A Fast Seismic Assessment Technique for Reinforced Concrete Buildings: Machine Learning-based Hassan Index

Fahri Baran Koroglu[a,b,*], Muhammet Fethi Gullu[a], Serdar Ciftci[c], Liam Pledger[d], Claudio Schill[d], Santiago Pujol[d]

[a] *Department of Civil Engineering, Harran University, Sanliurfa, Turkiye*

[b] *Department of Civil Engineering, Middle East Technical University, Ankara, Turkey*

[c] *Department of Computer Engineering, Harran University, Sanliurfa, Turkiye*

[d] *Department of Civil Engineering, University of Canterbury, Christchurch, New Zealand*

Abstract

Assessing large inventories of reinforced concrete structures in urban areas with high seismicity is a daunting task that requires tools that can be applied quickly to produce reliable results. The first goal should be to identify the most vulnerable structures requiring fast intervention. Existing assessment standards are often too complex for this purpose. Shiga et al. [1] and Hassan and Sozen [2] have proposed more efficient assessment options based on simple geometric parameters. The question addressed here is whether machine learning (ML) algorithms trained to use the same parameters can match field observations better. Survey data from 1,320 low- and mid-rise RC buildings are used to train and test an algorithm for seismic vulnerability classification. The algorithm was able to produce modest improvements in vulnerability classification, relative to the Hassan Index and for the dataset used in its training and testing. Yet, for a separate and independent dataset, the algorithm produced nearly the same quality of results as its empirical counterpart, which can be used using simple arithmetic. Nevertheless, the exercise presented suggests that it is worth compiling more data to keep exploring the possibilities ML offers.

Keywords: Reinforced Concrete, Seismic Vulnerability, Machine Learning, Support Point Method, Rapid Assessment

*Corresponding author.
Email address: fbkoroglu@harran.edu.tr (Fahri Baran Koroglu)

1. Introduction

Major earthquakes are still a serious risk to our society, as illustrated by recent earthquakes in Taiwan (2024), Türkiye (2023), Syria (2023), and Morocco (2023). Although seismic design codes and guidance have improved, seismically vulnerable buildings are still the leading cause of the economic destruction and loss of lives resulting from earthquakes, and quick assessment of the building stock is challenging. Large cities contain thousands of buildings, and current assessment methods to estimate building seismic vulnerability often involve convoluted analyses that can often require weeks of work per building.

The number of buildings and the effort associated with seismic assessment today make the cost of identifying the most critical buildings in a large urban area prohibitive. A less demanding alternative is to use simple building properties, easily derived from building plans or similar, to classify buildings in terms of their estimated need for strengthening. Such a pragmatic assessment method, if reliable, would drastically reduce the number of buildings that need to be assessed in detail.

A dataset of 1320 buildings surveyed after earthquake events around the world was used to train and test a machine learning (ML) model to differentiate between structures with and without severe damage. Damage reports and simple geometric quantities such as the number of stories, floor area, and the area of columns and walls were used to train the ML model. The model was then tested against a subset of the surveyed buildings and results were compared against results obtained using the Hassan Index [2], a more conventional assessment method used to classify vulnerable RC structures.

2. Background - RC structure vulnerability and damage classification

Hassan and Sozen [2] proposed one of the simplest methods to classify building structures according to their seismic vulnerability. Its simplicity does not detract from its reliability [3]. The data used to calibrate and test the method are open to the public and can be accessed at `https://datacenterhub.org`. The method proposed by Hassan and Sozen [2] is represented by Eq. (1). A threshold defines a boundary between the more vulnerable and less vulnerable buildings, for prioritization purposes.

$$\text{Building Status} = \begin{cases} \text{More vulnerable} & \text{if } PI < 0.25\% \\ \text{Less vulnerable} & \text{if } PI \geq 0.25\% \end{cases} \tag{1}$$

Here, the priority index, PI, is equal to $CI + WI$. The column index (CI) is equal to half of the total cross-sectional area of reinforced concrete (RC) columns at the base divided by the total floor area of the building above the ground, as shown in Eq. (2). Similarly, WI represents the wall index and is given in Eq. (3).

$$CI = \frac{A_{ct}}{2 \cdot \sum A_{ft}} \cdot 100 \tag{2}$$

$$WI = \frac{A_{wt}}{\sum A_{ft}} \cdot 100 \tag{3}$$

$\sum A_{ft}$ is the total floor area of the building above the ground, and A_{wt} is the sum of the cross-sectional areas of: (a) reinforced concrete walls, and (b) masonry infill walls, in one floorplan direction and at the base of the building (unless a higher level is deemed more critical). Hassan and Sozen [2] recommended dividing the term related to infill by 10.

The threshold on the right-hand side of Eq. (1) can be tuned depending on local perceptions of risk and availability of resources. Buildings above the threshold are classified as less vulnerable, and buildings below the threshold are classified as more vulnerable. The qualifiers "less" and "more" are needed because large swathes of towns often contain large numbers of buildings without the stiffness and the detailing needed to survive strong ground motion. An assessment system that requires fixing everything that does not meet current design norms often leads to inaction. The idea here is to reduce the problem to a manageable size by focusing first on the worst cases. Figure 1 illustrates the results of reconnaissance work after the Erzincan 1992 earthquake, plotted in the manner developed by Hassan and Sozen [2]. Each scatter point represents a building surveyed after the earthquake.

