Article

Deep Learning-Based Generalized Hysteresis Model Combining Physics-informed, Physics-encoded, and Data Augmentation Methods

Jaehwan Jeon [1], Oh-Sung Kwon [2], and Junho Song [1]

[1] Department of Civil and Environmental Engineering, Seoul National University, Seoul, Republic of Korea
[2] Department of Civil and Mineral Engineering, University of Toronto, Toronto, Canada
* Correspondence: junhosong@snu.ac.kr; Tel.: +82-2-880-8397

Abstract: Deep learning-based models have recently emerged as alternatives to traditional theory-driven hysteresis models, such as the Bouc-Wen class models, to simulate hysteretic behavior of structural elements. This paper proposes a deep learning-based generalized hysteresis model by: 1) proposing an architecture to emulate the solution process of conventional theory-based models, 2) incorporating an efficient physics-informed loss function to promote non-negative energy dissipation, and 3) augmenting the training dataset by resampling hysteresis data. The proposed deep learning-based generalized model can be trained on a small amount of hysteresis data while providing accurate predictions suitable for time history analysis. Additionally, the proposed model can account for stiffness and strength degradations and pinching effects. Testing across various theory-driven hysteresis models demonstrated that the proposed deep learning-based generalized hysteresis model can effectively reproduce arbitrary hysteresis behaviors. Additionally, the model is validated against experimental hysteresis data, confirming its capability to accurately represent real-world hysteresis behavior.

Keywords: Deep learning, Physics-based deep learning, Hysteresis, Bouc-Wen class models, Physics-informed loss function, Data augmentation, Physics-encoded technique

1. Introduction

Structural systems comprise numerous elements that exhibit nonlinear and hysteretic force-displacement relationships. Such elements may exhibit strength and stiffness degradations, along with pinching effect. These complex force-displacement behavior in structural elements is the main source of errors in predicting the dynamic response of a structure subject to random excitations. It is, therefore, important to model the hysteretic behaviors accurately to predict the dynamic response of a structural system and assess risks associated with failure.

Extensive research has been carried out to model the nonlinear force-displacement hysteresis of structural elements. The pioneering works by Bouc [1] and Wen [2] laid the groundwork for what is now recognized as the Bouc-Wen class hysteresis models. Because the original Bouc-Wen model could not effectively account for the long-term degradation of strength and stiffness or the pinching effects in structural elements, Bouc-Wen-Baber-Noori (BWBN) model added terms related to the dissipated hysteretic energy [3-5]. To represent asymmetric hysteretic behavior, Song and Der Kiureghian [6] further developed Bouc-Wen models, and Kim et al. [7] developed modified BWBN model to introduce yield strength as an explicit model parameter. Another commonly adopted model for nonlinear hysteresis is Modified-Ibarra-Medina-Krawinkler (m-IMK) model [8], which models hysteresis using piecewise polynomial functions and considers strength and stiffness deterioration by controlling four cyclic deterioration modes. Other popular models include the Prandtl-Ishlinskii operator [9] which uses a play and stop operator based on the constitutive relation of plastic materials, Preisach hysteresis model [10] which were developed to predict general hysteretic behavior including

structural hysteresis and magnetic hysteresis, nonlinear hysteretic model for steel material proposed by Menegotto and Pinto [11], and Ramberg-Osgood model [12].

Recently, the efforts to accurately predict various hysteretic behaviors led to an expansion from these conventional theory-driven models to data-driven approaches. This transition is largely attributed to the inherent complexities of structural hysteresis, which are often considered difficult to fully capture through the aforementioned modeling techniques. The existing theory-driven modeling methods involve defining mathematical or prescriptive rules to represent a specific mechanism for hysteresis. While these methods allow for parameter calibration to align with experimental results, their reliance on predefined mechanisms can be a significant limitation. In many cases, the nonlinear and complex nature of hysteretic behavior remains inadequately understood, further emphasizing the need for data-driven approaches to overcome these challenges.

In the data-driven approach, hysteretic behaviors are determined by training a 'black-box' model using a large dataset. As a result, it is not necessary to start the modeling process with mathematical or prescriptive rules, which often are not discovered for complex hysteretic behaviors. Ghaboussi et al. [13] pioneered the use of neural networks for predicting the rate-independent, memoryless stress-strain behavior of structural elements, while Jung and Ghaboussi [14] later developed a model applicable to rate-dependent materials. Kim et al. [15-16] integrated a neural network model with other structural elements for structural system-level analysis. Pei and Smyth [17-18] also developed feedforward neural networks for predicting nonlinear restoring forces. However, these artificial neural networks face limitations in accurately capturing memory-dependent hysteretic force-displacement relationships.

The advancement of computational capabilities in recent years has facilitated the development of more complex and generalized deep learning architectures for modeling hysteresis, such as the Long-Short Term Memory [19] (LSTM) model [20] and the Unrolled Attention (UA) model [21]. The LSTM model predicts forces at each time step by recursively feeding back the memory of past predictions into the model, enabling it to effectively capture hysteresis behaviors that depend on memory. Similarly, the attention model retains all past memories and generates new, significant memories through linear and non-linear combinations, making it well-suited for capturing memory-dependent hysteresis as well. Building on these models, Xu et al. [22] introduced a pyramid stacked architecture, and developed pyramid stacked gated recurrent unit (GRU) [23] attention model, pyramid transformer, and pyramid LSTM model, which were tested on diverse types of hysteresis behaviors, such as the Bouc-Wen model [1] and the Menegotto-Pinto model [11].

Parallel to this progression, efforts have been made to incorporate physical insights into data-driven models for modeling hysteresis. Integrating such insights can prevent models from producing unrealistic outcomes, such as ensuring that the system continues to dissipate energy with each progression of the hysteresis loops. These physics-incoporating techniques can be categorized into three main approaches: modifying deep learning architectures, predicting parameters of theory-driven models, and incorporating physics-informed loss functions.

First, one line of research focuses on modifying deep learning architectures to embed physical insignts directly into the model structure. For example, Joghataie and Farrokh [24] developed the Prandtl Neural Network (PNN), where the activation functions of neurons were modified to emultate the Prandtl-Ishlinskii stop-and-go operator. This architecture was later extended into the Generalized Prandtl Neural Network (GPNN) [25] and the Extended Preisach Neural Network [26] to improve congruency and account for rate dependency. Similarly, Borkowski et al. [27] applied regularization techniques to the GRU model to ensure a monotonically increasing plastic work in the stress-strain relationship.

Second, another approach involves using deep learning models to predict the parameters of theory-driven hysteresis models. Horton et al. [28] developed deep learning models to predict the parameters of the m-IMK model and Oh et al. [29] developed three CNN models for rapid estimation of modified BWBN model [7] and generate optimal loading histories. Gu et al. [30] developed an LSTM-based approach that predicts the parameters of three conventional hysteresis models,

including the Bouc-Wen model and Steel02 models available in OpenSees [31]. Their approach simulates hysteresis by taking a weighted sum of the outputs from each model.

Lastly, the third approach incorporates physics-informed loss functions to enhance the training of deep learning models. Zhang et al. [32], for example, developed a loss function related to the Bouc-Wen model, and Delgado-Trujillo et al. [33] introduced a loss function that enforces general governing equations inherent in structural hysteresis, such as Drucker's postulate [34] or Il'iushin's postulate [35].

Despite these progressions, significant challenges remain in utilizing physics-based data-driven hysteresis models for dynamic analysis. Two key requirements must be addressed for such models to be practically applicable: first, the model must accurately predict forces while learning from a limited dataset. Since conducting a large number of experiments to develop these models is often impractical, models that require excessive data for training may face significant limitations in real-world engineering applications. For instance, Wang et al. [21] and Xu et al. [22] trained their models using over 4,000 data points—a scale that may not be feasible in practice. Second, the model should account for long-term degradation effects caused by accumulated damage. Existing approaches, such as those by Pei & Smith [17-18], Ghaboussi et al. [13], and even advanced models like LSTM and PNN, fail to capture these long-term effects. While attention-based deep learning models [21-22] demonstrate petential in predicting long-term degradation, they encounter significant memory constraints, as their memory requirements scale quadratically with the prediction time steps. This limitation makes them unsuitable for dynamic analysis, which demands extensive memory capacity. Therefore, the development of a robust deep learning hysteresis model remains a pressing need—one that can accurately predict hysteresis from a limited dataset and account for long-term degradation effects, ensuring its applicability for dynamic analysis in practical engineering scenarios.

This paper introduces an LSTM-based neural network architecture capable of considering long-term degradation effects induced by general damage. The proposed model is robust enough to be trained with a limited dataset of cyclic loading or random force excitations, and can be used as an element in a structural model for dynamic simulations. Specifically, we employ a physics-encoded method to modify the LSTM cell architecture, enabling it to emulate the solution process of Bouc-Wen class models while accommodating general accumulated damage. Furthermore, a physics-informed loss function is used to enforce energy dissipation, alongside a data-augmentation method that resamples structural hysteresis data to increase training dataset.

The remainder of this paper is organized as follows. In Section 2, we briefly outlline the solution process of Bouc-Wen class models, as well as the deep learning methods used in the development of our proposed model. Section 3 provides a detailed explanation of the proposed physics-based LSTM model for hysteresis prediction, including the physics-encoded, physics-informed, and data-augmentation techniques. Numerical demonstrations are provided in Section 4, where the proposed method is applied to hysteresis data generated from theory-driven hysteresis models. Training results using experimental data are discussed Section 5, followed by a summary and directions for future research in Section 6.

2. Theoretical Background

2.1. Bouc-Wen Class Models

Let the displacement at time t as $u(t)$ and the corresponding resisting force as $F(t)$. In the Bouc-Wen class models, $F(t)$ is expressed as a linear combination of $u(t)$ and the auxiliary variable $z(t)$, which represents the nonlinear inelastic behavior. This relationship is written as

$$F(t) = ak_i u(t) + (1-a)k_i z(t) \tag{1}$$

where k_i denotes the initial stiffness of the system, and a denotes the post- to pre-yield stiffness ratio.

A governing equation of $z(t)$ is a first-order nonlinear ordinary differential equation, expressed as:

$$\dot{z}(t) = \frac{dz(t)}{dt} = \dot{u}(t)[A - \beta\, sgn(\dot{u})|z(t)|^{n-1}z(t) - \gamma|z(t)|^n \quad (2)$$

where $\dot{z}(t) = dz/dt$ represents the rate of change of the auxiliary variable, $\dot{u}(t) = du/dt$ denotes the velocity. The parameters A and n control the scale and sharpness of a hysteresis loop, respectively, while β and γ define its shape. Figure 1 (a) illustrates an example of a Bouc-Wen hysteresis curve under arbitrary loading.

