

Damage prediction for regular reinforced concrete buildings using the decision tree algorithm

A. Karbassi ^{a,1}, B. Mohebi ^b, S. Rezaee ^c, and P. Lestuzzi^a

^aApplied Computing and Mechanics Laboratory, Ecole Polytechnique Federale de Lausanne, Switzerland

^b Imam Khomeini International University of Qazvin, Iran

^c Blekinge Tekniska Högskola, Sweden

Abstract

To overcome the problem of outlier data in the regression analysis for numerical-based damage spectra, the C4.5 decision tree learning algorithm is used to predict damage in reinforced concrete buildings in future earthquake scenarios. Reinforced concrete buildings are modelled as single-degree-of-freedom systems and various time-history nonlinear analyses are performed to create a dataset of damage indices. Subsequently, two decision trees are trained using the qualitative interpretations of those indices. The first decision tree determines whether damage occurs in an RC building. Consequently, the second decision tree predicts the severity of damage as repairable, beyond repair, or collapse.

Keywords: damage prediction, decision tree, damage index, reinforced concrete, C4.5 algorithm

1. Introduction

Predicting damage in structures as a result of future earthquakes can be a very useful tool for seismic risk mitigation plans. A reliable estimation of damage has wide ranges of application in the seismic vulnerability evaluation of buildings that have not been designed to withstand earthquake loads. Such damage prediction can be used in scenario studies where effects of a single earthquake, often historically significant, on present-day portfolios in a region are evaluated [1].

¹ Corresponding author

Email address : amin.karbassi@epfl.ch (Amin Karbassi)

Postal address: IMAC ENAC EPFL, Station 18, CH-1015 Lausanne, Switzerland. Tel: +41 21 6936382

Equivalent single-degree-of-freedom (SDOF) systems have significant contribution in many research in the field of earthquake and structural engineering [2,3,4]. The response of the multi-degree-of-freedom (MDF) structure including regular RC buildings can be related to the response of an equivalent SDOF system, if the response is controlled by a single mode, determined from a high enough modal participation factor. Different methods also make use of equivalent SDOF systems to predict damage in structures [5]. One useful way to predict damage in scenario studies is to calculate a *damage index* (DI) which normally has a value close to zero if the structure remains elastic and close to 1.0 when the structure reaches complete damage or collapse. The available methodologies in the literature to calculate the damage index can be classified according to the number of parameters used (e.g., single-parameter [6,7,8] or double-parameter [9,10]), type of the concept used (e.g., fatigue-based [11,12,13], energy-based [14,15], or drift-based [16,17]) or according to the assessment level (local [18,19] or global [20,21,22]). The main problem with most of those methodologies, however, is the use of quantitative (numerical) representations of damage to replace the qualitative (nominal) meaning of different damage levels.

A very frequently-used damage index in the literature is the one proposed by Park and Ang [23] shown in Eq. (1).

$$DI_{Park\ and\ Ang} = (u_{max}/u_{mon}) + \beta \cdot E_H / F_y \cdot u_{mon} \quad (1)$$

The term u_{max} in this equation is the maximum deformations under earthquake loads (dynamic analysis), and the terms u_{mon} and F_y , also shown in Fig. 1, are ultimate deformation and the maximum base shear force from pushover analysis, respectively. Moreover, E_H is the non-recoverable dissipated hysteretic energy (Eq. 2), and β is a positive constant between about -0.3 to +1.2 (obtains from 250 experimental tests), which depends on structural characteristics and history of inelastic response [24].

$$E_H = F_y(u_{mon} - u_y) \quad (2)$$

An advantage of Park and Ang's equation is that it has been calibrated with experimental data. However, in some cases, when the system remains in the elastic mode ($E_H=0$), the equation gives DI values way bigger than zero which can be misleading towards the damage evaluation of the building. To overcome this problem, a modified version of Eq. (1) is proposed [25] as shown in Eq. (3).

$$DI_{Kunnath\ et\ al.} = \left(\frac{(u_{max} - u_y^*)}{(u_{mon} - u_y^*)} \right) + \beta \cdot E_H / F_y \cdot u_{mon} \quad (3)$$

The added term u_y^* in this equation is the displacement at yield of the equivalent SDOF system (Fig. 1) used to calculate the damage index.

The variation of damage index values for a series of single-degree-of-freedom (SDOF) systems with different structural properties subjected to multiple earthquakes with different characteristics forms *damage spectra* [26]. To accomplish that, a regression analyses is performed to express the damage index as a function of structural properties and earthquake characteristics [27]. The main problem in developing damage spectra, however, is the damage index *outlier* values. According to the definition of the damage index, once the DI exceeds 1.0, the building is assumed to be in complete damage state. In other words, values higher than 1.0 would not physically make sense as higher DI values do not indicate heavier collapse. However, as the calculation of the damage index from any equation presented in the literature including Eq. (3) is mathematical, the result can be theoretically any value bigger than 1.0. Such values would become outliers in the regression analysis to develop damage spectra. To overcome this problem, as the main novelty in this research, we have replaced the numerical damage spectra concept by a damage predictor algorithm (DPA) that uses the qualitative (nominal) meaning of the damage indices instead of the quantitative (numerical) representation. Consequently, the damage index values are translated into the corresponding damage description and grouped into 4 damage classes from no damage to collapse.

The main objective of this paper is to present a damage predictor algorithms in the form of decision trees for reinforce concrete buildings based on the qualitative meaning of the damage index, considering the soil class of the building's site. The proposed decision trees can be used as the first step of the seismic vulnerability assessment for a group of buildings to determine buildings in dangerous condition, for a more elaborated investigation. A machine learning procedure is applied to train two algorithms for each soil class using multiple nonlinear dynamic analyses performed on SDOF systems with different structural properties, using 612 ground motion records. To keep the calculations simple, the damage predictor algorithms are calculated as functions of the parameters shown in Eq. (4).

$$DIA = f(M, R, PGA, \mu, T, \frac{F_y}{W}) \quad (4)$$

M, R, and PGA in this equation are the magnitude, site-to-source distance, and the peak ground acceleration at the structure's site in g, respectively. The symbol μ denotes the global displacement ductility of the structure, T is the period of the mode of vibration with the highest modal participation factor (normally the first mode), and $\frac{F_y}{W}$ is the normalized yield strength of the structure. The accuracy of the damage predicting algorithms are later evaluated using results obtained from the nonlinear dynamic analyses done on a 3-D model of a 7-storey building. Finally, it is shown that the algorithms identify similar damage levels for reinforce concrete buildings damaged in two earthquake in Athens (1999) and in L'Aquila (2009).

2. Damage index for RC frame buildings

2.1. Structural properties of representing SDOF

The damage predictor algorithms in this article are developed for existing RC frame buildings with no significant vertical or horizontal irregularities. For this reason, equivalent SDOF systems (Fig. 1) with the structural properties shown in Table 1 are used in the non-linear dynamic analyses to calculate the

damage indices. The Takeda hysteresis model [28] is used for the numerical analyses that are performed with the computer program IDARC [29].

2.2. Ground motion characteristics

The proposed algorithm here is not supposed to take into account near-fault effects such as directivity and fling-step effects. Moreover, the number of data of ground motion records with a magnitude bigger than 7 is limited in a way that is difficult to properly develop an algorithm in that range. Consequently, earthquakes with a magnitude (M_s) between 5 and 6.9 that occurred in Europe since 1970 with a site-to-source distance between 10 to 100 km are selected from the European Strong-Motion Data [30]. For that purpose, 412 ground motion records at various stations, located on rock or stiff soil (Fig. 2), and another 200 recorded on soft and very soft soil (Fig.3) are chosen to perform the nonlinear dynamic analyses using the structural properties shown in Table 1.

2.3. Calculation of damage indices

Using the structural properties shown in Table 1, Eq. (3) is applied to calculate the damage indices through conducting 82 480 nonlinear dynamic analyses using the ground motion records shown in Fig. 2 for rock and stiff soil class, and other 40 000 analyses using the records shown in Fig. 3 for soft soil class. Examples of those damage index values are shown in Table 2. The values of F_y , u_y , u_{mon} , and E_H in Eq. (2), are calculated from pushover analysis on the equivalent SDOF systems, as shown in Fig. 1.

