An Attention-Augmented Long Short-Term Memory Network for PINN-based Structural Health Monitoring

Nikhil Mahar, Subhamoy Sen, Laurent Mevel

An Attention-Augmented Long Short-Term Memory Network for PINN-based Structural Health Monitoring

Nikhil Mahar[a], Subhamoy Sen[a,1], Laurent Mevel[b]

[a] *i4S Laboratory, Indian Institute of Technology Mandi, Mandi, HP, India*
[b] *Univ. Gustave Eiffel, Inria, Cosys-SII, I4S, Campus de Beaulieu, Rennes, France*

Abstract

Structural Health Monitoring (SHM) plays a vital role in maintaining the safety and functionality of structural and mechanical elements in different engineering sectors. Conventional model-based SHM techniques often struggle with the intricacies of complex systems due to oversimplified models. In contrast, data-driven, model-free methods provide speed and scalability but often fail to incorporate system physics, which leads to challenges in generalizing. Physics-informed neural networks (PINNs) have emerged as promising solutions by merging data-driven models with physical insights. Nevertheless, the majority of PINN applications in SHM concentrate on estimating dynamic system states, while inverse estimation remains less explored. Most existing inverse methods demand complete state measurements, which are rarely achievable in practical scenarios. This research presents a novel PINN framework—a physics informed attention-augmented Long Short-Term Memory network (Pi-Attn-LSTM)—designed for estimating parameters in structural systems without the need for comprehensive state data, thereby enhancing its feasibility. The model utilizes a temporal attention mechanism in an encoder-decoder setup to adaptively focus on critical features in sequential data, enhancing the accuracy of parameter estimation in dynamic environments. Validation is executed using both numerical simulations of multi-degree-of-freedom systems, including an IASC-ASCE-Benchmark problem, and experimental tests on data from a scaled aluminium frame. Moreover, comparative studies between the proposed Pi-Attn-LSTM and a typical Pi-LSTM model demonstrate the superior robustness and accuracy of Pi-Attn-LSTM. The extensive validation study shows that Pi-Attn-LSTM not only speeds up convergence but also reliably identifies localized structural degradation, highlighting its potential as a viable tool for SHM in real-world scenarios.

Keywords: *Parameter estimation, Structural system, Physics informed LSTM, Attention mechanism, Physics informed attention-augmented LSTM*

*Corresponding author; *E-mail address: subhamoy@iitmandi.ac.in*

1. Introduction

In structural engineering, the pursuit of accurately modeling and forecasting structural behavior generally falls into two broad categories: data-driven and physics-driven strategies. Data-driven methods involve statistical techniques used in both the time [1–4] and frequency domains [5–7], as well as latent space techniques like blind source separation (BSS) [8–11], which mostly depend on empirical data for developing models. Conversely, physics-driven methods presume a predetermined model structure, often expressed through parameterized models or governing differential equations, refined via optimization or Bayesian approaches [12–17]. While parameter estimation based on models, such as those employing Bayesian filtering algorithms [18–20], take advantage of inherent physics, model-independent approaches [8–11] typically depend on fault-sensitive statistical distances. Nevertheless, these methods generally overlook spatial correlations informed by physical understanding. Although support models can theoretically enhance precision, they can also introduce serious computational difficulties and potentially compromise robustness due to the sensitivity of statistical distances to model inaccuracies. However, recent advancements in machine learning (ML) and artificial intelligence (AI) algorithms provide encouraging solutions to these difficulties.

In the domain of vibration-based SHM and system identification, AI/ML have captured significant attention over the last decade [21, 22]. Artificial neural networks (ANNs) have been effectively utilized to determine various structural parameters, including mass, stiffness, damping coefficients, eigenvalues, and eigenmodes, by analyzing vibration and acceleration data [23, 24]. Recently, Wen et al. [25] effectively addressed the challenge of modeling high-dimensional spatiotemporal responses by integrating Graph Neural Networks (GNNs) with Temporal Convolutional Networks (TCNs). This approach leverages graph representations to capture spatial dependencies and enhances surrogate modeling performance compared to conventional methods such as ANNs and LSTM networks. By embedding spatial structure within the network architecture, the model improves accuracy and reduces data requirements relative to traditional black-box approaches. Similarly, Xu et al.[26] proposed a data-driven spatiotemporal model predictive control (MPC) framework to capture nonlinear distributed parameter system (DPS) dynamics. It integrates Kernel Principal Component Analysis (KPCA) and Least Squares Support Vector Machines (LS-SVM) to model complex nonlinear behavior across space and time. This extends earlier efforts [27–29], which primarily focused on temporal dynamics using tools such as Singular Value Decomposition (SVD), Support Vector Machines (SVM), and Principal Component Analysis (PCA). While these techniques are effective and computationally efficient, their precision, efficiency, and ability to generalize substantially depend on large volumes of training data, which restricts their use in situations with limited measurements. The mathematical models derived are often not aligned with first-principles-based formulations, reducing their physical consistency. As a result, these black-box models present challenges in system identification scenarios where understanding the underlying physics is crucial [30, 31].

Physical systems, often in the form of non-linear partial differential equations, are modeled using first principles. Thus, sophisticated techniques have been adopted in the past for solution and coefficient discovery of these physical systems [32–34]. However, recently, Physics-Informed Neural Networks (PINNs), introduced by [35], have emerged as a powerful tool for modeling and estimating physical systems, particularly when data availability is limited. By incorporating physics-driven loss functions alongside traditional data-driven loss functions, PINNs significantly enhance the quality of neural network estimates, even in scenarios with scarce data. Leveraging their versatility, PINNs have found applications across diverse fields of science and engineering, including physics [36, 37], solid mechanics [38], fluid mechanics [39, 40], biomedical engineering [41], geotechnical engineering [42], non-linear dynamics [43–46] and atmospheric sciences [47].

In structural engineering, researchers have explored the use of PINNs for solving both forward and inverse problems. Studies such as [48–50] have demonstrated the application of PINNs for dual state-parameter estimation in linear and nonlinear multi-degree-of-freedom (MDOF) systems under forced vibration. While these efforts showcase the potential of PINNs in system identification, their primary focus on modal parameter estimation poses limitations in practical scenarios. Modal properties are often challenging to extract in real-world settings and are highly sensitive to environmental noise, reducing the robustness and applicability of these methods.

Moreover, structural systems often encompass high-frequency ordinary or partial differential equations (ODEs/PDEs) and multiscale characteristics, where PINNs often face challenges in achieving convergence [51–53]. This challenge is further compounded by the conventional architecture of PINNs, which typically rely on multilayer perceptron (MLP) frameworks. Unlike methods such as finite element analysis (FEM), which inherently incorporate temporal dependencies by iteratively solving global states based on previous time steps, traditional PINNs treat the problem in a point-to-point manner. This lack of explicit modeling of temporal dependencies can result in degraded accuracy over time, particularly for dynamic systems.

To address these temporal challenges, Long Short-Term Memory (LSTM) networks offer a promising solution due to their ability to capture long-term dependencies in sequential data using gating mechanisms [54]. LSTM models have demonstrated success in various fields. For example, [55] introduced physics-informed LSTM (PI-LSTM) networks to predict nonlinear building responses during earthquakes, highlighting the potential of combining neural networks with physics-based constraints. Similarly, [56] used PI-LSTM models to predict complex dynamic systems, and [57] applied them to both single degree-of-freedom (SDOF) systems and multi-story buildings. [58] extended PI-LSTM models to predict seismic responses and estimate isolator parameters for frame structures. A notable limitation in [58], however, was utilization of full state data at all degrees of freedom (DOFs), which is often impractical in real-world scenarios.

While LSTM models have shown impressive performance in various domains, research continues to focus on enhancing their predictive accuracy. Recently, attention mechanisms [59] have been developed to improve neural networks by focusing on the most relevant portions of the input data. Attention mechanisms have

been widely adopted in fields such as machine translation [60], semantic segmentation [61], and time series prediction [62, 63], demonstrating significant improvements in accuracy and reliability.

In the field of civil engineering, recent studies have leveraged attention-based deep learning models for structural response prediction. For instance, [64] employed an attention-based Long Short-Term Memory (Attn-LSTM) network for predicting the response of bridge structures, while [65] integrated Transformer models with LSTMs to predict nonlinear seismic responses in multi-story buildings. Building on these advancements, this study introduces a novel physics informed attention-augmented Long Short-Term Memory network (Pi-Attn-LSTM) framework aimed at parameter estimation for framed civil structures. Unlike PI-LSTM [58], the proposed method relies solely on acceleration and forcing data, enhancing its practicality for real-world applications. The results of Pi-Attn-LSTM are also compared with a variant of PI-LSTM in discrete time settings termed Pi-LSTM to demonstrate the superior prediction accuracy and robustness of Pi-Attn-LSTM.

The paper is organized as follows: Section 2 outlines the detailed methodology, covering system dynamics, LSTM networks, attention-based LSTM models, Pi-LSTM, and Pi-Attn-LSTM. Section 3 presents numerical validations on a 4-degree-of-freedom (4-DOF) linear and non-linear structural system, a 5-DOF non-linear steel-framed building, and a 12-DOF IASC-ASCE benchmark model. Experimental validation of the proposed framework is demonstrated in Section 4, using an aluminum frame structure to showcase its practical application. Finally, Section 5 concludes with a summary of the key findings and implications of the study.

2. Methodology

The proposed method introduces a Pi-Attn-LSTM network for parameter identification of civil structures. A temporal attention mechanism is incorporated to enhance the performance of the traditional LSTM based encoder-decoder framework. The primary motivation behind using the attention mechanism is its ability to adaptively identify key features within the input time series, thereby improving predictive accuracy. This section begins with an introduction to the fundamental concepts of system dynamics. Following that, LSTM, LSTM based encoder-decoder model and attention mechanism are presented. After this the final illustration of the Pi-LSTM and Pi-Attn-LSTM are illustrated.

2.1. System Dynamics

The study of linear mechanical system dynamics is governed by foundational equations that describe the interplay between a system's motion, external forces, and boundary conditions. For complex systems with multiple dynamic interactions, these governing equations are typically discretized spatially into degrees of freedom (DOFs) via Finite Element (FE) modeling, which transforms the continuous system into a

manageable set of equations for numerical solutions as:

$$\mathbf{M}\ddot{\mathbf{x}}(t) + \mathbf{C}\dot{\mathbf{x}}(t) + \mathbf{K}\mathbf{x}(t) = \mathbf{u}(t) \tag{1}$$

Time integration methods are then applied to advance the system response through discrete time steps [66]. A linear time-invariant (LTI) system can be represented in state-space form as follows:

$$\dot{\mathbf{z}}(t) = \mathbf{F}_c\mathbf{z}(t) + \mathbf{E}_c\mathbf{u}(t) \tag{2}$$

where $\mathbf{z}(t)$ is the state vector comprising $\{\mathbf{x}, \dot{\mathbf{x}}\}^T$, $\mathbf{u}(t)$ denotes external forces, $\mathbf{F}_c$ is the state matrix characterizing intrinsic system dynamics, and $\mathbf{E}_c$ is the input matrix, both defined in continuous-time. The matrices $\mathbf{F}_c$ and $\mathbf{E}_c$ are given as:

$$\mathbf{F}_c = \begin{bmatrix} \mathbf{0}_{n\times n} & \mathbf{I}_{n\times n} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{C} \end{bmatrix}_{2n\times 2n} \tag{3}$$

$$\mathbf{E}_c = \begin{bmatrix} \mathbf{0}_{n\times n} \\ \mathbf{M}^{-1} \end{bmatrix}_{2n\times p} \tag{4}$$

Here, the mass ($\mathbf{M}$), damping ($\mathbf{C}$), and stiffness ($\mathbf{K}$) matrices characterize the structural properties of the system, with $\mathbf{I}_{n\times n}$ as the identity matrix and $\mathbf{0}_{n\times n}$ as the zero matrix, indicating no direct interaction between displacement and velocity in the system's inherent dynamics.

The state vector $\mathbf{z}(t) = \begin{bmatrix} \mathbf{x}(t) \\ \dot{\mathbf{x}}(t) \end{bmatrix}_{2n\times 1}$ includes both displacements $\mathbf{x}(t)$ and velocities $\dot{\mathbf{x}}(t)$ for each DOF. External forces $\mathbf{u}(t)$ act as system inputs, with size $p \times 1$, where p corresponds to the number of forces applied.

The system's outputs are observed through measurement equations that link the states to observable quantities, represented as:

$$\mathbf{y}(t) = \mathbf{H}_c\mathbf{z}(t) + \mathbf{L}_c\mathbf{u}(t) \tag{5}$$

In this equation, $\mathbf{y}(t)$ refers to measured outputs (e.g. accelerations at particular locations). The output matrix $\mathbf{H}_c = \begin{bmatrix} -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{C} \end{bmatrix}_{m\times 2n}$ maps the state vector $\mathbf{z}(t)$ to the measurement domain, while $\mathbf{L}_c = \begin{bmatrix} \mathbf{M}^{-1} \end{bmatrix}_{m\times p}$ captures the direct influence of external forces on measurements. Here, m is the number of outputs measured, typically fewer than the system's total states.

In practical applications, continuous-time equations are not directly observable; instead, measurements are obtained at discrete time intervals. Thus, the continuous-time equations must be discretized for numerical simulation. Using methods such as zero-order hold (ZOH) or bilinear transformation, the discrete-time state-space equations are formulated as follows:

$$\mathbf{z}_k = \mathbf{F}\mathbf{z}_{k-1} + \mathbf{E}\mathbf{u}_k \tag{6}$$

$$\mathbf{y}_k = \mathbf{H}\mathbf{z}_k + \mathbf{L}\mathbf{u}_k \tag{7}$$

where $\mathbf{z}_k$ and $\mathbf{u}_k$ are the state and input vectors at discrete time step k, with $\mathbf{F}$, $\mathbf{E}$, $\mathbf{H}$, and $\mathbf{L}$ as the discrete-time counterparts of $\mathbf{F}_c$, $\mathbf{E}_c$, $\mathbf{H}_c$, and $\mathbf{L}_c$. These discrete equations provide the basis for numerically simulating and estimating the dynamics of time-invariant systems, enabling future state predictions from sampled data.

