad (1)$$

Yearly costs were calculated to account for the fluctuating economy of Pakistan. The study collected data spanning 20 years (2000–2020). The data revealed that steel and cement prices had risen exponentially in recent years, while the costs of bricks, sand, and gravel had increased only slightly.

2.2. The Formula for Calculating CCI

Due to economic and political instability in Pakistan during the study period, and before a major earthquake that hit the country in late 2005 and jolted its economy, 2005 was chosen as the base year for this study. The weights of the construction materials were calculated by determining their quantities based on their unit prices and production values during the base year. Equations (2)–(5) provide the formulas used to calculate the weights of the construction elements for the base year.

Brick weight in 1000 bricks/PKR in the CCI (BW):

$$BW = \frac{BP_b \times 100 / (BP_b + SP_b + CP_b + GP_b)}{BU_b} \quad (2)$$

Steel weight in tons/PKR in the CCI (SW):

$$SW = \frac{SP_b \times 100 / (BP_b + SP_b + CP_b + GP_b)}{SU_b} \quad (3)$$

Cement weight in tons/PKR in the CCI (CW):

$$CW = \frac{CP_b \times 100 / (BP_b + SP_b + CP_b + GP_b)}{CU_b} \quad (4)$$

Sand and gravel weight in m³/PKR in the CCI (GW):

$$GW = \frac{GP_b \times 100 / (BP_b + SP_b + CP_b + GP_b)}{GU_b} \quad (5)$$

where,

BP_b : Production value of bricks at the base year in PKR.

SP_b : Production value of steel at the base year in PKR.
 CP_b : Production value of cement at the base year in PKR.
 GP_b : Production value of gravel and sand at the base year in PKR.
 BU_b : Unit rate of bricks at the base year in PKR/1000 bricks.
 SU_b : Unit rate of steel at the base year in PKR/ton.
 CU_b : Unit rate of cement at the base year in PKR/ton.
 GU_b : Unit rate of gravel and sand at the base year in PKR/m³.

After calculating the weights of each construction material, they are expressed in percentage by multiplying each material's weight by its unit cost within the base year. The index for the remaining years of the study period 2000–2020 is then calculated by multiplying the weight of each construction material by its unit price in the respective study year. The CCI obtained is a proportionate measure of the average cost change over time. The CCI for the study years, along with the corresponding actual price and predicted CCI of future years, can be used to estimate the cost of a product for any future year. This can be accomplished using Equation (6).

$$\text{Cost in year } A = \frac{\text{CCI of year } A}{\text{CCI of year } B} * \text{Cost in year } B \quad (6)$$

year A is the future year for which the product price is calculated. *year B* is the corresponding year for which the actual cost of the product is already known.

2.3. CCI Prediction Techniques

After calculating CCI for the study period using the derived formula, the future CCI was predicted using the three forecasting techniques. Modifications were made to the techniques to increase efficiency and accuracy. Further, a novel approach using the swish activation function was adopted to enhance the time series model. Machine learning studies commonly use three major activation functions: Tanh, Sigmoid, and Rectified Linear Unit (ReLU). ReLU is one of the most widely used techniques [39].

Swish activation function ($f(x) = x \cdot \text{sigmoid}(x)$) has been proposed recently [40]. It is a non-monotonic function that is smooth and consistently performs as well as or better than ReLU on deep networks. It has been successfully applied in various challenging domains, such as image classification and machine translation. Its non-monotonic property makes it stand out from other activation functions. Swish is simple and shares similarities with ReLU, making it easy for practitioners to substitute Swish units for ReLU in any neural network. In this study, Swish was used to improve the accuracy of the predicted results.

2.3.1. ANN

ANN is a popular technique for predicting future values using historical data, using units that function like neurons, and modeling correlations between inputs and outputs [41]. It comprises three layers: input, hidden, and output. Variables are assigned weights based on their impact on the output. This research uses the NeuroXL Clusterizer tool, a self-sorting neural network plugged into Microsoft Excel® (Microsoft Corporation, Redmond, Washington, DC, USA). The calculation process involves two steps: studying and testing the algorithms. To achieve this, the data from 2000–2020 were divided into two parts, with 80% of the data being utilized for training and the remaining 20% being used to test the model.

The Pseudocode for ANN adopted for this research is as follows:

1. Procedure ANN
 - Create an input database
2. Input ← database with all possible variable combinations
 - Train ANNs
3. for input = 1 to end of input, do

```

4   for neurons = 1 to 10 do
5   for repeat = 1 to 10 do
6   Train ANN
7   ANN storage ← save the highest test R2
8   end for
9   end for
10  ANN storage ← save the best prediction ANN depending on the input
11  end for
12  return ANN-Storage → Library with best predicting ANN for every variable combination
13  end procedure

```

It is important to note that this approach can be applied to any dataset worldwide; the model will select the number of neurons and training parameters based on the specific dataset. Thus, it will determine the best model parameters to achieve the highest possible accuracy. However, the model optimizers and building parameters will vary and must be adjusted according to the specific dataset.