2.4. Outlier data in damage index values

According to the basic definition of the damage index in Eq. (3), once the DI exceeds 1.0, the building is assumed to be in complete damage state. In other words, values higher than 1.0 would not make sense physically, and higher DI values do not indicate heavier collapse. However, for several cases, the calculated DI values (examples seen in Table 2) is computed to be much higher than 1.0. Such values

would become *outliers* when a regression analysis is performed to develop the damage spectra, which determine the relation among ground motion characteristics and the structural properties with the damage state in RC frame buildings. To overcome this problem, the qualitative aspect of the damage indices instead of the quantitative meaning is used in this paper. To this end, the relations between the damage levels with Park-Ang damage index values, shown in Table 3, are used to translate and group the calculated values into damage levels from no damage to collapse.

An observation of the “state of the building” column in Table 3 also indicates that the damage index values can be also classified differently from another perspective. Up to a DI value of 0.4, the building is considered to be repairable which leads to small economical loss. For DI’s from 0.4 to 1.0, the building would be in a “beyond repair” damage state which results in a high economical loss as a consequence of the obligation to destroy and replace the building after the earthquake. Having a DI bigger than one means that the building is collapsed (economical loss) leading to loss of life. These three classes of building damage states are important for insurance purposes and public safety management, and therefore are used as a part of the proposed damage predictor algorithms in this paper.

The challenge is to develop a classifier tool that can be trained with the qualitative (nominal) building damage state data instead of the quantitative (numerical) values. This is a supervised machine learning case in which, by definition, an algorithm (classifier) should be trained through analyzing a set of classified examples called *instances* (e.g., combination of ground motion parameters and structural characteristics) as input data to predict the correct class (e.g., the damage state of an RC building) for unseen instances. Each variable from the structural characteristics and the ground motion parameters at each instance are called the *attributes* [32]. Considering that our case is developing a decision support tool to identify vulnerable buildings, the decision tree concept is used in this paper to propose the damage predictor algorithms.

3. Decision trees

Decision trees are produced by developing algorithms which split a dataset (instances) into branch-like segments. These segments form an inverted decision tree that originates with a root node at the top of the tree. Example of such a tree, which has 5 nodes and 6 leaves, is shown in Fig. 4.

The example in Fig. 4 shows that the decision tree can consist of both continuous and discrete attributes. The relationship between the object of analysis that serves as the target field in the data and those which serve as input fields is used to create the decision rule to form the branches or segments underneath the root node. Once the relationship is configured, then one or more decision rules can be derived that describe the relationships between inputs and targets. Rules can be selected and used to display the decision tree, which provides means to visually examine and describe the tree-like network of relationships that characterize the input and target values. Decision rules predict, with a certain level of accuracy, the values of new or unseen observations that contain values for the inputs, but might not contain values for the targets.

3.1. C4.5 algorithm

To generate the decision tree in this paper, the C4.5 learning algorithm [33], also known as a statistical classifier, is used. The C4.5 algorithm, builds decision trees from a set of training data using the information entropy concept. In general, entropy in information theory is the measure of the uncertainty associated with the random variables. As this paper does not intent to go into the details of such mathematical concept, the reader is referred to a related reference [34] for more information.

The training data in supervised machine learning is a set S of already classified instances I_1, I_2, \dots, I_n . Each I_i consists of a number of attributes A_1, A_2, \dots, A_m . The training is augmented with values of C_1, C_2, \dots, C_n where a certain C_i represents the class to which a certain instance I_i belongs. Table 4, a repetition

of Table 2 but with the qualitative damage data, shows a set of 5 instances (I₁ to I₅) each having 6 attributes (A₁=M, A₂=R, etc.).

As seen in Table 4, a damage class (C₁=no damage, C₂= repairable, C₃=beyond repair, and C₄=collapse) is assigned to each instance based on the numerical value of its damage index. Such instances are used to train the decision tree in this paper. At each node of the tree, C4.5 algorithm chooses one attribute of the dataset that most effectively divides its set of samples into subsets enriched in one class or the other. The dividing efficiency is calculated according to the differences in entropy or the information gain that results from choosing an attribute for splitting the dataset. The information entropy for the set S over a chosen attribute (over which the split is being performed) is calculated as follows.

$$E(S) = -\sum_{k=1}^z f_s(k) \log_2 f_s(k) \quad (5)$$

In Eq. (5), z is the number of different values of the chosen attribute and $f_s(k)$ is the frequency of that attribute's k^{th} value in the set S . Based on the definition of entropy, an entropy equal to zero indicates a perfectly classified set. Consequently, attributes with higher entropies have higher potential to become the splitter node to improve the classification. This is measured through the calculation of the gain resulted in from a split over an attribute A_i .

$$G(S, A_i) = E(S) - \sum_{k=1}^z f_s(A_{ik}) \log_2 E(S_{A_{ik}}) \quad (6)$$

In Eq. (6), A_{ik} is the k^{th} value of the attribute A_i , $f_s(A_{ik})$ is the frequency of instances in the set S having A_{ik} as the value for the attribute A_i , and $S_{A_{ik}}$ is the subset of S containing all instances, where A_i is A_{ik} . Finally, the attribute with higher gain is picked as the splitter node to improve the entropy. The data mining task to divide the dataset to develop the damage predictor decision tree algorithms is conducted using WEKA [35] a free software developed at the University of Waikato, New Zealand for Machine Learning.

3.2. Decision tree algorithm for damage prediction

To reduce the prediction error in the proposed decision trees, the damage prediction procedure is developed as a two-phase algorithm using the relations between the damage index values and damage levels in RC buildings, shown in Table 3. In the first phase, a two-class decision tree classifies the status of the structure as *damaged* or *no damage*. Subsequently, in the second phase, a three-class decision tree determines whether the structure classified as damaged in the first phase is *repairable*, *beyond repair*, or has *collapsed*. Each pair of the decision trees has been developed for two soil classes: rock/stiff soil and soft soil. Fig. 5 demonstrates the steps to follow to use these decision tree algorithm for a regular RC building.

3.2.1 Damage or no damage: two-class decision tree algorithm

The distributions of the input attributes according to the damage or no damage outcome for each soil class are shown in Figs. 6 and 7. Using WEKA, a decision tree is developed for each soil class. For the rock and stiff soil class, the decision tree consists of 807 leaves and 1613 nodes. The decision tree for soft soil class, on the other hand, has 262 leaves with 523 nodes. The full two-class decision tree algorithm for each soil class can be downloaded from here².

3.2.2 Extent of damage: three-class decision tree algorithm

The distributions of the input attributes according to repairable, beyond repair, or collapse outcome for each soil class are shown in Figs. 8 and 9. Using WEKA, a decision tree is developed for each soil class. For the rock and stiff soil class, the decision tree consists of 647 leaves and 1293 nodes. The tree for the soft soil class, on the other hand, has 228 leaves and 455 nodes. The full three-class decision tree algorithm for each soil class can be downloaded from here¹.

² <https://documents.epfl.ch/users/k/ka/karbassi/www/>

4. Discussion of the results

4.1. Performance of the decision tree algorithms

The goal of a decision tree is to have a decision tool to predict future outcomes in the most possible precise manner. The performance of the decision tree is, therefore, evaluated according to the instances correctly classified for unseen inputs. For example, the performance of a decision tree with two outcomes as *positive* and *negative* is evaluated according to the ratio of the sum of the *true positive* and *true negative* outcomes to total outcomes. As shown in Table 5, the performance is usually shown as a 2-D matrix, known as the confusion matrix. Each class (outcome) has a row and a column in that matrix. The rows correspond to the actual outcome while the column corresponds to the classified outcome for the class. Elements in the matrix show how well the decision tree has correctly predicted a true outcome. A precise decision tool is one with large numbers on the main diagonal and small, ideally zero, off-diagonal elements.

The *tenfold cross-validation* technique [32] is used here to evaluate the decision trees. In this technique, the dataset is divided into 10 random sections and each time, the tree is trained with 90 percent of the data and tested with the remaining 10 percent. The process is done 10 times and the overall error of the decision tree is calculated by averaging the error from all of the ten folds. Tables 6 to 9 show the confusion matrix for each decision tree developed in the previous section using the tenfold cross validation technique.

