SECED 2023 Conference
Earthquake Engineering & Dynamics for a Sustainable Future
14-15 September 2023, Cambridge, UK

DEVELOPMENT OF A SEISMIC DAMAGE PREDICTION MODEL BY USING MACHINE LEARNING ALGORITHMS WITH AN ARTIFICIAL DATASET

Ali Talha ATICI[1], Onur ULKU[2] & Ufuk HANCILAR[3]

Abstract: *The determination of building damage and replacing or retrofitting the risky buildings provide to reduce financial loss and the number of casualties before a catastrophic event. Evaluating the seismic performance of existing buildings is a laborious process. However, using well-trained machine learning prediction models instead of thorough structural performance analysis can reduce computational demand. This article presents the development of a damage prediction model using machine learning methods from a dataset of reinforced concrete moment-resisting 2-D bare frame systems. The characteristics of representative buildings are obtained from the literature survey about building stock characteristics of the Marmara region. OpenSeesPy framework is employed for the parametric non-linear time history analysis of structures under actual seismic recordings. The selected seismic recordings are taken into account according to different intensity measures. The Maximum Inter-Story Drift Ratio (MIDR) has been used to classify the damage level of the buildings. Trials are conducted on four classification-related algorithms, including decision tree, k-nearest neighbors, support vector machine, and random forest. As a result of comparing the performance of these algorithms, the random forest algorithm outperformed.*

Introduction

It is essential to ensure that buildings do not exceed the designated damage level, when an earthquake occurs. Especially in seismic-prone regions, the behaviour of buildings must be known beforehand to safeguard human life and minimize losses. Unfortunately, two devastating sequence earthquakes (Mw= 7.8 and Mw=7.7) hit Turkey and Syria on 6 February 2023 along the East Anatolian Fault. Due to the inability to take adequate measures on time and poor performance of structures, more than 50,000 casualties were reported and severe damage was inflicted.

Following the occurrence of these two catastrophic earthquakes, attention has once again turned towards the expected earthquake in Istanbul. The expected earthquake in Istanbul is predicted to impact a significant portion of the Marmara region. Thus, this study considers both the building stock characteristics of the region and the expected seismic activity.

Determining the earthquake performance of existing structures is one of the problems that directly affect human life in earthquake engineering. Various methods have been developed to determine the level of damage in existing structures, which can be classified into three groups: (a) seismic performance assessment methods described in the modern seismic codes (b) rapid assessment methods that score buildings according to structural characteristics (c) machine learning prediction models (Stojadinović et al., 2022; Ulku et al., 2022; Wu and Sarno, 2022; Salmi et al., 2022; Mangalathu et al., 2020; Roeslin et al., 2020). The first method can estimate the seismic performance of individual buildings, which are mainly based on the Finite Element Method (FEM). Evaluating the buildings one by one with this method will minimize the loss that will occur after the disaster, but unfortunately, it is a very long and costly process to examine and perform performance analyses of many buildings with this method. The second method estimates earthquake damage to buildings by scoring the structural features that can be obtained by on-

[1] Department of Earthquake Engineering, Kandilli Observatory and Earthquake Research Institute (KOERI), Bogazici University, Istanbul, Turkey, ali.atici@boun.edu.tr
[2] Department of Earthquake Engineering, Kandilli Observatory and Earthquake Research Institute (KOERI), Bogazici University, Istanbul, Turkey
[3] Department of Earthquake Engineering, Kandilli Observatory and Earthquake Research Institute (KOERI), Bogazici University, Istanbul, Turkey

the-ground surveys of the buildings. This method provides a rougher prediction of structural performance compared to other methods.

This study develops machine learning classification models for estimating building performance based on datasets of building properties and performance to reduce analysis time and improve accuracy, with a focus on the Marmara region. Building properties were compiled from various studies and representative structures were simulated. Performance analysis was conducted using the Openseespy framework (McKenna et al., 2010) and non-linear time history (NLTH) analyses with actual earthquake records. The selected seismic recordings are taken into account according to different intensity measures, for instance, peak ground velocity (PGV) and peak ground acceleration (PGA) of the building. As a result of the analysis, MIDR values were obtained as an Engineering Demand Parameter (EDP) to classify the damage state of the buildings.

The dataset was obtained from properties of the modelled buildings and damage states of analyzed buildings in order to train classification algorithms. The best classification algorithm was selected by comparing evaluation metrics of the machine learning models.