To estimate the system health, Equation (6),(7) are parameterized with location based health parameters $\{\lambda_i\},\ i = 1, \cdots, n$, where n denotes the number of monitored locations. Each λ_i is defined to signify the localized stiffness deterioration level, with 0 indicating complete damage and 1 denoting a healthy state. For elucidation, Equation (6), (7) is here reformulated to reflect the dependence of the system matrices on the health indices λ:

$$\mathbf{z}_k = \mathbf{F}(\lambda)\mathbf{z}_{k-1} + \mathbf{E}\mathbf{u}_k \tag{8}$$

$$\mathbf{y}_k = \mathbf{H}(\lambda)\mathbf{z}_k + \mathbf{L}\mathbf{u}_k \tag{9}$$

This approach offers a robust framework for analyzing the dynamics of complex civil structures and facilitates the application of advanced data-driven methods, such as Physics-Informed Neural Networks (PINNs), in structural health monitoring contexts.

2.2. LSTM and Attention based LSTM encoder-decoder model

For the system with its dynamics defined in discrete time settings in the previous subsection, it becomes essential to select a framework capable of effectively handling time series data. Long-short-term memory (LSTM) neural network, a variant of Recurrent neural network (RNN) designed to eliminate the drawbacks of RNN such as preservation of long-term dependencies in sequential data using gate mechanism is therefore used as a reliable tool for time series modeling of structural systems [55, 57]. As shown in Figure 1(b), the standard architecture of a LSTM neural network contains a series of self-connected LSTM cells detailed in Figure 1(a), each of which comprises three interconnected components including an input gate, a forget gate, and an output gate. The input gate modulates the incorporation of new information into the cell state, enabling selective updating. The forget gate adaptively scales the long-dependent information from the previous time step, facilitating the retention or discarding of past information as required. The cell state undergoes updates based on the combined effects of the input and forget gates, thereby preserving only the relevant temporal information. The output gate modifies the hidden output of the LSTM cell according to the updated cell state.

In a discrete-time setup, at any arbitrary k^{th} time step t_k (where $k = 1, \ldots, l$), let the input state be x_k, the forget gate be f_k, the input gate be i_k, the output gate be o_k, the cell state memory be c_k, and the

hidden state output be h_k. For the previous time step $k-1$, the cell state memory is c_{k-1} and the hidden state output is h_{k-1}. The relationships among these variables are described by the following equations (also depicted in Figure 1(b)):

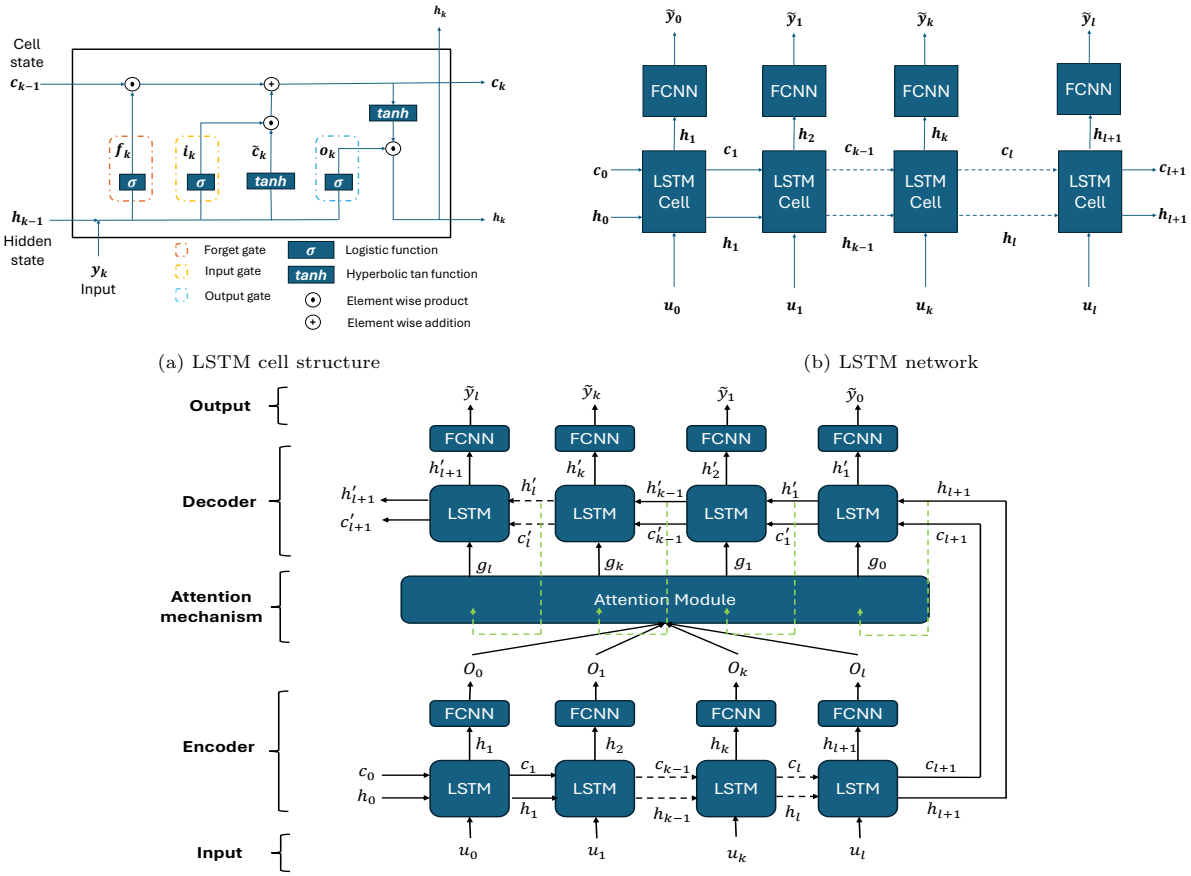
$$\begin{aligned} f_k &= \sigma\left(W_{xf}x_k + W_{hf}h_{k-1} + b_f\right), \\ i_k &= \sigma\left(W_{xi}x_k + W_{hi}h_{k-1} + b_i\right), \\ \tilde{c}_k &= \tanh\left(W_{xc}x_k + W_{hc}h_{k-1} + b_c\right), \\ o_k &= \sigma\left(W_{xo}x_k + W_{ho}h_{k-1} + b_o\right), \\ c_k &= f_k \odot c_{k-1} + i_k \odot \tilde{c}_k, \\ h_k &= o_k \odot \tanh\left(c_k\right) \end{aligned} \tag{10}$$

Here, $W_{\alpha\beta}$ (with $\alpha = \{x, h\}$ and $\beta = \{f, i, c, o\}$) are the weight matrices corresponding to different inputs (e.g. x_k or h_k) within different gates (e.g. input gate, forget gate, *tanh* layer or output gate), while b_β represents the corresponding bias vectors. For instance, W_{xf} and W_{hf} are the weight matrices for input vectors x_k and h_{k-1}, respectively, within the forget gate. The variable $\tilde{c}_k$ denotes a vector of intermediate candidate values created by a *tanh* layer, σ is the logistic sigmoid function, *tanh* is the hyperbolic tangent function, and $\odot$ denotes the Hadamard product (element-wise multiplication).

Although LSTM models have demonstrated remarkable performance across diverse domains, research continues to explore methods for enhancing their predictive accuracy. Recently, attention mechanisms [59] have emerged as a powerful tool for improving neural networks by selectively focusing on the most relevant portions of input data. The attention-based LSTM encoder-decoder model is an enhancement of the standard LSTM and sequence-to-sequence architecture [67], particularly effective for handling long and complex input sequences [58]. In a traditional LSTM encoder-decoder, the encoder processes an input sequence and encodes it into a fixed-length context vector. The decoder then uses this context vector to generate the output sequence. However, this approach struggles when dealing with long input sequences because the fixed-length vector may not adequately capture all relevant information [62, 63]. The introduction of an attention mechanism [60] overcomes this limitation by allowing the decoder to selectively focus on different parts of the input sequence during the decoding process, dynamically adjusting its focus at each time step. The attention mechanism enhances the standard LSTM encoder-decoder by allowing the decoder to access the entire sequence of encoder hidden states rather than compressing the input into a single vector.

As shown in Figure 1(c) encoder takes an input sequence $\mathbf{u} \in \mathbb{R}^l$, where l is the length of the sequence. The LSTM encoder processes the input sequence and produces hidden states $h_k \in \mathbb{R}^H$ at each time step k, where H is the hidden dimension size. Further output of the LSTM encoder is transformed through a fully connected layer neural network (FCNN),

$$o_k = \text{FCNN}(h_k), \tag{11}$$



(a) LSTM cell structure

(b) LSTM network

(c) Attention based LSTM encoder decoder model

Figure 1: Schematic of **(a) LSTM cell:** Detailed structure of an LSTM cell, highlighting the input, forget, and output gates that regulate information flow, enabling the cell to maintain and update information over time, **(b) LSTM neural network:** Architecture of LSTM neural network comprised of LSTM cell and fully connected neural network (FCNN), designed to capture long-term dependencies in sequential data, and **(c) Attention based LSTM encoder-decoder model:** Architecture of attention based LSTM encoder-decoder model comprised of encoder, attention module and decoder. Encoder processes the input sequence and the attention mechanism allows the decoder to focus on different parts of the input for generating accurate outputs.

Output o_k from FCNN is passed to attention module to form context vector $\mathbf{g}_k$. The attention module comprised of scaled dot-product attention as a core component, enabling the decoder to focus on different parts of the input sequence at each time step. In the attention mechanism the query matrix $\mathbf{Q} = (\mathbf{h}_1, \cdots, \mathbf{h}_{l+1})$ is the decoder's current hidden state. The keys matrix $\mathbf{K}$ and the values matrix $\mathbf{V}$ are the outputs from fully connected neural networks (FCNN) i.e. $\mathbf{K} = \mathbf{V} = (\mathbf{o}_0, \cdots, \mathbf{o}_l)$

For each decoding step k, the attention scores are computed by taking the dot product of the query $\mathbf{Q}$ and the keys $\mathbf{K}$, scaled by the square root of the hidden size H:

$$\delta_{km} = \frac{h_k^T \cdot o_m}{\sqrt{H}} \qquad \forall m \in \{0, \cdots, l\} \tag{12}$$

These attention scores δ_{km} are passed through a softmax function to generate attention weights $\mathbf{a}_{km}$,

which determine the relevance of each encoder hidden state with respect to the current decoder state:

$$\mathbf{a}_{km} = \text{softmax}\left(\delta_{km}\right) \qquad \forall m \in \{0, \cdots, l\} \tag{13}$$

The attention weights are then used to compute context vector $\mathbf{g}_k$ as a weighted sum of the values $\mathbf{V}$:

$$\mathbf{g}_k = \sum_m \mathrm{a}_{km} \cdot o_m \tag{14}$$

This context vector aggregates the relevant information from the encoder outputs based on the attention weights at the current decoding step. The decoder LSTM generates the output sequence iteratively, using both the context vector from the attention mechanism and the previous hidden state as input. Initially, the last hidden state from the encoder is used to initialize the decoder's hidden state.

For each time step k in the decoder, the context vector computed by the attention mechanism is passed to the decoder LSTM. The decoder LSTM updates its hidden state based on the context vector and its previous hidden state:

$$h_k^{'}, c_k^{'} = \text{LSTM}(\mathbf{g}_k, (h_{k-1}^{'}, c_{k-1}^{'})) \tag{15}$$

The output of the decoder LSTM is then passed through a fully connected layer to produce the final prediction for that time step:

$$\tilde{y}_k = \text{FCNN}(h_k^{'}) \tag{16}$$

where $\tilde{y}_k$ is the predicted output for time step k. The decoding process proceeds step-by-step, where at each step, the attention mechanism computes a new context vector $\mathbf{g}_k$ based on the current decoder state $\mathbf{o}_k$. This context vector allows the decoder to focus on different parts of the input sequence dynamically, ensuring that relevant information from the encoder is used to generate each output. The decoder's hidden state and the context vector are continuously updated, enabling the model to generate an entire sequence of predictions.

2.3. Physics-informed Attention LSTM (Pi-Attn-LSTM)

To estimate system health parameters, the system dynamics, as shown in equations (8) and (9), are combined with the attention-based LSTM framework introduced earlier to form a unified Pi-Attn-LSTM model. The proposed Pi-Attn-LSTM framework is also compared with the standard Pi-LSTM to highlight the performance improvements enabled by the attention mechanism. Denoting both networks by $N(\cdot)$ and their parameters by $\{\tilde{\theta}, \tilde{\lambda}\}$ — where $\tilde{\theta}$ represents the initial weights and biases, and $\tilde{\lambda}$ denotes the initial health parameters. System dynamics, used as a regularization term in Pi-Attn-LSTM and Pi-LSTM can be reformulated as:

$$\mathbf{z}_k = \mathbf{F}(\tilde{\lambda})\mathbf{z}_{k-1} + \mathbf{E}\mathbf{u}_k \tag{17}$$

$$\mathbf{y}_k = \mathbf{H}(\tilde{\lambda})\mathbf{z}_k + \mathbf{L}\mathbf{u}_k \tag{18}$$

The Pi-Attn-LSTM and Pi-LSTM frameworks utilize three distinct datasets during the training process: (i) input forces $[\mathbf{u}_k^c]_{m\times N_c}$, available at collocation points $[\mathbf{t}_k^c]_{1\times N_c}$, which serve as inputs to the network; (ii) high-fidelity measurement and force dataset pairs $\left[\mathbf{y}_k^h, \mathbf{u}_k^h\right]_{m\times N_h}$, available at N_h time points; and (iii) initial conditions $[\mathbf{z}_0]_{m\times 1}$, available at the initial time point t_0. These three data sources are integral to the training process of both the Pi-LSTM and Pi-Attn-LSTM models.