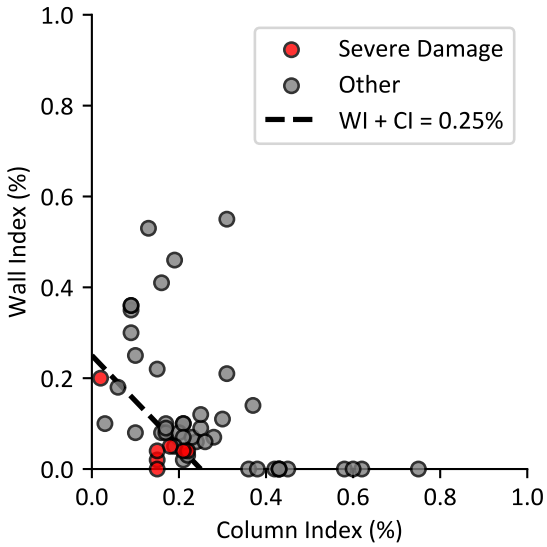

Figure 1: Relationship between column index (%), wall index (%), and reported damage to buildings after the 1992 Erzincan earthquake.

The exact shape or slope of the threshold in Eq. (1) and illustrated in Figure 1 does not appear to have a clear and direct impact on the balance between cost and benefits associated with using the classification method[4]. A more stringent threshold can help identify more critically vulnerable buildings, but it can also misclassify more buildings that may not require urgent intervention. At the same time, avoiding unnecessary retrofits by using a less stringent threshold results in a greater number of vulnerable buildings not classified as such.

The ideal classification method should minimize the number of misclassifications. The probability of misclassification is assessed using existing survey data. To organize the available data, two types of misclassifications are considered:

- **False Negative**: A building observed to have **severe damage** after an earthquake but classified as **less vulnerable**.
- **False Positive**: A building observed to **not** have **severe damage** after an earthquake but classified as **more vulnerable**.

Here, "risk" refers to the number of false negatives, while "cost" refers to the number of false positives. The terms "risk" and "cost" are expressed as

percentages of the total number of buildings, and values near zero indicate that the seismic assessment technique is effective. Substantial risk can lead to loss of lives, and excessive cost can lead to waste, unaffordability, and inaction.

Figure 2 illustrates the terms "cost" and "risk" as red quadrants in a 2x2 graphical "confusion" matrix. The quadrants along the opposite diagonal are related to true positives and negatives, i.e., the match between the assigned vulnerability level and damage observed in the field. The percent of buildings in the quadrants along this diagonal representing matching classifications and observations can be considered a measure of the reliability of the assessment method.

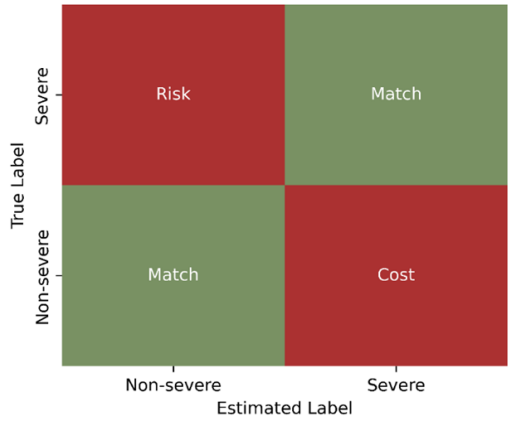

Figure 2: Definition of quadrants in confusion matrix regarding the seismic assessment classification.

The Hassan Index has been tested against field observations repeatedly [2–10]. Consistently, these tests have shown a correlation between the frequency of severe damage (presence of structural failures) and PI, with buildings represented by values of PI less than 0.1% classified as having "severe damage" in nearly 2 out of every 3 studied cases. Care must also be taken when assessing structures that did not sustain severe damage. The words of Howe [11] should be kept in mind: "Merely because a building has escaped injury in an earthquake is no guarantee that it is earthquake proof. One can be sure only if it is known to have the necessary strength and stiffness. It may only have been lucky. Not every soldier who goes into battle is hit. Because

[they] escaped does not prove [they were] bulletproof." There are numerous examples from previous reconnaissance efforts where seemingly poorly constructed buildings remained undamaged or where identical buildings such as school or apartment blocks in close proximity experienced large differences in structural damage from collapse to light damage [4–6, 10, 12].

Nevertheless, the point can still be made that too many buildings with low values of PI have been reported not to have experienced severe damage in zones inferred to have experienced intense ground motion. Reducing the number of false negatives ("cost"), may help redirect scarce resources required for strengthening.

The goal of this study is to use machine learning to try and organize the available field data better, reducing the number of misclassifications. In order to achieve the proposed aim, this is a leading and novel paper in the sense of estimating the seismic vulnerability of reinforced concrete buildings by combining an existing assessment method and deep neural networks, evaluating the reliability of the developed ML-based method from the perspective of risk and cost terms, utilizing a more appropriate splitting technique for the dataset. On top of these, the performance of the developed ML-based model is proven in the field by the data from the Taiwan earthquake [13].

3. Surveyed data description

Survey data from various regions of the world, ranging from Ecuador [9] to Taiwan [3], was compiled to calibrate and test a machine learning model. The surveyed data include eight input parameters, or features, that were used to train, validate, and test the ML model:

1. The number of floors of the building,
2. The typical floor area of the building,
3. The total floor area of the building above the ground,
4. The total cross-sectional area of RC columns in the first floor,
5. The total cross-sectional area of RC walls in the NS direction of the floor plan,
6. The total cross-sectional area of RC walls in the EW direction of the floor plan,
7. The total cross-sectional area of masonry infill in the NS direction of the floor plan,
8. The total cross-sectional area of masonry infill in the EW direction of the floor plan.

These features can be easily derived from building floor plans (formal or sketched in the field). Figure 3 shows an example of a form used in the field to collect information about a particular building.