2.3.2. Linear Regression

Linear regression is a mathematical technique used to examine the linear relationship between independent and dependent variables [15]. It is commonly used to model the relationship between scalar variables. It has been used to develop cost models since 1970. The current study used the least square linear regression method to forecast the CCI and calculate the associated R² value.

2.3.3. Time Series

The time series method predicts future values based on past data trends using time as the independent variable [18]. It consists of methods that recognize the fundamental setting of the information and make precise estimates. The absence of historical data can be problematic when applying this technique [42]. Univariate time series analysis is used in this study as it requires only one input variable to create the model. This research uses the autoregressive model, which was developed using Python, to predict CCI. The autoregressive model used in this study is shown in Equation (7).

$$X_t = c + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t \quad (7)$$

Here,

p = model's order

$\phi_1 \dots \phi_p$ = model's coefficients

c = constant

ε_t = white noise

The autoregressive model's order (p) represents the amount of historical data from preceding years used by the model to predict future CCI for each year. In this study, an autoregressive model of order 3 (AR³) was used. This was selected as it provided the highest spatial autocorrelation. Following AR³, the highest accuracy was achieved at an order of 6.

3. Results, Analysis, and Discussions

As explained in the method section, the weights of each building construction material considered in this study were calculated using Equations (2)–(5). The results of this calculation are given below:

BW = 0.1192 in 1000 bricks/PKR

SW = 0.0095 in tons/PKR

CW = 0.0427 in tons/PKR

$$GW = 0.126 \text{ in m}^3/\text{PKR}$$

The weights are then converted into percentage weights, as mentioned in the methodology section. The study results indicate that steel and cement have relatively higher percentage weights than sand and gravel. Moreover, bricks have the lowest percentage weight value among all the selected construction materials. The resulting CCI formula derived from the calculated weights of each building construction material is given in Equation (8).

$$\text{CCI (\%)} = 0.1192 \times BU_i + 0.0095 \times SU_i + 0.0427 \times CU_i + 0.126 \times GU_i \quad (8)$$

where,

BU_i : Unit rate of brick for the year i .

SU_i : Unit rate of steel for the year i .

CU_i : Unit rate of cement for the year i .

GU_i : Unit rate of gravel and sand for the year i .

The CCI for the remaining study years was calculated using Equation (8), and the results are presented in Table 1. In total, 960 data points were used in the current study to compute the CCI. This was the sum of values for the four building construction materials, calculated monthly over the 20-year analysis period ($20 \times 4 \times 12$). The CCI values for each year presented in Table 1 were utilized to develop and test the model using the three techniques, i.e., ANN, time series, and regression.

Table 1. The CCI values for the years 2000–2020.

Year	CCI (%)	Year	CCI (%)
2000	719	2011	1090
2001	756	2012	1180
2002	803	2013	1170
2003	840	2014	1185
2004	880	2015	1260
2005	775	2016	1275
2006	890	2017	1299
2007	975	2018	1341
2008	950	2019	1336
2009	998	2020	1345
2010	1022		

Table 1 shows CCI values for 2000–2020, in line with other studies [4], with inflation rates consistently rising except for 2005, 2008, 2013, and 2019, as was the case for most developing countries. The CCI decrease during these years was due to lower steel and cement prices and government policy changes aimed at gaining public support for re-election, with outgoing governments reducing prices in 2008, 2013, and 2018.

New governments formed in 2008 and 2019 introduced price control measures that slightly reduced CCI values. In 2019, these measures were aimed at creating job opportunities and stabilizing the local economy, while in 2005, a catastrophic earthquake caused a price reduction, and price control measures were introduced to help disaster-affected communities rebuild their homes and businesses. Table 1 highlights the fragile and constantly changing nature of Pakistan's economy, with an overall increase in CCI values indicating a gradual rise in inflation over the years, leading to higher construction material prices.

3.1. Calculating and Forecasting CCI

After calculating CCI values for the study period, the future CCI has been forecasted using the three prediction models used in this study. The associated results are discussed below.

3.1.1. Calculating and Forecasting CCI Using ANN

The ANN model achieved a mean error (ME) of 0.04 during the testing stage, making it capable of predicting CCI with acceptable accuracy. Figure 2 shows the CCI prediction profile generated by the ANN model, revealing a significant shift in trend in 2005 due to the earthquake disaster. This resulted in a dip in the CCI curve that year as inflation rates were artificially reduced to assist those affected. However, the curve gradually increased in the subsequent years as measures were taken to stabilize the economy after the artificial control measures were lifted. In 2008, another dip in the curve was observed due to the newly elected government's implementation of price control measures.