Preparation of the Dataset

Similar to deriving fragility and vulnerability models, data can be gathered to train machine learning algorithms for predicting earthquake damage using two approaches: (1) observing damage in the field after earthquakes, and (2) using analytical models to predict consequences. Both methods have strengths and weaknesses (Baker et al., 2021). The former method, the empirical approach, is directly validated by observational evidence but requires careful data collection. Buildings at each damage level must be inspected equally to obtain an unbiased estimate of the damage state that a building would experience given a particular level of ground motion. Otherwise, the observed dataset is going to be imbalanced. The imbalanced dataset can be handled in several ways, such as under-sampling or over-sampling methods (Kaur et al., 2020; Ulku et al., 2022). The latter one, the analytical approach, can also overcome the issue of imbalanced dataset, which is the main weakness of the empirical approach. However, the analytical approach requires that the analytical consequence predictions accurately represent reality (Baker et al., 2021). In this study, the analytical approach is preferred to generate a reliable and balanced dataset for developing a machine learning-based damage prediction model. The dataset is prepared from simulations of NLTH analyses of representative buildings to mimic reality.

Determination of Building Characteristics

The aim of generating a comprehensive dataset is to include most of the building characteristics in the Marmara region. The characteristics of representative buildings were obtained by conducting a literature survey of the building stock in the Marmara region and Istanbul (Bal et al. 2007, Azak et al. 2014, DEZIM, 2020).

The scope of this study is limited because considering all the structural features would be time-consuming. Figure 1 illustrates the joint probability distribution condition on the number of storeys and the construction year of reinforced concrete buildings in Istanbul, according to DEZIM (2020).

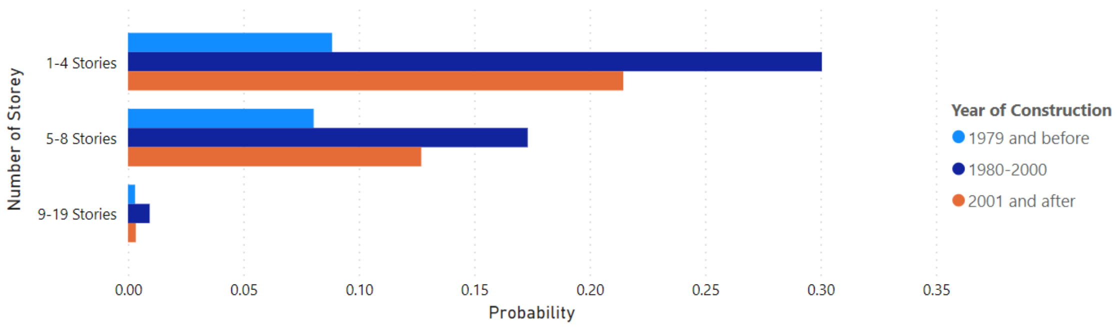


Figure 1: Joint probability condition on number of storey and construction year of reinforced concrete frame buildings in Istanbul (DEZIM, 2020)

As can be seen, in Istanbul, the largest city of the Marmara region, most of the reinforced concrete structures are low-rise (1 to 4 floors) and medium-rise (5 to 8 floors) buildings built between 1980 and 2000. For this study, to cover the majority of the buildings in the Marmara region, 3 and 5 storey buildings constructed between 1980 and 2000, respectively, were chosen for the examination of low-rise and mid-rise buildings. All the building characteristics, references of these parameters, and number of variations are listed in Table 1.

Modeling of Structures

The simulations of the structures were implemented by modeling two-dimensional (2D) finite element models in the OpenSeesPy framework. For simplicity, all buildings and column sections were assumed to a square. The middle axis of the buildings was modeled to determine the damage state under actual seismic loadings with NLTH analysis.