The compiled dataset classifies each surveyed building into one of five damage classes according to what was observed in the field: "None", "Light", "Moderate", "Severe", and "Collapse", as defined in [6]. Nevertheless, for simplicity, the damage classes were grouped into two classes: "Non-severe" and "Severe" damage. The non-severe damage class includes surveyed buildings that experienced no damage, light, or moderate damage. The severe damage class contains cases of severely damaged buildings and collapse. The damage class of each surveyed building is used to train, validate, and test a ML model to classify buildings as more or less vulnerable. Section 4 describes the ML model and the training process in detail.

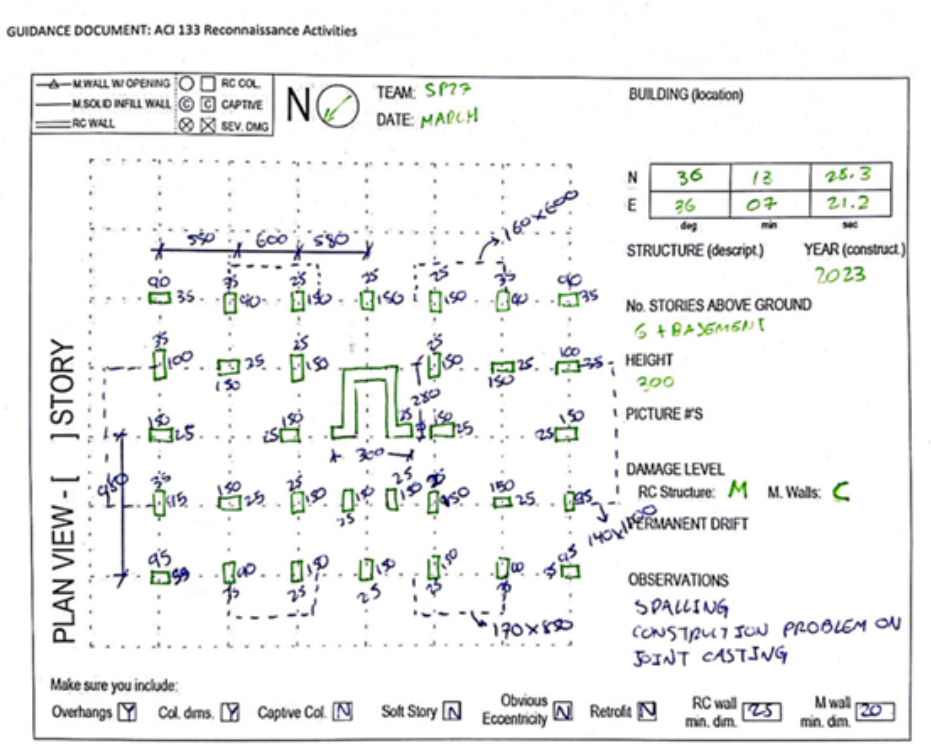

Figure 3: A filled-in typical floor plan form used in reconnaissance visits to gather building data.

The dataset contains 1320 surveyed buildings, of which 524 were classified

as severely damaged in the field. A more balanced dataset would require surveying more buildings with severe damage in future reconnaissance visits, but it is understood that severely damaged buildings can often be a) dangerous to survey and b) abandoned and inaccessible.

The dataset is divided into three subsets for training, validation, and testing purposes. A method called SPlit developed by [14], based on the support points method [15], was used to partition the dataset, preserving the statistical properties of the main dataset within the corresponding subsets. Table 1 lists the statistical properties of parameters used in the training, validation, and test datasets. The cross-sectional area of the lateral load system and the masonry infill is normalized with respect to the total floor area to facilitate the interpretation of the numbers in the table. The parameters (or features); "Minimum Wall Index [%]", "PI [%]", and "PI/Number of Floors" were derived from ad hoc feature engineering and resulted in improved performance of the DNN model.

Most surveyed buildings had fewer than 6 floors. The typical floor area of the surveyed buildings was smaller than 500 square meters, and in most cases, the floor area was uniform up to the height of the building. Over 90% of the surveyed structures had wall area ratios smaller than 0.2% in either direction. Findings from previous surveys highlight that buildings in Japan or Chile –which tend to have low seismic vulnerability [16, 17]– have, on average, a wall area ratio greater than 0.35% in each direction [1, 6, 18]. For the full dataset of surveyed structures, the average masonry infill area ratio in either direction is approximately 0.04%, with 80 % of the structures having less than 0.1% masonry infill.

4. Machine-Learning model development

4.1. Model Architecture

This study employs fully connected neural networks (FCNNs) to estimate a structure's earthquake vulnerability. This section provides a brief overview of FCNNs, followed by the model architecture and training hyperparameters used in this study. An FCNN is made up of an input layer, one or more hidden layers, and an output layer. Each layer, except for the input layer, consists of artificial neurons, each of which is fully connected to all neurons in the previous layer, as shown in Figure 4b. An artificial neuron computes a weighted sum of its inputs and applies an activation function to introduce

