buildings

MDPI

Article

Attention-Enhanced Progressive Transfer Learning for Scalable Seismic Vulnerability Assessment of RC Frame Buildings

Kaushik M. Gondaliya [1], Konstantinos Daniel Tsavdaridis [2,*], Aanal Raval [3], Jignesh A. Amin [1] and Komal Borisagar [4]

[1] Department of Civil (Structural) Engineering, School of Engineering and Technology, Gujarat Technological University, Chandkheda, Ahmedabad 382424, Gujarat, India; ap_kgondaliya@gtu.edu.in (K.M.G.)
[2] Department of Engineering, School of Science & Technology, City St George's, University of London, Northampton Square, London EC1V 0HB, England, UK
[3] Department of Computer Engineering, School of Engineering and Technology, Gujarat Technological University, Ahmedabad 382424, Gujarat, India
[4] Department of Electrical and Communication Engineering, School of Engineering and Technology, Gujarat Technological University, Ahmedabad 382424, Gujarat, India
* Correspondence: konstantinos.tsavdaridis@city.ac.uk

Abstract

Urban infrastructure in seismic zones demands efficient and scalable tools for damage prediction. This study introduces an attention-integrated progressive transfer learning (PTL) framework for the seismic vulnerability assessment (SVA) of reinforced concrete (RC) frame buildings. Traditional simulation-based vulnerability models are computationally expensive and dataset-specific, limiting their adaptability. To address this, we leverage a pretrained artificial neural network (ANN) model based on nonlinear static pushover analysis (NSPA) and Monte Carlo simulations for a 4-story RC frame, and extended its applicability to 2-, 8-, and 12-story configurations via PTL. An attention mechanism is incorporated to prioritize critical features, enhancing interpretability and classification accuracy. The model achieves 95.64% accuracy across five damage categories and an R^2 of 0.98 for regression-based damage index predictions. Comparative evaluation against classical and deep learning models demonstrates superior generalization and computational efficiency. The proposed framework reduced retraining requirements across varying building heights, shows potential adaptability to other structural typologies, and maintains high predictive fidelity, making it a practical AI solution for structural risk evaluation in seismically active regions.

Keywords: reinforced concrete frame; seismic vulnerability assessment; progressive transfer learning; attention mechanism; capacity spectrum-based method; structural damage prediction; nonlinear static pushover analysis

Academic Editor: Hugo Rodrigues

Received: 21 October 2025
Revised: 21 November 2025
Accepted: 1 December 2025
Published: 3 December 2025

Citation: Gondaliya, K.M.; Tsavdaridis, K.D.; Raval, A.; Amin, J.A.; Borisagar, K. Attention-Enhanced Progressive Transfer Learning for Scalable Seismic Vulnerability Assessment of RC Frame Buildings. *Buildings* **2025**, *15*, 4383. https://doi.org/10.3390/buildings15234383

1. Introduction

Earthquakes remain one of the most unpredictable natural disasters, with their intensity and impact on structures heavily influenced by factors such as soil type and fault distance [1,2]. Historically, researchers have endeavored to predict the behavior of structures under seismic loads, beginning with Kendy's pioneering work in 1980 on the lateral behavior of nuclear power plants [3]. Over the decades, studies have expanded to assess the seismic vulnerability and fragility of individual and grouped building structures,

particularly in the wake of catastrophic events like the Bhuj, Tarkey, and Nepal earthquakes. These disasters have underscored the need for accurate prediction models to mitigate the loss of life and infrastructure. In response, funding agencies such as FEMA [4] and ATC [5] have increasingly focused on reducing uncertainties, such as epistemic uncertainty, in structural behavior predictions. Open application programming integration advancements in structural software like CSI (https://www.csiamerica.com, 2 January 2025), Dlubal (https://www.dlubal.com/en, 2 January 2025), and MIDAS (https://www.midasuser.com/en, 2 January 2025), have enabled more precise modeling of material behavior. However, the true leap forward has come with the integration of artificial intelligence (AI) and machine learning (ML) algorithms. These models, particularly ANNs, are now capable of processing vast datasets, identifying patterns, and adapting to new data through techniques, i.e., transfer learning and fine-tuning.

Traditional ML models often struggle with performance where extensive task-specific datasets are required, or data acquisition is costly or limited. Additionally, the ability of traditional models to generalize across varying structural scales and environmental conditions remains constrained, therefore limiting their practical applicability in the real world [6–8]. Leveraging knowledge from pretrained models on similar tasks on new domain-specific tasks, transfer learning (TL) has emerged as a powerful approach to enhance performance, generalization, and adaptability [9,10]. The combination of advanced structural software and ML techniques represents a significant step forward in SVA.

In recent years, a substantial international research effort has focused on applying deep learning, transfer learning, and hybrid AI frameworks to seismic vulnerability assessments and post-earthquake damage recognition. Numerous studies have demonstrated the use of convolutional neural networks, multi-feature fusion, and deep transfer learning to detect structural damage from remote sensing, UAV imagery, and field photographs, significantly improving the speed of post-event reconnaissance in varied seismic regions [11–13]. Parallel developments have explored the use of data-driven models to predict nonlinear seismic responses, soil–structure interaction effects, and the time–history behavior of RC and steel structures, highlighting the increasing capability of ML-based approaches to capture complex structural phenomena traditionally addressed using computationally intensive numerical simulations [12–15]. TL is increasingly recognized for its ability to adapt pretrained models in cases where task-specific labeled datasets are scarce [16–18]. TL exploits the ability to retain learnt generic feature representations which can be then fine-tuned for a specific target domain. By using pretrained models to transfer knowledge, TL accelerates training without requirement of extensive target-domain data. Studies by Abdi et al. (2021) [19] and Dogan et al. (2023) [20], utilizing multi-feature fusion and deep transfer learning, respectively, signify the efficacy of TL in damage detection for tackling complex structural damage patterns and insufficient domain-specific data. Similarly, Xu et al. (2022) [21] and Jena et al. (2023) [22] demonstrated transfer learning's use for time–history predictions in structural systems and earthquake risk assessments. Further by leveraging a pretrained model's domain-invariant features, Lin et al. (2022) [23] illustrated how TL frameworks can improve the task of damage detection.

The integration of attention mechanisms with TL frameworks enriches the model's capability to focus on salient features present in complex data. In structural damage assessments, attention mechanisms allow models to prioritize critical regions. As demonstrated by Jena et al. (2023) [22] and Lin et al. (2022) [23], this approach is propitious while dealing with heterogeneous data sources or identifying subtle damage patterns. Moreover, attention layers effectively capture intricate feature relationships, such as inter-story drift and stress distribution, which are essential for accurate damage predictions in taller structures [6,7,24]. TL frameworks, when integrated with attention mechanisms, improve a model's interpretability and robustness with high-dimensional and temporally

dependent data. In the context of structural damage prediction, several gaps persist despite these advancements in the ML techniques. Existing works primarily focus on adapting pretrained models in a single transfer step, which may lead to challenges such as catastrophic interference (CI). CI is a phenomenon where a fine-tuned model loses its ability to generalize from source domain knowledge to target domain knowledge. Moreover, other traditional methods used for the assessment assume homogeneity between the source and target datasets, which may fail with real-world scenarios where domain shifts are prevalent.

To address these limitations, a promising alternative is offered through PTL, a specialized subset of TL [15,25,26]. PTL provides a novel way where knowledge transfer is performed from simpler tasks or structures to more complex ones. To adapt task-specific features, the model's general representation is fine-tuned incrementally. This gradual adaptation ensures effective generalization across diverse target datasets and mitigates CI. PTL works well in scenarios where data availability is limited or when datasets are highly specialized. This reduces the need for extensively labeled datasets without compromising performance.

The precise and dependable forecasting of structural damage relies on two essential types of uncertainty: epistemic and aleatoric [27–30]. The uncertainties are shaped by factors including the accuracy of structural models, comprehensive structural specifications, and the accessibility of historical earthquake ground-motion data. The challenges of uncertainty quantification are particularly pronounced in areas such as India, where there is substantial variation across various parameters [31–33]. The complexity of this issue is particularly evident in RC structures, where the spatial variability in material properties, such as differences in concrete strength, rebar distribution, and bond conditions, make it challenging to develop accurate and reliable predictions of structural performance under seismic loads. Whereas many prior CNN-based or hybrid deep models (often remote-sensing-driven) primarily address aleatoric variability for post-event classification, the present work advances mechanics-consistent SVA by coupling progressive transfer learning with an attention mechanism in a multi-configuration transfer setting and by tying predictions to fragility functions and standard damage-state thresholds used in HAZUS (https://www.fema.gov/flood-maps/products-tools/hazus, 10 October 2025) workflows.

This study focuses on implementing PTL to extend the capabilities of an existing ANN model that currently predicts structural damage for four-story RC frame buildings designed according to Indian seismic standards [32]. The primary objective is to scale up this model to accurately assess seismic vulnerability across a wider range of building heights, specifically 2-story, 8-story, and 12-story RC building frames, while maintaining the fixed bay width and number of bays across all building configurations. The study employs both classification-based models to categorize damage levels (No-damage, Slight, Moderate, Severe, and Complete) and regression-based models to provide continuous damage index predictions. By leveraging the hierarchical structure of PTL integrated with attention mechanisms, the research aims to establish a methodology where knowledge from the pretrained four-story building model is progressively transferred and adapted to buildings of varying heights. This approach addresses the limitations of conventional TL methods by minimizing CI while efficiently utilizing the limited available data for different building configurations. The study particularly emphasizes the application of these enhanced models in Indian regions, where material heterogeneity and regional seismic disparities present unique challenges for accurate structural damage prediction.

2. Proposed PTL-Based SVA Approach

The conceptual workflow of the proposed PTL-based framework is first outlined, followed by the corresponding implementation details. Figure 1 shows the proposed PTL-based framework, incorporating an attention mechanism to forecast structural damage levels in buildings of different heights (2-story, 4-story, 8-story, and 12-story) that adhere to Indian standards. This section offers an in-depth examination of the data preprocessing pipeline, model architecture, design rationale, training methodology, fine-tuning strategy, and hyperparameter optimization.

Figure 1. Architecture of the proposed progressive transfer learning (PTL) framework with attention mechanism for seismic damage prediction of RC frames.

The PTL framework systematically modifies an ANN model that has been trained on a foundational dataset (such as a four-story RC frame) to forecast damage levels across various building configurations. This organized modification facilitates effective knowledge exchange while preserving predictive precision. The fundamental elements of the base model consist of the following:

- An input layer alongside a bottleneck layer designed for effective feature extraction.
- Incorporating dropout layers and attention mechanisms to enhance generalization and focus on task-specific learning.
- A specialized output layer designed for predicting structural damage.

