infrastructures

MDPI

Article

Simplicial Complex-Enhanced Manifold Embedding of Spatiotemporal Data for Structural Health Monitoring

Nan Xu, Zhiming Zhang and Yongming Liu *

School for Engineering of Matter, Transport and Energy, Arizona State University, Tempe, AZ 85281, USA
* Correspondence: yongming.liu@asu.edu

Abstract: Structural Health Monitoring requires the continuous assessment of a structure's operational conditions, which involves the collection and analysis of a large amount of data in both spatial and temporal domains. Conventionally, both data-driven and physics-based models for structural damage detection have relied on handcrafted features, which are susceptible to the practitioner's expertise and experience in feature selection. The limitations of handcrafted features stem from the potential for information loss during the extraction of high-dimensional spatiotemporal data collected from the sensing system. To address this challenge, this paper proposes a novel, automated structural damage detection technique called *Simplicial Complex* Enhanced Manifold Embedding (SCEME). The key innovation of SCEME is the reduction of dimensions in both the temporal and spatial domains for efficient and information-preserving feature extraction. This is achieved by constructing a *simplicial complex* for each signal and using the resulting topological invariants as key features in the temporal domain. Subsequently, curvature-enhanced topological manifold embedding is performed for spatial dimension reduction. The proposed methodology effectively represents both intra-series and inter-series correlations in the low-dimensional embeddings, making it useful for classification and visualization. Numerical simulations and two benchmark experimental datasets validate the high accuracy of the proposed method in classifying different damage scenarios and preserving useful information for structural identification. It is especially beneficial for structural damage detection using complex data with high spatial and temporal dimensions and large uncertainties in reality.

Keywords: structural health monitoring; manifold learning; damage detection; simplicial complex; *Euler characteristic*

Citation: Xu, N.; Zhang, Z.; Liu, Y. Simplicial Complex-Enhanced Manifold Embedding of Spatiotemporal Data for Structural Health Monitoring. *Infrastructures* **2023**, *8*, 46. https://doi.org/10.3390/infrastructures8030046

Academic Editors: Carlo Rainieri, Andy Nguyen, You Dong and Dmitri Tcherniak

Received: 6 January 2023
Revised: 16 February 2023
Accepted: 17 February 2023
Published: 5 March 2023

1. Introduction

The implementation of Structural Health Monitoring (SHM) is a common practice in various engineering fields for enhancing the safety and reliability of structures. Vibration-based damage detection techniques are particularly prevalent in SHM applications for civil structures, utilizing the dynamic responses (displacement, velocity, acceleration, and strain) collected from sensors to assess the operational conditions of the structure. The field of vibration-based damage detection encompasses both physics-based and data-driven methods. In this study, our focus is on the latter, the data-driven approach. A comprehensive review of various methods for SHM can be found in the literature review by Sony et al. [1].

In SHM, data-driven methods rely on statistical learning techniques to recognize damage-sensitive characteristics that have been derived from structural measurement data, such as time series of acceleration. In statistical learning, there are two main approaches, namely supervised and unsupervised methods. For the purpose of SHM, supervised learning algorithms need data for both damaged and undamaged states for their training process. This enables them to offer insight into the existence, location, extent, and type of damage. Previous work has explored the use of artificial neural networks (ANNs) for post-earthquake damage detection in steel moment-frames, as proposed by Gonzalez and

Zapico [2]. A supervised learning approach for predicting seismic-induced damage was introduced by de Lautour and Omenzetter [3] using ANN for numerically simulated two-dimensional reinforced concrete (RC) frames. Alves et al. [4] evaluated the performance of Bayesian decision trees, ANN, and support vector machines (SVMs) for structural damage detection. Other researchers have applied SVM techniques to detect, estimate, or localize damage using time series models and damage-sensitive features [5–7]. Recently, SVM has been combined with Bayesian optimization for improved hyperparameter tuning, enabling the detection and localization of damage in various structures using cumulative intensity measures as damage-sensitive features [8,9]. Convolutional neural networks (CNNs) have also been used for real-time structural damage detection, including a one-dimensional CNN [10] and fully connected CNNs for detecting and localizing damage in large-scale structures [9,11–13]. Vision-based SHM methods, which utilize images of structures or components to identify and assess their condition, also employ deep learning techniques [9,11]. The main challenge of supervised learning in the context of SHM lies in acquiring data or labels for damaged conditions. This data can be obtained from physics-based modeling or experimental testing, but both methods have limitations. Building an accurate numerical model is difficult when the structure is complex in the context of geometry or composition, while experiments can be expensive or physically limited and are often tested on scaled subjects or components. Furthermore, the damage detection capability of supervised learning is limited to the types of damage considered in the training phase and may not be able to detect emergent damage types. For these reasons, unsupervised learning is more practical for SHM systems. Using classifiers trained on normal conditions of structures only is the most commonly used unsupervised learning method in SHM, as referenced in previous studies such as [14,15]. This approach eliminates the need to acquire data from damaged scenarios, whether from physical testing or numerical simulations, and makes data acquisition more feasible.

