



Earthquake damage and rehabilitation intervention prediction using machine learning

Sajan K C^a, Anish Bhusal^b, Dipendra Gautam^{c,*}, Rajesh Rupakhetty^c

^a Department of Civil and Environmental Engineering, University of Southern California, University Park Campus, Los Angeles, CA, United States

^b Department of Computer Science, University of Alabama, Huntsville, AL, United States

^c Earthquake Engineering Research Center, Department of Civil and Environmental Engineering, University of Iceland, Selfoss, Iceland



ARTICLE INFO

Keywords:

Earthquake damage
Damage grade
Rehabilitation intervention
Machine learning

ABSTRACT

Predicting damage grade and rehabilitation interventions is important, especially in the aftermath of moderate to strong earthquakes as prioritization of post-earthquake housing recovery needs information regarding the damage extent. Damage prediction is generally performed using fragility functions, which are generally associated with large uncertainties. Moreover, availability and representativeness of fragility functions for a region affected by an earthquake is not always a given. A more realistic prediction of damage might be obtained from methods that rely on relevant attributes of affected buildings. Artificial intelligence-based formulations have huge prospect in this regard. Using the ground shaking intensity measure and detailed building specific features of 549,251 buildings affected by the 2015 Gorkha earthquake in Nepal, this paper assesses efficacy of four common machine learning algorithms for damage grade and rehabilitation intervention prediction. Decision tree, random forest, XGBoost, and logistic regression algorithms are used to prepare machine learning models and test their performance. The XGBoost algorithm is found to predict building collapse and strengthening more accurately than the other algorithms. Moreover, feature importance from the XGBoost model identifies 19 of the top 20 most important features as relevant for both damage grade and rehabilitation intervention prediction.

1. Introduction

Earthquake damage to structures and infrastructures leads to functionality loss, economic loss, fatalities, and injuries. Losses, fatalities, and injuries are dominantly governed by the extent of damage to structural and non structural components. In the aftermath of major earthquakes, it is important to assess the severity and geospatial distribution of structure/infrastructure for proper post-disaster rescue, relief, and recovery operations [30]. First-responders, house owners, facility users, and local and state authorities depend on such information to make informed and timely decisions [28]. Rapid loss assessment after a major disaster is therefore very important for social resilience. Rapid and detailed loss assessment is a resource and time-consuming process.

Many studies have reported detailed damage mechanisms and vulnerability aspects of structures and infrastructures (see e. g. [2,34,20,16,18,39,3;6,41,42]). Such studies lay a strong foundation for resilience and reconstruction planning. For example, Bozza et al. [7] formulated a recovery framework to compare damage states in each recovery phase together with urban resilience scenario. To perform vulnerability, recovery, and resilience studies, a robust database is the primary requirement. Spatially distributed data with

* Corresponding author.

E-mail address: dig17@hi.is (D. Gautam).

the maximum possible details can help develop more realistic models. The Government of Nepal (GoN) launched Household Registration for Housing Reconstruction (HRHR) program after the 2015 Gorkha earthquake (see e.g. [40,36]). The survey collected information from more than one million buildings across the affected areas. It is worth noting that it takes a long time until the holistic outlook of the magnitude and dispersion of damage is obtained after completion of the post-earthquake damage inspection process. This duration depends on the severity of the damage. Interestingly, it took more than two months to complete damage assessment of about 100,000 buildings affected by the M_w 6.7 1994 Northridge earthquake and the reconstruction after the 2015 Gorkha earthquake took more than five years to conclude. Following the 2015 Gorkha earthquake (M_w 7.8), it took about nine months to initiate the holistic damage assessment targeted for reconstruction and repair prioritization for residential buildings. It took about 15 months to complete the damage assessment of over a million buildings in three phases [24]. It should be noted that though color tagging in buildings based on damage level was started immediately after the earthquake, the severity and widely distributed damages compelled the government to initiate the holistic damage assessment program, aiming to collect refined information using detailed damage assessment procedure. Undoubtedly, such programs accumulate detailed information regarding the performance of buildings subjected to a particular earthquake. As significantly large number of resources including manpower need to be deployed for house-to-house survey, it is a time and resource consuming process. Delays in damage/loss assessment translate to delays in repair, relief, and recovery operations.

For these reasons, a mechanism to predict the most likely damage state and repair intervention would be very useful. Such predictions are currently based on seismic fragility functions. Fragility functions are probabilistic models based on using experimental, analytical, and empirical observations, and to some extent expert judgments. These functions link ground shaking intensity to the extent of physical damage on buildings [37]. However, fragility functions rely on building taxonomy, which is an overall grouping of buildings with similar features. Buildings grouped under a given taxonomy class can however have very different attributes, which results in large uncertainties in fragility functions.. Many studies have highlighted the necessity of fragility functions for both pre-earthquake planning and post-earthquake interventions [12]. But, construction of fragility functions requires a large amount of data, which is not always available [4,3,14,19,15]. Apart from this, uncertainties in ground shaking intensities as well as variability in building attributes assigned to a class leads to additional uncertainties in fragility functions. Since most of the fragility functions are derived for global performance of structures and infrastructures, component level damages are hard to predict [17,1].

This prompts to the need for alternative methods, which make use of all relevant attributes of buildings in estimating expected damage during a shaking scenario. Artificial intelligence-based data mining has a promising scope in this endeavor. Though data from post-earthquake damage assessment is gradually growing, use of artificial intelligence in post-earthquake data mining is gaining momentum of late only. This could be because of the recent advancements in information technology that led to an increased computational efficiency and easy access to building-specific features that can be correlated with earthquake induced damage. Several studies have considered post-earthquake damage data to perform machine learning, deep learning, or hybrid learning based studies. Mangalathu et al. [31] used machine learning to classify building damage based on the ATC-20 color tagging scheme (red, yellow, and green) for the 2014 South Napa earthquake. Similarly, Roeslin et al. [38] used machine learning to classify building damage as negligible to slight damage and moderate damage to destruction for the 2017 Puebla-Morelos earthquake. Mangalathu and Burton [29] used a deep learning model to classify building damage using textual damage descriptions for the 2014 South Napa earthquake. Apart from empirical data-based models, several researchers have considered machine learning and deep learning for analytical investigation of earthquake effects in structures and infrastructures [31,30,11]. Artificial intelligence models can be used as an alternative or supplement to earthquake-induced damage assessments based on fragility functions [31]; however, earthquake damage data or experimental data are generally limited. For instance, Mangalathu et al. [31], Mangalathu and Burton [28], and Roeslin et al. [38] used 2276, 3423, and 216 damage data, respectively in their studies. Also, data from these studies were dominated by buildings with green tags or lower damage grades according to the European Macroseismic Scale (EMS-98). The existing literatures highlight that a larger dataset with more buildings would offer further opportunities to develop deeper insights on key building features most vulnerable to seismic damage and improve the accuracy of machine learning models [31,38]. Also, more detailed information about building typology and position, foundation, floor, and roof type, and site condition can be incorporated as input features in machine learning models to enhance efficacy and reliability. Assessing the performance of machine learning methods to predict the rehabilitation intervention based on physical features of buildings and intensity measures would be ideally instrumental in rapid reconstruction, retrofitting, or renovation planning. Although several artificial intelligence-based damage categorization studies are reported in literature, studies that cover rehabilitation prioritization do not exist to the best of the authors' knowledge. Since damage identification alone does not provide a rational way-out for further planning, rehabilitation prioritization is a must to guarantee housing for all earthquake victims.