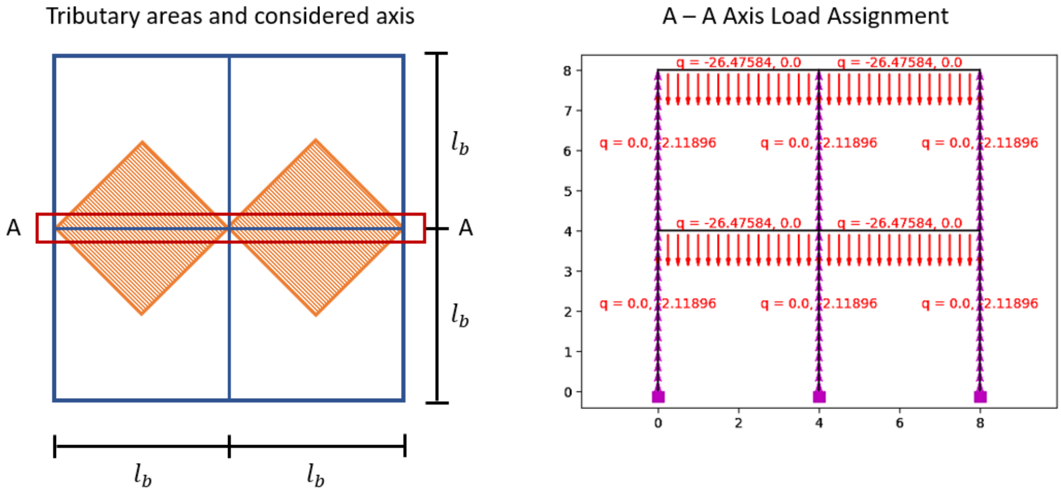
To ensure consistency in the mass of the structures, loads were imposed according to the dimensions of each structure. The loads on partition interior and exterior walls, assumed to be 2.5 kN/m^3, were imposed on the beams as distributed loads. The slabs were assumed to be reinforced concrete with a density of 25 kN/m^3 and a thickness of 15 cm. Both the superimposed dead loads and live loads were assumed to be 2 kN/m^2.

The structures were modelled as reinforced concrete moment-resisting 2D bare frame systems. Figure 2 is embedded to visualize the model with a representative building. Figure 2(a) shows the plan of the 2-span building, with the considered axis of the 2D frame and the loading tributary areas indicated. Figure 2(b) illustrates an example of the 2-storey building model scheme.

Structural Characteristics	**References**	**Parameters**	**Number of Variation**
Number of Storey	DEZIM (2020)	3 and 5	2
Storey Height	Bal et. al. (2007)	2.6, 2.8, and 3 m	3
Commercial Use of Ground Floor	Bal et. al. (2007) and Azak et. al. (2014)	‘Yes’ or ‘No’ If ‘Yes’, Ground Floor Height = 3.5 m.	2
Number of Span	Bal et. al. (2007) and Azak et. al. (2014)	3 and 4	2
Span Length	Azak et. al. (2014)	2.5, 3, and 4 m	3
Column Dimensions (For the simplicity, columns are assumed as square.)	ABYYHY-1975, Bal et. al. (2007) and Azak et. al. (2014)	3 storey: 25, 30, 35 cm 5 storey: 30, 35, 40 cm	3
Beam Width	Bal et. al. (2007)	25 cm	1
Beam Depth	Bal et. al. (2007)	60 cm	1
Concrete Compressive Strength	Bal et. al. (2007)	5, 12, 20, 28, and 35 MPa	5
Steel Yield Strength	Bal et. al. (2007)	370 MPa	1
Reinforcement Ratio for Column	ABYYHY-1975 (Minimum Requirement)	0.01	1
Reinforcement Ratio for Beam	ABYYHY-1975 (Minimum Requirement)	0.004	1
Soil Condition (V_{S30})		1130, 560, 270 cm/s	3
Total Number of Variation (Number of different buildings in this study)			3240

Table 1: Representative building characteristics, number of variation and total number of buildings

In OpenSeesPy, beam-column connections were assumed to be rigid, and columns are fixed at the base level in the FE models. The non-linear behavior of the structure was ensured by modeling all elements with “nonlinearBeamColumn” object. Distributed plasticity is provided by utilizing “RCSection2d” object which is an encapsulated fiber representation of a rectangular reinforced concrete section. Material properties of fiber section were defined by the Steel01 material for steel presence and Concrete04 for the concrete fibers from the OpenSeesPy framework. The mean value of the yield strength of S220 steel has been found approximately 370 MPa in Turkey (Akyuz et. al. 1999). For this reason, steel strength was adopted as 370 MPa instead of the nominal yield strength. Five different concrete strength between 5 and 35 are assumed to represent reality because of wide variation in concrete strength. Even “RCSection2d” object allows to model unconfined and confined concrete fibers, core and cover regions of column fibers are modeled as unconfined because of column stirrups of pre-code buildings were not folded properly and placed sparsely.