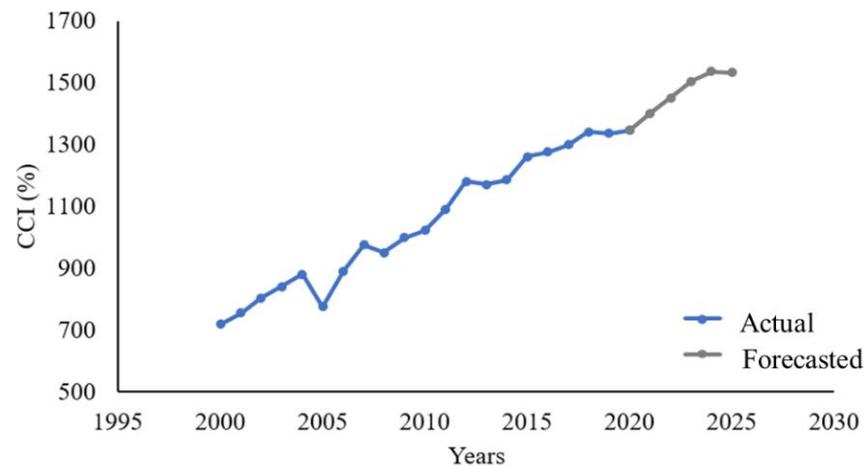


Figure 2. CCI predicted using ANN.

There was also a spike in the CCI pattern in 2011 and 2012 due to the devastating floods of 2010–2011 [43], reducing agricultural output and leading to higher inflation. The 2013 elections saw the government artificially decrease some costs to gain public support, resulting in hikes in CCI in 2015 when conditions normalized. Further, the preparatory cost of the elections and artificial price control negatively impacted the economy, causing price hikes and escalated inflation rates in subsequent years.

In 2020, the world was hit by the COVID-19 pandemic, and Pakistan was no exception. As a result, the country, as was the case for many other developing economies, faced a major economic crisis, leading to an unprecedented rise in inflation rates. This situation can also be seen in the CCI curve, where a sharp spike can be observed. The pandemic caused supply chain disruptions, reduced trade, and decreased demand, ultimately leading to increased inflation rates and adversely affecting the country. Additionally, the government's printing of more currency to support their expenses further contributed to inflation. However, the curve shows a declining trend in recent years, indicating that the economy is gradually stabilizing.

In Figure 2, the blue line shows the CCI values for 2000–2020, and the grey line displays the predicted future CCI values for 2021–2025. The CCI value for 2021 is 1405, which is predicted to increase steadily to 1533 by 2025. This upward trend is due to Pakistan's fragile economic situation, which the COVID-19 pandemic has further exacerbated. The government has taken on more foreign loans to stabilize the economy, but this has also led to increased inflation rates due to stricter debt conditions imposed by organizations such as the International Monetary Fund (IMF). The predicted CCI curve's upward slope reflects this trend. Immediate corrective actions are needed to prevent further deterioration of the economy and wider society.

3.1.2. Calculating and Forecasting CCI Using the Linear Regression Method

This study used linear regression to predict CCI with an R^2 value of 0.9269, indicating dependable outcomes. Figure 3 shows the predicted CCI values for 2021–2025, which

display a nearly linear trend. However, this trend is unexpected given the continuous fluctuations in the historical CCI record for the years 2000–2020, influenced by the ever-changing political and security situation in Pakistan. An entirely linear forecast of future economic conditions seems unrealistic, and thus, the linear regression technique only provides an approximate trend for the predicted future CCI, which may not be as reliable as the predictions made by the ANN model.

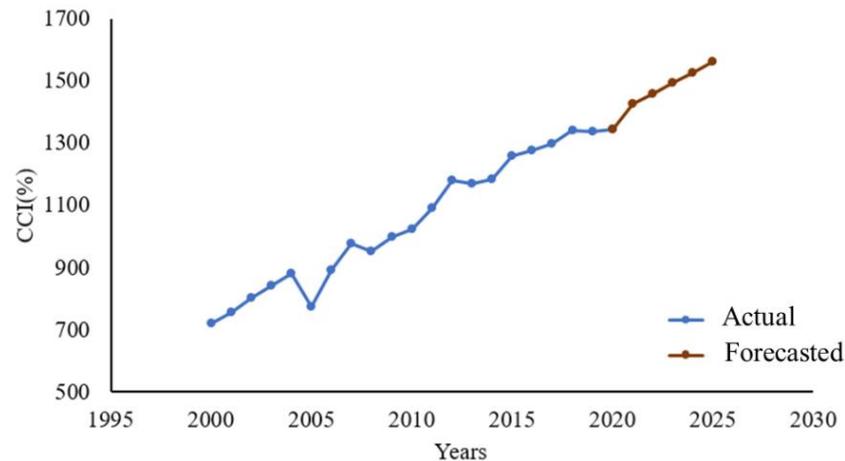


Figure 3. CCI predicted using the regression method.