This study is based on 549,251 building damage data (reinforced concrete, brick masonry, and stone masonry buildings) from the 2015 Gorkha earthquake to study the earthquake-induced building damage grade and repair solution prediction capacity of various machine learning models. The present study incorporates a broad range of data from 11 districts in central Nepal incorporating building level damage data and building specific features to construct machine learning models. Building specific features such as age, location (latitude and longitude), plinth area, height, land surface condition, house position, plan configuration, foundation type, ground floor type, roof type, superstructure type, and distance from the epicenter are considered as the predictor variables. Seismic intensity measures such as peak ground acceleration (PGA), and spectral acceleration at 0.3 s (SA0.3) are also tested. A multi-class EMS-98 [21] damage grade and repair solution for buildings are predicted using the same predictor variables.

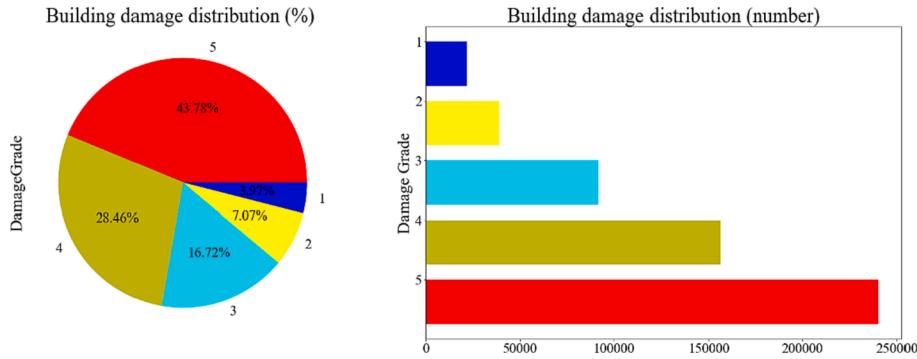


Fig. 1. Proportion of building damage grade for the screened data.

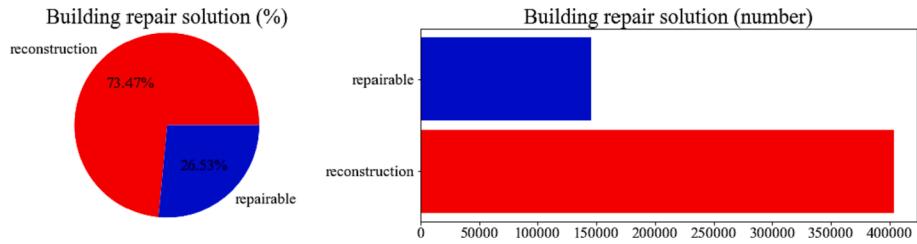


Fig. 2. Proportion of repair solutions for the screened data.

2. Materials and methods

2.1. Data

About 8 million people were affected by the 2015 Gorkha earthquake and the total economic losses related to the earthquake was estimated to be ~ US\$ 7 billion [32]. The earthquake severely damaged 498,852 buildings and minor or moderate damage in 256,697 buildings [32]. Most of the damaged buildings belonged to the low-strength stone masonry that is the most dominant construction system in rural Nepal [20]. Substandard and non-engineered reinforced concrete (RC) and brick masonry constructions were also damaged. Nine months after the earthquake, the Central Bureau of Statistics (CBS) of the Government of Nepal (GoN) started a detailed damaged assessment survey to determine eligibility for housing reconstruction grant. The survey was done using a census-based model in which each household was visited and assessed by a trained enumerator. Over a million houses were surveyed during three phases. Most of the data related to structural damage, building description, as well as household demography were acquired using structured survey forms. Each house surveyed in this program was assigned a damage grade according to the EMS-98 classification and a rehabilitation intervention was recommended. The rehabilitation intervention was divided into two categories: repairable and reconstruction. Repairable means the functionality of buildings can be restored by minor to major repair/retrofitting works. Whereas reconstruction was suggested for buildings damaged beyond repair.

The acquired data were screened to assure completeness of information on each variable. About half of the damaged houses did not have complete data of location coordinate (used as a predictor variable, which is also essential to obtain other predictor variables such as PGA and SA0.3) or had ambiguous information. Such records were excluded to obtain screened data with complete information for all selected variables. As a result, 549,251 building data are used in this study. Each of the selected buildings have a complete set of building-specific information that is pertinent to this study including building location (latitude, longitude, address), damage grade, rehabilitation strategy, age, plinth area, height before damage, ground condition (land surface condition), type of foundation, roof, ground floor, and superstructure, position of building (house position), and plan configuration. Of the 549,251 buildings, 21778(3.97 %), 38,821 (7.07 %), 91,849 (16.72 %), 156,329 (28.46 %), and 240,474 (43.78 %) were assigned with damage grades 1, 2, 3, 4, and 5 respectively as shown in Fig. 1. In addition, 403,525 (73.47 %) and 145,726 (26.53 %) buildings were assigned with repair and reconstruction rehabilitation interventions respectively, as illustrated in Fig. 2.

2.2. Selection of predictor variables

Predictor variables play key roles in predicting the desired target variable (EMS-98 damage grade and strengthening solution in this study). The average number of stories of buildings in the dataset is 3. For residential buildings up to 3 stories, SA0.3 is used as shaking intensity-related predictor variable. The SA0.3 and PGA corresponding to each building site were extracted using the location coordinates of the building from the USGS ShakeMap [45]. Building height is used as one of the predictor variables. To capture the spatial

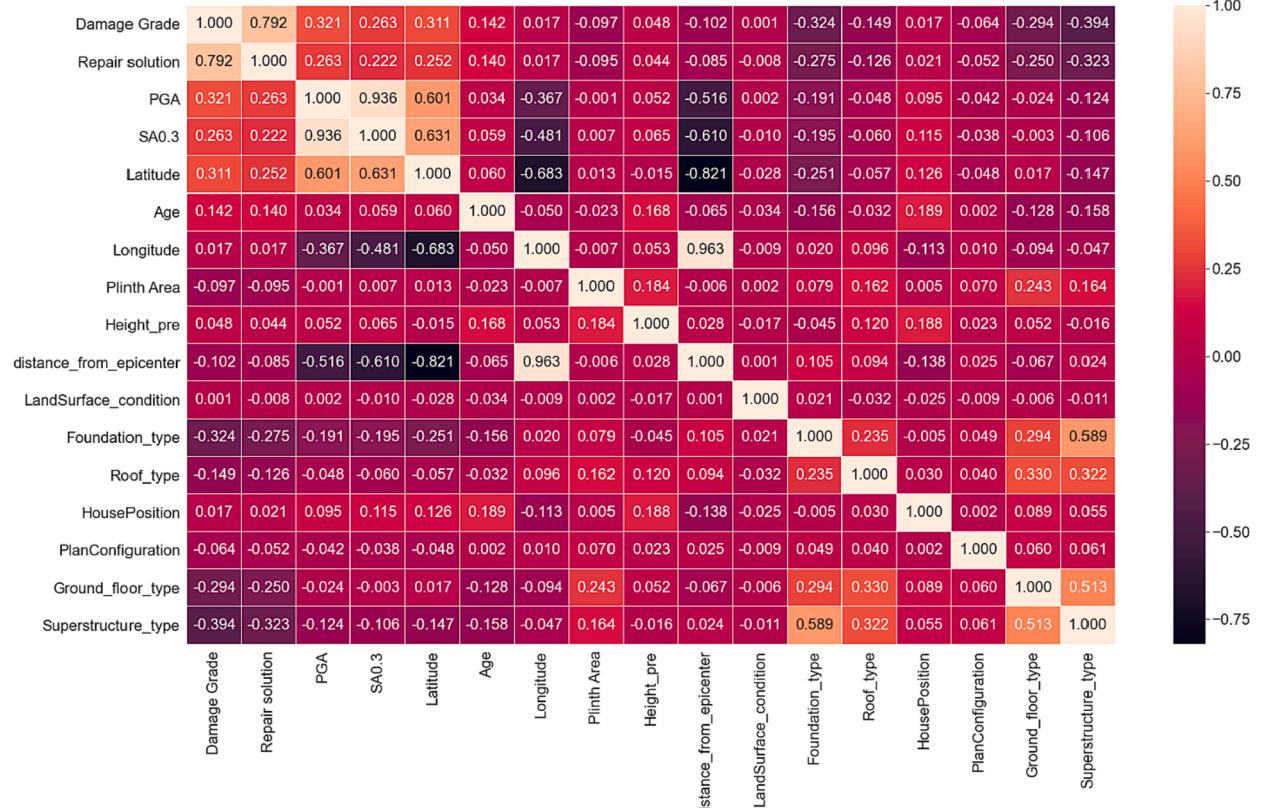


Fig. 3. Pearson's correlation coefficient plot of variables considered during feature selection.

