Model-based Damage Detection through Physics-guided Learning for Dynamic Systems

Ali I. Ozdagli and Xenofon Koutsoukos

Department of Computer Science, Vanderbilt University, Nashville, TN 37212 USA
Ali.I.Ozdagli@vanderbilt.edu
Xenofon.Koutsoukos@vanderbilt.edu

Abstract

Data-driven learning approaches have gained a lot of interest in evaluating and validating complex dynamic systems. One of the main challenges for developing a reliable learning model is the lack of training data covering a large range of various operational conditions. Extensive training data can be generated using a physics-based simulation model. On the other hand, the learning algorithm should be still tested with experimental data obtained from the actual system. Modeling errors may lead to a statistical divergence between the simulation training data and the experimental testing data, causing poor performance, especially for domain-agnostic black-box learning methods. To close the gap between the simulation and experimental domains, this paper proposes a physics-guided learning approach that combines the power of neural network with domain-specific physics knowledge. Specifically, the proposed architecture integrates physical parameters extracted from physics-based simulation data into the intermediate layers of the neural network to constrain the learning process. To demonstrate the effectiveness of the proposed approach, the architecture is adopted to a damage classification problem for a three-story structure. Our results show that the accuracy for localizing the damage correctly based on experimental data improves significantly over black-box models. We also use the physics-based intermediate layers to analyze the interpretability of the results.

1. Introduction

In the last decade, the use of machine learning (ML) algorithms gained a lot of interest within the community of condition monitoring for dynamic systems (Widodo & Yang, 2007; Farrar & Worden, 2012; Stetco et al., 2019). A majority of ML applications in this area exercise a data-driven black-box approach that utilizes a large volume of experimental data obtained directly from the actual dynamic system. Black-box methods are proven to be successful in diagnosing the system through characterization and localization of the damage (Bakhary et al., 2007). One of the obstacles for such methods is often the availability of sufficient training data (Z. Zhang & Sun, 2020). More specifically, access to a complete training dataset covering a wide range of conditions is costly and in some instances impossible without actually damaging the system prior to operation. This problem is a major roadblock in developing efficient data-driven algorithms for diagnostics of dynamic systems (Sadoughi & Hu, 2019).

For cases where training data captured from the field is limited, a data-driven black-box ML model could be trained with simulation data. In other words, to compensate for the lack of experimental training data, a representative analytical model can simulate the behavior of the system physics to some degree. While physics-based analytical models are capable of generating extensive training dataset, the resulting ML algorithm should still be evaluated with experimental testing data. Well-established analytical models are capable of simulating the dynamic response of the target system (Teughels & De Roeck, 2005; Jaishi & Ren, 2006). On the other hand, calibrating a large set of parameters for complex systems to achieve accurate physical behavior is often computationally exhaustive and at times infeasible (R. Zhang et al., 2020). Eventually, the analytical representation inherits modeling error. In this case, it is expected that the ML algorithm will fail to perform efficiently during testing since the simulation training data and experimental testing data are statistically divergent (Gardner et al., 2020). To address this drawback of data-driven black-box algorithms, the inference should incorporate domain-specific physical knowledge. The physics-guided learning (PGL) which is essentially a hybrid approach aggregating data-driven inference with physical parameters has the potential to leverage the performance of the condition monitoring further and to bridge the gap between simulation and experimental domains.

In recent years, a number of PGL approaches have been proposed (Karpatne et al., 2017; Jia et al., 2018; Sadoughi & Hu, 2019; R. Zhang et al., 2020; Z. Zhang & Sun, 2020; Yao et al., 2020). However, the variety of applications implies that the implementation of a proper PGL with domain-specific knowledge is non-trivial. Moreover, most of the existing work focuses on the prediction of system responses using PGL. In the area of diagnostics, little research effort has been devoted to incorporating of physical knowledge into the data-driven ML and exploiting deep learning architectures for damage classification.

In this paper, we propose a damage detection and localization architecture for dynamic systems, namely, physics-guided learning for structural health monitoring (PGL4SHM) that combines the power of neural networks with domain specific physics knowledge. In particular, PGL4SHM is a multi-task deep learning architecture which (i) utilizes the synthetic data simulated by a numerical physics-based representation of the target structure for training and (ii) incorporates domain-specific physical parameters derived from this representation into the loss function. The multi-tasking PGL4SHM is trained with simulated structural responses in time-series form which serve as the input to the deep network. Additionally, during the learning phase, the physical parameters such as natural frequencies and mode shapes that are known to be structural damage indicators (Kim et al., 2003) are used for training the intermediate layers of PGL4SHM (Muralidhar et al., 2019). The modal features (natural frequencies and mode shapes) can be extracted directly from the numerical model. Since the organic relationship between the physics-based model representation, structural responses, and damage state is embedded into the PGL4SHM, the architecture is capable of generalizing damage detection compared to black-box approaches. As PGL4SHM uses modal features derived from numerical model, this embedding is physics-based rather empirical.

To validate the PGL4SHM architecture, two case studies are considered. The first study is purely analytical and investigates the performance and efficiency of the proposed approach under ideal conditions since every aspect of the system is simulated. The second case study considers the experimental setup of a small-scale three-story structure tested at Los Alamos National Lab. The results from both the analytical and experiential case studies show that the proposed architecture has better generalization capability in localizing the damage compared to black-box models in the presence of modeling errors. The performance gain is more evident when the numerical representation deviates from the actual structure further. Lastly, we evaluated the explainability of the results by analyzing the relationship between structural responses, damage state, and the integrated physical parameters.

