Research Article

Computational Intelligence-Based Structural Health Monitoring of Corroded and Eccentrically Loaded Reinforced Concrete Columns

Somain Sharma, Harish Chandra Arora, Aman Kumar, Denise-Penelope N. Kontoni, Nishant Raj Kapoor, Krishna Kumar, and Arshdeep Singh

1Civil Engineering Department, Punjab Engineering College, Chandigarh 160012, India
2Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India
3Structural Engineering Department, CSIR-Central Building Research Institute, Roorkee 247667, India
4Department of Civil Engineering, School of Engineering, University of the Peloponnese, GR-26334, Patras, Greece
5School of Science and Technology, Hellenic Open University, GR-26335, Patras, Greece
6Architecture and Planning Department, CSIR-Central Building Research Institute, Roorkee 247667, India
7Department of Hydro and Renewable Energy, Indian Institute of Technology, Roorkee 247667, India

Correspondence should be addressed to Denise-Penelope N. Kontoni; kontoni@uop.gr

Received 6 November 2022; Revised 6 February 2023; Accepted 10 February 2023; Published 21 February 2023

Academic Editor: Ahmed Ibrahim

Copyright © 2023 Somain Sharma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Corrosion of embedded steel reinforcement is the prime influencing factor that deteriorates the structural performance and reduces the serviceability of reinforced concrete (RC) structures, especially during earthquakes. In structural elements, RC columns play a vital role in transferring the superstructure’s load to the substructure. The deterioration of RC columns can affect the structures’ overall performance. Hence, it becomes essential to estimate the remaining life of deteriorated RC columns. In the literature, only limited analytical models are available to calculate the remaining life of corroded and eccentrically loaded RC columns. As the number of dependent parameters increases, assessing the residual life of the structural elements and providing a practically applicable suitable model become very complex. Machine learning (ML)-based prediction models are beneficial in dealing with such complex databases. In this article, an ML-based artificial neural network (ANN), Gaussian process regression (GPR), and support vector machine (SVM) algorithms have been applied to estimate the residual strength of corroded and eccentrically loaded RC columns. The performance of the analytical and ML models is accessed using commonly used performance indices, namely, the coefficient of determination ($R^2$), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), a-20 index, and Nash–Sutcliffe (NS). The results of the proposed ANN model have been compared with the existing analytical model to identify the suitability of the best model. Based on performance analysis, the precision of the GPR and SVM models is lower than that of the ANN model. The processed results revealed that the $R^2$ value of the ANN model for training, testing, and validation datasets is 0.9908, 0.9757, and 0.9855, respectively. The MAPE, MAE, RMSE, NS, and a-20 index for all the datasets are 8.31%, 48.35 kN, 72.53 kN, 0.9886, and 0.8978, respectively. The precision of the ANN model in terms of the coefficient of determination is 225.77% higher than that of the analytical model. The sensitivity analysis demonstrates that the compressive strength of concrete plays the most significant role in the load-carrying capacity of corroded and eccentrically loaded RC columns. The proposed ANN model is reliable, accurate, fast, and cost effective. This model can also be used as a structural health-monitoring tool to detect the early damages in the RC columns.
1. Introduction

In RC structures, columns are essential to transfer the superstructure load to the foundation. The deterioration of columns may affect the overall performance and service life of the structure. The damages in the column may reduce the structures’ residual life, and as a result, the structure may show poor performance during seismic activities.

Rapid growth in industrialization and mobility of vehicles in developing South Asian countries is a matter of concern [1]. The leading gases produced by industries are carbon dioxide (CO₂) and sulphur. Carbon from the atmosphere tends to get into RC structures through the voids of the concrete. Impressed CO₂ reacts with iron in the reinforcement to form oxides of iron oxide (FeO₂), commonly referred to as rust or corrosion. The loss of alkalinity due to carbonation of steel bars, loss of alkalinity due to chloride attacks, cracks in concrete, and insufficient cover are some of the other factors that lead to corrosion of embedded reinforcement in concrete structures [2]. The utmost common cause of degradation and the sudden collapse of structures during their service life is corrosion of concrete-embedded reinforcement bars [3, 4].

Steel reinforcement corrosion can be categorized into two types: chloride-induced and carbon dioxide-induced corrosion [5, 6]. Carbonation-induced corrosion leads to uniform steel bar corrosion throughout its length, whereas chloride-induced corrosion leads to the formation of pits. Pits are the critical sections in the reinforcement bar where heavy cross-sectional losses take place due to corrosion [7, 8]. Under the action of loads, these sections form high-stress zones (as the effective area of steel is reduced at pit locations) and may lead to sudden failure of the reinforcement bar at that particular cross-section. In this way, the process of corrosion in RC columns leads to a reduction in the diameter of the reinforcement bar. Even a small loss in the cross-sectional area of a steel bar significantly reduces the load-carrying capacity of the structural element [9]. Corrosion of the stirrup diminishes the overall buckling resistance or load-carrying capacity of the column [10, 11]. Stirrups owing to their closer proximity to the outer atmosphere (compared with the main bars) and lesser cover bear more corrosion than main bars, significantly degrading the shear capacity of the column [12]. Also, the corrosion percentage of stirrups has higher significance than the corrosion level of main steel bars, as it governs the failure mode of columns [13]. The bond strength of concrete is also affected due to the volumetric expansion of corrosion as the formation of rust products, which is about 3–6 times more than the volume of steel-reinforcing bars [2]. This volumetric expansion leads to the cracking of concrete due to internal stresses and spalling of concrete, which further accelerates the corrosion process [2, 8]. More corrosion leads to more and wider cracks on the outer surface of concrete. Corrosion also changes the failure mode of the column from flexure-dominated failure to flexure-shear or pure shear-based failure [14, 15]. All these factors combined slowly degrade the structure, which leads to ultimate failure during the service life of the structure.

In general, concrete present around the steel bar protects it from corrosion as concrete creates an alkaline environment for embedded steel reinforcement which resists corrosion of the steel surface, and corrosion species such as O₂, Cl⁻, CO₂, and H₂O are prevented from entering. Despite these factors that subsidize the prevention of steel corrosion in concrete, cases of RC structure failure have been stated in alarming numbers from certain regions characterized by high humidity, weather, and high wet conditions [3].

Despite the recommendations provided by the codes of practice, steel corrosion is quite evident in structures. To fully understand the behavior of corroded structures, much effort is needed in the estimation of the remaining strength of corroded members varying the degree of corrosion. It is also important for the development of rehabilitation techniques/retrofitting techniques for the enhancement of the service life of the structure.

The available analytical model for the estimation of the remaining strength of eccentrically loaded and corroded RC columns is typically generated using a small number of column databases with constrained ranges of weight loss; thus, their accuracy is questionable. Moreover, analytical approaches are often complex and require additional assumptions, which makes them difficult to apply in actual practice.

Recently, soft computing tools such as artificial intelligence (AI) have gained much importance. ML algorithms have been tried and tested in different domains of civil engineering for optimization and prediction of results [16]. ML models have been successfully producing better results than already existing analytical models for prediction. The robustness of a model depends on several factors such as the problem domain, the size and quality of the dataset, and the choice of hyperparameters. In general, support vector machines (SVMs) and artificial neural networks (ANNs) have been known to perform well in a wide range of problems and have been widely used in various domains. Multivariate adaptive regression splines (MARS) is a non-parametric regression method that can handle nonlinear relationships between features and target variables. It has been shown to perform well in problems with complex relationships. Adaptive neuro-fuzzy inference systems (ANFISs) are a hybrid model that combines the strengths of both fuzzy systems and ANNs. It has been applied to various problems and has shown good performance in some cases. Extreme gradient boosting (XGBoost) is an ensemble learning method that has been widely used in civil engineering applications and has shown strong performance in many applications. The estimated probability of recurrence (EPR) is a method for predicting the probability of recurrence of an event, and it is typically used in earthquake engineering. Therefore, the robustness of a model ultimately depends on the problem domain and the quality of the data, and the model that works best for one problem may not be the best for another. Recently, many data-driven ML-based algorithms have been introduced in the civil engineering sector, and some of the studies are elaborated hereunder [17–22].
Xu et al. [13] predicted the residual load-carrying capacity of RC columns using the ML approach. In their study, 180 cyclic loading test specimens were collected from the literature. The study used six ML models to create the most effective model for the corroded columns’ failure mechanism and load-carrying capacity estimation. Among six algorithms, CatBoost and RF had the highest accuracy of 89% in estimating the seismic mode of failure of corroded RC columns. For the prediction of the axial capacity of a corroded RC column, the CatBoost model has been found to be best with an $R^2$ value of 0.92. Also, the RF algorithm showed a shift in the failure mode of the column during its service life, which was also mentioned by Al-Osta et al. [23].

