Journal of Innovations in Computer Science and Engineering
Shahid Beheshti Unversity

*January 2025, Special Issue on AI 4 All- 2*

# Predicting Damage States of RC Columns Using Machine Learning Algortithms

Amirali.Abdolmaleki,
*Faculty of Civil,Water and Environmental Engineering, Shahid Beheshti University,Tehran,Iran, mirali12701350@gmail.com*
Shima.Mahboubi✉, Code Orcid: 0000-0001-9059-871X
*Faculty of Civil,Water and Environmental Engineering, Shahid Beheshti University,Tehran,Iran.Sh_mahboubi@sbu.ac.ir*

Abstract— Performance-based design of bridges requires prediction of different damage states of components. (RC) piers are key components in bridge system, which may experience severe damages during earthquakes. Therefore, seismic damage assessment of RC bridges depends strongly on predicting failure modes RC piers. Using machine learning for damage evaluation of structures is becoming increasingly popular in earthquake engineering. This study implements three different machine learning techniques to capture different damage limit states of RC bridge piers under seismic loading. For this purpose, three machine learning techniques including K-Nearest Neighbors (KNN), Artificial Neural Networks (ANNs) and decision tree regressions were utilized for predicting four damage states of a RC bridge piers tested experimentally under seismic excitations based on drift limits. The efficiency of the three algorithms in damage prediction of RC piers were compared.

CC BY

***Keywords— Machine learning algorithm, RC Bridge pier, Damage states, Seismic performance.***

## I. Introduction

Bridge are key components in highway transportation systems. It is crucial that bridges continue to function and be serviceable during earthquakes. Seismic damage assessment of bridges in seismic zones are essential to preventing potential damage during an earthquake [1]. Therefore, understanding how seismic loads are transferred through different parts of the structural system of bridges, as well as the failure mechanisms associated with them are vital arteries [2]. Achieving optimal post-earthquake performance and predicting seismic performance across a range of earthquake scenarios requires identifying various structural component damage states. In recent years, many researchers have used machine learning techniques for predicting seismic responses of structures and bridges. For example, Mangalathu et al. [3] used ANN regression to create bridge specific fragility curves, without the need to group bridge classes in traditional regional risk assessment. Xie and DesRoches [4] applied stepwise and LASSO regression techniques to investigate the parameters that affect the seismic demands of different bridge components. Chaabene et al. [5] investigated several machine learning regression approaches for the prediction of mechanical properties of concrete. Figueiredo et al. [6] presented a hybrid finite element–based machine-learning method for damage detection in existing bridges. In addition to predicting structural responses to extreme loads, such as blasts, machine learning has also been used to predict earthquake damage. He et al. [7] employed nine ML algorithms to predict the maximum displacement of concrete-filled steel tubes (CFSTs) under blast loads. Almustafa and Nehdi [8] utilized a tree-based ensemble algorithm to predict the maximum displacement of RC columns under blast loads.The residual bearing capacity of RC columns was used in another study to estimate their damage indices, providing valuable insight into structural vulnerability under extreme conditions [9].
This paper utilized three machine learning algorithm including KNN, ANN to predict damage limit states based on RC column drift. RC columns need to be understood under blast and seismic loads, as recent studies have shown that factors such as reinforcement ratios, axial forces, and cross-section geometry significantly influence structural performance [10].

## II. RC Column Data base and damage states

To establish a comprehensive database of RC bridge columns, two experimental result databases of 203 circular and 253 rectangular column results have been utilized, respectively. These databases provided by the PEER structural performance database (SPD) [11].
The damage limit states proposed by Duta and Mander [12] for RC bridge columns were considered in this study. Dutta and Mandar [12] proposed five damage states for RC bridge columns based on displacement limits, which include almost none, slight, moderate, extensive, and complete, as listed in Table 1.