distribution of damage severity, location coordinates of buildings (latitude and longitude) and distance between epicenter and building location are considered as the possible predictor variables. Types of materials used in structural components (foundation, roof, and superstructure) affect expected damage and rehabilitation solutions available. Therefore, foundation type, roof type, ground floor type, and superstructure type based on construction materials are considered as predictor variables. Position of the house in relation to other nearby structures is considered for pounding effects. The land surface condition is another important variable considered as a predictor in this study. A feature is said to be redundant if one or more of the other features are highly correlated with it [22,23]. Some machine learning algorithms are negatively influenced by the highly correlated predictor variables. So, it is advised to remove highly correlated features [44]. Highly correlated features should be removed based on the best practice of machine learning modeling. Fig. 3 shows the relationship between selected variables for model preparation. Probable seismic intensity measures PGA and SA0.3 show a strong linear correlation as shown in Fig. 3. Similarly, distance from the epicenter is strongly correlated to latitude and longitude. This strong correlation is confirmed by Pearson's correlation coefficient as shown in Fig. 3. Therefore SA0.3 is used instead of PGA. Instead of epicentral distance, latitude and longitude of the building were also used by Roeslin et al. [38]. All the other features discussed above are retained for preparing the machine learning models. The description of the processed dataset and predictor variables is presented in Table 1. Pairplots of numerical predictor variables for damage grade and rehabilitation intervention are presented in Figs. 4 and 5, respectively. The figures depict the relationship between the variables and thus facilitate in assessing the interrelation between selected predictors.

2.3. Data processing

As shown in Table 1, the final dataset contains six numerical and seven categorical predictor variables. These categorical predictor variables are non-ordinal. Machine learning algorithms treat the order of encoded number of categorical features as an attribute of significance. In other words, algorithms assign a higher value for more importance or more weightage and vice versa. For the input data that does not have any ranking for categorical values, it can lead to inaccurate predictions and poor performance. Hence, one-hot encoding was employed to such categorical features. One-hot encoding creates a new binary feature for each category and assigns a value of 1 to the feature of each sample that corresponds to its original category. Three binary variables/columns for land surface condition, five for foundation type and ground floor type each, three for roof type, 11 for superstructure type, 10 for plan configuration, and four for house position were formed after employing one-hot encoding.

In machine learning classification, class imbalance refers to the situation when some classes are underrepresented compared to

Table 1

Description of features used in the final dataset.

Feature	Description	Unit	Data type	Categorical type description
Damage Grade	EMS-98 damage grade assigned to building	–	Categorical	1: Damage grade 1 2: Damage grade 2 3: Damage grade 3 4: Damage grade 4 5: Damage grade 5
Repair Solution	Repair solution suggested to damaged building	–	Categorical	1: Repairable 2: Reconstruction
SA0.3	Spectral acceleration at 0.3 s	g	Numerical	
Latitude	North-south position of building	Degree	Numerical	
Longitude	East-west position of building	Degree	Numerical	
Age	Age of building up to earthquake time	Years	Numerical	
Plinth area	Area covered by building at ground floor level	Sq. ft.	Numerical	
Height of the building before the earthquake (Height_pre)	Height of building before earthquake from lowermost point to highest point excluding parapet or gable wall.	ft.	Numerical	
Land surface condition (Landsurface_condition)	Slope of ground on which the building is constructed	–	Categorical	1: Flat 2: Moderate slope 3: Steep slope
Type of foundation (Foundation_type)	Material used to construct building foundation	–	Categorical	1: Mud mortar -stone/brick 2: Cement – stone/brick 3: RCC 4: Bamboo/Timber 5: Other
Ground floor type (Ground_floor_type)	Material used to construct building ground floor	–	Categorical	1: Mud 2: Brick/Stone 3: Timber 4: RCC 5: Other
Type of roof (Roof_type)	Material used to construct building roof	–	Categorical	1: Bamboo/Timber – light roof 2: Bamboo/Timber – heavy roof 3: RC (reinforced concrete)/RB (reinforced brick) /RBC (reinforced brick concrete) 1: Adobe/mud 2: Mud mortar - stone 3: Stone 4: Cement mortar - stone 5: Mud mortar - brick 6: Cement mortar - brick 7: Timber 8: Bamboo 9: RCC (Non-engineered) 10: RCC (Engineered) 11: Other
Type of superstructure (Superstructure_type)	Gravity and lateral load-carrying structural system	–	Categorical	1: Square 2: Rectangle 3: T-shape 4: L-shape 5: U-shape 6: E-shape 7: H-shape 8: Multi-projected 9: Building with central courtyard 10: Other
Plan configuration (PlanConfiguration)	Plan type of building	–	Categorical	1: Not attached 2: attached on 1 side 3: attached on 2 side 4: attached on 3 side
Position of house (HousePosition)	Position of building relative to the nearby building	–	Categorical	

others. In this study, data of damage grades 1, 2, and 3 classes are less than that for damage grades 4 and 5 due to dominance of highly vulnerable buildings in the affected areas. Accordingly, reconstruction as the rehabilitation intervention has a higher prevalence (see Figs. 1 and 2). The ratio of the number of buildings with damage grades 1, 2, 3, 4, and 5 is 11.03:7.17:4.21:1.78:1. For rehabilitation interventions, repairable and reconstruction, the ratio is 2.77:1. This type of skewed distribution can make many machine learning algorithms less effective, especially in predicting minority class examples [46]. Synthetic Minority Oversampling Technique (SMOTE) is one of the dominant methods used as a remedy for the class imbalance problem. It is based on sampling data from the minority class by simply generating data points on the line segment connecting a randomly selected data point and one of its K-nearest neighbors [13]. It over-samples minority classes by creating synthetic examples. Further information regarding SMOTE can be found in the