As seen from the sum of the numbers on the diagonals of Tables 6 and 7, the two-class decision trees correctly classify the instances in 95% (rock/stiff soil) and 97% (soft soil) of the cases. For the three-class algorithms the damage extent is predicted correctly in 75% of the cases for the rock/stiff soil classes and 81% for soft soil class (Tables 8 and 9).

4.2 Verification of the decision tree algorithms

4.2.1. Comparison with results from the numerical modelling of a multi-storey RC building

In order to verify the outcomes from the developed decision trees in this paper, the overall damage in a seven-storey buildings from the nonlinear dynamic analyses is compared with the predicted damage from those algorithms. The seven-storey RC building in Fig. 10 with a total mass of 892 Ton with the material properties shown in Table 10 is subjected to the series of ground motion records presented in Table 11. The pushover analysis of the building (Fig. 10.b) shows the shorter direction as the weaker direction; therefore, the nonlinear time-history analyses are performed in that direction using an Applied Element-based software [36].

When the algorithms are used to predict damage for the RC building, an equivalent moment of inertia of $0.5I_g$ is used to calculate the building's period of vibration, base shear coefficient, and ductility (Table 12) to incorporate the cracking effect in the concrete [37].

The comparison is shown in Figs 11 and 12. It is important to note that the distance of the points to the boundary of each region does not have any relation with the damage level in the building. In other words, e.g., a circle point in Fig. 11 at a further distance from the edge of the damage region inside that area does not necessarily indicate a higher damage level as a result of that ground motion record.

As seen in Figs. 11, if the nonlinear dynamic analysis results are taken as the benchmark (Fig. 11.a), the two-class algorithm miss-classifies the existence of damage for only one case (ground motion rec. 8). In this case (Fig. 11.b), a conservative error has happened and the two-class algorithm classifies the building as damaged for that record, while the nonlinear dynamic analysis doesn't show any global damage in the building for the same ground motion record.

The classification of the damage extent in the seven-storey RC building for those ground motion records that cause damage in the building (Fig. 11) is shown in Figure 12. It is assumed that yielding in a few columns from the nonlinear dynamic analysis is equivalent to repairable damage. Moreover, extensive yielding in the columns and the beams of the first floor is taken as “beyond repair” status of damage. With such assumptions, it is seen that the three-class algorithm (Figure 12.b) correctly classifies the damage extent in all cases.

4.2.2. Earthquake case in Athens

The Ms=5.8 earthquake in Athens Greece happened on September 7, 1999. It was observed that low-to mid-rise RC buildings with low ductility experienced the highest damage in the meizoseismal area (Table 13) with an epicentral distance between 10 to 20 km [38]. Example of this is the apartment building that suffered a major damage during the earthquake resulting the collapse of 60% of the structure [39]. The dynamic properties for such a building can be estimated as the values shown in Table 14 reflecting the absence of ductility provisions and the poor construction quality.

As shown in Table 15, the algorithms developed in this paper predict similar damage to the observed damage (collapse) for such buildings located in Nea Filadelfeia that was one of the worst affected municipalities.

4.2.3. Earthquake case in L’Aquila

The L’Aquila earthquake occurred in central part of Italy on April 6, 2009. It is reported that about 80% of the total collapse of the RC buildings in downtown of L’Aquila (11 out of 16) were in the southwestern part of the city founded on silt [40]. Those are mainly two- to six-storey residential RC buildings without seismic design that are classified as C according to the vulnerability class of EMS98 [41]. Dynamic properties shown in Table 16 are used to examine the damage prediction from the algorithms for that building class in a similar earthquake scenario with ground motion characteristics shown in

Table 17. The results of such damage prediction from the algorithms that are also similar to the observed damage after the earthquake in 2009 are shown in Table 18.

5. CONCLUSION

This paper presented a qualitative-based damage prediction tool for regular RC buildings using only a few earthquake parameters and structural properties: the magnitude, site-to-source distance, the peak ground acceleration (g) at the building's site, the displacement ductility of the structure, the fundamental period of vibration, and the normalized yield strength of the structure. A learning algorithm in the form of decision trees were trained using results from a large series of nonlinear dynamic analyses using ground motions on rock and stiff soil, and on soft soil. The first decision tree determines whether damage occurs in the target building or not. Consequently, if the first tree shows damage, the second decision tree is used to predict the severity of damage (repairable, beyond repair, or collapse). To verify the damage predicting ability of the developed algorithms, structural damage in real earthquake scenarios in Greece and in Italy and also from a 3-D numerical model of a 7-storey building were used. It was shown that the damage prediction from the algorithms coincided well with the damage observed in all of those benchmarks.

A good application of the developed algorithms in this paper depend on two conditions: (i) that the response of the building is controlled by one single mode, and (ii) that an accurate equivalent SDOF be developed (from pushover analysis) for the target building. Even if the torsional mode is dominant for a building, the first condition can be easily met. The challenge, however, is with the second condition which strongly depends on the shape of the loading vector used in the pushover analysis. If, for example, a bending load shape (which is normally used for regular buildings) is used to determine the equivalent SDOF for such irregular buildings, the algorithms will not provide reliable predictions. The second condition for such buildings, however, can be met through using proper loading vector shapes in the pushover analysis.

From a machine learning point of view, the accuracy of the developed decision trees is related to the availability of different training values for each attribute (i.e., M , R , a_{\max} , μ , T , $\frac{F_y}{W}$). As seen in Figs. 6 and 7, the number of instances for the magnitude, site-to-source, and peak ground acceleration attributes above 6.2, 60 km, and 0.2 g, respectively, are relatively lower because of fewer data in those ranges. Consequently, when the decision trees are used to assess damage for scenarios with a *combination* of the earthquake parameters in those ranges (e.g., $M=6.5$ at $D=65\text{km}$ with a $\text{PGA}=0.3\text{ g}$), the results should be considered with more prudence. Alternatively, when more data becomes available with time, the algorithms can be easily updated and improved.

Finally, from a design point of view, if the hazard de-aggregation information for a region is available, the decision tree algorithms can be used backwards for quick design purposes to determine the required ductility and yield strength for the credible earthquake in the region based on the acceptable damage level in the building code. The seismic hazard information for the region can be used to relate the acceptable level of risk to the peak ground acceleration at the building's site.

Acknowledgement

The authors are grateful for the technical advices received from Dr. Sandro Saitta on the evaluation of the decision tree algorithm performance in this research work, and the three anonymous reviewers for their helpful comments.

REFERENCES

- [1] A. Coburn, R. Spence, Earthquake protection, 2nd ed., Wiley, Chichester, U.K, 2002.
- [2] A.K. Chopra, C.H. Chintanapakdee, Comparing Response of SDF Systems to Near-Fault and Far-Fault Earthquake Motions in the Context of Spectral Regions, Earthquake Engng Struct Dynamics, 30(2001), 1769-1789.
- [3] R. Riddell, J.E. Garcia, E. Garces, Inelastic deformation response of SDOF systems subjected to earthquakes, Earthquake Engng Struct Dynamics, 31(2002), 515-538.
- [4] G.P. Mavroeidis, G. Dong, A.S. Papageorgiou, Near-fault ground motions and the response of elastic and inelastic single-degree-of-freedom (SDOF) systems, Earthquake Engineering and Structural Dynamics, 33(2004), 1023-1049.
- [5] ATC, Improvement of Nonlinear Static Seismic Analysis Procedures, FEMA 440 (2005). Redwood, CA: Federal Emergency Management Agency.
- [6] A. Ghobarah, H. Abou-Elfath, A. Biddah, Response-based damage assessment of structures, Earthquake Engng Struct Dynamics, 28(1999), 79–104.
- [7] R. Perera, C. Huerta, J. M. Orquin, Identification of damage in RC beams using indexes based on local modal stiffness, Construction and Building Materials, 22(2008), 1656-1667.
- [8] A. Massumi, E. Moshtagh, A new damage index for RC buildings based on variations of nonlinear fundamental period, The Structural Design of Tall and Special Buildings, 22(2013), 50-61.
- [9] G.W. Ma, H. Hao, Y. Zhou, Assessment of structure damage to blasting induced ground motions, Engineering Structures, 22(2000), 1378-1389.
- [10] M.E. Rodriguez, D. Padilla, A Damage Index for the Seismic Analysis of Reinforced Concrete Members, Journal of Earthquake Engineering, 13(2009), 364-383.
- [11] A.M. Reinhorn, R.E. Valles, Damage evaluation in inelastic response of structures: a deterministic approach, Technical Report NCEER, State University of New York at Buffalo, 1995.
- [12] R.A. Hindi, R. G. Sexsmith, A Proposed Damage Model for RC Bridge Columns under Cyclic Loading, Earthquake Spectra, 17(2001), 261-290.