For the BWBN model, the pinching function $h(z, \varepsilon)$, strength function $\nu(\varepsilon)$, and the stiffness function $\eta(\varepsilon)$ are introduced to capture the effects of pinching, as well as stiffness and strength degradation. Here, $\varepsilon(t)$ represents the amount of dissipated energy in the system. These functions are incorporated into the governing equation as shown in Eq. (3).

$$\dot{z}(t) = \dot{u}(t)\frac{h(z,\varepsilon)}{\eta(\varepsilon)}\{A(\varepsilon) - \nu(\varepsilon)[\beta\, sgn(\dot{u})|z|^{n-1}z - \gamma|z|^n]\} \quad (3)$$

Since the dissipated energy is a representative measure of damage [36-38], the aforementioned functions take $\varepsilon(t)$ as an input. This dependency ensures that the overall hysteresis degradation effects are directly influenced by the dissipated energy. Figure 1 (b) illustrates an example of a BWBN hysteresis curve under arbitrary loading.

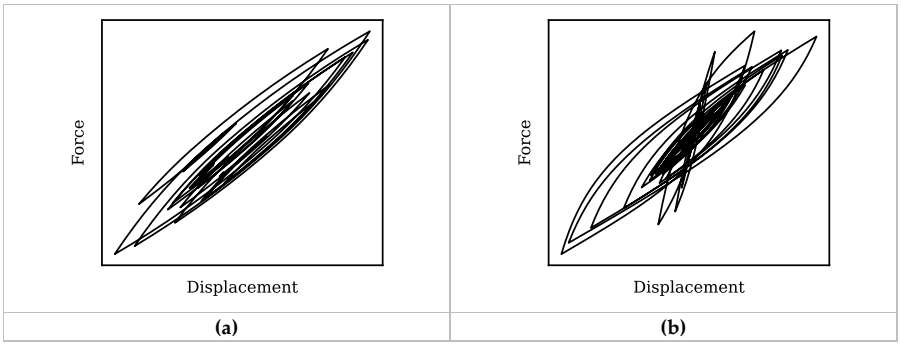


Figure 1. Examples of hysteresis curves generated by (a) the Bouc-Wen model, and (b) the Bouc-Wen-Baber-Noori model

2.2. Rules of Hysteresis

Drucker's postulate [34] states that the work done by external forces on incremental deformations should always be positive. Building on this principle, Delgado-Trujillo et al. [33] proposed a loss function based on Drucker's postulate. For an arbitrary time step t_i, where $t_i \in [0, t_n]$, the algorithm identifies two adjacent time steps, one in the forward direction (t_i^{nf}) and one in the backward direction (t_i^{pf}), by locating the nearest points along the hysteresis path with the same force value as in time t_i. Drucker's loss is then computed as:

$$\mathcal{L}_P = \sum_{i=0}^{n} \max\left(0, \varepsilon(t_i) - \varepsilon\left(t_i^{nf}\right)\right) + \max\left(0, \varepsilon\left(t_i^{pf}\right) - \varepsilon(t_i)\right). \quad (4)$$

메모 포함[OK1]: Shouldn't the index be i=0 to n?

The illustration of computing the Drucker's loss at an arbitrary time step t_i is shown in Figure 2.

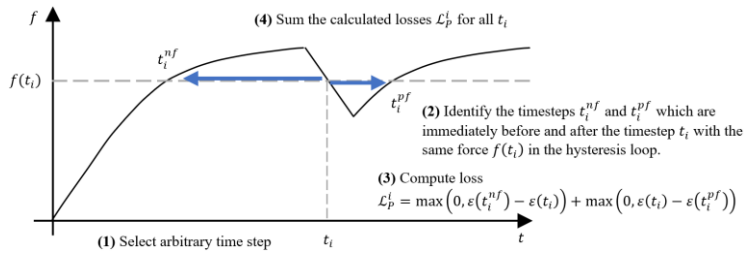


Figure 2. Illustration of computing the Drucker's loss proposed by Delgado-Trujillo et al. [33]

Delgado-Trujillo et al. [33] computed Drucker's loss by implementing a *for* loop over time, and for each time step, an additional nested *for* loop to identify the preceding and following time steps with the same force value. These identified steps were then used to calculate the energy difference. However, this approach introduces an excessive number of *for* loops, making the procedure highly inefficient and time-consuming. As a result, its application is limited to only a small number of hysteresis curves.

Many structural elements exhibit rate-independent behavior in their force-displacement relationships. Rate-independency refers to the property where the system's output response remains unchanged regardless of the speed at which the input changes. In other words, if the displacement paths are identical, the resulting force paths will also be identical. This study focuses on investigating such rate-independent structural elements, with the goal of improving the understanding and predictive accuracy of their force-displacement behavior.

2.3. Long-Short Term Memory (LSTM) Model

The Long-Short Term Memory (LSTM) model [19] is an advanced type of recurrent neural network (RNN), a class of artificial neural networks designed to process sequential data. Unlike conventional feedforward neural networks, RNNs allow data to flow back on itself, enabling the network to retain information from previous time steps.

Consider a single layer of a RNN, where the input vector at the k-th time step is defined as $\boldsymbol{x}(t_k)$. The hidden output $\boldsymbol{h}(t_k)$ of the layer then can be expressed as:

$$\boldsymbol{h}(t_k) = NN(\boldsymbol{x}(t_k), \boldsymbol{h}(t_{k-1}); \boldsymbol{w}) \tag{5}$$

where NN denotes a fully connected neural network layer, which consists of a linear transformation followed by a nonlinear activation function, and $\boldsymbol{w}$ represents the weight parameters of the neural network. As shown in Eq. (5), the weights of the LSTM model remain unchanged throughout each step of the prediction process. Figure 43 provides a basic graphical representation of RNNs.

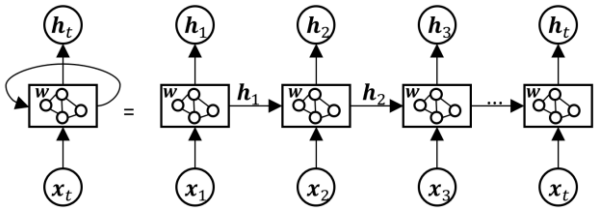


Figure 43. Graphical representation of RNNs

Traditional RNNs, which use a fully connected layer for neural networks as described in Eq. (5), face limitations in capturing long-term dependencies due to their reliance on immediate past outputs for information. This challenge, known as the "long-term dependency problem", significantly impacts hysteresis prediction, as hysteresis is influenced not only by its recent states but also by long-term states. To address this issue, LSTMs replace the typical fully connected neurons in RNNs with specialized units called LSTM cells. These cells introduce an additional long-term memory component, known as the cell state ($\boldsymbol{c}$), which is used alongside the hidden state ($\boldsymbol{h}$) and input ($\boldsymbol{x}$). The calculations for $\boldsymbol{h}(t_k)$ and $\boldsymbol{c}(t_k)$ at time t_k can be described as follows:

$$\boldsymbol{c}(t_k) = \boldsymbol{f}_k \otimes \boldsymbol{c}(t_{k-1}) + \boldsymbol{i}_k \otimes \tilde{\boldsymbol{c}}_k \tag{6a}$$

$$\boldsymbol{h}(t_k) = \boldsymbol{o}_k \otimes \tanh \boldsymbol{c}(t_k) \tag{6b}$$

where $\boldsymbol{f}_k$, $\boldsymbol{i}_k$, $\tilde{\boldsymbol{c}}_k$, and $\boldsymbol{o}_k$ denote the forget gate, input gate, candidate cell state, and output gate, respectively. These components are computed through linear operations on $\boldsymbol{h}(t_{k-1})$ and $\boldsymbol{x}(t_k)$, followed by activation functions. Specifically, $\tilde{\boldsymbol{c}}_k$ is obtained using a hyperbolic tangent activation function (tanh), while for $\boldsymbol{f}_k$, $\boldsymbol{i}_k$, and $\boldsymbol{o}_k$ use sigmoid activation functions (σ). The operator $\otimes$ represents element-wise multiplication. As shown in Eq. (6), the forget gate ($\boldsymbol{f}_k$) determines the amount of memory from the previous cell state to retain, while the input gate ($\boldsymbol{i}_k$) scales the contribution of the current candidate cell state ($\tilde{\boldsymbol{c}}_k$) to the updated cell state. The hidden state output is then computed as the element-wise multiplication of the cell state and the output gate. Figure ~~5~~4 illustrates the structure of an LSTM cell.

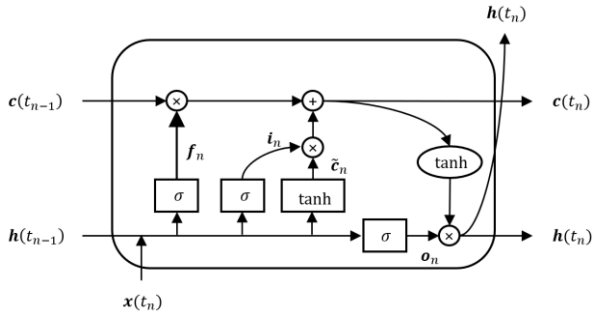


Figure ~~5~~4. Illustration of Long-Short Term Memory cell

메모 포함[OK2]: Please use identical font (bold face) that you used in the texts.

2.4. *Physics-based Methods in Deep Learning*

Faroughi et al. [39] categorized various deep learning methods that can incorporate underlying physics to accelerate the modeling of complex multiscale and multiphysics phenomena. Among these methods, the physics-encoded approach involves embedding known physical principles directly into the neural network architecture. This can be achieved by adopting the solution procedure of numerical models or physical laws, or by fixing specific weights in the neural network that are associated with the physics. For instance, as shown in Figure ~~6~~5 (a), instead of designing a neural network to independently predict each unknown response in the equation of motion (e.g., $\dot{u}_t$ and u_t), the neural network can leverage the relationship between u and $\dot{u}$. Specifically, the update equation $u_t = u_{t-1} + \dot{u}_{t-1}\Delta t$ can be encoded into the architecture, ensuring that the predictions adhere to the underlying physics.