Imam et al. [24] investigated the residual strength of corroded RC beams using ANNs. Four different algorithms were developed in this study; each algorithm was created using two hidden layer configurations, with two input and one output parameter. The two input parameters were the diameter of the bar and the corrosion activity index of the beam, while the output parameters defined each of the two outputs. The dataset was split into two sets, 70% and 30% for training and testing analyses, respectively. The output presented that the ANN algorithms that were utilized to estimate $C_f$ showed better results than the available analytical models. ANN algorithms had 92% accuracy when compared to experimental equations.

Gupta et al. [25] predicted the mechanical properties of rubberized concrete exposed to high temperatures using ANNs. The total dataset contains 324 specimens, divided into 70%, 15%, and 155 for training, validation, and testing datasets, respectively. The input layer consists of four input parameters, and the hidden layer consists of seven neurons and five output parameters. The statistical parameter of training shows an excellent result for the training dataset, and five output parameters. The statistical parameter of training shows an excellent result for the training dataset, which proves the good performance of the trained ANN dataset. The average correlation coefficient for the developed ANN model is 0.9923 and shows good precision. Tran et al. [26] predicted the punching shear strength (PSS) of a two-way RC slab using ANNs. The study included 218 datasets of the RC slabs that were collected from the literature to develop the ANN model. The dataset was split according to the Gupta et al. [25] study. Several ANN models were proposed (changing the neurons in the hidden layer), and the model with ten neurons in the hidden layer was found to be best according to the value of the MSE and $R$ value. The highest $R$ value was observed for the ANN model of 0.995, which is very close to 1, thus confirming the efficiency of the proposed model.

Mai et al. [27] proposed an ANN-based model for estimating the $f_{ek}$ of concrete with fly ash (FA) and blast furnace slag (BFS). The dataset was randomly split into two parts, training and testing, with a percentage of 70% and 30%, respectively. The best results were obtained at twenty-four neurons in the hidden layer. The proposed ANN model is easy to use and reduces the cost of practical experiments. The output presented that the ANN model is very effective in predicting the concrete compress strength with an $R^2$, MAE, and RMSE of 0.9285, 3.29 MPa, and 4.42 MPa, respectively. The feasibility and consistency of the ANN model in estimating the remaining strength of deteriorated RC columns are studied in this work. This study aims to calculate the remaining strength of deteriorated RC columns under axial loads through artificial intelligence-based and ML-improved ANNs. For this purpose, a database of 137 experimentally tested columns was collected from the literature. The collected columns were corroded up to variant levels of corrosion and tested under eccentric-loading conditions. All the test specimens were corroded using an accelerated corrosion process, which is widely used by researchers in the literature. The impressed currents of all the specimens in the collected database lie within an appropriate range of 0.1 mA/cm² to 4 mA/cm² as mentioned in [28, 29]. Quantification of corrosion levels was carried out based on the percentage weight loss of embedded steel-reinforcing bars, and the specimens were cast, followed by corrosion, and later experimentally tested. Corroded reinforcement bars were prepared according to ASTM G1-03 [30] to calculate the average percentage weight loss.

Najafzadeh et al. [31] predicted the scour in long contractions using ANFIS and SVM algorithms. The average flow velocity, critical threshold velocity of sediment movement, flow depth, median particle diameter, geometric standard deviation, and uncontracted and contracted channel widths are the input factors that influence the scour phenomenon in the modeling of ANFIS and SVM. The performance of the ANFIS model was 1.14% higher than that of the SVM model. According to the results of a sensitivity analysis, the ANFIS model’s scour depth modeling relies heavily on uncontracted channel widths and contracted channel widths. Additionally, the parametric analysis revealed that shear stress resulting from bed sediment motion at the contracted zone was a significant component in illuminating how input parameters affected the scour depth in protracted contractions. To estimate the scour depth around bridge piers, a comparison of the group method of data handling (GMDH) based on genetic programming (GP) and backpropagation (BP) systems was made by Najafzadeh and Barani [32]. The results showed that the GMDH-GP algorithm shows greater complexity and time-consuming nature than GMDH-BP, as GMDH-BP performed better at both the training and testing stages in predicting the scour depth. The sensitivity analysis revealed that the relationship between the pier diameter and flow depth is the most important factor when it comes to the scour depth. Najafzadeh and Azamathulla [33] predicted the scour of pile groups due to waves using the neuro-fuzzy (NF)-GMDH algorithm. The findings suggested that predictions made using NF-GMDH models could be more precise than those made using model trees and conventional equations. The maximum scour depth around piers with debris accumulation was estimated by Najafzadeh et al. [34] using GEP, EPR, and MT models. Comparisons were made between the performance of the testing findings for these models and the conventional methods based on regression techniques. Quantifying and contrasting the MT’s uncertainty prediction with those of other models was performed. In another study, the evaluation of neuro-fuzzyGMDH-based particle swarm optimization (PSO)
was used to predict the longitudinal dispersion coefficient in rivers by Najafzadeh and Tafarojnoruz [35]. The results of the differential evolutionary (DE), MT, genetic algorithm (GA), ANN, and conventional empirical equations were compared with the performance of the NF-GMDH-PSO model. The result analysis revealed that the DE and GA approaches outperformed the other procedures when applied to equations based on an AI methodology. The NF-GMDH-PSO network can be used as an alternative to the successful formulas discussed above because it accurately predicted the longitudinal dispersion coefficient.

For estimating the remaining strength of corroded and eccentrically loaded columns, there is no computer-based model that is currently accessible in the literature. This study is structured as follows: In Section 2, importance of this study has been discussed. The details of the collected database are explained in Section 3, with standardization of the collected database and the performance indices. In Section 4, an available analytical model for the calculation of the residual strength of the deteriorated column is described. Section 5 gives a thorough description of the ML algorithms (ANN, SVM, and GPR) used for the prediction and development of the ANN model. The findings of the proposed predictive model and comparison with the existing analytical model are discussed in Section 6. The conclusion of this study has been summarized in the last section.

2. Research Significance

It is important to determine the strength of the corroded column for the repair and rehabilitation of deteriorated columns. As such, there is no provision for the estimation of the remaining strength of corroded and eccentrically loaded columns in the current codes of practice. Thus, it becomes important to evaluate the strength of the corroded column with precision. However, the complexity of the eccentric loaded and corroded columns makes it very difficult to calculate the axial load-carrying capacity of the RC columns. To address this issue, an ML-based algorithm has been utilized in this study to calculate the axial load-carrying capacity of RC columns. According to the authors’ knowledge, this paper for the first time examines the axial capacity of the corroded and eccentrically loaded RC columns by utilizing ML algorithms (ANN, SVM, and GPR).