- [13] A. Bassam, A.H. Iranmanesh, F. Ansari, A simple quantitative approach for post earthquake damage assessment of flexure dominant reinforced concrete bridges, *Engineering Structures*, 33(2011), 3218-3225.
- [14] A. Benavent-Climent, A seismic index method for vulnerability assessment of existing frames: application to RC structures with wide beams in Spain, *Bulletin of Earthquake Engineering*, 9(2011), 491-517.
- [15] B.L. Wahalthantri, D.P. Thambiratnam, T.H.T. Chan, S. Fawzia, An improved method to detect damage using modal strain energy based damage index, *Advances in Structural Engineering*, 15(2012), 727-742.
- [16] M.A. Sozen, Review of earthquake response of reinforced concrete buildings with a view to drift control, *State-of-the-art in Earthquake Engineering*. Turkish National Committee on Earthquake Engineering: Istanbul, Turkey, (1981), 383–418.
- [17] A. Elenas, K. Meskouris, Correlation study between seismic acceleration parameters and damage indices of structures, *Engineering Structures*, 23(2001), 698-704
- [18] P.S. Rao, B.S. Sarma, N. Lakshmanan, F. Stangenberg, Damage model for reinforced concrete elements under cyclic loading, *ACI Materials Journal*, 95:6(1998), 682–690.
- [19] T. Sain, J.M. Chandra Kishen, Damage indices for failure of concrete beams under fatigue, *Engineering Fracture Mechanics*, 75:14(2008), 4036-4051.
- [20] A.D. Hanganu, E. Onate, A.H. Barbat, A finite element methodology for local/global damage evaluation in civil engineering structures, *Computers & Structures*, 80:20-21(2002), 1667-1687.
- [21] S. Amziane, J.F. Dube, Global RC Structural Damage Index Based on the Assessment of Local Material Damage, *Journal of Advanced Concrete Technology*, 6(2008), 459-468.
- [22] R. Scotta, L. Tesser, R. Vitaliani, A. Saetta, Global damage indexes for the seismic performance assesement of RC structures, *Earthquake Engng. Struct. Dyn.*, 38(2009), 1027–1049.
- [23] Y. J. Park, A. Ang, Mechanistic seismic model for reinforced concrete, *J. of Struct. Eng.* 111(1985), 722–739.
- [24] Y.J. Park, Seismic damage analysis and damage-limiting design of R/C structures, PhD Thesis, Department of Civil Engineering, University of Illinois, 1984.

- [25] S.K. Kunnath, A.M. Reinhorn, R.F., Lobo, IDARC Version 3.0: A program for the inelastic Damage Analysis of Reinforced Concrete Structures, Technical Report NCEER-92-0022, National Center for Earthquake Engineering Research, Buffalo, 1992.
- [26] P. Panyakapo, Evaluation of site-dependent constant-damage design spectra for reinforced concrete structures, *Earthquake Engng. Struct. Dyn.*, 33(2004), 1211–1231.
- [27] Y. Bozorgnia, V.V. Bertero, Damage Spectra: Characteristics and Applications to Seismic Risk Reduction, *Jour. of Struct. Engrg.*, 129(2003), 1330-1340.
- [28] T. Takeda, M.A. Sozen, N.N. Neilsen, Reinforced concrete response to simulated earthquakes, *ASCE J. Struct.*, 96:12(1970), 2557–2573.
- [29] A.M. Reinhorn, S.K. Kunnath, R. E. Valles, IDARC 2D Version 7.0: user manual, Department of Civil Engineering, State University of New York at Buffalo, 2010.
- [30] N. Ambraseys, P. Smit, R. Sigbjornsson, P. Suhadolc, B. Margaris, Internet-Site for European Strong-Motion Data, European Commission, Research-Directorate General, Environment and Climate Programme, 2002.
- [31] Y.J. Park, A. Ang, Y. Wen, Seismic Damage Analysis of Reinforced Concrete Buildings, *J. Struct. Eng.*, 111:4(1985), 740–757.
- [32] I.H. Witten, E. Frank, M.A. Hall, *Data mining: practical machine learning tools and techniques*, Morgan Kaufmann, Burlington, 3rd ed., 2011.
- [33] J.R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, 1993.
- [34] C.E. Shannon, A Mathematical Theory of Communication, *Bell System Technical Journal*, 27:3(1948), 379–423.
- [35] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA Data Mining Software: An Update, *SIGKDD Explorations*, 11:1(2009).
- [36] Applied Science International, *Extreme Loading® for Structures*, version 3.1 B58, 2010.
- [37] T. Paulay, J. N. Priestley, *Seismic Design of Reinforced Concrete and Masonry Buildings*, John Wiley & Sons, 1992.
- [38] A. Elenas, Athens Earthquake of 7 September 1999: Intensity Measures and Observed Damages, *ISET Journal of Earthquake Technology*, 40(2003), 77-97.

- [39] P.D. Kioussis, A.I. Karabinis, Investigation of Collapse Mechanism of an Apartment Building during the Earthquake of Athens on 7-9-1999, Proceedings of the 12th European Conference on Earthquake Engineering, London, (2002), paper 555.
- [40] A. Tertulliani, I. Leschiutta, P. Bordoni, G. Milana, Damage Distribution in L'Aquila City (Central Italy) during the 6 April 2009 Earthquake, Bulletin of the Seismological Society of America, 102(2012), 1543-1553.
- [41] G. Grünthal, European Macroseismic Scale, EMS-98, Centre Européen de Géodynamique et de Séismologie, Luxembourg, vol. 15, 1998.
- [42] ITACA Italian Accelerometric Archive, Database for the Italian strong motion records, <http://itaca.mi.ingv.it>, 2010.

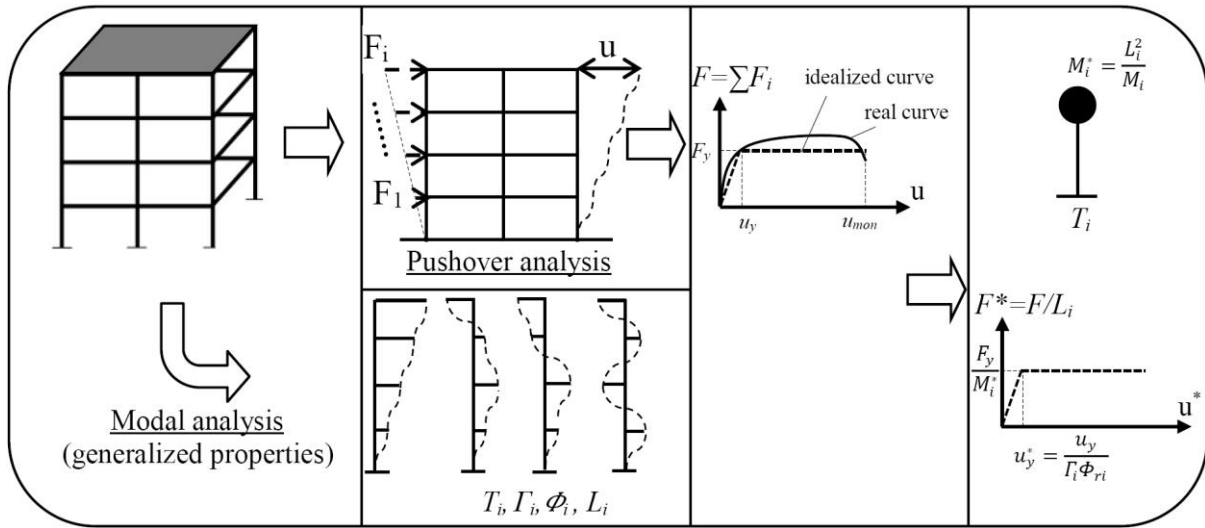


Fig. 1. Equivalent SDOF for the RC frame building

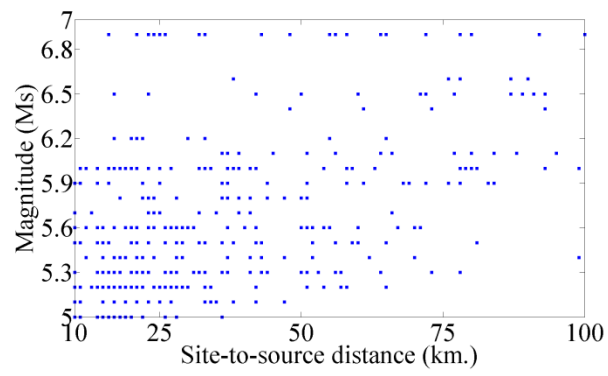


Fig 2. Distribution of the magnitude and site-to-source distance of ground motion records on rock and stiff soil