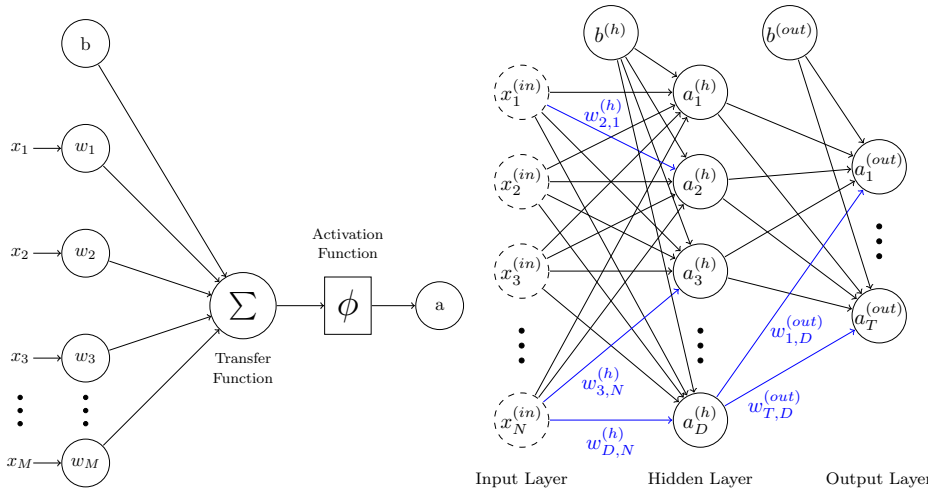

Figure 4: a) Schematic of an artificial neuron with M inputs b) Schematic of single hidden layer FCNN, with N inputs, D hidden units and T outputs.

nonlinearity. The output of an artificial neuron with M inputs is given by:

$$a = \phi\left(\sum_{i=1}^{M} w_i x_i + b\right) \tag{4}$$

where w_i is the weight corresponding to the i-th input x_i, b is the bias term, and ϕ is the activation function. A schematic of an artificial neuron is shown in Figure 4a. Forward propagation is the process of computing the FCNN's output layer by layer, where each layer's output serves as the input for the next. The output of the h-th layer in the FCNN is simply the activation of each of its D artificial neurons, $\mathbf{a}^{(h)} = [a_0^{(h)}, a_1^{(h)}, \ldots, a_i^{(h)}, \ldots, a_D^{(h)}]$.

Training an FCNN involves optimizing the weights and bias of each neuron. This is achieved iteratively by computing the model output, comparing it with the desired output, and calculating the error using the loss function. The loss function quantifies the difference between the model's predictions and ground truth. For regression tasks, mean squared error (MSE) is commonly used, while cross-entropy loss is standard for classification. Once the error is calculated, the model parameters are updated using the backpropagation algorithm. For further details on the training of neural networks, refer to Chapter 6 in [19].

In this study, the FCNNs uses 11 inputs (see Table 1) and four hidden layers

containing 16, 32, 64, and 128 neurons, respectively. Leaky ReLU [20] is used as the activation function for all neurons, except for the single neuron in the output layer, which uses a linear activation function. Structures are then classified as more or less vulnerable using a threshold of 0.5. Additionally, batch normalization [21] was applied to stabilize training, and dropout [22] (drop rate = 0.25) was used to reduce overfitting. A schematic diagram of the model architecture is shown in Figure 5.

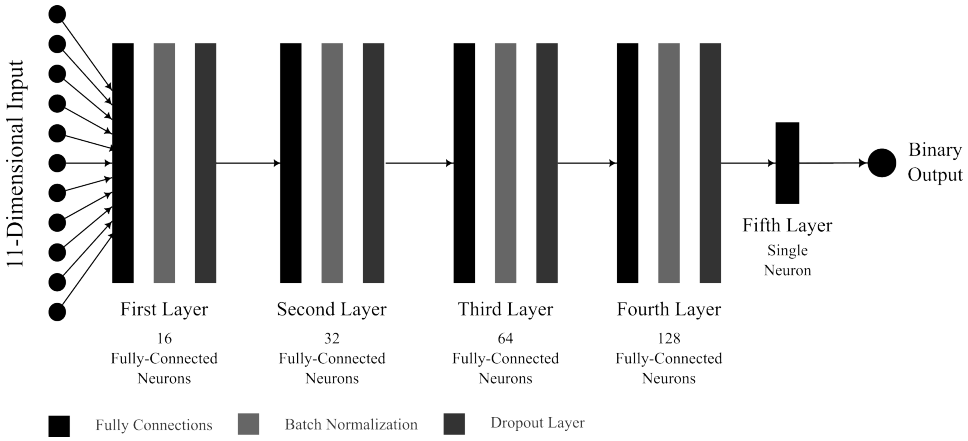

Figure 5: Implemented DNN architecture for RC building vulnerability estimation.

4.2. Dataset Split, Model Training, and Performance Evaluation

The full dataset is split into three different subsets called training, validation, and test datasets. The training subset is used to calibrate the model weights and parameters while the validation subset is used to evaluate the performance of the model during training. Hyperparameters such as the learning rate, dropout rate, etc., were fine-tuned based on the validation subset as fine-tuning based on the training performance can lead to overfitting. In order to prevent overfitting of the DNN models, three practices have been adopted in this work:

1. Learning rate scheduler,
2. Fine-tuning of hyperparameters and performance evaluation of the model based on the performance of the validation dataset,
3. Regularization by using dropout technique.

Once fine-tuning is complete on the validation subset, the test set, which is neither used for training nor validation purposes, is used to evaluate the final performance of the DNN without an obvious bias.

The full dataset is split with ratios of 72.6%, 12.8%, and 14.6% for training, validation, and test sets, respectively. The methodology followed for the dataset splitting and model training/evaluation for a single DNN model is illustrated in Figure 6.