Overall, PGL4SHM combines data-driven machine learning with knowledge of physics. As a result of this, PGL4SHM has the potential to improve damage detection and localization for SHM applications and promises more accurate decisions and prioritization for maintenance operations.

In summary, the major contributions of this paper are:

- A physics-guided learning architecture, PGL4SHM is proposed to generalize damage detection and location prediction for dynamic systems.
- The proposed architecture uses physics-constrained intermediate variable layers that rely on physical parameters known to be statistically important features for damage detection such as natural frequencies and mode shapes.
- For proof of concept, the proposed method is evaluated by comparing damage localization performance to black-box models for numerical and experimental cases.Results show that the new approach improves prediction accuracy in the presence of modeling error.
- PGL4SHM improves the explainability of the results since the intermediate layers expose valuable information that is highly relevant to the physics of the target structure.

2. PROBLEM FORMULATION

Structural systems can experience damage throughout their life-cycle. It is essential to detect and locate the damage earl-on before it progresses to a bigger failure. In this context, damage localization is a supervised classification problem.

In this study, we consider black-box deep neural networks that label raw input data in the form of acceleration time-series measurements according to the damage condition the structure is experiencing. The implicit assumption for the black-box model is that the training data is available for every expected damage condition. In reality, the training data is only available for no damage condition. To obtain training data for other damage conditions, the structure should be deliberately tarnished which is not practical. We can create a physics-based model of the structure and generate simulated data for various damage scenarios of interest. Accordingly, the black-box model can be trained with the simulated data and tested with experimental data after deployment. However, this approach is often not feasible since the physics-based simulation often has intrinsic modeling error. Due to the deviation between simulated training and experimental testing data, the black-box model will become ineffective in labeling the input correctly.

The problem considered in this paper is localizing the damage accurately in the presence of modeling errors. To address this problem, three challenges should be resolved. First, we need to create a physics-based representation of the target system based on the available data. This model should be used to generate simulated data. Secondly, we should design a deep learning architecture that is trained with simulated data but can generalize well for experimental data compared to the black-box model. Thirdly, we need to establish the physical parameters most relevant to the damage condition of the structure to integrated into intermediate layers of the architecture during inference. Ultimately, successful implementation of PGL4SHM should generalize the prediction well for the cases where field data is limited and physics-based simulation has some modeling error.

3. PHYSICS-GUIDED LEARNING

For a given set of structural response measurements in the time domain, we are interested in predicting the damage condition of the structure. Such a predictor can be trained with a supervised learning approach since for each input, x there is a label, y corresponding to the damage state. One way to learn the mapping from x to y is by training a black-box feed-forward neural network. By utilizing nonlinear activation functions within neurons, this network allows us to expose the complex relationship between the structural responses and the damage conditions. While the black-box networks are capable of learning the latent feature space, they can fail to generalize the predictions for unseen observations.

For many dynamic systems, the data labeled as *normal* is often available when they are deployed. However, access to data relevant to damage conditions is limited without harming the system. The absence of experimental data can be compensated by simulating damage on a finite element model of the system and obtaining new input/output pairs. However, due to the poor generalization of black-box models, the predictor will suffer from the presence of modeling errors and label the given inputs incorrectly. To address the limitation of the black-box models, this paper presents the physics-guided learning for structural health monitoring (PGL4SHM) which integrates the physical knowledge regarding the dynamic characteristics of the target structure into the deep learning architecture.

3.1. Overview of PGL4SHM Architecture

Damage occurring in a load-carrying member changes the dynamic characteristics of the structure (Balageas et al., 2010). Fundamental dynamic characteristics of a system can be described in terms of its modal parameters such as natural frequencies, f, and mode shapes, ϕ. These parameters can be obtained from time series data using frequency domain analysis techniques (Brincker et al., 2000). Prior research has shown that supervised black-box algorithms utilizing modal parameters in the input layer can predict damage detection and localization with success (Z. Wang et al., 1997; Hakim & Razak, 2014). A number of literature specifically focused on the use modal parameters such as natural frequencies and mode shapes to predict damages in a more refined manner (Kim et al., 2003; S.-Q. Wang & Li, 2012). On the other hand, majority of the aforementioned approaches depends on the existence of full-range experimental data. In this paper, to overcome the limitations of black-box models, we propose PGL4SHM where the domain-specific knowledge is embedded into the deep learning architecture through intermediate layers inspired from Muralidhar et al. (2019).

Here, we assume that the input is structural response measurements in the time domain obtained from a physics-based simulated model, the output is the damage condition associated with the input data. The intermediate layers utilize physical variables to improve supervised learning to enable a rich and generalized representation of the target system and to improve supervised learning. The physics-based model is developed as a representative finite element model (FEM) based on the available experimental data obtained from the undamaged structure. The simulated training data is generated using this FEM for various damage conditions of interest, including no damage case. The modal parameters f and ϕ can be extracted directly from the FEM. A simplified layout of the architecture for training is given in Figure 1. In this architecture,