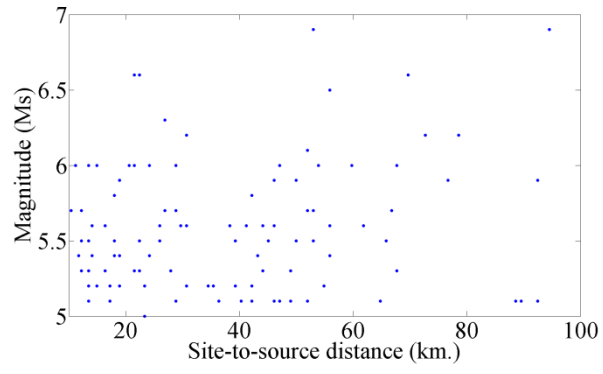


Fig. 3. Distribution of the magnitude and site-to-source distance of the ground motion records on soft soil

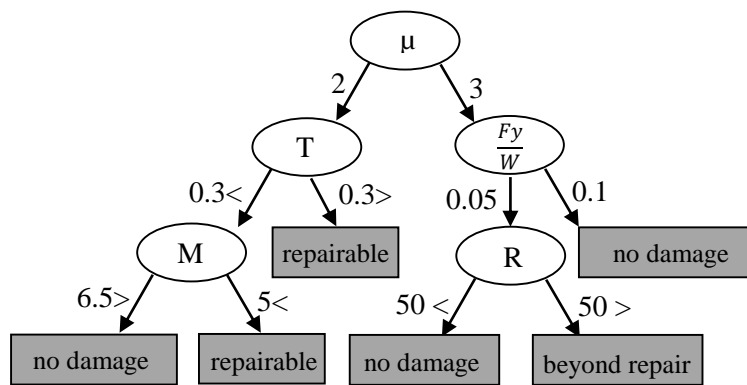


Fig. 4. Example of a decision tree

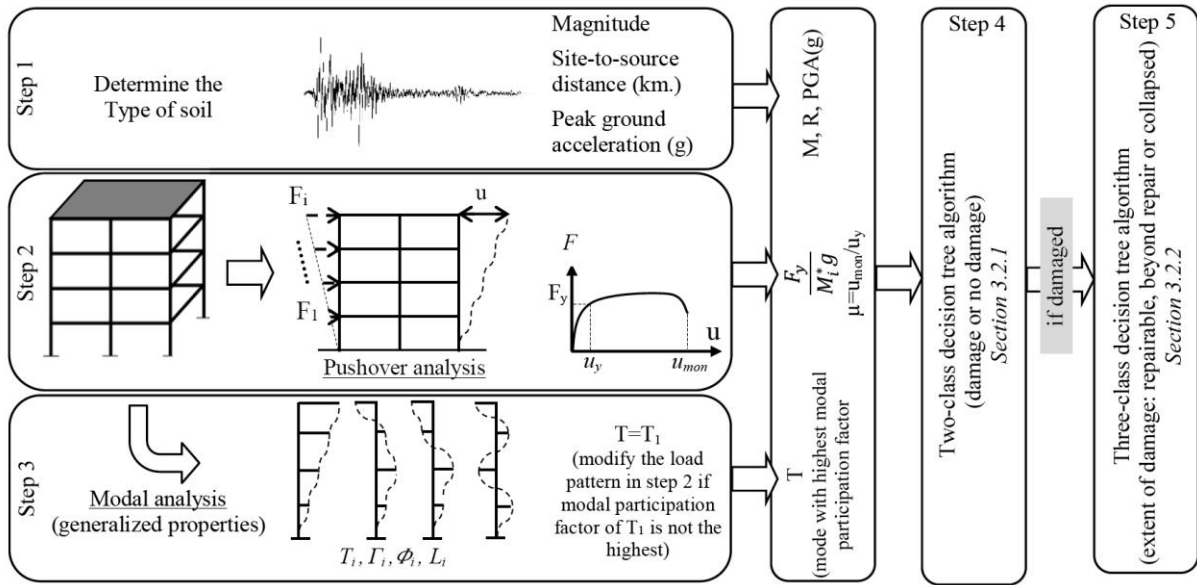


Fig. 5. Flowchart to indicate the step by step use of the proposed algorithms

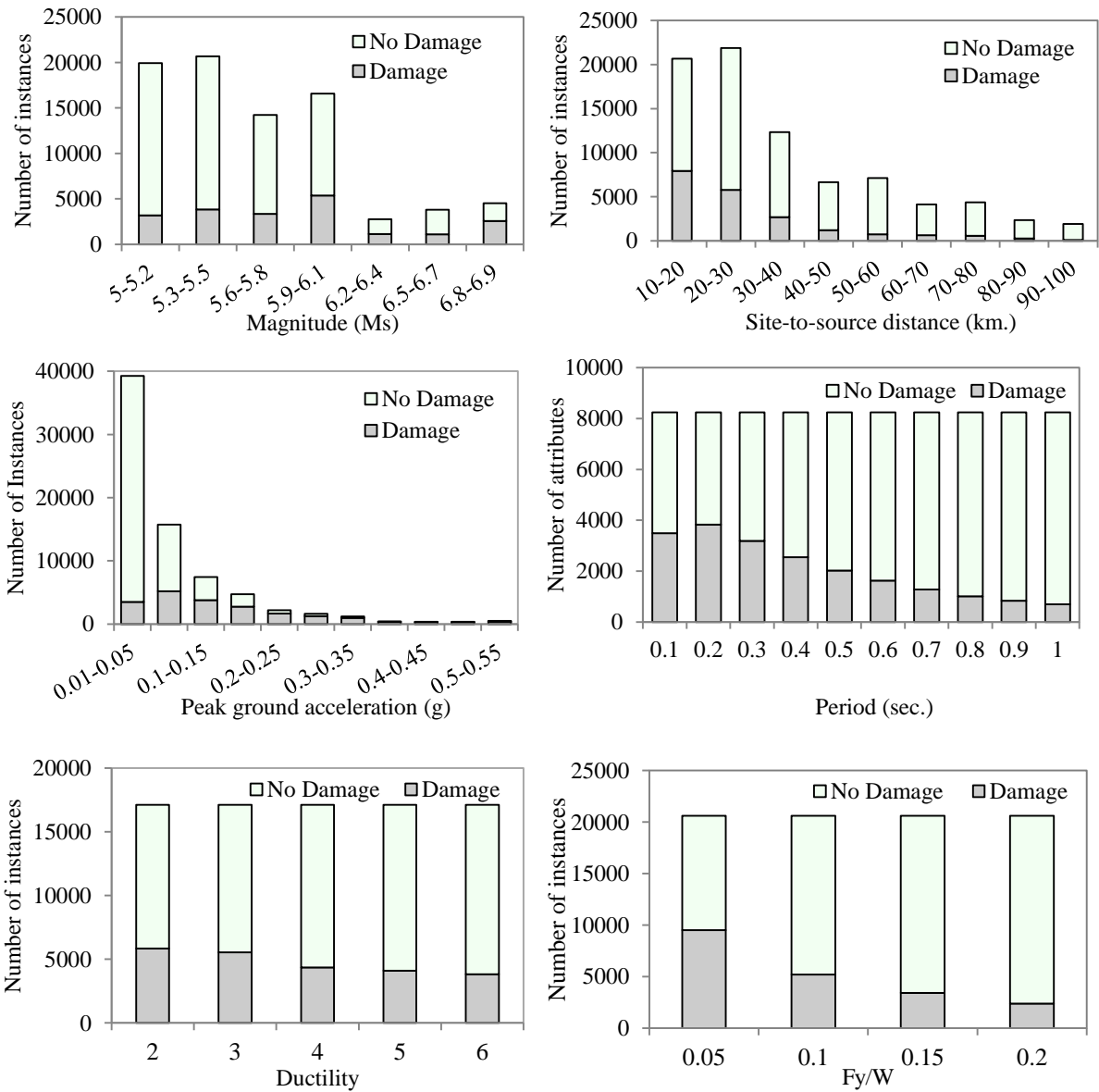


Fig. 6. Distribution of the six attributes according to the decision tree's two outcomes for rock/stiff soil class