The model is subjected to a methodical enhancement procedure, which includes the strategic freezing and unfreezing of layers to create tailored models for various building categories. This approach utilizes a structured hierarchy of grandparents, parents, and child elements, enabling the model to effectively generalize seismic damage evaluations while preserving essential knowledge.

The PTL framework facilitates effective transfer learning through layer-wise selective training, keeping the early layers frozen to preserve essential knowledge while allowing the higher layers to gradually adjust to new tasks. The latent space, located following the bottleneck layer, acts as a common feature representation, enabling smooth knowledge transfer among various building types. The process of adaptation is carried out in a systematic manner:

1. Initial layers (frozen) to secure essential representations, guaranteeing the preservation of knowledge.
2. Latent spaces (partially unfrozen) are progressively adjusting to heightened complexities through the selective unlocking of layers.
3. Output layers designed for specific tasks (completely trainable) are customized to predict structural damage in new types of RC frames.

Attention-based inter-task dependencies improve knowledge sharing across tasks, enhancing model adaptability, effectively maintaining relevant structural interactions while accommodating new features. Modifications tailored to specific tasks take place at the output layer, enabling accurate predictions for novel structural configurations. In the process of transitioning between tasks, the freezing–unfreezing mechanism plays a crucial role in facilitating fine-tuning while preventing CI, a prevalent challenge in TL where the acquisition of new information can interfere with previously established knowledge. The attention mechanism enhances latent space representations, allowing for adjustments based on specific task features while preserving previously acquired knowledge. This organized enhancement guarantees that every new assignment solely highlights the most pertinent characteristics while maintaining essential structural patterns. The present study maintains a consistent bay width and number of bays, while altering the story height. To broaden the model's capabilities to include more architectural features, certain higher latent space layers can be strategically unfrozen, facilitating a dynamic adjustment to increasingly intricate structures. The bottleneck layer modifies its compression ratio to adapt to increased variability in structural configurations.

In the context of extensive feature expansions, the implementation of batch normalization serves to reweight domain-specific features, ensuring stability during the fine-tuning process. This avoids redundant updates and guarantees that only relevant layers are trained according to the task's complexity. In cases where the model needs additional scalability, one can implement a hierarchical attention-based freezing strategy that focuses on feature-level adaptations while allowing for dynamic adjustments in the upper layers. The PTL framework in the present study, incorporating attention mechanisms, facilitates

effective TL for predicting seismic damage by preserving critical knowledge, avoiding CI, and flexibly adjusting to new structural characteristics.

3. Data Collection and PTL-Based Model

This section outlines the generation of an SVA dataset for 2-, 4-, 8-, and 12-story RC frame buildings using a probabilistic SVA methodology, as proposed by Gondaliya et al. (2022) [33]. In a subsequent study, Gondaliya et al. (2024) [32] developed an ML model trained on the probabilistic outputs of a four-story RC frame structure designed in accordance with Indian seismic codes. To further enhance the assessment, the capacity spectrum-based method is employed to derive seismic fragility curves, enabling a robust evaluation of the structural vulnerability of the selected RC building configurations.

3.1. Mathematical Modeling of RC Building Frames

The structural configurations and reinforcement detailing for the selected RC buildings, which include 2-, 4-, 8-, and 12-story special moment-resisting frames, are depicted in Figure 2 and summarized in Table 1. The model is based on M25 grade concrete, Fe500 grade reinforcement bars, and a modulus of elasticity of 200 GPa. A Monte Carlo simulation approach is utilized to address uncertainties in material strength properties, employing established probabilistic techniques to assess variability in mechanical properties. Material property limits were established by utilizing a 95% confidence interval in accordance with IS 456 (2000) [34] guidelines, reflecting the intrinsic variability in the characteristics of the concrete and steel sample datasets listed in Appendix A. Parameters, including the modulus of elasticity of concrete, were obtained through established empirical relationships, specifically $E_c = 5000\sqrt{f_c}$, ensuring consistency across simulations. The model incorporated the influence of infill wall mass, distributed uniformly across each floor level. To evaluate seismic fragility, a representative frame from each building type was chosen to establish damage state thresholds and produce corresponding fragility curves. The structural dimensions and load combinations adhered to the Indian design standards IS 875: Part-I and Part-II (1987), IS 1893: Part-I (2016), and IS 13920 (2016) [35–38]. The model incorporated slabs, internal walls, and external masonry walls with thicknesses of 125 mm, 115 mm, and 230 mm, respectively. Superimposed loads of 3 kN/m² for occupied floors and 0.75 kN/m² for roof levels were implemented to represent realistic usage scenarios.

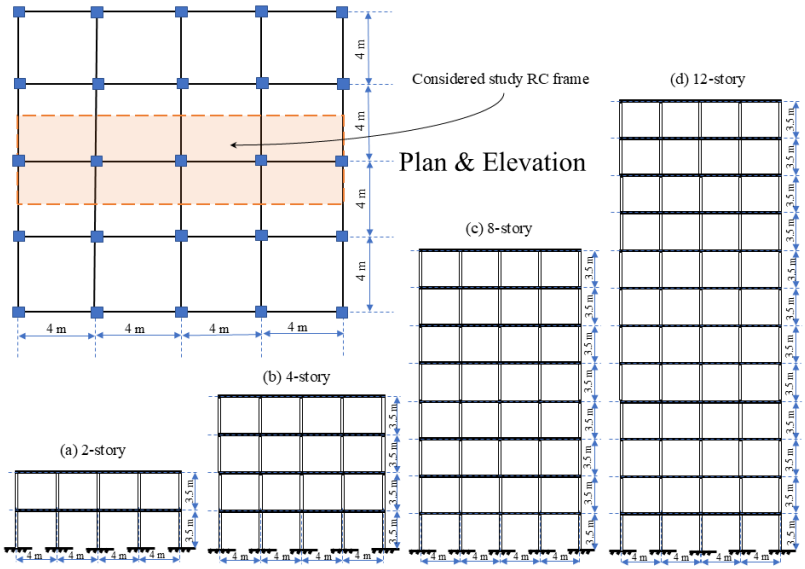

Figure 2. Plan and elevation of the selected RC frame buildings.

Table 1. Rebar detailing of the selected RC frame buildings in the present study.

Frame	Member	Floor	Width (mm)	Depth (mm)	Effective Section *
2-Storey	Beams	1	230	400	6–12 Φ (top) + 2–12 Φ (bottom)
	Columns	1	300	300	8–16 Φ
	Beams	2	230	400	3–16 Φ (top) + 2–12 Φ (bottom)
	Columns	2	300	300	4–16 Φ + 4–12 Φ
4-Storey	Beams	1–2	230	450	4–16 Φ (top) + 2–16 Φ (bottom)
	Columns	1–2	350	350	12–20 Φ
	Beams	3–4	230	400	4–16 Φ (top) + 2–16 Φ (bottom)
	Columns	3–4	300	300	12–16 Φ
8-Storey	Beams	1–4	300	450	2–20 Φ + 1–16 Φ (top) + 2–20 Φ (bottom)
	Columns	1–4	500	500	4–25 Φ + 8–20 Φ
	Beams	5–8	250	450	4–16 Φ (top) + 2–16 Φ (bottom)
	Columns	5–8	350	350	12–25 Φ
12-Storey	Beams	1–4	300	550	2–20 Φ (top) + 2–16 Φ (bottom)
	Columns	1–4	550	550	4–25 Φ + 8–20 Φ
	Beams	5–8	250	500	3–20 Φ (top) + 2–16 Φ (bottom)
	Columns	5–8	450	450	4–25 Φ + 8–20 Φ
	Beams	9–12	250	450	3–16 Φ (top) + 2–16 Φ (bottom)
	Columns	9–12	400	400	8–16 Φ

Note. * Φ represents the diameter of the rebar. Rebar is uniformly distributed in the column.

3.2. Capacity Spectrum-Based Method (CSM)

The CSM, as outlined in the HAZUS-MH MR5 guidelines, is employed to evaluate structural capacity and derive fragility curves that account for both epistemic and aleatoric uncertainties [39,40]. Widely adopted in performance-based seismic assessment, the CSM offers a robust approach to estimate the expected lateral deformation of RC building structures [41–43]. Gondaliya et al. (2023) [44] reviewed several SVA methodologies to derive the seismic fragility curves for the RC frame buildings. In the current study, the NSPA is performed using a gravity load combination of dead load (DL) plus 0.25 times the live load (LL), with lateral forces applied incrementally. The potential formation of plastic hinges is evaluated using the empirical relation [45], using Equation (1):

$$L_p = 0.08\ L + 0.022\ f_y \times d_b \quad (1)$$

where L is the member length, f_y is the yield strength of reinforcement (MPa), and d_b is the diameter of the longitudinal reinforcement (mm). For this study, plastic hinge locations are estimated to form at half the plastic hinge length from the structural member ends, aligning with expected flexural behavior patterns. The pushover curve is transformed into a capacity spectrum using SAP2000 (v25) [46], following ATC-40 and HAZUS guidelines [4,5].

The intersection point of the demand spectrum and capacity curve, termed the performance point, is determined via the equivalent linearization method (Procedure-A), reflecting the structure's expected seismic response under a specific ground motion level. In this study, the spectral displacement, S_d, used in the fragility formulation is obtained from a continuous set of performance points. This continuous treatment is adopted to better capture aleatoric uncertainty in seismic demand and to avoid instability associated with fitting fragilities from a small number of discrete intensities. The fragility curves express the conditional probability of exceeding a particular damage state DS_k at a given spectral displacement (S_d), modeled using a lognormal distribution by Equation (2).

$$p_k\,(S_d) = \mathrm{P}\,[\mathrm{DS} \geq DS_k \mid S_d] = \Phi\,[\frac{1}{\beta_k}\,\ln(\frac{S_d}{S_{d,ds}})] \quad (2)$$

$S_{d,ds}$ represents the mean spectral displacement at which a building reaches a specified damage state threshold. β_k denotes the standard deviation of the natural logarithm of spectral displacement corresponding to the kth damage state and Φ is the cumulative distribution function (CDF) of the standard normal distribution. Figure 3 shows the derived seismic fragility curves of the eight-story RC frame building with probabilistic SVA methodology. The statistical parameters used to generate these fragility curves, including uncertainty (β values) and corresponding damage state deformation limits, are summarized in Table 2.