Another widely used physics-based method in deep learning involves incorporating a physics-informed loss function alongside the conventional data-driven loss function. This approach utilizes loss functions specifically designed to constrain the model to adhere to physical principles. Defining

the data-driven loss as $\mathcal{L}_D$ and the i-th physics-informed loss as $\mathcal{L}_P^i$, the total loss function $\mathcal{L}$ for methods that incorporate physics-informed loss can be expressed as:

$$\mathcal{L} = \left(1 - \sum_{i=1}^{n_{pi}} \alpha_{pi}\right)\mathcal{L}_D + \sum_{i=1}^{n_{pi}} \alpha_{pi}\mathcal{L}_P^i. \tag{7}$$

Here, α_{pi} represents the weight coefficient that determines the extent to which the deep learning model focuses on the i-th physics-informed loss, and n_{pi} denotes the number of physics-informed loss functions. By incorporating $\mathcal{L}_P^i$, the model adjusts its output to align with physical principles. Furthermore, the physics-informed loss is designed to be differentiable with respect to the neural network's parameters ($\boldsymbol{w}$), ensuring compatibility with the backpropagation algorithm. Figure ~~6~~5 (b) illustrates the relationship between the total loss function, the physics-informed loss function, and the data-driven loss function.

Data augmentation is a machine learning used to increase the amount of training data by leveraging prior knowledge about the data, without the need to collect new data. By expanding the dataset, this approach can enhance the model's generalization ability and reduce the risk of overfitting to the original, smaller dataset. Data augmentation techniques have been widely adopted in fields such as image processing [40] and natural language processing [41]. For instance, in image classification tasks, images are augmented by being flipped, rotated, or adjusted for brightness and saturation, as shown in Figure ~~6~~5 (c). Since these augmented images are still considered equivalent to the originals, this effectively increases the dataset size. Traditionally, data augmentation was not considered a physics-based approach and was therefore not utilized as a physics-based technique. However, it can be regarded as a physics-based method when the augmented data is generated using underlying physical knowledge. In this paper, we apply this technique in Section 3 to enhance predictions for rate-independent hysteresis.

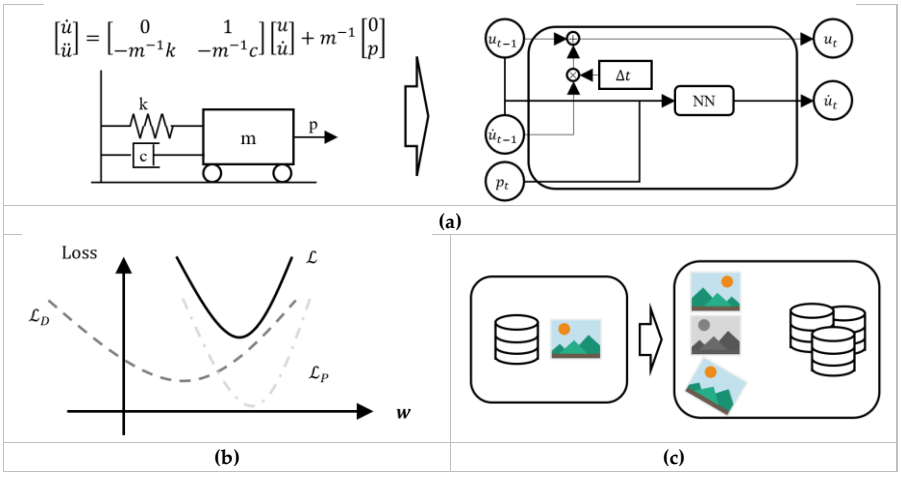


Figure ~~6~~5. Illustration of deep learning techniques which can apply knowledge of physics: (a) physics-encoded neural networks; (b) physics-informed loss function; (c) data augmentation techniques

메모 포함[OK3]: Make this figure in similar style with Figure 7.

3. Proposed Physics-based Long-Short Term Memory Model for Hysteresis Prediction

In this section, we propose an LSTM-based neural network model designed to emulate hysteresis curves by integrating the underlying laws and equations of hysteresis outlined in Sections 2.1 and 2.2, along with the deep learning techniques discussed in Sections 2.3 and 2.4.

3.1. Physics-encoded LSTM model Inspired by the Solution Process of Bouc-Wen Class Models

We encode the solution process of Bouc-Wen class models into the LSTM model. Although Bouc-Wen class models were not originally desinged to physically represent hysteresis phenomena, they have been extensively adapted to better capture such behaviors. With various modifications, these models have evolved to closely approximate many real-world hysteretic behaviors. Based on this, we treat the solution process of Bouc-Wen class models as "physics" in this study.

The encoding process consists of two key steps:

1) Replacing the auxiliary variable $z(t)$ in Eq. (1) with the hidden state $\boldsymbol{h}(t)$ in LSTM cells.

2) Incorporating long-term degradation effects by defining the accumulated energy of the hidden state as $\boldsymbol{\varepsilon_h} = \int \boldsymbol{h}(t)du$, which serves as an additional input to the deep learning model.

To elaborate on step 1), we first examine the solution process of the Bouc-Wen model.

3.1.1. Solution Process of Bouc-Wen class models

The first-order ordinary differential equations for Bouc-Wen class models, including Eqs. (2) and (3), can be reformulated in terms of dz/du, expressed as:

$$\frac{dz}{du} = \psi\big(\boldsymbol{\alpha}(u)\big) \tag{8}$$

where ψ is a nonlinear function that takes the transformation of u (i.e., $\boldsymbol{\alpha}(u)$) as its input vector. For instance, in the Bouc-Wen model, $\boldsymbol{\alpha}(u)$ can be defined as $\boldsymbol{\alpha}(u) = [u \quad \dot{u}]^T$, and $\psi\big(\boldsymbol{\alpha}(u)\big) = [A - \beta\, sgn(\dot{u})|z(u)|^{n-1}z(u) - \gamma|z(u)|^n]$. For the BWBN model, $\boldsymbol{\alpha}(u) = [u \quad \dot{u} \quad \varepsilon(u)]^T$ and $\psi\big(\boldsymbol{\alpha}(u)\big) = \frac{h(z(u),\varepsilon)}{\eta(\varepsilon)}\{A(\varepsilon) - \nu(\varepsilon)[\beta\, sgn(\dot{u})|z(u)|^{n-1}z(u) - \gamma|z(u)|^n]\}$.

Assuming $\boldsymbol{\alpha}\big(u(t_k)\big)$, $F(t_k)$ at the k-th time step (t_k) and Eqs. (1) and (8) as known, the force $F(t_{k+1})$ at the $(k+1)$-th time step can be obtained through the following two-step operation:

1. Nonlinear numerical integration: Solve the nonlinear ordinary differential equation given in Eq. (8).

$$z(t_{k+1}) = z(t_k) + \int_{u(t_k)}^{u(t_{k+1})} \psi(\boldsymbol{\alpha}(u))\, du \tag{9}$$

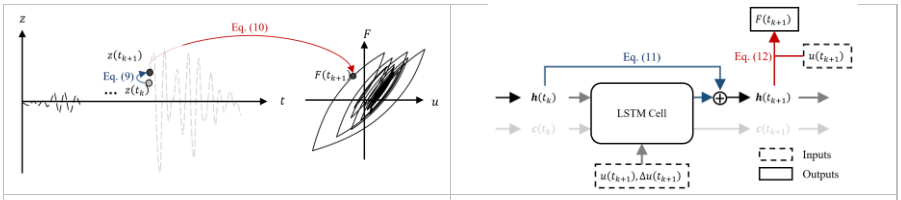
2. Linear operation: Convert to $F(t_{k+1})$ using Eqs. (1) and (9)

$$F(t_{k+1}) = ak_i u(t_{k+1}) + (1-a)k_i z(t_{k+1}) \tag{10}$$

The two-step process described above is employed in Sections 3.1.2 and 3.1.3 to develop the proposed deep learning architecture for hysteretic force prediction. The overall workflow of Eqs. (9) and (10) is illustrated in Figure ~~7~~6 (a).

3.1.2. Proposed LSTM Model Configuration emulating Bouc-Wen Class Models

By encoding the solution process of the Bouc-Wen class model into the LSTM model, various hysteresis behaviors can be more accurately represented with limited data. Consequently, the LSTM model is modified as shown in Figure ~~7~~6 (b).



(a) **(b)**

Figure 76. Graphical representation of (a) the solution process for Bouc-Wen class models and (b) the proposed physics-encoding technique in the LSTM model to emulate the solution process of Bouc-Wen class models.

First, the input of the deep learning model at time step t_{k+1}, denoted as $\boldsymbol{x}(t_{k+1})$,is defined as $\boldsymbol{x}(t_{k+1}) = [u(t_{k+1}) \quad \Delta u(t_k)]^T$, where $\Delta u(t_k) = u(t_{k+1}) - u(t_k)$. The displacement difference $\Delta u(t_k)$ is included as an input to represent the sign of $\dot{u}$, shown in Eqs. (2) and (3). Additionally, the integration step in Eq. (9) requires du, which is replaced by $\Delta u(t_k)$ for implementation.

In the proposed model, the hidden state $\boldsymbol{h}(t)$ of the LSTM operates as $z(t)$ in Eq. (9). This substitution is motivated by the observation that both $\boldsymbol{h}(t)$ in the LSTM and $z(t)$ in Bouc-Wen class models are nonobservable parameters that capture nonlinearity. To enable the proposed deep learning model to emulate the solution process of $z(t)$, we propose the following deep learning operation as a replacement for Eq. (6b).

$$\boldsymbol{h}(t_{k+1}) = \boldsymbol{h}(t_k) + \boldsymbol{o}_{k+1} \otimes \tanh \boldsymbol{c}(t_{k+1}) \tag{11}$$

Including the previous hidden state in Eq. (11) aligns the formulation with Eq. (9), where the previous auxiliary variable is incorporated into the nonlinear operation. This approach is analogous to adding a residual connection [42] in deep learning, a technique known to enhance gradient flow during training and improve network flexibility, ultimately leading to better prediction performance.

Additionally, in the computation of the output force, Bouc-Wen class models rely on both u and z, as described in Eq. (1). To emulate this process, $u(t_{k+1})$ is also used as an input to the final fully connected layer, formulated, as

$$F(t_{k+1}) = NN\big(\boldsymbol{h}(t_{k+1}), u(t_{k+1})\big). \tag{12}$$

The fully connected layer in Eq. (12) differs from typical LSTM models, where only the hidden states are used as inputs to the final fully connected layer.