3. Methods

The collection database is a very important step in developing machine-learning models. A database of 137 specimens was collected from previous studies to aid in the learning of the proposed ANN algorithm [36–46]. The details of the collected database are shown in Table 1. In total, eleven input parameters such as breadth \(b\), width \(h\), eccentricity \(e\), concrete compressive strength \(f_{ck}\), tensile strength of longitudinal bars \(f_{yL}\), concrete cover \(c\), longitudinal bar diameter \(d_{m1}\), the diameter of lateral bars \(d_{s}\), percentage of reinforcement \(\rho\), percentage weight loss (\(\eta\)), and stirrup spacing \((S_s)\) and the output parameter being the axial load \(P_u\) are collected from previous studies. The selection of appropriate input parameters becomes crucial in the formulation of a machine-learning model, as it significantly affects the results generated by the model [47]. The input parameters should be selected in such a way that they are easily measurable on-site with minimal destruction of the structure, and this allows the easy implication of the proposed model in real-life situations for best performance assessment [48, 49]. Figure 1 shows the distribution of the axial load \(P_u\) against the input parameters, and Figure 2 shows the correlation coefficient \((R)\) plot of all the selected parameters. The details of statistical parameters, such as the minimum, average, maximum, and standard deviation of all the specimens for the collected database, are shown in Table 2. The ranges of the input parameters such as \(b, h, e, f_{ck}, f_{yL}, c, d_{m1}, d_{s}, \rho, \eta, S_s\), and \(P_u\) are from 100 to 250 mm, 100 to 350 mm, 0 to 157 mm, 17.70 to 63.50 MPa, 354.44 to 550 MPa, 15 to 54 mm, 9.20 to 20 mm, 6 to 16 mm, 1 to 3.88, 0 to 20, 50 to 100 mm, and 42.02 to 3530 kN, respectively, as shown in Table 1. Of the eleven input parameters, five were concrete parameters, while the other six were steel parameters and the axial compressive load was the outcome. All these input parameters have been carefully chosen and they are also based on the availability of sufficient data for the best formulation of the proposed model. To propose any ML-based model, the number of required specimens should be greater than ten times the input parameters [50–52]; therefore, the development of the ANN algorithm for the calculation of the residual strength of the corroded column is satisfactory.

The methodology of the present work is shown in Figure 3, which first involves a literature review, followed by the collection and standardization of the database. The selected database was then split into three parts (training, testing, and validation); the values were predicted from the analytical model, and the ANN model was compared based on the performance indices.

3.1. Standardization of Data. Standardization is the process of making the data unitless, and hence, it is easily understood by artificial or machine algorithms. Standardization is the technique in which all values are ranged between two numbers, such as 0 to 1 used in this work. In the absence of normalization, large-value neurons have a much larger effect on training than small-value neurons; this may deviate from the training of the model and result in inaccurate outputs. Therefore, standardization of data was needed in our study. In this work, standardization has been performed as follows [53]:

\[
Z_{\text{normalized}} = \frac{\left( x - x_{\text{min}} \right)}{\left( x_{\text{max}} - x_{\text{min}} \right)},
\]

where \(Z_{\text{normalized}}\) is the normalized outcome, \(x\) is the value to be normalized in the selected dataset, \(x_{\text{min}}\) is the minimum value in the selected dataset, and \(x_{\text{max}}\) is the maximum value in the selected dataset.
<table>
<thead>
<tr>
<th>S. no</th>
<th>Reference</th>
<th>$b$</th>
<th>$h$</th>
<th>$E$</th>
<th>$f_{ch}$</th>
<th>$f_{yt}$</th>
<th>$c$</th>
<th>$d_m$</th>
<th>$d_s$</th>
<th>$\rho$</th>
<th>$\eta$</th>
<th>$s_v$</th>
<th>$P_u$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Xia et al. [36]</td>
<td>240.00</td>
<td>200.00</td>
<td>50–90</td>
<td>25.90</td>
<td>380.00</td>
<td>30.00</td>
<td>20.00</td>
<td>6.00</td>
<td>2.62</td>
<td>0–8.82</td>
<td>100–200</td>
<td>372–970</td>
</tr>
<tr>
<td>2</td>
<td>Azad and Al-Osta [37]</td>
<td>180–220</td>
<td>180–220</td>
<td>30–149.5</td>
<td>30–32</td>
<td>550.00</td>
<td>40.00</td>
<td>18–20</td>
<td>10.00</td>
<td>2.10–3.87</td>
<td>0–15.92</td>
<td>163.00</td>
<td>168–962</td>
</tr>
<tr>
<td>3</td>
<td>Wang and Liang [38]</td>
<td>200.00</td>
<td>200.00</td>
<td>46–157</td>
<td>43.1–45.3</td>
<td>397.50</td>
<td>30.00</td>
<td>18.00</td>
<td>8.00</td>
<td>2.54</td>
<td>0.86–13.52</td>
<td>100.00</td>
<td>223.9–756.5</td>
</tr>
<tr>
<td>4</td>
<td>Maaddawy [39]</td>
<td>125.00</td>
<td>125.00</td>
<td>47–116</td>
<td>28.50</td>
<td>550.00</td>
<td>15.00</td>
<td>10.00</td>
<td>6.00</td>
<td>2.00</td>
<td>0–4.25</td>
<td>125.00</td>
<td>91–215</td>
</tr>
<tr>
<td>5</td>
<td>Qiang and Jin [40]</td>
<td>100.00</td>
<td>100.00</td>
<td>0.00</td>
<td>57.80</td>
<td>354.44</td>
<td>15.00</td>
<td>9.20</td>
<td>6.00</td>
<td>2.65</td>
<td>0–1.20</td>
<td>60.00</td>
<td>545–710</td>
</tr>
<tr>
<td>6</td>
<td>Yi and Lei [41]</td>
<td>250.00</td>
<td>350.00</td>
<td>40–150</td>
<td>42.5–53.7</td>
<td>416–432</td>
<td>25.00</td>
<td>18.00</td>
<td>8.00</td>
<td>1.16</td>
<td>0–3.51</td>
<td>200.00</td>
<td>1310–3530</td>
</tr>
<tr>
<td>7</td>
<td>Shi-ping et al. [42]</td>
<td>150.00</td>
<td>120.00</td>
<td>35.00</td>
<td>52.00</td>
<td>427.00</td>
<td>20.00</td>
<td>10.00</td>
<td>6.50</td>
<td>1.74</td>
<td>0–4.01</td>
<td>100.00</td>
<td>409–580</td>
</tr>
<tr>
<td>8</td>
<td>Diao and Sun [43]</td>
<td>200.00</td>
<td>100.00</td>
<td>130.00</td>
<td>30.7–45.8</td>
<td>360</td>
<td>25.00</td>
<td>12.00</td>
<td>6.00</td>
<td>1.57</td>
<td>0–12.2</td>
<td>100.00</td>
<td>190–260</td>
</tr>
<tr>
<td>9</td>
<td>Chotickai et al. [44]</td>
<td>125.00</td>
<td>125.00</td>
<td>125.00</td>
<td>17.70</td>
<td>392.00</td>
<td>30.00</td>
<td>16.00</td>
<td>6.00</td>
<td>2.89</td>
<td>0–20</td>
<td>125.00</td>
<td>42–55.5</td>
</tr>
<tr>
<td>10</td>
<td>Wu et al. [45]</td>
<td>250.00</td>
<td>250.00</td>
<td>45.00</td>
<td>63.50</td>
<td>425–470</td>
<td>54.00</td>
<td>16.00</td>
<td>16.00</td>
<td>1.28</td>
<td>0–16.5</td>
<td>100.00</td>
<td>1320–2020</td>
</tr>
<tr>
<td>11</td>
<td>Al-Akhras and Al-Mashraqi [46]</td>
<td>250.00</td>
<td>250.00</td>
<td>120.00</td>
<td>47.00</td>
<td>420.00</td>
<td>25.00</td>
<td>10–12</td>
<td>6.00</td>
<td>1.00–1.44</td>
<td>0–13.3</td>
<td>50.00</td>
<td>1320–1586</td>
</tr>
</tbody>
</table>
3.2. Performance Criteria. It is important to assess the reliability of the ML model; for this purpose, six commonly used performance indices have been used. The performance indices used in this study are $R^2$ (correlation coefficient), MAE (mean absolute error), MAPE (mean absolute percentage error), a-20 index, NS (Nash–Sutcliffe) coefficient, and RMSE (root mean square error). These performance indices have been widely used in the literature [54–57].