(a) (b)

Figure 2: Representative model as 2-storey and 2-span. (a) Plan view, considered axis (A-A) for 2D frame, and considered loading tributary areas; (b) the 2D frame scheme from OpenSeesPy

Verification with SAP2000

A linear-elastic modeling and analysis procedure was followed in SAP2000 (Computers and Structures, Inc., 2000) and used for verification of the model created using OpenSeesPy. The fundamental vibration periods and modal participating mass ratios resulting from free vibration analyses of the finite element models of the building in SAP2000 and OpenSeesPy were found consistent. In Table 2, vibration periods and participating mass ratios are compared for the 3-storey buildings which both have same characteristics.

Number of Mode	Period (OpenSeesPy)	Period (SAP2000)	Participating Mass Ratios (OpenSeesPy)	Participating Mass Ratios (SAP2000)
1	0.4361	0.4194	0.900 (UX)	0.905 (UX)
2	0.1516	0.1479	0.386 (RY)	0.381 (RY)
3	0.1013	0.1005	0.014 (UX)	0.013 (UX)
4	0.0416	0.0413	0.708 (UZ)	0.729 (UZ)
5	0.0355	0.0339	0.312 (RY)	0.433 (RY)

Table 2: Linear model comparison of SAP2000 and OpenSeesPy

Selection of Ground Motion Records

Istanbul has been a very hazardous area in terms of not having experienced catastrophic earthquakes for over 200 years. After the 1999 earthquakes happened in Duzce province, many researchers drew attention to the problem of earthquakes bigger than 7 in Istanbul (Parsons et al. 2000; Parsons et al. 2004; Murru et al. 2016; Hubert-Ferrari et al. 2000). Since not only Istanbul has hosted many people but also the province is very important in terms of having many risky buildings, precautions should be taken in advance. Therefore, this paper focuses on Istanbul and the practices of the buildings in which most people dwell. We generated 3240 reinforced concrete moment-resisting 2D bare frame systems with various shapes and materials. Then we performed NLTH analyses under the selected ground motions to calculate the engineering demand parameter, and this parameter led us to the damage state of a building. Thus, ground motion selection is one of the challenges of this paper.

A new ground motion database is generated by combining recent catastrophic earthquakes that happened in Turkey (AFAD) with the CyberShake database (Baker et al., 2021). The new database allows us to pick up known worldwide ground motions from various stations, as well as earthquakes that occurred in Turkey. At the same time, we categorized the first vibration period of the buildings into ten groups (Table 3) to avoid the burden of the calculations because we assumed that there would be no considerable difference between the insignificant changes in the period of the buildings while selecting ground motion. Ground motions should be appropriate for seismological conditions for the area and match with conditional mean spectrum to find optimal list of ground motion records.

First Vibration Period Class	Period Boundries	Center of Period Intervals (Considered Period)	Number of Buildings
1	0.19 – 0.28	0.24	210
2	0.28 – 0.37	0.33	567
3	0.37 – 0.46	0.41	765
4	0.46 – 0.54	0.50	702
5	0.54 – 0.63	0.59	483
6	0.63 – 0.72	0.67	270
7	0.72 – 0.80	0.76	141
8	0.80 – 0.89	0.85	63
9	0.89 – 0.98	0.94	27
10	0.98 – 1.07	1.02	12

Table 3: List of period classes for ground motion selections

For the sake of simplicity, this paper focuses on a single rupture that was obtained from Murru et al. (2016). The authors proposed three probability models for single and multi-segment earthquakes, based on recurrence distributions. In this study, we used the median value (50th percentile) of magnitude, which is 7.56, assuming that the West Marmara, Central Marmara, and Cinarcik segments rupture together. These segments are characterized by a right-lateral strike-slip mechanism. We assumed that the buildings are located 10 km away from the rupture projection. Additionally, this paper takes into account the variability of soil conditions by considering three different values of V_{S30}: 1130, 560, and 270. Based on these pieces of information, 11 ground motion records and scale factors are selected by utilizing Baker and Lee's (2018) algorithm, for each V_{S30} value that matches the target distribution, which is conditioned on the first vibration period class. As an example, selected ground motions, scale factor, and intensity measures (PGA and PGV) of records are shown in Table 4, for free vibration period class 3 and 270 m/s time-averaged shear-wave velocity over the top 30 meters of the subsurface (V_{S30}). Moreover, the response spectra of selected ground motions are visualized in Figure 3.