3.1.3 Hidden State Energy Term to Consider Long Term Degradation Effects

In the Bouc-Wen class model discussed in Section 2.1, the dissipated energy $\varepsilon = \int F du$ is used to account for long-term degradation. Similarly, our deep learning model introduces the hidden state energy ($\boldsymbol{\varepsilon_h}$), a vector computed to capture long-term degradation effects. Unlike many other theory-driven hysteresis models that rely solely on dissipated energy to represent long-term degradation, $\boldsymbol{\varepsilon_h}$ is designed to account for general damage, inluding but not limited to dissipated energy. The hidden state energy $\boldsymbol{\varepsilon_h}(t_{k+1})$ at t_{k+1} can be computed as

$$\boldsymbol{\varepsilon_h}(t_{k+1}) = \int_{u(0)}^{u(t_{k+1})} \boldsymbol{h}(u)du = \boldsymbol{\varepsilon_h}(t_k) + \int_{u(t_k)}^{u(t_{k+1})} \boldsymbol{h}(u)du \doteq \boldsymbol{\varepsilon_h}(t_k) + \frac{\boldsymbol{h}(t_k) + \boldsymbol{h}(t_{k+1})}{2}\Delta u(t_k). \tag{13}$$

The computed $\boldsymbol{\varepsilon_h}(t_{k+1})$ is used as an additional hidden variable, alongside $\boldsymbol{h}(t_{k+1})$, in the LSTM model with input $\boldsymbol{x}(t_{k+1})$. It is important to note that $\boldsymbol{\varepsilon_h}(t_{k+1})$ is a cumulative value, meaning that it can exceed 1. If this value is directly fed into the deep learning model, it could cause an exponential increase in the model's outputs due to normalization issues. To address this, a hyperbolic tangent activation function is applied to $\boldsymbol{\varepsilon_h}(t_{k+1})$ before it is used as an input, ensuring that $\boldsymbol{\varepsilon_h} < 1$. Figure

87 illustrates the the proposed LSTM cell, which incorporates long-term degradation effects caused by general damage.

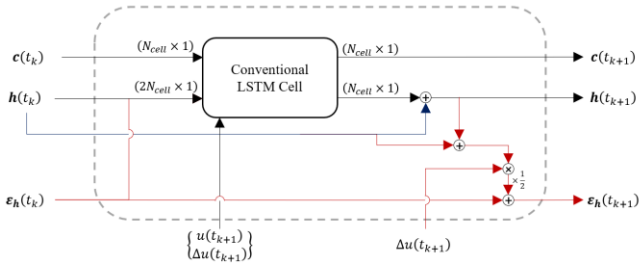


Figure 87. Proposed LSTM cell incorporating long-term degradation effects. The blue lines represent the operation described in Eq. (11), while the red lines corresponed to the operation in Eq. (13).

메모 포함[OK4]: Needs to explain why you used two different colors.

The proposed hidden state energy is designed to detect a general damage term that encompasses not only the dissipated energy but also other damage-related properties, such as maximum displacement, maximum load, and the number of cyclic loading occurrences. As an accumulative variable, the hidden state energy is particularly well-suited for capturing cumulative effects compared to conventional cell states or hidden states. This capability will be demonstrated in Section 4.3.5.

3.2. Physics-informed Loss Functions to Yield Drucker's Postulate

We implement a physics-informed loss to enforce Drucker's postulate, adopting the approach proposed by Delgado-Trujillo et al. [33]. However, as discussed in Section 2.2., the original formulation of Drucker's loss requires identifying force loops at every time step, which cannot be parallelized efficiently. This limitation makes the process computationally expensive, particularly since the loss is computed for every training sample. To address this issue, we propose an enhanced algorithm for computing the Drucker's loss. For an arbitrary force f_i, where $f_i \in [f_{min}, f_{max}]$, we identify time steps t_j that satisfy the condition $sgn\left(\left[f(t_j) - f(t_{j-1})\right] \cdot \left[f(t_{j+1}) - f(t_j)\right]\right) > 0$. For these identified time steps $t^{(j)}$, the loss is computed as:

$$\mathcal{L}_P = \sum_{t^{(i)} \in t^{(j)}} \max\left(0, \varepsilon\left(t^{(i+1)}\right) - \varepsilon\left(t^{(i)}\right)\right). \tag{14}$$

The proposed algorithm enhances computational efficiency in calculating Drucker's loss by enabling parallel computation. This allows Drucker's loss to be calculated simulatneously for multiple samples, significantly reducing the overall training time. Figure 98 illustrates the proposed enhanced algorithm for computing Drucker's loss.

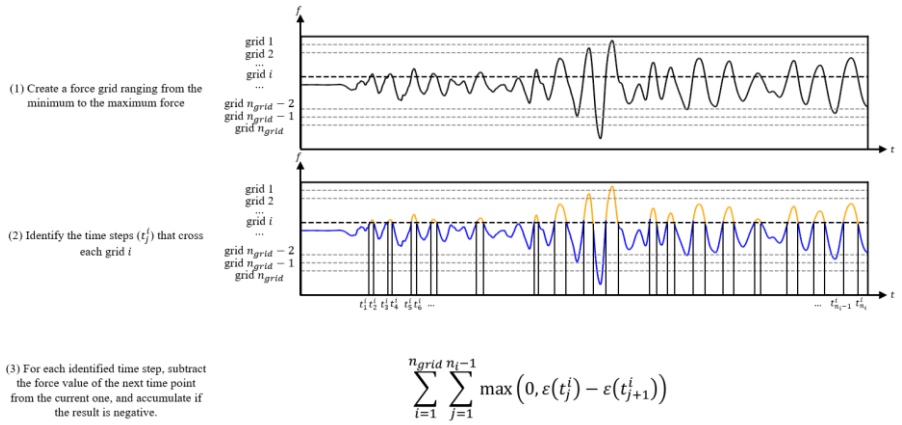


Figure ~~8~~. Algorithm of efficiently calculating the Drucker's loss

3.3. *Data Augmentation Utilizing the Rate-independency Property of the Hysteresis*

In this section, we propose a data augmentation technique for rate-independent materials. In many practical scenarios, there may not be sufficient hysteresis data available for training the deep learning hysteresis models. Due to the large number of parameters in deep learning models, this lack of data can lead to overfitting problems.

As described in Section 3.1.2, the proposed deep learning model uses displacement increments as inputs. When the same hysteresis curve is resampled at different intervals, different datasets with varying lengths and displacement increments can be generated. For instance, plotting the histogram of displacement differences yields different histograms depending on the sampling rate, as shown in Figure ~~10~~9. If the distribution of displacement increments for the hysteresis curve being predicted differs significantly from the distribution of the original training data, it may lead to poor prediction performance. Therefore, it is important to carefully consider the sampling rate to ensure that the distribution of displacement increments in the training data adequetely represents the target hysteresis behavior. By leveraging the rate-independency property described in Section 2.2, we expanded the dataset by randomly subsampling along the time axis with a sampling rate f_{aug}, as illustrated in Figure ~~11~~10. The sampling rate f_{aug} should be chosen carefully to aviod excessively distorting the actual force-displacement time history.

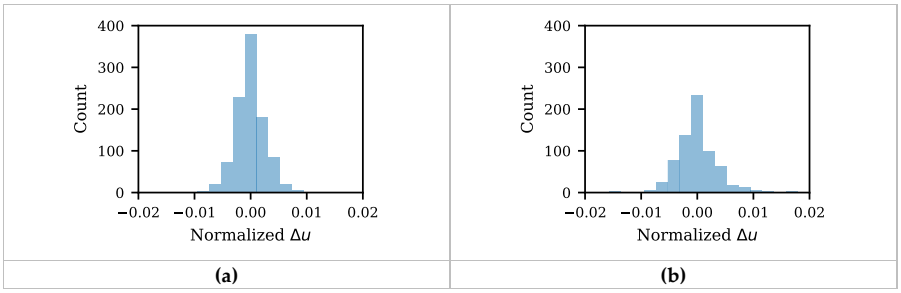


Figure ~~10~~9. Distribution of normalized displacement increment for an arbitrary hysteresis curve: (a) original sample; (b) when subsampled with $f_{aug} = 0.7$

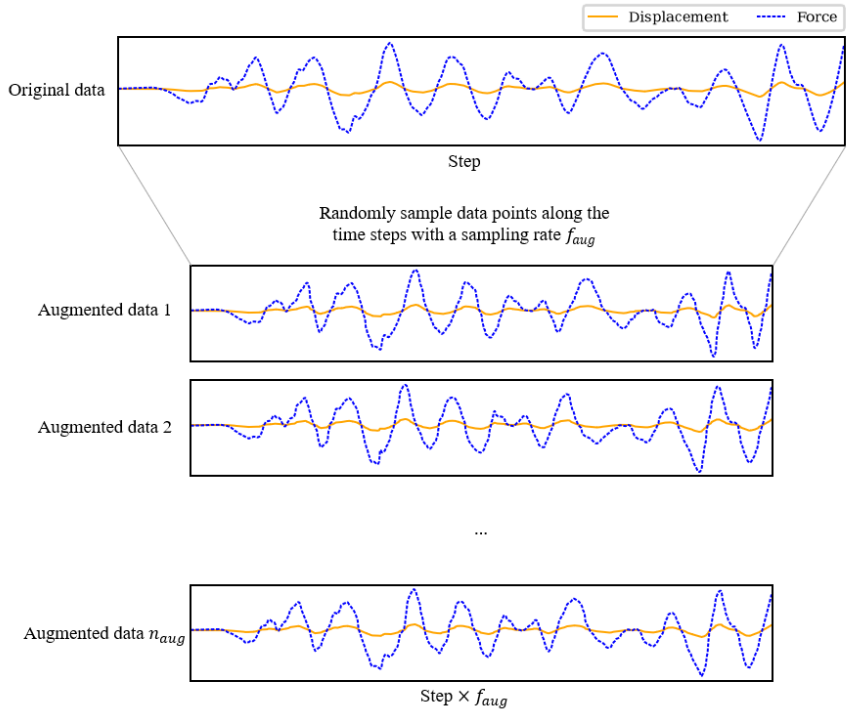


Figure ~~11~~10. Proposed data augmentation technique using the rate-independency property of hysteresis

4. Prediction of Hysteretic Behavior from Numerical Models

We utilized numerically simulated data (Section 4) and experimentally measured data (Section 5) to evaluate the capability of the proposed deep learning model in reproducing a wide range of hysteresis types under random excitations. The proposed methodology, as described in Sections 3.1, 3.2, and 3.3, is applied to demonstrate the efficiency of these approaches.