Submit Date: 2024-12-01
Accept Date: 2025-07-26
✉ Corresponding author

TABLE 1. Damage limits proposed by Dutta and Mander [12].

| Damage state | Description | Drift limits | Normalize drift |
|---|---|---|---|
| Almost no | First yield | 0.005 | 0.1 |
| Slight | Cracking, spalling | 0.007 | 0.14 |
| Moderate | Loss of anchorage | 0.015 | 0.3 |
| Extensive | Incipient column | 0.025 | 0.5 |
| Collapse | Column | 0.05 | 1 |

## III. machine learning techniques

Machine learning approaches are highly dependent on quality training data. The data used were initially 491 entries in the Sivaramakrishnan database [13], however, some data points that disrupted the machine learning processes were removed. For improved accuracy, only high-quality data was used to train the models. In this study, the goal is to utilize the best machine learning techniques to accurately predict damage states of RC bridge columns based on selected input parameters. Three different machine learning algorithms KNN, ANN, and Decision tree were used in this study are described in the following sections. Python was used to implement various machine learning methods.

## IV. K-Nearest Neighbors regression

K-Nearest Neighbors (KNN) regression is a non-parametric method to predict numerical values based on their distance from each other [14]. The value of K represents the number of nearest neighbors used in the decision-making process to predict the class or target value. For each new data point that requires a prediction, the distance between every training data point and it is calculated. Euclidean distance is commonly used, but other distance metrics may also be applied, such as Manhattan distancein this study, the data was loaded from an excel file, and the relevant features for modeling were selected. Afterward, the data was standardized using StandardScaler and then splited into training and testing sets so that the model can be trained on the training data and evaluated on the test data. In order to fine-tune the model, GridSearchCV was utlized to explore the best parameter values for the KNN model, which encompasses the number of nearest neighbors (n_neighbors) and the weighting method (weights). This exploration involved employing 5-fold cross-validation, which aids in assessing the model's accuracy across various data splits and pinpointing the most effective configurations. The optimized model was then evaluated on the test data, and the desired metrics were calculated. The choice of the number of nearest neighbors (K) significantly impacted the model's performance; lower values of K can reduce bias and increase variance, while higher values of K may decrease variance and increase bias. In this model, K values ranging from 1 to 30 were examined to determine the best settings for the model.

## V. Artificial Neural Networks

As a class of machine learning models, Artificial Neural Networks (ANNs) use interconnected layers of neurons to perform regression and classification tasks [15]. There are many advantages to using ANNs, including their high parallelism, tolerance of noise and errors, and ability to learn nonlinear relationships between input features and target variables. ANN can handle problems where traditional methods might fail. In addition, they can detect complex relationships in data that linear models cannot capture. Their ability to process multiple inputs simultaneously (parallel processing) enhances their performance, which make them robust in environments where data quality may vary. The objective of this study was to predict the variable Drift based on several input features using a feedforward neural network model, As a result, the model performance could be evaluated on unseen data, with three hidden layers of 256, 128, and 64 neurons, respectively, using ReLU (Rectified Linear Unit) activation functions. This study utilized deep networks using ReLU because it introduces non-linearity while reducing gradient vanishing, and was compiled using the Adam optimizer with 0.001 learning rate. Adam was selected because of its advantages over momentum-based optimization. Using a loss function based on Mean Absolute Error (MAE), it is highly effective for a variety of machine learning tasks, including SGD with Momentum and RMSprop. A neural network model was used in this study to learn complex relationships between the input features and the target values. Neural networks are commonly used in regression tasks to minimize absolute differences between predictions and true values.

## VI. Decision Tree regression

For classification and prediction, the Decision Tree (DT) algorithm divides data into a tree structure, with each node representing a feature (or attribute), and each branch representing its outcome. It is constructed by selecting features that provide the most information about a classification based on criteria such as Information Gain or Gini Index. As a result of implementing rules derived from various splits, the algorithm predicts the classification of new data. Decision trees are widely used because of their simplicity and interpretability. It is important to choose split features in a decision tree algorithm such that they optimize the criterion, such as Information Gain or Gini Index. This selection directly impacts the performance of the model.Additionally, optimizing the tree, such as pruning, is crucial to reduce complexity and enhance the model's ability to generalize to new data. Used in machine learning, decision trees are powerful techniques for modeling and prediction, which is why this method has been used in this study. Non-Linear Relationships and Use of the CART Algorithm:Due to the non-linear relationship between predictor features and the target variable, the CART (Classification and Regression Trees) algorithm has been employed. A recursive decision tree algorithm is used with the CART algorithm for modeling complex and non-linear relationships. This algorithm is suitable for modeling and simulating complex and non-linear relationships. It divides the input space into different sections and aims to minimize impurity in each section. A decision tree starts with a root node and is divided into internal nodes and terminal nodes through greedy splitting. Each internal node has only one parent and two children, while each terminal node has only one parent and makes a final prediction. The decision tree algorithm stops when one of the following conditions is met:

• All nodes have been created.
• All features have been fully covered.