Mathematical equations used to evaluate all performance indices are mentioned in equations (2)–(7). All these indices have also been used to compare the performance of analytical as well as ML-based models. The coefficient of the determination ($R^2$) value measures the correlation between outputs and targets; an $R^2$ value of 1 means a close relationship, and 0 indicates a random relationship. The mean absolute percentage error is a measure of the prediction error.
Table 2: Statistical properties of the selected dataset.

<table>
<thead>
<tr>
<th>Parameter type</th>
<th>Description</th>
<th>Symbol</th>
<th>Unit</th>
<th>Min</th>
<th>Max</th>
<th>Avg</th>
<th>Std</th>
<th>Input type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concrete</td>
<td>Breadth</td>
<td>b</td>
<td>mm</td>
<td>100.00</td>
<td>250.00</td>
<td>209.20</td>
<td>40.65</td>
<td>Input</td>
</tr>
<tr>
<td></td>
<td>Width</td>
<td>h</td>
<td>mm</td>
<td>100.00</td>
<td>350.00</td>
<td>201.75</td>
<td>55.26</td>
<td>Input</td>
</tr>
<tr>
<td></td>
<td>Eccentricity</td>
<td>e</td>
<td>mm</td>
<td>0.00</td>
<td>157.00</td>
<td>72.34</td>
<td>37.93</td>
<td>Input</td>
</tr>
<tr>
<td></td>
<td>Compressive strength of concrete</td>
<td>$f_{ck}$</td>
<td>MPa</td>
<td>17.70</td>
<td>63.50</td>
<td>38.09</td>
<td>13.43</td>
<td>Input</td>
</tr>
<tr>
<td></td>
<td>Cover</td>
<td>c</td>
<td>mm</td>
<td>15.00</td>
<td>54.00</td>
<td>35.00</td>
<td>10.85</td>
<td>Input</td>
</tr>
<tr>
<td>Steel</td>
<td>Yield strength of steel in tension</td>
<td>$f_{yt}$</td>
<td>MPa</td>
<td>354.44</td>
<td>550.00</td>
<td>462.24</td>
<td>76.71</td>
<td>Input</td>
</tr>
<tr>
<td></td>
<td>Diameter of the main bar</td>
<td>$d_m$</td>
<td>mm</td>
<td>9.20</td>
<td>20.00</td>
<td>16.88</td>
<td>3.58</td>
<td>Input</td>
</tr>
<tr>
<td></td>
<td>Diameter of the stirrup bar</td>
<td>$d_s$</td>
<td>mm</td>
<td>6.00</td>
<td>16.00</td>
<td>9.12</td>
<td>3.32</td>
<td>Input</td>
</tr>
<tr>
<td></td>
<td>% of reinforcement</td>
<td>$\rho$</td>
<td>—</td>
<td>1.00</td>
<td>3.88</td>
<td>2.32</td>
<td>0.78</td>
<td>Input</td>
</tr>
<tr>
<td></td>
<td>% weight loss of the main bar</td>
<td>$\eta$</td>
<td>—</td>
<td>0.00</td>
<td>20.00</td>
<td>5.53</td>
<td>4.90</td>
<td>Input</td>
</tr>
<tr>
<td></td>
<td>Stirrup spacing</td>
<td>$s_v$</td>
<td>mm</td>
<td>50.00</td>
<td>100.00</td>
<td>135.14</td>
<td>41.48</td>
<td>Input</td>
</tr>
<tr>
<td></td>
<td>Compressive load</td>
<td>$P_u$</td>
<td>MPa</td>
<td>42.02</td>
<td>3530.00</td>
<td>773.54</td>
<td>683.11</td>
<td>Output</td>
</tr>
</tbody>
</table>

Figure 2: Correlation coefficient plot (input and output parameters).

Figure 3: Methodology chart of the present work.
accuracy of a forecasting method in statistics [58]. Nash–Sutcliffe close to 1 means the good performance of the model and reduces the accuracy of the model towards zero [59]. In the case of a perfect model with a zero-estimate error variance, the Nash–Sutcliffe efficiency (NSE) equals one (NSE = 1) and vice versa. The a-20 index is a recently introduced statistical engineering measure that can be used to evaluate AI models by displaying the number of samples that suit the estimation values with a 20% variance from experimental values [60–62]. \( R^2 \), NS, and a-20 index values closer to 1 indicate the best correlation between the estimated and the experimental results. The lower the values of errors (MAPE, MSE, and RMSE), the better the performance of the model. These performance indices are based on the statistical assessment of the predicted values and the available experimental values. In addition to that, the scatter index (SI) is also used to assess the performance of developed ML models. The SI is a measure of the dispersion of data points in a dataset around their mean or average. It is a statistical tool used to assess the degree of spread of data and how far apart the data points are from one another. A scatter index of 0 indicates that data points are perfectly clustered around the mean, while a higher scatter index value implies a greater spread of data points and lower clustering around the mean. The specific calculation of the scatter index may vary depending on the type of data and the method used, but it is generally a useful tool in understanding the distribution of a dataset and making inferences based on it [63]. The SI can be computed by dividing the RMSE value by the observed dataset’s average values.

\[
R^2 = \frac{\sum_{i=1}^{N} (E_i - \bar{X})(S_i - \bar{S})^2}{\sqrt{\sum_{i=1}^{N} (E_i - \bar{X})^2 (S_i - \bar{S})^2}},
\]

\[
\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} |E_i - S_i/E_i| \times 100,
\]

\[
\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |E_i - S_i|,
\]

\[
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (E_i - S_i)^2}{N}},
\]

\[
\text{NS} = 1 - \frac{\sum_{i=1}^{N} (E_i - S_i)^2}{\sum_{i=1}^{N} (E_i - \bar{S})^2},
\]

\[
a20 = \text{index} = \frac{m20}{N},
\]

where \( E_i \) refers to the experimental value, \( S_i \) is the predicted value, \( \bar{X} \) and \( \bar{S} \) refer to the average of all experimental and estimated values, the m20 index is the number of specimens, for which the results of \( E_i/S_i \) lie between the range of 0.8 and 1.2, and \( N \) is the total number of specimens in the selected dataset.

### 4. Analytical Model

The model given by Azad for estimating the residual strength of corroded columns is simple to apply; however, the predicted results are inaccurate [37]. In this model, the author introduced a reduction factor, alpha, which is to be applied while calculating the strength of the corroded and eccentrically loaded column. The author proposed a two-step analytical method for the estimation of the strength of eccentrically and corroded loaded columns. The first step primarily involves the calculation of \( P^* \) based on conventional codes of practice, by using the reduced area of cross-sections \( A_r \). All other factors that influence the strength reduction, such as bond degradation, crack damage, and loss of yield strength, are adjusted in a single reduction factor \( \alpha \). The residual strength of the eccentrically loaded and corroded RC columns may be calculated using the equation as follows:

\[
P_{res} = \alpha P^*,
\]

where \( P_{res} \) is the residual compressive strength of deteriorated columns, \( \alpha \) is the combined reduction factor, and \( P^* \) is the strength of an uncorroded column calculated with the help of the conventional code of practice using the reduced area of cross-sections \( A_r \). The reduction factor \( \alpha \) is calculated using the following equation:

\[
\alpha = x_0 \left( \frac{e/h)^{1/3} (D'/D)^{x_2}}{(I_{corr}/I)^{x_1}} \right),
\]

where \( e \) is the eccentricity, \( h \) is the width of the column section, and \( D', D, I_{corr}, \) and \( T \) are the reduced diameter of the longitudinal reinforcing bars, the diameter of the uncorroded reinforcing bar, the impressed current, and the time period, respectively.

In this equation, \( x_0 = 1.323, x_1 = 0.14, x_2 = 1.48, \) and \( x_3 = 0.192 \) were adjusted so that the predicted residual strength could lie within the range of 70% to 110% of the experimental value so as to make the predictions reliable and practical to use.