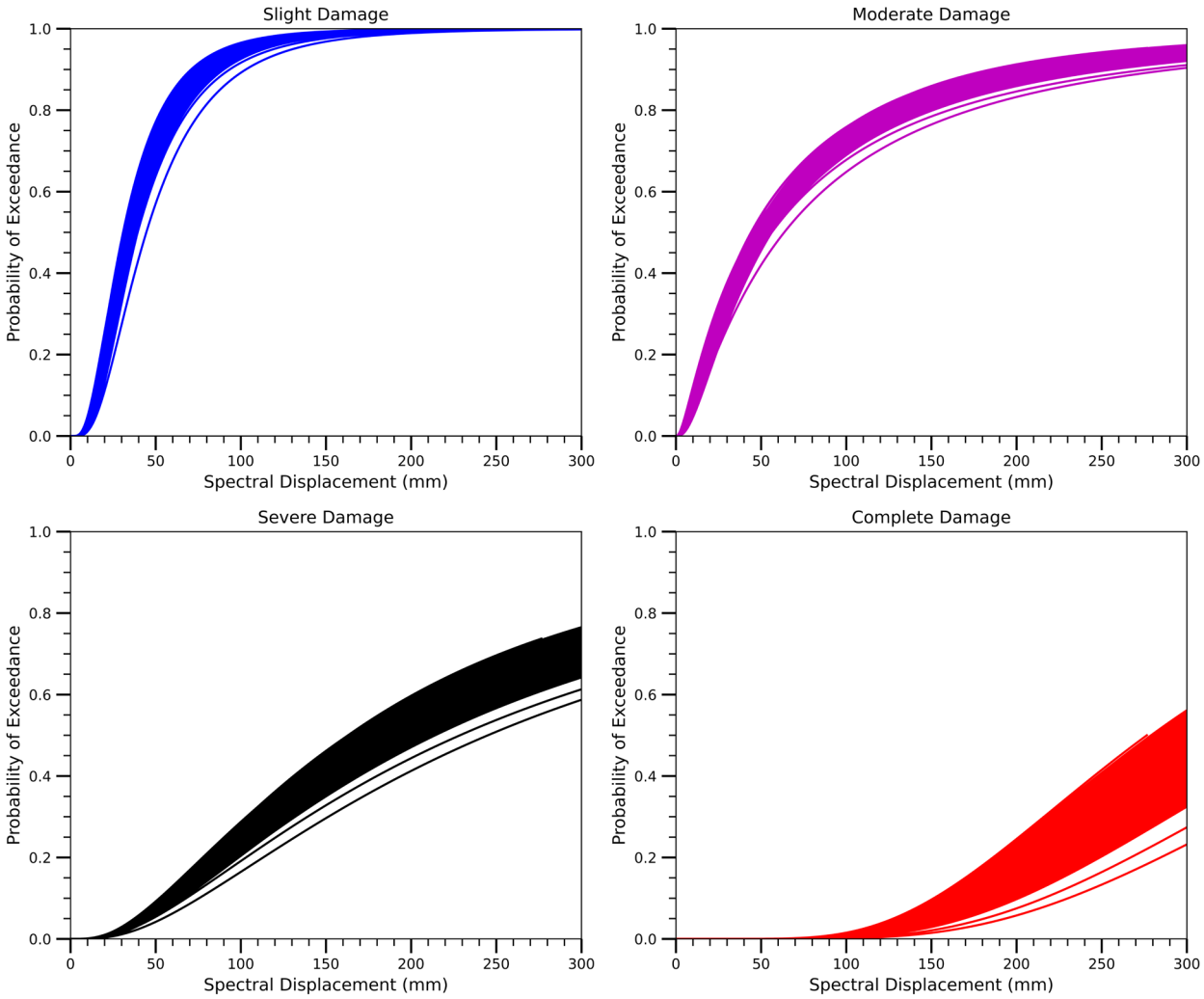

Figure 3. Seismic fragility curves are derived from various damage states for the 8-story RC frame building.

The weighted mean damage index, denoted as DS_m, serves as a single representative metric to indicate the expected damage condition of the building. This index corresponds to the discrete damage states and damage thresholds that are outlined in Table 3. Different hazard representations, demand–capacity models, and probability constructions can shift the resulting p_k and DS_m [47]. Accordingly, our damage probability matrix is constructed analytically at the performance point, which can be affected by the definitions of damage indices. The discrete probability p_k [N, d] quantifies the likelihood that a structure subjected to seismic demand d, and characterized by structural parameters *N*, will experience

the kth level of damage, categorized as the following: No-damage (k = 0), Slight (k = 1), Moderate (k = 2), Severe (k = 3), or Complete (k = 4). The most probable damage state is quantified using Equation (3). If DS_m = 2.0, it falls within the 1.5–2.5 interval [See Table 3]. Therefore, the Moderate damage state is expected in the RC frame building.

$$DS_m = \sum_{i=0}^{4} k\ p_k[N, d] \quad (3)$$

Table 2. Statical parameters of the collected simulation data for selected RC frame buildings.

Parameters	Mean	STD	Min	Max
β1 (Slight)	0.65	0.020	0.59	0.69
β2 (Moderate)	1.17	0.050	1.01	1.28
β3 (Severe)	0.91	0.030	0.82	0.97
β4 (Complete)	0.48	0.001	0.47	0.48
Slight Damage State (mm)	35.64	1.28	30.87	44.53
Moderate Damage State (mm)	50.92	1.82	44.10	63.62
Severe Damage State (mm)	192.42	9.63	162.98	245.09
Complete Damage State (mm)	333.92	18.52	276.51	426.56

Table 3. More likely damage states, corresponding mean damage index, and damage threshold intervals.

Mean Damage Index Intervals	More Likely Damage State (ML Class)	Damage Threshold
0.0–0.5	No-damage (Class 3)	—
0.5–1.5	Slight (Class 5)	$S_{d1} = 0.7D_y$
1.5–2.5	Moderate (Class 4)	$S_{d2} = D_y$
2.5–3.5	Severe (Class 2)	$S_{d3} = D_y + 0.25\left(D_u - D_y\right)$
3.5–4.0	Complete (Class 1)	$S_{d4} = D_u$

Note. Mean damage index intervals thresholds are the deterministic criteria for class assignment.

3.3. Transfer Learning-Based ML Model for SVA

The transfer learning framework utilizes an attention mechanism and begins with training on a four-story dataset. The dataset was first randomly partitioned into 80% development and 20% held-out test. Within the development set, 20% was used as a validation subset for hyperparameter tuning; the test set was not used for model selection. Following this, the pretrained model was fine-tuned using 2-story, 8-story, and 12-story datasets to evaluate its ability to generalize across different structural configurations. To address potential domain shift and ensure proper feature alignment across datasets, MinMax normalization was applied. The entire process is organized into the following main parts:

- Data processing;
- Base model;
- Progressive transfer learning;
- Attention mechanism;
- Hyperparameter optimization and fine-tuning.

3.3.1. Data Preprocessing

Several preprocessing steps are applied on the dataset to ensure the efficiency and stability of model. Preprocessing involves categorical encoding and feature scaling to ensure consistent and normalized data. For computational suitability, qualitative variables are converted into numerical representations using categorical encoding. For effective optimization and model stability, Standard Scaling is applied to normalize all numeric

features to have a mean of 0 and a standard deviation of 1. Scaling parameters are calculated from the training set to prevent data leakage and then applied to the validation sets.

3.3.2. Base Model

Based on the ANN, the architecture is designed for structural health predictions across different building configurations. The base model utilizes a fully connected ANN with 128 neurons in the input layer. This design choice stems from the need to efficiently capture high-dimensional input features. The architecture of the model starts with an input layer configured to accept multi-dimensional input features. A feature vector $x \in \mathbb{R}_d$, with d number of features, is fed to the input layer. Dense layers with 128 and 64 neurons are used as hidden layers. Hidden layers have dense layers with ReLU activation functions. The ReLU activation function is applied to introduce nonlinearity and ensure effective learning of intermediate representations. A dense layer with a 128-neuron layer is designed to extract high-dimensional feature interactions that balance the model's complexity and generalization. This layer is followed by dropout layer with probability p to prevent overfitting.

This layer is followed by dropout to prevent overfitting. A dense layer with 64 neurons combined with a dropout forms a bottleneck layer that condenses task-specific features and retains critical information while reducing redundancy. Finally, the output layer is configured to classify structural damage into four categories: Complete, Severe, Moderate, and Slight. The output layer uses the SoftMax activation function for probabilistic output given by Equation (4). P ($y_i = k \mid x_i$) denotes the probability of inputs that belong to class k. W_k and b_k are the weight of vector and bias corresponding to class k. The term $\sum_j exp(W_j h + b_j)$ performs summation over all possible classes that normalize the probability, ensuring a total of one. K is a set of classes: k ∈ {Complete, Severe, Moderate, and Slight}. For the model to perform regression tasks for severity prediction, linear activation is used by modifying the final dense layer as per Equation (5). W_{out} and b_{out} are the weight and bias of the output layer.

$$P(yi = k|xi) = \frac{exp(W_k h + b_k)}{\sum_j exp(W_j h + b_j)} \quad (4)$$

$$y = W_{out} \cdot h + b_{out} \quad (5)$$

The base model contains approximately 20,000 trainable parameters. This architecture strikes a balance between computational efficiency and model complexity. With these modular components, this model allows for the seamless integration of progressive transfer learning.

3.3.3. Progressive Transfer Learning (PTL)

The PTL approach is adapted to improve performance across multi-story building datasets by progressively transferring learned representations. The progressive transfer learning framework fine-tunes the model incrementally to ensure smooth knowledge transfer from the source to target domain. The base model is pretrained on a comprehensive four-story building damage dataset to capture generic feature representations, as per Equation (6). L is a loss function (Sparse Categorical Cross-Entropy). This loss function L(y, y′) is defined by Equation (7). y is a true label, y′ is the predicted probability for class i, and C is total number of classes. For the case in regression task, this loss function is replaced with mean squared error (MSE), provided by Equation (8), where N is the number of observations in the dataset. The initial layers, which capture fundamental features, remain frozen by Equation (9). Deeper layers are fine-tuned for the new task by Equation (10). This unfreezing of layers is performed progressively using an annealing function

(AF) controlling layer l given by Equation (11). L_o is the initially frozen layer count. σ is the sigmoid function for smooth transitions. τ is the temperature parameter that controls the unfreezing rate. The annealing function starts with a high temperature (explore solution space) and gradually reduces it to refine the solution. Temperature determines how the probabilities are distributed and controls randomness or exploration. A higher temperature produces more uniform probabilities.

$$\theta_S^* = \arg\min_\theta L(f(x;\theta), y) \quad (6)$$

$$L(y, y') = -\sum_{i=1}^{C} y_i log(y'_i) \quad (7)$$

$$LMSE = \frac{1}{N}\sum_{i=1}^{N}(yi - y'i)^2 \quad (8)$$

$$\theta_S^* = \theta frozen \quad (9)$$

$$\theta_{trainable} = \arg\min_\theta L(f(x;\theta_{trainable},\theta_{frozen}),y) \quad (10)$$

$$AF = \sigma\left(\frac{l-L_0}{\tau}\right) \quad (11)$$

Starting with the deeper layers, fine-tuning is performed gradually, unfreezing earlier layers. Layer freezing ensures that the core knowledge is preserved while allowing for incremental adaptation. This hierarchy in learning gradually refines feature extraction, avoiding drastic changes to the model's underlying knowledge. During this stage, the attention mechanism is refined to optimize the importance assigned to key features of the new dataset. The dense layer of 64 neurons with dropout acts as a new classification head to adapt the model to the new damage-classification task.

