Research Article

Machine Learning-Based Fragility Assessment of Reinforced Concrete Buildings

Abdur Rasheed 1, Muhammad Usman 2, Muhammad Zain 2, and Nadeem Iqbal 3,4

1Department of Civil Engineering, MY University, Islamabad, Pakistan
2School of Civil and Environmental Engineering, National University of Sciences and Technology, Sector H-12, 44000 Islamabad, Pakistan
3Department of Computer Science, Abdul Wali Khan University Mardan, Mardan, Khyber Pakhtunkhwa 23200, Pakistan
4Division of Computer Science, Mathematics and Science, Collins College of Professional Studies, St. John’s University New York, New York City, NY 11439, USA

Correspondence should be addressed to Muhammad Usman; m.usman@nice.nust.edu.pk and Nadeem Iqbal; nikhan@awkum.edu.pk

Received 26 May 2022; Revised 20 July 2022; Accepted 22 July 2022; Published 25 August 2022

1. Introduction

Earthquake is one of the major natural disasters which has caused large destruction in many parts of the world in terms of economic and human loss [1, 2]. According to World Health Organization (WHO), earthquakes killed about 747,234 people and caused an economic loss of nearly 661 billion dollars worldwide from 1998 to 2017 [3]. Most of the fatalities in earthquakes are due to the collapse of buildings and infrastructures in seismic-prone countries [4]. Pakistan is also located in a highly active earthquake belt where many devastating earthquakes have jolted the region in the past. Some of the disastrous earthquakes include the 1945 Makran earthquake (Magnitude >8.0), the 1931 Mach earthquake (Magnitude = 7.3), the 1935 Quetta earthquake (Magnitude = 7.4), the 1974 Pathan earthquake (Magnitude = 6.0), and the 2005 Muzaffarabad earthquake (Magnitude = 7.6) [5–7]. In Pakistan, the poor infrastructure design and construction have caused dead consequences, especially the collapse of school buildings in the 2005 earthquake which has caused the death of 19,000 students and hugely affected the economy of the country [8]. Nearly 2/3rd of the academic buildings were completely collapsed [9]. In Pakistan before the 2005 earthquake, the majority of the constructions were designed without incorporating the detailed seismic design provisions and use of...
nonengineered construction. Old practices were carried out in the design of buildings and bridges [10–14]. After the 2005 earthquake, the government emphasized the establishment of seismic codes which led to the Building Code of Pakistan [15, 16]. Researchers have performed analytical studies that a large magnitude earthquake can happen in Pakistan in the future [17]. Therefore, it is important to evaluate the structures against possible earthquakes.

The fragility curves, like many other techniques, can be used for the evaluation of the seismic vulnerability of buildings [18]. Fragility is the probability that the structural response (demand) due to earthquake ground motion exceeds the capacity of the structure [19]. These curves can be beneficial for retrofitting the structures and planning for future possible earthquakes in highly seismic zones such as Pakistan. Fragility curves can be established using judgmental, empirical, or analytical procedures [20, 21]. However, analytical techniques are the most commonly practiced methods using available past recorded earthquake data and structural details [22]. Researchers have established the fragility curves for 55-story core-wall RC buildings in the Philippines using two damage limit states: damage control and collapse prevention [23].

Extensive research has already been carried out on the development of fragility curves for buildings, bridges, and dams using numerical techniques [24]. However, limited literature is available to develop fragility curves using soft computing techniques. Moreover, it has also been observed that nonlinear fragility curves demand substantial analysis time and itself is a resource extensive job [19]. This article addresses the previously mentioned issues, machine learning techniques; i.e., ANNs and GEP were deployed to predict the results of conventional approaches [25, 26].

2. ANN and GEP Models

2.1. ANN Model. Artificial Neural Network (ANN) is a machine learning (ML) technique that works on the concept analogous to the human brain system. It was developed as a simplified model of brain function. There are billions of neurons (brain cells) that are interconnected in the human neural system, and these cells allow conveying signals from one neuron to the neighbouring ones [27]. It is used for solving complex problems as a powerful regression tool in various fields. The important feature of ANN is its ability to learn from experience and examples and to predict meaningful solutions. There are numerous applications of ANNs such as prediction, classification, data association, and filtering of data [28]. In the field of engineering, it has got a significant role in terms of time saving and computational cost [29, 30]. Limited research has been done in the field of structural engineering based on Artificial Intelligence (AI) techniques. However, researchers are interested to incorporate AI techniques in structural engineering due to their low computational cost and time saving. Huang and Huang have conducted a study on seismic fragility analysis of RC bridges using ANN [31]. Liu and Zhang have employed the ANN-based methodology for developing the fragility curves of steel frames [29]. Similarly, Neves et al. have assessed steel bridges based on structural health monitoring and optimization of civil engineering structures [32, 33].

The ANN architecture consists of three or more layers: an input layer, hidden layer, and output layer. Input data (theoretical, empirical, experimental, or combination of the three) is fed into the input layer on which the network learns and recognizes, and then it is transmitted to the hidden layer. The neuron has a particular weight, bias value, and an activation function. The output layer predicts the target or final results based on various algorithms [34, 35]. The neuron in the hidden layer collects the input data from the input layer and transmits it to the output layer as given in the following equation [26]:

\[ h_i = \left\{ \sum_{j=1}^{m} w_{ij} x_i + b_j \right\} \quad j = 1, 2 \ldots n \]

\[ i = 1, 2, \ldots m. \]

Here \( w_{ij} \) is the weight coefficient, \( x_i \) indicates the \( i \)th input variable, \( h_j \) shows the output of the \( j \)th hidden neuron in the hidden layer, \( b_j \) is the bias of the \( j \)th neuron in the hidden layer, and \( f^p \) represents an activation function.