• Impurity reduction in the nodes has been minimized.
A parameter for adjusting or stopping the tree can also be set as its maximum depth (max depth). By default, the tree is considered to have infinite depth. Grid Search is used to find the best parameters for the tree's maximum depth (max_depth) and the minimum number of samples required for a split (min_samples_split). In this search, tree depths (ranging from 12 to 20) are examined in order to determine which model is best suited for making more accurate predictions.

## VII. Comparison of the Results

A total of 491 entries were originally in the Sivaramakrishnan database [13], but some data points interfered with machine learning. Several parameters that affect damage limit states of RC columns were investigated. including the spacing of ties (s), volumetric stirrup ratio (ρs), diameter of longitudinal reinforcing bars (db), compressive strength of concrete (fc'), yield stress of transverse reinforcement (fys), axial load (P), yield stress of longitudinal reinforcement (fy), total cross-sectional area (Ag), longitudinal reinforcement ratio (ρl), depth (D), column length (L), effective cover (d'), cover measured from the outer surface of the column to the center of closest longitudinal reinforcement and their interactions to identify the most significant parameters affecting limit states. They identified the axial load ratio (P/Agfc') and the column aspect ratio (L/D) as key parameters influencing concrete spalling and proposed a regression equation for predicting displacement at the onset of concrete spalling. In this paper, five parameters are taken as input data for machine learning approaches as presented in Table 2.

TABLE 2. Statistical ranges of rectangular column input parameters.

| Input Parameter | Input Parameter symbol | Minimum | Maximum | Mean |
|---|---|---|---|---|
| Concrete compressive strength (MPa) | $fc'$ | 21.1 | 102.7 | 41.4 |
| Longitudinal reinforcement yield strength (MPa) | $f_{yl}$ | 331.0 | 517.1 | 443.8 |
| Longitudinal reinforcement ratio | $\rho_l$ | 1.010 | 3.550 | 2.070 |
| Transverse reinforcement yield strength (MPa) | $f_{yt}$ | 255.0 | 793.0 | 439.5 |
| Transverse reinforcement ratio | $\rho_s$ | 0.000 | 6.700 | 1.582 |

Fig.1 to Fig.4 show the Heatmap chart Pearson R, R2, MAPE , and RMSE, which represent the performance of KNN, ANN and DT algorithm in slight, moderate, extensive, collapse damage states. As seen, in the slight damage state, the KNN algorithm performed best because the MAPE was obtained 2.33, which is lowest value compared to those of other two algorithms. Moreover, R2, for KNN, ANN and DT is 0.9909, 0.9953, 0.8238 respectively, which shows KNN perfomes better. Also, the Pearson r for KNN algorithm is 0.9955 that is about 0.0416 higher than ANN algorithm and 0.0862 higher than DT algorithm, KNN excels in this scenario because it relies on local distance-based relationships, which work well when data is well-structured with minimal noise. ANN also performs well because of its ability to model nonlinear relationships effectively. However, DT performs poorly because tree-based models partition the feature space into discrete segments and may struggle with fine-grained variations in data. Since the slight damage state has relatively small variations, DT fails to capture subtle dependencies effectively. In a similar manner, RMSE shows the best performance of the KNN algorithm. According to Fig. 2, the KNN algorithm in the moderate damage state shows the best performance, where Pearson r, R2, MAPE and RMSE are 0.9790, 0.9582, 3.18, and 0.2783, respectively. For the extensive damage state, the errors are slightly higher than other damage states, but still the KNN algorithm performs well and for Pearson r, R2, MAPE and RMSE shows 0.9853, 0.9708, 4.78, 0.3512 respectively. In collapse damage state, KNN algorithm shows the best performance, this outcome aligns with the nature of distance-based algorithms like KNN, which depend on feature similarity. A slight increase in complexity and variation in the dataset does not hinder KNN ability to find nearest neighbors despite moderate damage. In addition, since ANN are capable of learning complex nonlinear functions, they may require more training data before generalizing effectively. In comparison, DT has not been as successful due to its oversimplified partitioning approach, which makes it less effective when dealing with moderately varying data.

