On the Effectiveness of Dimensionality Reduction for Unsupervised Structural Health Monitoring Anomaly Detection

Mohammad Hesam Soleimani-Babakamali

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Computer Science and Application

Ismini Lourentzou, Chair

Mohammed Fawzi Seddik Farghally

Rodrigo Sarlo

April 19, 2022

Blacksburg, Virginia

Keywords: Unsupervised SHM; Generative Adversarial Networks; Anomaly Detection, Dimensionality Reduction; Autoencoder; Regularization

On the Effectiveness of Dimensionality Reduction for Unsupervised Structural Health Monitoring Anomaly Detection

Mohammad Hesam Soleimani-Babakamali

(ABSTRACT)

Dimensionality reduction techniques (DR) enhance data interpretability and reduce space complexity, though at the cost of information loss. Such methods have been prevalent in the Structural Health Monitoring (SHM) anomaly detection literature. While DR is favorable in supervised anomaly detection, where possible novelties are known a priori, the efficacy is less clear in unsupervised detection. In this work, we perform a detailed assessment of the DR performance trade-offs to determine whether the information loss imposed by DR can impact SHM performance for previously unseen novelties. As a basis for our analysis, we rely on an SHM anomaly detection method operating on input signals' fast Fourier transform (FFT). FFT is regarded as a raw, frequency-domain feature that allows studying various DR techniques. We design extensive experiments comparing various DR techniques, including neural autoencoder models, to capture the impact on two SHM benchmark datasets exclusively. Results imply the loss of information to be more detrimental, reducing the novelty detection accuracy by up to 60% with autoencoder-based DR. Regularization can alleviate some of the challenges though unpredictable. Dimensions of substantial vibrational information mostly survive DR; thus, the regularization impact suggests that these dimensions are not reliable damage-sensitive features regarding unseen faults. Consequently, we argue that designing new SHM anomaly detection methods that can work with high-dimensional raw features is a necessary research direction and present open challenges and future directions.

On the Effectiveness of Dimensionality Reduction for Unsupervised Structural Health Monitoring Anomaly Detection

Mohammad Hesam Soleimani-Babakamali

(GENERAL AUDIENCE ABSTRACT)

Structural health monitoring (SHM) aids the timely maintenance of infrastructures, saving human lives and natural resources. Infrastructure will undergo unseen damages in the future. Thus, data-driven SHM techniques for handling unlabeled data (i.e., unsupervised learning) are suitable for real-world usage. Lacking labels and defined data classes, data instances are categorized through similarities, i.e., distances. Still, distance metrics in high-dimensional spaces can become meaningless. As a result, applying methods to reduce data dimensions is currently practiced, yet, at the cost of information loss. Naturally, a trade-off exists between the loss of information and the increased interpretability of low-dimensional spaces induced by dimensionality reduction procedures. This study proposes an unsupervised SHM technique that works with low and high-dimensional data to assess that trade-off. Results show the negative impacts of dimensionality reduction to be more severe than its benefits. Developing unsupervised SHM methods with raw data is thus encouraged for real-world applications.

Dedication

This article is dedicated to my parents, the holy spectrum of Lord my earthly eyes can grasp.

Acknowledgments

I would like to express my sincerest gratitude to Dr. Ismini Lourentzou. I learned more than I had imagined, whether brain-learned academic matters or soul-learned perspectives for discerning the reality out of this fictional World; I am grateful for this miraculous sense.

Contents

List of Figures

List of Tables

List of Abbreviations

FAR False alarm ratio

$SFAR$ Stern false alarm ratio

TAR True alarm ratio

AE Autoencoder

AUC Area under the receiver operating characteristic curves

CNN Convolutional neural network layer

DR Dimensionality reduction

FFT Fast Fourier transform

FPR False Positive Rate

GAN Generative adversarial network

LSTM Long Short-Term Memory units

PCA Principal component analysis

ROC Receiver operating characteristic

SHM Structural health monitoring

DR is a transformation that represents the data in a space with a lower number of dimensions while preserving the variation in the data.

SHM concerns monitoring a system over time by analyzing its responses to detect possible abnormal states.

FFT is an algorithm to compute the discrete Fourier transform of a sequence.

PCA is a linear dimensionality reduction technique.

GAN is a class of machine learning frameworks consisting of two sub-neural networks, generator, and discriminator, trained while competing.

AE stands for an autoencoder network, an instance of deep dimensionality reduction techniques.

LSTM stands for Long Short-Term Memory units, an instance of deep neural network layers.

CNN stands for convolutional neural network layers.

TAR stands for true alarm ratio equals the detected novelties ratio.

FAR stands for false alarm ratio equal to the ratio of false alarms to the total data instances.

$SFAR$ stands for stern false alarm ratio equals the ratio of classes with false alarms detected to the total number of classes.

ROC analysis shows the accuracy of a binary classifier as its detection threshold varies.

AUC measures a binary classifier performance from 0 to 1, with 1 depicting the best performance and 0.5 denoting a random guess.

Chapter 1

Introduction

Aging infrastructure, stern safety requirements, limited natural resources, and post-hazard condition assessments call for structural health monitoring of infrastructures [5]. This need has become more tangible with the increasing population and the growing concerns on climate change associated with global pollutants, such as construction and demolition of non-habitable damaged infrastructure [25, 53]. SHM assesses infrastructure functionality to identify possible global and local defects and damages. Local failures are then curated through maintenance, avoiding total demolition of infrastructure. Deterioration, fatigue, or extraordinary events are prominent causes of such progressive defects. SHM, though, comes at the cost of equipment and human labor. Whether SHM can save costs in the life-cycle of infrastructures such as bridges or more common residential buildings is an active field of research [40].

SHM detects novelties in the behavior of structures. Inspections and maintenance are planned upon SHM alerts. Structures are designed to satisfy specific conditions in response to loads and their target usages. Hence, monitoring infrastructures' mechanical properties are the target of SHM. Mechanics of Structures and vibrational data translate such properties into a mathematical, human-understandable format. Transducers, such as accelerometers or strain gauges, capture structures' responses in the format of high-dimensional time signals. Those signals are then translated to human-understandable representations, such as power-spectrum or low-dimensional vibrational modes. Those representations are then assessed

with system identification mechanisms, such as subspace identification [50], and novel states are identified.

With the emergence of deep learning in 2012, various disciplines, such as Civil engineering, devised and improved their data interpretability problems. In SHM, deep learning coupled with low-cost, smart wireless sensors [3] is a recent research area to enable the long-lasting goal of developing SHM tools for real-life structures [4] and ultimately cities with networks of numerous infrastructures. Modes and signal-processing-based representative features are categorized as hand-crafted features. They are shown to not be scalable across different structures [2]. Thus, deep learning, a representative learning tool, is the resolution to obtain damage-sensitive features from the incoming high-dimensional data. Though, the blind use of machine learning tools while neglecting the domain knowledge, data structure, and target applications is not safe. First, the nature of data may differ from well-known machine learning problems. SHM data can be far different from video, text, or image datasets while being supervised, unsupervised, or even anomaly detection in both categories. In SHM, the target is to identify novel structural states in the future. With anticipated damage events, supervised learning is a reliable tool to teach the machine the patterns of unsafe structural conditions. Still, in real-life problems, novelties are unknown. Supervised studies are of high importance. They form the basis of developing deep SHM tools; still, unsupervised SHM tools are the sought solution for unknown damages in the future. Consequently, the current SHM research direction is to move from the supervised setting to real-time, semi-supervised, and unsupervised SHM [12, 30, 33, 38, 39], demanding no labeled data.

With high-dimensional signals as data, machine interpretability issues arise due to the curse of dimensionality. Dimensionality-reduction techniques have gained much attention recently to enable unsupervised SHM [48]. Still, the question remains whether such practices are safe in SHM or not? What are the differences between SHM and text data, in which la-

tent semantic analysis [23], in which Principal Component Analysis (PCA), a simple linear dimensionality reduction technique or autoencoder-based features [20] have shown great potential in natural language processing problems? This study studies several SHM datasets to answer that question to identify best practices for unsupervised SHM for real-life usage.

Chapter 2

On the effectiveness of dimensionality reduction for unsupervised structural health monitoring anomaly detection

Abstract

Dimensionality reduction techniques (DR) enhance data interpretability, though at the cost of information loss. Regarding anomaly detection in Structural Health Monitoring (SHM), there is a performance trade-off between the more interpretable DR-based features and the information loss. While DR is favorable in supervised detection where possible novelties are known a priori, the trade-off is less clear in unsupervised detection. This study performs a detailed assessment of this trade-off to determine whether the information loss imposed by DR can impact SHM performance for previously unseen novelties. A novelty detection method operating with only the fast Fourier transform (FFT) of input signals is developed to perform that assessment. FFT is regarded as a raw, frequency-domain feature that allows studying various DR techniques. We design several experiments to exclusively capture the impact of information loss while assessing the trade-off on two benchmark SHM datasets. Results imply the loss of information to be more detrimental than complications in data interpretability caused by the curse of dimensionality, reducing the novelty detection accuracy

by up to 60% with autoencoder-based DR. Regularization is shown to alleviate the accuracy decline, though in an unpredictable manner. Dimensions of substantial vibrational information mostly survive DR, and the regularization impact suggests that those dimensions are not reliable damage-sensitive features regarding unknown faults. Moreover, the unpredictable regularization impact makes the safe selection of regularization strategies challenging, urging to avoid the DR in unsupervised novelty detection with SHM data. Consequently, developing novel methods to maintain raw features, such as the one devised herein, is encouraged to avoid the negative impacts of DR on SHM novelty detection performance.

2.1 Introduction

Novelty detection in Structural Health Monitoring (SHM) is a pattern recognition process [47], in which structures are monitored for irregularities via signals from embedded sensors. SHM data are often high-dimensional. For example, a one-second time-series data instance has 1000 dimensions when sampled with a frequency of 1000 Hz. Dimensions are further expanded with multiple input channels in operation. The curse of dimensionality [24] and the noise in SHM data [21, 41] pose challenges to pattern recognition tasks. Dimensionality reduction (DR) can address several problems related to the curse of dimensionality [8, 17, 36]. DR can render more meaningful distance metrics, remove the co-linearity in the data, speed up data analysis, reduce data storage requirements, and enhance interpretability [46]. DR, however, induces loss of information. Hence, there exists a trade-off between the impact of information loss and the benefits of DR. Several factors, e.g., supervised versus unsupervised modeling, dependencies between variables, etc., impact this trade-off.

DR techniques remove redundant and less informative dimensions from training data. If the

training set is representative of the population it is sampled from, then DR applies to that data population (i.e., generalizability). The applications of DR techniques in supervised SHM problems are well-studied. Seminal works include the application of Principal Component Analysis (PCA) and Kernel-PCA for system identification [16, 31]. Autoencoders, a type of deep neural network, are employed to carry out the DR [9, 48]. With specific neural network architectures, such as Generative Adversarial Networks [19], raw features were employed directly for SHM of structures [35] or machining tools [27, 52, 54]. Unlike supervised settings, in unsupervised novelty detection tasks such as SHM, the training data, i.e., already observed data, is intrinsically non-representative of the global data population. SHM is innately a novelty detection process in which unseen structural states are identified through time. Consequently, SHM training sets and associated trained DR may not generalize well to unseen data. DR deletes dimensions that are not informative in the training set, yet these can be damage-sensitive features in an upcoming novel structural state.