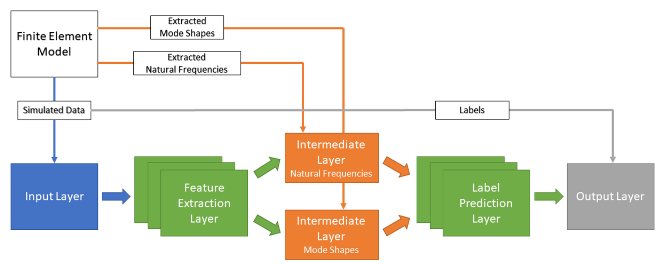

Figure 1. Simplified layout for training PGL4SHM

the input later takes the simulated time-series data obtained from FEM. Each piece of simulated data is associated with a label designating the damage condition. The feature extraction layers are a set of layers designed as convolutional neural networks (CNN). Additionally, there are two individual intermediate variable layers in parallel. The output of each intermediate layer corresponds to a modal parameter (f and ϕ). While the modal parameters are simply extracted from the FEM using eigenvalue problem (Craig Jr & Kurdila, 2006), it can be also derived from time series using domain-specific frequency-domain analysis processes (Ghanem & Shinozuka, 1995). The intermediate layers are *physics-guided* and are directly associated with physically meaningful modal parameters which are known to be good damage indicators. In this regard, this architecture exploits the feature extraction as a modal analysis step to compute intermediate variables which essentially blends domain-specific knowledge with the learning process. For this study, we assumed the intermediate variable layers are densely connected following a flattening layer after CNN based feature extraction layers. Next, the label prediction layers are tasked to extract features from modal properties to determine the damage condition for the given input. For training, PGL4SHM requires simulated time-series data, associated modal parameters, and labels. During testing, the architecture needs only experimental data to the input layer and predicts the relevant damage condition accordingly.

A black-box architecture is very similar to the PGL in nature with a main difference. Since black-box does not utilize intermediate layers and physical parameters associated with it, these layers are simply not implemented.

3.1.1. Physics-based Modeling

A physics-based presentation of the structure can be often achieved by modeling the target structure using a finite element model. In FEMs, the structural systems are modeled as a set of discrete elements (known as finite elements) that are related to the physical properties of the structure such as stiffness, area of the member section, etc. A finite element model can be idealized as a set of mass (M), stiffness (K), and damping matrix (C) which can be written in terms of equations-of-motions (EOMs).

$$M\ddot{x} + C\dot{x} + Kx = F \tag{1}$$

where F is the input excitation such as ground motion, x, $\dot{x}$, and x are the acceleration, velocity, and displacement obtained from the system, respectively. A finite element model can be generated either manually by representing each structural element faithful to its physical properties or by extracting the EOM matrices from the experimental data (Fritzen, 1986; Chen et al., 1996). Complex FEMs involve large dimensional matrices which complicate the modeling and make the calibration process cumbersome. As a result, the modeling errors are inevitable but often acceptable for many engineering applications.

Once the matrices are obtained, the training data can be simulated using Eq 1. By modifying M or K depending on the damage type, various damage conditions can be simulated. For every damage type, an eigenvalue problem can be applied to extract modal parameters as follows:

$$\lambda M\phi = K\phi \tag{2}$$

where λ is the diagonal eigenvalue matrix and can be written also as $\lambda = \mathrm{diag}(2\pi f^2)$. It should be noted that for every combination of K and M pair, a unique pair of f and ϕ can be generated.

In addition to eigenvalue analysis, the modal parameters can be obtained from structural responses using sophisticated time and frequency analysis techniques (Ghanem & Shinozuka, 1995). A clear relationship between physics-based EOM matrices, structural responses, and modal features is obvious as all of them are related to the dynamic characteristics of the structure. Integration of modal parameters into the learning process as domain-specific knowledge is a promising tool for generalizing damage detection compared to black-box approaches.

3.1.2. Learning Process

This network is typically trained with structural response data obtained from a representative FEM. Additionally, the architecture utilizes physics-based modal parameters also obtained from FEM. Accordingly, the empirical loss function that needs to be minimized during learning can be formalized as follows:

$$Loss = Loss_{\text{DMG}} + \lambda_{\text{PGL}} Loss_{\text{PGL}} \tag{3}$$

Eq. 3 implies that the network utilizes a multi-task learning scheme, where $Loss_{\text{DMG}}$ corresponds to the categorical cross-entropy loss between the actual damage condition, y and predicted label, $\hat{y}$; $Loss_{\text{PGL}}$ represents mean square error (MSE) for the physics-guided learning parameters; and λ_{PGL} is the trade-off parameter. Please note that a black-box model will only use $Loss_{\text{DMG}}$ for the training and disregard the physics related loss, $Loss_{\text{PGL}}$.

The physics-guided loss, $Loss_{\text{PGL}}$ given in Eq. 3 can be described as:

$$Loss_{\text{PGL}} = Loss_{\text{PGL}}(f, \hat{f}) + Loss_{\text{PGL}}(\phi, \hat{\phi}) \tag{4}$$

Here, $Loss_{\text{PGL}}(f, \hat{f})$ is the MSE between the actual natural frequencies, f and predicted frequencies, $\hat{f}$; and $Loss_{\text{PGL}}(\phi, \hat{\phi})$ is the MSE between the actual mode shapes, ϕ and predicted ones, $\hat{\phi}$. Since $Loss_{\text{PGL}}$ is a regression loss, we assume the neurons of the intermediate layers are linearly activated.

4. EVALUATION

For this study, we evaluated the PGL4SHM architecture by comparing it to the black-box model performance. We have considered two case studies. The first case study focuses on a finite element model of a simply supported beam, and the second case investigates experimental testing of a three-story structure.