3.3.4. Attention Mechanism

Following the hidden layers, an attention mechanism is integrated to dynamically weigh the importance of the input features. The task of this layer is to enhance the interpretability and robustness of the model. Given extracted feature $h_{extracted}$, attention score $A_{attention}$ is calculated as per Equations (12) and (13).

$$E_i = v^T \tanh(W_a h_{extracted} + b_a) \quad (12)$$

$$A_{attention} = \frac{exp(e_i)}{\sum_j exp(e_j)} \quad (13)$$

$A_{attention}$ determines the importance of the features. W_a and b_a are learnable parameters. Output features from the frozen layer are used as inputs to the attention mechanism. The attention mechanism operates on these extracted features and assigns higher weights to critical inputs. This higher weight ensures that relevant structural attributes receive more focus. This in turn, refines the features dynamically, reduces redundant information, and improves feature extraction quality, improving adaptability to more complex datasets. Refined feature representation h_{att} is given by Equation (14). The output of the attention layer is then fed into the newly added classification head. The last dense layer with 64 neurons and ReLU is adapted as new classification head.

$$H_{att} = \sum_{attention=1}^{N} A_{attention} h_{extracted} \quad (14)$$

3.3.5. Hyperparameter Optimization and Fine-Tuning

To optimize performance, the learning rate is selected as 0.001, with a decay factor of 0.1 every 10 epochs, as defined in Equation (15). To balance efficiency and prevent overfitting, a batch size of 32 and a dropout rate of 0.2 are used. The multi-head attention

mechanism consists of eight heads to capture diverse feature relationships. Experiments are conducted with Adam, RMSprop, and SGD optimizers to determine the best optimization strategy. RMSprop (Root Mean Square Propagation) adjusts learning rates based on recent gradients to stabilize training. Stochastic Gradient Descent (SGD) updates weights using individual samples to reduce computational costs, requiring careful tuning of the learning rate. To ensure adaptive learning, the Adam optimizer is used, provided by Equation (16). θ_t denotes the current model's parameters at time t, θ_{t+1} denotes the model's parameters at the next time stamp, η is the learning rate, and ϵ is a small constant added to the denominator to prevent division by zero. Adaptive Moment Estimation (Adam optimizer) maintains a separate learning rate for each parameter. This optimizer uses estimates of the first and second moments of the gradients and computes individual adaptive learning rates for various existing parameters. m_t and v_t are the first- and second-moment estimates. The biased first-moment estimate (m_t) is an exponentially weighted moving average of the gradients defined by Equation (17). g_t is the gradient of the loss function at time step t. β_1 controls how much past gradients influence the current step. The biased second-moment estimate (v_t) is the exponentially weighted moving average of squared gradients with β_2, which controls how much past squared gradients influence the update by Equation (18). Fine-tuning used a smaller learning rate of 0.0001. The attention-enhanced classification head has a rate of 0.3. The base model is trained on 50 epochs and the PTL fine-tuning process is conducted for 20 epochs. For multi-class classification, Sparse Categorical Cross-Entropy loss is used.

$$\eta_t = \eta_0 \cdot 0.1^{\lfloor t/10 \rfloor} \ (\eta_0 = 0.001) \tag{15}$$

$$\theta_{t+1} = \theta_t - \eta \ \frac{m_t}{\sqrt{v_t + \varepsilon}} \tag{16}$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{17}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{18}$$

The proposed architecture uses the discussed ANN as a base model. For the complex task of predicting data for multi-story buildings, sequential learning via fine-tuning is enabled to implement progressive transfer learning combined with attention mechanisms. The approach begins with a base ANN model that contains an input layer of 128 neurons to process high-dimensional interactions among the input data. This enables them to capture the complex relationships between features and capabilities to learn intricate patterns. For feature compression and refined task-specific features, a bottleneck layer with 64 neurons was included. The integration of the bottleneck layer with dropout avoids overfitting and regularizes the network, therefore enabling generalized and efficient learning. Attention is incorporated by prioritizing critical features (e.g., material and load distribution) to enhance performance on task-specific features. To emphasize the load distribution of taller buildings with higher layers, attention weights shift based on story complexity. Initially, this model was trained on four-story data, saving its learned embeddings and weights.

The total loss incorporates the loss terms for both the base model and the progressively transferred knowledge model and attention. For base model loss L_{base}, Sparse Categorical Cross-Entropy loss is used for multiclass classification, and it is replaced by MSE in severity score. Progressive transfer loss is given as Equation (19). α controls knowledge retention. Attention regularization is given by Equation (20). This ensures the stability of attention weights. $A'_{attention}$ is the mean attention weight across samples. Finally, the total loss function is given by Equation (21); λ_1, λ_2, and λ_3 are balancing hyperparameters.

$$L_{ptl} = -\sum_{i=1}^{C} \text{yilog(y'i)} + \alpha||\theta_{new} - \theta_{old}||^2 \tag{19}$$

$$L_{att} = ||A_{attention} - A'_{attention}|| \quad (20)$$

$$L_{total} = \lambda_1 L_{base} + \lambda_2 L_{ptl} + \lambda_3 L_{att} \quad (21)$$

The gradient-based adaptation provided via attention dynamically learns to weigh structural features. This step ensures that highly relevant features receive stronger updates while redundant information receives less. The gradient-based update for attention is given in Equation (22).

$$\frac{\partial L_{total}}{\partial W_a} = \sum_{attention=1}^{N} A_{attention} \frac{\partial L_{total}}{\partial x_{attention}} \quad (22)$$

The implementation of the PTL framework with attention mechanisms was carried out using a suite of Python-based (v3.13) libraries. Data preprocessing and transformation were performed using Pandas (v2.3.3) for data handling, NumPy (v2.3.4) for numerical computations, and Scikit-learn (v1.7.2) for scaling (StandardScaler), label encoding (LabelEncoder), and dataset partitioning. Neural network models were developed using TensorFlow (v2.20.0) and Keras (v3.12), leveraging both Sequential and Functional APIs to construct flexible architectures. Core components such as Dense layers, Dropout (for regularization), BatchNormalization (for training stability), and MultiHeadAttention (for feature prioritization) were integrated into the model. KerasTuner (v1.4.8) was employed for automated hyperparameter optimization, while Lambda layers ensured compatibility between custom input–output transformations and attention mechanisms. For performance evaluation, Scikit-learn's metrics—mean squared error (MSE), R^2 score, classification report, and confusion matrix—were used to assess regression and classification outcomes. Finally, Principal Component Analysis (PCA) was applied to visualize class-wise separability in reduced dimensions, and Matplotlib (v3.10.7) and Seaborn (v0.12.0) supported the generation of performance plots and visual analytics. The algorithmic breakdown of the purposed present approach is presented in Algorithm 1.

Algorithm 1: Structural Damage Prediction using Progressive Transfer Learning and Attention Mechanism

Input:

1. Preprocessed dataset $D_{Source} = \{X_s, Y_s\}$;
2. Target dataset $D_{target} = \{X_t, Y_t\}$;
3. Where X represents structural parameters and Y denotes damage levels; base model f_{base}; learning rate η; Epochs E; Batch size B.

Output:

1. Final Trained model final for structural damage classification and damage index prediction.

Step 1: Base Model Definition and Training

1. Define an ANN architecture:
   (1) Input layer $h_o \in \mathbb{R}^d$;
   (2) Dense layers: 128 and 64 neurons, with ReLU activation;
   (3) Dropout layers to mitigate overfitting;
   (4) Output layer: SoftMax for classification and linear activation for regression.
2. Compile the model:
   (1) Use Sparse Categorical Cross-Entropy loss for classification;
   (2) For regression, replace with MSE (Equation (5)).
3. Train the base model f_{base} on a four-story dataset D_{source} for E_{base} epochs.

Step 2: Progressive Transfer Learning (PTL)

1. Load the pretrained base model f_{base} and freeze initial layers: $\theta_s^* = \theta_{frozen}$;
2. Fine-tune deeper layers on the target dataset D_{target}: (Equation (7));

3. Knowledge Retention Loss (progressive transfer loss L_{ptl}): To ensure that transferred knowledge is retained: L_{ptl} (Equation (16));
4. Implement an AF to gradually unfreeze layers: $AF = \sigma(\frac{l-L_0}{\tau})$, where L_o is the initially frozen layer count and σ is the sigmoid function;
5. Train for E_{ptl} epochs with a reduced learning rate η_{ptl}.

Step 3: Attention Mechanism Integration

1. Extract latent features $h_{extracted}$ from shared layers;
2. Compute attention scores and distribution (Equations (9) and (10));
3. Generate refined feature representations: h_{att} (Equation (11));
4. Attention Regularization Loss L_{att} (Equation (17));
5. Pass refined features to the new classification head (dense layer with 64 neurons, ReLU activation, and dropout).

Step 4: Total Loss Function and Gradient-Based Adaptation

1. Define total loss (Equation (15)); λ_1, λ_2, and λ_3 are balancing hyperparameters.
2. Update attention weights dynamically to focus on critical features (Equations (13)–(15).

Step 5: Hyperparameter Optimization and Fine-Tuning

1. Set learning rate decay: (Equation (12));
2. Use Adam optimizer for weight updates: (Equation (13));
3. Train for an additional E_{final} epoch.

4. Results and Discussion

The proposed PTL framework, combined with attention mechanisms, was evaluated for its effectiveness in predicting the seismic damage of RC frame buildings with varying heights (2-, 4-, 8-, and 12-story). The model was initially trained on a four-story dataset using a SVA approach, incorporating Monte Carlo simulation, NSPA, and the CSM, the model was subsequently fine-tuned for other building configurations through transfer learning.