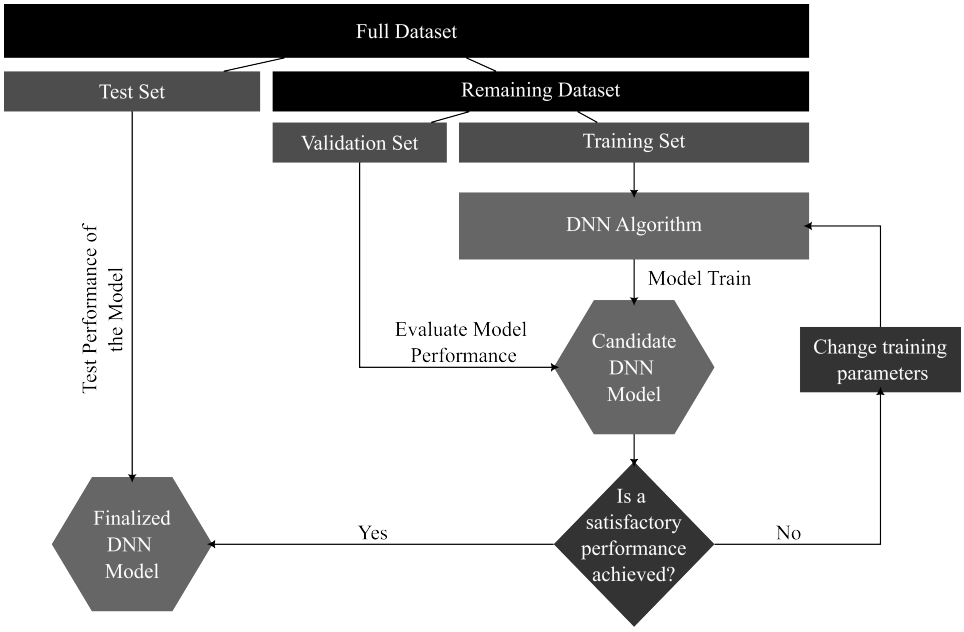

Figure 6: Flowchart of the dataset splitting and DNN model selection.

One of the critical parameters during the training of DNN models is the learning rate. A high learning rate can lead to overshooting and failure to converge to a solution. Alternatively, a low learning rate increases the computation cost, and can lead to an underfitted model that has converged a local minima. A learning rate drop scheduler is used to gradually reduce or "drop" the learning rate of the model during training and avoid overfitting of the model. DNN models can quickly overfit if the training set is small. Ad-hoc fine-tuning of the algorithm was used in this study. During the model training, the training and validation error histories were monitored to ensure that the model was not overfitting.

4.3. Handling of Imbalanced Dataset

The full dataset is slightly biased with 524/1320 (40%) of the surveyed buildings classified as having severe damage and this bias can impact the

algorithm. For example, the DNN model may be more likely to classify the buildings as less vulnerable due to having a higher amount of non-severely damaged samples. Nevertheless, it is possible to prevent this by introducing a penalty factor, which has been calculated as the ratio of the total amount of non-severely damaged buildings to the severely damaged buildings. The penalty factor is used to increase the weight of the "error" during training related to the misclassification of underrepresented class (in this case, severely damaged buildings).

4.4. DNN Senate Concept

When developing a neural network, each connection between neurons has an initial weight and bias. Different initial weights can lead to a different DNN model after training. For example, one model, with certain initial weights, may classify a building as "more vulnerable", while another model, with different initial weights, may classify the same building as "less vulnerable". A good DNN model, with well-tuned training parameters, tends to be less receptive to changes to the initial weights. Nevertheless, cherry-picking the best-performing model carries a risk of producing formulations that perform poorly for unseen data, especially when the model is trained and validated with small datasets.

Rather than using a single DNN model with set initial weights, this study applies the idea of using a group or "senate" of DNN models, each with different initial weights. The average output of all of the DNN models is used to classify the seismic vulnerability of each structure. This method helps to reduce any aleatoric uncertainty associated with the model. 100 DNN models with the architecture illustrated in Figure 5 were trained, each with different initial weights. The five models with the highest accuracy and lowest "risk" -as described in Section 2- were used to classify structures as more or less vulnerable. All of the results in the following sections related to the DNN model refer to an average output of the 5 models.

4.5. DNN-based Hassan Index Method

As mentioned in Section 2, more than two thirds of surveyed buildings with a PI of less than 0.1 % were classified in the field as having "severe damage". The DNN-based Hassan Index approach proposed in this study only uses the DNN results for structures with a $PI \geq 0.1$ % and classifies all structures with a PI of less than 0.1 % as vulnerable.

4.6. Limitations and Reproducibility

The ML model developed in this study was trained on a dataset of building parameters with certain statistical ranges described in Table 1. Deep neural networks have no extrapolation capacity, and results from the DNN model are not applicable for buildings that fall outside of the ranges of the defined statistical properties.

Different CPUs and GPUs or package versions can converge to different model parameters for a DNN due to differences in numerical precision. Therefore, it is not guaranteed that the same results can be reproducible for a DNN model in different local machines or online environments because of the possibility of being assigned to different CPUs in different logins. A Google Colab link has been created, including the trained DNN model to provide users with reproducible results regardless of the capabilities of local machines.

5. Results

The trained DNN model was used to classify the vulnerability of structures in the validation and test datasets. The DNN-based model results, in terms of overall accuracy and the percentage of true positives, true negatives, false positives ("cost"), and false negatives ("risk") are shown in Table 2. The results of the DNN-based model are compared with results obtained from the Hassan Index classification method described in Section 2 along with two variants of the Hassan Index method proposed by Ozcebe et. al [23] & Shah et al. [7]. The results of the various methods are shown in Table 2.

The DNN-based model outperforms the Hassan Index with better overall accuracy (74% vs 61% for the test subset), and also leads to a reduced "cost" and a reduced "risk". The similarity between the validation and test results of the DNN-based model demonstrates that the model has learned underlying trends and patterns within the dataset and provides reassurance that the model is not overfitting.