4.1. Implementation

The FEM (simply supported beam) and experimental structure (three-story structure) are excited with white noise under various damage conditions and the resulting dynamic responses are collected from all available sensors in terms of accelerations for some amount of time. Then, the accelerations are divided into 1-second chunks and each of these chunks are categorized according to the relevant damage state. The data obtained from this process is the *reference data* and used for testing.

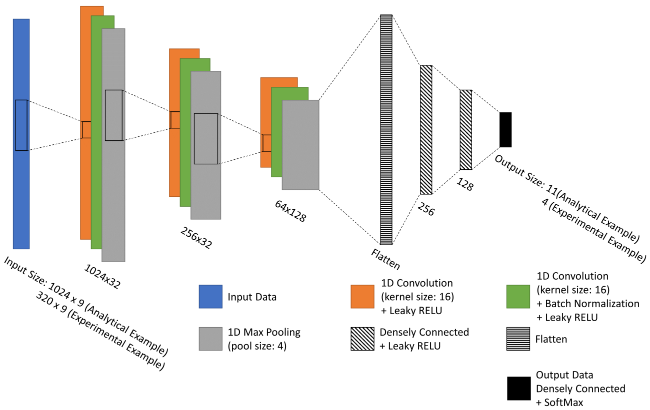

Figure 2. Black-box architecture adopted from Lin, Nie, & Ma, 2017

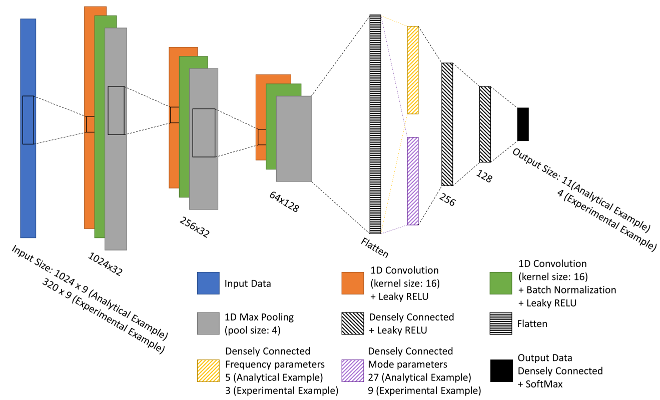

Figure 3. PGL4SHM architecture

In parallel, for each case, another FEM model is developed to replicate the original structure. This model is intentionally misrepresented to some degree in order to introduce modeling errors that occasionally occur during the design process. Using this FEM, the structural responses and corresponding damage labels are generated. In addition, the modal parameters are extracted from this FEM and vectorized. This data is then divided into training, validation, and testing with a ratio of $0.6 : 0.2 : 0.2$, respectively. The training and validation data is used during the training phase of PGL4SHM. The testing data and the *reference data* are used for performance evaluation. Before training, all available data is standardized by removing the mean and scaling to unit variance with respect to training data. All FEM and experimental data is standardized with scikit-learn toolbox.

Next, two neural network models are trained for each case. The first neural network is a black-box model that learns end-to-end relationship between the time series input and the damage condition (see Figure 2). The network is structured as prescribed in Lin et al. (2017). This model does not utilize physics-guded variables, mode shapes and frequencies at all. The dimension of the input depends on the number of the sensors and the sampling number. The feature extractor and label prediction layers are CNN and DNN, respectively. All neurons have leaky RELU activation functions. The size of the output layer changes with respect to the number of damage conditions considered for the case study. The neurons of this layer are activated with softmax. To generalize the predictions and mitigate the internal covariate shift, batch normalization layers are also inserted to the black-box model. Lastly, to reduce the number of trainable parameters, every batch normalization layer is followed by a max pool layer. The second model, PGL4SHM architecture is trained with the training data to minimize the loss function given in Eq. 3 (see Figure 3). This network structure follows an arrangement similar to the black-box model with the addition of intermediate value layers which employs the physics-guided modal parameters. The intermediate layers are densely connected and the neurons are linearly activated. Both black-box model and PGL4SHM architectures are trained using Keras running on TensorFlow 2.0 in Python 3.7. The performance of both architectures is evaluated by computing the classification accuracy.

4.2. Case 1: Analytical Example

This case focuses on the effectiveness of the proposed model where modeling errors relevant to environmental, operational, and material uncertainties are controlled more precisely. Here, we consider a simply supported beam studied by Lin et al. (2017). The beam has a span length of $L = 10.0\ m$ and a rectangular section with $0.1m$ width and $0.25m$ (see Figure 4). The beam is assumed to be made of steel with the elastic modulus of $206\ GPa$ and density of $7,900\ kg/m^3$. The damping is simulated with classic Rayleigh damping where mass matrix (M) proportional factor, α is $1.0\ s^{-1}$ and stiffness matrix (K) proportional factor, β is $1.15 \times 10^{-6}\ s$. The beam is modeled using FEM tool, Open System for Earthquake Engineering Simulation - OpenSees (McKenna et al., 2010). The beam is discretized into ten equally long members that have linear elastic-beam column element properties. Excluding support nodes, the beam has 9 nodes. To generate acceleration responses, the beam is excited at each of the nine nodes vertically with a random noise. This excitation has a Gaussian distribution with a mean of $200\ N$ and standard deviation. To simulate finite features of the environmental noise, the random excitation is filtered with an eighth-order Butterworth filter that has a cutoff of $512\ Hz$. The sampling rate for the simulation is selected $8192\ Hz$. To reduce the volume of the data, the simulation data is downsampled to $1024\ Hz$ and only vertical accelerations at nine nodes are considered. For each loading case, the size of one simulation instance is $(9\ \text{nodes} \times 1024\ Hz)$. To simulate damage conditions, each of the ten members is damaged individually by reducing the member stiffness by 10% to 50% with 10% increments. Including *no damage state*, 11 damage conditions are simulated at nine loading positions across ten members. In addition to time-series data, for each data simulation, first 5 natural frequencies and 27 three modes shape points (3 modes $\times$ 9 nodes) are extracted from the FEM analytically using OpenSees. The resulting data is categorized according to the damage location independent of the magnitude of the damage and the location of the excitation. All the data generated so far constitutes the reference data for testing. In parallel, another set of simulation data are generated with