Figure 4a presents the training, validation, and fine-tuning loss trends. The training loss (blue) rapidly decreases and stabilizes, while the validation loss (orange) remains consistently lower, indicating that the model generalizes well without overfitting. The flat and low fine-tuning loss curve (green) reflects successful adaptation to new structural configurations, suggesting that minimal additional learning was required. The regression phase of the model, aimed at predicting the damaged index, demonstrated high precision. As shown in Table 4, the PTL model achieved an MSE of 0.0002 and an R^2 value of 0.98, indicating highly accurate predictions with near-complete variance capture. This is visually confirmed by Figure 4b, where predicted and actual values show near-perfect correlation, validating the model's robustness.

Subsequently, the continuous damage index was discretized into five ordinal categories—No-damage, Slight, Moderate, Severe, and Complete—via label encoding, transforming the regression output into a classification problem. Table 5 summarizes the precision, recall, and F1-scores. Furthermore, Figure 5 illustrates the confusion matrix reflecting the classification performance across various damage categories. Our new PTL-based model demonstrates strong performance in correctly identifying "Complete" damage cases, with 20,620 instances accurately classified. Similarly, the "No-damage" category shows high accuracy, with 4523 correct classifications. The model also performs reasonably well for "Slight" damage (2419 correct classifications), "Moderate" damage (1674 correct classifications), and "Severe" damage (1670 correct classifications). However, several misclassification patterns are evident. Notably, 76 instances of "Complete" damage were incorrectly classified as "No-damage", which could have significant implications in

disaster-response scenarios where complete structural failures might be overlooked. The model also misclassified 114 "Moderate" damage cases as "Severe" and 104 as "Slight", indicating difficulties in distinguishing between adjacent damage severity levels. For "Severe" damage, the model incorrectly classified 155 instances as "Complete" and 95 as "Moderate", suggesting a tendency to either overestimate or underestimate damage in borderline cases. Similarly, 209 "Slight" damage cases were misclassified as "No-damage", and 91 as "Moderate", revealing challenges in detecting minimal structural damage. Precision was highest for the "Complete" category (0.993), vital for post-earthquake interventions where accurate identification of total collapse is critical. However, the "Moderate" (0.900) and "Severe" (0.899) categories exhibited lower precision, reflecting occasional mislabeling. Recall followed a similar trend: "Complete" and "No-damage" achieved 0.996 and 0.989, respectively, while "Severe" had the lowest recall at 0.871, indicating some under-detection of significant damage. F1-scores ranged from 0.885 ("Severe") to 0.994 ("Complete"), providing a balanced view of classification reliability. The macro-averaged precision, recall, and F1-scores were all 0.945, indicating excellent overall performance. The implications of misclassifications between certain categories (e.g., "Severe" classified as "Moderate") warrant careful consideration in the context of practical applications, where such errors might influence resource allocation and response planning in post-disaster scenarios.

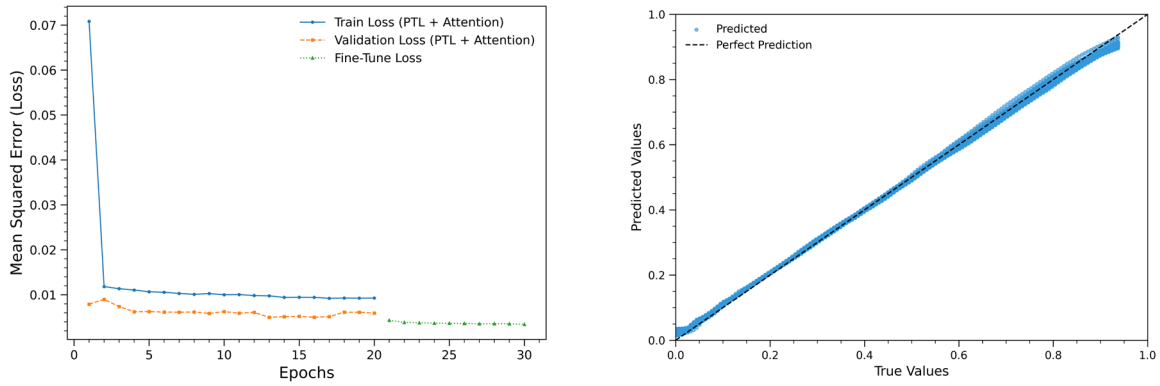

Figure 4. (**a**) Training, validation, and fine-tuning loss curves; (**b**) predicted vs. actual damage index from regression model.

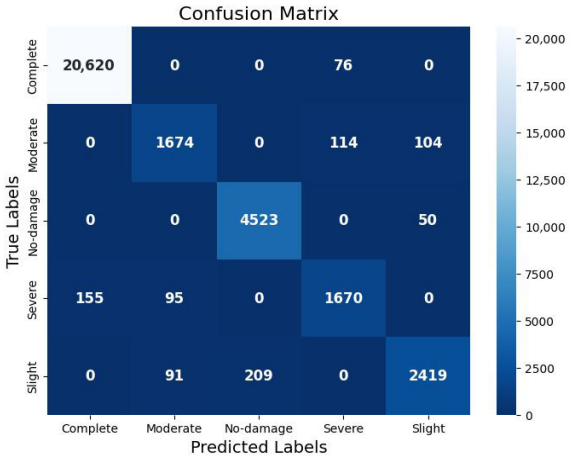

Figure 5. Confusion matrix shows the classification performance of the PTL model across five structural damage categories.

Table 4. Comparative evaluation of ML models with mean squared error (MSE), R^2 score performance metrics, and regression model.

Model	Mean Squared Error (MSE)	R^2 Score
Progressive TL with Attention	0.002	0.98
Traditional Deep Learning	0.005	0.92
Basic Transfer Learning	0.002	0.94
Random Forest Regression	0.010	0.89
Gradient Boosting	0.007	0.91
Linear Regression	0.050	0.75

Table 5. Precision, recall, and F1-score metrics derived from the confusion matrix for each damage category.

Structural Damage	Precision	Recall	F1-Score
Class 1—Complete	0.98	0.99	0.99
Class 2—Severe	0.89	0.90	0.89
Class 3—No-damage	0.94	0.99	0.97
Class 4—Moderate	0.89	0.84	0.86
Class 5—Slight	0.95	0.86	0.90

Figure 6 demonstrates the correlation between actual and predicted damage classes, indicating minor discrepancies, especially within the "Slight" and "Moderate" categories. Figure 7 presents a PCA-based 2D projection that explains the classification behavior of the proposed model. Reducing the high-dimensional feature space to two principal components effectively visualizes class separability. In the plot, blue dots represent correctly classified samples, whereas red "X" marks indicate misclassifications. The "Complete" damage category constitutes a distinct and closely clustered region, thereby affirming high model accuracy and feature separability. The "Slight" damage class exhibits a greater density of misclassifications, presumably resulting from overlapping features with adjacent classes. Instances of "moderate" damage are notably more dispersed, particularly around the central low-variance area, underscoring the difficulty in identifying subtle inter-class distinctions. The "Severe" category demonstrates a lower rate of misclassifications and maintains clearer boundaries, suggesting improved feature representation. Simultaneously, the "No-damage" class exhibits significant overlap with low-damage states, indicating challenges in distinguishing it from early stage damage patterns. PCA visualization indicates that the model successfully captures dominant decision boundaries.

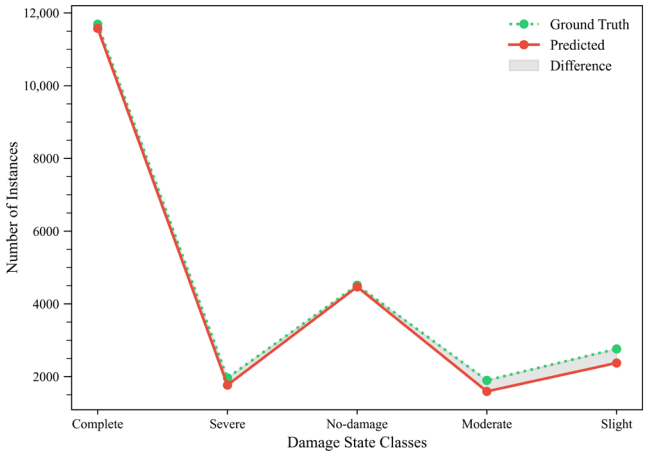

Figure 6. Comparison between actual and predicted class distribution of structural damage states using PTL classifier.