With the lack of labels in the unsupervised setting, distance metrics between data instances are the basis of decision-making per novelty detection. Clustering [11, 12, 55] and threshold-models [42, 45] are the most employed approaches in the unsupervised SHM. The main drawback of these methods is their intolerance of high-dimensional data. More specifically, distance metrics become less precise in high-dimensional spaces (i.e., curse of dimensionality). Consequently, extracting low-dimensional representations from the high-dimensional data is typical before clustering. However, an underlying assumption exists that informative dimensions in the baseline data (i.e., observed data) are damage-sensitive features while their representative values are extracted. That assumption is *explicitly* made when calculating statistical or vibrational parameters, i.e., modes, [10, 11, 12, 15, 51]. On the other hand, that assumption is *implicitly* made while extracting damage-sensitive features with data-driven DR-based methods [13, 45], as informative dimensions survive the application of DR. Limited

attention is given in assessing how reliable that assumption is, for example, by comparing the performance of DR approaches with models operating with raw high-dimensional features. The incorporation of raw high-dimensional features is reported in supervised SHM [1, 55] with the aid of representation learning tools (e.g., deep architectures). In contrast, utilizing high-dimensional features for unsupervised SHM tasks has received limited attention.

The purpose of practicing DR techniques is to create an "embedding" in which similar data are embedded in their vicinity, away from other data populations. The embedded representation of data makes the class separation more uncomplicated. With such embeddings, for instance, in the case of machinery data, few-shot or zero-shot learning has become possible [44, 49]. A training set representing the global data population is vital. Concerning unsupervised SHM, such general datasets are not available since unseen novelties in the future are of unknown data populations, and diversity in infrastructure, e.g., frames, bridges, or dams, is far greater than, for example, machinery tools. Still, the discrete Fourier transform of a signal embeds responses from various structural states into the frequency domain while preserving their similarities and differences. That transformation is efficiently carried out with the fast Fourier transform (FFT) algorithm. Theoretically, FFT can separate observed data from unseen novelties due to their different frequency content, even without witnessing them. FFT is a raw frequency-domain representation of structural responses and is thus high-dimensional and vulnerable to the curse of dimensionality and noise. Therefore, FFT is a convenient feature for building an unsupervised novelty detection framework to assess the trade-off above.

In this work, we design an unsupervised SHM approach operating on FFT features of input signals and delve into an empirical comparison of novelty detection with the raw FFT features and features derived through various DR techniques, such as PCA [18], Kernel-PCA [34], and Autoencoders [9]. In summary, the contributions of this study are summarized as follows.

- We present a detailed analysis of the discussed trade-off. We observe the negative impact of DR-induced loss of information to be more detrimental to novelty detection accuracy than the data interpretability complications caused by the curse of dimensionality due to non-generalizable unsupervised SHM datasets. Those detrimental effects become stronger as the DR method becomes more complex, with the simplest employed DR, PCA, suffering the least reduction in novelty detection accuracy.

- We show that the preservation of data dimensions by employing raw features, such as the FFT of response signals, is an effective resolution to avoid the non-generalizability of unsupervised SHM data. Consequently, we suggest developing SHM techniques operating with raw features to avoid the application of DR.

- We evaluate the assumption that the most informative dimensions in the observed data are damage-sensitive features by applying regularization to DR methods. Regularization reduces the negative impact of DR-induced loss of information, denoting the assumption mentioned above to be invalid regarding unsupervised SHM. Novelties may manifest in noisy data dimensions that are often not preserved by DR or by manually selecting representative features such as modes.

2.2 Methodology

An unsupervised SHM framework with high-dimensional raw inputs is necessary to study the impact of DR. Generative Adversarial Networks (GAN) can handle high-dimensional inputs in an unsupervised setting. GANs transforms the unsupervised problem into a binary classification. Consequently, deep neural networks with billions of parameters make representation learning possible. The desired unsupervised SHM framework is developed and presented in this section by employing GAN.

2.2.1 High-dimensional Raw Features

SHM novelty detection frameworks collect acceleration or strain data from several data channels. The data stream is then windowed to form time-series data instances, i.e., signals. As outlined in Section 2.1, the FFT of said signals is the high-dimensional feature, and a GAN-based unsupervised SHM framework is developed to detect novelties from that feature.

Presuming a window length of W, the FFT, a windowed data point, has W dimensions. FFT of real numbers, i.e., spectral lines, are mirrored around the $(W/2)^{\text{th}}$ spectral line. Thus, magnitudes of spectral lines from zero till $W/2$ (H) are selected, i.e., half-spectrum FFT. Each data channel's half-spectrum FFT is normalized, dividing over its mean value while capping the maximum value at 10. The final feature, F, is the concatenation of all normalized half-spectrum FFT of all channels. With N data channels, F has $H \times N$ dimensions.

2.2.2 Novelty Detection Framework

The GAN-based framework consists of a discriminator $\mathcal{D}$ and a generator $\mathcal{G}$, both trained with the F features. The trained discriminator is a binary classifier, distinguishing between normal and novel data instances. The trained generator implicitly defines the training data distribution P_{data} while sampling from a latent space P_z. A trained $\mathcal{G}$ can generate additional data to tune novelty detection thresholds. A high-level overview of the framework is shown schematically in Fig. 2.1. The framework includes the Training, Tuning, and Detection phases. The utility of this framework is for comparison; thus, DR is applicable to the high-dimensional raw feature F before the GAN training.

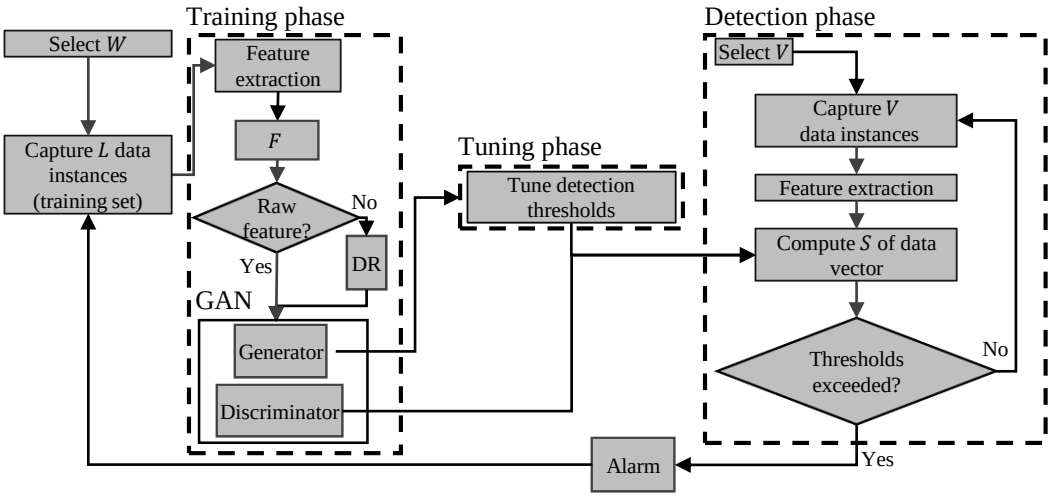


Figure 2.1: GAN-based unsupervised SHM framework.

2.2.3 Training Phase

A pictorial outline of a GAN training routine is shown in Fig. 2.2. The GAN's training objective is to train $\mathcal{G}$ that generates synthetic ("fake") realistic data (i.e., resembling the training set) that $\mathcal{D}$ fails to discriminate between generated "fake" and "real" training data. $\mathcal{G}$ randomly samples from P_z to generate the fake data. P_z is constant, and thus $\mathcal{G}$ is implicitly defining the training data distribution. With the described objectives, the following loss functions for $\mathcal{D}$ and $\mathcal{G}$ are defined as follows

$$\ell_{\mathcal{D}} = -\mathbb{E}_{x\sim P_{\mathrm{x}}}\left[\log \mathcal{D}(x)\right] - \mathbb{E}_{z\sim P_z}\left[\log\left(1-\mathcal{D}(\mathcal{G}(z))\right)\right], \tag{2.1}$$

where x is a data instance of the training set and

$$\ell_{\mathcal{G}} = -\mathbb{E}_{z\sim P_z}\left[\log \mathcal{D}\left(\mathcal{G}(z)\right)\right]. \tag{2.2}$$

Note that $\ell_{\mathcal{D}}$ penalizes $\mathcal{D}$ when synthetic data are classified as real. On the contrary, $\ell_{\mathcal{G}}$ penalizes $\mathcal{G}$ when $\mathcal{D}$ is not misled to label synthetic data as real. The overall training

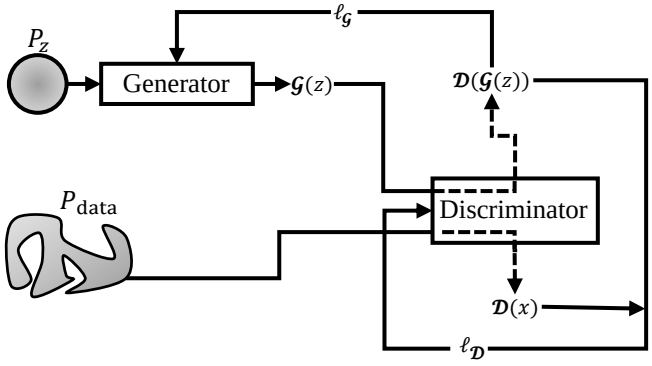


Figure 2.2: GAN training: $\ell_{\mathcal{D}}$ and $\ell_{\mathcal{G}}$ are $\mathcal{D}$ and $\mathcal{G}$ training losses.

procedure characterizes a zero-sum game [29].

2.2.4 Tuning and Detection

As explained in Section 2.2.3, the utility of $\mathcal{D}$ is detecting novel data instances. By designing $\mathcal{D}$ to have sigmoid activations in the last layer, the novelty detection score of the data signal x is formulated as

$$s(x) = -\log(\mathcal{D}(x)). \tag{2.3}$$

Here, $s\colon x \to [0, \infty]$. The limits of the range specify model predictions that map to "real" (i.e., normal) and "fake" (i.e., abnormal) data. In SHM, a single signal x can be noisy, especially if its duration is short. In the literature, batches of time-series data instances are set as the basis of novelty detection in various studies [16, 43]. Similarly in this work, a vector of V data instances is considered at each novelty detection iteration. We denote $\mathbf{s}$ as the output of $\mathcal{D}$ given those V data instances. We propose a threshold model to perform the novelty detection, given $\mathbf{s}$, formulated as

$$P_{50}(\mathbf{s}) > T, \tag{2.4}$$

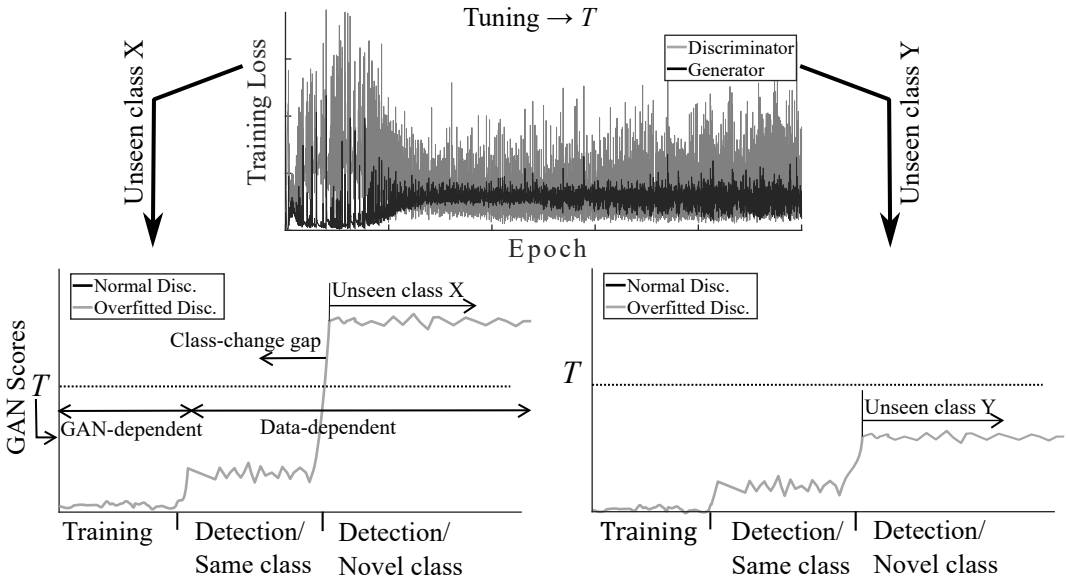


Figure 2.3: Impact of imperfect tuning of T on novelty detection performance.

where P_{50} is the 50th percentile of $\mathbf{s}$, an outlier-resistant representative value. In the proposed framework (Fig. 2.1), P_{data}, defined by $\mathcal{G}$, tunes the detection threshold. Since P_{data} is the training set (i.e., non-novel data) distribution, threshold T in Eq. (2.4) is tuned so that no novelties are identified from data instances generated by $\mathcal{G}$. Threshold T is set based on the highest score that $\mathcal{D}$ outputs for $\mathcal{G}$-generated data instances. Methods such as Monte Carlo histogram sampling (MCHS) [14] can estimate the distribution of those scores to decide the detection threshold. Still, generator $\mathcal{G}$ might generate data instances that are not representative of the training set, as noted in the literature [37, 38, 55]. The reason is that P_{data} is a continuous function, and the neural network may not be adequately trained for the whole range of that distribution. We removed the non-representative generated data by fitting a two-class Gaussian Mixture model [32] to the distribution. The Gaussian class with the higher mean value (i.e., less-representative data instances) is discarded. The maximum novelty detection score of the Gaussian class with the lower mean value is assigned to the detection threshold T.