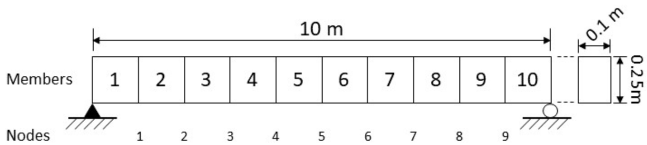

Figure 4. Simply supported beam model used for analytical case

an inaccurate FEM model. To account for environmental, operational, and material uncertainties, for each data instance, the stiffness of the inaccurate model is perturbed with a log-normal distribution. Four inaccurate models are developed where the maximum error of all sampled elastic modulus varies between 5% and 20% with 5% increments. This data is used for training, validation, and testing of PGL4SHM. In addition to time-series data, for each data simulation, the first 5 natural frequencies and 3 modes shapes are extracted from the FEM analytically using OpenSees.

Following the deep-learning architecture provided by Lin et al. (2017), the black-box model and PGL4SHM counterpart are developed, yielding about 1,072,267 and 621,739 trainable parameters to optimize, respectively. The PGL4SHM has intermediate layers between feature extraction and label prediction layers. These intermediate layers act as a choke point, decreasing the number of trainable parameters. To make up for the capacity of the PGL4SHM, two more convolutional layers (a regular convolutional layer and one with batch normalization and max pooling) are added before the flattening. This model, namely PGL4SHM - Extended, has 1,097,387 trainable parameters.

Two versions of PGL4SHM (regular and extended) are compared to the black-box architecture. Table 1 summarizes the classification accuracy and the improvement over black-box architecture with respect to the maximum modeling error in percentage. In addition, Figure 5 visualized the accuracy of all architectures. For no modeling error (ME 0%), while black-box outperforms the regular PGL4SHM, the performance of extended PGL4SHM surpasses all of them. When there is a small modeling error (ME 5%), black-box is the best among the three, resulting in to 94 percent accuracy. On the other hand, the difference between black-box and extended PGL4SHM (84.98 vs 84.55 percent) is negligible. The power of PGL4SHM shines when the modeling error is above 5 percent. For the cases ME 10%, ME 15%, and ME 20%, the performance of both PGL4SHM architectures succeeds black-box significantly. Overall, the improvement of prediction accuracy increases progressively with the modeling error.

In general, the black-box model is a good choice when the modeling fidelity is ensured. Both black-box and extended PGL4SHM have about the same amount of trainable parameters, and their prediction accuracies are similar. The extended PGL4SHM is successful for almost every case except the case ME 5%, however, compared to black-box, the performance loss is negligible. When the modeling error is small, compared to the extended PGL4SHM, the regular PGL4SHM is, in general, less effective due to the small number of trainable parameters. The results clearly show that especially when the numerical model does not represent the actual system properly, blending physical parameters with data-driven machine learning has a positive impact in improving the damage localization.

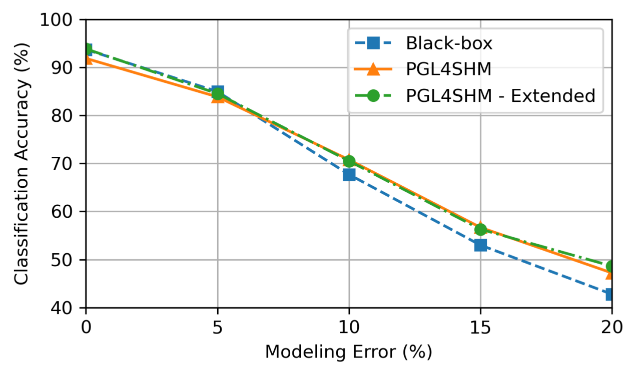

Figure 5. Visualization of classification accuracy for analytical case

4.3. Case 2: Experimental Example

The performance of the PGL4SHM approach is also evaluated on a small-scale three-story structure tested by Figueiredo et al. (2009). An electromagnetic shaker is attached to the base of the structure (see Figure 6). The structure was excited with a band-limited white noise and the resulting horizontal acceleration responses and the excitation force were measured at a sampling rate of 320 Hz for about 25 s. For this study, including undamaged state, four damage conditions are considered. The damage states are established by reducing the stiffness of one or two columns at each floor by 87.5 percent. Each response data instance is categorized according to its respective damage condition. After the input force is removed from the measurements, time series data are divided into 1-second chunks. Each chunk is categorized according to its respective damage condition. The data collected in this phase is the reference data for testing. In addition, a high-fidelity lumped-mass model is generated

	Black-box	PGL4SHM		PGL4SHM - Extended	
Modeling Error	Accuracy (%)	Accuracy (%)	Improvement (%)	Accuracy (%)	Improvement (%)
ME 0%	93.75	91.90	-1.97	93.89	0.15
ME 5%	84.98	83.88	-1.29	84.55	-0.50
ME 10%	67.71	70.78	4.55	70.52	4.15
ME 15%	52.98	56.67	6.96	56.29	6.24
ME 20%	42.75	47.18	10.35	48.60	13.68