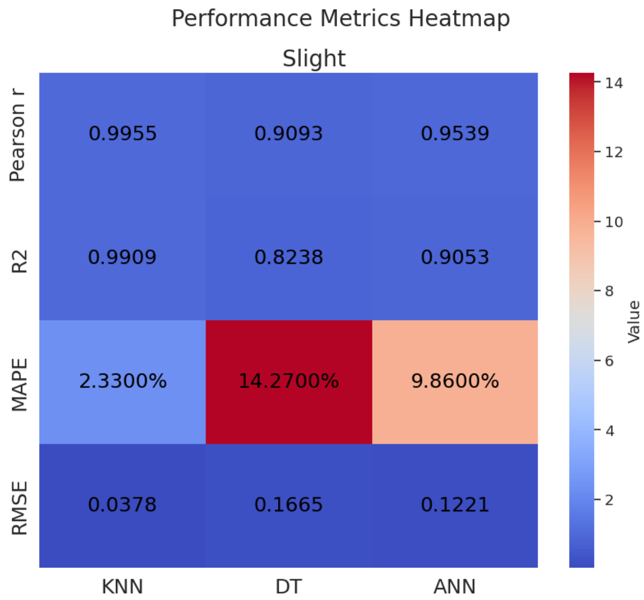


Fig. 1 Results of training and testing ML models for slight damage state.

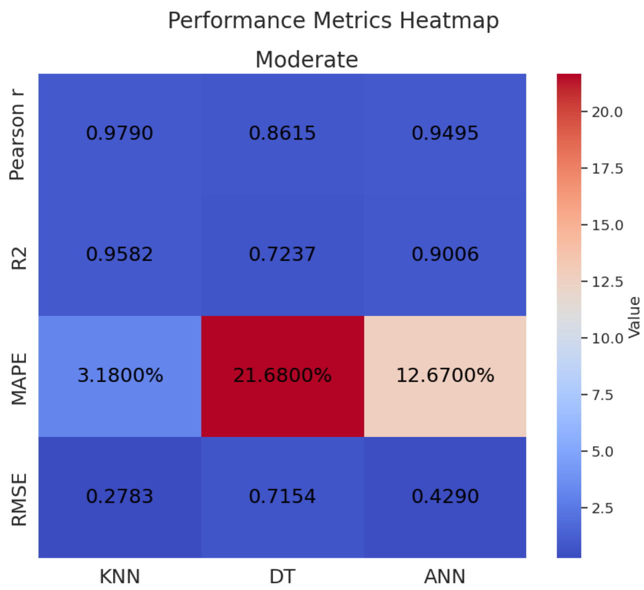


Fig. 2. Results of training and testing ML models for moderate damage state.