Other studies, such as those by Adeli and Wu [44], have also concluded that linear regression is not the best approach for estimating CCI. They argue that this technique is unsuitable for developing cost models as it lacks a clear approach to help estimators choose the best model that fits the historical data. Additionally, the regression technique requires identifying all input variables upfront, which can be challenging and may lead to higher errors [45].

3.1.3. Calculating and Forecasting CCI Using Time Series

In this research, the autoregressive model shown in Equation (7) was used to predict CCI through time series analysis. An order of 10 was selected for this model, which means that the current model uses data from the preceding ten years to predict future CCI. Python libraries, including NumPy, pandas, Seaborn, Matplotlib, sklearn.model_selection, sklearn.linear_model, sklearn.metrics, and sklearn.preprocessing, were used to develop all the models in the study. The libraries were used for data handling, visualization, importing, training, testing, splitting the datasets, and calculating the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The linear regression model was imported using sklearn.linear_model library. Further, the conventional ARIMA model was used for time series analysis. The models were verified by calculating ME, MAE, and U-Stat.

Figure 4 shows the predicted CCI trend for 2021–2025 using the autoregressive time series technique. The predicted CCI values start at 1376 for 2021 and increase steadily to 1545 by 2024 due to Pakistan’s unstable economy and high inflation rate. The instability can be attributed to factors such as corruption, repayments of foreign remittances, and aid in fighting the pandemic. The time series model’s prediction of a decline in the CCI value in 2025 is unrealistic, as it does not align with the current situation in Pakistan. There are no scheduled events to suggest a sudden reduction in inflation, and the country’s economy is currently struggling. However, the overall upward trend in the prediction curve seems more realistic, except for the unexpected dip in the last year of the forecast.

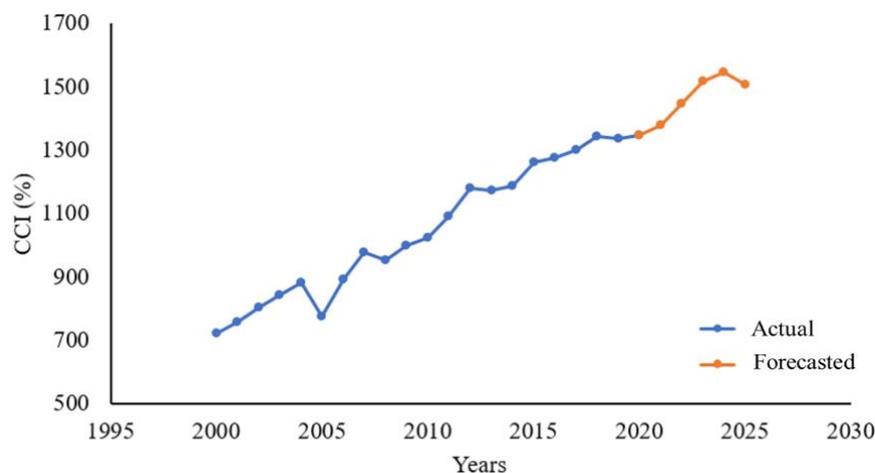


Figure 4. CCI predicted using the time series method.

3.2. Assessment of Forecasting Precision

Results from all prediction models were compared in Table 2 and Figure 5. CCI values for all models showed an increasing trend after 2020. However, the ANN and time series models predict a decrease in 2025, with the time series model showing a significant drop that appears unrealistic. The regression model predicts a continuous increase in CCI values, with a rise from 1345 in 2020 to 1566 in 2025.

Table 2. Forecasted values of CCI using ANN, regression, and time series.

Year	ANN	Linear Regression	Time Series
2021	1405	1423	1376
2022	1462	1467	1444
2023	1509	1487	1517
2024	1540	1529	1545
2025	1533	1566	1506