The calculation of \( P^* \) involves the use of the reduced area of the cross-section \( A_r \) which is calculated using the following equation:

\[
A_r = A_r (1 - \gamma),
\]

where \( A_r \) is the cross-sectional area of the original bar and \( \gamma \) is a variable that may be calculated using the following equation:

\[
\gamma = 2P_rT/D,
\]

where \( P_r \) is the penetration rate of the corroded bars, \( T \) is the time period of corrosion, and \( D \) is the original uncorroded diameter of the reinforcing bars. The penetration rate \( P_r \) can be given by the following equation:

\[
P_r = 0.0003185l_{corr},
\]
where \( I_{\text{corr}} \) refers to the impressed current density.

With the help of equations (11) and (12), \( D \), which is the reduced diameter of the corroded reinforcement bar, can be calculated using the following equation:

\[
D = D - 2P_T T. \tag{13}
\]

5. Artificial Intelligence

5.1. Artificial Neural Network (ANN). ANNs are computational models that use supervised ML algorithms to adjust and self-program in order to produce certain output parameters. Neural networks are advanced computational tools with the capability of solving multidimensional nonlinear problems [64]. The network is mainly comprised of individual elements called nodes linked to each other in a particular pre-defined architecture. ANNs attempt to simulate the logical decision-making capability of the brain to create a correlation between all the input parameters and the corresponding output. Developers have arranged processors or neurons in layers that function in parallel to allow them to accomplish the desired results. ANNs mainly comprise \((n + 2)\) layer structures, where \( n \) represents the number of hidden layers in the network, while the other two layers are the input and output layers [58]. The number of hidden layers and neurons in the hidden layer can be adjusted according to the demand from the neural network. A neural network stands out because of its capability to complete tasks based on logical decisions with infinite permutations and combinations, just like the human brain. There are a few attributes of neural networks, such as adaptive learning, self-organization, real-time operation, and fault tolerances, which set them apart and make them extremely powerful [65]. Neural network essential programs themselves are based on learning datasets and thus do not require performing each and every step manually with human intervention as in the case of conventional computers. Artificial neurons in the hidden layer receive inputs depending on the synaptic weight associated with that neuron, which is nothing but the amplitude or intensity of a connection between two particular neurons [66].

5.1.1. Introduction to ANN. Here, in equation (14), \( z_i \) may be compared with any one input parameter, for instance, the breadth of the column, \( W_j \) and \( k \) are the internal weight and bias through the training process in the hit-and-trial method. The ANN was developed in 1958 by Frank Rosenblatt and was called a perceptron; this formed the basics of ANNs [67]. In the literature, there exist many variations of artificial neural networks such as the recurrent neural network (RNN), feedforward neural network (FFNN), and spiking neural network (SNN). The FFNN algorithm is the simplest among all the other algorithms, which is based on connecting the inputs and outputs through one-way connections between the two. FFNNs can be classified into the following two types: the single-layer perceptron (SLP) and multilayer perceptron (MLP), while an SLP is simpler but is only capable of dealing with nonlinear problems. Multilayer perceptrons are widely used due to their capability to solve complex problems and ability to deal with nonlinear problems.

It is primarily based on the biological neural network, which is responsible for the functioning of the human brain. Biological neurons comprise soma cells, axons, synapses, and dendrites, which are replaced by nodes, inputs, weights, and outputs, respectively. Multiple nodes work together in parallel to estimate the output. The input is obtained in each and every neuron if the input layer, adjusted by bias and weight, is sent to each neuron in the hidden layer one by one. Similarly, the output signal from the hidden layer is sent to every neuron in the output layer. The essential part of the neuron is the learning procedure, which is mainly part of supervised learning. In the case of supervised learning, a dataset consisting of inputs along with the corresponding outputs is fed into the neural network for learning purposes.

Multiple inputs, \( z_1, z_2, z_3, \ldots, z_j \), are sent to the summing junction along with the corresponding weights from each neuron. The summing junction receives the inputs in the form of \( w_1 z_1, w_2 z_2, w_3 z_3, \ldots, w_j z_j \), which is nothing but the dot product of the weight and input matrices. The neuron has its own bias/offset, which is summed up with the dot product of inputs and weights, ultimately forming the predicted output \( y_j \). The basic mathematical equation of working neurons is described in the following equation [24, 68, 69]:

\[
y_j = f \left( \sum_{j=1}^{N} w_j z_j + k \right), \tag{14}\n\]

where \( y_j \) is the desired output of the neuron, \( f(.) \) is a unit step function or transfer function, \( w_j \) is the weight connected with the \( j^{th} \) input to the neuron, \( z_j \) is the input to the neuron, and \( k \) is the offset, bias, or threshold.
models; therefore, only one hidden layer is adopted. At the same time, it increases the complexity of the projected layers. In general, the accuracy of the ANN model is improved with a single hidden layer in between the input and output layers. In comparison to the other two gradient descent approaches, the LM technique is more powerful, fast, and accurate.

5.1.2. Development of the ANN Model. In this study, the dataset used to train the ANN algorithm is already explained in Section 2. The ANN model is trained in the MATLAB 2021a environment on Intel(R) Xeon(R) W-2145 @3.70 GHz, 32 GB RAM system. The input layer consists of eleven input parameters, and hidden layer neurons are adjusted between the range of three and fifteen, as shown in Figure 4. It shows the structure of the developed ANN model. Only one hidden layer gave satisfactory results in terms of R and MSE, and it is easy to develop the ANN model with a single hidden layer in between the input and output layers. In general, the accuracy of the ANN model is improved with each increasing neuron in the hidden layer, but at the same time, it increases the complexity of the projected models; therefore, only one hidden layer is adopted.

The standardized dataset has been randomly divided into three different portions, training, testing, and validation. In this study, the training dataset comprises 70% of the total dataset, which is 95 samples out of a total of 137 total samples. Validation datasets were utilized to measure network generalization, and the validation dataset comprises 15% of the total dataset, which is 21 samples, and similarly, the testing dataset has the same number of samples. The network output was repeatedly altered by changing the neurons’ number from 3 to 15 in the hidden layer.

For training ANNs, three different algorithms are available in MATLAB, namely, scaled conjugate gradient (SCG), Bayesian regularization (BR), and Levenberg–Marquardt (LM). In this case, only the LM algorithm was used because in this algorithm, training automatically stops when generalization stops refining, as specified by an increase in the MSE of the validation samples. In comparison to the other two gradient descent approaches, the LM technique is more powerful, fast, and accurate [70, 71]. The LM backpropagation algorithm has been most widely utilized by researchers to train the ANN model in the past [72, 73]. As this algorithm requires less time than the other two, it was adopted for the training of the neural network.

Based on these outputs, three different values of the correlation coefficient and mean square error for training, validation, and testing were established, as shown in Table 3. With the help of the obtained data, rank analysis was performed for the selection of the best ANN model. Although the rank analysis was performed, other performance indices were calculated to ascertain the selection of the best ANN model.

Performance indices for the unnormalized results were calculated for all thirteen ANN models obtained through training of the ANN model by altering the neurons’ number in the hidden layer, as shown in Table 4. The performance of the different neurons based on MSE for training, testing, and validation datasets is shown in Figure 5. It is depicted in Figure 5(a) that the least MSE was observed when the neurons’ number in the hidden layer was 5 and that the maximum MSE was observed when the neurons’ number in the hidden layer was 8. Figure 5(b) shows the MSE of testing; in the plot, the least MSE was seen when the neurons’ number in the hidden layer was 15 and the maximum MSE was observed when the neurons’ number in the hidden layer was 11. Figures 5(c) and 5(d) show the MSE of validation and all datasets. The ANN model with 13 neurons in the hidden layer was selected as the best ANN model.

5.2. Gaussian Process Regression (GPR). GPR is a Bayesian nonparametric machine-learning method used for regression analysis. It was introduced in the mid-90s by Carl Edward Rasmussen and Christopher K. I. Williams. GPR is based on the idea of modeling a function as a Gaussian process, which is a collection of random variables where any finite number of them has a joint Gaussian distribution. This allows for the modeling of the uncertainty in predictions rather than just the mean value as in traditional linear regression. Formally, GPR involves defining “a prior over functions,” which is then updated based on the observed data in order to obtain a posterior distribution over functions. This posterior is used.