Collocation points represent uniformly spaced time instances at which the physics-informed constraints are applied, while high-fidelity measurement data are obtained at time points where experimental measurements are available. Generally, the number of high-fidelity data points N_h equals the number of collocation points N_c. However, to account for real-world challenges such as sensor malfunctions or low resolution sensor data, number of high fidelity points N_h can be less than collocation points N_c. This dual approach of inducing data and physics enhances the robustness and generalizability of the Pi-LSTM/Pi-Attn-LSTM models, allowing for accurate predictions and deeper insights into the underlying system dynamics, even under uncertain or incomplete data conditions.

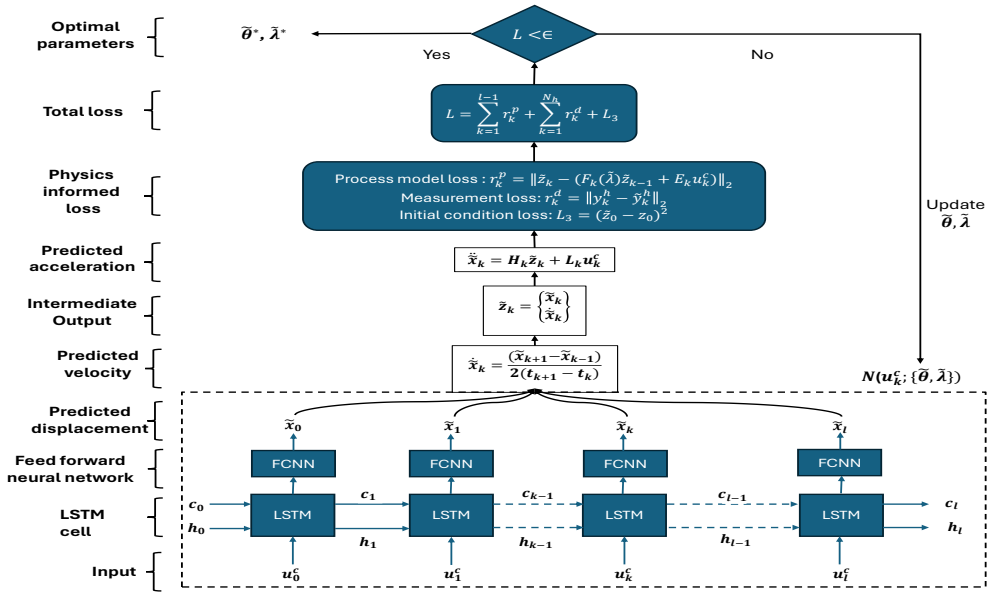
The proposed Pi-Attn-LSTM and Pi-LSTM takes time $[\mathbf{t}_k^c]_{m\times 1}$ as the primary input. For (i) Pi-LSTM as shown in Figure 2(a) the network architecture consists of an LSTM layer followed by a fully connected neural network layer (FCNN), and for (ii) Pi-Attn-LSTM shown in Figure 2(b), the network architecture comprised of LSTM layer followed by FCNN termed as encoder, output from encoder is passed through attention module to form a context vector $\mathbf{g}_k$, which is further passed to decoder LSTM network followed by FCNN to predict displacement $\tilde{\mathbf{x}}$. Prediction from network $N(\cdot)$ at time instant k can be formulated as follows:

$$\tilde{\mathbf{x}}_k = N(\mathbf{u}_k^c, \{\tilde{\theta}, \tilde{\lambda}\}) \tag{19}$$

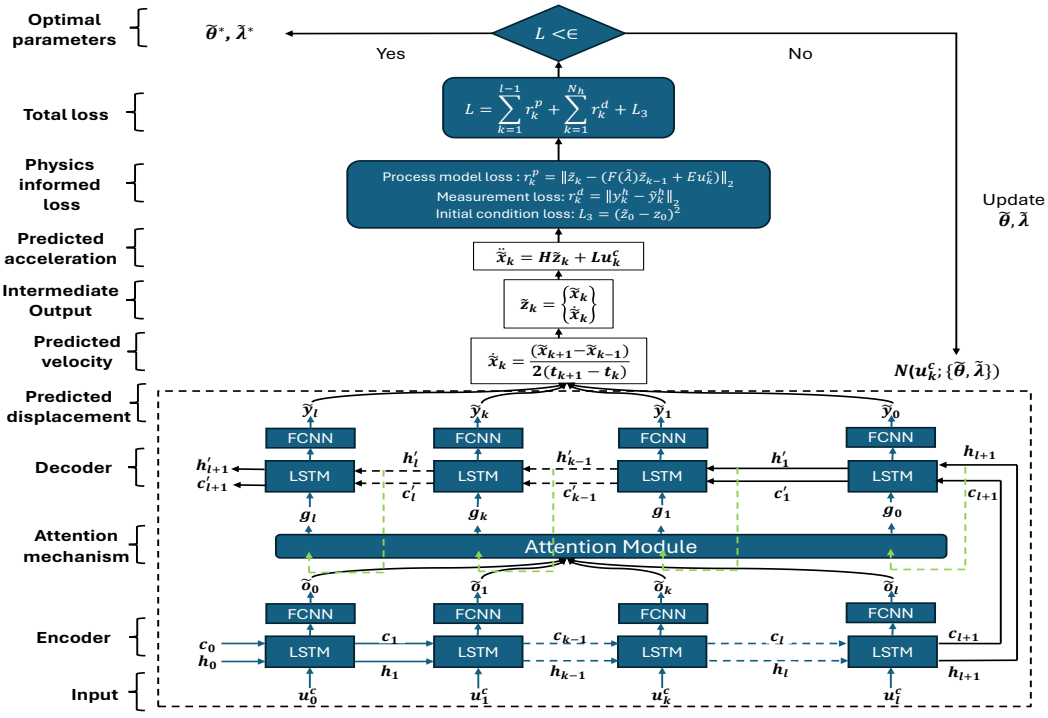
where, $\tilde{\mathbf{x}}_k$ represents the predicted displacement in the time step k. The goal is to optimize these parameters to enhance the accuracy of the network's predictions. By refining $\tilde{\theta}$ and $\tilde{\lambda}$, both the physics-based loss (related to adherence to physical laws) and the data-based loss (related to discrepancies between predictions and actual data) can be minimized.

To validate data-based predictions against physical laws and real response data, a collective loss function is employed. For comparison between model predictions and either data or physics, the predicted displacement, $\tilde{\mathbf{x}}_k$, is numerically differentiated using a central finite difference scheme to obtain the corresponding predicted velocity, $\dot{\tilde{\mathbf{x}}}_k$. Based on these predictions, a residual r_k^p is defined as:

$$r_k^p = \left\| \tilde{\mathbf{z}}_k - (\mathbf{F}(\tilde{\lambda})\tilde{\mathbf{z}}_{k-1} + \mathbf{E}(\tilde{\lambda})\mathbf{u}_k^c) \right\|_2 \tag{20}$$



(a) Pi-LSTM



(b) Pi-Attn-LSTM

Figure 2: Schematic diagram of **(a) Physics-informed LSTM (Pi-LSTM):** The model architecture uses a feedforward neural network combined with LSTM cells to predict displacement and **(b) Physics informed attention LSTM (Pi-Attn-LSTM):** The architecture combines an encoder-decoder structure with attention mechanisms to predict displacement. Both networks utilizes a physics-informed loss function, incorporating process model loss, measurement loss, and initial condition loss, to estimate the optimal parameters ($\tilde{\theta}$, $\tilde{\lambda}$).

Here, $\mathbf{F}(\tilde{\lambda})$ and $\mathbf{E}(\tilde{\lambda})$ are computed based on current estimates of λ (i.e., $\tilde{\lambda}$). The vectors $\tilde{\mathbf{z}}_k$ and $\tilde{\mathbf{z}}_{k-1}$ are formed by combining displacement and velocity estimates at time steps k and $k-1$, respectively:

$$\tilde{\mathbf{z}}_i = \begin{bmatrix} \tilde{\mathbf{x}}_i \\ \dot{\tilde{\mathbf{x}}}_i \end{bmatrix} \tag{21}$$

The physics loss is computed at all collocation points N_c, with residuals from these points integrated as:

$$\mathcal{L}_1(\{\tilde{\theta}, \tilde{\lambda}\}) = \frac{1}{N_c} \sum_{k=1}^{N_c} r_k^p \tag{22}$$

Further, the measurement model is used to estimate the predicted acceleration $\ddot{\tilde{\mathbf{x}}}_k$ at each time step as follows:

$$\ddot{\tilde{\mathbf{x}}}_k = \mathbf{H}(\tilde{\lambda})\tilde{\mathbf{z}}_k + \mathbf{L}\mathbf{u}_k^c \tag{23}$$

To assess data loss, the predicted acceleration response is compared with the true acceleration response. Since true measurements are available at N_h time points ($N_h \subset N_c$), the predicted measurements $\tilde{\mathbf{y}}_k^h$ are sampled from $\ddot{\tilde{\mathbf{x}}}_k$, and the data loss, $\mathcal{L}_2(\{\tilde{\theta}, \tilde{\lambda}\})$, is calculated as follows:

$$r_k^d = \left\| \mathbf{y}_k^h - \tilde{\mathbf{y}}_k^h \right\|_2 \tag{24}$$

An L2 norm is applied to quantify discrepancies across different degrees of freedom (DOFs) between the predicted and true measurements, yielding a single scalar value. The total data loss is computed as:

$$\mathcal{L}_2(\{\tilde{\theta}, \tilde{\lambda}\}) = \frac{1}{N_h} \sum_{k=1}^{N_h} r_k^d \tag{25}$$

Additionally, an initial condition loss is defined as:

$$\mathcal{L}_3(\{\tilde{\theta}, \tilde{\lambda}\}) = \left\| \mathbf{z}_0 - \tilde{\mathbf{z}}_0 \right\|_2 \tag{26}$$

The total loss function is thus formulated as:

$$L(\{\tilde{\theta}, \tilde{\lambda}\}) = \alpha\mathcal{L}_1(\{\tilde{\theta}, \tilde{\lambda}\}) + \beta\mathcal{L}_2(\{\tilde{\theta}, \tilde{\lambda}\}) + \gamma\mathcal{L}_3(\{\tilde{\theta}, \tilde{\lambda}\}) \tag{27}$$

where α, β, and γ are hyperparameters that weight the different loss terms. Although various studies have explored optimal selection of these hyperparameters [61, 68, 69], in this study, they are set to unity. A self-adaptive weighting strategy could be employed in future work to expedite the training process. The optimization problem to determine the neural network and structural parameters $\{\tilde{\theta}, \tilde{\lambda}\}$ is thus formulated as follows:

$$\{\tilde{\theta}^*, \tilde{\lambda}^*\} = \arg\min_{\{\tilde{\theta},\tilde{\lambda}\}} L(\{\tilde{\theta}, \tilde{\lambda}\}) \tag{28}$$

3. Numerical validation

To validate the proposed framework in structural engineering applications, two case studies are considered, each representing civil structures with distinct characteristics. The first case study examines a 4-story steel frame previously investigated in [70], providing a reference of a laboratory-scaled framed structure. The second case study uses a numerical model of a 5-story real steel frame building acting as a reference for the real structure, as detailed in [71]. The third case study employs the IASC-ASCE benchmark model acting as a standard to test the efficacy of many SHM algorithms as referenced in [72]. The performance of the Pi-LSTM model alongside a Pi-Attn-LSTM model is evaluated to highlight the benefits of the attention mechanism. Specifically, the robustness of Pi-Attn-LSTM is checked under various conditions, including differing initialization schemes, noise levels, and force inputs, as well as under conditions of temporal sparsity. This analysis is intended to explore the adaptability and enhanced performance of the attention-augmented model in handling real-world structural monitoring challenges.

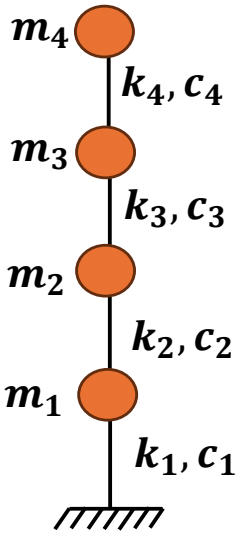
3.1. 4-DOF linear system

The proposed approach is validated using a linear 4-*dof* structural system represented by a spring-mass-dashpot model adopted in [70]. A schematic diagram of the same is presented in Figure 3. The generalized equation of motion for the dynamics of the system can be perceived from the generalized equation in Equation (2) that imposes the assumed dimension. Accordingly, the mass matrix can be idealized with $\mathbf{M} = diag([m_1; m_2; m_3; m_4])$, and other system matrices can be defined as follows:

$$\mathbf{K} = k_0 \begin{bmatrix} \lambda_1 + \lambda_2 & -\lambda_2 & 0 & 0 \\ -\lambda_2 & \lambda_2 + \lambda_3 & -\lambda_3 & 0 \\ 0 & -\lambda_3 & \lambda_3 + \lambda_4 & -\lambda_4 \\ 0 & 0 & -\lambda_4 & \lambda_4 \end{bmatrix}_{4\times4}, \quad \mathbf{F}_c = \begin{bmatrix} \mathbf{0}_{4\times4} & \mathbf{I}_{4\times4} \\ \mathbf{M}^{-1}\mathbf{K} & \mathbf{M}^{-1}\mathbf{C} \end{bmatrix}_{4\times4}, \quad \mathbf{E}_c = \begin{bmatrix} \mathbf{0}_{4\times4} \\ \mathbf{M}^{-1} \end{bmatrix}_{4\times2}$$

In this system, the elements of the diagonal mass matrix are each set to $m_1, m_2, \ldots, m_4 = 84.3$ kg. The initial undamaged stiffness of all springs is assumed to be $k_{0,i} = 126.3$ kN/m for each i. Damping coefficients are determined using Rayleigh damping with a damping ratio of $\zeta = 2\%$. The system is considered to have deteriorated in its second, third, and fourth degrees of freedom (DOFs). These damages are represented through a set of health indices (HIs), denoted by λ, which quantify the level of deterioration in the spring stiffness on a scale from 0 to 1. Here, $\lambda = 1$ represents a fully healthy spring, while $\lambda = 0$ indicates complete failure. The health indices modify the initial stiffness according to $k_i = \lambda_i k_{0,i}$, where k_i is the effective stiffness of the i-th spring, λ_i is the corresponding health index, and $k_{0,i}$ is the initial undamaged stiffness.