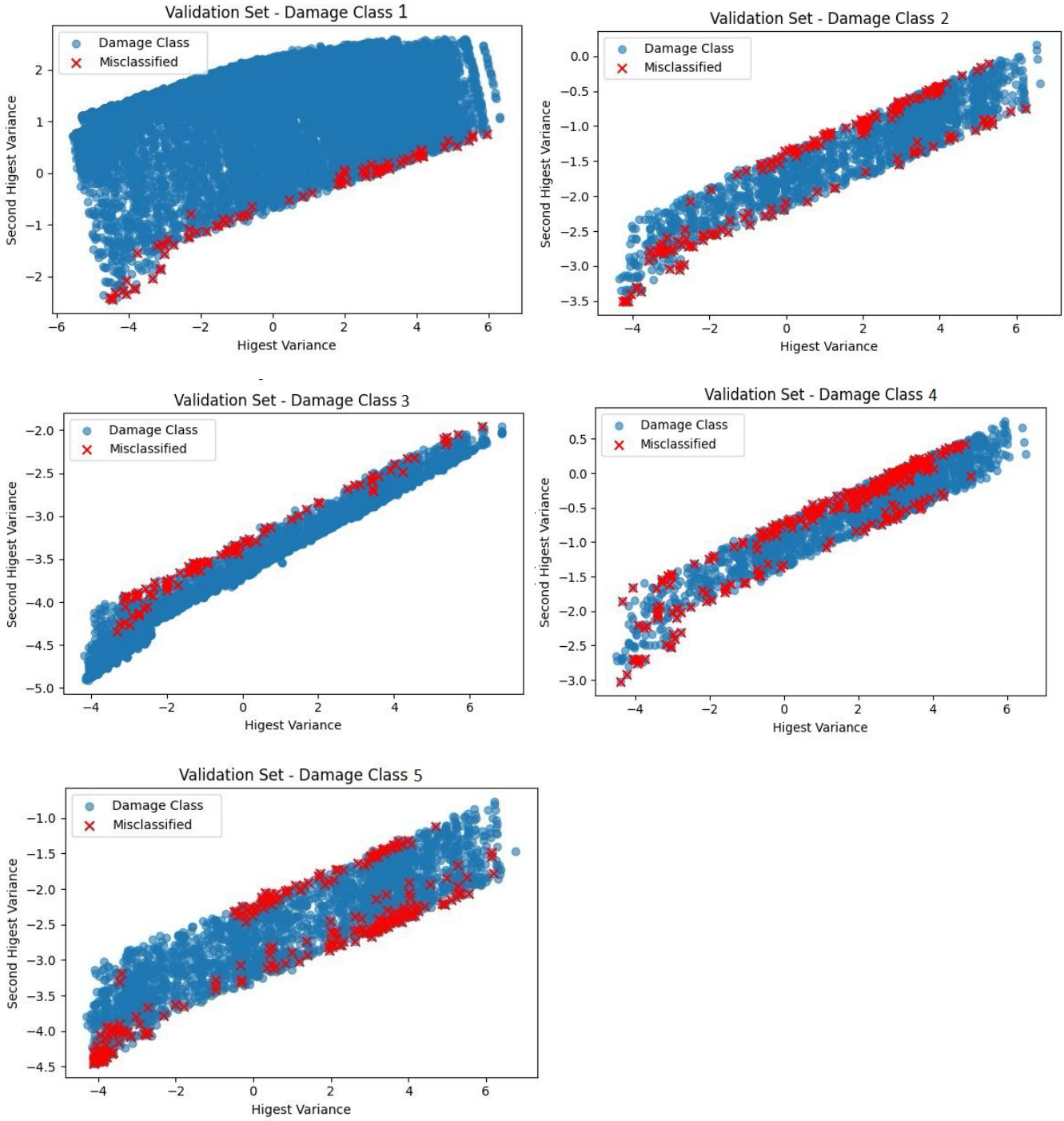

Figure 7. Principal Component Analysis (PCA) plot for visualizing correct and incorrect classifications of damage states.

To evaluate the relative effectiveness of the proposed model, a comparative analysis was conducted against a range of traditional and deep learning-based approaches, as summarized in Table 6. Classical ML algorithms—Random Forest (RF), Gradient-Boosted Trees (GBT), and Support Vector Machine (SVM)—achieved classification accuracies of 70.82%, 75.18%, and 65.02%, respectively, indicating moderate performance in seismic damage classification. In contrast, deep learning models demonstrated substantial improvements, with a standard Convolutional Neural Network (CNN) achieving 85.11% accuracy. The application of basic TL to the CNN further improved accuracy to 90.25%. The proposed PTL model, augmented with attention mechanisms, achieved the highest

classification accuracy of 95.64%, underscoring its superior capacity to extract and generalize complex patterns from structural data.

Table 6. Comparative analysis of different ML models on the given dataset in terms of accuracy (%) for a given classification task of identifying the structural damage.

Model	Accuracy (%) of Classification
RF	70.82
GBT	75.18
SVM	65.02
CNN	85.11
CNN with basic TL	90.25
Progressive TL with attention	95.64

Furthermore, the influence of optimization algorithms on model performance was evaluated, as presented in Table 7. Among the tested optimizers—Adam, RMSprop, and SGD—Adam yielded the best performance across both classification and regression tasks. Specifically, it achieved the lowest MSE of 0.0002, the highest R^2 score of 0.99, and the peak classification accuracy of 95.64%. RMSprop followed closely, delivering a classification accuracy of 93.12%, an MSE of 0.0003, and a higher classification loss of 0.1725. In contrast, SGD demonstrated the least effective performance with a classification accuracy of 88.65%, a higher MSE of 0.0004, and a loss of 0.2201, indicating slower convergence and reduced learning efficiency. Figure 8 shows through SHAP analysis that the spectral displacement at the performance point (S_d) is the dominant driver of predicted damage index, consistent with fragility formulations that express exceedance probabilities as functions of displacement demand. Higher ductility and lateral stiffness yield predominantly negative SHAP values, reducing predicted damage, because ductile members dissipate hysteretic energy and adequate stiffness limits inter-story drift, the primary damage proxy [25]. Similarly, low ductility or insufficient stiffness shifts predictions toward higher damage states. Strength parameters (f_y and f_{ck}) also reduce predicted damage, aligning with mechanics-based expectations, though their effect sizes are smaller than those of S_d, ductility, and stiffness. The comparatively minor contributions of story and material type, after conditioning on S_d and mechanical properties, indicate that the learned representation prioritizes demand–capacity interactions at the performance point over nominal configuration labels, coherent with the PTL + attention architecture's focus on task-salient features. Overall, these attributions align with capacity–design principles and prior observations for RC frames, underscoring the importance of ductile detailing and drift control for damage mitigation. To assess the contribution of different fine-tuning schedules and attention mechanisms, a comprehensive ablation study was performed. Table 8 summarizes the comparative performance of several configurations in terms of the mean squared error (MSE), coefficient of determination (R^2), and overall classification accuracy. The PTL + attention configuration outperforms all baselines, attaining an MSE of 0.002 and R^2 of 0.98, with an overall classification accuracy of 95.64%. Compared with single-head attention or conventional transfer learning, the proposed scheme improves accuracy by up to 3–5%, confirming the benefit of progressive unfreezing coupled with feature-level attention regularization.

The present study improves the precision of structural damage prediction and provides a scalable, adaptable method for evaluating various types of RC building frames. The findings validate that the proposed PTL framework, combined with attention mechanisms, provides a data-efficient and reliable method for predicting seismic damage, in accordance with Indian seismic design standards. The proposed PTL–attention approach offers clear advantages in terms of scalability and computational efficiency, as a model

trained on a single four-story configuration can be progressively adapted to other building heights with minimal retraining, while the attention mechanism enhances interpretability by prioritizing structurally meaningful features. At the same time, the methodology has notable limitations, as it relies entirely on simulation-based fragility data and does not yet include validation against irregular configurations, diverse structural typologies, nonlinear time–history analyses, or real post-earthquake observations. A broader and more heterogeneous dataset, together with real-world validation, will therefore be essential to establish the generalizability and practical applicability of the framework.

Table 7. Adam, RMSprop, and SGD optimizers performance to predict the output as structural damage (classification-based) and damage index (regression-based).

Optimizer	Classification Accuracy (%)	Loss (Sparse Categorical Cross-Entropy: Classification Based)	Regression (R^2)	Loss (MSE—Regression Based)
Adam	95.64	0.1353	0.9970	0.0002
RMSprop	93.12	0.1725	0.9920	0.0003
SGD	88.65	0.2201	0.9850	0.0004

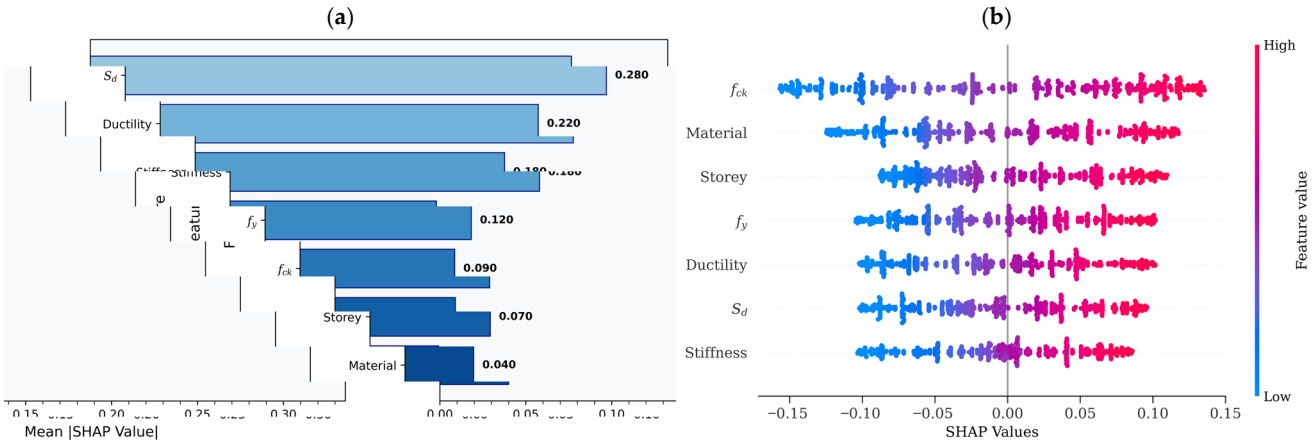

Figure 8. SHAP plot of input features for damage index detection dataset. (**a**) Mean |SHAP value|. (**b**) Bee-swarm of feature attributes.

Table 8. Comparative evaluation of transfer learning and attention configurations for SVA damage state prediction.

Category	Model	Hyperparameters	MSE	R^2	Accuracy (%)
Unfreeze/AF Schedule	Head-only fine-tune	LR head = 0.001, Adam	0.0012	0.960	88.50
	Last block unfrozen	LR base = 0.0001	0.0050	0.975	91.20
	Full retrain (no TL)	LR = 0.0001, Adam	0.0090	0.970	90.40
	Random unfreeze	same LR	0.0150	0.962	89.00
Attention	No attention	LR head = 0.001, Adam	0.0450	0.974	90.80
	Attention without PTL	LR head = 0.0001, Adam	0.0060	0.969	89.60
	Single-head attention	LR head = 0.001, Adam	0.0028	0.978	92.30
Proposed present study	PTL + attention	LR head = 0.001, Adam	0.0020	0.980	95.64

5. Conclusions

This study introduces a comprehensive and scalable methodology for assessing the seismic vulnerability of RC frame buildings, employing a PTL approach integrated with

attention mechanisms. The study employed probabilistic techniques, including Monte Carlo simulation, the CSM, and NSPA, and various structural responses to various seismic hazard demands accounting for epistemic and aleatoric uncertainties. An initial ANN model developed for four-story RC frames was systematically refined to include 2-, 8-, and 12-story configurations, resulting in enhanced prediction accuracy and reduced computational demands. The incorporation of attention mechanisms enhanced both model interpretability and performance across diverse structural typologies. Based on the detailed analysis and observed model performance, the following are the key findings of the proposed PTL-based seismic damage prediction framework:

- The proposed PTL framework effectively generalizes learned features from the source model (four-story RC frame) to target domains (2-, 8-, and 12-story RC frames) without requiring extensive retraining.
- The classification model achieved 95.64% accuracy with strong F1-scores across all damage states, while the regression model reached an R^2 of 0.98 with a minimal MSE = 0.0002.
- Attention mechanisms improved the model's focus on critical input features (e.g., stiffness, material strength, and ductility), boosting performance and reducing misclassification.
- The model was able to distinguish between damage classes (No-damage to Complete) effectively, though minor confusion was observed in borderline cases of the explicitly moderate vs. severe damage states.
- Comparative evaluation confirmed the superiority of the PTL-based model over traditional ML techniques (e.g., RF and SVM) and basic TL approaches.