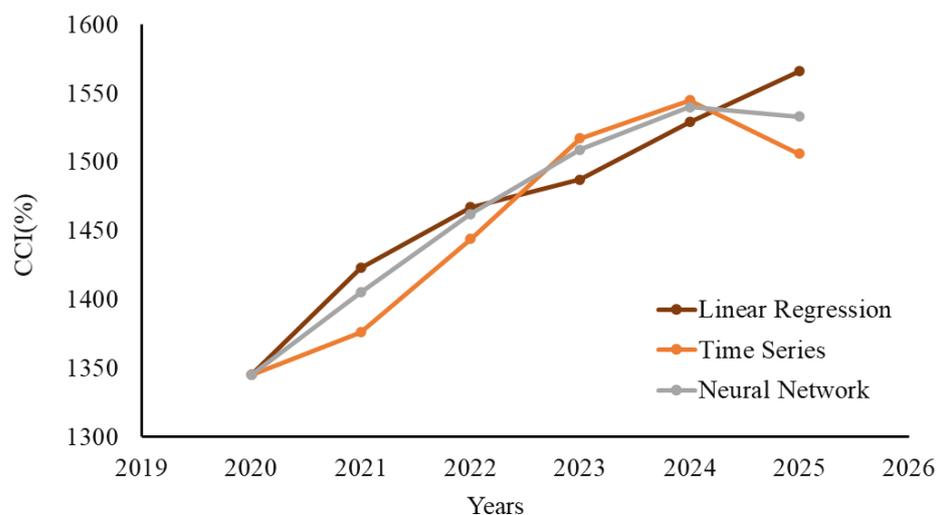


Figure 5. Comparison of predicted CCI obtained from all models.

Figure 5 compares the predicted trends of the CCI for Pakistan from 2021 to 2025 using the regression, ANN, and time-series models. The brown line represents the regression model, the grey line represents the ANN model, and the orange line represents the time-series model. The regression model predicts a nearly linear trend post-2023, which is unlikely given the historical fluctuations in CCI due to the country’s uncertain political and

security situations. In contrast, the ANN and time-series models' predictions align with the past CCI trend from 2000–2020, where the CCI value does not show a consistently linear trend for a long time. The ANN model's results are more accurate and reliable than the time series and regression models, predicting a more plausible dip in CCI values for 2025. The regression model predicts the highest CCI value of 1560 for 2025, while the time-series model predicts the lowest CCI value.

ME, MAE, and U-Stat values were calculated to evaluate the accuracy of the three prediction models used in this study. These measures are commonly used in cost prediction studies and can help determine the best model [46]. ME is the average of all errors in a dataset, while MAE is the average of absolute forecast errors. The smaller the values of these measures, the better the model's prediction ability. U-Stat is a relative accuracy measure that compares forecasted results with minimal historical data, with deviations squared to give more weight to and exaggerate large errors. A U-Stat value of less than 1 suggests that the forecasting technique is superior to guessing, equal to 1 suggests it is as good as guessing, and greater than 1 suggests it is worse than guessing.

Table 3 shows the ME, RMSE, MAE, and U-Stat values for the three prediction models. All models have low error values, indicating the effectiveness of the accuracy-enhancing techniques used in the study. The U-Stat values are less than 1 for all three models, indicating that they are better than guessing. The ANN model has the lowest error values, with a ME of 0.04, RMSE of 34.91, MAE of 28.3, and a U-Stat of 0.62, making it the most accurate model for predicting future CCI compared to the other two methods. Therefore, it can be concluded that the ANN model provides the most reliable results.

Table 3. RMSE and MAE for the three prediction models.

Error	ANN	Regression	Time Series
ME	0.04	0.3	0.22
RMSE	34.91	36.19	36.82
MAE	28.3	28.3	30.07
U-Stat	0.62	0.64	0.65

Table 3 confirms that the time series and regression models have higher error values than the ANN model, indicating that they provide less accurate results. However, the regression model is as reliable as ANN due to their similar MAE values. Overall, the ANN model offers the most dependable results in this study as it has MAE and U-Stat values similar to the regression model but outperforms both in other error measures.

The current study shows superior results compared to published articles. Wang and Ashuri [30] used K-NN and PERT techniques and obtained higher MAE values than the current models indicating superior results of the current study. Similarly, Moon [31] used the VECM technique and obtained an MAE value greater than the current research. The results of this study are aligned with Jafarzadeh et al. [47], who found the ANN approach to be best fitted for predicting the seismic retrofit construction cost. Further, Wilmot and Mei [48] and Wang et al. [49] also identified ANN as the best-fitted model for predicting construction cost. Therefore, it can be inferred that the approach adopted in developing the models in the current study has successfully increased their accuracy.

The swish activation function and other modifications led to highly accurate models in this study, with the ANN model providing the most precise predictions of CCI values. This is useful for contractors and project owners in accurately estimating budgets and avoiding delays and cost overruns, especially in long-term projects undertaken in developing countries. In addition, the ANN model for CCI prediction can also contribute to stabilizing Pakistan and other developing countries' volatile economic situation during and after the pandemic, promoting economic growth and development in the construction industry and megaprojects.