(a) University of Connecticut laboratory-scale structure[70]



(b) Equivalent lumped mass model

Figure 3: Schematic diagram of steel frame for numerical experiment

With this configuration, the true health indices at each location are represented by the vector $\lambda = \{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}\}$, indicating varying levels of induced damage across the system. The lumped spring-mass-dashpot model is subsequently simulated under two excitation scenarios: (i) earthquake excitation using the El Centro earthquake data from May 18, 1940 (North-South Component) [73], and (ii) ambient vibration modeled as stochastic white Gaussian noise (SWGN), sampled at a rate of 50 Hz and applied to all nodes. For parameter identification in these cases, training data consisting of acceleration time histories is generated over a 5-second interval, sampled at a frequency of 100 Hz.

3.1.1. Network Topology

The network architecture for the Pi-LSTM model begins with an input layer containing 4 neurons, followed by a single-layer LSTM cell with 128 hidden units. The output of the LSTM cell is passed to a fully connected neural network (FCNN) comprising 128 input neurons, 4 hidden layers with 128 neurons each, and a final output layer with 4 neurons.

The Pi-Attn-LSTM model features three main components: encoder, attention mechanism, and decoder. The encoder includes an LSTM cell followed by an FCNN. The LSTM cell takes input from a 4-neuron input layer and produces hidden and cell states, each of 128 dimensions. These hidden states are fed into an FCNN with two hidden layers, each containing 128 neurons, to produce a transformed output vector, $\mathbf{o}_k$, of 128 dimensions. The attention mechanism processes these outputs to generate a context vector, $\mathbf{g}_k$, also of 128 dimensions.

The context vector is then passed to the decoder, which consists of an LSTM cell followed by an FCNN.

The LSTM cell takes the 128-dimensional context vector as input and produces output, which is subsequently processed by an FCNN with two hidden layers (128 neurons each) and a final output layer of 4 neurons.

Both Pi-LSTM and Pi-Attn-LSTM models are optimized using the Adam optimizer with a learning rate of 1×10^{-4}, trained for 1×10^5 epochs. For consistency in comparison, weights and biases (θ) are initialized from a normal distribution, and the structural parameters ($\tilde{\lambda}$) are initialized to 1. Initial hidden ($\mathbf{h}_0$) and cell ($\mathbf{c}_0$) states are set to zero in both models.

3.1.2. Comparitive study between Pi-Attn-LSTM and Pi-LSTM

An initial convergence study of the Pi-LSTM and Pi-Attn-LSTM models was performed, assuming 100% data availability with a noise level of 1%. As illustrated in Figure 4(a), the convergence histories reveal that Pi-LSTM and Pi-Attn-LSTM reach minimum loss values of 0.012 and 0.0079, respectively, after approximately 1×10^5 epochs. This indicates that both models converge, but Pi-Attn-LSTM achieves a lower loss, highlighting its improved accuracy.

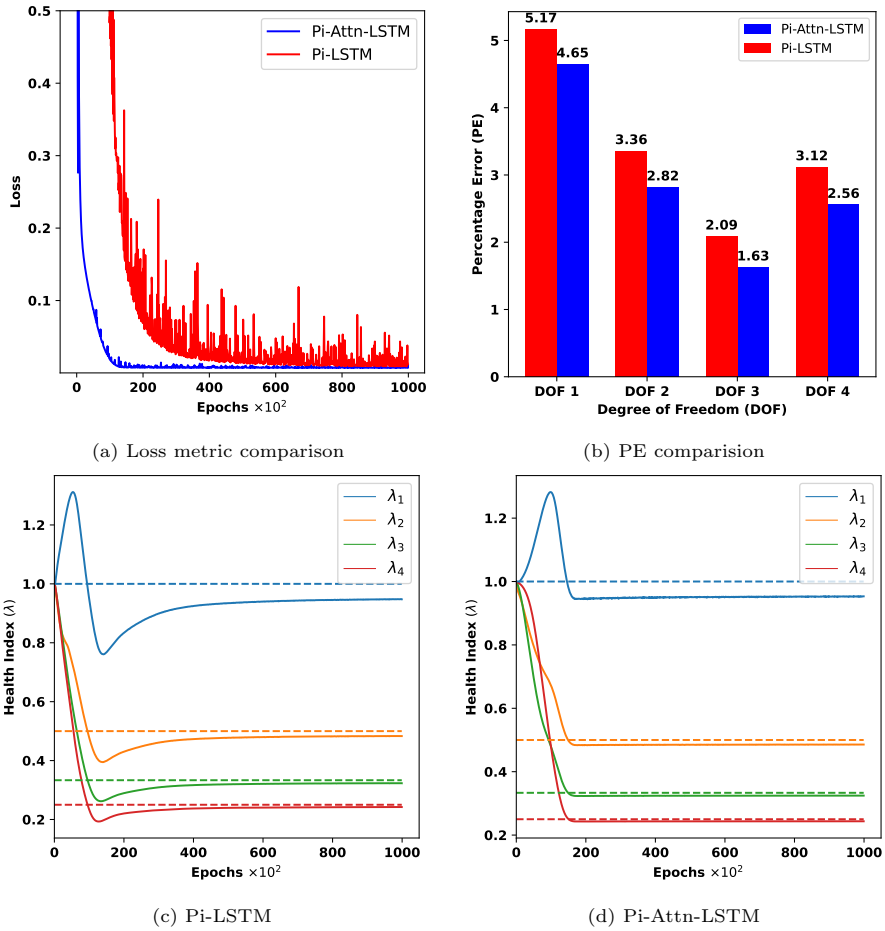
Further analysis of the percentage error (PE) in parameter estimation, shown in Figure 4(b, c, d), demonstrates that Pi-Attn-LSTM maintains a PE below 5% in the estimation of critical model parameters, while also converging faster than Pi-LSTM to achieve optimal health parameters before 2×10^4 epochs. The observed advantage in convergence speed and precision can be attributed to the attention mechanism embedded in Pi-Attn-LSTM, which allows the model to dynamically weigh input features across time steps, prioritizing information that contributes most to prediction accuracy.

3.1.3. Adaptability against forcing and initialization variability

The robustness of the proposed Pi-Attn-LSTM framework was further evaluated by examining its performance under varying input forces and different initialization schemes. As shown in Figure 5(a, b), the model maintained its predictive accuracy despite changes in the initial parameter settings. This consistency indicates that Pi-Attn-LSTM is less affected by the randomness inherent in parameter initialization, which is a critical attribute for applications where precise initialization may not always be feasible. Additionally, the model's stable performance under different excitation types, including both deterministic (earthquake) and stochastic (SWGN) inputs, highlights its robustness. This stability suggests that Pi-Attn-LSTM can effectively generalize across a range of real-world conditions, making it a reliable tool for structural health monitoring (SHM) applications that involve varying environmental and operational conditions.

3.1.4. Sensitivity against noise and data sparisity

The robustness of the Pi-Attn-LSTM model was also examined under conditions of varying noise levels and temporal and spatial data sparsity, both of which are common challenges in real-world data. In scenarios with reduced signal-to-noise ratio (SNR), as illustrated in Figure 6(a), the model demonstrated strong resilience; even at an SNR of 5%, the parameter estimation error (PE) remained below 10%. This outcome



(a) Loss metric comparison

(b) PE comparision

(c) Pi-LSTM

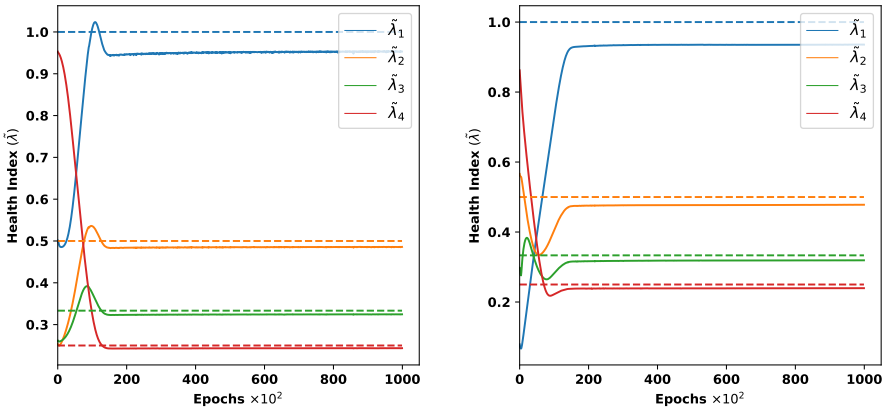
(d) Pi-Attn-LSTM

Figure 4: Comparative performance analysis of Pi-Attn-LSTM and Pi-LSTM frameworks in structural health monitoring tasks. **(a) Loss metric comparison:** The loss metric over training epochs for both Pi-Attn-LSTM (blue) and Pi-LSTM (red) models, showing faster convergence and stability for Pi-Attn-LSTM. **(b) Percentage error (PE) Comparison:** Bar plot illustrating the percentage error (PE) across different degrees of freedom (DOF) for both models, with Pi-Attn-LSTM (blue) generally achieving lower PE than Pi-LSTM (red) across all DOFs. **(c) Health index ($\tilde{\lambda}$) evolution for Pi-LSTM:** The evolution of the health index parameters $\tilde{\lambda}_1$ through $\tilde{\lambda}_4$ over epochs, showing convergence trends for the Pi-LSTM model. **(d) Health Index ($\tilde{\lambda}$) Evolution for Pi-Attn-LSTM:** Similar evolution for the Pi-Attn-LSTM model, indicating smoother convergence with the attention mechanism in place.

highlights the Pi-Attn-LSTM's ability to maintain predictive accuracy despite high noise levels, which is crucial in applications where data quality may fluctuate.

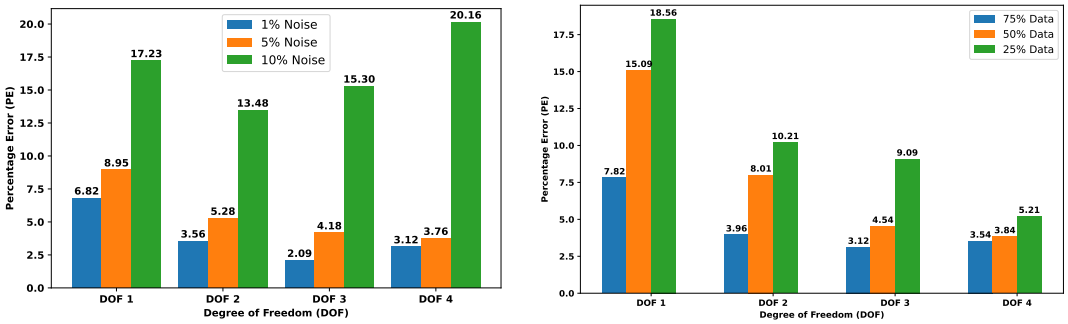
As illustrated in Figure 6(b), while increased data sparsity led to a rise in PE in estimated parameters from 7.82% to 18.56%, this effect highlights a valuable insight: certain health parameters, particularly in the first story, are more responsive to data density. This sensitivity may stem from the complex dynamic response at the first degree of freedom, where stiffness contributes significantly to the signal characteristics. Consequently, understanding and addressing these effects could lead to enhanced prediction quality in such scenarios.

A promising way forward involves leveraging modal information derived from existing measurement



(a) El-Centro (b) SWGN

Figure 5: Convergence of randomly initialized health index $\tilde{\lambda}$ of 4-DOF steel frame subjected to El-Centro and ambient vibration with Pi-Attn-LSTM. **(a) Health Index ($\tilde{\lambda}$) with El-centro:** Convergence of the health index parameters $\tilde{\lambda}_1$ through $\tilde{\lambda}_4$ under random initialization conditions, illustrating the robustness of the proposed framework to initialization variability. **(b) Health Index ($\tilde{\lambda}$) with ambient vibration:** Health index evolution when subjected to ambient vibration modeled as SWGN demonstrating the robustness of the framework to handle forcing variability.