The present study establishes a foundation for extending SVA models beyond the tested regular RC frames with uniform bay configurations, using a PTL–attention framework that shows promising scalability across different building heights. However, the current results are conditional on several modeling assumptions. First, all frames are regular in plan and elevation with fixed bay configurations; irregular layouts, vertical discontinuities, and soft-storey systems are not yet represented and may exhibit different transferability behavior. Second, the damage probabilities, p_k, and mean damage index, DS_m, are computed from an analytically constructed damage probability matrix at the capacity–spectrum performance point using deterministic damage thresholds derived from a bilinear idealization. Different choices of hazard representation, demand–capacity modeling, and damage-index definitions would inevitably shift the resulting p_k and DS_m, and this sensitivity is an inherent limitation of the present framework. Third, the dataset is fully simulation-based and has not yet been calibrated against real seismic recordings, field damage surveys, or monitoring data. Within these bounds, the proposed PTL architecture remains modular and adaptable, and the attention mechanism provides a flexible way to reweigh features when moving towards more complex applications, including irregular plan geometries, alternative materials, and soil–structure interactions. This approach shows potential for incorporating soil–structure interaction effects and adjusting to diverse seismic code requirements in various geographic areas. These directions present considerable potential for enhancing the model's generalizability and impact, facilitating the development of a comprehensive, scalable solution for seismic risk assessments in complex urban environments.

Author Contributions: Conceptualization, K.M.G. and K.D.T.; Methodology, K.M.G.; Software, A.R. and K.M.G.; Validation, J.A.A. and K.B.; Formal Analysis, K.M.G. and A.R.; Investigation, K.M.G.; Resources, J.A.A. and K.B.; Data Curation, A.R.; Writing—Original Draft Preparation, K.M.G.; Writing—Review and Editing, K.D.T., J.A.A. and K.B.; Visualization, A.R.; Supervision, K.D.T. and

J.A.A.; Project Administration, K.M.G.; Funding Acquisition, K.M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data, simulation codes, and trained models that support the findings of this study are available from the corresponding authors upon reasonable request.

Acknowledgments: The authors acknowledge the Ministry of Education, Government of India, for its continued initiatives supporting academic research and institutional capacity development under which this work was conceptualized. The authors also thank the Gujarat Technological University—School of Engineering and Technology and the City, University of London, for providing research resources and collaborative facilities. During the preparation of this work, the authors used ChatGPT (v5.1) to improve language clarity and readability. After using this tool, the authors reviewed and edited the content as necessary and take full responsibility for the final text.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANN	Artificial Neural Network
ADRS	Acceleration–Displacement Response Spectra
ATC	Applied Technology Council
CSM	Capacity Spectrum-Based Method
CI	Catastrophic Interference
CM	Confusion Matrix
DL	Dead Load
DS_m	Weighted Mean Damage Index
FEM	Finite Element Method
GBT	Gradient-Boosting Trees
HAZUS	Hazards U.S. Multi-Hazard Methodology
LL	Live Load
MAE	Mean Absolute Error
ML	Machine Learning
MSA	Multiple Stripe Analysis
MSE	Mean Squared Error
NSPA	Nonlinear Static Pushover Analysis
OAPI	Open Application Programming Interface
PCA	Principal Component Analysis
PTL	Progressive Transfer Learning
ReLU	Rectified Linear Unit
RF	Random Forest
SCS	Structural Control Systems
SVA	Seismic Vulnerability Assessment
SVM	Support Vector Machine
TL	Transfer Learning

Appendix A

Data collection is a crucial step in the implementation of a purposeful knowledge transfer approach to assess seismic vulnerability. In Table A1, a few sample datasets are presented, which will be utilized to train the progressive transfer learning (PTL) model.

Table A1. Forty data points are shown randomly from eight hundred ninety-three thousand, one hundred six data points.

Sr. №	Story	Material	Structural Type	f_y	f_{ck}	Ductility	Stiffness	S_d	Damage State
1	8	RC	SMRF	24.53	452.44	5.13	6029.91	59	Moderate
2	8	RC	SMRF	25.77	494.00	5.10	6033.48	59	Moderate
3	8	RC	SMRF	25.04	510.94	5.23	5992.86	59	Moderate
4	8	RC	SMRF	24.21	512.24	5.06	6005.80	59	Moderate
5	8	RC	SMRF	24.75	487.05	5.22	5934.82	59	Moderate
6	8	RC	SMRF	25.39	515.68	5.32	6070.54	59	Moderate
7	8	RC	SMRF	24.56	513.04	5.02	6147.77	59	Moderate
8	8	RC	SMRF	24.96	479.66	5.06	5959.38	59	Moderate
9	8	RC	SMRF	24.94	509.66	5.05	6091.52	59	Moderate
10	8	RC	SMRF	22.73	511.29	5.23	5895.54	59	Moderate
11	8	RC	SMRF	26.05	507.04	5.15	6232.59	59	Moderate
12	8	RC	SMRF	25.41	494.78	5.20	5994.64	59	Moderate
13	8	RC	SMRF	23.92	463.92	5.08	6022.77	59	Moderate
14	8	RC	SMRF	23.94	491.37	5.26	5991.52	59	Moderate
15	2	RC	SMRF	24.51	477.44	5.33	5979.02	75	Severe
16	2	RC	SMRF	24.31	469.85	5.33	5913.39	75	Severe
17	2	RC	SMRF	25.02	523.37	5.09	5937.50	75	Severe
18	2	RC	SMRF	25.16	496.55	5.27	5985.71	75	Severe
19	12	RC	SMRF	25.11	507.52	5.00	6161.61	128	Complete
20	12	RC	SMRF	24.38	517.14	4.99	5942.41	128	Complete
21	12	RC	SMRF	25.33	524.54	5.03	5920.09	128	Complete
22	8	RC	SMRF	24.67	504.15	5.07	6073.66	28	Slight
23	8	RC	SMRF	24.92	477.09	5.14	6014.29	28	Slight
24	8	RC	SMRF	24.68	516.32	5.07	6225.00	28	Slight
25	8	RC	SMRF	24.10	490.70	5.14	6020.09	28	Slight
26	8	RC	SMRF	25.81	491.24	5.07	6105.80	28	Slight
27	8	RC	SMRF	25.38	487.27	5.25	6084.82	28	Slight
28	8	RC	SMRF	25.58	500.08	5.01	6297.77	28	Slight
29	8	RC	SMRF	25.25	491.25	5.04	6205.80	28	Slight
30	8	RC	SMRF	25.26	489.74	5.13	6175.89	28	Slight
31	8	RC	SMRF	25.89	549.48	4.93	6222.77	28	Slight
32	8	RC	SMRF	24.14	477.44	5.33	5979.02	28	Slight
33	8	RC	SMRF	24.31	469.85	5.33	5913.39	28	Slight
34	8	RC	SMRF	25.02	523.37	5.09	5937.50	28	Slight
35	2	RC	SMRF	24.90	489.12	4.98	6037.05	8	No-damage
36	2	RC	SMRF	25.21	490.92	5.13	6159.82	8	No-damage
37	2	RC	SMRF	24.48	520.70	5.10	6233.93	8	No-damage
38	2	RC	SMRF	25.08	505.25	5.77	5949.55	8	No-damage
39	2	RC	SMRF	25.76	511.70	5.25	5832.14	8	No-damage
40	2	RC	SMRF	24.99	496.73	5.00	6026.34	8	No-damage

References

1. Öztürk, M.; Karan, M.A. Impact of Near-Fault Seismic Inputs on Building Performance: A Case Study Informed by the 2023 Maras Earthquakes. *Appl. Sci.* **2025**, *15*, 10142. https://doi.org/10.3390/app151810142.
2. Palanci, M.; Demir, A.; Kayhan, A.H. The Investigation of Displacement Demands of Single Degree of Freedom Models Using Real Earthquake Records Compatible with TBEC-2018. *Pamukkale J. Eng. Sci.* **2021**, *27*, 251–263. https://doi.org/10.5505/pajes.2020.47936.