The formulations proposed by Ozcebe et al. [23] & Shah et al. [7] have the lowest accuracy, close to 50%, however, they also lead to the lowest "risk" or the number of false negatives at 2-3 %. While both of these formulations lead to a low "risk", they do so by being highly conservative with a "cost" of approximately 45 %, more than double that of the DNN-based model. This high "cost" would require many buildings to be retrofitted and classified as more vulnerable even though many of these buildings were not severely

damaged after an earthquake. A high "cost" makes the classification method unfeasible to implement in practice and difficult to decide which building to prioritize.

The DNN-based model leads to the lowest value of "cost" while also keeping the "risk" low, at 5%.

5.1. 2024 Taiwan Earthquake

The Taiwan Earthquake on April 3rd, 2024 provided a perfect opportunity to test the DNN-based model against an entirely new dataset not included in the initial training, testing, or development of the model. The data collected by Shegay et al. [13] was used to test the DNN-based model. The full dataset consisted of 66 RC buildings, however, 3 of the samples were eliminated since they fell outside the statistical bounds of the model training set. Full details about the dataset are described in Shegay et al. (2025) [13]. Table 3 shows the DNN-based model results for the 2024 Taiwan Earthquake dataset along with the Hassan Index and the indices proposed by Ozcebe et al. [23] & Shah et al. [7].

For the Taiwan dataset, the DNN-based model has an accuracy of 73%, a "cost" of 24% and 3% "risk". The similarity between the DNN-based model results for the Taiwan dataset, and the test and validation sets emphasize that the model is not overfitting and capable of producing reliable results for new data.

Nevertheless, the Hassan Index leads to the same level of accuracy from the Taiwan dataset (73%) and has a "risk" of 0%. That is, the Hassan Index threshold correctly identifies all of the severely damaged structures surveyed in Taiwan. The "cost" is 27%, slightly higher than the DNN model. The results suggest that the Hassan Index is as effective as the DNN model in the case of surveyed buildings in Taiwan.

The Hassan Index variations proposed by Ozcebe et al. [23] & Shah et al. [7] have slightly lower accuracy and a larger number of false positives ("cost") than the DNN-based model. Both formulations lead to 0% risk, the same as the Hassan Index.

6. Discussion and limitations

Overall -and on the basis of the available data- the DNN-based model tends to produce more reliable results. Nevertheless, the proposed DNN-based model should be used to produce relative vulnerability measures. That is,

the method would be best used to decide which building to strengthen first instead of trying to classify some buildings as "earthquake-proof," which is, of course, futile. In fact, it is prudent to admit that in the field data, there are abundant instances of buildings that would be deemed vulnerable to earthquakes through any objective measure that, for unclear reasons, have survived earthquakes without significant damage. In that sense, the complete elimination of "cost" through DNN may cause unreasonable increases in "risk." Another potential shortcoming of the use of DNN is that it requires the use of software, while existing methods require only arithmetic. To address this problem, the ready-to-use source code of the deep neural network described and trained as explained in this article can be openly accessed via this Google Colab link.

6.1. Why 100% accuracy is unattainable

Cases demonstrating wide variations in damage among comparable buildings close to one another abound Pujol et al. [4] as mentioned in Section 2. The differences can be attributed to slight variations in strength or deformability among nominally similar but brittle buildings subjected to relatively similar demands 'hovering' near their thresholds. It is also plausible that some buildings benefit from foundation uplift, sliding, and rocking while others do not. Classification methods such as Hassan Index or the DNN-based model would not differentiate between two geometrically identical structures, however, dozens of identical structures in the field and within the datasets have exhibited varying quantities of damage [4, 5, 10].

Radical differences in the performance of comparable buildings have also been attributed to plausible sharp variations in ground motion occurring within relatively short distances [12]. This study does not explore the effects of varying ground motion intensity on the likelihood of damage, and only considers geometrical properties. Some of the surveyed cities had few or no accelerometers nearby to record the ground shaking, for example: Hassan and Sozen [2]. Additionally, the intensity of shaking varied widely among the different surveys, with a recorded peak ground velocity (PGV) of nearly 200 cm/s in Türkiye (2023), and a PGV of approximately 40 cm/s in the areas surveyed after the 2016 Taiwan earthquake.

As mentioned in Section 2, these classification methods employ simple geometrical properties of the structure and do not consider aspects such as, the age of the building, construction quality, detailing of reinforcement, the location of lap splices, concrete cover, or the layout of the lateral load resisting system. Additionally, this study focuses on surveyed buildings that have

sustained severe damage and does not distinguish between other structures that sustained moderate, light, or no damage.

Because of the listed reasons, the work presented here was done understanding that complete elimination of false negatives, i.e., reducing "cost" to zero, may neither be possible nor desirable without detailed information about each building and the intensity of the motion imparted to the foundations of each building. From that point of view, it should be clear that reducing "risk" (i.e., reducing "false negatives") is much more crucial than reducing cost.

6.2. The black-box nature of machine learning and neural networks

Lastly, machine learning can be perceived as a threat to the ability of the engineer to understand the mechanics of the problem at hand. At least two responses are in order: 1) mechanics remains the best means to identify parameters that can be used within machine learning algorithms, and 2) the engineer shall be required to investigate how results obtained with an ML model change with assumed variations in input parameters and judge whether the inferred variations can be justified from first principles. In short, machine learning shall not prevent thinking. In that sense, machine learning represents a new type of 'calculator,' not a new type of engineering.

7. Conclusions

This paper introduces a promising and fast seismic assessment technique, the DNN-based Hassan Index Method, which combines deep neural networks (DNNs) with the assessment method proposed by Hassan and Sozen [2] for reinforced concrete buildings. The performance of the proposed technique is compared with the performance of the original formulation by Hassan and Sozen [2] based on three parameters:

- **Accuracy:** The percentage of cases classified in agreement with observation as less or more prone to severe damage.
- **Risk:** The percentage of false negatives, buildings that are considered "safe" but surveys show that they experienced severe damage.
- **Cost:** The percentage of false positives, buildings that are considered "more vulnerable" but surveys show that they did not experience severe damage.