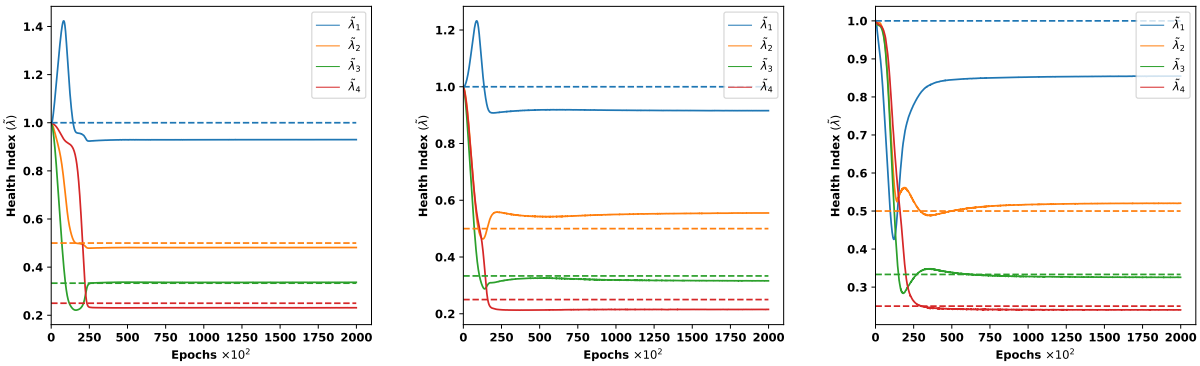
(a) Noise sensitivity analysis (b) Data spasticity analysis

Figure 6: Analysis of percentage error (PE) in structural parameter estimation under varying noise levels and data availability conditions for each degree of freedom (DOF). **(a) Noise sensitivity analysis:** Percentage error (PE) across different degrees of freedom (DOF 1 to DOF 4) under varying noise conditions (1%, 5%, and 10% noise levels). The results show that higher noise levels lead to increased estimation error across all DOFs, with the largest error observed at the 10% noise level. **(b) Data sparsity analysis:** Percentage error (PE) across DOFs under varying data sparcity conditions (75%, 50%, and 25% of the data retained). Increase in data scarcity leads to error increment, with the highest errors at 25% data, indicating the model's sensitivity to data sparsity. *This figure depicts the robustness of Pi-Attn-LSTM under different noise levels and data sparsity, highlighting the impact of both factors on parameter estimation accuracy.*

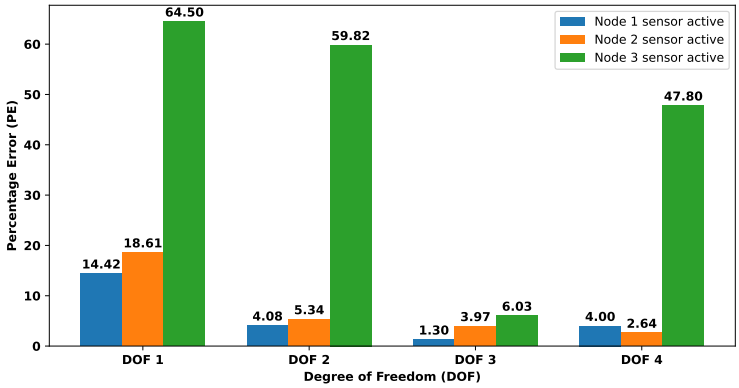
data. Modal insights capture fundamental dynamic behaviors, offering a robust supplement to data-limited predictions. This approach suggests an exciting future direction: developing methods to extract and apply modal information within the physics-informed neural network (PINN) framework could bolster the Pi-Attn-LSTM's predictive robustness. Integrating such data-driven modal information could allow the model to perform consistently well, even when data is sparse, potentially expanding its effectiveness in real-world applications with limited data availability.

Furthermore, sensor malfunctions in space can significantly impact the accuracy of parameter estimation. To evaluate the effect of sensor sparsity, a systematic analysis is conducted, as shown in Figure 7(a-c). The

results reveal that a reduction in the number of active sensors leads to increased estimation errors, with the highest inaccuracies observed when only a single sensor is operational. However, the estimation errors remain within acceptable limits for damaged elements even in the presence of extreme sensor sparsity. Furthermore, optimal sensor location test is also conducted as shown in Figure 7(d) highlighting the importance of optimal sensor placement in achieving high accuracy, as also suggested in [74]. These findings demonstrate the robustness of the PINN-based framework in handling data sparsity scenarios, reinforcing its applicability to real-world structural health monitoring challenges.



(a) 3-sensors active (b) 2-sensors active (c) 1-sensor active



(d) Optimal sensor location analysis

Figure 7: Analysis of the impact of sensor configuration on structural parameter estimation accuracy under 90% data availability subjected to El-Centro ground motion under 1% noise. **(a–c) Sensor activation analysis:** Convergence behavior of estimated health index (λ_i) under different sensor configurations (three sensors, two sensors, and one sensor active). The results indicate that reducing the number of active sensors affects the convergence rate and accuracy of parameter estimation. **(d) Optimal sensor location analysis:** Percentage error (PE) across different degrees of freedom (DOFs) under varying sensor location configurations, showing that with the largest errors observed when sensor placed away from 1st node. This figure highlights the effect of sensor placement on parameter estimation accuracy, emphasizing the importance of optimal sensor positioning for robust performance of Pi-Attn-LSTM.

3.1.5. 4-DOF cubic non-linear system

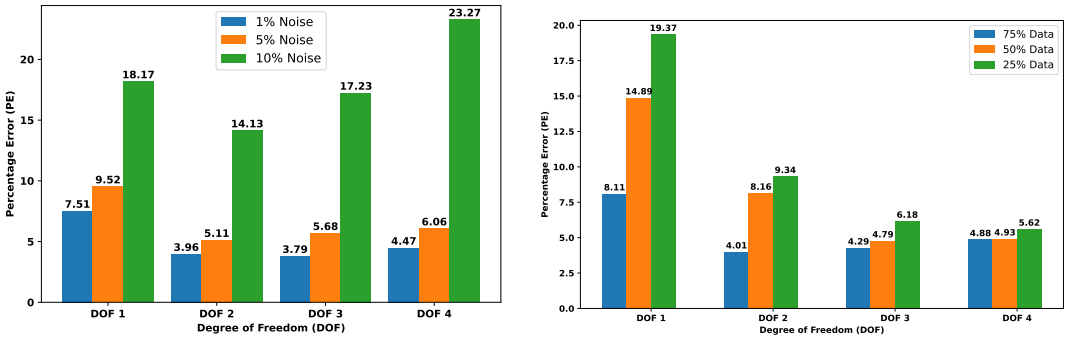
To further evaluate the efficacy of the proposed algorithm, the previously analyzed 4-DOF system is modified to include cubic nonlinearity at the base while maintaining the same system matrices. The generalized

equation of motion in Equation 1 modifies to:

$$\mathbf{M}\ddot{\mathbf{x}}(t) + \mathbf{C}\dot{\mathbf{x}}(t) + \mathbf{K}\mathbf{x}(t) + \mathbf{D}\mathbf{x}^3(t) = \mathbf{u}(t) \tag{29}$$

where additional cubic stiffness matrix is $\mathbf{D} = \text{Diag}([k_0, 0, 0, 0])$. Nonlinear systems are important, as most real-world physical systems exhibit nonlinear behavior. Therefore, validating the proposed algorithm under such conditions is essential.

The system is subjected to the El-Centro earthquake, with the recorded force and acceleration response as the input-output pair in Pi-Attn-LSTM. Maintaining the same network topology and hyperparameters, the robustness of the Pi-Attn-LSTM model is further assessed under varying noise levels and temporal sparsity. As shown in Figure 8(a), the model demonstrates strong resilience to low SNR conditions. Even at an SNR of 5%, the PE remains below 10%, showing the model's ability to maintain predictive accuracy despite high noise levels. Additionally Figure 8(b) depicits that increasing data sparsity raises PE from 8.11% to 19.37%, with the first story exhibiting greater sensitivity due to its complex dynamic response introduced with cubic stiffness. This shows the importance of data availability in enhancing predictive accuracy, reinforcing the proposed framework's effectiveness in structural health monitoring.



(a) Noise sensitivity analysis

(b) Data spasticity analysis

Figure 8: Analysis of percentage error (PE) in structural parameter estimation under varying noise levels and data availability conditions for each degree of freedom (DOF). **(a) Noise sensitivity analysis:** Percentage error (PE) across different degrees of freedom (DOF 1 to DOF 4) under varying noise conditions (1%, 5%, and 10% noise levels). The results show that higher noise levels lead to increased estimation error across all DOFs, with the largest error observed at the 10% noise level. **(b) Data sparsity analysis:** Percentage error (PE) across DOFs under varying data sparcity conditions (75%, 50%, and 25% of the data retained). Increase in data scarcity leads to error increment, with the highest errors at 25% data, indicating the model's sensitivity to data sparsity. *This figure depicts the robustness of Pi-Attn-LSTM under different noise levels and data sparsity, highlighting the impact of both factors on parameter estimation accuracy.*

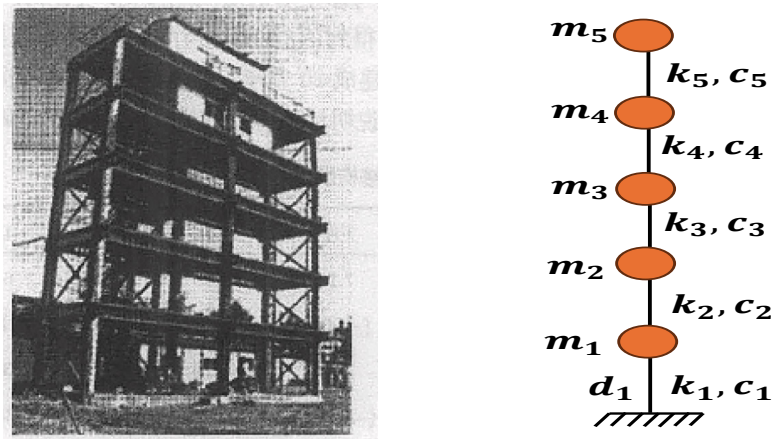
3.2. 5-storey steel building with non-linear damping

The proposed Pi-Attn-LSTM is further verified against an actual 5-story steel building built in 1970 [71] shown in Figure 9(a). The story stiffness and mass of the each floor is $\{77.89, 47.94, 52.57, 43.37, 60.35\}$ MN/m, and $\{30, 32, 30, 35, 34\}$ tonnes respectively (cf. [71]). Damping coefficient considered for each floor are $\{40, 40, 40, 40, 40\}$ KNs/m. Along with this a non-linear damper is also added at base storey having

damping coefficient of 40 KNs/m. The generalized equation of motion in Equation 1 modifies to:

$$\mathbf{M}\ddot{\mathbf{x}}(t) + \mathbf{C}\dot{\mathbf{x}}(t) + \mathbf{K}\mathbf{x}(t) + \mathbf{D}\dot{\mathbf{x}}^2(t) = \mathbf{u}(t) \tag{30}$$

where additional quadratic damping matrix is $\mathbf{D} = \text{Diag}([40, 0, 0, 0])$. These nonlinear dampers helps in dissipating energy and are commonly used in steel structural systems. Therefore, validating the proposed algorithm under such non-linear conditions is essential. Again Pi-Attn-LSTM framework is utilized for parameter identification by considering $k_{0,i}$ as 77.89 MN/m and initializing $\tilde{\lambda}_i$ with 1.



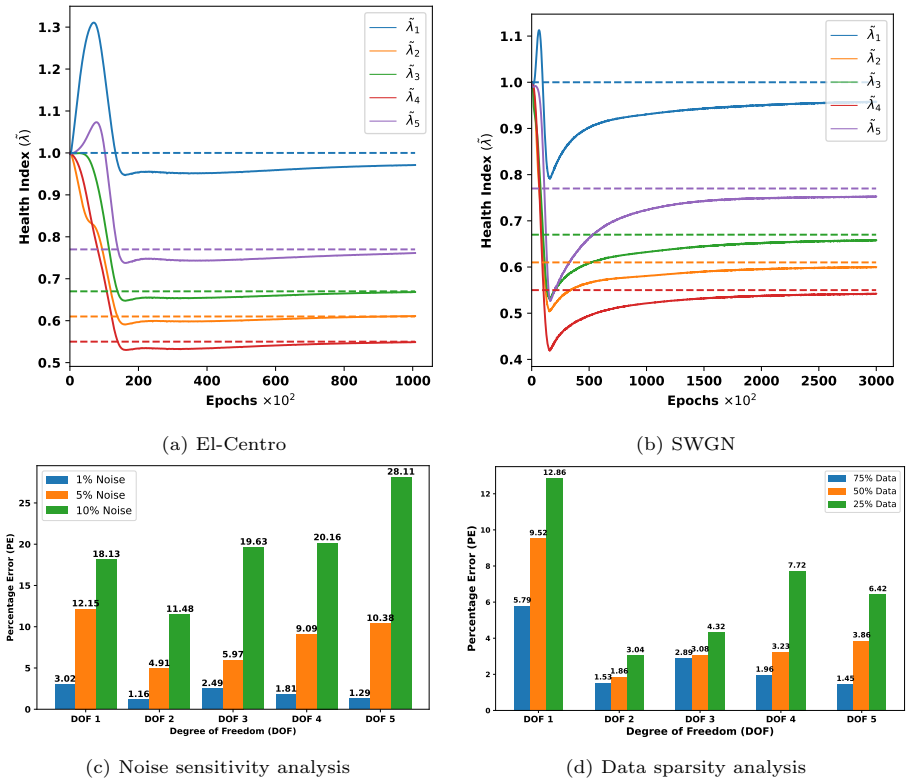
(a) 5 storey steel framed building [71] (b) Equivalent lumped mass model

Figure 9: Schematic diagram of building model for numerical experiment

The lumped spring-mass-dashpot model shown in Figure 9(b) was simulated under two distinct excitation conditions: (i) seismic excitation using the North-South component of the El Centro earthquake record from May 18, 1940 [73], and (ii) ambient vibration modeled as stochastic white Gaussian noise (SWGN), sampled at 500 Hz and applied uniformly across all nodes. To closely emulate real-world conditions, the training dataset comprised 5-second acceleration time histories, to which 1% SWGN noise was added, while 10% of the data was intentionally omitted to simulate sensor malfunctions. In this setup, the input and output layers of Pi-Attn-LSTM models were modified to accommodate 5 nodes instead of the original 4.The overall architecture of model remained consistent, only the stopping criteria set to 3×10^5 epochs for structure subjected to SWGN to handle non-linearities effectively.

As shown in Figure 10(a,b), the Pi-Attn-LSTM model demonstrates reliable parameter identification under earthquake excitation and ambient vibrations modeled as stationary white Gaussian noise (SWGN). This finding highlights the model's robustness to different types of force inputs, showing its versatility. In practical applications, structures are subjected to diverse vibration sources, not limited to seismic events. Furthermore, noise sensitivity and data sparsity analysis demonstrated strong resilience; even at an SNR of 5%, the parameter PE remained below 15%. Data sparsity analysis depicts Pi-Attn-LSTM effectiveness under severe data sparsity scenarios. The model's capacity to accurately identify parameters under varying

force and noise and data sparisity conditions strengthens its suitability for real-world structural health monitoring tasks across different environmental and operational contexts.