There are various types of Artificial Neural architecture. The most common types of ANN are comprised of Single Layer Perceptron, Multilayer Perceptron (MLP), Radial Basis-Network (RBN), and Recurrent Neural Network (RNN), the Probabilistic Neural Network (PNN), and Cascade Correlation Neural Network [36].

One of the most commonly used algorithms is backpropagation, which can reduce the error function. Several error matrices are used for solving engineering problems, and selecting the appropriate error matrix is mandatory [37]. The predicted value minus the actual output value is known as the error function, and the mean square error (MSE) is employed for the determination of this difference as follows mathematically:

\[ \text{MSE} = \frac{\sum_{i=1}^{N} (O_i - \bar{O}_i)^2}{N}, \]  

where \( N \) shows the number of test data, \( O_i \) is the \( i \)th predicted value, and \( \bar{O}_i \) is the \( i \)th desired target value. ANN model halts once the MSE is lower than the target value; otherwise the training continues and the values of weights and biases will keep being updated. Another important decision is the selection of an appropriate activation function while creating an ANN model. The most widely used activation functions are log-sigmoid and tan-sigmoid functions, which can be either linear or nonlinear. The nonlinear function enhances the nonlinear behaviour of the available data, so sigmoid, nonlinear function is adopted in this study as shown in the following equation [31]:

\[ f(x) = \frac{1}{1 + e^{-ax}}. \]

Here \( a \) is the sigmoid slope, \( a = 1 \). The output data can be computed in the input layer as follows:
\[
O_k = \left( \sum_{j=1}^{n} X_j \cdot Y_j + C_j \right),
\]

where \(X_j\) shows the weight component between the output layer and the \(j^{th}\) neuron in the hidden layer, while \(C_j\) is the bias coefficient in the output layer.

The Levenberg-Marquardt (LM), also known as the damped least squares method, is used for the computation of nonlinear least-square problems [38]. LM can mathematically be represented by the following equation [39]:

\[
y_{k+1} = y_k - \left[ \Delta^T (y_k) \Delta (y_k) \mu I \right]^{-1} \Delta^T (y_k) \nu(y_k),
\]

where \(y_{k+1} =\) learning function, \(\Delta(y_k) =\) Jacobian matrix, and \(V(y_k) =\) nonlinear function.

Bayesian Regularization, also known as backpropagation, is another ANN training algorithm that updates the weights and biases as per the LM optimization [40]. One-step secant (OSS) algorithm connects the quasi-Newton approach and conjugate gradient algorithms. It does not require computing the inverse matrix while calculating the new search direction. However, more computation is needed using the OSS algorithm as compared to conjugate gradient methods [41].

### 2.2. GEP Model

Genetic Programming (GP) is another machine learning technique that is an extended form of Genetic Algorithm (GA), introduced by Koza [42]. The GA was developed by Holland based on inspiration from Darwin’s theory of evolution [43]. GP and GA are two different approaches based on solution representation. The GA gives linear strings or chromosomes of fixed length, while GP gives nonlinear strings of different sizes and shapes (parse trees). In 1999, Candida Ferreira introduced Genetic Programming, a technique that produces computer programs for modelling any phenomenon by mimicking the biological evolution [44].

In the GEP technique, the entities are used as linear strings having a fixed length (genome or chromosome), which are subsequently represented as nonlinear strings of different shapes and sizes (expression trees). Despite its fixed length, the GEP chromosome can encode expression trees (ETs) of varying sizes and shapes [44]. There are two major components in GEP, namely, chromosome or genome and expression tree (ETs). The chromosomes comprise a linear string of fixed length having one or more genes. Each gene itself is a string of fixed length, including arithmetic operations, fitted length parameters as a function sets, and a terminal set of constants. The genes in GEP consist of a head and a tail. The head includes symbols that show both functions and terminals while the tail comprises terminals only [45, 46].

The significance of the research is to employ the ML approach for the prediction and comparison of drift ratios of school buildings with the results obtained through the conventional approach. The optimized AI architecture and functions were used to develop the fragility curves. Similarly, the GEP model, including the expression trees, and equations were utilized for forecasting the drift values, which were then used in the development of fragility curves. The performance of ANN and GEP models was evaluated using MSE.

### 3. Methodology

This study focuses on the ANN and GEP-based fragility assessment of reinforced concrete school buildings located in seismic zone-4 of Pakistan that were designed after Muzaffarabad earthquake. Field data were collected from the school buildings through field visits and professional interviews with the concerned authorities. Field data containing structural information like the number of spans, beams, columns, the thickness of slabs, and location of stairs was collected by Ferreira from district Muzaffarabad for the existing RC school buildings and was verified with the structural drawings [44]. The databases containing building dimensions are tabulated in Table 1 [21]. The nonlinear incremental dynamic analysis was employed to determine the structural behaviour against the gradually scaled 20 past earthquake records depending upon the site history. The Peak Ground Acceleration (PGA) was selected as seismic intensity measure (IM). Three damage limit states: serviceability damage state (DS1), damage control limit state (DS2), and collapse prevention damage state (DS3), were defined. The fragility curves were plotted for each damage state using the IDA procedure in FEM software (PERFORM -3D).