Table 1. Classification accuracy of black-box and PGL4SHM for analytical case

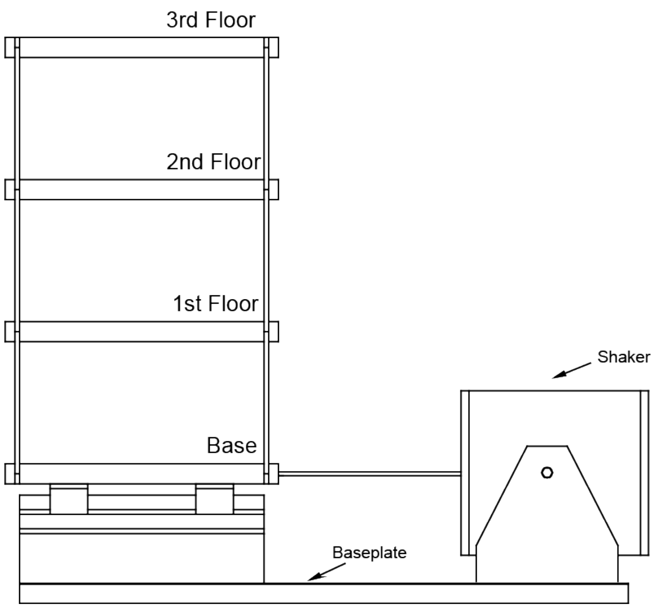

Figure 6. Three-story structure used for experimental case

in the form of mass-stiffness-damping matrices using the parameters provided by Hernandez-Garcia et al. (2010) and Sun & Betti (2015). A 10 percent error is introduced into the stiffness matrix to simulate the modeling uncertainties. Using this imperfect model, data for all damage conditions are produced. In addition, three natural frequencies and 9 mode shapes points (3 modes $\times$ 3 stories) are obtained using this model. The data from the imperfect FEM model is used for training and validation of PGL4SHM. Black-box and PGL4SHM architectures have 707,844 and 557,459 trainable parameters, respectively. No further layers are added to PGL4SHM to extend the capacity of PGL4SHM. We also ensured there is no overfitting by validating the models against their respective numerical datasets. The performance of trained architectures and the improvement of PGL4SHM over black-box architecture for the corresponding modeling error are provided in Table 2. The classification performance of the black-box for no modeling error (96.18%) is greater than that of PGL4SHM (90.74%) and the performance loss reaches up to 6 percent. For a moderate level of modeling error (ME 10%), the black-box model yields a poor performance (38.06%) compared to PGL4SHM (70.82%).

	Black-box	PGL4SHM	
Modeling Error	Acc. (%)	Acc. (%)	Impr. (%)
ME 0%	96.18	90.74	-5.66
ME 10%	38.06	70.82	86.07

Table 2. Classification accuracy of black-box and PGL4SHM for experimental case

Additionally, Table 2 presents the averaged F1-scores for the experimental case. The results and improvements are in parallel with the classification accuracies.

	Black-box	PGL4SHM	
Modeling Error	F1 (%)	F1 (%)	Impr. (%)
ME 0%	96.21	90.48	-5.96
ME 10%	31.09	67.36	116.66

Table 3. Averaged F1-score of black-box and PGL4SHM for experimental case

Lastly, Fig. 7 illustrates averaged ROC for the experimental case. Blackbox and PGL4SHM have similar ROC performance when there is no modeling error. Under moderate level of modeling error, PGL4SHM has a better classification performance compared to the blackbox model.

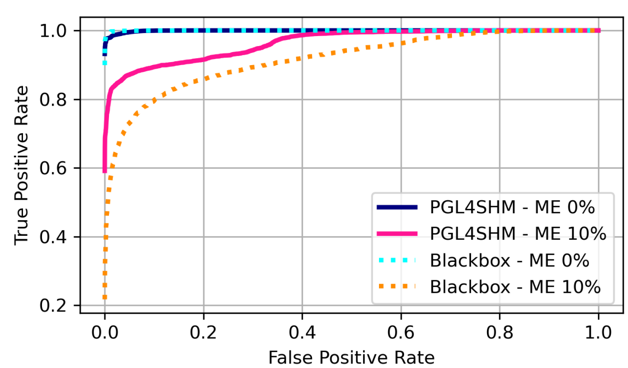

Figure 7. Averaged ROC curves for experimental case

From the results, it is evident that the black-box overfits the numerical data such that the latent features of the experimental data cannot be perceived. As a result, without the integration of the physical parameters, the data-driven black-box architecture fails to predict the damage classes correctly. The results indicate that in the presence of modeling error, the generalization of PGL4SHM is much more successful.

4.4. Effect of Hyper-parameters

Here, we investigated the effect of the trade-off parameter, λ_{PGL} on the prediction accuracy to understand the generalization of PGL4SHM under no modeling error. Table 4 summarizes the performance of both models and the weights for classification loss and physics-based loss for no modeling error. Trade-off parameter, in general, does not affect the performance of PGL4SHM, except for the case ($\lambda_{PGL} = 2.0$) where the weight for physical parameter loss is larger than the one for the classification loss. PGL4SHM with no weights to the physical parameters ($\lambda_{PGL} = 0.0$) is similar to the black-box model, but it still contains the intermediate layers. It is clear from the results that the introduction of intermediate layers degrades the performance of PGL4SHM when there is no modeling error. The small dimension of intermediate layers after the label prediction layer (see Figure 1) causes the learning to be *under-complete* leading to decrease in accuracy. For larger models, the number of physical parameters can be increased and the label prediction layer will have a more complete basis for learning. For general purposes, weighting the losses equally ($\lambda_{PGL} = 1.0$) is a good starting point in training the PGL4SHM.