The problem of classifying and grouping similar subjects for SHM has proven to be a challenging task by using unsupervised learning techniques. Most existing methods face several limitations such as (a) the difficulty in automating the high dimensional spatiotemporal data representation without relying on handcrafted feature selection; (b) the lack of robustness in detecting damage under large uncertainties and/or noise; and (c) the requirement of large amounts of data for training in order to obtain stable and accurate results. To address these challenges, this paper presents a novel approach called simplicial complex-enhanced topological manifold embedding for structural health monitoring. The proposed approach involves two stages: (i) using a simplicial complex-based method for feature extraction in the temporal domain, and (ii) applying topological manifold embedding for dimension reduction in the spatial domain.

A *simplicial complex*-based method, more specifically topological invariants, is a mathematical tool that can be used to distinguish if two topological subjects are the same. This paper uses the *Euler characteristics (ECs)* as the topological invariant for characterizing temporal behaviors. *EC* is a general topological invariant that is utilized to capture the geometric properties of complex datasets with a reduced dimension and complexity while retaining the most significant information [16–18]. Given a time series, its topological property can be distilled into an *EC curve* through the application of a manifold *filtration* outlined in Section 3.2.2. One important benefit of the *simplicial simplex* is that it is intrinsically robust with respect to noise. For example, squeezed and unsqueezed water balls may appear very differently (e.g., due to random squeezing forces). However, their topological geometries are the same and can be identified from the *simplicial complex*.

The visualization of data with many parameters often involves using dimension reduction techniques. One popular linear method is principal component analysis (PCA) [19]. PCA has been employed to visualize genetic distances within intricate manifolds. Nonlinear dimension reduction techniques, such as t-distributed stochastic neighborhood embedding (t-SNE) [20] and uniform manifold approximation and projection (UMAP) [21], are frequently employed to group high-dimensional vectors with respect to low-dimensional

embeddings. Meanwhile, PCA has the drawback of requiring linear independence among all variables, and t-SNE mainly concentrates on local structure with specific parameter settings, while UMAP may not produce meaningful embeddings. A novel nonlinear dimension reduction approach, referred to as curvature-augmented manifold embedding and learning (CAMEL) [22], has recently been introduced, incorporating information about curvatures in high-dimensional topology. CAMEL demonstrates superior clustering abilities compared to many existing manifold embedding techniques and has been used to convert the spatial information of sensor arrays to a lower-dimensional representation. This study integrates the use of *ECs* and CAMEL to classify various damage scenarios (flowchart shown in Figure 1) through numerical simulations and evaluation using two benchmark datasets. One major benefit of this research for engineering applications is its computational efficiency and resilience to noise in analyzing high-dimensional data.