**Table 3: Selection of the best neuron based on R and MSE.**

<table>
<thead>
<tr>
<th>S. no</th>
<th>Neuron</th>
<th>MSE</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tr</td>
<td>Val</td>
<td>Te</td>
<td>Tr</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>0.000345</td>
<td>0.001025</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.000447</td>
<td>0.001617</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0.000294</td>
<td>0.000972</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>0.000508</td>
<td>0.000634</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>0.000470</td>
<td>0.002255</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>0.001054</td>
<td>0.000671</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>0.000353</td>
<td>0.000737</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>0.000365</td>
<td>0.000589</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
<td>0.000329</td>
<td>0.000284</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>0.000822</td>
<td>0.001790</td>
</tr>
<tr>
<td>11</td>
<td>13</td>
<td>0.000365</td>
<td>0.000691</td>
</tr>
<tr>
<td>12</td>
<td>14</td>
<td>0.000332</td>
<td>0.000120</td>
</tr>
<tr>
<td>13</td>
<td>15</td>
<td>0.000534</td>
<td>0.002312</td>
</tr>
</tbody>
</table>

Tr = training, Val = validation, and Te = testing.
to make predictions by computing the expected value of the function at any new input points along with the uncertainty of the prediction. The prediction is a Gaussian distribution with a mean and covariance, which encode the uncertainty in the prediction. In summary, GPR is a powerful method for function approximation and uncertainty quantification and has applications in various fields such as geostatistics, civil engineering, time series analysis, and robotics.

Table 4: Performance indices of the selected neuron and selection of the best neuron.

<table>
<thead>
<tr>
<th>S. no</th>
<th>Neuron</th>
<th>$R^2$</th>
<th>MAPE (%)</th>
<th>MAE (kN)</th>
<th>RMSE (kN)</th>
<th>NS</th>
<th>a-20 index</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>0.9846</td>
<td>10.7206</td>
<td>58.0916</td>
<td>84.8553</td>
<td>0.9845</td>
<td>0.8686</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.9846</td>
<td>10.7206</td>
<td>58.0916</td>
<td>84.8553</td>
<td>0.9845</td>
<td>0.8686</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0.9859</td>
<td>9.4113</td>
<td>50.8701</td>
<td>81.8401</td>
<td>0.9855</td>
<td>0.8759</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>0.9842</td>
<td>11.1574</td>
<td>60.6778</td>
<td>88.9969</td>
<td>0.9829</td>
<td>0.8175</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>0.9740</td>
<td>16.8373</td>
<td>67.5812</td>
<td>114.7180</td>
<td>0.9716</td>
<td>0.8102</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>0.9811</td>
<td>17.8737</td>
<td>83.0355</td>
<td>110.8595</td>
<td>0.9735</td>
<td>0.7883</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>0.9828</td>
<td>11.4321</td>
<td>58.0129</td>
<td>91.4494</td>
<td>0.9819</td>
<td>0.8248</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>0.9880</td>
<td>10.9737</td>
<td>52.6464</td>
<td>74.6723</td>
<td>0.9880</td>
<td>0.8248</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
<td>0.9726</td>
<td>11.8004</td>
<td>51.4994</td>
<td>114.5017</td>
<td>0.9717</td>
<td>0.8613</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>0.9611</td>
<td>34.4218</td>
<td>82.3478</td>
<td>139.0856</td>
<td>0.9583</td>
<td>0.8102</td>
</tr>
<tr>
<td>11</td>
<td>13</td>
<td>0.9887</td>
<td>8.3102</td>
<td>48.3474</td>
<td>72.5273</td>
<td>0.9886</td>
<td>0.8978</td>
</tr>
<tr>
<td>12</td>
<td>14</td>
<td>0.9849</td>
<td>10.9992</td>
<td>51.0480</td>
<td>83.9931</td>
<td>0.9848</td>
<td>0.8686</td>
</tr>
<tr>
<td>13</td>
<td>15</td>
<td>0.9804</td>
<td>23.6002</td>
<td>68.6299</td>
<td>98.2182</td>
<td>0.9792</td>
<td>0.7810</td>
</tr>
</tbody>
</table>

Figure 5: Mean square error plots of (a) training datasets, (b) validation datasets, (c) testing datasets, and (d) all datasets.
5.3. Support Vector Machine (SVM). SVM is a type of supervised learning algorithm used for classification and regression analysis. It was first introduced by Vladimir N. Vapnik and Alexey Ya. Chervonenkis in the mid-90s. SVM is based on the idea of finding a hyperplane that separates the data into different classes in the case of classification or predicts the target value in the case of regression. The hyperplane is chosen such that it maximizes the margin or the distance between the hyperplane and the closest data points, known as support vectors. These support vectors are the key elements that determine the position of the hyperplane. Formally, SVM solves a quadratic optimization problem to find the optimal hyperplane. In the case of linear separability, the hyperplane can be represented by a simple equation, while in nonlinear separability, the data are transformed into a high-dimensional feature space using techniques such as the kernel trick, and the problem is solved in this space. In summary, SVM is a robust and effective machine-learning algorithm, particularly in high-dimensional data and when the classes are well separated. It has found applications in various fields such as text classification, bioinformatics, and computer vision.

6. Results and Discussion

The predicted results obtained from ANNs and analytical models are presented in this section. Performance indices have been used to analyze the performance of the individual models.

The number of datasets has been passed through the analytical model presented by Najafzadeh and Barani [32], and the processed results are depicted in Figure 6. Figure 6(a) shows the graph between the predicted and experimental values calculated from Azad’s model. Most of the specimens lie below the linear-fitting line, which reflects that the values predicted by the analytical model are conservative. For about 64% of the specimens, the predicted residual strength was less than half of the actual experimental value. Figure 6(b) shows the line plot variation in the experimental and predicted results along with errors. The blue line shows the experimental results, the pink-dotted line shows the predicted values, and the diamond symbol on the top red lines accounts for the observed errors. It also shows that the line variation of the predicted results lies below the experimental results, confirming that the analytical model is too conservative in terms of the predicted results. Figure 6(b) also shows that most of the errors lie above the zero line and have a larger deviation in the predicted values.

Figure 7(a) shows the range of errors in the analytical model. This figure shows the number of samples vs. the error in the predicted results through the analytical model. The higher the variation between the experimental data and predicted values, the more will be the errors and vice-versa. Most of the errors observed are positive errors, confirming the conservativeness of the predicted results. While the maximum error value is observed as 2802.65 kN, the lowest observed error value is −108.64 kN. Figure 7(b) shows the frequency distribution histogram of the errors in the analytical model.

The scatter plot, the line plot to show the variation in the experimental and predicted values with respect to their errors, the range of the errors in the analytical model, and the frequency distribution plot of the respective errors of the SVM model are shown in Figures 8(a)–8(d), respectively. In the SVM model, the range of errors is between −327 kN and 399 kN, as shown in Figure 8(d). The coefficient of determination of the SVM model is 0.9828, and the other performance indices are shown in Table 5. Similarly, the scatter plot, the line plot to show the variation in the experimental and predicted values with respect to their errors, the range of the errors in the analytical model, and the frequency distribution plot of the respective errors of the GPR model are shown in Figures 9(a)–9(d), respectively. The coefficient of determination of the GPR model is 0.47%.