Furthermore, the seismic performance was predicted using the ANN model. C. Global drift (Y) is chosen as the output value in the output layer. The accuracy of the target values using the trained ANN model requires some important tasks such as the definition of a suitable ANN architecture, selection of training, testing data sets, and training and testing the network. The same architecture is trained and tested for each damage limit state. The ANN model was trained and tested in MATLAB. One of the important steps in solving the problem is choosing the optimum number of neurons and hidden layers. The percentage of error can be reduced by optimization of the number of hidden layers and the number of neurons. ANN architecture comprises a selection of training algorithms, several hidden layers, and a number of neurons and activation functions. This study utilizes a network architecture of 5 neurons in the input layer, 25 neurons in hidden layer, and one neuron in the output layer. An appropriate training algorithm is required to recognize the relationship between the input layer and output layer. In this study, three training algorithms (Levenberg-Marquardt, Bayesian Regularization, One-Step Secant) were compared, and the best one was chosen for the establishment of fragility curves. The ANN flow diagram including the collection data, selection of architecture, training and testing the network, and finally obtaining the results is shown in Figure 1.

The drift ratios were also predicted using GEP model. The GEP model was created using three groups of fitting parameters such as general parameters, numeric constants, and genetic operators. The general parameters consist of the
number of chromosomes, the number of genes, head size, and the linking functions. The numeric constants comprise constants per gene, data type, lower bound, and upper bound. The genetic operators include rate of mutation, function insertion, gene transposition, and gene recombination. Figure 2 depicts the sequential explanation of the GEP flowchart, i.e., collection of data, distribution of data into the training and testing set, construction of chromosomes, displaying and execution of expression trees (ETs), and the measurement of fitness.

Table 1: Structural parameters for the selected buildings [23].

| Building ID | No. of stories | No. of bays X-direction | No. of bays Y-direction | Total building height (m) | Typical floor area (m²) | Area of slab opening (m²) |
|-------------|----------------|-------------------------|-------------------------|--------------------------|-------------------------|--------------------------|
| BLR-1       | 1              | 6                       | 3                       | 3.8                      | 328.6                   | —                        |
| BLR-2       | 1              | 3                       | 3                       | 3                        | 101.5                   | —                        |
| BLR-3       | 1              | 4                       | 2                       | 3.8                      | 143.1                   | —                        |
| BLR-4       | 1              | 4                       | 2                       | 3.8                      | 146.6                   | —                        |
| BLR-5       | 1              | 5                       | 2                       | 3.6                      | 197                     | —                        |
| BLR-6       | 1              | 6                       | 3                       | 3.6                      | 328                     | —                        |
| BLR-7       | 1              | 6                       | 3                       | 3                        | 306.6                   | —                        |
| BLR-8       | 1              | 3                       | 2                       | 3.8                      | 80.8                    | —                        |
| BLR-9       | 2              | 4                       | 2                       | 7.6                      | 145.2                   | 9.3                      |
| BLR-10      | 2              | 4                       | 2                       | 7.6                      | 125.5                   | 9.3                      |
| BLR-11      | 2              | 11                      | 2                       | 6.1                      | 344.5                   | 21.6                     |
| BLR-12      | 2              | 2                       | 2                       | 6.1                      | 66.4                    | 6.8                      |
| BLR-13      | 2              | 5                       | 2                       | 6.7                      | 192.7                   | 16.3                     |
| BLR-14      | 2              | 8                       | 3                       | 6.7                      | 253.6                   | 23.5                     |
| BLR-15      | 2              | 3                       | 2                       | 6.1                      | 58                      | 5.7                      |
| BLR-16      | 2              | 5                       | 2                       | 6.1                      | 223                     | 7.7                      |
| BLR-17      | 3              | 4                       | 2                       | 11                      | 390.2                   | 24.8                     |
| BLR-18      | 3              | 5                       | 2                       | 11                      | 209.2                   | 15.3                     |
| BLR-19      | 3              | 12                      | 2                       | 9.1                      | 377.6                   | 21.4                     |
Lastly, the results of IDA, ANNs, and GEP were compared for the fragility assessment of RC buildings.

4. Configuration of Building Topology

The buildings considered in this study comprise 1–3 stories’ buildings with different number of bays in both orthogonal directions as shown in Table 1. The size of columns and beams are different for all structures considered in this study. The height and floor area of the selected buildings are also tabulated in Table 1.

5. Structural Modelling

To perform the nonlinear analysis, a 3D nonlinear static model was created in CSI-Perform 3D that captures various aspects of structural behaviour such as hinge rotation and material strain. The characteristics of locally available materials were considered to achieve the actual structural behaviour as per the research done by Rafi and Nasir [47], Grade-40 reinforcing steel was used in the analytical model as suggested by professional experts during the interviews. The nonlinear trend of the structure can be achieved if the nonlinear inelastic fibre sections are used for modelling the reinforcing steel and concrete. For this purpose, the Mander model was assigned to concrete material, and complete confinement effects were counted in the analysis [19, 48]. For steel bars, the nonbuckling steel model was used. The material and loading values are tabulated in Table 2.

6. Selection of Ground Motion

An important step in the development of fragility curves is the selection of appropriate past ground motion records. The target spectrum is provided by the Building Code of Pakistan (BCP) but it is not reliable. Therefore, in this study, a total of 20 different earthquake ground motions were adopted with magnitude (here 6.5–8.0), source to site distance, and the shear-wave velocity, Vs (here 175–300 m/sec), as shown in Table 3 [29]. These earthquake records were selected which resonate the site characteristics such as the fault mechanism of the Kashmir region.