(a) El-Centro

(b) SWGN

(c) Noise sensitivity analysis

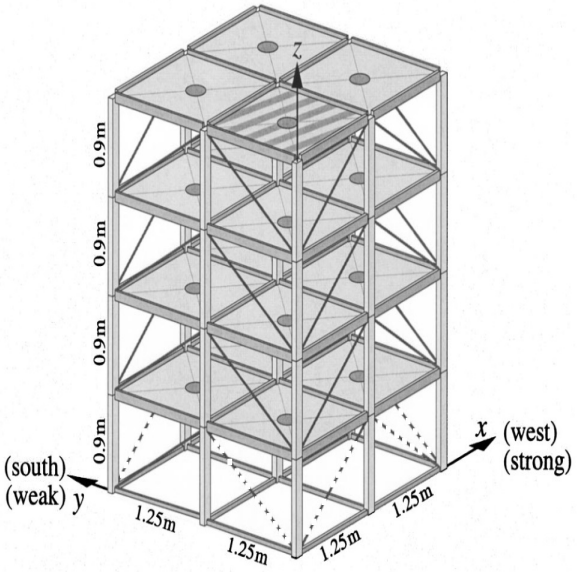
(d) Data sparsity analysis

Figure 10: Convergence and sensitivity study of health index $\tilde{\lambda}$ of 5-DOF steel building subjected to El-Centro and ambient vibration with Pi-Attn-LSTM. **(a) Health index ($\tilde{\lambda}$) evolution subjected to El-Centro ground motion:** Health index estimation ($\tilde{\lambda}_i$) over epochs using Pi-Attn-LSTM model with 90% accelerometer data and 1% noise. **(b) Health index ($\tilde{\lambda}$) evolution subjected to ambient vibration:** Health index estimation ($\tilde{\lambda}_i$) over epochs using the SWGN as force with 90% data available. **(c) Noise sensitivity analysis:** Percentage error (PE) across different degrees of freedom (DOF 1 to DOF 5) under varying noise conditions (1%, 5%, and 10% noise levels). **(d) Data sparsity analysis:** Percentage error (PE) across DOFs under varying data sparcity conditions (75%, 50%, and 25% of the data retained). Increase in data scarcity leads to error increment, with the highest errors at 25% data, indicating the model's sensitivity to data sparsity. *This figure depicts the robustness of Pi-Attn-LSTM under different loadings, noise levels and data sparsity, highlighting the impact of factors on parameter estimation accuracy.*

3.3. IASC-ASCE benchmark structure model

The proposed Pi-Attn-LSTM framework is further applied to identify structural damage in the IASC-ASCE steel-frame quarter-scale model structure which is a benchmark to test SHM algorithms. As depicted in Fig. 11, the IASC-ASCE benchmark structure is a 4-story, 2 × 2 bay steel frame model with eight bracing members per floor. The building has a total height of 3.6 m and a plan dimension of 2.5 × 2.5 m located at the Earthquake Engineering Research Laboratory of the University of British Columbia (UBC). It is constructed using hot-rolled steel with a nominal yield strength of 300 MPa. The cross-sectional properties of the scaled structural members are provided in Table 1 [72].

(a) Steel-frame quarter-scale model structure [72]



(b) Benchmark model with no stiffness in 1^{st} floor braces [72]

Figure 11: Schematic digram of benchmark model for numerical experiment

Table 1: Properties of the structure members [72]

Property	**Braces**	**Beams**	**Columns**
Section type	L25 × 25 × 3	S75 × 11	B100 × 9
Cross-section area A (m^2)	0.141×10^{-3}	1.43×10^{-3}	1.133×10^{-3}
Young's modulus E (Pa)	2×10^{11}	2×10^{11}	2×10^{11}
Moment of inertia (strong) I_x (m^4)	0	1.22×10^{-6}	1.97×10^{-6}
Moment of inertia (weak) I_y (m^4)	0	0.249×10^{-6}	0.664×10^{-6}
Torsion constant J (m^4)	0	38.2×10^{-9}	8.01×10^{-9}
Shear modulus G (Pa)	0.77×10^{11}	0.77×10^{11}	0.77×10^{11}
Mass per unit volume ρ (kg/m^3)	7800	7800	7800

The IASC-ASCE benchmark building is modeled as a 12-DOF system, represented as a shear building with 3-DOFs per floor. The floor masses in the x, y and θ directions are $\{3.45, 2.65, 2.65, 1.80\}$ tonnes, $\{3.45, 2.65, 2.65, 1.80\}$ tonnes and {3.81,2.98,2.98,2.05} tonnes, respectively. For the undamaged structure, the horizontal stiffness of each floor in the x (strong) direction is 106.60 MN/m, in y (weak) direction it is 67.90 MN/m while in θ direction it is 232.02 MN/m. In the first phase of the IASC-ASCE benchmark study, six loading scenarios were considered. For the first loading scenario first floor is assumed to have no braces as illustrated in Fig. 11(b). The structure is then subjected to El-Centro ground motion, and acceleration data is sampled at a frequency of 300 Hz for 5 seconds. The removal of the first-floor braces causes a significant reduction in stiffness: in the y-direction, the stiffness decreases from 67.90 MN/m to 19.67 MN/m, in the x-direction, it decreases from 106.60 MN/m to 58.37 MN/m and in θ-direction it reduces from 232.02 MN/m to 81.30 MN/m.

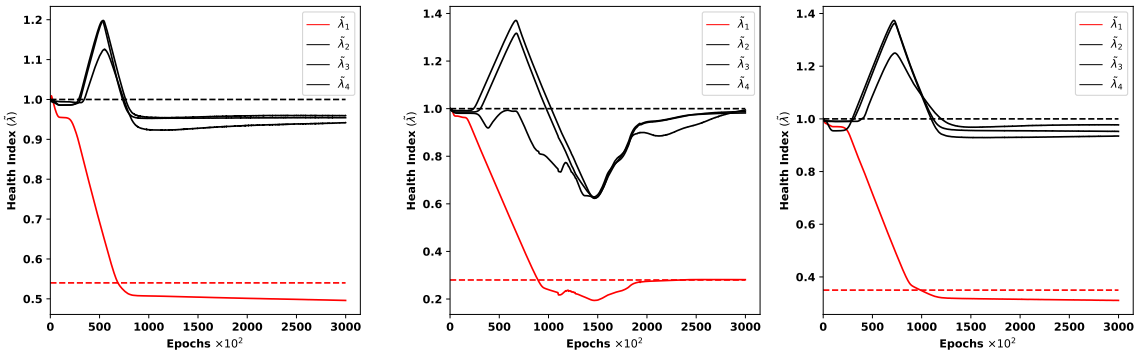
To reduce computational effort during estimation with the Pi-Attn-LSTM framework, the structure is decoupled into x, y and θ directions, treating them as independent entities. Consequently, three separate neural networks are employed to estimate the health index $\tilde{\lambda}$ for the x, y, and θ directions. The overall network topology remains unchanged, except for the input and output layers, which consist of 4 neurons each for x, y, and θ directions. Adam optimiser is used with 1×10^{-5} learning rate till reaching the maximum 3×10^5 epochs. With initial stiffness values $k_{x0,i} = 106.60$ MN/m, $k_{y0,i} = 67.90$ MN/m and $k_{\theta 0,i} = 232.02$ MN/m and by initialising $\tilde{\lambda}_i = 1$, the Pi-Attn-LSTM framework is utilised for parameter identification. Training data is comprised of 1350 acceleration data points, and it is corrupted with 1% SWGN to meet real-life scenarios. It can be observed from Figure 12 (a,b,c) that the proposed algorithm converged to a reliable health index ($\tilde{\lambda}$) in all 3 directions. Additionaly, noise sensitivity analysis and data sensitivity analysis in Figure 12(d,e) also depicts that PE in parameter estimation does not exceed 15% even with high 5% snr and only 25% data. Furthermore, five additional damage scenarios from [72] are considered, and the percentage error in each estimated health index is summarized in Table 2. The results indicate that the proposed algorithm performs robustly across various damage patterns and noise conditions on the benchmark structure.

Table 2: PE in estimated health indices under different damage patterns

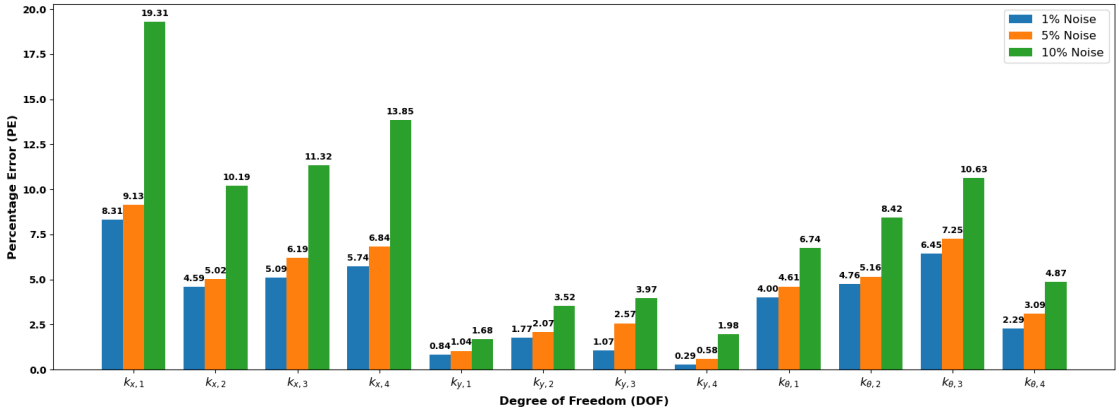
Stiffness	Damage Pattern 1	Damage Pattern 2	Damage Pattern 3	Damage Pattern 4	Damage Pattern 5	Damage Pattern 6
k_{x1}	8.31	7.13	9.25	8.74	7.89	6.43
k_{x2}	4.59	5.12	4.78	5.92	4.32	3.88
k_{x3}	5.09	6.27	5.45	4.82	5.97	4.39
k_{x4}	5.74	4.98	6.83	5.12	6.41	5.23
k_{y1}	0.84	1.35	1.76	1.01	1.52	0.79
k_{y2}	1.77	2.03	2.41	1.77	2.92	1.39
k_{y3}	1.07	1.89	1.45	2.21	1.38	0.91
k_{y4}	0.29	0.67	0.68	0.42	0.76	0.35
$k_{\theta 1}$	4.00	4.95	3.78	5.42	3.99	4.15
$k_{\theta 2}$	4.76	5.52	5.11	4.32	5.78	4.09
$k_{\theta 3}$	6.45	5.88	7.53	6.02	6.97	5.71
$k_{\theta 4}$	2.29	2.91	3.12	2.48	3.63	2.07

4. Experimental validation: Three-story shear frame

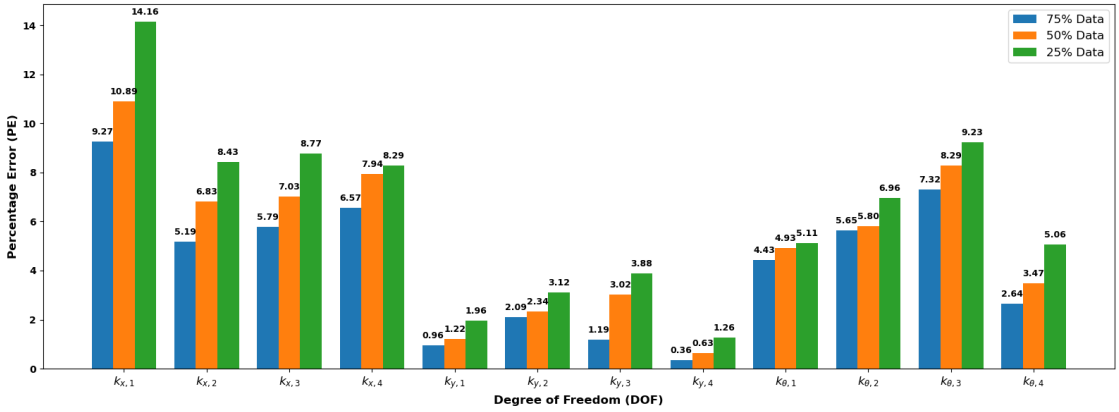
In this study, a laboratory experiment was conducted on a three-story aluminum frame, serving as a shear-building model (see Figure 13), with ambient temperatures controlled between 20°C and 25°C. The experiment aimed to assess the performance of the proposed Pi-Attn-LSTM framework on a scaled structural model under controlled conditions. Real-life structures, inherently possessing infinite degrees of freedom (*i-dof*), introduce complexities reflective of real-world conditions, where the governing system physics may be only partially understood. Estimating the health of such systems based on noisy and incomplete, discrete measurements provides a rigorous test of any proposed algorithm's robustness. In this experiment, the behavior of the *i-dof* system was inferred through the Pi-Attn-LSTM framework, supported by a simplified



(a) x-direction (b) y-direction (c) θ-direction



(d) Noise sensitivity analysis



(e) Data Sparsity analysis

Figure 12: **Health index convergence and noise sensitivity analysis:** Convergence of health indices $\hat{\lambda}_i$ of a 12-DOF IASC-ASCE benchmark model under EL-Centro ground motion using the Pi-Attn-LSTM model. Panels (a), (b), and (c) show the convergence of health indices in the x-, y-, and θ-directions, respectively, over training epochs with 90% accelerometer data and 1% noise. Panel (d) and (e) illustrates the noise sensitivity and data sparsity analysis as percentage error (PE) across x-, y-, and θ-directions for noise levels of 1%, 5%, 10%. and 75%, 50%, and 25% data availability.

physics-based conceptual model. The primary objective was to evaluate the framework's ability to converge towards a stable and accurate estimation of the system's health, thereby validating its potential for real-world structural health monitoring.

Figure 13: **Experimental Setup for Damage Detection:** (a) Experimental setup of a 3-story aluminum frame model subjected to EL-Centro earthquake excitation using a shake table. The setup includes a DEWESoft DAQ system with a 4-channel Krypton (ICP) module, accelerometers, and the frame mounted on the shake table. Panels (b) and (c) show close-up views of the frame in the undamaged and damaged conditions, respectively, highlighting the targeted damage location.