Selected Earthquake	SF_1	PGA (g)	PGV (cm/s)
Golcuk Earthquake 1999 (M_w = 7.8)	1.18	0.38	63
Pazarcik Earthquake 2023 (M_w = 7.7)	0.91	0.60	170
Pazarcik Earthquake 2023 (M_w = 7.7)	0.65	0.51	87
Duzce Earthquake 1999 (M_w = 7.2)	1.30	1.06	86
Pazarcik Earthquake 2023 (M_w = 7.7)	1.01	0.61	114
San Jacinto 1899 (M_w = 7.15)	1.01	0.75	96
San Jacinto 1 (M_w = 7.25)	3.90	2.30	479
Simi-Santa Rosa 1 (M_w = 6.65)	2.29	0.66	126
San Andreas 1 (M_w = 8.05)	4.99	0.65	79
San Andreas 1 (M_w = 7.95)	1.55	0.69	96
San Andreas 1 (M_w = 7.85)	4.46	0.76	168

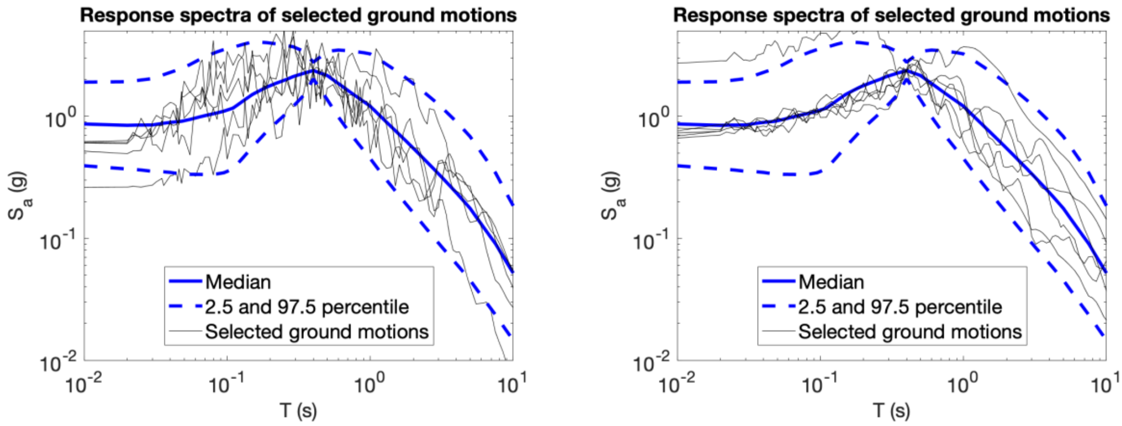
Table 4: An example of selected ground motions, scale factors and intensity measures for period class 3 (V_{S30} = 270 m/s)

Structural Analyses

The NLTH analyses are performed to obtain an artificial dataset for machine learning algorithms. At the beginning, eleven earthquake records are used for non-linear analysis for 3240 various buildings. MIDR values are retrieved from 35640 completed analyses for classifying the performance of buildings as damage state. The results of the analyses, which is the target value of the dataset, were not distributed equally along the determined damage states. For instance, more than 20 thousand analyses (approximately 57% of the dataset) were performed as collapse damage states (Figure 4, light blue). This consequence was interpreted as the rupture scenario produces high ground motion intensity measures. To overcome the issue of imbalanced dataset, two different scale factors are used in terms of varying intensity measures. The first scale factor comes from the Baker and Lee's (2018) algorithm itself, and the second one comes from the Equation 1:

$$SF_2 = SF_1 \times \frac{IDR}{Mean_{MIDR}} \tag{1}$$

where SF_1 and SF_2 represent the first and the second scale factors, respectively. IDR denotes the inter-story drift ratio corresponding to the collapse limit (Table 5), and $Mean_{MIDR}$ indicates the mean MIDR obtained from NLTH analysis while using the first scale factor. IDR is 0.01, and $Mean_{MIDR}$ is 0.046, which is pretty higher than collapse limit. As a result, the second scale factor is calculated by reducing the first scale factor by multiplying with 0.22.