4. Conclusions

This study used ANN, regression, and time series to develop models for predicting CCI in developing countries, using Pakistan as an example. The study found that the regression and time series models had high error values and showed slightly unrealistic predictions. However, the ANN technique was the most effective at predicting CCI, with the lowest error values and the most efficient and accurate predictions. The study suggests that recurrent networks, specifically LSTM networks, would improve accuracy further.

The study's models showed better outputs with less MAE than previous studies due to the introduction of the swish activation function and algorithm modifications during model development. However, the study is limited to Pakistan, and the forecasts were made based on available data for the last 20 years, i.e., 2000–2020. The real-time application of the proposed solution can be checked based on the specific examples from the available data for the year 2021 onwards. Also, field experiments may be conducted. Similar analyses may also be conducted to predict the cost index in other industries. Further, the study did not consider the impact of the 2005 earthquake, assuming that the earthquake was limited to a specific area and that the impact on the entire country's construction material prices may have been minimal. Nevertheless, future studies can cater to such impacts. Despite these limitations, the proposed models are expected to perform well and give reliable predictions in other countries based on theoretical validations in the study.

The study provides a foundational approach for developing CCIs to aid in determining material construction cost fluctuations and benefits projects in the volatile economies of developing countries. Future research can build upon this by implementing deep learning algorithms for CCI prediction and expanding the model to include more construction materials and developed countries. Further testing and validation through field experiments could improve the reliability of the model.

Author Contributions: Conceptualization, B.A. and A.M.; methodology, B.A., A.M., H.I. and M.u.B.; software, B.A., A.M. and H.I.; validation, B.A., A.M., H.I., M.u.B. and F.U.; formal analysis, B.A., A.M., H.I. and M.u.B.; investigation, B.A. and A.M.; resources, F.U.; data curation, B.A., A.M., H.I. and M.u.B.; writing—original draft preparation, B.A., A.M., H.I., M.u.B. and F.U.; writing—review and editing, A.M. and F.U.; visualization, B.A., A.M., H.I., M.u.B. and F.U.; supervision, A.M. and F.U.; project administration, A.M. and F.U.; funding acquisition, F.U. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data, models, or codes supporting this study's findings are available from the lead author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

ANN	Artificial Neural Network
AR ³	Autoregressive model of order 3
ARIMA	Autoregressive integrated moving average (ARIMA)
BW	Weight of Bricks
CCI	construction cost index
CW	Weight of Cement
ENR	Engineering News Record
GW	Weight of Gravel
IMF	International Monetary Fund
K-NN	K-Nearest Neighbors algorithm
LSSVM	Least Squares SVM

ME	Mean Error
MAE	Mean Absolute Error
PBS	Pakistan Bureau of Statistics
PERT	Program Evaluation and Review Technique (PERT)
PKR	Pakistani Rupees (PKR)
ReLU	Rectified Linear Unit (ReLU)
SSRIM	Self-adaptive structural radial basis neural network intelligence
SVM	Support Vector Machine
SW	Weight of Steel
U-Stat	Theil's U statistic
VECM	Vector Error Correction Model (VECM)