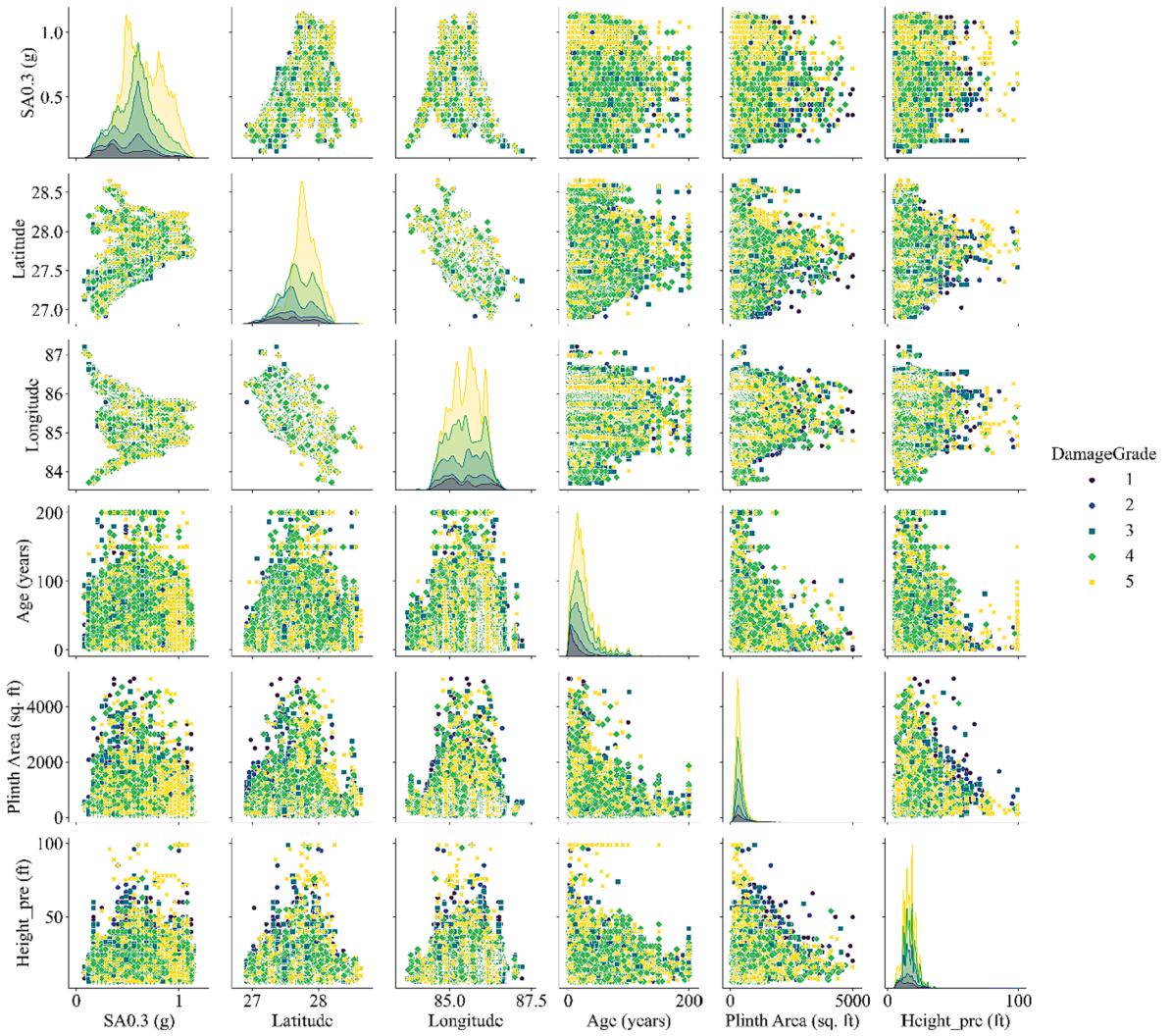


Fig. 4. Pairplot of numerical features for damage grade prediction.

papers by Chawla et al. [9] and Elreedy and Atiya [13].

The final dataset is split into 67 % and 33 % fractions as training and testing datasets, respectively. In other words, 367,998 and 181,253 building data were used for training and testing, respectively for both target variables (damage grade and rehabilitation intervention). The SMOTE method for removing class imbalance is employed on the training dataset. This training data is used to build various machine learning models for each target variable. The performance of these models is evaluated based on their prediction accuracies on testing data.

3. Machine learning application

In recent years AI techniques have been rapidly and widely used in various domains of engineering. Among them, machine learning, deep learning, and pattern recognition have recently acquired considerable attention in the field of structural and earthquake engineering [43]. Machine learning is a major subfield of artificial intelligence. Machine learning algorithms are of three types: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning makes use of labeled input and output datasets whereas unsupervised learning is used to analyze and cluster unlabeled datasets. This study uses supervised machine learning classification algorithms. Each of the algorithms used in this study is briefly described in the following sections.

3.1. Decision tree

A decision tree is a flow-chart-like tree structure, where each internal node denotes a test in an attribute, each branch represents an outcome of the test, and leaf nodes represent classes or class distributions. The topmost node in a tree is the root node. The basic



Fig. 5. Pairplot of numerical features for strengthening solution prediction.

algorithm for decision tree induction is a greedy algorithm that constructs decision trees in a top-down recursive divide-and-conquer manner. The tree starts as a single node representing the training samples. If the samples are all of the same class, then the node becomes a leaf and is labeled with that class. Otherwise, the algorithm uses an entropy-based measure known as information gain as a heuristic to select the attribute that will best separate the samples into individual classes. This attribute becomes the test or decision attribute at the node. A branch is created for each known value of the test attribute, and the samples are partitioned accordingly. The algorithm uses the same process recursively to form a decision tree for the samples at each partition. Once an attribute has occurred at a node, it needs not be considered in any of the node's descendants. In the development of the decision tree, the training set space ($X_1, X_2, X_3, \dots, X_p$) is first divided into I distinct and non-overlapping regions R_1, R_2, \dots, R_I . The splitting to divide the regions is carried out based on the value of the Gini Index (GI) and Entropy (E). The GI measures the frequency at which any element of the dataset will be mislabeled when it is randomly labeled. A feature with a lower value of the GI is chosen for the split. Entropy (E) is a measure of information that indicates the disorder of the features with the target. Similar to the GI, the optimum split is chosen by the feature with less entropy value and are defined as follows:

$$GI = \sum_{i=1}^k \hat{p}_{mi}(1 - \hat{p}_{mi}) \quad (1)$$

$$E = - \sum_i^k \hat{p}_{mi} \log \hat{p}_{mi} \quad (2)$$

Where \hat{p}_{mi} represents the portion of training observations in the m^{th} region from the i^{th} class. Then, the average of the response

values of the training dataset is considered to predict the observation within the same region; the same process is iterated to grow the decision tree [25].

3.2. Random forest

The random forest consists of an ensemble of decision trees grown from a randomized variant of trees. The random forest algorithm was proposed by Breiman [8]. With some exceptions, a random forest classifier has all the hyperparameters of a decision tree classifier. It introduces extra randomness when growing trees. In other words, instead of searching for the best feature when splitting a node, it searches for the best feature among a random subset of features which yields a better model. The training of random forest is done via the bagging method where majority class voting is used as the final class. Further details regarding random forest algorithm can be found in the papers by Breiman [8] and Parmar et al. [35].

3.3. Xgboost (*eXtreme gradient boosting*)

XGBoost is a decision tree-based ensemble algorithm that uses a gradient boosting framework. It was developed by Tianqi Chen. After its unveiling in 2016, it has been widely used across many disciplines. It is known to have improved performance and speed than other algorithms. Further details regarding XGBoost can be found in the contribution by Chen and Guestrin [10].

3.4. Logistic regression

Logistic regression is applicable for both binary class and multi-class classification problems. For the binary class, it is defined as:

$$\hat{y} = \begin{cases} 0 & \text{if } \hat{p} < 0.5 \\ 1 & \text{if } \hat{p} \geq 0.5 \end{cases} \quad (3)$$

$$\hat{p} = \frac{x^T \theta}{1 + e^{-t}} \quad (4)$$

where \hat{y} is the prediction, \hat{p} is the probability, x is a vector, θ is the model parameter vector, and σ is the logistic function. For multi-class problems with classes $C > 2$, training data with N samples such that $\{(x_i, y_i)\}_i^N \subset \mathbb{R}^d \times \{1, 2, 3, \dots, C\}$, where x_i is independent and identically distributed from an unknown probability distribution over the random vectors (X, Y). In multi-class penalized logistic regression, the conditional class probabilities are estimated via logit stochastic models as follows:

$$\left\{ \begin{array}{l} P(Y = 1|X = x; w) = \frac{e^{(\beta_1^T x)}}{1 + \sum_{c=1}^{C-1} e^{(\beta_c^T x)}} \\ \vdots \\ P(Y = C|X = x; w) = \frac{1}{1 + \sum_{c=1}^{C-1} e^{(\beta_c^T x)}} \end{array} \right\} \quad (5)$$

where $w = [\beta_1^T; \beta_2^T; \dots; \beta_{C-1}^T]$, $w \in \mathbb{R}^{(C-1)d}$ is a collection of different parameter vectors of m and is equal to $C - 1$ linear models [26]. Details regarding the binary and multi-class logistic regression can be found in the papers by Karshmakers et al. [26] and Bewick et al. [5].