(a) (b)

Figure 3: Response spectra of selected ground motions (a) AFAD database (b) CyberShake database [Figures were plotted by using Baker and Lee's (2018) MATLAB Algorithm]

Damage State	Inter-story Drift Ratio (%)
No Damage	0 – 0.1
Slight Damage	0.1 – 0.2
Moderate Damage	0.2 – 0.5
Extensive Damage	0.5 – 1.0
Collapse	>1.0

Table 5: Inter-story drift ratio (%) boundries (Ghobarah, 2004)

After calculating SF_2, we performed NLTH again by multiplying records with this scale factor. In the end, 71280 analyses are performed to provide a balanced dataset for training machine learning classification algorithms (Figure 4).

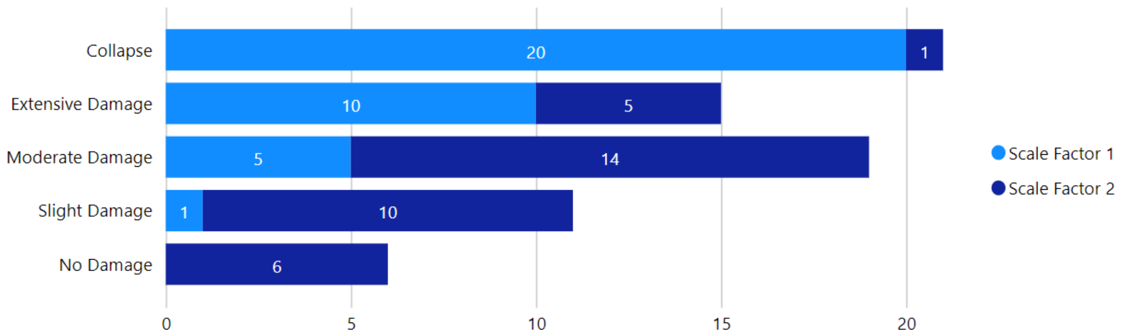


Figure 4: NLTH analysis results as damage states for SF_1 and SF_2 (in thousand)

Application of machine learning algorithms

Machine learning is a broad field that includes various types of training system, and the choice of algorithm type depends on the model's purpose and the available dataset (Alpaydin, 2014). In this study, the objective of the machine learning model is to predict the building damage state based on the input data [building and seismic characteristics (Table 6)]. We developed two prediction models to represent seismic characteristics using two different intensity measures, PGV and PGA. Specifically, this study focuses on classification-related machine learning algorithms, including k-nearest neighbors, support vector machines (SVMs), decision trees, and random forests.

Data Label	**Data Type**	**Data - 1**	**Data - 2**	**…**	**Data - 71280**
Number of Storey	Numerical	3	3	…	5
Number of Span	Numerical	3	3	…	4
Span Length (m)	Numerical	2.5	2.5	…	4
Storey Height (m)	Numerical	2.6	2.6	…	3
Column Area (m^2)	Numerical	0.0625	0.0625	…	0.16
Concrete Strength (MPa)	Numerical	5	5	…	35
First Storey - Commercial Use	Categorical	Yes	Yes	…	No
PGA (g)	Numerical	0.361	0.472	…	0.176
PGV (cm/s)	Numerical	17.668	44.784	…	43.298
Damage State	Categorical	Collapse	Extensive	…	Slight

Table 6: Preview of the dataset

Pre-processing techniques were applied to ensure optimal performance. Normalization is a common best practice in data pre-processing, as it can improve the performance and interpretability of machine learning models. In this study, normalizing numerical data and converting categorical data into numerical values are involved as pre-processing techniques.

In order to evaluate the effectiveness of machine learning techniques in predicting damage states, the entire dataset is divided into two subsets: a training set and a testing set. The training set is used to construct the predictive model, while the testing set is utilized to evaluate the performance of the model. For this study, 80% of the data is allocated to the training set, and the remaining data is allocated to the testing set.

Furthermore, we split the training set into 10 cross-validation sets using the Stratified K-Folds method. Additionally, the machine learning algorithms used in this study have many hyperparameters, so we tuned them to find the best solution for the given dataset. For instance, in the KNN algorithm, we tuned the number of neighbors, weights (the weight of data in each neighborhood), and the distance metric (Euclidean or Manhattan) between neighbors.