4.1. Data description

We set up a numerical example to evaluate whether the proposed deep learning model can effectively learn and predict the hysteresis behavior of a single degree of freedom system subjected to earthquake excitations (Sections 4.3.1, 4.3.2, 4.3.3, and 4.3.4) or a pre-defined displacement history (Section 4.3.5), as used in Xu et al. [22]. The description of the entire dataset and the training process of the model are illustrated in Figure 11. For earthquake excitations, we utilized 80 ground motion records, selected from a larger dataset compiled by Kim et al. [43] based on the NGA-West 2 Database [44].

To ensure consistent predictions from the deep learning model, we normalized earthquake ground motion records to have the same Peak Ground Acceleration (PGA). Since the range of possible earthquake magnitudes is virtually infinite, normalizing the input data helps the model focus on learning hysteresis behavior without being influenced by variations in ground motion intensity. Although this step is not strictly required for the model, it was adopted to make the training process more efficient. Among the 80 ground motions analyzed, results from 40 were used for training and validation, while the remaining 40 were used to evaluate the model's performance.

메모 포함[OK5]: I suggest you make a table summarizing hysteretic models (i.e. BW, BWBN, test results, etc.) and training data (i.e. earthquake response, displacement history, dynamic analysis) that you will present in Sections 4 and 5.

메모 포함[JJ6R5]: 그림으로 그리는 것이 좋을 것 같다!!!!

In Sections 4.3.1, 4.3.2, 4.3.3, and 4.3.4, we performed nonlinear time history analyses on a single-degree-of-freedom oscillator modeled using the Bouc-Wen, BWBN, bilinear, and Ramberg-Osgood models. Each system was designed with a normalized mass and a natural frequency of 1Hz. The specific parameters adopted for the target hysteresis models within OpenSeesPy are summarized in Tables 1 and 2. These numerical simulations generated hysteresis curves corresponding to each earthquake excitation. And in Section 4.3.5, we utilized the results generated by Xu et al. [22], where the Giurffré-Menegotto-Pinto model was analyzed using a predefined displacement history. Additionally, to prevent exploding gradients, which can occur when the inputs of the deep learning model exceed 1, the displacement input (u) was normalized to be bounded between -0.9 and 0.9., we used the same normalization factor for the displacement increment (Δu) as u to ensure the same unit as u.

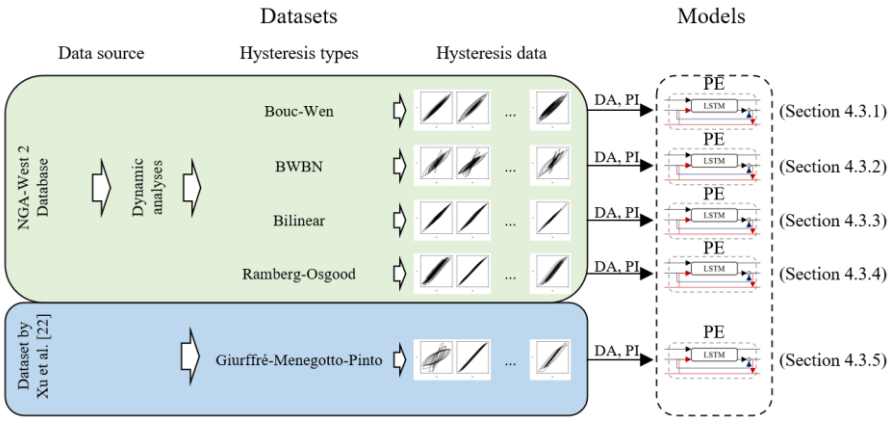


Figure 11. Illustration of the generated and collected datasets and the schematic of each model's training process.

Table 1. Parameters of the testing Bouc-Wen class hysteresis models

Type of hysteresis	α	k_0	n	γ	β	A_0	q	ζ	p	ξ	$\Delta\xi$	λ
Bouc-Wen model	0.1	30.5	1	-0.5	1.5	1	-	-	-	-	-	-
Bouc-Wen-Baber-Noori model	0.1	30.5	1	-0.5	1.5	1	0.1	0.97	1	0.2	0.002	0.1

Table 2. Parameters of the testing non Bouc-Wen class hysteresis models

Type of hysteresis	F_y	E_0	b	a	n
Bilinear model	7	30.5	0.7	-	-
Ramberg-Osgood model	7.5	30.5	-	0.002	5

4.2. Model description

We utilized a single layer of physics-encoded cells with 64 neurons, resulting in a total of 37,825 trainable parameters. While this is significantly larger than the number of model parameters in conventional hysteresis models, it remains relatively small compared to other deep learning models, such as those proposed in [21-22]. For the physics-informed loss, α_P was set to 0.2, and the losses were normalized using the initial predictions of the untrained deep learning model. For the data

메모 포함[OK7]: Please move the table to the main body.

메모 포함[OK8]: unclear

메모 포함[JJ9R8]: 이거는 Data description 에 들어가야할 것 같다

서식 있음: 2_2text_no_indent

서식 지정함: 글꼴: 굵게

서식 있음: 5_1★figure_caption, 가운데

augmentation technique, the sampling rate f_{aug} was set to 0.8. This ensured that the original force-displacement curve remained undistorted. Training was conducted over 1,000 epochs, and the model weights corresponding to the lowest validation loss were used to evaluate the test dataset.

4.3. Training and Prediction Results

서식 있음: 2_1★text

4.3.1. Bouc-Wen model

Figure ~~13~~12 illustrates the results of three randomly selected hysteresis curves from the 40 testing samples, representing prediction results for small, moderate, and large displacement inputs. The proposed model successfully captured the Bouc-Wen hysteresis behavior with high accuracy. Across all testing 40 hysteresis samples, the MSE was 2.2435×10^{-5}, and the physics-informed loss was 1.4197×10^{-7}.

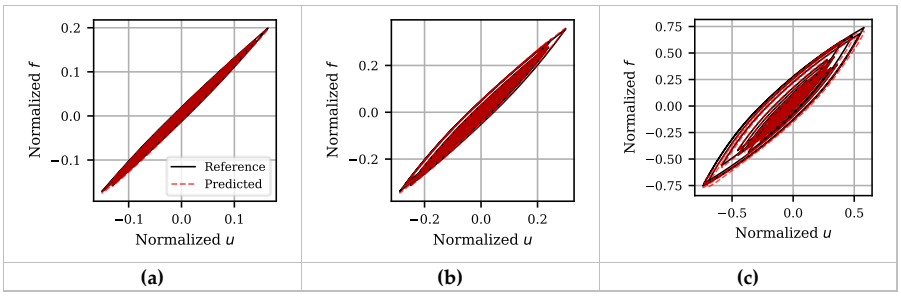


Figure ~~13~~12. Prediction results for various testing hysteresis samples of the Bouc-Wen model: prediction results for (a) small displacement; (b) moderate displacement; and (c) large displacement

In this test, we evaluated whether the model, trained solely on earthquake-induced excitations, could generalize to predict hysteresis not only for the testing earthquake-induced excitations but also for cyclic loadings. To this end, cyclic displacement loading with various amplitudes was applied to the deep learning model, as illustrated in Figure ~~14~~13, and the results were compared to those of the reference Bouc-Wen model. Figure ~~15~~14 presents the prediction results for cyclic loading. Overall, the curves align well with the reference. However, for cyclic loading with a maximum normalized amplitude of 1, the predicted hysteresis curve appears slightly thinner than the reference curve (Figure ~~15~~14 (d)). This discrepancy is likely due to the limited training data available for such large amplitudes, resulting in slightly less accurate predictions in this case.

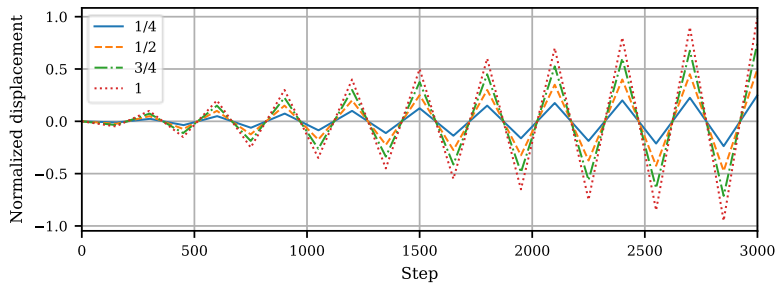


Figure ~~14~~13. Cyclic loading displacement history for hysteresis generation

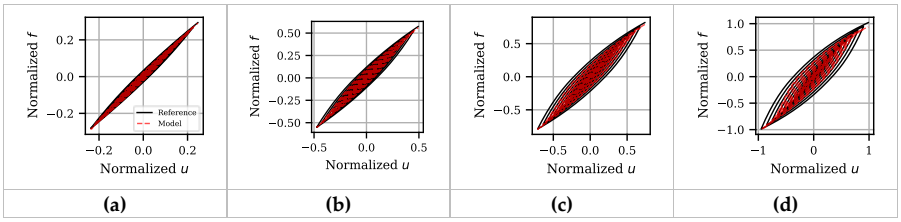


Figure ~~15~~14. Hysteresis prediction results under the cyclic loading shown in Figure ~~14~~13

서식 지정함: (한글) 한국어

4.3.2. BWBN model

The Bouc-Wen model presented in Section 4.3.1 does not account for stiffness or strength degradation. To test whether the proposed LSTM model could be trained on hysteretic behaviors exhibiting these degradation effects, we used the BWBN model as a reference for generating target hysteresis data.

The proposed LSTM model was trained and testing following the same procedure described in Sections 3.1, 3.2, and 3.3. Compared to a conventional LSTM model with a single layer with LSTM cells followed by a fully connected layer, our physics-informed technique achieved a 10.3% reduction in MSE, while the physics-encoded technique resulted in a 94.8% reduction in MSE. When physics-encoded, physics-informed, and data augmentation techniques were combined, the proposed model demonstrated a 95.3% reduction in MSE. The physics-encoded methodology was particularly effective in this case, as the BWBN model incorporates long-term degradation effects. However, the best prediction performance was achieved when all proposed techniques were applied simultaneously.

메모 포함[OK10]: Confusing wording. Do you men single fully-connected layer?

For performance comparison, we also tested the Pyramid-stacked LSTM model (PyLSTM) proposed by Xu et al. [22] as an additional baseline model alongside with the conventional LSTM model. As a recent state-of-the-art approach with publicly available source code, PyLSTM served as a suitable benchmark for our study. We implemented the PyLSTM model with two layers and a fully connected layer. The Attention-based models proposed by Xu et al. [22] were excluded from this comparison due to their significant memory requirements as the time steps progress, which led to memory overflow issues that prevented practical use in our experiments and pose challenges for dynamic analysis.