3.5. Performance evaluation of machine learning models

The performance of machine learning models is evaluated by prediction accuracy along with other metrics such as precision, recall, F_1 score, and confusion matrix. The percentage of predicted classes of target variables (damage grades and rehabilitation intervention) that are correctly assigned by the machine learning algorithms is called the precision. Whereas recall is defined as the percentage of actual classes of target variables that are correctly assigned by the machine learning algorithm. The ability of machine learning models to accurately predict damage grades/rehabilitation intervention is indicated by a high value of precision and recall. Eqs. (6) and (7) depict precision and recall of a classification algorithm used in machine learning.

$$\text{Precision} = \frac{\text{Number of true positive}}{\text{Sum of number of true positive and false positive}} \quad (6)$$

$$\text{Recall} = \frac{\text{Number of true positive}}{\text{Sum of number of true positive and false negative}} \quad (7)$$

F_1 score calculated as the harmonic mean of precision and recall is a single representative metric that accounts for both precision

Table 2

Damage grade prediction accuracy of various machine learning algorithms.

Algorithm	Data set	Target feature	Precision	Recall	F ₁ score	Data set accuracy
Decision Tree	Training	Grade 1	0.8	0.64	0.71	
		Grade 2	0.33	0.46	0.38	
		Grade 3	0.44	0.41	0.43	0.538
		Grade 4	0.4	0.46	0.43	
		Grade 5	0.72	0.67	0.69	
	Testing	Grade 1	0.38	0.58	0.46	
		Grade 2	0.25	0.16	0.19	
		Grade 3	0.34	0.43	0.38	0.535
		Grade 4	0.46	0.43	0.45	
		Grade 5	0.73	0.70	0.72	
Random Forest	Training	Grade 1	0.80	0.63	0.71	
		Grade 2	0.29	0.48	0.36	
		Grade 3	0.42	0.41	0.42	0.530
		Grade 4	0.39	0.44	0.41	
		Grade 5	0.75	0.61	0.67	
	Testing	Grade 1	0.35	0.66	0.46	
		Grade 2	0.25	0.13	0.17	
		Grade 3	0.35	0.32	0.33	0.528
		Grade 4	0.44	0.41	0.42	
		Grade 5	0.69	0.74	0.71	
XGBoost	Training	Grade 1	0.80	0.68	0.73	
		Grade 2	0.41	0.50	0.45	
		Grade 3	0.42	0.48	0.44	0.583
		Grade 4	0.49	0.50	0.49	
		Grade 5	0.79	0.70	0.74	
	Testing	Grade 1	0.43	0.60	0.50	
		Grade 2	0.33	0.17	0.22	
		Grade 3	0.43	0.34	0.38	0.577
		Grade 4	0.49	0.50	0.49	
		Grade 5	0.70	0.78	0.74	
Logistic Regression	Training	Grade 1	0.73	0.60	0.66	
		Grade 2	0.28	0.37	0.32	
		Grade 3	0.25	0.33	0.29	0.440
		Grade 4	0.23	0.34	0.27	
		Grade 5	0.71	0.44	0.55	
	Testing	Grade 1	0.33	0.59	0.42	
		Grade 2	0.16	0.10	0.13	
		Grade 3	0.24	0.15	0.18	0.435
		Grade 4	0.35	0.25	0.29	
		Grade 5	0.53	0.71	0.61	

Table 3

Rehabilitation intervention prediction accuracy of various machine learning algorithms.

Algorithm	Data set	Target feature	Precision	Recall	F ₁ score	Data set accuracy
Decision Tree	Training	Repairable	0.79	0.85	0.82	0.827
		Reconstruction	0.86	0.80	0.83	
		Repairable	0.59	0.59	0.59	0.783
		Reconstruction	0.85	0.85	0.85	
		Repairable	0.76	0.84	0.80	0.809
	Testing	Reconstruction	0.86	0.78	0.82	
		Repairable	0.58	0.58	0.58	0.777
		Reconstruction	0.85	0.85	0.85	
		Repairable	0.83	0.92	0.87	0.879
		Reconstruction	0.93	0.84	0.88	
Random Forest	Training	Repairable	0.73	0.54	0.62	0.824
		Reconstruction	0.85	0.93	0.89	
		Repairable	0.64	0.87	0.74	0.774
		Reconstruction	0.91	0.72	0.80	
		Repairable	0.60	0.38	0.46	0.767
	Testing	Reconstruction	0.80	0.91	0.85	
		Repairable	0.60	0.38	0.46	
		Reconstruction	0.85	0.93	0.89	
		Repairable	0.64	0.87	0.74	
		Reconstruction	0.91	0.72	0.80	
XGBoost	Training	Repairable	0.64	0.87	0.74	0.774
		Reconstruction	0.91	0.72	0.80	
		Repairable	0.60	0.38	0.46	0.767
		Reconstruction	0.80	0.91	0.85	
		Repairable	0.60	0.38	0.46	
	Testing	Reconstruction	0.85	0.93	0.89	
		Repairable	0.64	0.87	0.74	
		Reconstruction	0.91	0.72	0.80	
		Repairable	0.60	0.38	0.46	
		Reconstruction	0.80	0.91	0.85	
Logistic Regression	Training	Repairable	0.64	0.87	0.74	0.774
		Reconstruction	0.91	0.72	0.80	
		Repairable	0.60	0.38	0.46	0.767
		Reconstruction	0.80	0.91	0.85	
		Repairable	0.60	0.38	0.46	
	Testing	Reconstruction	0.85	0.93	0.89	
		Repairable	0.64	0.87	0.74	
		Reconstruction	0.91	0.72	0.80	
		Repairable	0.60	0.38	0.46	
		Reconstruction	0.80	0.91	0.85	

		Predicted Class					
		1	2	3	4	5	R
Actual Class	1	129432 (16.1%)	21132 (2.6%)	6588 (0.8%)	2969 (0.3%)	997 (0.1%)	68%
	2	43983 (5.5%)	66618 (8.3%)	31775 (3.9%)	14730 (1.8%)	4012 (0.5%)	50%
	3	13118 (1.6%)	31066 (3.9%)	67168 (8.3%)	36247 (4.5%)	13519 (1.7%)	48%
	4	3775 (0.5%)	12363 (1.5%)	29459 (3.7%)	78618 (9.8%)	36903 (4.6%)	50%
	5	1320 (0.2%)	1283 (0.1%)	6376 (0.8%)	24383 (3%)	127756 (15.9%)	70%
	P	80%	41%	42%	49%	79%	58%

		Predicted Class					
		1	2	3	4	5	R
Actual Class	1	4329 (2.4%)	998 (0.6%)	782 (0.4%)	634 (0.4%)	444 (0.2%)	60%
	2	2536 (1.4%)	2144 (1.2%)	3500 (1.9%)	2972 (1.64%)	1659 (0.9%)	17%
	3	1670 (0.9%)	1747 (1%)	10217 (5.6%)	10576 (5.8%)	6100 (3.4%)	34%
	4	825 (0.5%)	996 (0.5%)	6004 (3.3%)	25704 (14.2%)	18060 (10.0%)	50%
	5	706 (0.4%)	643 (0.3%)	3244 (1.8%)	12660 (7.0%)	62103 (34.3%)	78%
	P	43%	33%	43%	49%	70%	58%

Fig. 6. Performance of XGBoost algorithm for prediction of building damage grade: (a) confusion matrix for training dataset and (b) confusion matrix for testing dataset. P and R stand for precision and recall values, respectively, whereas numbers 1, 2, 3, 4, and 5 denote damage grades 1, 2, 3, 4, and 5 respectively.