![Figure 6](image-url) **Figure 6**: Results of the analytical model. (a) The scatter plot and (b) line plot to show the variation in the experimental and predicted values with respect to their errors.
The highest error accumulation is observed at −70 kN for about 42 specimens, followed by 33 for 190 kN. As only about 28% of the specimens, the predicted values lie within 0.8 to 1.2 times the experimental results. The predicted results also show that the discussed model has high errors and very low accuracy in the prediction of results. Thus, on the basis of the predicted results, this analytical model is very conservative and is responsible for the uneconomical solution for repair and rehabilitation purposes. Thus, it is not helpful to apply this model for practical engineering purposes. The ANN model with 13 neurons in the hidden layer was selected as the best model for the prediction, as this model had the highest coefficient of determination \( R^2 \) of 0.9887, as shown in Table 4. Figure 10(a) represents the plot between the experimental and estimated results for the training dataset, which contains 95 samples from the selected dataset. It also shows a line plot to show variations between the experimental and predicted results. The predicted line (pink-dotted line) almost coincides with the experimental line (blue line), which shows that the predicted values for the training dataset are very close. Similarly, Figures 10(b) and 10(c) show validation and testing datasets, respectively. However, as shown in Figure 10(d), the predicted value line does not completely coincide with the experimental value line for the testing dataset and also shows higher errors than the training and validation datasets. The blue- and pink-dotted lines for the experimental and predicted results almost coincide with each other, showing that the estimated values are very adjacent to the measured values.

The red error line in all the plots also lies very close to the zero-error line, showing minimal errors in the predicted values. Figure 10(d) shows the results for all datasets; the red error line lies close to the zero-error baseline, and the experimental value line and the predicted value line almost overlap each other, showing a close relationship. Figure 11(a) shows the line variation of the error and the number of samples. The red line shows the error in the predicted value and the experimental value. The highest value of the error observed is 231.29 kN, while the lowest observed error is −219.19 kN. Figure 11(b) shows the frequency histogram for the error.

The highest frequency of the error is observed at −20 kN, followed by +20 kN. The red line shows the normal distribution of errors across the plot. All of these plots show that the ANN model has been successful in predicting the results accurately. This model shows minimal errors and a high correlation with experimental values. Now, for the selection of the best model and for the calculation of the remaining strength of corroded and eccentrically loaded columns, the analytical model is compared with the formulated ANN model. The analytical model has performed very poorly in all the performance indices when compared with that of the ANN model. The empirical model showed high errors in the predicted results, doubting its ability for strength prediction in practical engineering. On the contrary, the ANN model showed much confidence in the predicted results with minimal results. The performance indices of the proposed model outperformed those of the existing analytical model. For the analytical model, the \( R^2 \) value is 0.3035, which is 69.30% less than that of the proposed ANN model, which is 0.9887. Similarly, the \( R^2 \) value of the ANN model is 0.12% higher than that of the GPR model. For the proposed ANN model, the MAE value is 8.31%, which is 83.27% less than that for the analytical model, which is 51.94%. Similarly, the MAE value of the ANN model is 7.12% lower than that of the GPR model. The MAE and RMSE values of the ANN are 48.34 and 72.52, respectively, which are 89.85% and 90.4% less than those of the analytical model. The analytical model values of the NS and a-20 index are 0.1653 and 0.2700, respectively, which are 83.27% and 69.92% less than those of the ANN model, which are 0.9886 and 0.8978, respectively. The scatter index of the ANN model is 18.97% and 6% lower than that of the SVM and GPR models, respectively. The details of performance indices for both the analytical model and the proposed ANN model are presented in Table 5. The output of the performance indices depicts that the proposed ANN model has astonishingly outperformed the analytical model. The developed model has better-predicted results, and negligible errors are observed. Thus, it is reliable to use this model in the practical engineering field, as it produces good results.

Figure 12 shows the comparison of the models on the basis of the frequency of errors. The rectangular box shows the frequency of the errors presented in the analytical model (AM). The other four boxes show the frequency of the errors presented in the SVM model (SM), GPR model, and AN model. The highest frequency of the error is observed at −20 kN, followed by +20 kN. The red line shows the normal distribution of errors across the plot. All of these plots show that the ANN model has been successful in predicting the results accurately. This model shows minimal errors and a high correlation with experimental values. Now, for the selection of the best model and for the calculation of the remaining strength of corroded and eccentrically loaded columns, the analytical model is compared with the formulated ANN model. The analytical model has performed very poorly in all the performance indices when compared with that of the ANN model. The empirical model showed high errors in the predicted results, doubting its ability for strength prediction in practical engineering. On the contrary, the ANN model showed much confidence in the predicted results with minimal results. The performance indices of the proposed model outperformed those of the existing analytical model. For the analytical model, the \( R^2 \) value is 0.3035, which is 69.30% less than that of the proposed ANN model, which is 0.9887. Similarly, the \( R^2 \) value of the ANN model is 0.12% higher than that of the GPR model. For the proposed ANN model, the MAE value is 8.31%, which is 83.27% less than that for the analytical model, which is 51.94%. Similarly, the MAE value of the ANN model is 7.12% lower than that of the GPR model. The MAE and RMSE values of the ANN are 48.34 and 72.52, respectively, which are 89.85% and 90.4% less than those of the analytical model. The analytical model values of the NS and a-20 index are 0.1653 and 0.2700, respectively, which are 83.27% and 69.92% less than those of the ANN model, which are 0.9886 and 0.8978, respectively. The scatter index of the ANN model is 18.97% and 6% lower than that of the SVM and GPR models, respectively. The details of performance indices for both the analytical model and the proposed ANN model are presented in Table 5. The output of the performance indices depicts that the proposed ANN model has astonishingly outperformed the analytical model. The developed model has better-predicted results, and negligible errors are observed. Thus, it is reliable to use this model in the practical engineering field, as it produces good results.
of SVM, GPR, and ANN models. Green-, blue-, and turquoise-colored rectangular boxes show the errors in the SVM, GPR, and ANN models, respectively. This graphical representation of the frequency error plot also shows that the ANN model is quite accurate and efficient in calculating the axial load-carrying capacity of the eccentrically loaded and corroded columns.

6.1. Proposed Formulation. Based on the ANN model, the following formulation can be utilized to estimate the axial capacity of the eccentrically loaded and corroded RC columns. The generalized equation of the ANN algorithm is expressed as

\[
P_a = \text{activation function}(W_{HO}Y_i + B_{HO}),
\]

where the activation function used to model this ANN algorithm is "purelin" and expressed as

\[
f(x) = x,
\]

where \(W_{HO}\) are weights between the hidden layer and the output layer and \(B_{HO}\) is the bias between the hidden layer and the output layer.

\[
Y_i = \text{activation function}(W_{IO}X^* + B_{IO}),
\]

where \(X^*\) is the input data and \(W_{IO}\) and \(B_{IO}\) are weights and biases, respectively.

**Table 5: Results of the analyzed models.**

<table>
<thead>
<tr>
<th>Model</th>
<th>(R^2)</th>
<th>MAPE (%)</th>
<th>MAE (kN)</th>
<th>RMSE (kN)</th>
<th>NS</th>
<th>a20-index</th>
<th>SI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Azad and Al-Osta [32]</td>
<td>0.3035</td>
<td>51.9426</td>
<td>476.3429</td>
<td>756.1248</td>
<td>0.1653</td>
<td>0.2700</td>
<td>0.977</td>
</tr>
<tr>
<td>Proposed model (SVM)</td>
<td>0.9828</td>
<td>14.3146</td>
<td>61.5274</td>
<td>89.8912</td>
<td>0.9826</td>
<td>0.8248</td>
<td>0.116</td>
</tr>
<tr>
<td>Proposed model (GPR)</td>
<td>0.9875</td>
<td>8.9473</td>
<td>49.6251</td>
<td>77.4057</td>
<td>0.9871</td>
<td>0.8975</td>
<td>0.100</td>
</tr>
<tr>
<td>Proposed model (ANN)</td>
<td>0.9887</td>
<td>8.3102</td>
<td>48.3474</td>
<td>72.5273</td>
<td>0.9886</td>
<td>0.8978</td>
<td>0.094</td>
</tr>
</tbody>
</table>

**Figure 8:** Results of the SVM model. (a) The scatter plot, (b) the line plot to show the variation in experimental and predicted values with respect to their errors, (c) the range of the errors in the analytical model, and (d) the frequency distribution plot of the respective errors.
Figure 9: Results of the GPR model. (a) The scatter plot, (b) the line plot to show the variation in experimental and predicted values with respect to their errors, (c) the range of the errors in the analytical model, and (d) the frequency distribution plot of the respective errors.
Figure 10: Results of the ANN model: (a) training datasets; (b) validation datasets; (c) testing datasets; (d) all datasets.
Figure 11: (a) Range of the errors in the ANN model. (b) The frequency distribution plot of the respective errors.