Results and Conclusion

Each algorithm was evaluated based on performance metrics such as precision, recall, accuracy, and F1 score. The Random Forest algorithm outperformed the other algorithms in every performance metric for both intensity measures (Table 7). The Random Forest algorithm produced remarkably favorable outcomes with the machine learning model, exhibiting a proportional prediction accuracy of more than 90% across the administered tests.

Classification Algorithm	IM	Precision	Recall	Accuracy	F1 Score
Decision Tree (DT)	PGV	0.90	0.90	0.90	0.89
	PGA	0.89	0.90	0.90	0.89
Support Vector Machine (SVM)	PGV	0.59	0.57	0.61	0.57
	PGA	0.70	0.69	0.70	0.70
Random Forest (RF)	PGV	0.92	0.92	0.92	0.92
	PGA	0.92	0.92	0.92	0.92
K-Nearest Neighbor (KNN)	PGV	0.46	0.45	0.49	0.45
	PGA	0.58	0.58	0.59	0.58

Table 7: Evaluation metrics of machine learning algorithms

Interpretation of the confusion matrix will give detailed information about the result of the machine learning algorithms. Table 8 and Table 9 shows the confusion matrixes for models trained by using datasets with different intensity measures.

		Predicted Class					Total	Recall
		No Dam.	Slight	Moderate	Extensive	Collapse		
Actual Class	No Dam.	1137	70	1	0	0	1208	0.94
	Slight	74	1961	125	0	0	2160	0.91
	Moderate	0	145	3440	115	4	3704	0.93
	Extensive	0	0	146	2568	222	2936	0.87
	Collapse	0	0	3	167	4078	4248	0.96
Total		1211	2176	3715	2850	4304	Accuracy: 0.92	
Precision		0.94	0.90	0.93	0.90	0.95		

Table 8: Confusion matrix of random forest algorithm, dataset with PGV

		Predicted Class					Total	Recall
		No Dam.	Slight	Moderate	Extensive	Collapse		
Actual Class	No Dam.	1137	70	1	0	0	1208	0.94
	Slight	67	1973	119	1	0	2160	0.91
	Moderate	0	140	3436	121	7	3704	0.93
	Extensive	0	3	158	2597	178	2938	0.88
	Collapse	0	2	5	188	4053	4248	0.95
Total		1204	2188	3719	2907	4238	Accuracy: 0.93	
Precision		0.94	0.90	0.92	0.89	0.96		

Table 9: Confusion matrix of random forest algorithm, dataset with PGA

The present study demonstrates the applicability of a machine learning-based model for rapid damage prediction. Such models, developed using the data sets created by realistic three-dimensional structural analyses and post-disaster reconnaissance data, can accelerate the taking of necessary precautions by learning the performance of the existing structures. Furthermore, the resulting fragility functions derived from these models can facilitate more accurate loss calculations.

References

ABYYHY. (1975). Regulation on Buildings to be Built in Disaster Areas, Ministry of Construction and Settlement. (In Turkish: Afet Bölgelerinde Yapılacak Yapılar Hakkında Yönetmelik, Imar ve Iskan Bakanlığı) Ankara, Turkey.

AFAD. Earthquake Department of the Disaster and Emergency Management Presidency, Ankara, Turkey. https://tadas.afad.gov.tr

Akyuz S, Uyan M, Yildirim H. A statistical evaluation on the concrete steel bars used in Turkey. Tech J Assemb Readymix Concrete 1999;35:93–100 [in Turkish].

Alpaydin, E. (2014). Introduction to Machine Learning. Amsterdam University Press.

Azak T.E., Ay B.Ö., Akkar S. (2014). A statistical study on geometrical properties of turkish reinforced concrete building stock, "2nd European Conference on Earthquake Engineering and Seismology" Istanbul, Turkey, 25-29 August 2014.

Baker, J., Bradley, B., & Stafford, P. (2021). Seismic Hazard and Risk Analysis. Cambridge: Cambridge University Press. doi:10.1017/9781108425056

Baker, J. W., Goulet, C., Luco, N., Rezaeian, S., and Teng, G. (2021). "A Subset of CyberShake Ground Motion Time Series for Response History Analysis." Earthquake Spectra. https://doi.org/10.1177/8755293020981970

Baker, J. W., and Lee, C. (2018). "An Improved Algorithm for Selecting Ground Motions to Match a Conditional Spectrum." Journal of Earthquake Engineering, 22(4), 708–723.