The experimental structure replicates an idealized shear building with rigid floors and lightweight columns, with input excitation and response measurements designed to approximate real-world conditions. This setup consists of a three-story, single-bay moment frame, standing 1.64 meters tall, with continuous columns of cross-section 25 mm $\times$ 4 mm. Each floor is represented by aluminum plates (300 mm $\times$ 150 mm $\times$ 4 mm) bolted to the columns, modeling a lumped mass system. The base of the frame is secured to a shake table using bolts and lock washers, providing a stable, rigid connection. The floors are spaced at intervals, giving a nominal story height of 0.41 meters (410 mm).

The experimental assembly is mounted on a unidirectional shake table that simulates ground motion. The shake table platform, measuring 1000 mm $\times$ 1000 mm, can accommodate loads up to 400 kg, generating base excitations with a maximum acceleration of 1 g and a maximum displacement stroke of $\pm$100 mm. The frequency of input waveforms can range from 0 to 15 Hz. Due to the rigid floor connections, the columns at each level exhibit fixed boundary conditions, allowing calculation of the equivalent stiffness and enabling the use of a 3-*dof* reduced-order lumped mass model. Since the mass at each floor level significantly exceeds that of the columns, a lumped-mass assumption at each floor is appropriate. For data collection, four uni-axial Integrated Circuit Piezoelectric (ICP) accelerometers are placed at each floor level, including the ground level. Earthquake excitation is applied to the frame through a servohydraulic actuator controlled via a control panel on the shake table. Data from the accelerometers is sampled at 100 Hz using a 4-channel data acquisition (DAQ) system, with DewesoftX software employed for real-time viewing and recording. Detailed

Table 3: Comparison of modal frequencies between the experimental (real undamaged) model and the calibrated numerical model. The table presents the first three modal frequencies (ω_1, ω_2, ω_3) in Hz, alongside their corresponding relative errors. The relative error illustrates the accuracy of the numerical model calibration in replicating the real system's dynamic properties.

Frequency (Hz)	**Experimental**	**Numerical**	**Percentage error**
ω_1	3.04	3.05	0.32
ω_2	8.39	8.55	1.90
ω_3	12.69	12.35	2.67

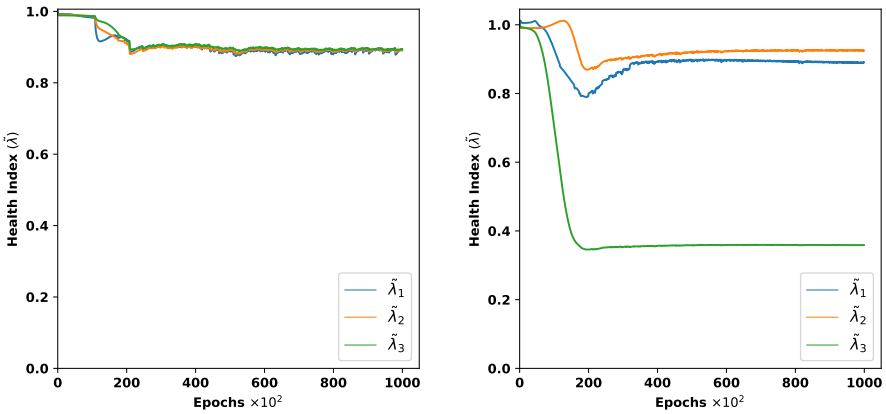
specifications of instruments can be found in [75]

To integrate the known physics of the real frame structure into the estimation algorithm, it is essential to develop a numerical model that accurately replicates the experimental structure's behavior. This involves initially creating a 3-*dof* lumped mass model of the system and calibrating the properties of aluminum material so that the dynamic characteristics of the model, such as natural frequency, are closely aligned with those of the real structure. This calibration step ensures that the initial health index values are near their true values, thus reducing the computational effort required for stable convergence in the Pi-Attn-LSTM-based estimation process. Consequently, the undamaged frame, mounted on a shake table, is subjected to earthquake excitation (El-Centro earthquake) using a servo-hydraulic actuator. The time-domain data is converted to the frequency domain using the Frequency Domain Decomposition (FDD) approach, allowing for the extraction of experimental natural frequencies.

Aluminum frame properties were tuned to align the experimental modal properties with the numerical ones. This tuning resulted in a mass density of $\rho = 2700$ kg/m^3 and a modulus of elasticity of $E = 68$ GPa. Based on these assumptions, the calibrated values for mass and stiffness in the numerical model were calculated as $m_1 = m_2 = m_3 = 1.70$ kg and $k_1 = k_2 = k_3 = 3.15$ kN/m, respectively. The damping coefficients were derived using the Rayleigh damping criteria with a damping ratio of 1%. Table 3 presents a comparison between the modal parameters from the real experiment and the numerical model. With the calibrated stiffness values denoted as $k_{0,i}$ and the prepared support model, the Pi-Attn-LSTM framework was used to estimate the health indices of the test structure. In this approach, the calibrated story stiffnesses were assumed to represent the undamaged state, with the health indices λ initially set at unity. The architecture of the LSTM model consists of an LSTM cell with 3 input neurons, followed by the same encoder-attention-decoder architecture with 3 output neurones as briefed in previous sections.

The estimation process yielded convergent health indices that reflect the healthy state of the structure. The updated modal properties, compared to experimental and initial values, are shown in Table 3. The estimated health indices ($\tilde{\lambda}$) are illustrated in Figure 14(a). Contrary to the initial assumption, $\tilde{\lambda}$ values for each story level did not converge to 1, this deviation can be largely attributed to simplistic assumptions made during modeling and also in the boundary conditions. Therefore, the updated stiffness of a healthy frame can be considered as $k_1 = 2.874$ kN/m, $k_2 = 2.861$ kN/m, and $k_3 = 2.867$ kN/m.

The Pi-Attn-LSTM algorithm's effectiveness in detecting structural damage was validated by analyzing response data from a test structure with damage. Damage was introduced by loosening bolts on diagonal columns at the third story, reducing stiffness to simulate a compromised state. For estimation, structural parameters were maintained as same as the initial numerical model assuming no prior information on the health state is available for the system, and measurement data from this damaged setup were processed in the Pi-Attn-LSTM framework. As shown in Figure 14(b), and Table 4 this induced damage resulted in a 60.75% reduction in third-floor stiffness, while estimation errors for k_1 and k_2 remained below 10% when compared with estimates obtained for updated healthy numerical model, supporting that the prior numerical assumptions do not affect results much. Additionally, frequency changes between experimental and numerical data, detailed in Table 5, further confirm the model's accuracy. These results highlight the Pi-Attn-LSTM's strength in structural health assessment and its potential applicability to real-world SHM tasks.



(a) Healthy (b) Damaged

Figure 14: Convergence of health index $\tilde{\lambda}$ using Pi-Attn-LSTM framework for 3-DOF aluminium frame subjected to El-Centro ground motion. **(a) Health index $\tilde{\lambda}$ evolution under healthy condition:** Convergence plots showing oscillations around 0.92 as the measurement data being noisy, and **(b) Health index $\tilde{\lambda}$ evolution under damaged condtion:** Health index $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ converged near to 0.92 while $\tilde{\lambda}_3$ showed substantial reduction depicting the damage condition.

Table 4: Comparison of story stiffness values in initial and estimated (healthy and damaged) states for an aluminium frame. The table presents the numerical values of stiffness for each story level (k_1, k_2, and k_3) under two conditions: the estimated (healthy) state and the artificially induced damaged state. The % change in stiffness acts as an indicator of the severity of the damage at each story.

Stiffness (KN/m)	**Numerical (Initial)**	**Estimated (Healthy)**	**Estimated (Damaged)**	**% change in stiffness**
k_1	3.15	2.874	2.641	8.10
k_2	3.15	2.861	2.913	1.81
k_3	3.15	2.867	1.125	**60.75**

Table 5: Comparison of modal frequencies between experimental and numerical measurements for the damaged condition. The table lists the first three modal frequencies (ω_1, ω_2, ω_3) in Hz, along with their respective relative errors. The relative error values indicate the accuracy of the Pi-Attn-LSTM in capturing the dynamic characteristics of the structure under the damaged state.

Frequency (Hz)	**Experimental**	**Numerical**	**Percentage error**
ω_1	2.88	2.78	3.47
ω_2	7.46	7.16	4.02
ω_3	12.18	11.63	4.51

5. Conclusion

This study presents a novel Pi-Attn-LSTM model, an advancement over the Pi-LSTM model for parameter estimation in civil engineering structures. The inclusion of an attention mechanism in the LSTM encoder-decoder structure enables it to adaptively identify important features in temporal data, thus ultimately enhancing its parameter estimation accuracy. Evaluation of the proposed Pi-Attn-LSTM was done with 3 numerical and 1 experimental studies. Numerical studies included a 4-story steel frame, a 5-story numerical steel building, and an IASC-ASCE benchmark model to assess the model's robustness against varying noise levels, an initialization scheme, input forces, and temporal sparsity scenarios. Results demonstrated that Pi-Attn-LSTM outperformed conventional Pi-LSTM, achieving lower estimation errors (<10% under 5% noise) and faster convergence by dynamically prioritizing relevant time steps. Notably, the model exhibited good resilience to data sparsity, maintaining accuracy even with only 25% of data. This adaptability is critical for real-world applications where data acquisition is often compromised by sensor failures or economic constraints. An experimental validation was conducted using a three-story aluminium frame subjected to earthquake excitation to verify the efficacy of Pi-Attn-LSTM. The proposed framework accurately estimated the frame's health indices, under undamaged and intentionally induced damage conditions, demonstrating its ability to localize and quantify stiffness deterioration. By combining acceleration measurements with a simplified physics-based model, Pi-Attn-LSTM successfully identified localized stiffness reductions (e.g., 60.75% degradation at the third story) despite noisy data and idealized assumptions. The alignment between numerical predictions and experimental modal properties (errors <5%) highlighted the model's ability to bridge gaps between theoretical models and real-system behavior, a recurring challenge in SHM. This experimental validation supports the framework's applicability to real-world SHM tasks, highlighting its potential as a robust and adaptable tool for structural damage detection and assessment. Overall, the Pi-Attn-LSTM framework significantly advances physics-informed deep learning in SHM by addressing key limitations in traditional deep learning models, such as dependency on extensive data, sensitivity to measurement noise, and limitations under data sparsity. Its ability to leverage both physics-based constraints and data-driven learning, combined with the flexibility of the attention mechanism, positions Pi-Attn-LSTM as a powerful tool for enhancing the accuracy and resilience of structural health assessments in civil engineering applications.

Future work could explore extending this framework to a wider variety of structural types such as bridges, aircraft components, and wind turbines subjected to different environmental conditions, further enhancing its adaptability and robustness in SHM applications. Another promising direction involves addressing the prevalent issue of missing input force information in structural systems; making the framework resilient to unknown or unmeasured input forces would broaden its applicability in scenarios where force data is unavailable or incomplete, thus utilizing output-only data for SHM. A further extension could incorporate Bayesian approaches within the Pi-Attn-LSTM framework, enabling a Bayesian physics-informed neural network or neural network embedded in the Bayesian framework that accounts for uncertainty. Such an approach would allow the model to provide probabilistic estimates, offering a more comprehensive and uncertainty-aware solution for structural health monitoring. These extensions collectively would expand Pi-Attn-LSTM's utility, paving the way for more resilient, adaptable, and precise structural health monitoring applications.

6. Funding

This study was funded by Science & Engineering Research Board (SERB), New Delhi, India, through grant file No. $CRG/2023/005267$

7. Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT to correct grammatical mistakes. After using ChatGPT, authors reviewed and edited the content as needed and takes full responsibility for the content of the published article.

References

[1] Michèle Basseville, Laurent Mevel, and Maurice Goursat. Statistical model-based damage detection and localization: subspace-based residuals and damage-to-noise sensitivity ratios. *Journal of sound and vibration*, 275(3-5):769–794, 2004.

[2] Michele Basseville, Albert Benveniste, Maurice Goursat, Luc Hermans, Laurent Mevel, and Herman Van der Auweraer. Output-only subspace-based structural identification: from theory to industrial testing practice. *Journal of Dynamic Systems Measurement and Control-Transactions of the Asme*, 123(4), 2001.

[3] Michael Döhler, Laurent Mevel, and Falk Hille. Subspace-based damage detection under changes in the ambient excitation statistics. *Mechanical Systems and Signal Processing*, 45(1):207–224, 2014.

[4] Smriti Sharma and Subhamoy Sen. One-dimensional convolutional neural network-based damage detection in structural joints. *Journal of Civil Structural Health Monitoring*, 10(5):1057–1072, 2020.

[5] Rune Brincker, Lingmi Zhang, and Palle Andersen. Modal identification of output-only systems using frequency domain decomposition. *Smart materials and structures*, 10(3):441, 2001.

[6] Jer-Nan Juang and Richard S Pappa. An eigensystem realization algorithm for modal parameter identification and model reduction. *Journal of guidance, control, and dynamics*, 8(5):620–627, 1985.

[7] Smriti Sharma and Subhamoy Sen. Bridge damage detection in presence of varying temperature using two-step neural network approach. *Journal of Bridge Engineering*, 26(6):04021027, 2021.

[8] Adel Belouchrani, Karim Abed-Meraim, J-F Cardoso, and Eric Moulines. A blind source separation technique using second-order statistics. *IEEE Transactions on signal processing*, 45(2):434–444, 1997.

[9] Gaëtan Kerschen, Fabien Poncelet, and J-C Golinval. Physical interpretation of independent component analysis in structural dynamics. *Mechanical Systems and Signal Processing*, 21(4):1561–1575, 2007.

[10] Yongchao Yang and Satish Nagarajaiah. Time-frequency blind source separation using independent component analysis for output-only modal identification of highly damped structures. *Journal of Structural Engineering*, 139(10):1780–1793, 2013.

[11] Satish Nagarajaiah and Yongchao Yang. Modeling and harnessing sparse and low-rank data structure: a new paradigm for structural dynamics, identification, damage detection, and health monitoring. *Structural Control and Health Monitoring*, 24(1):e1851, 2017.

[12] Jann N Yang, Silian Lin, Hongwei Huang, and Li Zhou. An adaptive extended kalman filter for structural damage identification. *Structural Control and Health Monitoring: The Official Journal of the International Association for Structural Control and Monitoring and of the European Association for the Control of Structures*, 13(4):849–867, 2006.