Figure ~~16~~15 shows the prediction results for a randomly selected test hysteresis curve. As illustrated, the proposed model effectively captured stiffness degradation (Figures ~~16~~15 (a), (b), and (c)), whereas the conventional LSTM model (Figures ~~16~~15 (d), (e) and (f)) and the PyLSTM model (Figure ~~16~~15 (g), (h), and (i)) failed to do so. Figure ~~17~~16 presents a scatter plot of the reference maximum peak force from 40 testing hysteresis curves and the predicted maximum peak force. Consistent with the results in Figure ~~16~~15, the proposed model achieved the highest R^2 value (Figure ~~17~~16 (a)), indicating superior prediction performance. In constrast, the PyLSTM model (Figure ~~17~~16 (c)) had the lowest R^2 value, even lower than that of the standard LSTM model. This performance drop is likely due to the PyLSTM model's weighted stacking approach, which requires at least two layers. Consequently, the PyLSTM model has more layers compared to other single-layer models, leading to an excessive number of parameters that may have caused overfitting.

메모 포함[OK11]: Check figure labels.

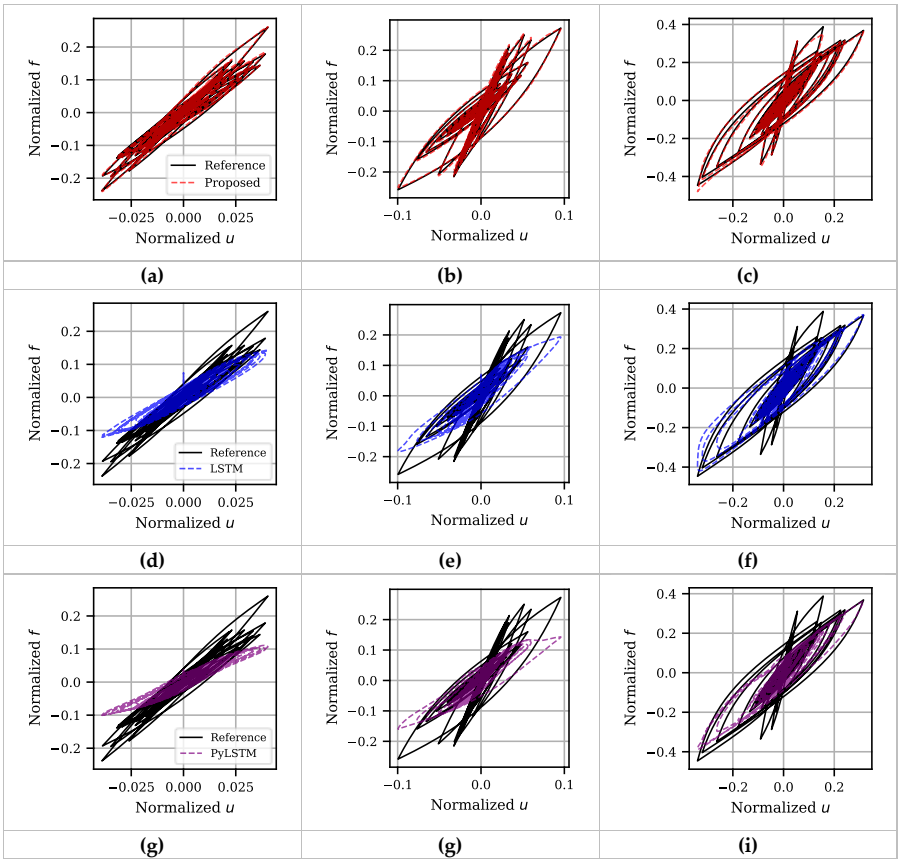


Figure 1615. Prediction results for an arbitrary testing BWBN hysteresis curves: proposed LSTM model prediction results for (a) small displacement, (b) moderate displacement, and (c) large displacement; LSTM prediction results for (d) small displacement, (e) moderate displacement, and (f) large displacement; and PyLSTM model prediction results for (g) small displacement, (h) moderate displacement, and (i) large displacement

메모 포함[전12]: Abc 를 우리꺼로 하자

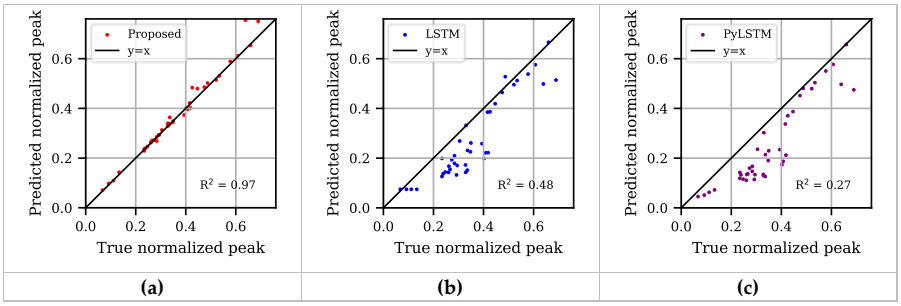


Figure 1716. Prediction results for peak values across 40 testing hysteresis curves: (a) the proposed LSTM model, (b) LSTM model, and (c) PyLSTM model.

4.3.3. Bilinear Model

We tested whether the proposed model could be trained on and accurately predict bilinear hysteretic behavior. While the Bouc-Wen model introduced in Section 4.3.1 theoretically represents bilinear behavior as n approaches infinity, it often struggles to accurately capture bilinear hysteresis in practice. In constrast, the proposed deep learning model is designed to generalize beyond such limitations. To evaluate this, we tested whether the proposed model could successfully predict bilinear hysteretic behavior, even when it emulates the solution process of the Bouc-Wen class models.

Figure ~~18~~17 shows the force-displacement prediction results for a randomly selected earthquake-induced testing hysteresis. The results demonstrate that, while the proposed model emulates the solution process of the Bouc-Wen model, it successfully overcomes the practical limitations of the Bouc-Wen model in capturing bilinear hysteresis. This highlights the model's ability to predict hysteresis across a broad range of behaviors, including cases where the Bouc-Wen model struggles to perform accurately. Additionally, Figure ~~19~~18 presents the prediction results for the testing dataset as a time history. The predictions closely align with the reference data, further validating the robustness of the proposed model in handling bilinear hysteresis.

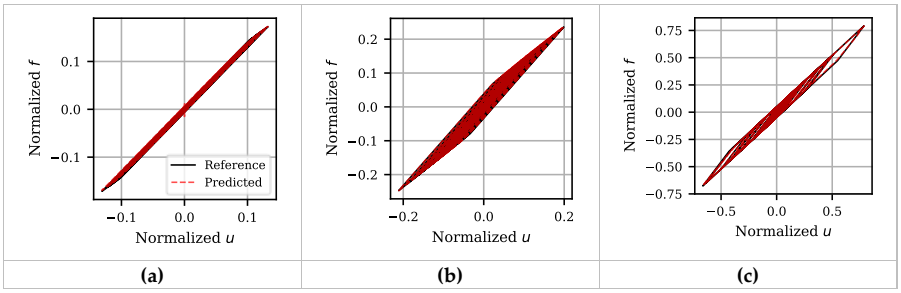
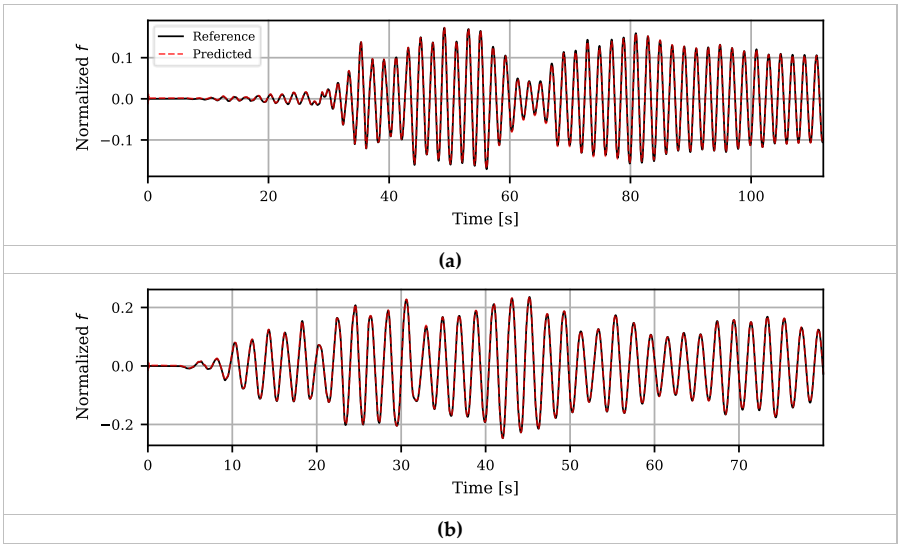
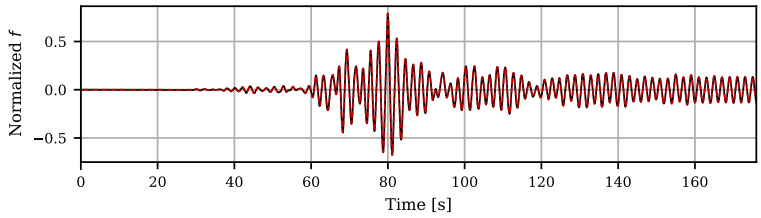


Figure ~~18~~17. Prediction results for various testing hysteresis samples of the bilinear model: prediction results for (a) small displacement; (b) moderate displacement; and (c) large displacement





(c)

Figure ~~19~~18. Predicted force time series and reference force time series for the testing hysteresis samples in Figure ~~18~~17: prediction results for (a) small displacement; (b) moderate displacement; and (c) large displacement

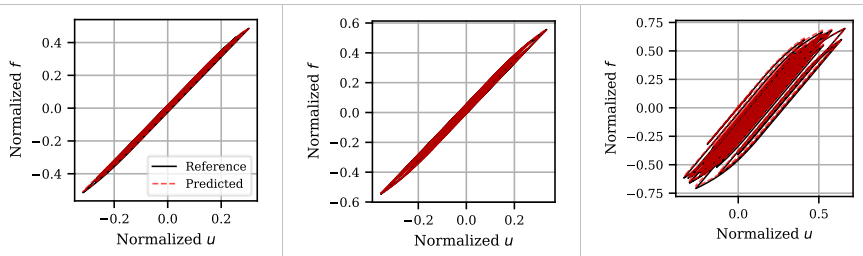
4.3.4. Ramberg-Osgood model

The Ramberg-Osgood model, unlike bilinear hysteresis, represents a class of hysteresis that cannot be captured by the Bouc-Wen model. To evaluate the versatility of the proposed model, we tested its ability to predict hysteresis behaviors fundamentally different from those representable by the Bouc-Wen model.