Detection thresholds directly impact the performance of classifiers. An imperfect tuning of

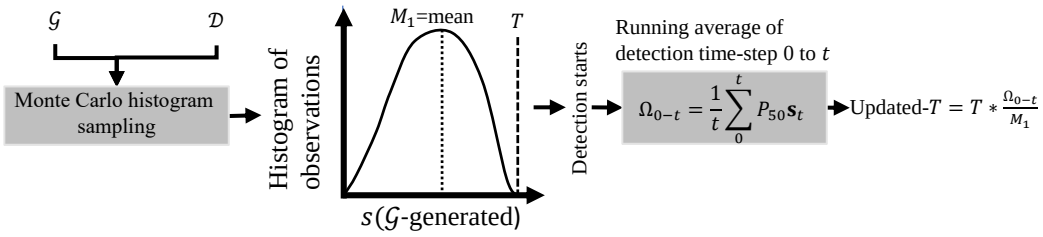
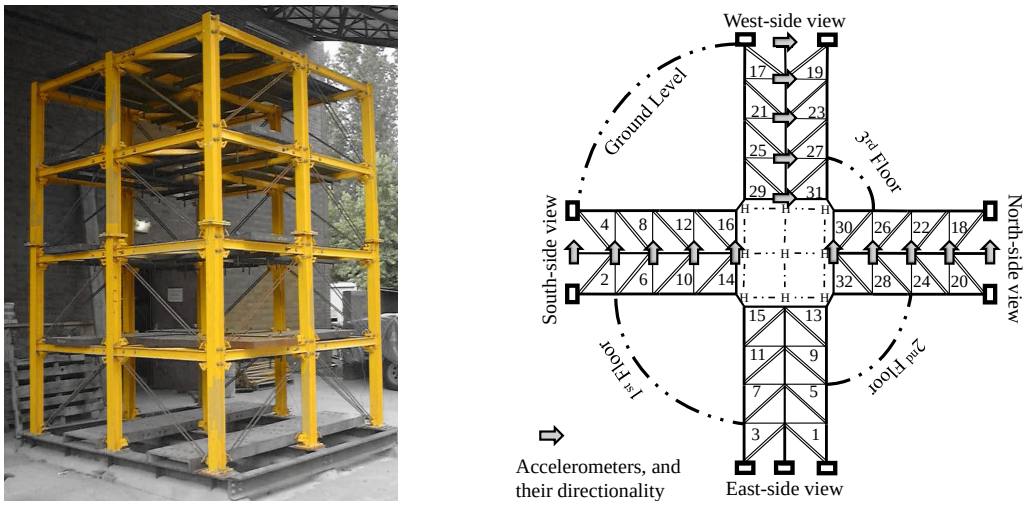


Figure 2.4: Updating the tuned detection threshold T with the incoming data detection scores' statistics.

the detection threshold T results in novelties remaining undetected even though the novelty detection score might have captured the difference, as shown in Fig. 2.3. The objective of novelty detection is to capture the sudden rise of novelty detection scores while moving from "Detection/Same Class" to "Detection/Novel Class" (Fig. 2.3). To this end, the threshold T obtained from the Tuning phase is updated, as shown in 2.4. The rationale behind the update is that the tuned threshold imperfect from the Tuning phase shows the extent of a normal data instance detection score's deviation from zero, i.e., ideal novelty detection score of normal data. Since the detection score is not zero for the incoming data (Fig. 2.3), the distribution of T is updated with their running average. Ultimately, while assessing the trade-off, comparative analyses, such as Receiver Operating Characteristic analysis, are introduced in Section 2.3.4 to remove the possible impacts of imperfectly tuned T on the trade-off assessment.

2.3 Experimental Details

This study employs the developed framework (Fig. 2.1) with the initial FTT-based feature and features obtained from the application of various dimensionality reduction techniques, such as PCA, KernelPCA, and Autoencoders, briefly described in this section. We perform experiments on two benchmark SHM datasets.



(a) (b)

Figure 2.5: Yellow Frame: (a) structure (b) brace and sensor configuration.

2.3.1 Datasets

Yellow Frame Dataset

The Yellow Frame structure is a well-established several benchmark SHM dataset. The three-story, one-third scale steel frame, shown in Fig. 2.5a, contains modular elements, such as masses, braces, beams, and columns [28]. The adopted dataset has brace removal as damage. The brace and sensor configurations are shown in Fig. 2.5, with details of brace removal reported in Table 2.1.

Qatar University Grandstand Simulator Dataset

The Qatar University Grandstand Simulator dataset (QUGS) is established to perform SHM on stadium models [6]. The structure and the configuration of sensors are shown in Fig. 2.6. Additional details on the structure and its instrumentation are available in Avci *et al.* [7]. In QUGS, damages are induced with loosened bolts at joints. An accelerometer is installed

Table 2.1: Yellow Frame data classes.

Data class	Brace removal (Fig. 2.5)	Data class	Brace removal (Fig. 2.5)
DC1	None	DC7	2(II)
DC2	(2, 4)(II)	DC8	(2, 4)(I)
DC3	(2, 4, 18, 20)(II)	DC9	(25, 27)(I)
DC4	(1, 2, 3, 4, 17, 18, 19, 20)(II)	DC10	(29, 31, 8, 6)(I)
DC5	(2, 4, 17, 19)(II)	DC11	(21, 23 ,29 , 31)(I)
DC6	(2, 4)(II) + (18, 20)(I)		

[I] One brace is removed
[II] Both braces are removed

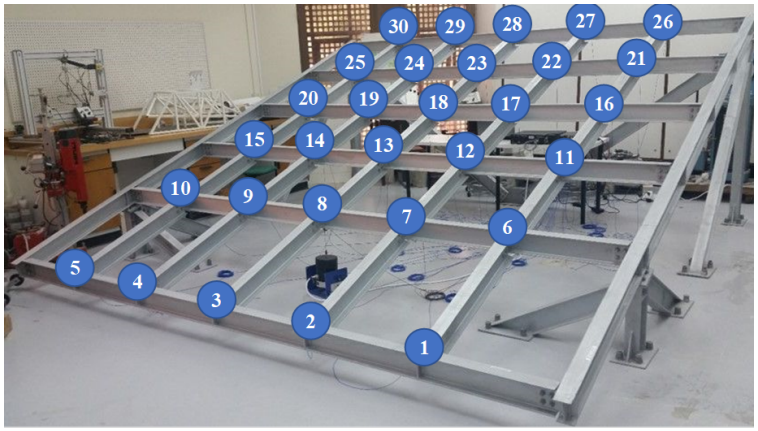


Figure 2.6: QUGS structure [6].

in the vicinity of all joints. Eleven data classes are selected from thirty-one damage scenarios in the original dataset. The employed dataset contains one normal and ten damaged states. Ten sensors (i.e., data channels) per selected defects' locations are considered. The dataset is summarized in Table 2.2.

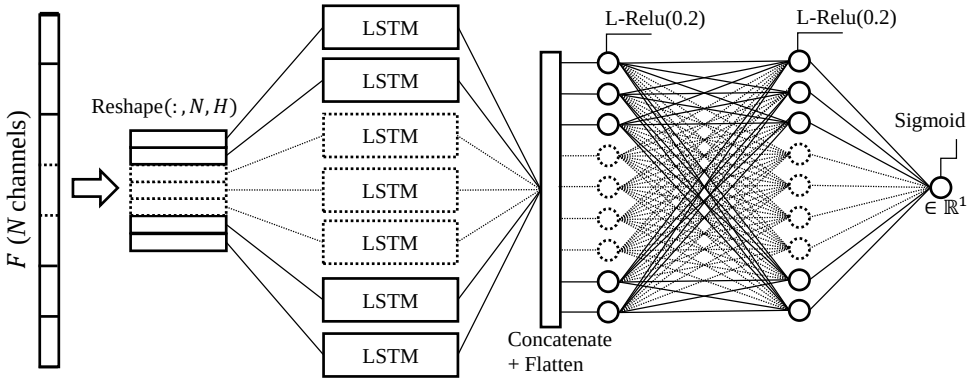
2.3.2 Training and Hyper-parameter Details

The generator $\mathcal{G}$ architecture is a three-layer, fully-connected neural network. The initial layer samples from P_z with 100 dimensions, followed by layers with 256 and 512 neurons.

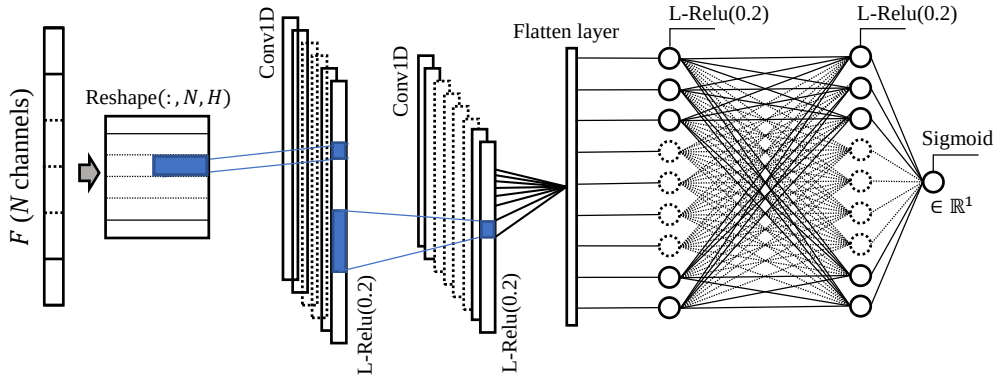
Table 2.2: QUGS data classes.