The seismic intensity measure (IM) for the establishment of fragility curves is an important parameter; however, there is no clear method to decide what intensity measures to be selected. Researchers have used various IMs such as Peak Ground Velocity (PGV), Peak Ground Acceleration (PGA), and Peak Ground Displacement (PGD), and spectral response acceleration (Sa) [49]. The intensity measures can be achieved either from earthquake records directly or by using the response spectrum of recorded earthquakes in the past which match the site history.

If the actual earthquake data is unavailable for a particular location, then a synthetic ground motion can also be developed [50]. The most frequently used intensity measure for the fragility analysis is PGA; however, it is considered that Sa is comparatively better [51]. Similarly, Seo et al. did the fragility analysis of curved steel bridges using PGA as an intensity measure. Tavares et al. have used PGA values for the establishment of fragility curves to evaluate the vulnerability of Canadian highway bridges [52]. The fundamental principles in the selection of accurate IMs depend on how well the hazard level of ground motion correlates the damage level of the structures. Therefore, the closeness of this correlation results in the accuracy of the fragility curves.

7. Structural Limit State Definition

A structure cannot fulfil its intended function once the limit state is exceeded. The structural capacity can be determined by establishing suitable limit states. Different limit states have been used by researchers in the past for the evaluation of structures. Avşar et al. employed three damage limit states: serviceability, damage control, and collapse prevention limit states for the assessment of ordinary highway bridges [49]. Researchers used four damage limit states (slight, moderate, extensive, and collapse) based on inter-story drift ratio to assess steel buildings against earthquake loading [53, 54]. In this study, the global drift values are used as damage states adopted from the research conducted by Zain et al. as shown in Table 4.

8. Development of Fragility Curve

Fragility analysis can be defined as the probability of occurrence of damage in a structure or structural member. The probability of failure can be represented by a lognormal distribution which can be mathematically expressed as shown in the following equation [55]:

\[ P_f[D \geq \text{CIM}] = \Phi \left( \frac{\ln(\text{IM}) - \ln(m)}{\beta} \right), \]

where \( P_f \) = probability of damage occurrence, \( \Phi \) = standard normal distribution function, \( D \), \( C \) = earthquake demand and capacity of the structure, respectively, and IM = earthquake ground motion intensity measures (PGA, as in this study).

The terms “m” and “\( \beta \)” are the median value and dispersion ratio of the lognormal distribution, respectively. The lognormal distribution was used by different researchers for the prediction of structural damage due to earthquake ground motion [56].

9. Incremental Dynamic Analysis

IDA is one of the most reliable and effective approaches for the assessment of the structural performance of buildings subjected to earthquake ground motion. In this method, the ground motion records are scaled and then applied to the
structure incrementally. The nonlinear response of the structure is captured in each step. Bayar and Bavaghar [57] adopted the IDA technique for the establishment of fragility curves for highway skewed bridges. The damage assessment of structures can be done by computing the likelihood of damage caused due to earthquake ground excitation. For this purpose, the fragility curves can be employed to quantify the probability of such damage.

10. Results and Discussions

10.1. IDA-Based Fragility Curves. The results of numerical analysis (NA), i.e., incremental dynamic analysis, were utilized to develop the seismic fragility curves for the building topology under consideration. The lognormal distribution function was considered for the conditional probabilities using MATLAB code. The fragility curves for three limit states: serviceability (DS-1), damage control (DS-2), and collapse prevention (DS-3), were established by plotting the IMs on axis while the probability of exceedance on the Y-axis is shown in Figure 3. The probability of exceedance of any limit state can be selected from the fragility curves.

| Sr. no | Earthquake       | Year | Magnitude | PGA (g) | $R_{rup}$ (kms) | Vs30 (m/sec) |
|--------|------------------|------|-----------|---------|----------------|--------------|
| 1      | San Fernando     | 1971 | 6.61      | 0.225   | 22.77          | 316.46       |
| 2      | Gazli, ISSR      | 1976 | 6.80      | 0.864   | 5.47           | 259.59       |
| 3      | Tabas, Iran      | 1978 | 7.35      | 0.106   | 28.79          | 324.53       |
| 4      | Spitak Armenia   | 1988 | 6.77      | 0.20    | 24.0           | 343.53       |
| 5      | Loma Prieta      | 1989 | 6.93      | 0.1695  | 24.57          | 239.69       |
| 6      | Loma Prieta      | 1989 | 6.93      | 0.331   | 9.30           | 347.90       |
| 7      | Northridge-01    | 1994 | 6.69      | 0.345   | 8.66           | 297.71       |
| 8      | Northridge-01    | 1994 | 6.69      | 0.309   | 12.51          | 326.47       |
| 9      | Chi-Chi Taiwan   | 1999 | 7.62      | 0.137   | 24.96          | 235.13       |
| 10     | Chi-Chi Taiwan   | 1999 | 7.62      | 0.273   | 16.04          | 233.14       |
| 11     | Chi-Chi Taiwan   | 1999 | 7.62      | 0.1448  | 25.42          | 297.86       |
| 12     | Chi-Chi Taiwan   | 1999 | 7.62      | 0.165   | 17.11          | 272.67       |
| 13     | Chi-Chi Taiwan   | 1999 | 7.62      | 0.1918  | 11.57          | 212.72       |
| 14     | Kashmir earthquake | 1999 | 7.60  | 0.2517  | 26.0           | 223.04       |
| 15     | St. Elias, Alaska | 1999 | 7.54  | 0.1759  | 26.45          | 306.37       |
| 16     | Niigata, Japan   | 1979 | 6.63      | 0.4764  | 12.81          | 274.17       |
| 17     | Montenegro, Yugoslavia | 2004 | 7.1  | 0.2928  | 5.76           | 318.74       |
| 18     | Chuetsu-oki, Japan | 1979 | 6.8  | 0.176   | 29.45          | 334.0        |
| 19     | Iwate, Japan     | 2007 | 6.9      | 0.2194  | 8.43           | 279.36       |
| 20     | Iwate, Japan     | 2008 | 6.9      | 0.2057  | 16.67          | 348.98       |