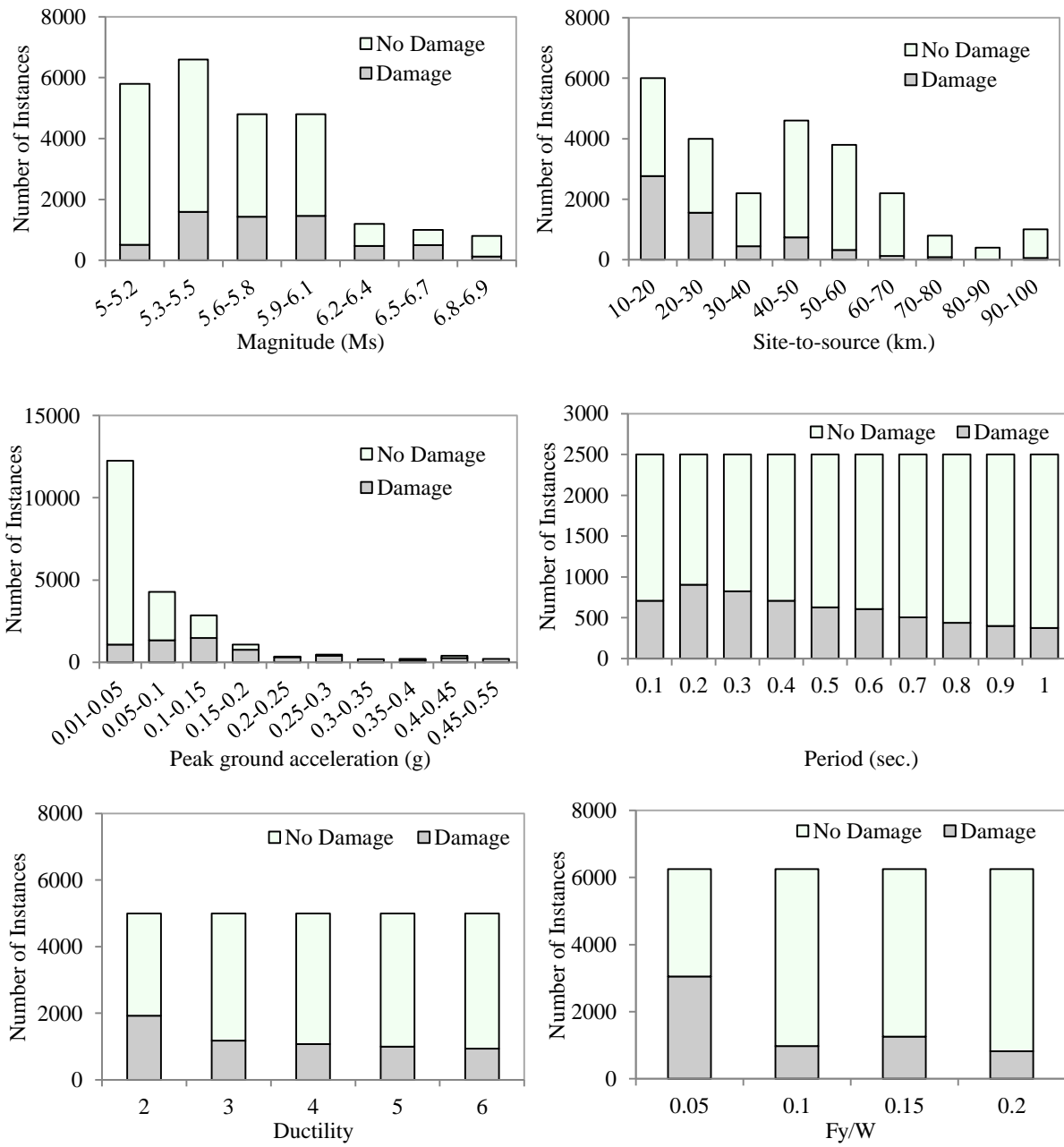


Fig. 7. Distribution of the six attributes according to the decision tree's two outcomes for soft soil class

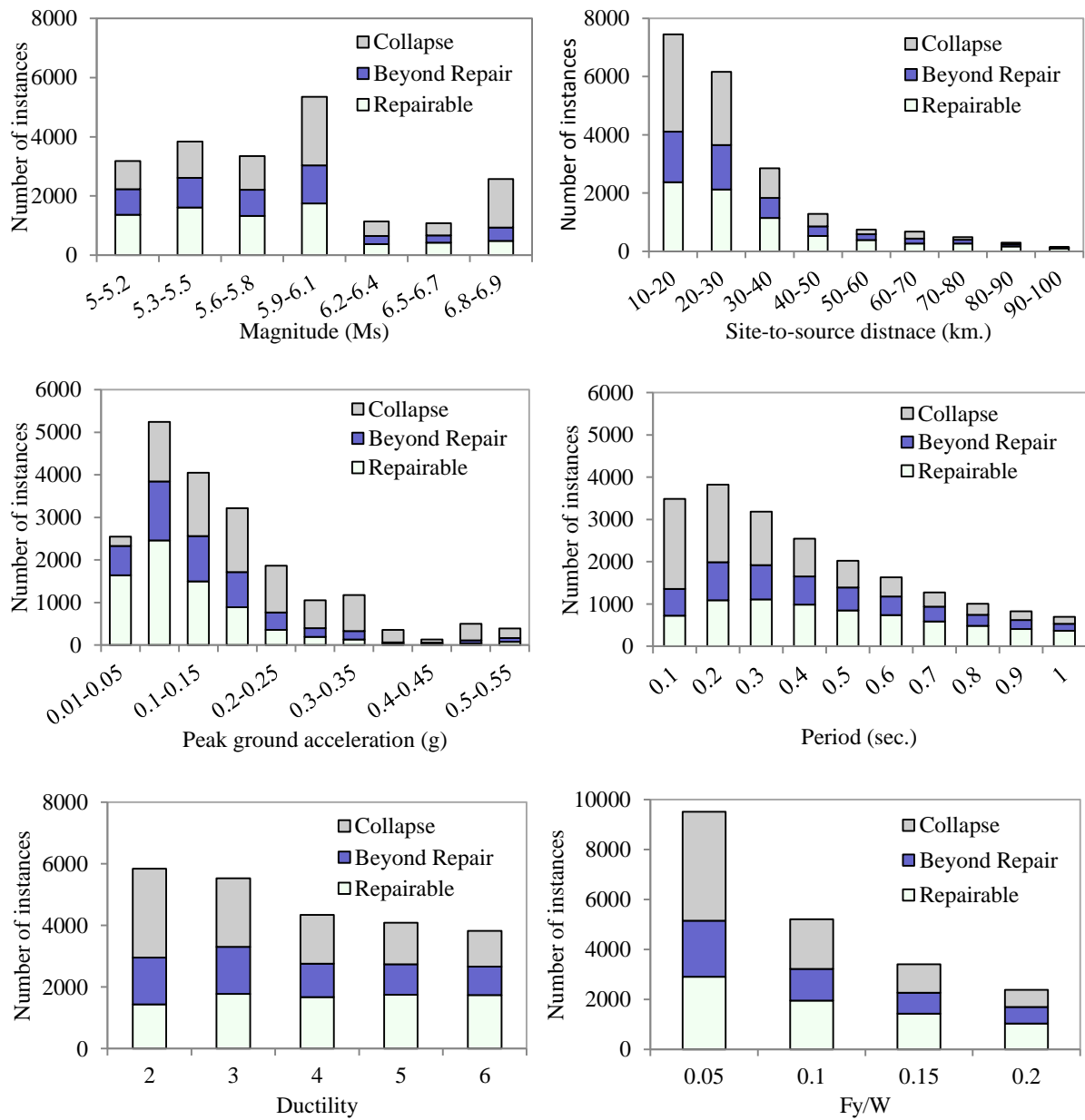


Fig. 8. Distribution of the six attributes according to the decision tree's three outcomes for rock/stiff soil class