References

1. Marandi Alamdari, A.; Jabarzadeh, Y.; Samson, D.; Sanoubar, N. Supply chain risk factors in green construction of residential mega projects—interactions and categorization. *Eng. Constr. Archit. Manag.* **2023**, *30*, 568–597. [[CrossRef](#)]
2. Chen, P.-H.; Nguyen, T.C. A BIM-WMS integrated decision support tool for supply chain management in construction. *Autom. Constr.* **2019**, *98*, 289–301. [[CrossRef](#)]
3. Cao, Y.; Ashuri, B. Predicting the volatility of highway construction cost index using long short-term memory. *J. Manag. Eng.* **2020**, *36*, 04020020. [[CrossRef](#)]
4. Elfahham, Y. Estimation and prediction of construction cost index using neural networks, time series, and regression. *Alex. Eng. J.* **2019**, *58*, 499–506. [[CrossRef](#)]
5. Cao, M.-T.; Cheng, M.-Y.; Wu, Y.-W. Hybrid computational model for forecasting Taiwan construction cost index. *J. Constr. Eng. Manag.* **2015**, *141*, 04014089. [[CrossRef](#)]
6. Saar, C.C.; Chuing, L.S.; Yusof, A.M.; Zakaria, R.; Chuan, T.M. Construction cost index: A case study in malaysia. In Proceedings of the IOP Conference Series: Materials Science and Engineering; IOP Publishing: Bristol, UK, 2019; p. 012059.
7. Wong, J.M.; Chan, A.P.; Chiang, Y.H. Time series forecasts of the construction labour market in Hong Kong: The Box-Jenkins approach. *Constr. Manag. Econ.* **2005**, *23*, 979–991. [[CrossRef](#)]
8. Wong, E.; Swei, O. New Construction Cost Indices to Improve Highway Management. *J. Manag. Eng.* **2021**, *37*, 04021030. [[CrossRef](#)]
9. Johnston, R.; Jones, K.; Manley, D. Confounding and collinearity in regression analysis: A cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Qual. Quant.* **2018**, *52*, 1957–1976. [[CrossRef](#)]
10. Holm, L.; Schaufelberger, J.E. *Construction Cost Estimating*; Routledge: London, UK, 2021.
11. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 4768–4777.
12. Büyükaşahin, Ü.Ç.; Ertekin, Ş. Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition. *Neurocomputing* **2019**, *361*, 151–163. [[CrossRef](#)]
13. Lee, C.; Won, J.; Lee, E.-B. Method for predicting raw material prices for product production over long periods. *J. Constr. Eng. Manag.* **2019**, *145*, 05018017. [[CrossRef](#)]
14. Amigó, J.M.; Hirata, Y.; Aihara, K. On the limits of probabilistic forecasting in nonlinear times series analysis. *Chaos: Interdiscip. J. Nonlinear Sci.* **2016**, *26*, 123114. [[CrossRef](#)] [[PubMed](#)]
15. Zhang, N.; Lin, A.; Shang, P. Multidimensional k-nearest neighbor model based on EEMD for financial time series forecasting. *Phys. A Stat. Mech. Its Appl.* **2017**, *477*, 161–173. [[CrossRef](#)]
16. Ashuri, B.; Lu, J. Time series analysis of ENR construction cost index. *J. Constr. Eng. Manag.* **2010**, *136*, 1227–1237. [[CrossRef](#)]
17. Chakraborty, D.; Elhegazy, H.; Elzarka, H.; Gutierrez, L. A novel construction cost prediction model using hybrid natural and light gradient boosting. *Adv. Eng. Inform.* **2020**, *46*, 101201. [[CrossRef](#)]
18. Verm, A.; Verma, I. Use of artificial neural network in design of fly ash blended cement concrete mixes. *Int. J. Recent Technol. Eng.* **2019**, *8*, 4222–4233. [[CrossRef](#)]
19. Shiha, A.; Dorra, E.M.; Nassar, K. Neural networks model for prediction of construction material prices in Egypt using macroeconomic indicators. *J. Constr. Eng. Manag.* **2020**, *146*, 04020010. [[CrossRef](#)]
20. Williams, T.P. Predicting changes in construction cost indexes using neural networks. *J. Constr. Eng. Manag.* **1994**, *120*, 306–320. [[CrossRef](#)]
21. Kim, G.-H.; An, S.-H.; Kang, K.-I. Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Build. Environ.* **2004**, *39*, 1235–1242. [[CrossRef](#)]
22. Kim, B.-C.; Kwak, Y.H. Improving the accuracy and operational predictability of project cost forecasts: An adaptive combination approach. *Prod. Plan. Control* **2018**, *29*, 743–760. [[CrossRef](#)]
23. Fan, M.; Sharma, A. Design and implementation of construction cost prediction model based on SVM and LSSVM in industries 4.0. *Int. J. Intell. Comput. Cybern.* **2021**, *14*, 145–157. [[CrossRef](#)]
24. Wang, J.; Ashuri, B. Predicting ENR construction cost index using machine-learning algorithms. *Int. J. Constr. Educ. Res.* **2017**, *13*, 47–63. [[CrossRef](#)]