10.2. ANN-Based Fragility Curves. The numerical analysis procedure is computationally time taking and costly; therefore, machine learning techniques might be helpful. In the current study, the fragility curves were predicted using the ANN approach.

The fragility curves were predicted using the ANN technique for all three damage limit states as shown in Figure 4. All three damage limit states of the ANN model show the same trends as depicted in the IDA-based approach. In order to obtain an accurate model, 70% of the data was selected for training and 15% for validation, and 15% data was used for testing the model.

Three backpropagation algorithms, LM, OSS, and BR, are compared based on their performance, i.e., mean square error (MSE) as shown in Table 5. It is vivid that the Levenberg-Marquardt (TRAINLM) is the most efficient out of all algorithms with the least MSE equal to 0.0257. Therefore,
Table 5: Comparison of training algorithms for ANN optimization.

| Sr. no. | Algorithms | Hidden layer | Performance (MCE) |
|---------|------------|--------------|-------------------|
| 1       | TRAINLM    | 1            | 0.0257            |
| 2       | TRAINOSS   | 1            | 0.0738            |
| 3       | TRAINBR    | 1            | 0.2810            |

Table 6: Optimized ANN architecture was used for training.

| Sr. no. | Parameters | Selection |
|---------|------------|-----------|
| 1       | Training function | Levenberg—Marquardt |
| 3       | Performance function | Mean square error (MSE) |
| 4       | No. of hidden layers | 1 |
| 5       | Number of neurons in hidden layers | 25 |
| 6       | Input parameters | 5 |
| 7       | Output parameters | 1 |

Table 7: Optimized GEP parameters.

| Sr. no. | Parameters | Selection |
|---------|------------|-----------|
| 1       | Head size | 10 |
| 3       | No. of chromosomes | 100 |
| 4       | No. of genes | 3 |
| 5       | Linking function | Multiplication |
| 6       | Fitting function | MSE |
| 7       | Constant per gene | 10 |
| 8       | Data type | Floating point |
| 9       | Lower bound: upper bound | -10:10 |
| 10      | Mutation rate | 0.00138 |
| 11      | Inversion rate | 0.00546 |

Figure 4: Fragility curves for three limit states using ANN approach.

Figure 5: Comparison plot between target and predicted values of ANN model.

Figure 6: Fragility curves for three limit states using GEP model.

Figure 7: Comparison plot between target and predicted values of GEP model.
the LM training algorithm has been adopted for the ANN model.

Table 6 shows the optimized ANN architecture (5-25-1) which is selected based on hits and trials. A single hidden layer with 25 neurons is selected based on the lowest epochs and MSE values. MSE is chosen as the performance function and LM as a training function.

In order to predict the drift values using ANN model, a relationship is developed between the target values and the predicted values as shown in Figure 5. The performance of the model can be indicated by the correlation coefficient, $R^2$, which is equal to 0.9384.

10.3. GEP-Based Fragility Curves. Selection of an appropriate GEP architecture is necessary; hence the suitable architecture was selected in this study with the optimized GEP parameters; for example, head size, number of chromosomes, number of genes, functions, and data type parameters are shown in Table 7. The number of genes and

Figure 8: Distribution of errors of ANN and GEP models.

Figure 9: Expression trees (ETs) of gene expression programming model.
number of chromosomes are limited to 3 and 100, respectively.

The fragility curves for all three limit states were plotted as shown in Figure 6, which follow almost the same trend as in the case of IDA-based fragility curves; however, compared to the ANN model, the accuracy of the GEP model is lower in terms of coefficient of correlations, $R^2$.

In order to predict the GEP model, the target values and predicted values of GEP model were compared as shown in Figure 7. The results obtained from the GEP model are in the acceptable range with a coefficient of correlation, $R^2$, equal to 0.87.

The error distribution between the target and predicted values is also shown in Figure 8. From the error plot, the maximum error and average error are 1.451 and 0.1953, respectively, for the ANN model. The GEP model gives the maximum error equal to 1.478 and the average error of 0.133. The maximum error in the GEP model is greater than that of the ANN model.

The global drift values ($Y$) can also be predicted using GEP-based empirical equation. Here equation (7) is used to forecast the global drift values. The variables $A$, $B$, and $C$ of equation (7) can be determined using equations (8)–(10), respectively.

Global drift values = $Y$

$$= A \times B \times C,$$

$$A = D - (M \times P^2) \times (2.504 - D) + \frac{8.14}{N},$$

$$R^2 = \frac{\sum (y_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}.$$
\[ B = \frac{M \times P}{(N \times D \times M^2) + (D \times V \times P)}, \]
\[ C = M^2 - \left( \frac{P + P \times P}{(-5.87 + N)/( -7.89 + D) \times (M - N)} \right), \]

where \( Y \) = global drift values, \( N \) = earthquake number, \( D \) = duration of earthquake, \( M \) = magnitude of earthquake, \( P \) = Peak Ground Acceleration (PGA), and \( V \) = average seismic shear-wave velocity up to a depth of 30 meters from the surface (Vs30).