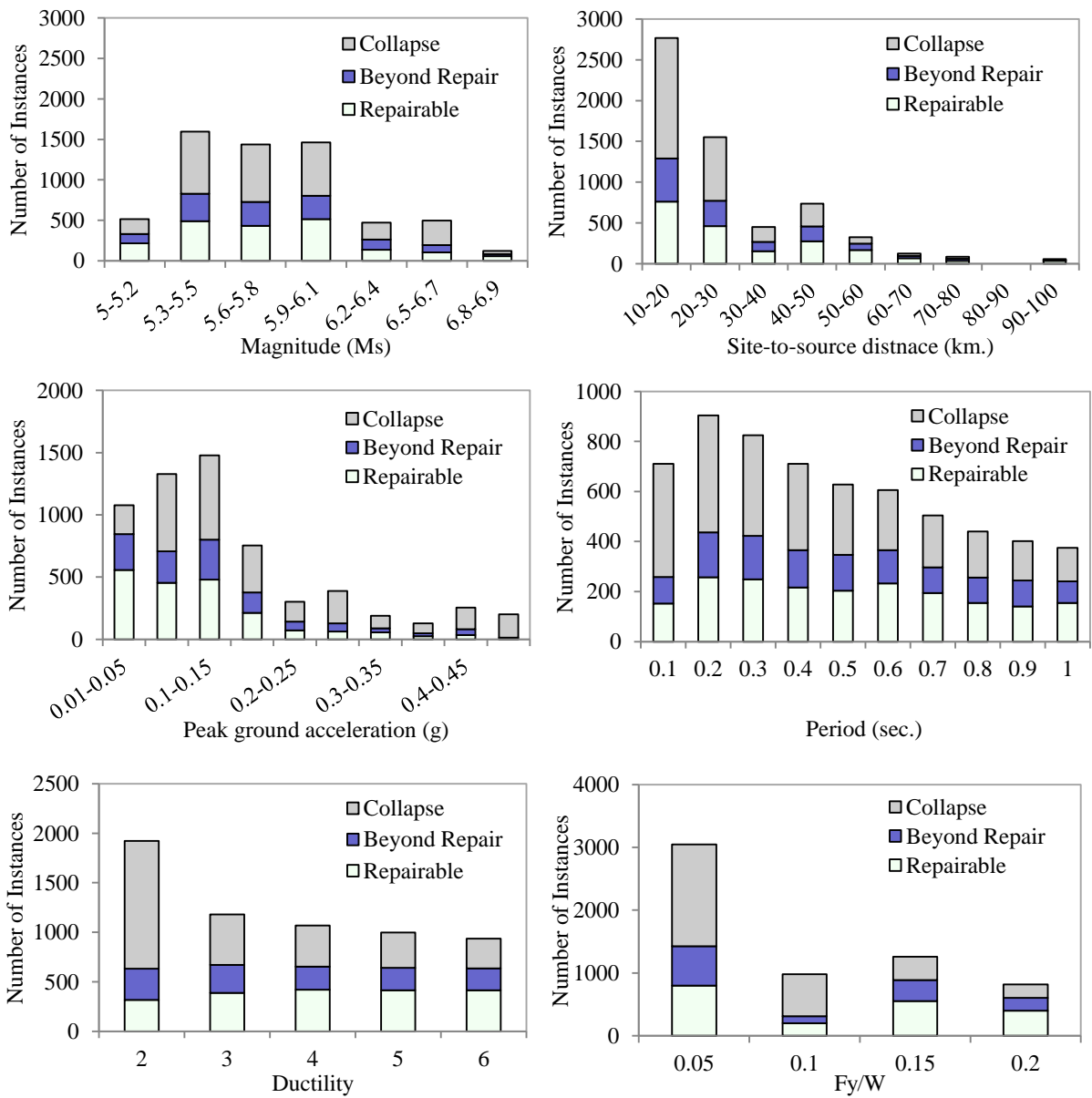


Fig. 9. Distribution of the six attributes according to the decision tree's three outcomes for soft soil class

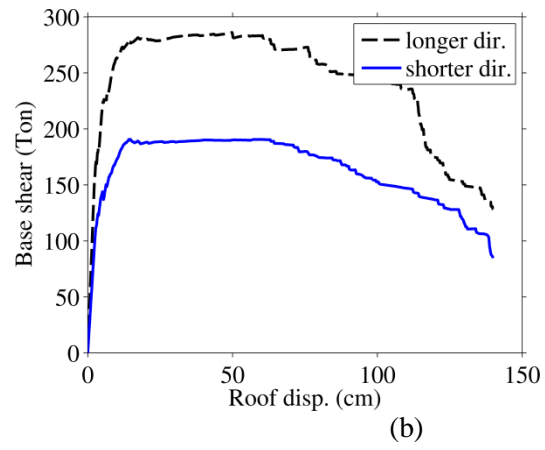
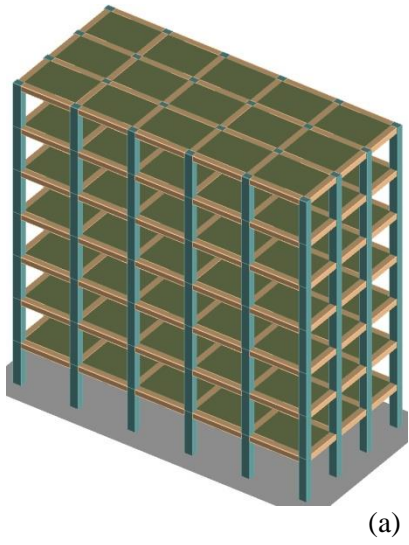


Fig. 10.a. 7-storey RC building used to verify the accuracy of the decision tree algorithms b. pushover curves in the longer and shorter directions of the building

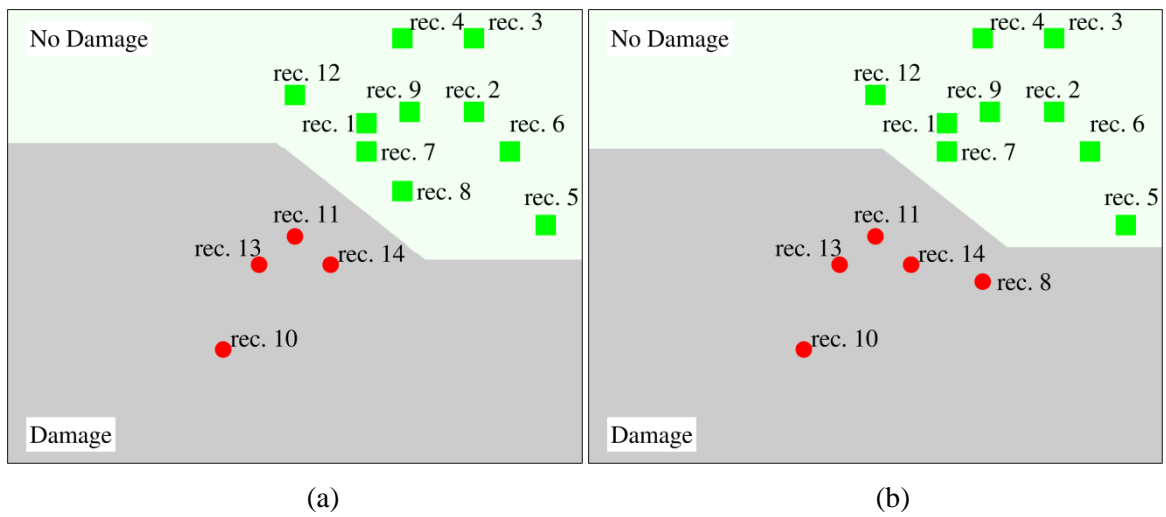


Fig. 11. Comparison of damage prediction for the seven-storey RC building as a result of the earthquake records in Table 11 from a. nonlinear dynamic analysis (used as the benchmark) and b. the two-class algorithm