Figure ~~20~~19 shows the prediction results for the force-displacement relationship of a randomly selected testing hysteresis dataset. The results demonstrate that the proposed model can accurately learn and predict hysteresis behaviors even for models that cannot be represented by the Bouc-Wen model.

To further validate the model's applicability, we conducted additional testing by applying the learned hysteresis model in a dynamic analysis. The structural properties used for the dynamic analysis are the same as those described in Section 4.1. Figure ~~21~~20 presents the results of the dynamic analysis using the proposed deep learning hysteresis model under testing earthquake excitations. The results indicate that the proposed model is robust enough to accurately represent hysteresis behavior in dynamic analysis.

메모 포함[OK13]: Needs to explain more about the structure: period, mass, damping ratio, etc. for three different cases in the figure.



(a) (b) (c)

Figure ~~20~~19. Prediction results for various testing hysteresis samples of the Ramberg-Osgood model: prediction results for (a) small displacement; (b) moderate displacement; and (c) large displacement

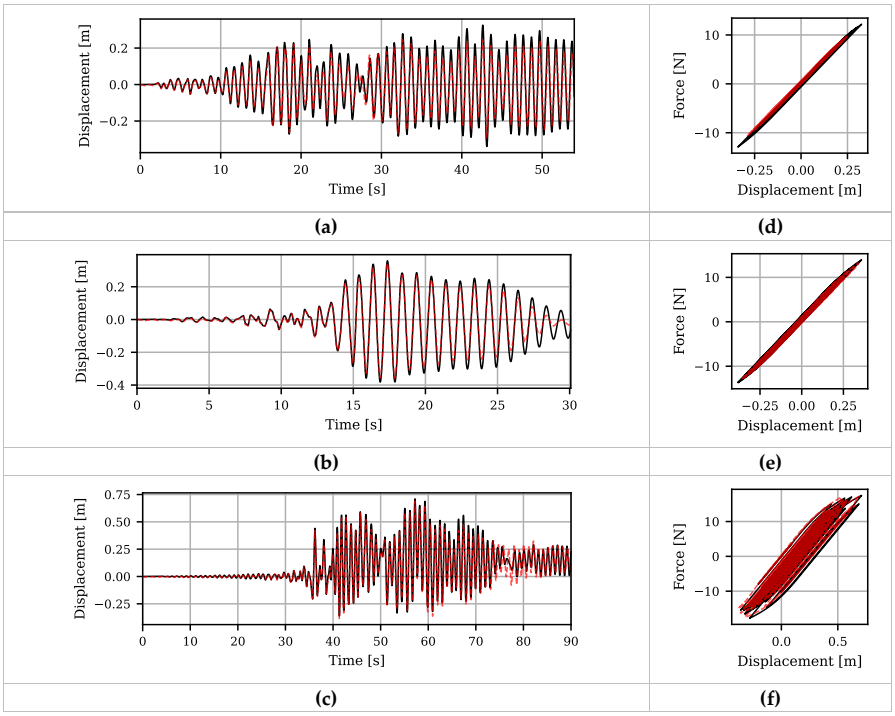


Figure ~~21~~20. Dynamic response prediction results subject to testing earthquakes in Figure ~~20~~19: displacement time histories for (a) small, (b) moderate, and (c) large displacement, and the corresponding hysteresis curve for (d) small, (e) moderate, and (f) large displacement

메모 포함[OK14]: At the right side of each figure, add force-deformation relationship.

4.3.5. Giurffré-Menegotto-Pinto model

The Giuffré-Menegotto-Pinto model represents the stress-strain relationship of steel materials, effectively capturing their nonlinear behavior and hysteresis characteristics. In this model, isotropic hardening is linked to maximum stress and maximum strain, rather than hysteretic energy in the BWBN model. To evaluate whether the proposed model can capture degradation effects caused not only by hysteretic energy but also by general accumulated damage, we conducted a prediction test using this model. The data for this experiment were sourced from the study by Xu et al. [22], with the number of hysteresis datasets reduced to one-fourth. This reduction was implemented to demonstrate that the model can be trained effectively with a smaller dataset.

Figure ~~22~~21 presents a portion of the testing hysteresis data from Xu et al. [22] alongside the corresponding prediction results from the proposed model. Overall, the proposed model performed well in predicting hysteresis behavior. However, in some cases, the predictions were less accurate, likely due to data loss during the sampling process used to align the data into sequences of consistent length.

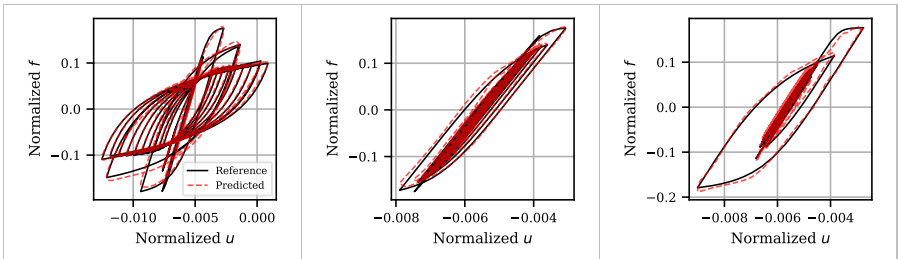


Figure ~~22~~21. Prediction results for various testing hysteresis samples of the Giurffré-Menegotto-Pinto model

5. Prediction of Hysteretic Behavior from Physical Experiments

5.1. Data description

We used a cast steel replaceable modular yielding link, designed by Mortazavi et al. [45-47], to function as a yielding fuse in eccentrically braced frames (EBFs). The configuration of the modular yielding link is shown in Figure @LICON(22). For training data, we utilized force-displacement histories from the first-floor yielding link in three different hybrid simulations of a four-story office building, referred to as HS-1, HS-2, and HS-3 [45]. Additionally, a single cyclic loading history of the yielding link was used as testing data [46-47]. Figures ~~24~~23 and ~~25~~24 illustrate the force-displacement relationships obtained from the three hybrid simulation results and the single cyclic loading result, respectively.

However, the data shown in Figure ~~24~~23 was insufficient for training the deep learning model, necessitating an increase in the dataset size to achieve accurate predictions. To address this, we used the Bouc-Wen model to generate additional data. First, we Bouc-Wen model parameters for the dataset were identified using the genetic algorithm developed by Oh et al. [48]. The orange line in Figure ~~24~~23 represents the result of fitting the dataset using the Bouc-Wen model. Based on these identified parameters, an auxiliary training dataset was generated using 40 random earthquake histories from Section 4. The model was initially trained on the auxiliary data generated by the Bouc-Wen model and subsequently trained on the actual training data. We believe that with a larger amount of training data, this intermediate step may not have been unnecessary.

메모 포함[전15]: Pedram 에게 요청하는 것이 좋을 것 같다

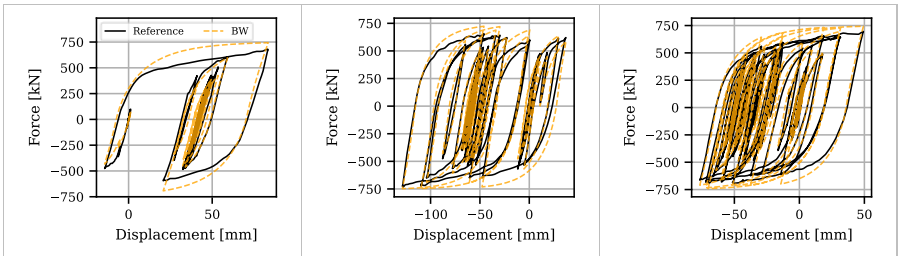


Figure ~~24~~23. Hysteresis results of the yielding link and fitted curves using the Bouc-Wen model with optimized parameters for (a) hybrid simulation test 1 (HS-1); (b) hybrid simulation test 2 (HS-2); and (c) hybrid simulation test 3 (HS-3)

5.2. Testing result

Figure ~~25~~ 24 shows the prediction results for the testing cyclic hysteresis of a yielding link. The experimental results demonstrate that the proposed model fits the Link dataset more accurately than the Bouc-Wen model. Specifically, the loss of the proposed model on the testing hysteresis curve was reduced by 48% compared to that of the Bouc-Wen model. A slight wiggling pattern is observed in the upper right corner of Figure ~~25~~24, which is likely to diminish with the availablility of additional training data.

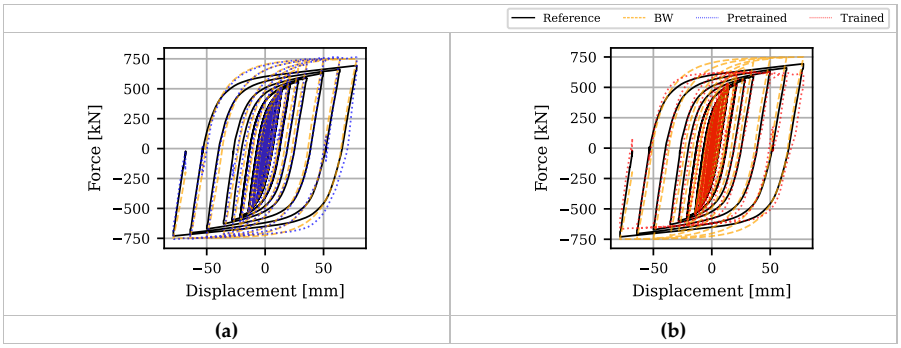


Figure ~~25~~24. Hysteresis prediction results for testing cyclic loading: (a) before training with actual data (fitted to the Bouc-Wen model); (b) the model pretrained on the Bouc-Wen model, then trained with actual data

6. Conclusions

In this study, we developed a deep learning model capable of being trained on general hysteretic behaviors. First, we designed a model structure that emulates the solution process of the Bouc-Wen model. Second, we introduced a modified physics-informed loss that satisfies Drucker's postulate which is adapted for parallel computation. Finally, we proposed a data augmentation technique applicable to rate-independent materials. The proposed methodology enables the deep learning model to be trained with a small volume of data while considering degradation effects and remaining robust enough for use in dynamic analysis.

The developed deep learning model was applied to various reference hysteresis models, including the Bouc-Wen, BWBN, bilinear, Ramberg-Osgood, and Giuffré-Menegotto-Pinto models. We confirmed that the model accurately predicts hysteretic behavior, even in cases where the hysteretic behavior cannot be well-represented by the Bouc-Wen model—such as when long-term degradation effects are not directly linked to hysteretic energy. Furthermore, by applying the model to experimentally obtained hysteresis data, we demonstrated that it can accurately produce test results.