Figure 9 shows the graphical representation of the GEP model which can be decoded in the form of equations.

The fragility curves established using incremental dynamic analysis are compared with ANN and GEP techniques corresponding to all three limit states as shown in Figure 10. In Figure 10(a), the probability of exceedance of the ANN model occurs at a lower PGA value compared to the GEP model, where the same probability of exceedance occurs at a higher PGA value. Corresponding to PGA that equals 0.4 g, the probability of exceedance is approximately 58%, 40%, and 20% using the ANN model, IDA approach, and GEP model, respectively. However, the trend reverses at PGA of nearly 0.6 g, the GEP model predicts the highest probability of exceedance followed by the IDA approach, and the lowest probability of exceedance is predicted using the ANN model. At PGA of 1.0 g and above, the fragility curves of all three models coincide, which shows the 100% probability of exceedance of the serviceability limit state.

For the damage control limit state in Figure 10(b), the ANN model follows almost the same trend as the NA model; however, the GEP model shows a high probability of exceedance at PGA greater than 0.8 g. At PGA 0.6 g, the probability of exceedance is nearly 33%, 18%, and 4%, using ANN, IDA, and GEP approaches, respectively. There is an abrupt change in the curvature of the GEP-based fragility curve at the damage limit state; however, the ANN model predicts very well at PGA 1.0 g and above.

For the collapse prevention limit state, the results of both ANN and GEP approaches forecast almost the same result as the numerical analysis-based approach as shown in Figure 10(c).

The results obtained can be used for the fragility assessment of buildings.

11. Conclusion

This study presents the applications of machine learning approaches such as ANN and GEP to predict the global drift values for the fragility analysis of RC school buildings in Pakistan. The results of ANN and GEP were compared with the analysis results of the incremental dynamic analysis (IDA) technique. Using soft computing techniques, the computation cost can be reduced, and the complex modelling process as in the numerical approaches can be avoided. Therefore, the machine learning techniques are useful in predicting the seismic behaviour of buildings using fragility curves.

(i) Out of the two ML approaches, the ANN model gives more efficient results in terms of coefficient of determination compared to the GEP model corresponding to all three limit states. The ANN model predicts the global drift values with \( R^2 \) equal to 0.938 compared to the GEP model having \( R^2 \) equal to 0.87.

(ii) Both the ML models produced very similar fragility curves as were obtained using numerical modelling. Unlike the conventional numerical approaches such as incremental dynamic analysis which requires huge computational cost and is time-consuming, the ML techniques, especially the ANN approach, can be utilized to establish the fragility curves for the seismic assessment of buildings.

The performance of ML models can further be enhanced by improving the quality data and increasing the number of datasets.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] S. A. Memon, M. Zain, D. Zhang, S. K. U. Rehman, M. Usman, and D. Lee, "Emerging trends in the growth of structural systems for tall buildings," Journal of Structural Integrity and Maintenance, vol. 5, no. 3, pp. 155–170, 2020.
[2] B. L. Khan, M. Azeem, M. Usman, S. H. Farooq, A. Hanif, and M. Fawad, "Effect of near and far field earthquakes on performance of various base isolation systems," Procedia Structural Integrity, vol. 18, pp. 108–118, 2019.
[3] Cred The Center of Research on Epidemiology of Disasters, "Economic Losses, Poverty & DISASTERS," CRED, 2017.
[4] M. Lew and S. A. L. Mario, Why Buildings Fall Down: How Structures Fail, Norton - Library of Congress visual sourcebooks in architecture, design, and engineering, 2002.
[5] N. Mahmood, K. Khan, Z. Rafi, and D. Finn Lovholt, "Mapping of Tsunami Hazard along Makran Coast of Pakistan," Report, 2012.
[6] M. Zare and S. Karimi-Paridari, "Field observations on geological aspects," in Proceedings of the The 14th World Conference on Earthquake Engineering, IEEE, Beijing, China, October 2008.
[7] S. Rehman, Z. Rafi, and C. Lindholm, "Probabilistic seismic hazard analysis for Quetta city," Acta Geophysica, vol. 49, pp. 737–761, 2012.
[8] N. Ahmad, Q. Ali, H. Crowley, and R. Pinho, "Earthquake loss estimation of residential buildings in Pakistan," Natural Hazards, vol. 73, no. 3, pp. 1889–1955, 2014.
[9] Eefit, "The Kashmir, Pakistan Earthquake of 8 October 2005 a Field Report by Eefit," A FIELD REPORT BY EEFIT, 2005.
[10] S. M. Ali, A. N. Khan, S. Rahman, and A. M. Reinborn, "A survey of damages to bridges in Pakistan after the major earthquake of 8 october 2005," Earthquake Spectra, vol. 27, no. 4, pp. 947–970, 2011.
[11] M. A. Hassan, M. Usman, A. Hanif, S. H. Farooq, and J. Ahmed, “Improving structural performance of timber wall panels by inexpensive FRP retrofitting techniques,” Journal of Building Engineering, vol. 27, Article ID 101004, 2020.

[12] J. Ahmad, M. Usman, M. A. Hassan, S. H. Farooq, and A. Hanif, “Enhancing lateral load performance of traditional timber wall (Dhajji-Devari) by strengthening of joints,” IOP Conference Series: Materials Science and Engineering, vol. 431, no. 7, Article ID 072002, 2018.