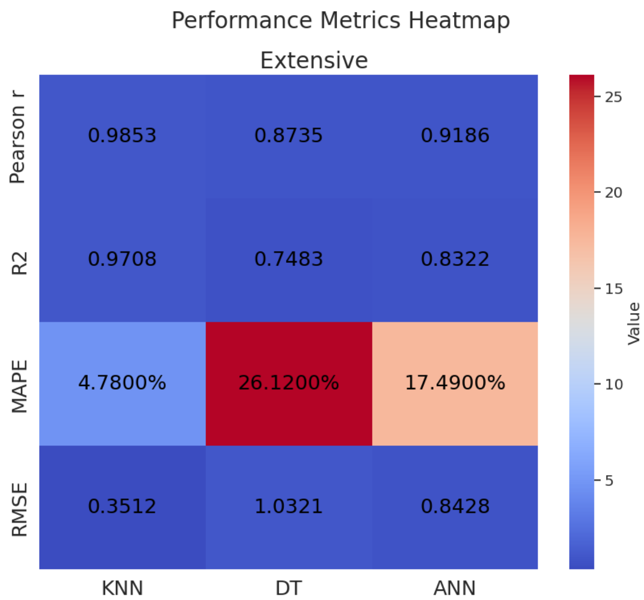


Fig. 3. Results of training and testing ML models for extensive damage state

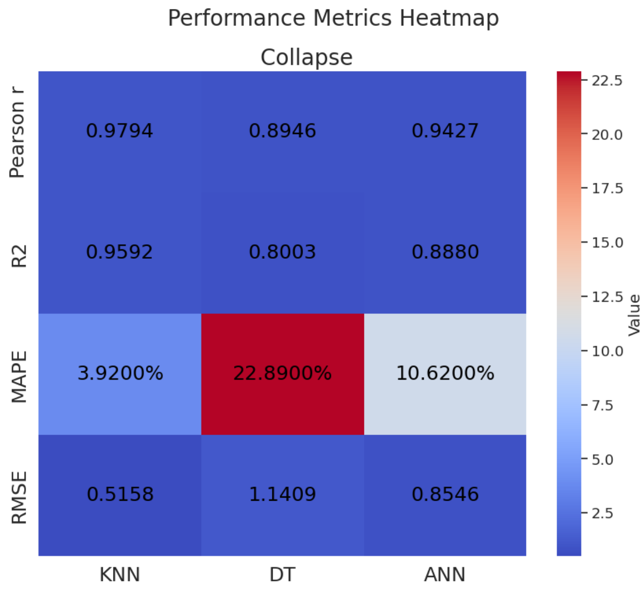


Fig. 4 Results of training and testing ML models in Collapse damage state

Fig.5 to Fig.8 presents show the final comparison between the actual data and the predictions from the optimize machine learning models. In these figures, the distance between the points compare to the ideal red line shows the amount of the model's prediction error. Points closer to this line indicate more accurate predictions. For each model's predictions, a linear trend line is drawn to show the general trend of the predictions. These lines merely represent the general distribution of predictions relative to the actual data.

Generally, the model's accuracy is close to 1:1 in the moderate damage state, indicating good performance. In all three models, the trend lines are in good agreement with the slope of the ideal line, which suggests that the predictions are in agreement with the data. In the extensive damage state, the dispersion of predictions is higher than the 1:1 line, which indicates the increase of uncertainty or challenge in predicting this category of data. The trend lines of the models differ from the ideal line in some cases, which may indicate a decrease in prediction accuracy compare to moderate damage state. In slight damage state, machine learning models have shown their accurate and stable performance. The low scatter of the points and the conformity of the trend lines with the ideal line confirm this. These results indicate the ability of the models to provide accurate predictions for data with small changes. The collapse case has a few scattered points, indicating complicated algorithms. Data noise and complexity are contributing to the decreased performance. Due to the increasing variety of data and similar training samples, KNN models are less sensitive to small-scale variations and generalize more readily than DT models. ANN remains competitive, as its deep learning structure effectively captures highly nonlinear trends; however, it is prone to overfitting when the dataset becomes overly complex. In contrast, DT’s rule-based structure and discrete decision boundaries make it the least adaptable model, as it struggles to accommodate a broader range of variations, leading to higher errors in highly variable data.