λ_{PGL}	Accuracy (%)
0.0	86.26
0.5	89.24
1.0	87.48
2.0	68.32

Table 4. Effect of hyper-parameters on the classification accuracy under no modeling error (ME 0%)

4.5. Interpretability of Intermediate Layer Outputs

We evaluated the explainability of the PGL4SHM by analyzing the relationship between the damage condition and intermediate layer outputs. Specifically, we focused on the interpretability of natural frequency, as it is more human-comprehensible and easier to visualize. Figure 8 illustrates *predicted* natural frequencies from intermediate layers, along with the *experimental* (true) and *simulated* (training) counterparts for four damage cases, where the modeling error is %10. Here, PGL4SHM is evaluated with experimental data. For each damage case, the intermediate layers in PGL4SHM predict three natural frequencies around 30, 55, and 70 Hz with some variance. Compared to the *experimental* true frequencies of the structure (square markers), *simulated* values extracted from FEM (cross markers) always undershoot. This is expected since the modeling error is introduced to the FEM

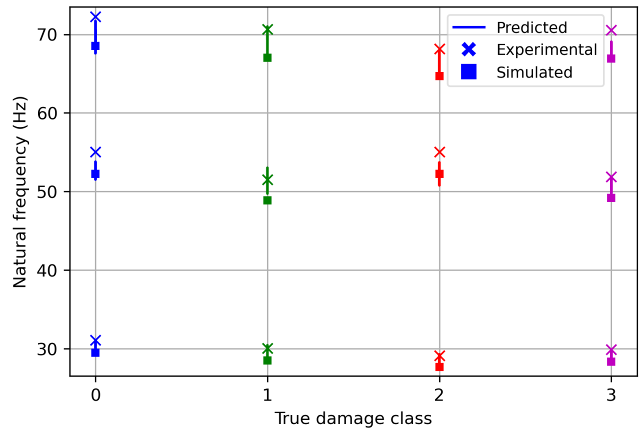

Figure 8. Interpretability of intermediate layer outputs

by reducing the stiffness matrix by 10 percent which causes the *simulated* frequencies to decrease. In general, the *predicted* frequencies range from *simulated* to *experimental* values.

During training, the simulated modal parameters are used for physics-based loss function. On the other hand, integration of physics-based parameters into the training also constrains the inference such that PGL4SHM favors to predict the modal parameters towards the experimental true counterparts. There are some cases where the predicted values do not distribute uniformly between experimental and simulated values. The distortion is substantial especially for the second modes (50 Hz) of damage class 1 and 2. This error causes some of the intermediate value outputs from class 1 and 2 to overlap with the damage class 0 (no damage class) leading to mislabeling. Due to the explainability of results, such problematic instances can be in theory captured algorithmically and corrected at testing time.

5. CONCLUSION

In this paper, we have presented a physics-based deep learning architecture, PGL4SHM to detect and localize the damage in mechanical systems. The proposed approach incorporates physical parameters such as natural frequencies and mode shapes, which are known to be statistically meaningful features for damage detection, into the intermediate layers of deep neural networks. To accommodate the intermediate layers, the architecture introduced physics-based loss into empirical loss function. To evaluate the proposed approach, we considered analytical and experimental cases. Both examples show that physics-guided learning improves the accuracy of the damage localization compared to black-box models in the presence of modeling errors. Our empirical study shows that weighting the classification and physical loss equally is an effective starting point for training. Lastly, we discussed the interpretability of intermediate layer output by analyzing the relationship between predicted modal parameters and classification performance. Our findings indicate that the misclassified instances could be explained through the characterization of predicted natural frequencies.

ACKNOWLEDGMENT

This work is supported in part by DARPA through contract number FA8750-20-C-0537. Any opinions, findings, and conclusions or recommendations expressed are those of the authors and do not necessarily reflect the views of the sponsor.

REFERENCES

Bakhary, N., Hao, H., & Deeks, A. J. (2007). Damage detection using artificial neural network with consideration of uncertainties. *Engineering Structures*, *29*(11), 2806–2815.

Balageas, D., Fritzen, C.-P., & Güemes, A. (2010). *Structural health monitoring* (Vol. 90). John Wiley & Sons.

Brincker, R., Zhang, L., & Andersen, P. (2000). Modal identification from ambient responses using frequency domain decomposition. In *Proc. of the 18*'international modal analysis conference (imac), san antonio, texas.*

Chen, S., Ju, M.-S., & Tsuei, Y. (1996, 01). Estimation of mass, stiffness and damping matrices from frequency response functions. *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, *118*(1), 78-82. Retrieved from `https://doi.org/10.1115/1.2889638` doi: 10.1115/1.2889638

Craig Jr, R. R., & Kurdila, A. J. (2006). *Fundamentals of structural dynamics*. John Wiley & Sons.

Farrar, C. R., & Worden, K. (2012). *Structural health monitoring: a machine learning perspective*. John Wiley & Sons.