[13] Eleni N Chatzi and Andrew W Smyth. The unscented kalman filter and particle filter methods for nonlinear structural system identification with non-collocated heterogeneous sensing. *Structural Control and Health Monitoring: The Official Journal of the International Association for Structural Control and Monitoring and of the European Association for the Control of Structures*, 16(1):99–123, 2009.

[14] AE Charalampakis and CK Dimou. Identification of bouc–wen hysteretic systems using particle swarm optimization. *Computers & structures*, 88(21-22):1197–1205, 2010.

[15] Saeed Eftekhar Azam, Eleni Chatzi, and Costas Papadimitriou. A dual kalman filter approach for state estimation via output-only acceleration measurements. *Mechanical systems and signal processing*, 60:866–886, 2015.

[16] Neha Aswal, Subhamoy Sen, and Laurent Mevel. Switching kalman filter for damage estimation in the presence of sensor faults. *Mechanical Systems and Signal Processing*, 175:109116, 2022.

[17] Neha Aswal, Baidurya Bhattacharya, and Subhamoy Sen. Joint and dual estimation of states and parameters with extended and unscented kalman filters. In *Recent Developments in Structural Health Monitoring and Assessment–Opportunities and Challenges: Bridges, Buildings and Other Infrastructures*, pages 223–252. World Scientific, 2022.

[18] Neha Aswal, Subhamoy Sen, and Laurent Mevel. Estimation of local failure in tensegrity using interacting particle-ensemble kalman filter. *Mechanical Systems and Signal Processing*, 160:107824, 2021.

[19] Eshwar Kuncham, Neha Aswal, Subhamoy Sen, and Laurent Mevel. Bayesian monitoring of substructures under unknown interface assumption. *Mechanical Systems and Signal Processing*, 193:110269, 2023.

[20] Neha Aswal, Subhamoy Sen, and Laurent Mevel. Strain-based joint damage estimation approach robust to unknown non-stationary input force. *Structural Control and Health Monitoring*, 29(10):e2999, 2022.

[21] Md Sazzad Hossain, Zhi Chao Ong, Zubaidah Ismail, Siamak Noroozi, and Shin Yee Khoo. Artificial neural networks for vibration based inverse parametric identifications: A review. *Applied Soft Computing*, 52:203–219, 2017.

[22] Calebe Paiva Gomes de Souza, Paulo Roberto Gardel Kurka, Romulo Goncalves Lins, and Jose Medeiros de Araujo Junior. Performance comparison of non-adaptive and adaptive optimization algorithms for artificial neural network training applied to damage diagnosis in civil structures. *Applied Soft Computing*, 104:107254, 2021.

[23] Dehao Liu and Yan Wang. Multi-fidelity physics-constrained neural network and its application in materials modeling. *Journal of Mechanical Design*, 141(12), 2019.

[24] Bin Xu, Zhishen Wu, Genda Chen, and Koichi Yokoyama. Direct identification of structural parameters from dynamic responses with neural networks. *Engineering Applications of Artificial Intelligence*, 17(8):931–943, 2004.

[25] Ziming Wen, Yu Li, Hu Wang, and Yong Peng. Data-driven spatiotemporal modeling for structural dynamics on irregular domains by stochastic dependency neural estimation. *Computer Methods in Applied Mechanics and Engineering*, 404: 115831, 2023.

[26] Bowen Xu and Xinjiang Lu. A data-driven spatiotemporal model predictive control strategy for nonlinear distributed parameter systems. *Nonlinear Dynamics*, 108(2):1269–1281, 2022.

[27] Yinlai Jiang, Isao Hayashi, and Shuoyu Wang. Knowledge acquisition method based on singular value decomposition for human motion analysis. *IEEE Transactions on Knowledge and Data Engineering*, 26(12):3038–3050, 2014.

[28] Gilberto Fernandes Jr, Joel JPC Rodrigues, and Mario Lemes Proenca Jr. Autonomous profile-based anomaly detection system using principal component analysis and flow analysis. *Applied Soft Computing*, 34:513–525, 2015.

[29] Zhen Liu and Han-Xiong Li. A spatiotemporal estimation method for temperature distribution in lithium-ion batteries. *IEEE Transactions on Industrial Informatics*, 10(4):2300–2307, 2014.

[30] YC Liang, DP Feng, and JE Cooper. Identification of restoring forces in non-linear vibration systems using fuzzy adaptive neural networks. *Journal of sound and vibration*, 242(1):47–58, 2001.

[31] Luca Facchini, Michele Betti, and Paolo Biagini. Neural network based modal identification of structural systems through output-only measurement. *Computers & Structures*, 138:183–194, 2014.

[32] Xing Lü, Liang-Li Zhang, and Wen-Xiu Ma. Oceanic shallow-water description with (2+ 1)-dimensional generalized variable-coefficient hirota–satsuma–ito equation: Painlevé analysis, soliton solutions, and lump solutions. *Physics of Fluids*, 36(6), 2024.

[33] Di Gao, Wen-Xiu Ma, and Xing Lü. Wronskian solution, bäcklund transformation and painlevé analysis to a (2+ 1)-dimensional konopelchenko–dubrovsky equation. *Zeitschrift für Naturforschung A*, 79(9):887–895, 2024.

[34] Liang-Li Zhang, Xing Lü, and Sheng-Zhi Zhu. Painlevé analysis, bäcklund transformation and soliton solutions of the (2+ 1)-dimensional variable-coefficient boussinesq equation. *International Journal of Theoretical Physics*, 63(7):160, 2024.

[35] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

[36] Yu Zhang and Xing Lü. Data-driven solutions and parameter discovery of the extended higher-order nonlinear schrödinger equation in optical fibers. *Physica D: Nonlinear Phenomena*, 468:134284, 2024.

[37] Che Han and Xing Lü. Variable coefficient-informed neural network for pde inverse problem in fluid dynamics. *Physica D: Nonlinear Phenomena*, 472:134362, 2025.

[38] Ehsan Haghighat, Maziar Raissi, Adrian Moure, Hector Gomez, and Ruben Juanes. A physics-informed deep learning framework for inversion and surrogate modeling in solid mechanics. *Computer Methods in Applied Mechanics and*

Engineering, 379:113741, 2021.
[39] Xiaowei Jin, Shengze Cai, Hui Li, and George Em Karniadakis. Nsfnets (navier-stokes flow nets): Physics-informed neural networks for the incompressible navier-stokes equations. *Journal of Computational Physics*, 426:109951, 2021.
[40] Zan Yang, Dan Li, Wei Nai, Lu Liu, Jingjing Sun, and Xiaowei Lv. Compressible non-newtonian fluid based road traffic flow equation solved by physical-informed rational neural network. *IEEE Access*, 12:12992–13009, 2024.
[41] Jan Oldenburg, Finja Borowski, Wiebke Wollenberg, Alper Öner, Klaus-Peter Schmitz, and Michael Stiehm. Augmentation of experimentally obtained flow fields by means of physics informed neural networks (pinn) demonstrated on aneurysm flow. In *Current Directions in Biomedical Engineering*, volume 9, pages 519–523. De Gruyter, 2023.
[42] Peng Lan, Jing-jing Su, Xin-yan Ma, and Sheng Zhang. Application of improved physics-informed neural networks for nonlinear consolidation problems with continuous drainage boundary conditions. *Acta Geotechnica*, 19(1):495–508, 2024.
[43] Cheng-Ze Li, Xing Lü, Jia-Jun Gong, and Yu Lei. Extended seir model of covid-19 spread focusing on compartmental flow in england. *Nonlinear Dynamics*, 113(1):971–988, 2025.
[44] Che Han and Xing Lü. Novel patterns in the space variable fractional order gray–scott model. *Nonlinear Dynamics*, 112 (18):16135–16151, 2024.
[45] Feng Cao, Xing Lü, Yi-Xuan Zhou, and Xi-Yu Cheng. Modified seiar infectious disease model for omicron variants spread dynamics. *Nonlinear Dynamics*, 111(15):14597–14620, 2023.
[46] Xue Peng, Yi-Wei Zhao, and Xing Lü. Data-driven solitons and parameter discovery to the (2+ 1)-dimensional nlse in optical fiber communications. *Nonlinear Dynamics*, 112(2):1291–1306, 2024.
[47] Björn Lütjens, Catherine H Crawford, Mark Veillette, and Dava Newman. Pce-pinns: Physics-informed neural networks for uncertainty propagation in ocean modeling. *arXiv preprint arXiv:2105.02939*, 2021.
[48] Sarvin Moradi, Burak Duran, Saeed Eftekhar Azam, and Massood Mofid. Novel physics-informed artificial neural network architectures for system and input identification of structural dynamics pdes. *Buildings*, 13(3):650, 2023.
[49] Tong Liu and Hadi Meidani. Physics-informed neural networks for system identification of structural systems with a multiphysics damping model. *Journal of Engineering Mechanics*, 149(10):04023079, 2023.
[50] Zhilu Lai, Charilaos Mylonas, Satish Nagarajaiah, and Eleni Chatzi. Structural identification with physics-informed neural ordinary differential equations. *Journal of Sound and Vibration*, 508:116196, 2021.
[51] Aditi Krishnapriyan, Amir Gholami, Shandian Zhe, Robert Kirby, and Michael W Mahoney. Characterizing possible failure modes in physics-informed neural networks. *Advances in neural information processing systems*, 34:26548–26560, 2021.
[52] Levi McClenny and Ulisses Braga-Neto. Self-adaptive physics-informed neural networks using a soft attention mechanism. *arXiv preprint arXiv:2009.04544*, 2020.
[53] Olga Fuks and Hamdi A Tchelepi. Limitations of physics informed machine learning for nonlinear two-phase transport in porous media. *Journal of Machine Learning for Modeling and Computing*, 1(1), 2020.
[54] Klaus Greff, Rupesh K Srivastava, Jan Koutník, Bas R Steunebrink, and Jürgen Schmidhuber. Lstm: A search space odyssey. *IEEE transactions on neural networks and learning systems*, 28(10):2222–2232, 2016.
[55] Ruiyang Zhang, Yang Liu, and Hao Sun. Physics-informed multi-lstm networks for metamodeling of nonlinear structures. *Computer Methods in Applied Mechanics and Engineering*, 369:113226, 2020.
[56] Shubhendu Kumar Singh, Ruoyu Yang, Amir Behjat, Rahul Rai, Souma Chowdhury, and Ion Matei. Pi-lstm: Physics-infused long short-term memory network. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 34–41. IEEE, 2019.
[57] Fangyu Liu, Junlin Li, and Linbing Wang. Pi-lstm: Physics-informed long short-term memory network for structural response modeling. *Engineering Structures*, 292:116500, 2023.
[58] Yangyang Liao, Hesheng Tang, Rongshuai Li, Lingxiao Ran, and Liyu Xie. Seismic response prediction and parameters estimation of the frame structure equipped with the base isolation-fluid inerter system (fs-bifi) based on the pi-lstm model. *Engineering Structures*, 309:118077, 2024.
[59] Dzmitry Bahdanau. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.
[60] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
[61] Yude Wang, Jie Zhang, Meina Kan, Shiguang Shan, and Xilin Chen. Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12275–12284, 2020.
[62] Yao Qin, Dongjin Song, Haifeng Chen, Wei Cheng, Guofei Jiang, and Garrison Cottrell. A dual-stage attention-based recurrent neural network for time series prediction. *arXiv preprint arXiv:1704.02971*, 2017.
[63] Youru Li, Zhenfeng Zhu, Deqiang Kong, Hua Han, and Yao Zhao. Ea-lstm: Evolutionary attention-based lstm for time series prediction. *Knowledge-Based Systems*, 181:104785, 2019.
[64] Yuchen Liao, Rong Lin, Ruiyang Zhang, and Gang Wu. Attention-based lstm (attlstm) neural network for seismic response modeling of bridges. *Computers & Structures*, 275:106915, 2023.
[65] Qingyu Zhang, Maozu Guo, Lingling Zhao, Yang Li, Xinxin Zhang, and Miao Han. Transformer-based structural seismic response prediction. In *Structures*, volume 61, page 105929. Elsevier, 2024.
[66] Saifon Chaturantabut and Danny C Sorensen. Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific Computing*, 32(5):2737–2764, 2010.
[67] I Sutskever. Sequence to sequence learning with neural networks. *arXiv preprint arXiv:1409.3215*, 2014.
[68] Levi D McClenny and Ulisses M Braga-Neto. Self-adaptive physics-informed neural networks. *Journal of Computational Physics*, 474:111722, 2023.

[69] Dehao Liu and Yan Wang. A dual-dimer method for training physics-constrained neural networks with minimax architecture. *Neural Networks*, 136:112–125, 2021.

[70] Charles DeVore, Zhaoshou Jiang, Richard E Christenson, Gannon Stromquist-LeVoir, and Erik A Johnson. Experimental verification of substructure identification for damage detection in shear buildings. *Journal of Engineering Mechanics*, 142 (1):04015060, 2016.

[71] Xin-Yu Guo and Sheng-En Fang. Structural parameter identification using physics-informed neural networks. *Measurement*, 220:113334, 2023.

[72] Erik A Johnson, Heung-Fai Lam, Lambros S Katafygiotis, and James L Beck. Phase i iasc-asce structural health monitoring benchmark problem using simulated data. *Journal of engineering mechanics*, 130(1):3–15, 2004.

[73] vibrationdata. URL `https://www.vibrationdata.com/elcentro.htm`.

[74] Marcus Haywood-Alexander, Giacomo Arcieri, Antonios Kamariotis, and Eleni Chatzi. Response estimation and system identification of dynamical systems via physics-informed neural networks. *arXiv preprint arXiv:2410.01340*, 2024.

[75] Md Armanul Hoda, Eshwar Kuncham, and Subhamoy Sen. Response and input time history dataset and numerical models for a miniaturized 3d shear frame under damaged and undamaged conditions. *Data in Brief*, 45:108692, 2022.