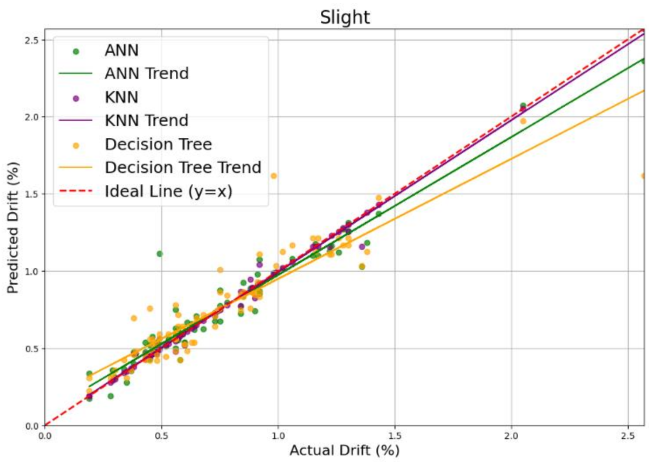


Fig. 5. Results of training and testing ML models slight limit

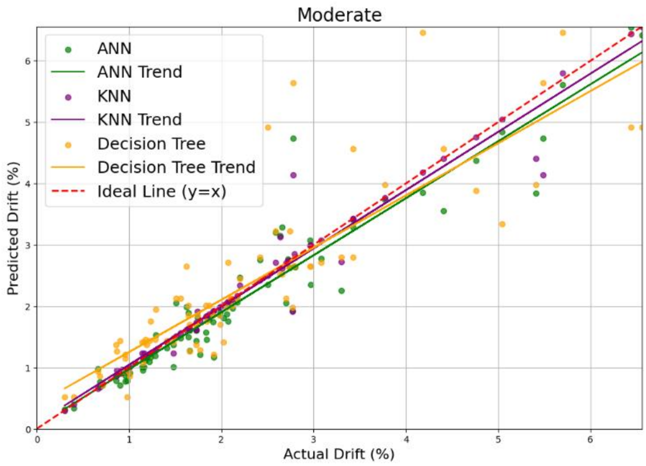


Fig. 6 Results of training and testing ML models moderate limit

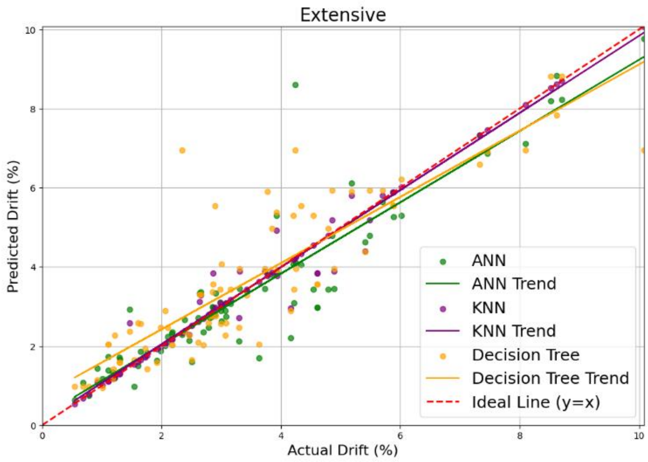


Fig. 7. Results of training and testing ML models extensive limit

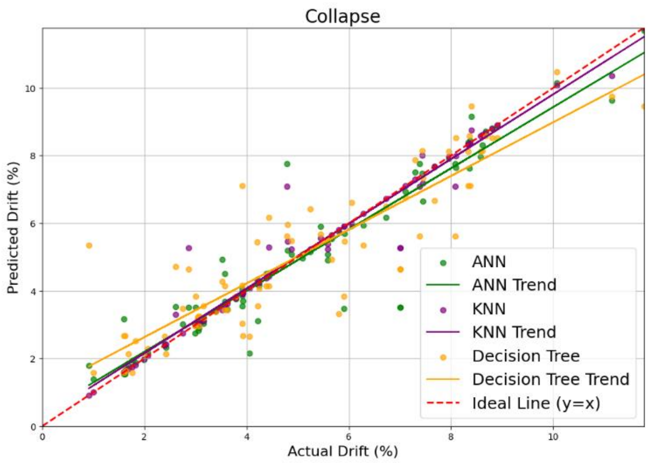


Fig. 8. Results of training and testing ML models collapse limit.

## VIII. Conclusion

This study explores the performance of three machine learning regression techniques including ANN, KNN, DT in predicting RC bridge column damage states based on drift ration. An extensive experimental database of RC bridge columns with varying material and geometrical properties were used. The machine learning regression methods were trained using 80% of the established data set. For each damage state, a second round of evaluation among the proposed models suggests the best performer, with the remaining 20% of the dataset used to test efficiency. The following conclusions can be drawn from this study:

- The machine learning models allow for the rapid prediction of drift-based damage states with good accuracy in RC bridge columns.
- The best model for prediction of cracking of concrete and buckling of steel reinforcement, which are corresponding to slight, moderate and extensive damage state were the KNN regression.

- The KNN regression shows better performance in predicting slight, moderate, and extensive and collapse damage states in comparison to ANN and DT.
- KNN is superior to ANN and DT across all damage levels because it effectively utilizes local data structures. Due to its rigid decision boundaries, DT struggles to capture nonlinear relationships because it requires more data.
- As evidenced by low data scatter, machine learning models are stable and accurate even when damage is mild or moderate. Predicting collapse states is more challenging due to increased variability, which emphasizes the need for advanced methods like ensemble learning or deep learning.

***References***

[1] Z. Liu, A. Guo, C. Zhao, A. Sextos, "Seismic response of bridges employing knowledge-enhanced neural networks for the lumped plasticity modelling of RC piers" 09 December 2023.