Figueiredo, E., Park, G., Figueiras, J., Farrar, C., & Worden, K. (2009). Structural health monitoring algorithm comparisons using standard data sets. *Los Alamos National Laboratory, Los Alamos, NM, Report No. LA-14393*.

Fritzen, C.-P. (1986, 01). Identification of Mass, Damping, and Stiffness Matrices of Mechanical Systems. *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, *108*(1), 9-16. Retrieved from `https://doi.org/10.1115/1.3269310` doi: 10.1115/1.3269310

Gardner, P., Liu, X., & Worden, K. (2020). On the application of domain adaptation in structural health monitoring. *Mechanical Systems and Signal Processing*, *138*, 106550.

Ghanem, R., & Shinozuka, M. (1995). Structural-system identification. i: Theory. *Journal of Engineering Mechanics*, *121*(2), 255–264.

Hakim, S., & Razak, H. A. (2014). Modal parameters based structural damage detection using artificial neural networks-a review. *Smart Structures and Systems*, *14*(2), 159–189.

Hernandez-Garcia, M. R., Masri, S. F., Ghanem, R., Figueiredo, E., & Farrar, C. R. (2010). An experimental investigation of change detection in uncertain chain-like systems. *Journal of Sound and Vibration*, *329*(12), 2395–2409.

Jaishi, B., & Ren, W.-X. (2006). Damage detection by finite element model updating using modal flexibility residual. *Journal of sound and vibration*, *290*(1-2), 369–387.

Jia, X., Karpatne, A., Willard, J., Steinbach, M., Read, J., Hanson, P. C., ... Kumar, V. (2018). Physics guided recurrent neural networks for modeling dynamical systems: Application to monitoring water temperature and quality in lakes. *arXiv preprint arXiv:1810.02880*.

Karpatne, A., Watkins, W., Read, J., & Kumar, V. (2017). Physics-guided neural networks (pgnn): An application in lake temperature modeling. *arXiv preprint arXiv:1710.11431*.

Kim, J.-T., Ryu, Y.-S., Cho, H.-M., & Stubbs, N. (2003). Damage identification in beam-type structures: frequency-based method vs mode-shape-based method. *Engineering structures*, *25*(1), 57–67.

Lin, Y.-z., Nie, Z.-h., & Ma, H.-w. (2017). Structural damage detection with automatic feature-extraction through deep learning. *Computer-Aided Civil and Infrastructure Engineering*, *32*(12), 1025–1046.

McKenna, F., Scott, M. H., & Fenves, G. L. (2010). Nonlinear finite-element analysis software architecture using object composition. *Journal of Computing in Civil Engineering*, *24*(1), 95-107. Retrieved from `https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29CP.1943-5487.0000002` doi: 10.1061/(ASCE)CP.1943-5487.0000002

Muralidhar, N., Bu, J., Cao, Z., He, L., Ramakrishnan, N., Tafti, D., & Karpatne, A. (2019). Physics-guided design and learning of neural networks for predicting drag force on particle suspensions in moving fluids. *arXiv preprint arXiv:1911.04240*.

Sadoughi, M., & Hu, C. (2019). Physics-based convolutional neural network for fault diagnosis of rolling element bearings. *IEEE Sensors Journal*, *19*(11), 4181–4192.

Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., ... Nenadic, G. (2019). Machine learning methods for wind turbine condition monitoring: A review. *Renewable energy*, *133*, 620–635.

Sun, H., & Betti, R. (2015). A hybrid optimization algorithm with bayesian inference for probabilistic model updating. *Computer-Aided Civil and Infrastructure Engineering*, *30*(8), 602–619.

Teughels, A., & De Roeck, G. (2005). Damage detection and parameter identification by finite element model updating. *Revue européenne de génie civil*, *9*(1-2), 109–158.

Wang, S.-Q., & Li, H.-J. (2012). Assessment of structural damage using natural frequency changes. *Acta Mechanica Sinica*, *28*(1), 118–127.

Wang, Z., Lin, R., & Lim, M. (1997). Structural damage detection using measured frf data. *Computer methods in applied mechanics and engineering*, *147*(1-2), 187–197.

Widodo, A., & Yang, B.-S. (2007). Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical systems and signal processing*, *21*(6), 2560–2574.

Yao, H., Gao, Y., & Liu, Y. (2020). Fea-net: A physics-guided data-driven model for efficient mechanical response prediction. *Computer Methods in Applied Mechanics and Engineering*, *363*, 112892.

Zhang, R., Liu, Y., & Sun, H. (2020). Physics-guided convolutional neural network (phycnn) for data-driven seismic response modeling. *Engineering Structures*, *215*, 110704.

Zhang, Z., & Sun, C. (2020). Structural damage identification via physics-guided machine learning: a methodology integrating pattern recognition with finite element model updating. *Structural Health Monitoring*, 1475921720927488.

BIOGRAPHIES

Ali I. Ozdagli is a graduate student in the Department of Electrical Engineering and Computer Science at Vanderbilt University. Prior to this, he got his M.S. in civil engineering from the University of Notre Dame and his Ph.D. from Purdue University. His research focuses on adapting machine learning approaches to structural health monitoring applications.

Xenofon Koutsoukos is a professor with the Department of Electrical Engineering and Computer Science and a senior research scientist with the Institute for Software Integrated Systems, Vanderbilt University. His research work is in the area of cyber-physical systems with an emphasis on security and resilience, control, diagnosis and fault tolerance, formal methods, and adaptive resource management. He received the Ph.D. degree in electrical engineering from the University of Notre Dame. He is a Fellow of the IEEE.