Data class	Loosened bolt (Fig. 2.6)	Data class	Loosened bolt (Fig. 2.6)
SDC1	None	SDC2	1
SDC3	5	SDC4	7
SDC5	9	FDC6	13
SDC7	18	SDC8	22
SDC9	24	SDC10	26
SDC11	30		

The final layer has F neurons. For the discriminator $\mathcal{D}$, the model is an LSTM neural network, shown in Fig. 2.7a. Additionally, we perform ablation studies with three types of discriminator neural networks, dense, convolutional (CNN), and recurrent (RNN). The CNN-based discriminator $\mathcal{D}$ is built with 1D-CNN layers, shown in Fig. 2.7b. Both LSTM and CNN-based $\mathcal{D}$ are channel-based, in which data from each channel are inputs to separate LSTM or CNN layers. The dense discriminator $\mathcal{D}$ has four layers with neurons equal to F dimension, 1024, 512, and 1. The final layer has a sigmoid activation function. The first three layers have Leaky ReLU activation functions with a negative slope of 0.02 [26]. The application of DR results in features of fewer dimensions. Consequently, $\mathcal{D}$ and $\mathcal{G}$ architectures (i.e., number of neurons) are updated regarding the new input dimensions, maintaining a constant ratio of parameters in $\mathcal{G}$ and $\mathcal{D}$. The GAN model is optimized via the ADAM optimizer [22]. This optimizer has three hyperparameters: learning rate, the exponential decay rate for the first (β_1), and second moment (β_2) estimations. A learning rate of 1e-4, $\beta_1 = 0.9$, and $\beta_2 = 0.999$ are set. The target epoch is 20000 with a batch size equal to L, i.e., the number of data instances in the training set.



(a) Neurons are updated based on data dimensionality.



(b) Conv1D is a one-dimensional convolutional neural network layer; neurons are updated based on data dimensionality.

Figure 2.7: Channel-based $\mathcal{D}$ architectures: (a) LSTM-based $\mathcal{D}$ (b) CNN-based $\mathcal{D}$

2.3.3 Dimentionality Reduction Methods

We experiment with three DR techniques, PCA, Kernel-PCA, and Autoencoders (AE) (Section 2.3.3). Each sensor is a separate data source, and DR techniques are applied individually to each sensor. At the onset of detection, or once a novel class is detected, the framework captures the first L data instances from all sensors to train the GAN model (Fig. 2.1). The sensor-wise application of DR is included in the training phase with the following four steps. First, training data instances from each sensor are fed to DR to have features with the target dimension of d_t. Second, features obtained from the training set are normalized

to have a mean of zero and a unit standard deviation. Third, GAN is trained with those features. Finally, during the Detection phase, incoming data is fed to the DR, normalized, and novelty detection is carried out.

We apply DR to the FFT feature to observe its impact on novelty detection performance. We additionally study the impact of regularization. Non-regularized DR strictly preserves informative data dimensions while removing non-informative and noisy ones. Regularization hinders DR from overfitting informative dimensions, letting non-informative ones survive. Since novel states are unseen, we argue that removing noisy dimensions is sub-optimal as they might contain the most distinctive features of unseen novelties. Regularization is both explicitly and implicitly applicable. L2 regularization, for instance, explicitly regularizes the loss function with the λ parameter. Implicit regularization occurs by reducing DR models' complexity, e.g., fewer target dimensions. Different target dimensions provoke implicit regularization for all DR models, the same as the less-complex AE-II compared to AE-I. Explicit regularization is tried for AE-based DR.

Acceleration data points are turned into signals with the window size of $W = 1000$ for both datasets. Consequently, F has 7500 and 5000 dimensions for Yellow Frame and QUGS, respectively, with each channel producing 500 dimensions of F. Three target dimensions, d_t, of 100, 50, and 10 are considered. Details of AE architectures are shown in Figs. 2.8a and 2.8b. AE-I (Fig. 2.8a) is the more complex AE with a higher chance to overfit its training set compared to AE-II (Fig. 2.8b). No pre-training is intended for the autoencoders. The same ADAM optimizer as for the GAN training is used (Section 2.3.2). The L2 regularization is applied to AE-I with λ of 0.01 and 0.025. No regularization is defined for AE-II as the less complex architecture of AE-II acts as an implicit regularization.

The application of DR has two consequences. The first is the loss of information, and the second is the change in dimensions. Investigating the first consequence is the objective of this

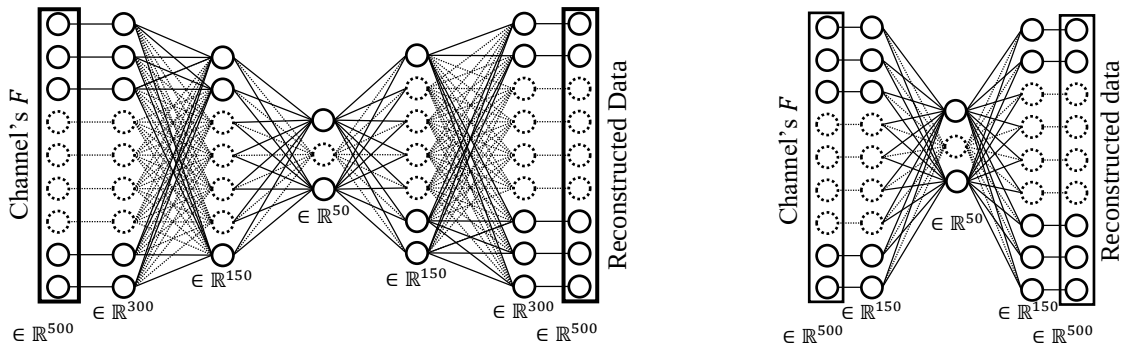


Figure 2.8: AE architectures for the target dimension of 50: (a) AE-I (b) AE-II.

study, whereas the latter consequence might impact observations. A change in dimensions alters neural network architectures, layers, and neurons, which can, in turn, change the detection results. The initial feature and the ones obtained through DR are used for the trade-off assessment. Aside from those features, F is built from fewer sensors and is fed to the detection framework to ensure the framework performance is not impacted by the number of dimensions, or if it does, it helps interpret the results. To this end, only three and two sensors are chosen from Yellow Frame and QUGS, respectively. Those sensors emulate the reduction in dimensionality from 500 to 100, but no data dimensions are lost. F, i.e., the input of GAN, has 5000 dimensions with ten sensors in QUGS, dropping to 1000 dimensions with two sensors. Moreover, Considering FFT as the main feature, and regardless of sensor quantities, frequencies from 0 to $F_s/2$ are captured. On the contrary, some dimensions are not represented in DR-based features. Three sensors are on the first three floors on the west side of the frame (fig. 2.5a), and two sensors are at positions 12 and 13 of the QUGS structure (Fig. 2.6). To sum up, PCA, Kernel-PCA, and AE-I ($\lambda = 0.001, 0.025$) are utilized for target dimensions (i.e., d_t) of 100, 50, and 10. AE-II is examined only for $d_t = 100$ to assess the impact of model complexity.

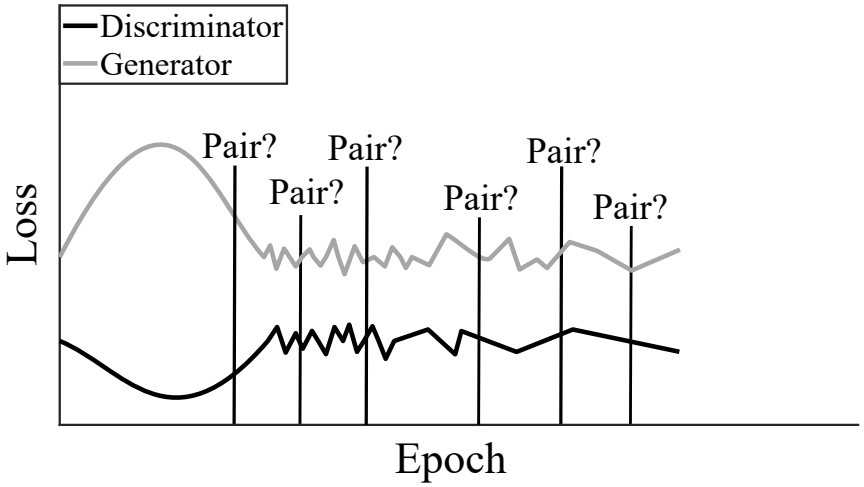


Figure 2.9: Variation in $\mathcal{G}$-$\mathcal{D}$ pairs.

2.3.4 Comparative Analyses

The core of the novelty detection framework is a GAN model, from which a $\mathcal{D} - \mathcal{G}$ pair is required to initiate the novelty detection process (Fig. 2.1). A schematic of a GAN's training loss is shown in Fig. 2.9. According to this figure, a trained GAN contains diverse $\mathcal{G}$-$\mathcal{D}$ pairs equal to the number of its training iterations, which is $20,000$ herein (Section 2.2.3). Training neural networks optimize a loss function, and the best termination iteration is unknown with unseen testing samples. For instance, a fast termination in a supervised setting leads to underfitting, whereas overfitting occurs with excessive training. In this study, and to have a thorough examination of the framework's performance, deterministic and stochastic $\mathcal{G}$-$\mathcal{D}$ pair selections are practiced.

Before presenting comparative analyses, the concept of permutation detection needs to be specified. The order of data classes in an *online*, unsupervised novelty detection process alters the detection accuracy. The reason is that data classes have different extents of similarities and differences. For example, DC7 (Table 2.1) might be identified from DC6 but not from DC9. The order of classes impacts detection results and is excluded by permuting

data classes. Total permutation equals 11! for utilized datasets with eleven data classes. For online novelty detection starting from DC1 to DC11, the true alarm ratio (TAR) is defined as detected novelties divided by ten, and the false alarm ratio (FAR) equals false alarms divided by total detection iterations performed in the Detection phase. Regarding permutation detection, $\overline{TAR}$ and $\overline{FAR}$ are defined as the mean of TAR and FAR across those permutations.

The first comparative analysis includes a deterministic selection of $\mathcal{G}$-$\mathcal{D}$ pairs. Accordingly, iterations 7500, 10000, 12500, 15000, and 17500 are selected results in $5^{11} * 11! = 1.94e + 15$ instances of online novelty detection. Suppose the analysis starts from the first data class, and one of the five pairs is selected to initiate the novelty detection process. The same process is re-iterated whenever a novel class is identified. GAN models are trained for each data class on their training portion beforehand to schedule the massive analysis. Those previously trained $\mathcal{D}$-$\mathcal{G}$ pairs are called at the moment their corresponding class is identified. For this analysis and concerning F and DR-based features under the variety of target dimensions, $\overline{TAR}$ and $\overline{FAR}$ are the comparative metrics to assess the trade-off. This analysis, hereafter, is called Deterministic Analysis.

The novelty detection process can be idealized as a binary classification with the TAR and FAR scores to draw Receiver operating characteristic (ROC) curves. TAR closely mimics the true alarm ratio in ROC curves. For the false alarm ratio, the Stern false alarm ratio ($SFAR$) is defined as the number of data classes with at least one false alarm divided by 11. The ratio is named "Stern" since a whole class is labeled as false per a single false alarm to discern differences between different data (i.e., with and without DR). ROC curves are built upon $\overline{TAR}$ and $\overline{SFAR}$ across the five selected epochs. The ROC curve is essential as it excludes T from the accuracy estimation. The novelty detector's sensitivity to novel data is measured with AUC, the area under ROC curves. This comparative analysis is termed as

ROC Analysis hereafter.

The third analysis includes a stochastic selection of $\mathcal{G}$-$\mathcal{D}$ pairs through a Monte Carlo histogram sampling scheme. The accuracy metric is TAR for each stochastic selection of $\mathcal{D}$-$\mathcal{G}$ pairs. Henceforward, this analysis is called Stochastic Analysis.