		Predicted Class				
		0	1	R		
Actual Class	0	223173 (41.3%)	47189 (8.7%)	92%		
	1	18825 (3.5%)	251537 (46.5%)	84%		
	P	83%	93%	88%		

		Predicted Class				
		0	1	R		
Actual Class	0	25915 (14.3%)	22175 (12.2%)	54%		
	1	9907 (5.5%)	123256 (68%)	93%		
	P	72%	85%	82%		

Fig. 7. Performance of XGBoost algorithm for prediction of building repair solution: (a) confusion matrix for training dataset (b) confusion matrix for testing dataset. P and R stand for precision and recall values, respectively, whereas numbers 0 and 1 denote repairable and reconstruction as rehabilitation interventions.

and recall. Similarly, accuracy is defined as the ratio of the number of correct predictions made by the machine learning model to the total number of input samples. The F_1 score and prediction accuracy can be represented as follows:

$$F_1 \text{ Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Sum of precision and recall}} \quad (8)$$

$$\text{Prediction accuracy} = \frac{\text{Number of true positive} + \text{Number of true negative}}{\text{sum of true positive, true negative, false positive, and false negative}} \quad (9)$$

A confusion matrix is a tabular way of representing the performance of machine learning models. Each element in a confusion matrix denotes the number of predictions for each class of target variable made by the model that classified the classes correctly or incorrectly. Each diagonal element in the confusion matrix denotes the classes of target variables that are classified correctly by the machine learning model whereas the off-diagonal elements represent the classes that are incorrectly predicted. Decision tree, random forest, XGBoost, and logistic regression classifiers are used for both target variables (damage grade and rehabilitation intervention).

4. Results and discussions

The performance of all four machine learning algorithms used to predict multi-class damage grade and binary class rehabilitation

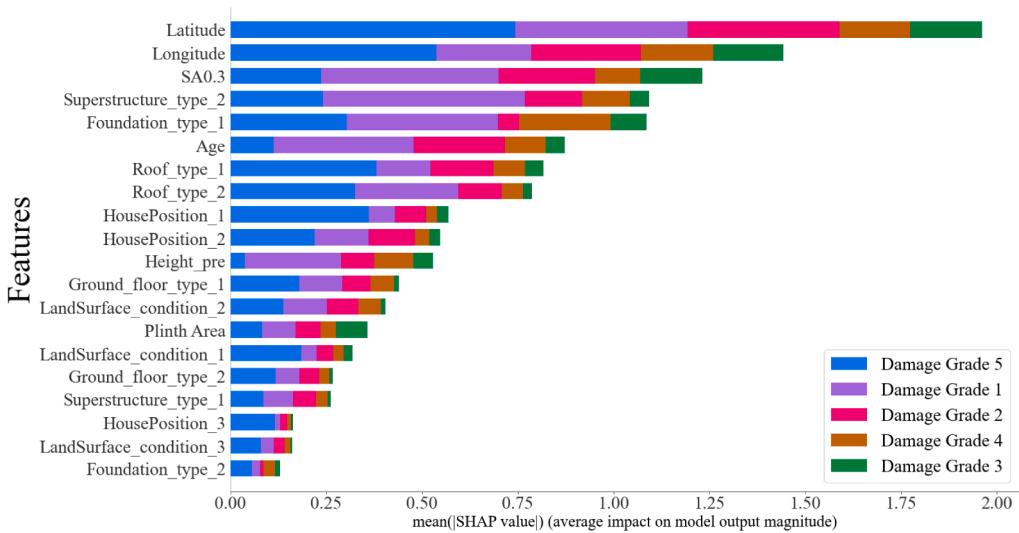


Fig. 8. Feature importance based on mean SHAP value for damage grade in test data using XGBoost algorithm.

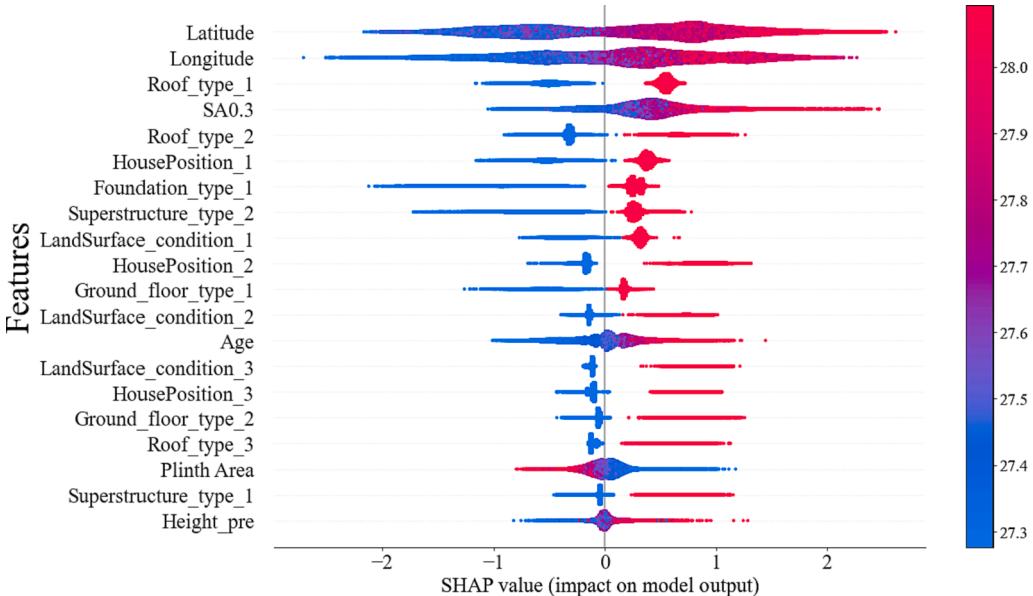


Fig. 9. Feature importance based on SHAP value for rehabilitation intervention in test data using XGBoost algorithm.

intervention for buildings in terms of precision, recall, F1 score, and accuracy is shown in Tables 2 and Table 3, respectively.

As shown in Tables 2 and 3, XGBoost has the best training and testing accuracy for both target variables followed by the decision tree algorithm. The XGBoost algorithm, although never tested before for similar applications, shows a great promise in predicting damage and rehabilitation interventions for residential buildings in high seismicity region. The XGBoost has F₁ score of 0.5, 0.22, 0.38, 0.49, and 0.74; precision value of 0.43, 0.33, 0.43, 0.49, and 0.70; and recall value of 0.60, 0.17, 0.34, 0.50, and 0.78, for damage grades 1, 2, 3, 4, and 5, respectively. It infers that when the XGBoost model predicts that a building is of damage grade 5, it is correct 70 % of the time, and it correctly predicts 78 % of the building within damage grade 5. Moreover, this model has around 50 % precision and recall values for damage grades 1 and 4. However, damage grades 2 and 3 have low precision and recall values. This is due to the lack of a specific pattern for these damage grades apropos to damage grades 1, 4, and 5. Usually the field assessment staff fail to accurately discern between grades 2 and 3 so grade 2 is usually converged to grade 1 and grade 3 is converged to grade 4. In this way, it is likely that the actual number of grade 2 and grade 3 buildings would be rather underestimated. Enhancing recording efficacy by an experienced assessor can improve prediction accuracies for both damage grades.