Figure 12: Comparison between the errors of analytical and ANN models.

Figure 13: The significance of the input variables in relation to $P_u$. 

Shock and Vibration
where the activation function used to model this ANN algorithm is "tansig" and expressed as

$$\text{Tansig} = \frac{2}{1 + e^{-\alpha x}} - 1,$$

(18)

where $W_{10}$ is the weight between the input layer and the hidden layer and $B_{10}$ is the bias between the input layer and the hidden layer.

The final formulation for the estimation of the deteriorated capacity of the columns is expressed in the following equation:

$$P_u = 0.4886D_1 - 0.0104D_2 + 0.2075D_3 - 0.1986D_4 + 0.9196D_5 - 0.1666D_6 - 0.3422D_7 - 0.8613D_8 - 0.3005D_9 - 0.1212D_{10} + 0.6598D_{11} - 0.1423D_{12} + 0.1837D_{13} - 0.2320.$$  

(19)

The values of $D_1$ to $D_{13}$ are calculated from the following equation:

\[
\begin{bmatrix}
D_1 \\
D_2 \\
D_3 \\
D_4 \\
D_5 \\
D_6 \\
D_7 \\
D_8 \\
D_9 \\
D_{10} \\
D_{11} \\
D_{12} \\
D_{13}
\end{bmatrix} = \begin{bmatrix}
0.1609 \\
0.5684 \\
-0.4059 \\
0.0720 \\
0.7595 \\
-0.6211 \\
-0.5914 \\
0.3182 \\
0.3868 \\
-0.1002 \\
-0.4422 \\
0.6469 \\
0.4266
\end{bmatrix} - \begin{bmatrix}
-0.3467 \\
0.1548 \\
-0.6082 \\
0.4319 \\
-0.4567 \\
-0.6156 \\
-0.2734 \\
-0.7178 \\
-0.0235 \\
-0.3263 \\
-0.8329 \\
-0.2164 \\
0.0889
\end{bmatrix} \cdot \begin{bmatrix}
0.3556 \\
0.4674 \\
0.5498 \\
0.7572 \\
0.5973 \\
0.7328 \\
0.5518 \\
-0.6134 \\
0.2601 \\
0.9599 \\
0.4719 \\
0.2235 \\
0.0896
\end{bmatrix} + \begin{bmatrix}
0.9528 \\
0.3802 \\
0.4971 \\
0.6112 \\
0.4718 \\
0.3196 \\
0.8246 \\
0.5186 \\
0.1749 \\
0.4497 \\
0.4999 \\
0.5956 \\
0.2235
\end{bmatrix} = \begin{bmatrix}
0.5728 \\
0.6666 \\
0.0235 \\
-0.2923 \\
-0.0237 \\
0.3151 \\
-0.2828 \\
0.1749 \\
0.1494 \\
-0.4347 \\
-0.6238 \\
0.6350 \\
-0.0281
\end{bmatrix} = \begin{bmatrix}
0.4247 \\
0.2838 \\
0.0884 \\
0.7382 \\
0.7039 \\
0.6758 \\
-0.6633 \\
0.8939 \\
0.1878 \\
0.5533 \\
-0.6211 \\
-0.6350 \\
0.3019
\end{bmatrix}.
\]

(20)

6.2. Sensitivity Analysis. The sensitivity analysis was conducted to assess the influence of input variables on the output of the ANN model. This was performed by analyzing the current weights in the neural network, as suggested by Milhe [74]. The relative impact of each input variable on the network output was determined by using the linking weights between input neurons, hidden neurons, and output neurons. The equation used for this purpose takes into account the weights ($w_{ij}$ and $v_{jk}$) for each hidden neuron in the network.

$$Q_{jk} = \frac{\sum_{i=1}^{L} w_{ij}/\sum_{r=1}^{N} |w_{ij}| v_{jk}}{\sum_{i=1}^{L} \sum_{r=1}^{N} |w_{ij}| v_{jk}}.$$  

(21)

In the equation, $\sum_{r=1}^{N} |w_{ij}|$ is the sum of the weights linking $N$ input neurons to hidden neuron $j$. $Q_{jk}$ represents the proportion of the influence of the input variable $x_i$ on the output variable in relation to the other remaining inputs. It is important to note that the total sum of the $Q_{jk}$ index for all input variables must be equal to 100%. Figure 13 illustrates the relative impact of the input variables on $P_u$ in this study.

As depicted in Figure 13, the variable compressive strength of concrete and tensile strength of longitudinal bars have the greatest and least impact on $P_u$, with a relative importance of 11.69% and 7.56%, respectively. The input variables that come next in terms of influence on $P_u$ are stirrup spacing ($S_s$), diameter of lateral bars ($d_l$), percentage of reinforcement ($\rho$), concrete cover ($c$), width ($h$), longitudinal bar diameter ($d_m$), eccentricity ($e$), percentage weight loss ($\eta$), and breadth ($b$), respectively.

7. Conclusions

This paper develops three machine learning-based models, namely, SVM, GPR, and ANN models, to forecast the residual strength of the corroded and eccentrically loaded RC columns. A total of 137 experimental datasets of corroded RC columns were gathered from the previous studies to establish the SVM, GPR, and ANN models. The number of neurons in the hidden layer of the ANN model was changed in order to find the most effective ANN-based model. The developed model consisted of eleven input parameters such as breadth ($b$), width ($h$), eccentricity ($e$), concrete compressive strength...
(f_{c, k}), tensile strength of longitudinal bars (f_{y, k}), concrete cover (c), diameter of longitudinal bars (d_{m}), diameter of lateral bars (d_{l}), percentage of reinforcement (\rho), percentage weight loss (g), and stirrup spacing (s_{v}) and one output parameter (eccentric compressive load). The reliability of the selected ANN algorithm for the estimation of the residual strength of the deteriorated RC column is compared with that of the existing analytical model and developed SVM and GPR models to obtain the best model:

(i) The analytical model has poor precision in estimating the residual strength of the deteriorated RC columns, as only 28% of the predicted values lie in the range from 0.8 to 1.2 times the predicted value.

(ii) SVM and GPR models have coefficients of determination of 0.9829 and 0.9875, respectively. The MAPE value of the GPR model is greater than that of the SVM model, demonstrating the GPR model’s superiority.

(iii) The ML-based ANN model has good precision in estimating the residual strength of corroded and eccentrically loaded columns, as over 88% of the predicted values lie in the range from 0.8 to 1.2 times the predicted value.

(iv) The analytical model was found to be highly conservative in predicting the values; as for about 64% of specimens, the predicted values were less than 50% of the experimental values.

(v) The performance of the ANN model is superior to that of the SVM, GPR, and current analytical models, according to performance indices and graphical representation.

(vi) The analysis of sensitivity illustrates that the compressive strength of concrete has the most substantial impact (11.69%) on the ability of corroded and eccentrically loaded RC columns.

The proposed ANN model in this work is efficient in estimating the residual strength of deteriorated RC columns, but more datasets should be employed in future research works to achieve the outstanding precision of machine-learning algorithms. Also, a majority of the test specimens are scaled-down columns and may not accurately reflect the parameters involved in real-sized specimens; therefore, the experimental outcomes of large-scale corroded RC columns might be utilized to encourage the use of machine learning-based models. Other ML techniques should be researched, and the impact of input parameters on model performance should be analyzed for the development of more resilient and efficient models. The results of the proposed ANN model are only valid for the data falling in the input and output parameters, which is the limitation of this study. This model can also be used as a structural health-monitoring tool to detect the early damages in RC columns.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

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

References