The fourth analysis is designed to remove T from novelty detection altogether. Tuned thresholds may lead to zero false alarms. Such situations suggest lowering thresholds to estimate the framework's sensitivity to novelties more accurately. Suppose a class has zero false alarms in Stochastic Analysis with a known threshold. That threshold is reduced while no false alarm is detected, with the novelty detection re-iterated for that class on that particular permutation. Likewise, the threshold is increased until no false alarm is detected if false alarms are present. That analysis completely removes all threshold-related impacts on novelty detection. It reveals novelty detection accuracy's sensitivity to the loss of information in the fairest manner. In other words, this analysis aims to show how results vary when threshold is tuned such that the false alarm rate, otherwise termed false positive rate (FPR) is zero. The analysis is thereafter termed FPR0 Analysis. FPR0 is no longer unsupervised since Stochastic Analysis results are needed for carrying it out.

ROC Analysis and FPR0 measure two different aspects of novelty detection. FPR0 measures novelty detection accuracy with zero false alarms, which is ideal for SHM, or what we call "novelty distinguishing capability". In contrast, ROC Analysis displays novelty detection accuracy at different false alarm regimens, or what we call "sensitivity to novelties". As explained, $\mathcal{G} - \mathcal{D}$ pairs selection and threshold T directly impact the novelty detection accuracy. The impact of threshold T is not present in those analyses per their design. The $\mathcal{G} - \mathcal{D}$ pair selection is removed for FPR0 with the stochastic selection. For ROC Analysis, five epochs spanning training iterations are selected to alleviate the impact of epoch selection. All comparative studies are conducted for the LSTM-based $\mathcal{D}$, with only FPR0, i.e., the most

comprehensive analysis, performed for Dense and CNN-based $\mathcal{D}$ to inspect the consistency in results.

2.4 Results and Discussion

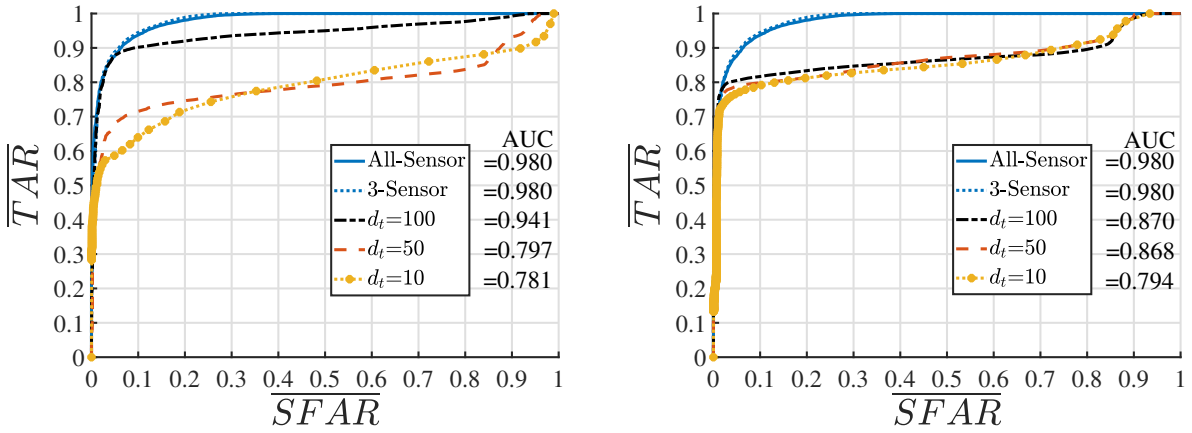
Comparative analyses are performed, and observations are discussed to provide insightful answers to the research questions. The first question is whether the loss of information due to DR or the curse of dimensionality is more detrimental to unsupervised SHM (i.e., the trade-off)? The second question concerns the validity of deeming informative dimensions in the observed data to be damage-sensitive features. In this Section, "All-Sensor" refers to F without DR, while "2-Sensor" and "3-Sensor" refer to Yellow Frame and QUGS data with 3 and 2 sensors, respectively.

2.4.1 Yellow Frame Deterministic Analysis and ROC Analysis

Deterministic Analysis of All-Sensor, 3-Sensor, and PCA-based DR is carried out. Resulting $\overline{TAR}$ and $\overline{FAR}$ reported in Table 2.3. Results indicate a reduction in $\overline{TAR}$ with an increasing degree of DR. At the same time, the 3-Sensor scenario has almost the same accuracy as All-Sensor. 3-Sensor accuracy indicates the loss of information and not the number of dimensions that cause the decline. ROC Analysis is conducted as shown in Fig. 2.10a. A low AUC indicates the detection score fluctuation in the observed data to be comparable to the rise of scores at the onset of novelties. Thus, AUC in Fig. 2.10a demonstrates the loss of information (i.e., DR) to lessen the framework's novelty detection sensitivity, while 3-Sensor maintains almost the same sensitivity as All-Sensor. These conclusions are made for only the Deterministic Analysis and ROC Analysis with the PCA-based DR.

Table 2.3: Yellow Frame Deterministic Analysis. PCA versus All-Sensor and 3-Sensor.

Accuracy metric	All-Sensor ($d_t = 500$)	3-Sensor ($d_t = 100$)	PCA ($d_t = 100$)	PCA ($d_t = 50$)	PCA ($d_t = 10$)
$\overline{TAR}$	0.940	0.944	0.870	0.709	0.648
$\overline{FAR}$	0.0047	0.0527	0.0098	0.0272	0.0048



(a) (b)

Figure 2.10: Yellow Frame ROC Analysis: (a) PCA versus All-Sensor and 3-Sensor (b) Kernel-PCA versus All-Sensor and 3-Sensor.

Deterministic Analysis results of Kernel-PCA and AE-based DR are reported in Table 2.4 and 2.5. Likewise, ROC Analysis results are shown in Fig. 2.10b, and 2.11. There are three observations in the results concerning the research questions. The first observation regards model complexities. AE-based DR has the lowest novelty detection sensitivity compared to Kernel-PCA and PCA-based DR. Also, the less-complex autoencoder network, AE-II, outperforms the non-regularized AE-I. The second observation regards d_t. A lower d_t incites both regularization (i.e., less complexity) and additional loss of information. A smaller d_t in PCA-based DR reduces the accuracy. However, the impact of lower d_t on Kernel-PCA and AE-based DR's accuracy is mixed. Accuracy is decreased moving to $d_t = 50$ then increased with further reduction of d_t to 10 in Table 2.4. The third observation purely interests the regularization in AE-based DR. Studying the impacts of regularization on novelty detection

Table 2.4: Yellow Frame Deterministic Analysis. Kernel-PCA versus All-Sensor and 2-Sensor.

Accuracy metric	All-Sensor ($d_t = 500$)	2-Sensor ($d_t = 100$)	Kernel-PCA ($d_t = 100$)	Kernel-PCA ($d_t = 50$)	Kernel-PCA ($d_t = 10$)
$\overline{TAR}$	0.940	0.944	0.756	0.789	0.771
$\overline{FAR}$	0.0047	0.0527	0.0028	0.0136	0.0153

sensitivity benefits the interpretation of the second observation, distinguishing between the loss of information and regularization. Regularization lessens the accuracy decline. Nonetheless, intensifying the regularization with a higher λ is not always accompanied by a better novelty detection accuracy (e.g., AE-I with d_t of 10, Table 2.5).

The first observation indicates the non-generalizability of SHM data. It also questions the assumption that the observed data's most informative dimensions also extend to novel classes. The reason is that adding more complexity to a non-regularized DR increases the chance of overfitting. Regularization improves the accuracy while a lower d_t declines it. Lowering d_t incites both regularization and more information loss. Consequently, the loss of information dominates the benefits of DR. Generally speaking, regularization's impact on generalizability is nonlinear as excess regularization invokes underfitting. Instances of underfitting are discernible in Table 2.3 when increasing λ (i.e., excessive regularization), which makes regularization a non-reliable solution to improving DR performance for SHM problems.

2.4.2 Yellow Frame Stochastic Analysis

The most unbiased comparison is through Stochastic Analysis and FPR0. The Monte Carlo histogram sampling includes the variation of $\mathcal{G}$-$\mathcal{D}$ pairs in the estimation of accuracy. Stochastic Analysis of LSTM-based $\mathcal{D}$ is shown in Fig. 2.12 for all scenarios. FPR0

Table 2.5: Yellow Frame Deterministic Analysis AE.

DR	$d_t = 100$ $\overline{TAR}$ ($\overline{FAR}$)	$d_t = 50$ $\overline{TAR}$ ($\overline{FAR}$)	$d_t = 10$ $\overline{TAR}$ ($\overline{FAR}$)
AE-I	0.534 (0.0048)	0.534 (0.0043)	0.589 (0.0203)
AE-I, $\lambda = 0.01$	0.650 (0.0050)	0.607 (0.0005)	0.725 (0.0152)
AE-I, $\lambda = 0.025$	0.703 (0.0038)	0.679 (0.0024)	0.717 (0.0127)
AE-II	0.775 (0.0103)	-	-

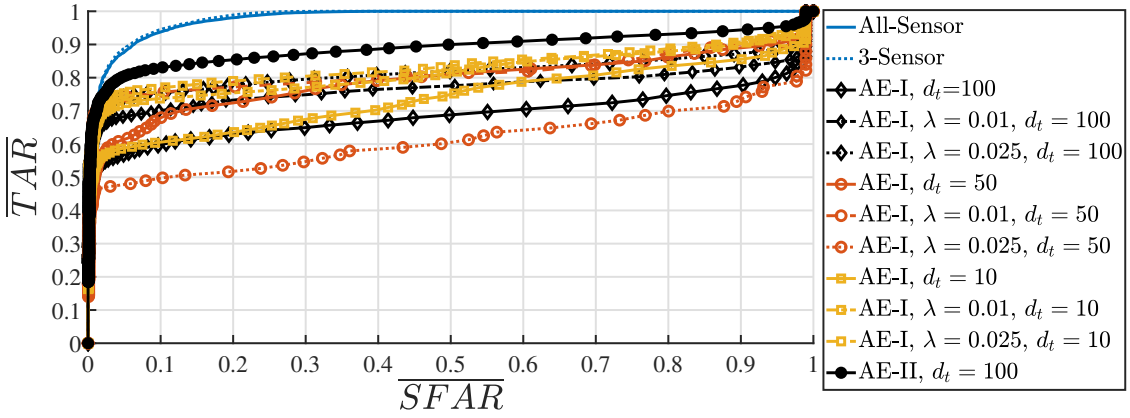


Figure 2.11: Yellow Frame ROC Analysis. AE versus All-Sensor and 3-Sensor.

accuracy histograms are built utilizing Stochastic Analysis outputs, depicted in Fig. 2.13.

We observe similar outcomes as in the previous analysis. PCA-based DR maintains the highest accuracy among DR methods. In contrast, 3-Sensor maintains the same novelty detection performance as All-Sensor. All previous conclusions are valid with the same trends in the stochastic assessment regarding the novelty detection performance and the trade-off. Those conclusions are further evaluated by completing the same analyses on QUGS in the following sections.