Although the proposed model successfully captured various theory-based hysteresis behaviors, applying it to a more complex rule-based models may remain challenging. Developing a more general model capable of handling such complicated rule-based hysteresis models is an area for future research. Additionally, future studies will focus on extending the application of the model to the force-displacement relationships of multi-degree-of-freedom structural systems.

Another important direction is enabling the proposed model to output predictions as a predictive distribution. This would allow engineers to evaluate both the accuracy of the predictions and the credibility of the deep learning model. Ongoing research aims to leverage Bayesian Neural Networks to quantify the predictive uncertainty of the deep learning hysteresis model.

Acknowledgment: This research is funded by the project "Development of Life-cycle Engineering Technique and Construction Method for Global Competitiveness Upgrade of Cable Bridges" of the Ministry of Land, Infrastructure and Transport (MOLIT) of the Korean Government, grant number 17SCIP-B119964-02.

618 **Conflicts of Interest:** The authors declare no conflict of interest.
619

7. References

1. Bouc, R. (1967). Forced vibrations of mechanical systems with hysteresis. *Proc. of the Fourth Conference on Nonlinear Oscillations, Prague, 1967*.
2. Wen, Y.-K. (1976). Method for random vibration of hysteretic systems. *Journal of the Engineering Mechanics Division*, *102*(2), 249–263.
3. Baber, T. T., & Wen, Y.-K. (1981). Random vibration of hysteretic, degrading systems. *Journal of the Engineering Mechanics Division*, *107*(6), 1069–1087.
4. Baber, T. T., & Noori, M. N. (1985). Random vibration of degrading, pinching systems. *Journal of Engineering Mechanics*, *111*(8), 1010–1026.
5. Noori, M., Choi, J.-D., & Davoodl, H. (1986). Zero and nonzero mean random vibration analysis of a new general hysteresis model. *Probabilistic Engineering Mechanics*, *1*(4), 192–201.
6. Song, J., & der Kiureghian, A. (2006). Generalized Bouc–Wen model for highly asymmetric hysteresis. *Journal of Engineering Mechanics*, *132*(6), 610–618.
7. Kim, T., Kwon, O., & Song, J. (2023). Deep learning based seismic response prediction of hysteretic systems having degradation and pinching. *Earthquake Engineering & Structural Dynamics*, *52*(8), 2384–2406.
8. Ibarra, L. F., Medina, R. A., & Krawinkler, H. (2005). Hysteretic models that incorporate strength and stiffness deterioration. *Earthquake Engineering & Structural Dynamics*, *34*(12), 1489–1511.
9. Visintin, A., & Visintin, A. (1994). Plays, Stops and Prandtl-Ishlinskiĭ Models. *Differential Models of Hysteresis*, 59–96.
10. Preisach, F. (1935). Über die magnetische Nachwirkung. *Zeitschrift Für Physik*, *94*(5), 277–302. https://doi.org/10.1007/BF01349418
11. Menegotto, M., & Pinto, P. E. (1973). Method of analysis for cyclically loaded reinforced concrete plane frames including changes in geometry and non-elastic of elements under combined normal force and bending. *Proceedings, IABSE Symposium*, 20–102.
12. Ramberg, W. (1943). Description of Stress-Strain Curves by Three Parameters.
13. Ghaboussi, J., Pecknold, D. A., Zhang, M., & Haj-Ali, R. M. (1998). Autoprogressive training of neural network constitutive models. *International Journal for Numerical Methods in Engineering*, *42*(1), 105–126.
14. Jung, S., & Ghaboussi, J. (2006). Neural network constitutive model for rate-dependent materials. *Computers & Structures*, *84*(15–16), 955–963.
15. Kim, J., Ghaboussi, J., & Elnashai, A. S. (2010). Mechanical and informational modeling of steel beam-to-column connections. *Engineering Structures*, *32*(2), 449–458.
16. Kim, J., Ghaboussi, J., & Elnashai, A. S. (2012). Hysteretic mechanical–informational modeling of bolted steel frame connections. *Engineering Structures*, *45*, 1–11.
17. Pei, J.-S., & Smyth, A. W. (2006). New approach to designing multilayer feedforward neural network architecture for modeling nonlinear restoring forces. I: Formulation. *Journal of Engineering Mechanics*, *132*(12), 1290–1300.
18. Pei, J.-S., & Smyth, A. W. (2006). New approach to designing multilayer feedforward neural network architecture for modeling nonlinear restoring forces. II: Applications. *Journal of Engineering Mechanics*, *132*(12), 1301–1312.
19. Hochreiter, S. (1997). Long Short-term Memory. *Neural Computation MIT-Press*.
20. Ni, X., Xiong, Q., Kong, Q., & Yuan, C. (2022). Deep HystereticNet to predict hysteretic performance of RC columns against cyclic loading. *Engineering Structures*, *273*, 115103.
21. Wang, C., Xu, L., & Fan, J. (2020). A general deep learning framework for history-dependent response prediction based on UA-Seq2Seq model. *Computer Methods in Applied Mechanics and Engineering*, *372*, 113357.
22. Xu, Y., Lu, X., Fei, Y., & Huang, Y. (2023). Hysteretic behavior simulation based on pyramid neural network: Principle, network architecture, case study and explanation. *Advances in Structural Engineering*, *26*(13), 2359–2374.
23. Chung, J. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *ArXiv Preprint ArXiv:1412.3555*.
24. Joghataie, A., & Farrokh, M. (2008). Dynamic analysis of nonlinear frames by Prandtl neural networks. *Journal of Engineering Mechanics*, *134*(11), 961–969.
25. Farrokh, M., Dizaji, M. S., & Joghataie, A. (2015). Modeling hysteretic deteriorating behavior using generalized Prandtl neural network. *Journal of Engineering Mechanics*, *141*(8), 04015024.

26. Farrokh, M., Dizaji, F. S., & Dizaji, M. S. (2022). Hysteresis identification using extended preisach neural network. *Neural Processing Letters*, 1–25.
27. Borkowski, L., Sorini, C., & Chattopadhyay, A. (2022). Recurrent neural network-based multiaxial plasticity model with regularization for physics-informed constraints. *Computers & Structures*, *258*, 106678.
28. Horton, T. A., Hajirasouliha, I., Davison, B., & Ozdemir, Z. (2021). Accurate prediction of cyclic hysteresis behaviour of RBS connections using deep learning neural networks. *Engineering Structures*, *247*, 113156.
29. Oh, S., Song, J., & Kim, T. (2024). Deep learning-based modularized loading protocol for parameter estimation of Bouc-Wen class models. *ArXiv Preprint ArXiv:2411.02776*.
30. Gu, Y., Lu, X., & Xu, Y. (2023). A deep ensemble learning-driven method for the intelligent construction of structural hysteresis models. *Computers & Structures*, *286*, 107106.
31. McKenna, F., & Fenves, G. L. (2001). The OpenSees command language manual. *University of California, Berkeley (Opensees. Ce. Berkeley. Edu)*.
32. Zhang, R., Liu, Y., & Sun, H. (2020). Physics-informed multi-LSTM networks for metamodeling of nonlinear structures. *Computer Methods in Applied Mechanics and Engineering*, *369*, 113226.
33. Delgado-Trujillo, S., Alvarez, D. A., & Bedoya-Ruíz, D. (2023). Hysteresis modeling of structural systems using physics-guided universal ordinary differential equations. *Computers & Structures*, *280*, 106988.
34. Drucker, D. C. (1950). Some implications of work hardening and ideal plasticity. *Quarterly of Applied Mathematics*, *7*(4), 411–418.
35. Il'Iushin, A. A. (1961). On the postulate of plasticity. *Journal of Applied Mathematics and Mechanics*, *25*(3), 746–752.
36. Kunnath, S. K., & Chai, Y. H. (2004). Cumulative damage-based inelastic cyclic demand spectrum. *Earthquake Engineering & Structural Dynamics*, *33*(4), 499–520.
37. Benavent-Climent, A. (2007). An energy-based damage model for seismic response of steel structures. *Earthquake Engineering & Structural Dynamics*, *36*(8), 1049–1064.
38. Hernandez, E. M., & May, G. (2013). Dissipated energy ratio as a feature for earthquake-induced damage detection of instrumented structures. *Journal of Engineering Mechanics*, *139*(11), 1521–1529.
39. Faroughi, S. A., Pawar, N., Fernandes, C., Raissi, M., Das, S., Kalantari, N. K., & Mahjour, S. K. (2022). Physics-guided, physics-informed, and physics-encoded neural networks in scientific computing. *ArXiv Preprint ArXiv:2211.07377*.
40. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C. A., Cubuk, E. D., Kurakin, A., & Li, C.-L. (2020). Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in Neural Information Processing Systems*, *33*, 596–608.
41. Bari, M. S., Mohiuddin, T., & Joty, S. (2020). Multimix: A robust data augmentation framework for cross-lingual nlp. *ArXiv Preprint ArXiv:2004.13240*.
42. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.
43. Kim, T., Kwon, O.-S., & Song, J. (2019). Response prediction of nonlinear hysteretic systems by deep neural networks. *Neural Networks*, *111*, 1–10.
44. Power, M., Chiou, B., Abrahamson, N., Bozorgnia, Y., Shantz, T., & Roblee, C. (2008). An overview of the NGA project. *Earthquake Spectra*, *24*(1), 3–21.
45. Mortazavi, P., Kwon, O., & Christopoulos, C. (2023). Pseudo-dynamic hybrid simulations of steel eccentrically braced frames equipped with cast steel replaceable modular yielding links. *Earthquake Engineering & Structural Dynamics*, *52*(12), 3622–3648.
46. Mortazavi, P., Lee, E., Binder, J., Kwon, O.-S., & Christopoulos, C. (2023). Large-scale experimental validation of optimized cast steel replaceable modular yielding links for eccentrically braced frames. *Journal of Structural Engineering*, *149*(7), 04023071.
47. Mortazavi, P., Binder, J., Kwon, O.-S., & Christopoulos, C. (2023). Ductility-targeted design of cast steel replaceable modular yielding links and their experimental validation through large-scale testing. *Journal of Structural Engineering*, *149*(7), 04023080.
48. Oh, S., Kim, T., & Song, J. (2023). Bouc–Wen class models considering hysteresis mechanism of RC columns in nonlinear dynamic analysis. *International Journal of Non-Linear Mechanics*, *148*, 104263.