Compared with the use of the criterion proposed by Hassan and Sozen [2], which states that $PI \leq 0.25\%$ indicates higher vulnerability (a threshold that can and should be adjusted depending on local conditions and risk tolerance), the use of the DNN-based method helped:

1. Increase **accuracy** by a factor of 1.25 for the test set (from approximately 61% to 75%),
2. Reduce **cost** by a factor of 1.50 for the test set (from approximately 31% to 20%), and
3. Decrease **risk** by factor of 1.50 (from 8% to 5%).

The caveats presented in Section 2 and Section 6.1 about apparently vulnerable buildings that have been observed to survive earthquakes should be kept in mind in relation to **3**. It could be argued that the decrease in **"risk"** implies a potential reduction in loss of life compared with what can be obtained from the simple arithmetic used to compute the index proposed by Hassan and Sozen [2]. At the same time, the reduction in **"cost"** may represent an opportunity to use resources more effectively. Nevertheless, given all the variables left out of the studied formulations, it is prudent to use the proposed technique to prioritize building strengthening efforts or to schedule structures for a more detailed study. As a result, a lower bound of $PI < 0.1\%$ is proposed to identify vulnerable structures, irrespective of the output from the DNN model.

The results show that the formulated DNN-based model has potential as an initial seismic assessment tool that can help prioritize mitigation efforts. Nevertheless, despite large increases in computational effort, only modest increases in performance were obtained from DNN in comparison with simpler prioritization tools. That result should not surprise given

1. the complexity of the earthquake problem,
2. the simplicity of the information available,
3. the limited number of cases with which we operate.

It is possible, however, that concerted efforts to collect more data will lead to improvements in the studied classification method. DNN-based algorithms work best if trained with large amounts of data. DNN algorithms trained using small datasets are prone to "overfitting". In this study, overfitting was kept in check by using the "dropout" technique and the "learning-rate scheduler", Nevertheless, the results presented suggest that use of DNN in seismic assessment still needs to be guided by experience and engineering judgment.

The data and codes used in this study have been made public to benefit governmental institutions and industry. The proposed technique's applicability should be restricted to conditions similar to those in which the data used for calibration were collected.

Acknowledgement

The authors greatly acknowledge the efforts of those who surveyed and gathered the data after previous earthquakes.

References

[1] T. Shiga, Earthquake damage and wall index of reinforced concrete buildings, Architectural Institute of Japan (1968) 29–32.

[2] A. F. Hassan, M. A. Sozen, Seismic vulnerability assessment of low-rise buildings in regions with infrequent earthquakes, ACI Structural Journal 94 (1997) 31–39.

[3] A. Y. Puranam, A. Irfanoglu, S. Pujol, T. C. Chiou, S. J. Hwang, Evaluation of seismic vulnerability screening indices using data from the taiwan earthquake of 6 february 2016, Bulletin of Earthquake Engineering 17 (2019) 1963–1981.

[4] S. Pujol, L. Laughery, A. Puranam, P. Hesam, L. H. Cheng, A. Lund, A. Irfanoglu, Evaluation of seismic vulnerability indices for low-rise reinforced concrete buildings including data from the 6 february 2016 taiwan earthquake, Journal of Disaster Research 15 (2020) 9–19.

[5] C. Dönmez, S. Pujol, Spatial distribution of damage caused by the 1999 earthquakes in turkey, Earthquake Spectra 21 (2005) 53–69.

[6] S. Pujol, I. Bedirhanoglu, C. Donmez, J. D. Dowgala, M. Eryilmaz-Yildirim, K. Klaboe, F. B. Koroglu, R. D. Lequesne, B. Ozturk, L. Pledger, E. Sonmez, Quantitative evaluation of the damage to rc buildings caused by the 2023 southeast turkey earthquake sequence, Earthquake Spectra 40 (2024) 505–530.

[7] P. Shah, S. Pujol, M. Kreger, A. Irfanoglu, 2015 nepal earthquake, Concrete International 39 (2017) 42–49.

[8] C. Sim, C. Song, N. R. Skok, A. Irfanoglu, S. Pujol, M. Sozen, Database of low-rise reinforced concrete buildings with earthquake damage, `https://datacenterhub.org/resources/` (2016).

[9] E. Villalobos, C. Sim, J. P. Smith-Pardo, P. Rojas, S. Pujol, M. E. Kreger, The 16 april 2016 ecuador earthquake damage assessment survey, Earthquake Spectra (2018) 1201–1217.

[10] W. Zhou, W. Zheng, S. Pujol, Seismic vulnerability of reinforced concrete structures affected by the 2008 wenchuan earthquake, Bulletin of Earthquake Engineering 11 (2013) 2079–2104.

[11] G. E. Howe, Requirements for buildings to resist earthquakes, American Institute of Steel Construction (1936).

[12] S. Alcocer, A. Behrouzi, S. Brena, K. J Elwood, A. Irfanoglu, M. Kreger, R. Lequesne, G. Mosqueda, S. Pujol, A. Puranam, et al., Observations about the seismic response of rc buildings in mexico city, Earthquake Spectra 36 (2020) 154–174.