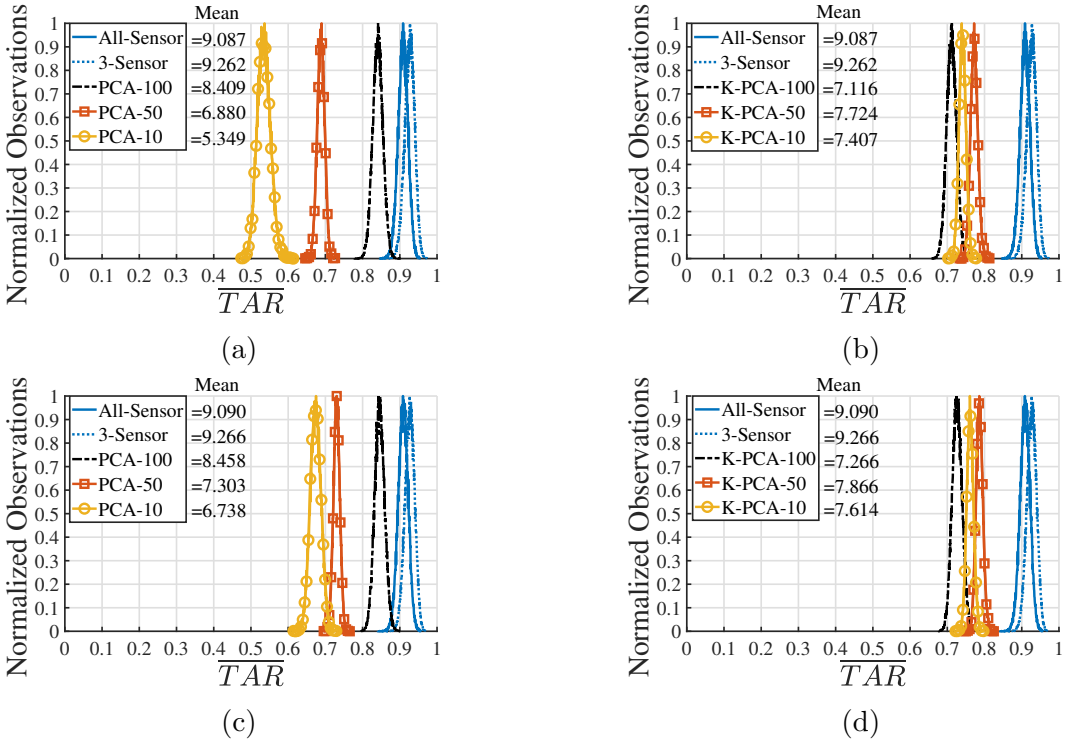


Figure 2.12: Yellow Frame Stochastic Analysis: (a) PCA,(b) Kernel-PCA, (c) (FPR0) PCA,(d) (FPR0) Kernel-PCA

2.4.3 QUGS Deterministic Analysis and ROC Analysis

QUGS Deterministic Analysis is reported in Tables 2.6, 2.7, and 2.8 for PCA, Kernel-PCA, and AE-based DR, respectively. Results imply the same trend regarding the trade-off and regularization. Results show (1) DR-based novelty detection to lower novelty detection sensitivity, with the higher reduction with a more complex DR, (2) regularization dwindles the decline in novelty detection sensitivity, while 2-Sensor maintains the novelty detection sensitivity of All-Sensor. An interesting observation regarding the regularization-induced underfitting happens for DR with lower d_t. In Table 2.8 it occurs with d_t of 50 and 10, and for Yellow Frame, it happened only with d_t of 10 (Table 2.5). Fewer d_t conveys a less complex DR (i.e., AE with fewer neurons), which is more prone to underfitting. This observation further confirms the earlier conclusions on DR. The noted reason in Section 2.3.4 to design

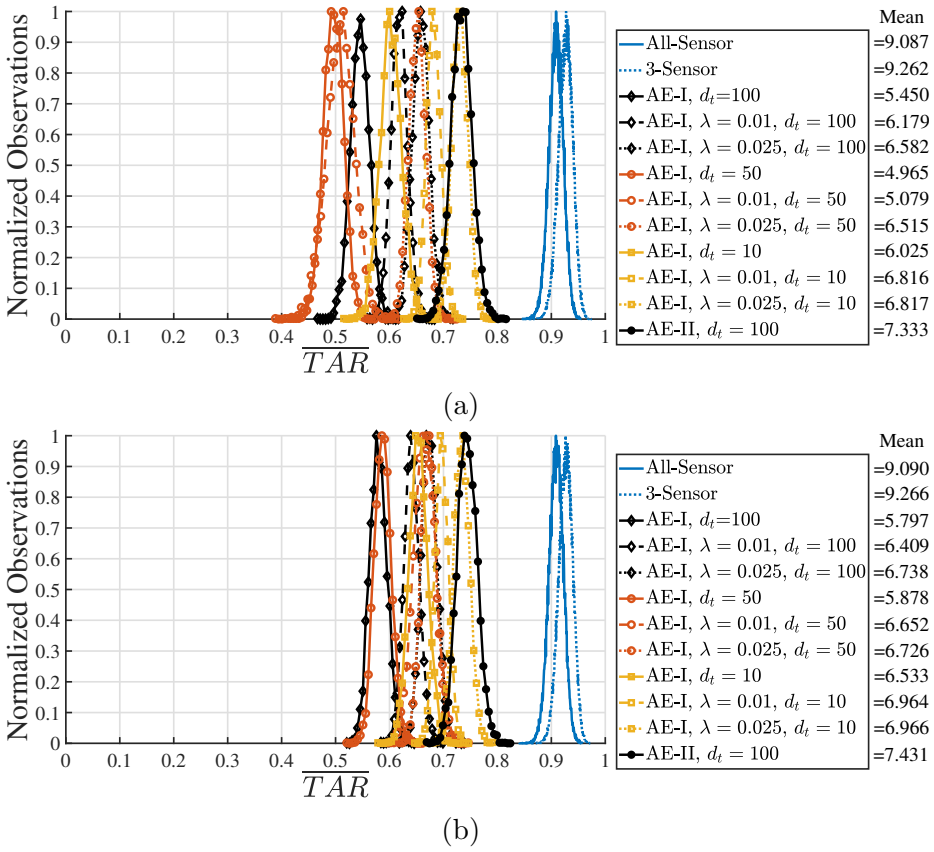


Figure 2.13: Yellow Frame Stochastic Analysis (top) and FPR0 (bottom) results; AE-based DR.

Table 2.6: QUGS Deterministic Analysis, PCA versus All-Sensor and 2-Sensor.

Accuracy metric	All-Sensor ($d_t = 500$)	2-Sensor ($d_t = 100$)	PCA ($d_t = 100$)	PCA ($d_t = 50$)	PCA ($d_t = 10$)
$\overline{TAR}$	1.000	0.996	0.718	0.744	0.790
$\overline{FAR}$	0.0000	0.0000	0.0000	0.0000	0.0000

the FPR0 is observable in Table 2.6, 2.7, and 2.8. $\overline{FAR}$ is zero across the selected epochs. Zero $\overline{FAR}$ may indicate the imperfect threshold tuning and the possibility of lowering them to more precisely estimate TAR and the framework's sensitivity to novelties.

ROC Analysis results are shown in Fig. 2.14a, 2.14b, and 2.15, for PCA, Kernel-PCA, and AE-based DR, respectively. Regarding the ROC curves in Figs. 2.10a and 2.14a an

Table 2.7: QUGS Deterministic Analysis, Kernel-PCA versus All-Sensor and 2-Sensor.

Accuracy metric	All-Sensor ($d_t = 500$)	2-Sensor ($d_t = 100$)	Kernel-PCA ($d_t = 100$)	Kernel-PCA ($d_t = 50$)	Kernel-PCA ($d_t = 10$)
$\overline{TAR}$	1.000	0.996	0.652	0.698	0.753
$\overline{FAR}$	0.0000	0.0000	0.0000	0.0000	0.0000

Table 2.8: QUGS Deterministic Analysis AE.

DR	$d_t = 100$ $\overline{TAR}$ ($\overline{FAR}$)	$d_t = 50$ $\overline{TAR}$ ($\overline{FAR}$)	$d_t = 10$ $\overline{TAR}$ ($\overline{FAR}$)
AE-I	0.549 (0.0000)	0.404 (0.0000)	0.504 (0.0000)
AE-I, $\lambda = 0.01$	0.803 (0.0000)	0.926 (0.0000)	0.967 (0.0000)
AE-I, $\lambda = 0.025$	0.907 (0.0000)	0.789 (0.0000)	0.946 (0.0000)
AE-II	0.778 (0.0006)	-	-

interesting phenomenon is observable. As before with the Yellow Frame dataset, the highest novelty detection performance has resulted without the application DR. However, in the QUGS data set, decreasing the number of PCA and Kernel-PCA dimensions improved the sensitivity to novelties and increased accuracy. The opposite behavior was observed in the Yellow Frame data set. Yellow Frame F has 7500 dimensions compared to 5000 dimensions in QUGS. Thus, PCA in QUGS has a higher chance of overfitting than Yellow Frame, with the regularization side of lower d_t wins over the information loss.

2.4.4 QUGS Stochastic Analysis

Stochastic Analysis and FPR0 results are shown in Figures 2.16 for PCA and Kernel-PCA. Likewise, results are shown in Fig. 2.17 for AE-based DR. Stochastic Analysis supports the observation of novelty detection performance made for Yellow Frame. Results follow the

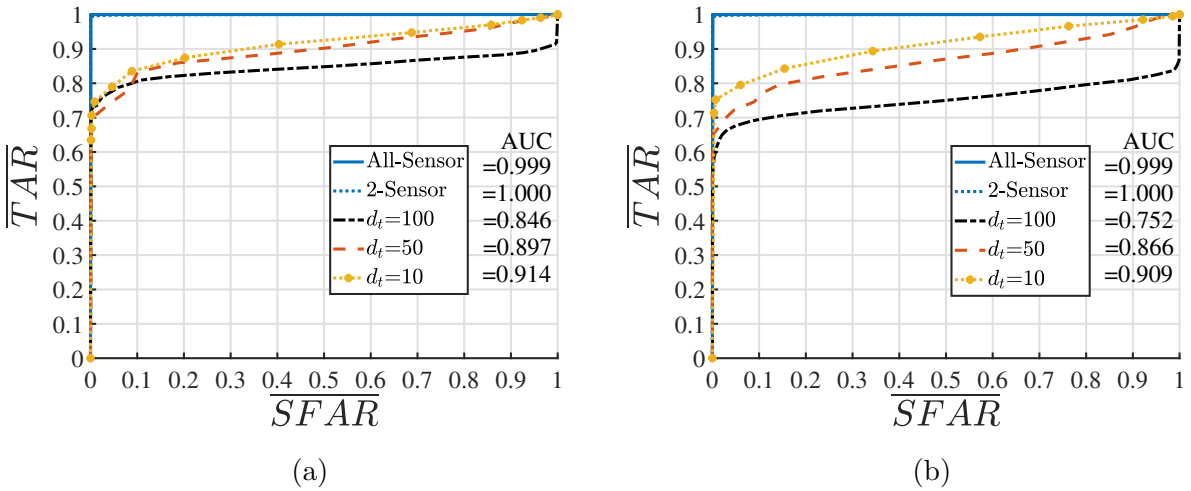


Figure 2.14: QUGS ROC Analysis: (a) PCA versus All-Sensor and 2-Sensor (b) Kernel-PCA versus All-Sensor and 2-Sensor.

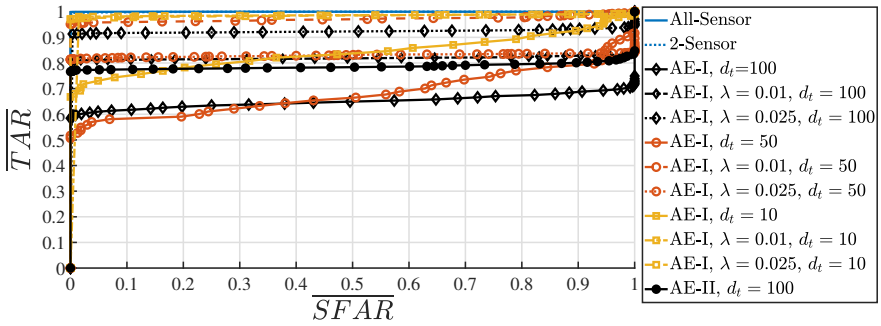


Figure 2.15: QUGS ROC Analysis AE versus All-Sensor and 2-Sensor.

same fluctuation of accuracy due to the application of DR, with regularization to mitigate its impact. The most complex DR, AE, maintains the lowest accuracy as with the Yellow Frame. Nevertheless, the implicit regularization increases AE novelty detection performance to up to 95% of All-Sensor, higher than the best performance of PCA. Still, the underfitting occurs even more intensely. It now occurs for two d_t of 50 and 10 where λ of 0.01 performs better than λ of 0.025. This observation is also present for AE-based DR Deterministic Analysis, noted beforehand. QUGS initial feature has fewer dimensions than Yellow Frame. Accordingly, AE has fewer neurons (i.e., less complex) and is more prone to underfitting.