25. Moon, T. Forecasting construction cost index using interrupted time-series. *KSCE J. Civ. Eng.* **2018**, *22*, 1626–1633. [[CrossRef](#)]
26. Maqsoom, A.; Zulfarnain, M.; Irfan, M.; Ullah, F.; Alqahtani, F.K.; Khan, K.I.A. Drivers of, and Barriers to, the Adoption of Mixed Reality in the Construction Industry of Developing Countries. *Buildings* **2023**, *13*, 872. [[CrossRef](#)]
27. Ullah, F.; Thaheem, M.J.; Sepasgozar, S.M.; Forcada, N. System dynamics model to determine concession period of PPP infrastructure projects: Overarching effects of critical success factors. *J. Leg. Aff. Disput. Resolut. Eng. Constr.* **2018**, *10*, 04518022. [[CrossRef](#)]
28. Kamal, A.; Abas, M.; Khan, D.; Azfar, R.W. Risk factors influencing the building projects in Pakistan: From perspective of contractors, clients and consultants. *Int. J. Constr. Manag.* **2022**, *22*, 1141–1157. [[CrossRef](#)]
29. Ahmed, A.; Mehdi, M.R.; Baig, M.A.U.; Arsalan, M. The assessment of sustainability of freight transportation in Pakistan. *Iran. J. Sci. Technol. Trans. Civ. Eng.* **2022**, *46*, 2593–2608. [[CrossRef](#)]
30. Maqsoom, A.; Khan, M.U.; Khan, M.T.; Khan, S.; Ullah, F. Factors influencing the construction time and cost overrun in projects: Empirical evidence from Pakistani construction industry. In Proceedings of the 21st International Symposium on Advancement of Construction Management and Real Estate, The University of Hong Kong, Hong Kong, China, 14–16 December 2016; pp. 769–778.
31. Iftikhar, R.; Müller, R.; Ahola, T. Crises and Coping Strategies in Megaprojects: The Case of the Islamabad–Rawalpindi Metro Bus Project in Pakistan. *Proj. Manag. J.* **2021**, *52*, 394–409. [[CrossRef](#)]
32. Ullah, F.; Al-Turjman, F. A conceptual framework for blockchain smart contract adoption to manage real estate deals in smart cities. *Neural Comput. Appl.* **2023**, *35*, 5033–5054. [[CrossRef](#)]
33. Kim, S.; Abediniangerabi, B.; Shahandashti, M. Pipeline construction cost forecasting using multivariate time series methods. *J. Pipeline Syst. Eng. Pract.* **2021**, *12*, 04021026. [[CrossRef](#)]
34. Kim, S.; Choi, C.-Y.; Shahandashti, M.; Ryu, K.R. Improving accuracy in predicting city-level construction cost indices by combining linear ARIMA and nonlinear ANNs. *J. Manag. Eng.* **2022**, *38*, 04021093. [[CrossRef](#)]
35. Shrestha, K. Cost Comparison of Highway Rest Area Operations: In-House Workforce versus Outsourcing Methods. *J. Constr. Eng. Manag.* **2023**, *149*, 04023062. [[CrossRef](#)]
36. Grogan, T. Dramatic drop in fuel prices starts to affect the markets. *Eng. News Rec.* **2015**, *1*, 1–16.
37. Khan, N.A.; Monti, G.; Nuti, C.; Vailati, M. Effects of Infills in the Seismic Performance of an RC Factory Building in Pakistan. *Buildings* **2021**, *11*, 276. [[CrossRef](#)]
38. Tufail, M.; Ahmad, N.; Mirza, S.M.; Mirza, N.M.; Khan, H.A. Natural radioactivity from the building materials used in Islamabad and Rawalpindi, Pakistan. *Sci. Total Environ.* **1992**, *121*, 283–291. [[CrossRef](#)]
39. Ramachandran, P.; Zoph, B.; Le, Q.V. Searching for activation functions. *arXiv* **2017**, arXiv:1710.05941.
40. Prajit, R.; Zoph, B.; Quoc, V.L. Swish: A self-gated activation function. *arXiv* **2017**, arXiv:1710.059417.
41. Ahmad, M.W.; Mourshed, M.; Rezgui, Y. Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy Build.* **2017**, *147*, 77–89. [[CrossRef](#)]
42. Wu, N.; Green, B.; Ben, X.; O’Banion, S. Deep transformer models for time series forecasting: The influenza prevalence case. *arXiv* **2020**, arXiv:2001.08317.
43. Atif, S.; Umar, M.; Ullah, F. Investigating the flood damages in Lower Indus Basin since 2000: Spatiotemporal analyses of the major flood events. *Nat. Hazards* **2021**, *108*, 2357–2383. [[CrossRef](#)]
44. Adeli, H.; Wu, M. Regularization neural network for construction cost estimation. *J. Constr. Eng. Manag.* **1998**, *124*, 18–24. [[CrossRef](#)]
45. Bode, J. Neural networks for cost estimation: Simulations and pilot application. *Int. J. Prod. Res.* **2000**, *38*, 1231–1254. [[CrossRef](#)]
46. Aslam, A.; Rana, I.A.; Bhatti, S.S. The spatiotemporal dynamics of urbanisation and local climate: A case study of Islamabad, Pakistan. *Environ. Impact Assess. Rev.* **2021**, *91*, 106666. [[CrossRef](#)]
47. Jafarzadeh, R.; Ingham, J.M.; Wilkinson, S.; González, V.; Aghakouchak, A.A. Application of artificial neural network methodology for predicting seismic retrofit construction costs. *J. Constr. Eng. Manag.* **2014**, *140*, 04013044. [[CrossRef](#)]
48. Wilmot, C.G.; Mei, B. Neural network modeling of highway construction costs. *J. Constr. Eng. Manag.* **2005**, *131*, 765–771. [[CrossRef](#)]
49. Wang, Y.-R.; Yu, C.-Y.; Chan, H.-H. Predicting construction cost and schedule success using artificial neural networks ensemble and support vector machines classification models. *Int. J. Proj. Manag.* **2012**, *30*, 470–478. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.