For rehabilitation intervention prediction, XGBoost has F₁ scores of 0.62 and 0.89, precision values of 0.73 and 0.85, recall values

of 0.54 and 0.93 for repairable and reconstruction interventions, respectively. This indicates that when XGBoost model states that a building should be reconstructed, it is correct 85 % of the time. XGBoost correctly predicts 93 % of buildings that need to be reconstructed after an earthquake for the given shaking intensity. As an example, the confusion matrices for damage grade and rehabilitation interventions (repair solutions) prediction using the XGBoost algorithm is shown in Figs. 6 and 7, respectively.

One helpful way to interpret the prediction of machine learning models is to use the SHAP (SHapley Additive exPlanations) approach [27]. This approach provides Shapley values for the input features based on which feature importance for the model can be obtained. Figs. 8 and 9 show the SHAP feature importance on the test dataset for damage grade and rehabilitation intervention for the XGBoost algorithm. In Figs. 8 and 9, the number after the features corresponds to the corresponding type of category of each feature as shown in the last column of Table 1. Figs. 8 and 9 highlight the most important 20 features in descending order of importance for the corresponding target variable. It is interesting to note that out of the 20 most important features, 19 of them are common in both target variables. Latitude, longitude, SA0.3, superstructure_type_2 (Mud mortar -stone typology, see table 2), and foundation_type_1 (Mud mortar -stone/brick type, see Table 2) are the top five most important features for damage grade prediction. Similarly, for rehabilitation intervention prediction, latitude, longitude, roof_type_1 (bamboo/timber -light roof), SA0.3, and roof_type_2 (bamboo/timber heavy roof) are the top five most important features. Position features of the building (latitude and longitude) were also the most important top two features in the study by [38].

5. Conclusions

This paper assesses the effectiveness of four common machine learning algorithms to predict earthquake induced damage grade and rehabilitation intervention in residential buildings. For this, a dataset of 549,251 buildings collected after the 2015 Nepal earthquake is used. Non-ordinal categorical features are encoded using one-hot encoding, and SMOTE is used to handle the class-imbalance issue. Data used to develop machine learning models contains seismic and building-specific detail (building typology/superstructure type, plinth area, height, type of foundation, ground floor, roof based on construction material, house position, plan configuration, land surface condition, age, building location (latitude and longitude), and spectral acceleration) features. Thus, this study tested four common machine learning algorithms and the best accuracy is found to be at 58 % using the XGBoost algorithm, which is superior to decision tree, random forest, and logistic regression algorithms. The XGBoost algorithm therefore seems promising for multi-class prediction problems. The XGBoost is equally efficient in binary class prediction problems too. The prediction accuracy for repair solutions is found to be satisfactory (82 % using XGBoost). Key building features which govern damage during earthquakes have been identified using the SHAP approach. This information can be useful in building more refined taxonomy for vulnerability classification. It can also be useful in preparing guidelines for building design and construction, vulnerability assessment, and formulating seismic risk reduction policies. With the large dataset complied in this study, deep learning-based models can be explored in the future. Similarly, other algorithms that have shown promising results for similar database can be explored to assess the efficacy of machine learning models.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to thank National Reconstruction Authority (NRA) for providing the dataset for this study. Rajesh Rupakhetty acknowledges support from the University of Iceland Research Fund. The two anonymous reviewers are acknowledged for their constructive feedbacks and discussions, which greatly enhanced the quality of the manuscript.

References

- [1] Adhikari, Rabindra, and Dipendra Gautam. 2019. "Component Level Seismic Fragility Functions and Damage Probability Matrices for Nepali School Buildings." *Soil Dynamics and Earthquake Engineering* 120: 316–19. 10.1016/j.soildyn.2019.02.009.
- [2] Augenti, Nicola, Edoardo Cosenza, Mauro Dolce, Gaetano Manfredi, Angelo Masi, and Linda Samela. 2004. "Performance of School Buildings during the 2002 Molise , Italy , Earthquake" 20 (July): 257–70. 10.1193/1.1769374.
- [3] B. Bessason, J.Ó. Bjarnason, R. Rupakhetty, Statistical Modelling of Seismic Vulnerability of RC, Timber and Masonry Buildings from Complete Empirical Loss Data, *Engineering Structures* 209 (2020), <https://doi.org/10.1016/j.engstruct.2019.109969>.
- [4] B. Bessason, J.Ó. Bjarnason, R. Rupakhetty, Statistical Modelling of Seismic Vulnerability of RC, Timber and Masonry Buildings from Complete Empirical Loss Data, *Engineering Structures* (2020), <https://doi.org/10.1016/j.engstruct.2019.109969>.
- [5] V. Bewick, L. Cheek, J. Ball, Statistics Review 14: Logistic Regression, *Critical Care*. BioMed Central. (2005), <https://doi.org/10.1186/cc3045>.
- [6] M. Biglari, A. Formisano, B.H. Hashemi, Empirical Fragility Curves of Engineered Steel and RC Residential Buildings after Mw 7.3 2017 Sarpol-e-Zahab Earthquake, *Bulletin of Earthquake Engineering* 19 (6) (2021), <https://doi.org/10.1007/s10518-021-01090-4>.
- [7] A. Bozza, D. Asprone, F. Parisi, G. Manfredi, Alternative Resilience Indices for City Ecosystems Subjected to Natural Hazards, *Computer-Aided Civil and Infrastructure Engineering* 32 (7) (2017), <https://doi.org/10.1111/mice.12275>.