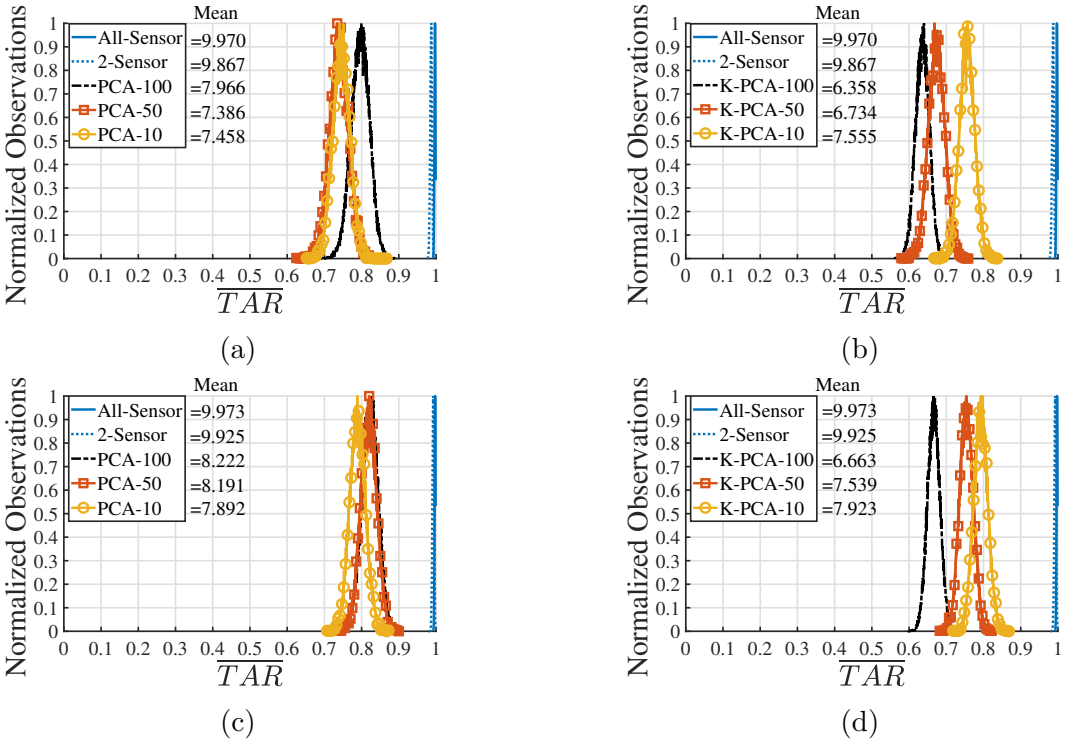


Figure 2.16: QUGS Stochastic Analysis: (a) PCA, (b) Kernel-PCA, (c) (FPR0) PCA, (d) (FPR0) Kernel-PCA

2.4.5 Yellow Frame versus QUGS

Comparing the impact of DR on both datasets, observations from Yellow Frame that QUGS supports are listed as follows. (1) Change in the target dimension alters the DR complexity, serving as regularization if the complexity is reduced. Excluding the regularization from the fewer target dimensions, accuracy declines with decreasing target dimensions (i.e., a higher loss of information). This observation emphasizes that patterns in observed data classes are not generalizable to unseen novelties. It further shows that the undesirable impacts of the curse of dimensionality (e.g., noise and co-linearity in data) that DR mitigates are preferable to the loss of information that DR causes. (2) With the same target dimension, a more complex DR results in a larger accuracy decline. This observation questions the discarded less informative dimensions under the assumption that informative dimensions in observed

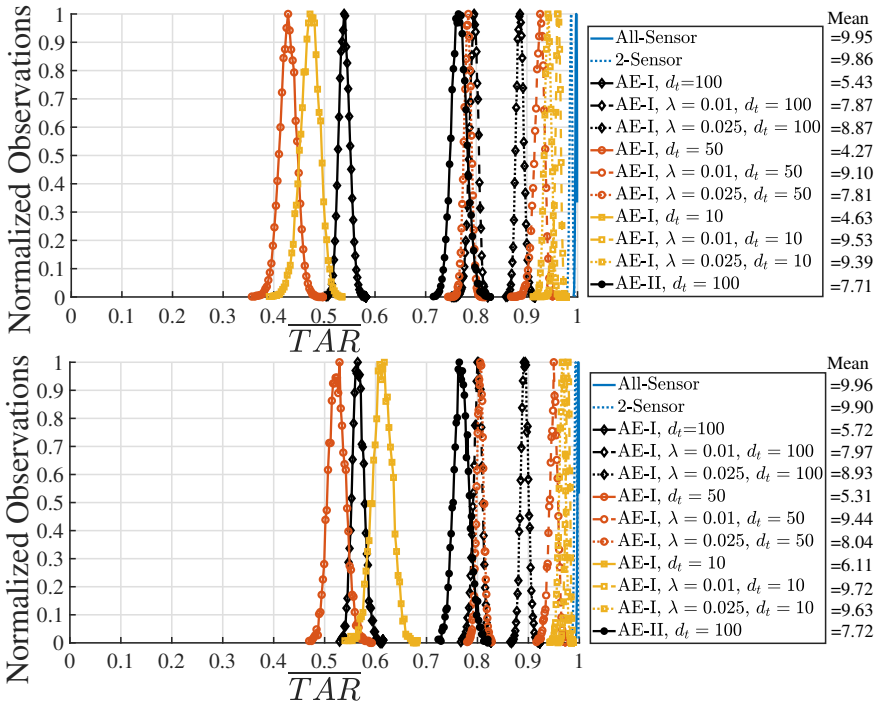
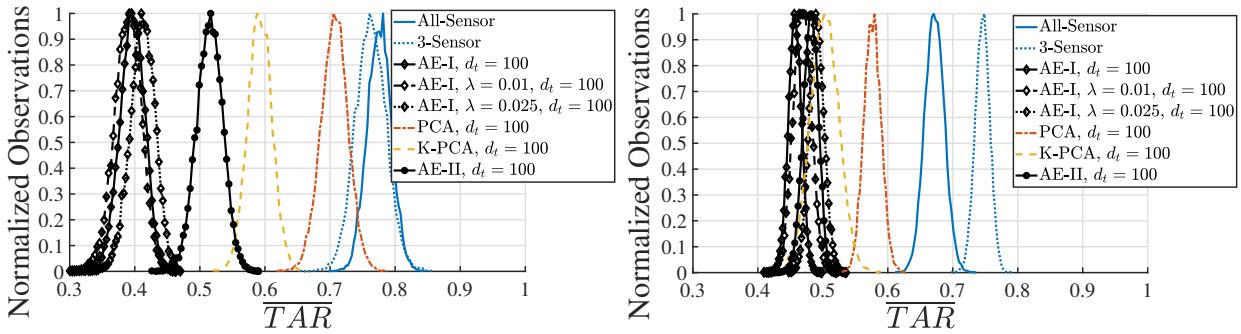


Figure 2.17: QUGS Stochastic Analysis (top) and FPR0 (bottom) results; AE-based DR.

data classes are better novelty indicators than other dimensions. (3) Regularization lessens the negative impacts of DR. This observation likewise negates discarding less-informative dimensions and suggests employing regularized DR while developing novelty detectors for real-world implementations. However, the optimal choice of regularization hyperparameters is not practical in an unsupervised setting with a regularized DR susceptible to causing even less favorable outcomes.

2.4.6 Stochastic Analysis; Dense and CNN-based Discriminator Models

Dense and CNN-based $\mathcal{D}$ were designed in Section 2.3.2 to investigate whether results generalize to other model architectures. FPR0 is performed exclusively considering Yellow Frame. Results are shown in Fig. 2.18b and 2.18a for Dense and CNN-based architectures,



(a) (b)

Figure 2.18: Yellow Frame Stochastic Analysis: (a) CNN-based $\mathcal{D}$ (b) Dense-$\mathcal{D}$.

respectively. The same conclusions as with prior analyses are drawn from these figures. PCA-based DR has superior accuracy, but even the best case cannot match the full data dimension's accuracy. Regularization also improves the decline in the accuracy of AE-I, though unpredictably. For instance, λ of 0.025 can result in decreased accuracy than λ of 0.01. Also, the less complex AE-II achieves higher accuracy than AE-I.

An interesting remark on these figures is the impact of the curse of dimensionality on Dense-based $\mathcal{D}$. Compared to CNN and LSTM, Dense architecture has the least potential for understanding sequential features. Approximating structures with modes, the FFT of their response is unique for each structural state. The sequence of spectral lines (i.e., frequencies) and their magnitudes define those unique structural states. That behavior makes F a sequence. CNN and LSTM better comprehend All-Sensor and 3-Sensor sequential features than Dense networks. Those layers are designed to understand spatial (CNN) and temporal (LSTM) dependencies. Consequently, Dense layers face data interpretability issues caused by the curse of dimensionality, and with fewer data dimensions obtained from only three sensors, accuracy increases. Nevertheless, DR is more detrimental to accuracy than the curse of dimensionality, even with the Dense network.

Chapter 3

Conclusion

Dimensionality reduction (DR) is a tool to alleviate intricacies associated with high-dimensional features, i.e., the curse of dimensionality. Complications of high-dimensional data include sparse search areas or distance metrics that become less distinctive across data classes. DR, however, comes at the cost of information loss. There exists a trade-off between the lost information and the benefits DR. The balance of that trade-off is task-dependent. The target task can be supervised, unsupervised, or anomaly detection on both categories interacting with unseen data distributions. This study explores the trade-off regarding unsupervised novelty detection with structural health monitoring (SHM) data. We examine two benchmark SHM datasets to answer two essential questions. First, whether DR-induced loss of information or the curse of dimensionality is more detrimental to novelty detection accuracy. And second, whether supposing unseen novelties to be detectable with the informative data dimensions of observed data classes to be taken as damage-sensitive features. That supposition is currently practiced in the literature in the lack of a general unsupervised SHM approach operating with raw data due to the shortcomings of methods such as clustering functioning high-dimensional features (i.e., meaningless distance metric in high-dimensional spaces). An unsupervised SHM method operating with high-dimensional features is compared with three DR methods, PCA, Kernel PCA, and Autoencoders. Several comparative analyses are designed to assess the resulting novelty detection accuracy of the low-dimensional features versus the original high-dimensional data.

We conclude that DR-induced loss of information is more detrimental than the curse of dimensionality. For example, features derived from AE can reduce novelty detection accuracy by up to 60%. Regularization and the complexity of DR impact the change of accuracy. For AE models, explicit L2-regularization and reducing the number of neural network layers (i.e., implicit regularization) reduce the accuracy decline. The simplest DR, PCA, has the best accuracy across DR models, still inferior to the accuracy obtained from the initial raw feature. Regularization alleviates DR's negative impacts on the accuracy, though a non-proper hyper-parameter selection can further reduce it. Regularization hinders DR from overfitting the highly informative dimensions, allowing more low-informational ones to survive. Thus, it questions the assumption above, indicating that highly-informative data dimensions in the baseline (i.e., already observed) data class are not reliable for building novelty detectors. However, optimizing regularization hyperparameters is not practical in an unsupervised setting with unknown novel data classes. Thus, this study recommends that more effort be invested in developing unsupervised SHM techniques that function with raw (or close to raw), high-dimensional features in the absence of information about novel classes.