[13] M. Haseeb, A. Bibi, J. Z. Khan, I. Ahmad, and R. Malik, “Construction of earthquake resistant buildings and infrastructure implementing seismic design and building code in northern Pakistan 2005 earthquake affected area,” International Journal of Business and Social Science, vol. 2, no. 4, pp. 168–177, 2011.

[14] A. Rasheed, S. H. Farooq, M. Usman, A. Hanif, N. A. Khan, and R. A. Khushnood, “Structural reliability analysis of superstructure of highway bridges on China-Pakistan economic corridor (CPEC): a case study,” Journal of Structural Integrity and Maintenance, vol. 3, no. 3, pp. 197–207, 2018.

[15] BCP, “Seismic Provisions, Government of Islamic Republic of Pakistan,” Ministry of Housing and Works, Islamabad, Pakistan, 2007.

[16] A. Mustaq, S. A. Khan, H. F. Gabriel, and S. Haider, “Seismic vulnerability assessment of deficient RC structures with bar pullout and joint shear degradation,” Advances in Civil Engineering, vol. 2015, Article ID 537405, 10 pages, 2015.

[17] R. Bilham, S. Lodhi, S. Hough, S. Bukhari, A. Murtaza Khan, and S. F. A. Rafeeqi, “Uncertain past,” Uncertain Future, vol. 78, pp. 601–613, 2007.

[18] S. Ahmad, N. Kyriakides, and K. Pilakoutsos, “Seismic fragility assessment of existing sub-standard low strength reinforced concrete structures abstract,” Earthquake Engineering and Engineering Vibration, vol. 14, pp. 439–452, 2015.

[19] M. Zain, M. Usman, and S. Hassan Farooq, “A framework with reduced computational burden for seismic fragility assessment of reinforced concrete buildings in high-intensity seismic zones,” Structures, vol. 33, no. October, pp. 3055–3065, 2021.

[20] J. W. Baker, “Measuring bias in structural response caused by ground motion scaling,” Pacific Conference on Earthquake Engineering, vol. 056, pp. 1–6, 2007.

[21] M. Zain, N. Anwar, F. A. Najam, and T. Mehmoood, “Seismic fragility assessment of reinforced concrete high-rise buildings using the uncoupled modal response history analysis (UMRHA),” Proceedings of the International Conference on Earthquake Engineering and Structural Dynamics, vol. 47, pp. 201–218, 2019a.

[22] Y. Pan, A. K. Agrawal, M. Ghosh, and S. Alampani, “Seismic fragility of multistep simply supported steel highway bridges in New York state. II: fragility analysis, fragility curves, and fragility surfaces,” Journal of Bridge Engineering, vol. 15, no. 5, pp. 462–472, 2010.

[23] M. Zain, M. Usman, and T. Mehmoood, “Seismic vulnerability assessment of school buildings in seismic zone 4 of Pakistan,” Advances in Civil Engineering, vol. 2019, Article ID 5808256, 14 pages, 2019b.

[24] D. E. Choe, P. Gardoni, and D. Rosowsky, “Fragility increment functions for deteriorating reinforced concrete bridge columns,” Journal of Engineering Mechanics, vol. 136, no. 8, pp. 969–978, 2010.

[25] R. Segura, J. E. Padgett, and P. Paultr, “Metamodel-based seismic fragility analysis of concrete gravity dams,” Journal of Structural Engineering, vol. 146, no. 7, Article ID 04020121, 2020.

[26] S. Mangalathu, G. Heo, and J. S. Jeon, “Artificial neural network based multi-dimensional fragility development of skewed concrete bridge classes,” Engineering Structures, vol. 162, no. August 2017, pp. 166–176, 2018.

[27] D. T. Akomolafe, “Scholarly research library comparative study of biological and artificial neural networks,” Of Applied Engineering and Scientific Research, vol. 2, no. 1, pp. 36–46, 2013.

[28] M. Shokri and K. Tavakoli, “A review on the artificial neural network approach to analysis and prediction of seismic damage in infrastructure,” International Journal of Hydroinformatics, vol. 2, no. 4, p. 178, 2019.

[29] Z. Liu and Z. Zhang, “Artificial neural network based method for seismic fragility analysis of steel frames,” KSCE Journal of Civil Engineering, vol. 22, no. 2, pp. 708–717, 2018.

[30] M. M. Javidan, H. Kang, D. Isobe, and J. Kim, “Computationally efficient framework for probabilistic collapse analysis of structures under extreme actions,” Engineering Structures, vol. 172, pp. 440–452, 2018.

[31] C. Huang and S. Huang, “Predicting capacity model and seismic fragility estimation for RC bridge based on artificial neural network,” Structures, vol. 27, no. October 2020, pp. 1930–1939, 2020.

[32] A. C. Neves, I. González, J. Leander, and R. Karoumi, “Structural health monitoring of bridges: a model-free ANN-based approach to damage detection,” Journal of Civil Structural Health Monitoring, vol. 7, no. 5, pp. 689–702, 2017.

[33] C. Neves, “Structural health monitoring of bridges: data-based damage detection method using machine learning,” Thesis, Structural Engineering and Bridges, 2020.

[34] S. Yousaf, S. Imran, and J. Ahmad, “Features selection and optimized neural network architecture for modelling flows in solar collectors,” in Proceedings of the - 2017 International Conference on Frontiers of Information Technology, December 2018.