- [8] L. Breiman, Random Forests, *Machine Learning* 45 (1) (2001) 5–32.
- [9] Chawla, Nitesh V, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. “SMOTE: Synthetic Minority over-Sampling Technique.” *Journal of Artificial Intelligence Research* 16: 321–57.
- [10] Chen, Tianqi, and Carlos Guestrin. 2016. “XGBoost: A Scalable Tree Boosting System.” In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13–17-Aug:785–94. 10.1145/2939672.2939785.
- [11] H.V. Dang, M. Raza, T.V. Nguyen, T. Bui-Tien, H.X. Nguyen, Deep Learning-Based Detection of Structural Damage Using Time-Series Data, Structure and Infrastructure Engineering (2020), <https://doi.org/10.1080/15732479.2020.1815225>.
- [12] A.S. Elnashai, L. Di Sarno, Fundamentals of Earthquake Engineering, From Source to Fragility, Wiley (2008). 10.1002/9780470024867.
- [13] D. Elreedy, A.F. Atiya, A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for Handling Class Imbalance, *Information Sciences* 505 (December) (2019) 32–64, <https://doi.org/10.1016/j.ins.2019.07.070>.
- [14] C. Del Gaudio, M.T. De Risi, P. Ricci, G.M. Verderame, Empirical Drift-Fragility Functions and Loss Estimation for Infills in Reinforced Concrete Frames under Seismic Loading, *Bulletin of Earthquake Engineering* (2019), <https://doi.org/10.1007/s10518-018-0501-y>.
- [15] D. Gautam, Observational Fragility Functions for Residential Stone Masonry Buildings in Nepal, *Bulletin of Earthquake Engineering* (2018), <https://doi.org/10.1007/s10518-018-0372-2>.
- [16] D. Gautam, R. Adhikari, R. Rupakhetty, Seismic Fragility of Structural and Non-Structural Elements of Nepali RC Buildings, *Engineering Structures* 232 (January) (2021), 111879, <https://doi.org/10.1016/j.engstruct.2021.111879>.
- [17] D. Gautam, R. Adhikari, R. Rupakhetty, Seismic Fragility of Structural and Non-Structural Elements of Nepali RC Buildings, *Engineering Structures* 232 (2021), 111879, <https://doi.org/10.1016/j.engstruct.2021.111879>.
- [18] D. Gautam, N. Chettri, K. Tempa, H. Rodrigues, R. Rupakhetty, Seismic Vulnerability of Bhutanese Vernacular Stone Masonry Buildings: From Damage Observation to Fragility Analysis, *Soil Dynamics and Earthquake Engineering* 160 (2022), 107351, <https://doi.org/10.1016/j.soildyn.2022.107351>.
- [19] Gautam, Dipendra, Giovanni Fabbrocino, and Filippo Santucci de Magistris. 2018. “Derive Empirical Fragility Functions for Nepali Residential Buildings.” *Engineering Structures*. 10.1016/j.engstruct.2018.06.018.
- [20] D. Gautam, H. Rodrigues, K.K. Bhetwal, P. Neupane, Y. Sanada, Common Structural and Construction Deficiencies of Nepalese Buildings, *Innov Infrastruct Solut* 1 (1) (2016) 1, <https://doi.org/10.1007/s41062-016-0001-3>.
- [21] Grunthal, G. 1998. “European Macroseismic Scale 1998.” Luxembourg.
- [22] Isabelle Guyon, André Elisseeff, An Introduction to Variable and Feature Selection, *Journal of Machine Learning Research* 3 (2003) 1157–1182.
- [23] Mark A. Hall. Correlation-Based Feature Selection for Machine Learning, Department of Computer Science, University of Waikato, Hamilton, New Zealand, 1999.
- [24] Housing Recovery and Reconstruction Platform (HRRP). 2020. “The path to housing recovery, Nepal earthquake 2015: Housing reconstruction, Government of Nepal.
- [25] Gareth James, Daniela Witten, Trevor Hastie. *An Introduction to Statistical Learning with Applications in R*, Springer Texts in Statistics, 2013.
- [26] P. Karsmakers, K. Pelckmans, J.A.K. Suykens, Multi-Class Kernel Logistic Regression: A Fixed-Size Implementation, In *IEEE International Conference on Neural Networks - Conference Proceedings* 1756–61 (2007), <https://doi.org/10.1109/IJCNN.2007.4371223>.
- [27] Lundberg, Scott M, and Su-In Lee. 2017. “A Unified Approach to Interpreting Model Predictions.” In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–77.
- [28] Mangalathu, Sujith, and Henry V. Burton. 2019a. “Deep Learning-Based Classification of Earthquake-Impacted Buildings Using Textual Damage Descriptions.” *International Journal of Disaster Risk Reduction* 36 (March). 10.1016/j.ijdr.2019.101111.
- [29] ———. 2019b. “Deep Learning-Based Classification of Earthquake-Impacted Buildings Using Textual Damage Descriptions.” *International Journal of Disaster Risk Reduction* 36 (March). 10.1016/j.ijdr.2019.101111.
- [30] S. Mangalathu, J.-S. Jeon, Regional Seismic Risk Assessment of Infrastructure Systems through Machine Learning: Active Learning Approach, *Journal of Structural Engineering* 146 (12) (2020), [https://doi.org/10.1061/\(asce\)st.1943-541x.0002831](https://doi.org/10.1061/(asce)st.1943-541x.0002831).
- [31] S. Mangalathu, H. Sun, C.C. Nweke, Z. Yi, H.V. Burton, Classifying Earthquake Damage to Buildings Using Machine Learning, *Earthquake Spectra* 36 (1) (2020) 183–208, <https://doi.org/10.1177/8755293019878137>.
- [32] National Planning Commission (NPC). 2015. “Nepal Earthquake 2015 - Post Disaster Needs Assessment. Vol. B: Sector Reports.” Government of Nepal.
- [33] F. Parisi, N. Augenti, Earthquake Damages to Cultural Heritage Constructions and Simplified Assessment of Artworks, *Engineering Failure Analysis* 34 (2013), <https://doi.org/10.1016/j.engfailanal.2013.01.005>.
- [35] Parmar, Akash, Rakesh Kataria, and Vatsal Patel. 2019. “A Review on Random Forest: An Ensemble Classifier.” In *Lecture Notes on Data Engineering and Communications Technologies*, 26:758–63. Springer Science and Business Media Deutschland GmbH. 10.1007/978-3-030-03146-6_86.
- [36] Platt, Stephen, Dipendra Gautam, and Rajesh Rupakhetty. 2020. “Speed and Quality of Recovery after the Gorkha Earthquake 2015 Nepal.” *International Journal of Disaster Risk Reduction* 50 (November). 10.1016/j.ijdr.2020.101689.
- [37] K. Porter, R. Kennedy, R. Bachman, Creating Fragility Functions for Performance-Based Earthquake Engineering, *Earthquake Spectra* 23 (2) (2007) 471–489, <https://doi.org/10.1193/1.2720892>.
- [38] S. Roesslin, Q. Ma, H. Juárez-García, A. Gómez-Bernal, J. Wicker, L. Wotherspoon, A Machine Learning Damage Prediction Model for the 2017 Puebla-Morelos, Mexico, Earthquake, *Earthquake Spectra* 36 (2_suppl) (2020) 314–339, <https://doi.org/10.1177/8755293020936714>.
- [39] X. Romão, A.A. Costa, E. Paupério, H. Rodrigues, R. Vicente, H. Varum, A. Costa, Field Observations and Interpretation of the Structural Performance of Constructions after the 11 May 2011 Lorca Earthquake, *Engineering Failure Analysis* 34 (2013), <https://doi.org/10.1016/j.engfailanal.2013.01.040>.
- [40] R. Rupakhetty, S. Olafsson, B. Halldorsson, The 2015 Mw 7.8 Gorkha Earthquake in Nepal and Its Aftershocks: Analysis of Strong Ground Motion, *Bulletin of Earthquake Engineering*. (2017), <https://doi.org/10.1007/s10518-017-0084-z>.
- [41] R. Rupakhetty, R. Sigbjörnsson, Quantification of Loss and Gain in Performance Using Survey Data: A Study of Earthquake-Induced Damage and Restoration of Residential Buildings, *Natural Hazards* 74 (2014) (3), <https://doi.org/10.1007/s11069-014-1279-0>.
- [42] R. Rupakhetty, R. Sigbjörnsson, S. Ólafsson, Damage to Residential Buildings in Hveragerði during the 2008 Ölfus Earthquake: Simulated and Surveyed Results, *Bulletin of Earthquake Engineering* 14 (7) (2016), <https://doi.org/10.1007/s10518-015-9783-5>.
- [43] Salehi, Hadi, and Rigoberto Burgueño. 2018. “Emerging Artificial Intelligence Methods in Structural Engineering.” *Engineering Structures* 171 (November 2017): 170–89. 10.1016/j.engstruct.2018.05.084.
- [44] Tang, Jiliang, Salem Alelyani, and Huan Liu. 2014. “Feature Selection for Classification: A Review.” In *Data Classification: Algorithms and Applications*. 10.1201/b17320.
- [45] United States Geological Survey. 2017. “M 7.8—36 Km E of Khudi, Nepal.” 2017. <https://earthquake.usgs.gov/earthquakes/eventpage/us20002926/executive>.
- [46] S. Wang, X. Yao, Multiclass Imbalance Problems: Analysis and Potential Solutions, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 42 (4) (2012) 1119–1130, <https://doi.org/10.1109/TSMCB.2012.2187280>.