Bibliography

[1] Osama Abdeljaber, Onur Avci, Serkan Kiranyaz, Moncef Gabbouj, and Daniel J Inman. Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *Journal of Sound and Vibration*, 388:154–170, 2017.

[2] Osama Abdeljaber, Onur Avci, Mustafa Serkan Kiranyaz, Boualem Boashash, Henry Sodano, and Daniel J Inman. 1-d cnns for structural damage detection: Verification on a structural health monitoring benchmark data. *Neurocomputing*, 275:1308–1317, 2018.

[3] JP Amezquita-Sanchez, M Valtierra-Rodriguez, and H Adeli. Wireless smart sensors for monitoring the health condition of civil infrastructure. *Scientia Iranica*, 25(6):2913–2925, 2018.

[4] Juan Pablo Amezquita-Sanchez and Hojjat Adeli. Signal processing techniques for vibration-based health monitoring of smart structures. *Archives of Computational Methods in Engineering*, 23(1):1–15, 2016.

[5] Farhad Ansari. *Sensing issues in civil structural health monitoring*, volume 1. Springer, 2005.

[6] Onur AVCI. Qatar university grandstand simulator (qugs), Dec 2018. URL http://onur-avci.com/benchmark/qugs/.

[7] Onur Avci, Osama Abdeljaber, Serkan Kiranyaz, Mohammed Hussein, and Daniel J Inman. Wireless and real-time structural damage detection: A novel decentralized method for wireless sensor networks. *Journal of Sound and Vibration*, 424:158–172, 2018.

[8] Shaeela Ayesha, Muhammad Kashif Hanif, and Ramzan Talib. Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Information Fusion*, 59:44–58, 2020.

[9] Pierre Baldi and Kurt Hornik. Neural networks and principal component analysis: Learning from examples without local minima. *Neural networks*, 2(1):53–58, 1989.

[10] Abd Ennour Bouzenad, Mahjoub El Mountassir, Slah Yaacoubi, Fethi Dahmene, Mahmoud Koabaz, Lilian Buchheit, Weina Ke, et al. A semi-supervised based k-means algorithm for optimal guided waves structural health monitoring: A case study. *Inventions*, 4(1):17, 2019.

[11] LA Bull, TJ Rogers, C Wickramarachchi, EJ Cross, K Worden, and N Dervilis. Probabilistic active learning: An online framework for structural health monitoring. *Mechanical Systems and Signal Processing*, 134:106294, 2019.

[12] LA Bull, K Worden, and N Dervilis. Towards semi-supervised and probabilistic classification in structural health monitoring. *Mechanical Systems and Signal Processing*, 140:106653, 2020.

[13] Francesco Cadini, Luca Lomazzi, Marc Ferrater Roca, Claudio Sbarufatti, and Marco Giglio. Neutralization of temperature effects in damage diagnosis of mdof systems by combinations of autoencoders and particle filters. *Mechanical Systems and Signal Processing*, 162:108048, 2022.

[14] Russel E Caflisch. Monte carlo and quasi-monte carlo methods. *Acta numerica*, 7:1–49, 1998.

[15] Rharã de Almeida Cardoso, Alexandre Cury, and Flavio Barbosa. Automated real-time

damage detection strategy using raw dynamic measurements. *Engineering Structures*, 196:109364, 2019.

[16] Pascal De Boe and Jean-Claude Golinval. Principal component analysis of a piezosensor array for damage localization. *Structural health monitoring*, 2(2):137–144, 2003.

[17] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification (pt. 1)*, volume 2. Wiley-Interscience, 2000.

[18] Karl Pearson F.R.S. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2 (11):559–572, 1901. doi: 10.1080/14786440109462720.

[19] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[20] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.

[21] XG Hua, YQ Ni, ZQ Chen, and JM Ko. Structural damage detection of cable-stayed bridges using changes in cable forces and model updating. *Journal of structural engineering*, 135(9):1093–1106, 2009.

[22] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[23] Thomas K Landauer and W Kintsch. Latent semantic analysis. *Automated Essay Scoring: A Cross-disciplinary Perspective*, page 87, 2003.

[24] John A Lee and Michel Verleysen. *Nonlinear dimensionality reduction*. Springer Science & Business Media, 2007.

[25] Qiming Luo, Lepeng Huang, Xuanyi Xue, Zengshun Chen, Fengbin Zhou, Lihao Wei, and Jianmin Hua. Occupational health risk assessment based on dust exposure during earthwork construction. *Journal of Building Engineering*, 44:103186, 2021.

[26] Andrew L Maas, Awni Y Hannun, Andrew Y Ng, et al. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, page 3. Citeseer, 2013.

[27] Wentao Mao, Wushi Feng, Yamin Liu, Di Zhang, and Xihui Liang. A new deep auto-encoder method with fusing discriminant information for bearing fault diagnosis. *Mechanical Systems and Signal Processing*, 150:107233, 2021.

[28] A Mendler, CE Ventura, and S Allahdadian. The yellow frame: experimental studies and remote monitoring of the structural health monitoring benchmark structure. In *Special Topics in Structural Dynamics, Volume 5*, pages 233–244. Springer, 2019.

[29] John F Nash et al. Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 36(1):48–49, 1950.

[30] Mohammad Hossein Rafiei and Hojjat Adeli. A novel unsupervised deep learning model for global and local health condition assessment of structures. *Engineering Structures*, 156:598–607, 2018.

[31] Edwin Reynders, Gersom Wursten, and Guido De Roeck. Output-only structural health monitoring in changing environmental conditions by means of nonlinear system identification. *Structural Health Monitoring*, 13(1):82–93, 2014.

[32] Douglas A Reynolds. Gaussian mixture models. *Encyclopedia of biometrics*, 741(659-663), 2009.

[33] TJ Rogers, K Worden, R Fuentes, N Dervilis, UT Tygesen, and EJ Cross. A bayesian

non-parametric clustering approach for semi-supervised structural health monitoring. *Mechanical Systems and Signal Processing*, 119:100–119, 2019.

[34] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Kernel principal component analysis. In *International conference on artificial neural networks*, pages 583–588. Springer, 1997.

[35] Zhiqiang Shang, Limin Sun, Ye Xia, and Wei Zhang. Vibration-based damage detection for bridges by deep convolutional denoising autoencoder. *Structural Health Monitoring*, page 1475921720942836, 2020.

[36] Ehsan Sheybani and Giti Javidi. Dimensionality reduction and noise removal in wireless sensor network datasets. In *2009 Second International Conference on Computer and Electrical Engineering*, volume 2, pages 674–677. IEEE, 2009.

[37] Mohammad Hesam Soleimani-Babakamali, Roksana Soleimani-Babakamali, and Rodrigo Sarlo. A general framework for supervised structural health monitoring and sensor output validation mitigating data imbalance with generative adversarial networks-generated high-dimensional features. *Structural Health Monitoring*, page 14759217211025488, 2021.

[38] Mohammad Hesam Soleimani-Babakamali, Reza Sepasdar, Kourosh Nasrollahzadeh, Ismini Lourentzou, and Rodrigo Sarlo. Toward a general unsupervised novelty detection framework in structural health monitoring. *Computer-Aided Civil and Infrastructure Engineering*, 2022.

[39] Mohammad Hesam Soleimani-Babakamali, Reza Sepasdar, Kourosh Nasrollahzadeh, and Rodrigo Sarlo. A system reliability approach to real-time unsupervised structural health monitoring without prior information. *Mechanical Systems and Signal Processing*, 171:108913, 2022.

[40] Michela Torti, Ilaria Venanzi, Simon Laflamme, and Filippo Ubertini. Life-cycle management cost analysis of transportation bridges equipped with seismic structural health monitoring systems. *Structural Health Monitoring*, 21(1):100–117, 2022.

[41] Sarehati Umar, Norhisham Bakhary, and ARZ Abidin. Response surface methodology for damage detection using frequency and mode shape. *Measurement*, 115:258–268, 2018.

[42] William Soo Lon Wah, John S Owen, Yung-Tsang Chen, Ahmed Elamin, and Gethin Wyn Roberts. Removal of masking effect for damage detection of structures. *Engineering Structures*, 183:646–661, 2019.

[43] William Soo Lon Wah, Yung-Tsang Chen, and John S Owen. A regression-based damage detection method for structures subjected to changing environmental and operational conditions. *Engineering Structures*, 228:111462, 2021.

[44] Duo Wang, Ming Zhang, Yuchun Xu, Weining Lu, Jun Yang, and Tao Zhang. Metric-based meta-learning model for few-shot fault diagnosis under multiple limited data conditions. *Mechanical Systems and Signal Processing*, 155:107510, 2021.

[45] Zilong Wang and Young-Jin Cha. Unsupervised deep learning approach using a deep auto-encoder with an one-class support vector machine to detect structural damage. *Structural Health Monitoring*, page 1475921720934051, 2020.

[46] Ian H Witten and Eibe Frank. Data mining: practical machine learning tools and techniques with java implementations. *Acm Sigmod Record*, 31(1):76–77, 2002.

[47] Keith Worden, Charles R Farrar, Graeme Manson, and Gyuhae Park. The fundamental axioms of structural health monitoring. *Proceedings of the Royal Society A: Mathemat-*

ical, Physical and Engineering Sciences, 463(April):1639–1664, 2007. ISSN 1364-5021. doi: 10.1098/rspa.2007.1834.

[48] Xinya Wu, Yan Zhang, Changming Cheng, and Zhike Peng. A hybrid classification autoencoder for semi-supervised fault diagnosis in rotating machinery. *Mechanical Systems and Signal Processing*, 149:107327, 2021.

[49] Saibo Xing, Yaguo Lei, Shuhui Wang, Na Lu, and Naipeng Li. A label description space embedded model for zero-shot intelligent diagnosis of mechanical compound faults. *Mechanical Systems and Signal Processing*, 162:108036, 2022.

[50] Ai-Min Yan and Jean-Claude Golinval. Null subspace-based damage detection of structures using vibration measurements. *Mechanical Systems and Signal Processing*, 20(3): 611–626, 2006.

[51] Wang-Ji Yan, Dimitrios Chronopoulos, Ka-Veng Yuen, and Yi-Chen Zhu. Structural anomaly detection based on probabilistic distance measures of transmissibility function and statistical threshold selection scheme. *Mechanical Systems and Signal Processing*, 162:108009, 2022.

[52] Xiaoan Yan, Ying Liu, and Minping Jia. Health condition identification for rolling bearing using a multi-domain indicator-based optimized stacked denoising autoencoder. *Structural Health Monitoring*, 19(5):1602–1626, 2020.

[53] Shaochen Yang, Jinling Liu, Xiangyang Bi, Yongqiang Ning, Shengying Qiao, Qianqian Yu, and Jie Zhang. Risks related to heavy metal pollution in urban construction dust fall of fast-developing chinese cities. *Ecotoxicology and Environmental Safety*, 197:110628, 2020.

[54] Wennian Yu, Il Yong Kim, and Chris Mechefske. Analysis of different rnn autoencoder

variants for time series classification and machine prognostics. *Mechanical Systems and Signal Processing*, 149:107322, 2021.

[55] Masoumeh Zareapoor, Pourya Shamsolmoali, and Jie Yang. Oversampling adversarial network for class-imbalanced fault diagnosis. *Mechanical Systems and Signal Processing*, 149:107175, 2021.