Bal, E.I., Crowley H., Pinho, R., Gülay, F.G., Detailed assessment of structural characteristics of Turkish RC building stock for loss assessment models, Soil Dynamics and Earthquake Engineering, Volume 28, Issues 10–11, 2008, Pages 914-932, ISSN 0267-7261, https://doi.org/10.1016/j.soildyn.2007.10.005

Computers and Structures, Inc. (SAP2000). Static and dynamic finite element analysis of structures; 2000.

DEZIM. (2020). The project of updating probable earthquake losses estimation for Istanbul province. Istanbul Metropolitan Municipality, Directorate of Earthquake and Ground Research. https://depremzemin.ibb.istanbul/wp-content/uploads/2020/02/DEZiM_KANDiLLi_DEPREM-HASAR-TAHMiN_RAPORU.pdf (PDF) [in Turkish].

Ghobarah, A. (2004) On drift limits associated with different damage levels. In: International work shop on performance-based seismic design. Department of Civil Engineering, McMaster University; June28–July 1, 2004.

Hubert-Ferrari, A., A. Barka, E. Jacques, S. Nalbant, B. Meyer, R. Armijo, P. Tapponnier and G.C.P.King, Seismic hazard in the Marmara Sea following the 17 August 1999 Izmit earthquake, Nature, 404, 269-272, 2000

Kaur, H., Pannu, H. S., & Malhi, A. K. (2020). A systematic review on imbalanced data challenges in machine learning. ACM Computing Surveys, 52(4), 1–36. https://doi.org/10.1145/3343440

Mangalathu, S., Sun, H., Nweke, C. C., Yi, Z., & Burton, H. V. (2020). Classifying earthquake damage to buildings using machine learning. Earthquake Spectra, 36(1), 183–208. https://doi.org/10.1177/8755293019878137

McKenna, F., Scott, M. H., and Fenves, G. L. (2010). Nonlinear finite-element analysis software architecture using object composition. Journal of Computing in Civil Engineering, 24(1):95-107. doi:10.1061/(ASCE)CP.1943-5487.0000002

Murru, M., Akinci, A., Falcone, G., Pucci, S., Console, R., and Parsons, T. (2016), M ≥ 7 earthquake rupture forecast and time-dependent probability for the sea of Marmara region, Turkey, J. Geophys. Res. Solid Earth, 121, 2679– 2707, https://doi.org/10.1002/2015JB012595

Parsons, T. (2004). Recalculated probability of M >= 7 earthquakes beneath the Sea of Marmara, Turkey. Journal of Geophysical Research. 109. https://doi.org/10.1029/2003JB002667

Parsons, T., S. Toda, R. S. Stein, A. Barka, and J. H. Dieterich (2000), Heightened odds of large earthquakes near Istanbul: An interaction-based probability calculation, Science, 288, 661– 665. DOI: 10.1126/science.288.5466.661

Roeslin, S., Ma, Q., Juárez-Garcia, H., Gómez-Bernal, A., Wicker, J., & Wotherspoon, L. (2020). A machine learning damage prediction model for the 2017 Puebla-Morelos, Mexico, earthquake. Earthquake Spectra, 36(2_suppl), 314–339. https://doi.org/10.1177/8755293020936714

Salmi, Z.J., Khodakarami, M.I. & Behnamfar, F. Development of seismic fragility curves for RC/MR frames using machine learning methods. Asian J Civ Eng 24, 823–836 (2023). https://doi.org/10.1007/s42107-022-00533-w

Stojadinović, Z., Kovačević, M., Marinković, D., & Stojadinović, B. (2021). Rapid earthquake loss assessment based on machine learning and representative sampling. Earthquake Spectra, 38(1), 152– 177. https://doi.org/10.1177/87552930211042393

Ulku O., Atici A.T., Yesilirmak, F., and Hancilar U. (2022). Classifying and predicting earthquake damage by using machine learning after the 2020 Elazığ, Turkey, earthquake, "3rd European Conference on Earthquake Engineering and Seismology" Bucharest, Romania, 4-9 September 2022.

Wu J., Sarno L., (2023). A machine-learning method for deriving state-dependent fragility curves of existing steel moment frames with masonry infills. Engineering Structures, Volume 276, 2023, 115345, ISSN 0141-0296, https://doi.org/10.1016/j.engstruct.2022.115345