[35] R. Dastres and M. Soori, “Artificial neural network systems,” International Journal of Imaging & Robotics, 2021.

[36] M. Y. Rafiq, G. Bugmann, and D. J. Easterbrook, “Neural network design for engineering applications,” Computers & Structures, vol. 79, no. 17, pp. 1541–1552, 2001.

[37] M. Z. Naser and A. Amir, “Insights into performance fitness and error metrics for machine learning,” 2020, https://arxiv.org/abs/2006.00887.

[38] D. T. Akomolafe, “Scholars research library comparative study of biological and artificial neural networks,” Of Applied Engineering and Scientific Research, vol. 2, no. 1, pp. 36–46, 2013.

[39] Z. Liu and Z. Zhang, “Artificial neural network based method for seismic fragility analysis of steel frames,” KSCE Journal of Civil Engineering, vol. 22, no. 2, pp. 708–717, 2018.

[40] M. M. Javidan, H. Kang, D. Isobe, and J. Kim, “Computationally efficient framework for probabilistic collapse analysis of structures under extreme actions,” Engineering Structures, vol. 172, pp. 440–452, 2018.

[41] C. Huang and S. Huang, “Predicting capacity model and seismic fragility estimation for RC bridge based on artificial neural network,” Structures, vol. 27, no. October 2020, pp. 1930–1939, 2020.

[42] A. C. Neves, I. González, J. Leander, and R. Karoumi, “Structural health monitoring of bridges: a model-free ANN-based approach to damage detection,” Journal of Civil Structural Health Monitoring, vol. 7, no. 5, pp. 689–702, 2017.

[43] C. Neves, “Structural health monitoring of bridges: data-based damage detection method using machine learning,” Thesis, Structural Engineering and Bridges, 2020.

[44] S. Yousaf, S. Imran, and J. Ahmad, “Features selection and optimized neural network architecture for modelling flows in solar collectors,” in Proceedings of the - 2017 International Conference on Frontiers of Information Technology, December 2018.

[45] R. Dastres and M. Soori, “Artificial neural network systems,” International Journal of Imaging & Robotics, 2021.
B. H. Sumida, A. I. Houston, J. M. McNamara, and W. D. Hamilton, “Genetic algorithms and evolution,” *Journal of Theoretical Biology*, vol. 147, no. 1, pp. 59–84, 1990.

C. Ferreira, *Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence*, Springer, Berlin, Germany, 2006.

M. A. Khan, A. Zafar, F. Farooq et al., “Geopolymer concrete compressive strength via artificial neural network, adaptive neuro fuzzy interface system, and gene expression programming with K-fold cross validation,” *Frontiers in Materials*, vol. 8, p. 66, 2021.

P. D. Kilmer, “Gene expression programming: a survey,” *Journalism: Theory, Practice & Criticism*, vol. 11, no. 3, pp. 369–373, 2010.

M. M. Rafi and M. M. Nasir, “Experimental investigation of chemical and physical properties of cements manufactured in Pakistan,” *Journal of Testing and Evaluation*, vol. 42, no. 3, Article ID 20130158, 2014.

J. B. Mander, M. J. N. Priestley, and R. Park, “Conducted providing the stress-strain relation for the concrete and steel are- known. The moments and curvatures associated with increasing flexural deformations of the column may be computed for various column axial loads by incrementing the curvature A,” *Journal of Structural Engineering*, vol. 114, no. 8, pp. 1804–1826, 1989.

Ö Avşar, A. Yakut, and A. Caner, “Analytical fragility curves for ordinary highway bridges in Turkey,” *Earthquake Spectra*, vol. 27, no. 4, pp. 971–996, 2011.

E. Choi, “Seismic Analysis and Retrofit of Mid-America Bridges,” *Earthquake Retrofitting*, p. 384, 2002.

Y. Pang and L. Wu, “Seismic fragility analysis of multispans reinforced concrete bridges using mainshock-aftershock sequences,” *Mathematical Problems in Engineering*, vol. 2018, Article ID 1537301, 12 pages, 2018.

D. H. Tavares, J. E. Padgett, and P. Paultre, “Fragility curves of typical as-built highway bridges in eastern Canada,” *Engineering Structures*, vol. 40, pp. 107–118, 2012.

T. T. Nahar, M. M. Rahman, C. A. Tuan, and D. K. Kim, “Seismic risk assessment based on drift ratio for steel frame using high dimensional model representation and incremental dynamic analysis,” in *Proceedings of the International Conference on Planning, Architecture and Civil Engineering*, Rajshahi University of Engineering & Technology, Rajshahi, Bangladesh, February 2019.

L. Jiang, H. Zheng, and Y. Hu, “Experimental seismic performance of steel- and composite steel-panel wall strengthened steel frames,” *Archives of Civil and Mechanical Engineering*, vol. 17, no. 3, pp. 520–534, 2017.

J. W. Baker and M. Eeri, “Efficient analytical fragility function fitting using dynamic structural analysis,” *Earthquake Spectra*, vol. 31, no. 1, 2015.

C. A. Kircher, A. A. Nassar, O. Kustu, and W. T. Holmes, “Development of building damage functions for earthquake loss estimation,” *Earthquake Spectra*, vol. 13, no. 4, 1997.

M. Bayat and Y. Bavaghar, “Seismic fragility curves for highly skewed highway bridges,” *Journal of Vibroengineering*, vol. 19, no. 4, pp. 2749–2758, 2017.