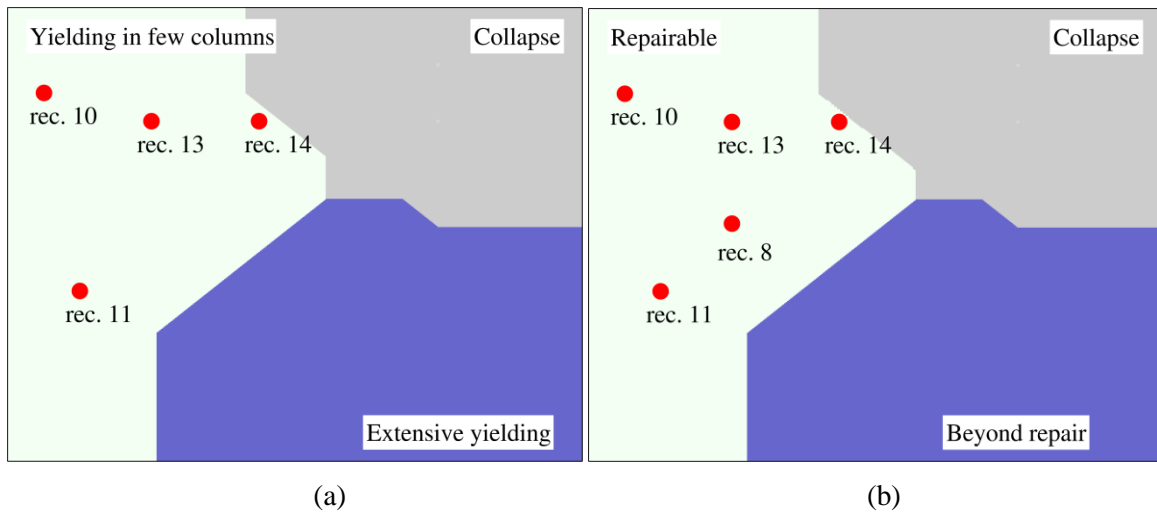


Fig. 12. Comparison of damage extent prediction for the seven-storey RC building as a result of the earthquake records of Table 11 which cause damage in the building (Fig. 11). a. nonlinear dynamic analysis (benchmark) and b. three-class algorithm

Table 1. Structural properties of the RC frame buildings used to develop the damage predictor algorithms

Period	$\frac{Fy}{W}$	Ductility
0.1-1.0	0.05-0.2	2-6

Table 2. Examples of damage index values for different ground motions and structural parameters

M	R (km.)	$a_{max}(g)$	μ	$\frac{F_y}{W}$	T	DI
6.5	42	0.3	2	0.1	0.5	0.76
5.3	18	0.21	3	0.05	0.1	4.23
5.5	15	0.4	3	0.05	0.1	8.84
5.6	26	0.1	4	0.15	0.3	0
5.9	19	0.09	2	0.2	0.2	0.34

Table 3. Relations between damage index and damage levels in RC buildings [31]

Damage extent	Damage index	State of building
Slight	< 0.1	no damage
Minor	0.1-0.25	minor damage
Moderate	0.25-0.4	repairable
Severe	0.4-1.0	beyond repair
Collapse	> 1.0	loss of building

Table 4. Classified instances according to the damage extent for the combination of different ground motions and structural parameters as input attributes

instance	M	R (km.)	$a_{max}(g)$	μ	$\frac{F_y}{W}$	T	damage class
I ₁	6.5	42	0.3	2	0.1	0.5	beyond repair
I ₂	5.3	18	0.21	3	0.05	0.1	collapse
I ₃	5.5	15	0.4	3	0.05	0.1	collapse
I ₄	5.6	26	0.1	4	0.15	0.3	no damage
I ₅	5.9	19	0.09	2	0.2	0.2	repairable

Table 5. Confusion matrix for a two-class decision tree algorithm

Actual \ Classified as	Class 1	Class 2
Class 1	<i>True positive</i> (instances correctly classified as Class 1)	<i>False negative</i> (instances from Class 1, incorrectly labelled as Class 2)
Class 2	<i>False positive</i> (instances from Class 2, incorrectly labelled as Class 1)	<i>True negative</i> (instances that were correctly classified as Class 2)

Table 6. Confusion matrix for the 2-class decision tree for rock/stiff soil

Actual \ Classified as	no damage	damaged
	no damage	73%
damaged	3%	22%

Table 7. Confusion matrix for the 2-class decision tree for soft soil

Actual \ Classified as	no damaged	damaged
	no damaged	damaged
no damaged	74%	1%
damaged	2%	23%

Table 8. Confusion matrix for the 3-class decision tree for rock/stiff soil

Actual \ Classified as	repairable	beyond Repair	collapse
repairable	11%	5%	2%
beyond repair	8%	29%	5%
Collapse	1%	4%	35%

Table 9. Confusion matrix for the 3-class decision tree for soft soil

Actual \ Classified as	Repairable	Beyond Repair	Collapse
repairable	27%	4%	1%
beyond repair	6%	11%	4%
Collapse	1%	3%	43%

Table 10. Material properties for the multi-storey frames shown in Fig. 10

	Concrete	Reinforcement Steel
Modulus of elasticity (GPA)	25	200
Unconfined compression strength (MPA)	23.5	-
Yield strength (MPA)	-	39
Tensile strength (MPA)	2.8	-
Density (kN/m ³)	24	78

Table 11. M-R of the ground motion records used to verify the decision tree algorithms

record no.	M_s	R (km.)	$a_{max}(g)$	soil class
1	6.8	24	0.17	stiff
2	5.8	29	0.16	soft
3	5.6	21	0.13	stiff
4	6.1	16	0.25	soft
5	6.1	28	0.17	soft
6	6.8	29	0.17	stiff
7	5.4	22	0.35	stiff
8	6.5	25	0.35	soft
9	6.3	20	0.15	stiff
10	6.8	15	0.36	soft
11	6.4	13	0.32	soft
12	6.3	12	0.38	stiff
13	6.3	10	0.4	soft
14	6.3	12.5	0.37	soft

Table 12. Structural features of the seven-storey building in the weaker direction

$T_{\text{cracked}}(\text{sec.})$	Fy/W	ductility
0.73	0.2	6.0

Table 13. Characteristic for ground excitation at Nea Filadelfeia [38]

M_s	R (km.)	PGA(g)
5.8	12	0.27

Table 14. Typical structural features of RC buildings which experienced the highest damage in Athens 1999 earthquake

no. of storeys	T (sec.)	Fy/W	ductility
3-7	0.3-0.4	0.05-0.10	2-3

Table 15. Predicting damage from the algorithms for Athens 1999 earthquake scenario using the inputs of Tables 13 and 14

soil class	two-class algorithm	three-class algorithm
rock/stiff	damaged	collapse
Soft	damaged	collapse($\mu=2$) beyond repair ($\mu=3$)

Table 16. Typical structural features of RC buildings which experienced the highest damage in L'Aquila 2009 earthquake

no. of storeys	T (sec.)	Fy/W	ductility
2-6	0.3-0.5	0.15	3

Table 17. Characteristic for ground excitation at seismic station AQK [42]

M_s	R (km.)	PGA(g)
6.3	10.5	0.36

Table 18. Damage predictions from the algorithms for L'Aquila 2009 earthquake scenario using the inputs of Tables 16 and 17

soil class	two-class algorithm	three-class algorithm
Soft	damaged	collapse($T \leq 0.4$) beyond repair ($T = 0.5$)