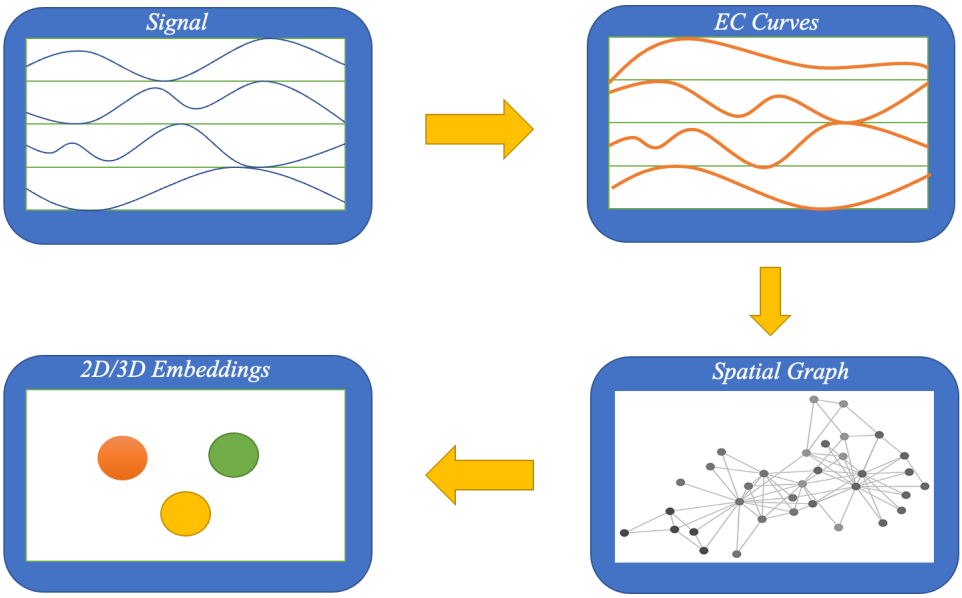


Figure 1. A flowchart of the proposed study.

2. Mathematical Preliminary

In this section, we first discuss some necessary mathematical background materials. Specifically, in Section 2.1, we briefly review simplicial complex, and in Section 2.2 we recall certain topological invariants that will be applied in subsequent sections. The remainder of the section is devoted to setting up the necessary background on differential geometry, focusing on Ricci curvature and Ricci flow. See [23–27] for more detailed discussions.

2.1. Simplicial Complex

In mathematics, a *simplicial complex* (or *simplex*) is a set of points, line segments, triangles, and their n-dimensional counterparts [28]. As shown in Figure 2, one vertex can generate a point, two points can generate a line segment (by connecting the two points), three points generate a triangle (by connecting all pairs of points with segments and filling the space among them), and then we can generate an n-simplicial complex [29]. It should be noted that at least $n+1$ points are required in order to generate a *simplicial complex* in dimension n.

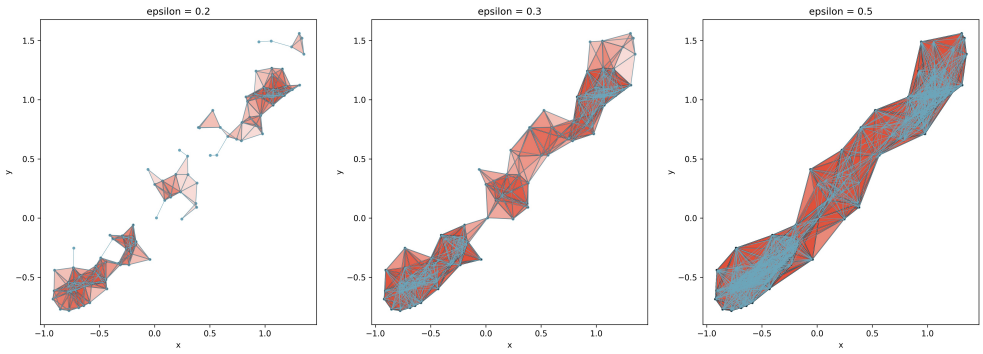


Figure 2. Examples of simplicial complexes. The color of triangles indicate solid surface; epsilon indicates the maximum edge length between any two points.

Now, we give the specific definitions:

1. The *standard k-simplex*, denoted by Δ^k, is the convex span of the elementary basis of $\mathbb{R}^{k+1}$, i.e.,
$$\Delta^k := \left\{ (t_0, t_1, \cdots, t_k) \in \mathbb{R}^{k+1} : \sum_{i=0}^{k} t_i = 1, t_i \geqslant 0, \forall i \right\}.$$
For example, the standard 1-simplex Δ^1 is the (closed) line segment in $\mathbb{R}^2$ connecting the two points $(1, 0)$ and $(0, 1)$.
2. A *k-simplex* is the convex span of $k+1$ geometrically independent points $x_0, x_1, \cdots, x_k$ in $\mathbb{R}^n$ (i.e, $x_1 - x_0, \cdots, x_k - x_0$ are linearly independent); we denote it by $[x_0, x_1, \cdots, x_k]$, and we call these generating points the *vertices* of $[x_0, x_1, \cdots, x_k]$.
3. Deleting any vertex from a k-simplex $[x_0, x_1, \cdots, x_k]$, the $(k-1)$-simplex spanned by the remaining vertices is called a *face* of $[x_0, x_1, \cdots, x_k]$. Moreover, if x_i is the deleted vertex, then the face is denoted by $[x_0, \cdots, \widehat{x_i}, \cdots, x_k]$.
4. A *singular k-simplex* in a topological space X is a continuous map $\sigma : \Delta^k \to X$.

Given the definitions above, if we let $e_0, e_1, \cdots, e_k$ denote the standard basis vectors for $\mathbb{R}^{k+1}$, it is not hard to see that $\Delta^k = [e_0, e_1, \cdots, e_k]$.