[2] R. Ventura, B. Barabino, G. Maternini, " Estimating the frequency of traffic overloading on road bridges" August 2024: 11 (4): 776-796

[3] Mangalathu S, Heo G, Jeon J-S. Artificial neural network based multi-dimensional fragility development of skewed concrete bridge classes. Eng Struct 2018;162: 166–76.

[4] Xie Y, DesRoches R. Sensitivity of seismic demands and fragility estimates of a typical California highway bridge to uncertainties in its soil-structure interaction modeling. Eng Struct 2019;189:605–17.

[5] Ben Chaabene W, Flah M, Nehdi ML. Machine learning prediction of mechanical properties of concrete: Critical review. Constr Build Mater 2020;260:119889.

[6] Figueiredo E, Moldovan I, Santos A, Campos P, Costa JCWA. Finite Element-Based Machine-Learning Approach to Detect Damage in Bridges under Operational and Environmental Variations. J Bridge Eng 2019;24(7):04019061.

[7] J. He, L. Jiang, L. Jiang, T. Wen, Y. Hu, W. Guo, and J. Sun, "Estimation of blast-induced peak response of concrete-filled double-skin tube columns by intelligence-based technique," *Thin-Walled Struct.*, vol. 186, p. 110670, 2023, https://doi.org/10.1016/j.tws.2023.110670.

[8] M. K. Almustafa and M. L. Nehdi, "Machine learning model for predicting structural response of RC columns subjected to blast loading," *Int. J. Impact Eng.*, vol. 162, p. 104145, 2022, https://doi.org/10.1016/j.ijimpeng.2021.104145.

[9] Y. Shi, H. Hao, and Z.-X. Li, "Numerical derivation of pressure–impulse diagrams for prediction of RC column damage to blast loads," *Int. J. Impact Eng.*, vol. 35, no. 11, pp. 1213–1227, 2008, https://doi.org/10.1016/j.ijimpeng.2007.09.001.

[10] D. Rajkumar, R. Senthil, B. Bala, K. AkshayaGomathi, and S. Mahesh Velan, "Numerical study on parametric analysis of reinforced concrete column under blast loading," *J. Perform. Constr. Facil.*, vol. 34, no. 1, 2020, https://doi.org/10.1061/(ASCE)CF.1943-5509.0001382.

[11] Berry M, Eberhard MO. Performance Models for Flexural Damage in Reinforced Concrete Columns 2003:162.

[12] Dutta, A., and J. B. Mander. "Seismic fragility analysis of highway bridges." Proceedings of the INCEDE-MCEER center-to-center project workshop on earthquake engineering Frontiers in transportation systems. 1998.

[13] Sivaramakrishnan B. Non-Linear Modeling Parameters for Reinforced Concrete Columns Subjected to Seismic Loads. University of Texas 2010

[14] S. Uddin, I. Haque, H. Lu, M. Moni, E. Gide, "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction" 15 April 2022.

[15] I.A Basheer, M Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application" 1 December 2000.

Yang D, Yang J, Shi J. Prediction and analysis of damage to RC columns under close-in blast loads based on machine learning and

Environmental Engineering at Shahid Beheshti University in Tehran, Iran. She earned her Ph.D. in Civil Engineering with a specialization in Earthquake Engineering from the Shahid Beheshti University.

**Research Interests**

Dr. Mahboubi's research focuses on:

- Structural and Earthquake Engineering
- Seismic assessment of Steel and RC buildings and bridges
- Damage evaluation and fragility analysis of structures
- She has authored several publications on seismic performance and damage assessment of reinforced concrete bridges and structural components.

**Academic and Professional Activities**

In addition to her teaching and research roles, Dr. Mahboubi serves as the faculty advisor for the Civil Engineering Scientific Association at Shahid Beheshti University.

**Contact Information**

- **Email**: sh_mahboubi@sbu.ac.ir
- **University Profile**: https://cwe.sbu.ac.ir/~sh_mahboubi

Amirali Abdolmaleki is a third-year undergraduate student at the School of Civil, Water, and Environmental Engineering, Shahid Beheshti University. His academic interests lie in the fields of structural engineering and sustainable infrastructure. He is actively engaged in student-led projects and scientific associations within the department.