[13] A. Shegay, T. Suzuki, T.-C. Chiou, L. Hogan, J. Monical, H. Aydogdu, Z. Yi, J. Byrne, K. Elwood, S.-J. Hwang, K. Jiang, S.-J. Jhuang, H.-J. Lee, J.-K. Lee, C.-C. Lin, M.-L. Lin, Y.-C. Ling, H. Liu, M. Maeda, M. Otani, W.-C. Shen, D. Shearer, H. Shoda, L.-W. Song, M. Stephens, P.-W. Weng, T.-P. Yuen, C.-H. Yang, C.-H. Yang, J.-H. Zhang, Y.-J. Zeng, 2024 hualien earthquake reconnaissance of reinforced concrete buildings, Tech. Rep. NCREE-25-XXX, National Center for Research on Earthquake Engineering, Taiwan (2025, Preprint).

[14] V. R. Joseph, A. Vakayil, Split: An optimal method for data splitting, Technometrics 64 (2022) 166–176.

[15] S. Mak, V. R. Joseph, Support points, The Annals of Statistics 46 (2018) 2562–2592.

[16] R. Riddell, S. L. Wood, J. C. de la Llera, The 1985 chile earthquake: Structural characteristics and damage statistics for the building inventory in viña del mar, Civil Engineering Studies SRS-534 (1987).

[17] M. Sarrafzadeh, K. J. Elwood, R. P. Dhakal, H. Ferner, D. Pettinga, M. Stannard, M. Maeda, Y. Nakano, T. Mukai, T. Koike, Performance of reinforced concrete buildings in the 2016 kumamoto earthquakes and seismic design in japan, Bulletin of the New Zealand Society for Earthquake Engineering 50 (2017) 394–435.

[18] L. Pledger, S. Pujol, R. Chandramohan, Investigating the effect of stiffness on the seismic performance of rc structures, in: NZSEE 2023 Annual Conference, New Zealand Society for Earthquake Engineering, Auckland, New Zealand, 2023.

[19] S. Raschka, Y. Liu, V. Mirjalili, Machine Learning with PyTorch and Scikit-Learn, Packt Publishing, Birmingham, UK, 2022.

[20] A. L. Maas, A. Y. Hannun, A. Y. Ng, et al., Rectifier nonlinearities improve neural network acoustic models, in: Proc. icml, Vol. 30, Atlanta, GA, 2013, p. 3.

[21] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: Proceedings of the 32nd International Conference on Machine Learning, 2015.

[22] N. Srivastava, G. Hinton, A. Krizhevsky, R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, Journal of Machine Learning Research 15 (2014) 1929–1958.

[23] G. Özcebe, J. Ramirez, S. T. Wasti, 1 may 2003 bingol earthquake - engineering report, `http://www.seru.metu.edu.tr/archives.htmlhttp://www.anatolianquake.org` (2003).

Table 1: Statistical Summary of Building Features

Feature	**Dataset**	**Min**	**Max**	$\boldsymbol{\mu}$	$\boldsymbol{\sigma}$
Number of Floors	Training	0.5	16	4.22	2.42
	Validation	1	15	4.24	2.50
	Test	1	16	4.26	2.57
Typical Floor Area [m^2]	Training	22	4880	336	322
	Validation	37	1450	330	270
	Test	35	2070	321	282
Total Floor Area [m^2]	Training	46	24400	1600	2090
	Validation	86	13400	1610	2020
	Test	52	16400	1590	2100
Column Area Ratio [%]	Training	0	2.35	0.39	0.26
	Validation	0	1.61	0.39	0.24
	Test	0	1.70	0.39	0.25
RC Wall Area Ratio (NS) [%]	Training	0	1.66	0.07	0.16
	Validation	0	1.39	0.07	0.18
	Test	0	1.39	0.07	0.16
RC Wall Area Ratio (EW) [%]	Training	0	1.52	0.06	0.14
	Validation	0	1.05	0.06	0.15
	Test	0	1.14	0.06	0.14
Masonry Infill Area Ratio (NS) [%]	Training	0	0.40	0.04	0.06
	Validation	0	0.38	0.04	0.06
	Test	0	0.33	0.04	0.06
Masonry Infill Area Ratio (EW) [%]	Training	0	0.70	0.04	0.06
	Validation	0	0.48	0.04	0.06
	Test	0	0.54	0.04	0.06
Minimum Wall Index [%]	Training	0	1.12	0.06	0.09
	Validation	0	1.05	0.06	0.10
	Test	0	0.71	0.06	0.09
PI [%]	Training	0	1.29	0.25	0.15
	Validation	0	1.05	0.25	0.15
	Test	0	1.15	0.25	0.16
PI/Number of Floors [%]	Training	0	1.78	0.10	0.14
	Validation	0	0.88	0.10	0.12
	Test	0	0.94	0.10	0.13

Table 2: Results for structural damage classification based on Hassan Index and the proposed DNN model (TP = True positive, TN = True negative, FP = False psotive, FN = False negative).

	Dataset	Accuracy	TP	TN	FP "Cost"	FN "Risk"
DNN-based model	Validation	75%	40%	35%	20%	5%
	Test	74%	40%	34%	21%	5%
Hassan Index [2]	Validation	63%	31%	32%	30%	8%
	Test	61%	30%	31%	31%	8%
Ozcebe et al. [23]	Validation	50%	37%	13%	47%	2%
	Test	53%	38%	15%	46%	2%
Shah et al. [7]	Validation	53%	37%	16%	44%	3%
	Test	52%	36%	16%	45%	3%

Table 3: Results for structural damage classification of buildings surveyed after the 2024 Taiwan Earthquake (TP = True positive, TN = True negative, FP = False positive, FN = False negative).

	Accuracy	TP	TN	FP "Cost"	FN "Risk"
DNN-based model	73%	44%	29%	24%	3%
Hassan Index [2]	73%	48%	25%	27%	0%
Ozcebe et al. [23]	67%	48%	19%	33%	0%
Shah et al. [7]	70%	48%	22%	30%	0%