3. Kennedy, R.P.; Cornell, C.A.; Campbell, R.D.; Kaplan, S.; Perla, H.F. Probabilistic Seismic Safety Study of an Existing Nuclear Power Plant. *Nucl. Eng. Des.* **1980**, *59*, 315–338. https://doi.org/10.1016/0029-5493(80)90203-4.
4. Federal Emergency Management Agency (FEMA); National Institute of Building Sciences (NIBS). *HAZUS-MH MR5: Advanced Engineering Building Module—Technical and User's Manual*; Federal Emergency Management Agency: Washington, DC, USA, 2003.
5. Applied Technology Council (ATC). *Seismic Evaluation and Retrofit of Concrete Buildings (Vols. 1 & 2)*; Applied Technology Council: Redwood City, CA, USA,1996.
6. Gao, Y.; Mosalam, K.M. Deep Transfer Learning for Image-Based Structural Damage Recognition. *Comput. Aided Civ. Eng.* **2018**, *33*, 748–768. https://doi.org/10.1111/mice.12363.
7. Ogunjinmi, P.D.; Park, S.-S.; Kim, B.; Lee, D.-E. Rapid Post-Earthquake Structural Damage Assessment Using Convolutional Neural Networks and Transfer Learning. *Sensors* **2022**, *22*, 3471. https://doi.org/10.3390/s22093471.
8. Avci, O.; Abdeljaber, O.; Kiranyaz, S.; Hussein, M.; Gabbouj, M.; Inman, D.J. A Review of Vibration-Based Damage Detection in Civil Structures: From Traditional Methods to Machine Learning and Deep Learning Applications. *Mech. Syst. Signal Process.* **2021**, *147*, 107077. https://doi.org/10.1016/j.ymssp.2020.107077.
9. Liuzzi, M.; Aravena Pelizari, P.; Geiß, C.; Masi, A.; Tramutoli, V.; Taubenböck, H. A Transferable Remote Sensing Approach to Classify Building Structural Types for Seismic Risk Analyses: The Case of Val d'Agri Area (Italy). *Bull. Earthq. Eng.* **2019**, *17*, 4825–4853. https://doi.org/10.1007/s10518-019-00648-7.
10. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In Proceedings of the The Thirty-First Annual Conference on Neural Information Processing Systems (NIPS) 2017, Long Beach, CA, USA, 4–9 December 2017.
11. Rasheed, A.; Usman, M.; Zain, M.; Iqbal, N. Machine Learning-Based Fragility Assessment of Reinforced Concrete Buildings. *Comput. Intell. Neurosci.* **2022**, *2022*, 5504283. https://doi.org/10.1155/2022/5504283.
12. Kazemi, F.; Asgarkhani, N.; Jankowski, R. Machine Learning-Based Seismic Fragility and Seismic Vulnerability Assessment of Reinforced Concrete Structures. *Soil Dyn. Earthq. Eng.* **2023**, *166*, 107761. https://doi.org/10.1016/j.soildyn.2023.107761.
13. Ghosh, S.; Chakraborty, S. Seismic Fragility Analysis of Bridges by Relevance Vector Machine Based Demand Prediction Model. *Earthq. Eng. Eng. Vib.* **2022**, *21*, 253–268. https://doi.org/10.1007/s11803-022-2082-7.
14. Gondaliya, K.; Kumari, S.; Amin, J.; Ansari, A.; Bush, R. A Novel Machine Learning Framework for Seismic Vulnerability Assessment of Urban Infrastructure. In Proceedings of the 2025 International Conference for Artificial Intelligence, Applications, Innovation and Ethics (AI2E), Muscat, Oman, 3 February 2025; IEEE: Piscataway, NJ, USA; pp. 1–5.
15. Chen, Y.; Sun, Z.; Zhang, R.; Yao, L.; Wu, G. Attention Mechanism Based Neural Networks for Structural Post-Earthquake Damage State Prediction and Rapid Fragility Analysis. *Comput. Struct.* **2023**, *281*, 107038. https://doi.org/10.1016/j.compstruc.2023.107038.
16. Asgarkhani, N.; Kazemi, F.; Jankowski, R.; Formisano, A. Dynamic Ensemble-Learning Model for Seismic Risk Assessment of Masonry Infilled Steel Structures Incorporating Soil-Foundation-Structure Interaction. *Reliab. Eng. Syst. Saf.* **2026**, *267*, 111839. https://doi.org/10.1016/j.ress.2025.111839.
17. Wang, J.; Xie, Y.; Guo, T.; Du, Z. Predicting the Influence of Soil–Structure Interaction on Seismic Responses of Reinforced Concrete Frame Buildings Using Convolutional Neural Network. *Buildings* **2023**, *13*, 564. https://doi.org/10.3390/buildings13020564.
18. Mata, R.; Nuñez, E.; Hernández, M.; Correa, C.; Bustamante, G. Seismic Performance of RC Moment Frame Buildings Considering SSI Effects: A Case Study of the New Venezuelan Seismic Code. *Buildings* **2023**, *13*, 1694. https://doi.org/10.3390/buildings13071694.
19. Abdi, G.; Jabari, S. A Multi-Feature Fusion Using Deep Transfer Learning for Earthquake Building Damage Detection. *Can. J. Remote Sens.* **2021**, *47*, 337–352. https://doi.org/10.1080/07038992.2021.1925530.
20. Dogan, G.; Hakan Arslan, M.; Ilki, A. Detection of Damages Caused by Earthquake and Reinforcement Corrosion in RC Buildings with Deep Transfer Learning. *Eng. Struct.* **2023**, *279*, 115629. https://doi.org/10.1016/j.engstruct.2023.115629.
21. Xu, Y.; Lu, X.; Fei, Y.; Huang, Y. Iterative Self-Transfer Learning: A General Methodology for Response Time-History Prediction Based on Small Dataset. *J. Comput. Des. Eng.* **2022**, *9*, 2089–2102. https://doi.org/10.1093/jcde/qwac098.
22. Jena, R.; Shanableh, A.; Al-Ruzouq, R.; Pradhan, B.; Gibril, M.B.A.; Ghorbanzadeh, O.; Atzberger, C.; Khalil, M.A.; Mittal, H.; Ghamisi, P. An Integration of Deep Learning and Transfer Learning for Earthquake-Risk Assessment in the Eurasian Region. *Remote Sens.* **2023**, *15*, 3759. https://doi.org/10.3390/rs15153759.

23. Lin, Q.; Ci, T.; Wang, L.; Mondal, S.K.; Yin, H.; Wang, Y. Transfer Learning for Improving Seismic Building Damage Assessment. *Remote Sens.* **2022**, *14*, 201. https://doi.org/10.3390/rs14010201.
24. Samuel, M.A.; Xiong, E.; Haris, M.; Lekeufack, B.C.; Xie, Y.; Han, Y. Assessing Seismic Vulnerability Methods for RC-Frame Buildings Pre- and Post-Earthquake. *Sustainability* **2024**, *16*, 10392. https://doi.org/10.3390/su162310392.
25. Yu, Z.; Shen, D.; Jin, Z.; Huang, J.; Cai, D.; Hua, X.-S. Progressive Transfer Learning. *IEEE Trans. Image Process.* **2019**, *31*, 1340–1348.
26. Iman, M.; Miller, J.A.; Rasheed, K.; Branch, R.M.; Arabnia, H.R. EXPANSE: A Deep Continual/Progressive Learning System for Deep Transfer Learning. *arXiv* **2022**, https://doi.org/10.48550/arXiv.2205.10356.
27. Bulleit, W.M. Uncertainty in Structural Engineering. *Pract. Period. Struct. Des. Constr.* **2008**, *13*, 24–30. https://doi.org/10.1061/(ASCE)1084-0680(2008)13:1(24).
28. Dolšek, M. Simplified Method for Seismic Risk Assessment of Buildings with Consideration of Aleatory and Epistemic Uncertainty. *Struct. Infrastruct. Eng.* **2011**, *8*, 939–953. https://doi.org/10.1080/15732479.2011.574813.
29. Nafeh, A.M.B.; O'Reilly, G.J. Fragility function uncertainty quantification in infilled RC frame buildings. In Proceedings of the 9th ECCOMAS Thematic Conference on Computational Methods in Dynamics and Earthquake Engineering, Athens, Greece, 12–14 June 2023; pp. 2168–2181.
30. Franchin, P.; Pinto, P.E.; Rajeev, P. Confidence Factor? *J. Earthq. Eng.* **2010**, *14*, 989–1007. https://doi.org/10.1080/13632460903527948.
31. Crowley, H.; Bommer, J.J.; Pinho, R.; Bird, J. The Impact of Epistemic Uncertainty on an Earthquake Loss Model. *Earthq. Engng. Struct. Dyn.* **2005**, *34*, 1653–1685. https://doi.org/10.1002/eqe.498.
32. Gondaliya, K.M.; Vasanwala, S.A.; Desai, A.K.; Amin, J.A.; Bhaiya, V. Machine Learning-Based Approach for Assessing the Seismic Vulnerability of Reinforced Concrete Frame Buildings. *J. Build. Eng.* **2024**, *97*, 110785. https://doi.org/10.1016/j.jobe.2024.110785.
33. Gondaliya, K.; Bhaiya, V.; Vasanwala, S.; Desai, A. Probabilistic Seismic Vulnerability of Indian Code-Compliant RC Frame. *Pract. Period. Struct. Des. Constr.* **2022**, *27*, 04022028. https://doi.org/10.1061/(ASCE)SC.1943-5576.0000708.
34. *IS 456*; Plain and Reinforced Concrete—Code of Practice. Bureau of Indian Standards: New Delhi, India, 2000.
35. *IS 875 (Part 1)*; Design Loads (Other Than Earthquake) for Buildings and Structures—Dead Loads. Bureau of Indian Standards: New Delhi, India, 1987.
36. *IS 875 (Part 2)*; Design Loads (Other Than Earthquake) for Buildings and Structures—Imposed Loads. Bureau of Indian Standards: New Delhi, India, 1987.
37. *IS 1893 (Part 1)*; Criteria for Earthquake Resistant Design of Structures—General Provisions and Buildings. Bureau of Indian Standards: New Delhi, India, 2016.
38. *IS 13920*; Ductile Detailing of Reinforced Concrete Structures Subjected to Seismic Forces—Code of Practice. Bureau of Indian Standards: New Delhi, India, 2016.
39. Barbat, A.H.; Pujades, L.G.; Lantada, N. Performance of Buildings under Earthquakes in Barcelona, Spain. *Comput. Aided Civ. Eng* **2006**, *21*, 573–593. https://doi.org/10.1111/j.1467-8667.2006.00450.x.
40. Barbat, A.H.; Carreño, M.L.; Pujades, L.G.; Lantada, N.; Cardona, O.D.; Marulanda, M.C. Seismic Vulnerability and Risk Evaluation Methods for Urban Areas. A Review with Application to a Pilot Area. *Struct. Infrastruct. Eng.* **2010**, *6*, 17–38. https://doi.org/10.1080/15732470802663763.
41. Choudhury, T.; Kaushik, H.B. Seismic Fragility of Open Ground Storey RC Frames with Wall Openings for Vulnerability Assessment. *Eng. Struct.* **2018**, *155*, 345–357. https://doi.org/10.1016/j.engstruct.2017.11.023.
42. Vargas Alzate, Y.F.; Pujades Beneit, L.G.; Barbat, A.H.; Hurtado Gomez, J.E.; Diaz Alvarado, S.A.; Hidalgo Leiva, D.A. Probabilistic Seismic Damage Assessment of Reinforced Concrete Buildings Considering Directionality Effects. *Struct. Infrastruct. Eng.* **2018**, *14*, 817–829. https://doi.org/10.1080/15732479.2017.1385089.
43. Pujades, L.G.; Barbat, A.H.; González-Drigo, R.; Avila, J.; Lagomarsino, S. Seismic Performance of a Block of Buildings Representative of the Typical Construction in the Eixample District in Barcelona (Spain). *Bull Earthq. Eng* **2012**, *10*, 331–349. https://doi.org/10.1007/s10518-010-9207-5.
44. Gondaliya, K.; Amin, J.; Bhaiya, V.; Vasanwala, S.; Desai, A. Generating Seismic Fragility Curves of RC Frame Building Using NSPA and IDA. *Asian J. Civ. Eng.* **2023**, *24*, 523–538. https://doi.org/10.1007/s42107-022-00516-x.
45. Priestley, M.J.N.; Seible, F.; Calvi, G.M. *Seismic Design and Retrofit of Bridges*, 1st ed.; Wiley: Hoboken, NJ, USA, 1996; ISBN 978-0-471-57998-4.

46. Computers and Structures, Inc. (CSI). *SAP2000—Structural Analysis Program—Advanced Static and Dynamic Finite Element Analysis of Structures*; Computers and Structures, Inc. (CSI): Berkeley, CA, USA, 2011.
47. Palanci, M.; Senel, S.M. Earthquake Damage Assessment of 1-Story Precast Industrial Buildings Using Damage Probability Matrices. *Bull. Earthq. Eng.* **2019**, *17*, 5241–5263. https://doi.org/10.1007/s10518-019-00660-x.