The construction of *simplicial complexes* provides topological insights into the spaces where the data points are located. The reason for using *simplicial complexes* is that it helps us to better understand high-dimensional spaces and structures that usually cannot be visualized in a direct way. The theoretical guarantee that we can regard a *simplicial complex* as the same as some high-dimensional space is a so-called *homeomorphism*, which is also called a continuous transformation. If two topological spaces or geometric figures are homeomorphic, then there is a continuous bijection between them that has a continuous inverse map. We do not distinguish homeomorphic spaces, as they are topologically equivalent. Figure 3 shows a famous example of the *homeomorphism* of donuts and mugs.

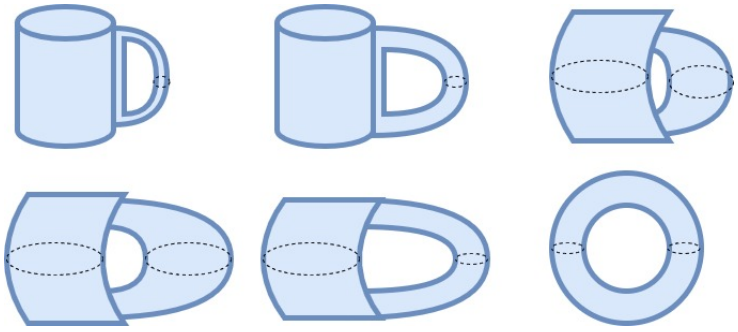


Figure 3. Example of homeomorphism: a mug is topologically equivalent to a donut.

The following properties are met for any *simplicial complexes*.

- Rigid motions, such as rotation, translating, and reflection. We can move or rotate the simplex anywhere, and it still is regarded as the same *simplex*.
- Stretch. We can stretch out any points away from each other and change the connected structures as well.
- We cannot crush a *simplex* from n dimension to $n-1$ dimension by deformation.

2.2. Topological Invariants

One way to study topological spaces is to use a topological invariant, which is a mathematical object (e.g., a number, a polynomial, or a group) assigned to a space. Topological invariants are invariant under *homeomorphisms* and thus can be used to distinguish topological spaces. Here, we discuss three invariants: *singular homology groups*, *Betti numbers*, and the *Euler characteristic*.

We begin with *singular homology groups*. Given a topological space X, recall that a singular k-simplex in X is a continuous map $\sigma : \Delta^k \to X$, where Δ^k can be identified with $[e_0, e_1, \cdots, e_k]$. For such a k-simplex, we set

$$\partial_n(\sigma) := \sum_{i=0}^{k} (-1)^i \sigma|_{[e_0,\cdots,\widehat{e_i},\cdots,e_k]}.$$

We define the *singular group of k-chains* in X, denoted by $C_k(X)$, as the free abelian group generated by the singular k-simplices in X, i.e.,

$$C_k(X) := \left\{ \sum_i n_i \sigma_i : n_i \in \mathbb{Z}, \sigma_i : \Delta^k \to X \text{ continuous}, \forall i \right\},$$

where each $\sum_i n_i \sigma_i$ is a finite formal sum and is called a *singular k-chain* in X. Then, we can extend ∂_n linearly to all singular k-chains and thus define the boundary map

$$\partial_k : C_k(X) \to C_{k-1}(X).$$

One important property of boundary maps, which can be proved using the definition, is that

$$\partial_k \circ \partial_{k+1} = 0.$$

Then, the image $\mathrm{im}(\partial_{k+1})$ is a subgroup of the kernel $\ker(\partial_k)$, which is itself a subgroup of $C_k(X)$. We call $(C_k, \partial_k)_{k\in\mathbb{N}}$ the *singular chain complex* of X, and for each $k \in \mathbb{N}$, we define the k^{th} *singular homology group* of X to be the quotient

$$H_k(X) := \ker(\partial_k)/\mathrm{im}(\partial_{k+1}).$$

The k^{th} *Betti number* of X, denoted by $b_k(X)$, is the rank of the k^{th} singular homology group of X, i.e.,

$$b_k(X) := \mathrm{rank}(H_k(X)).$$

For example, the closed surfaces of the topological space are first introduced and expressed as the first *Betti number* shown in Figure 4. Figure 4 is a good example of the topological invariant, which shows that this invariant is not affected much by high noise.

Then, the *Euler characteristic* of X, denoted by $\chi(X)$, is what we used in this paper introduced in Section 3.2.1.

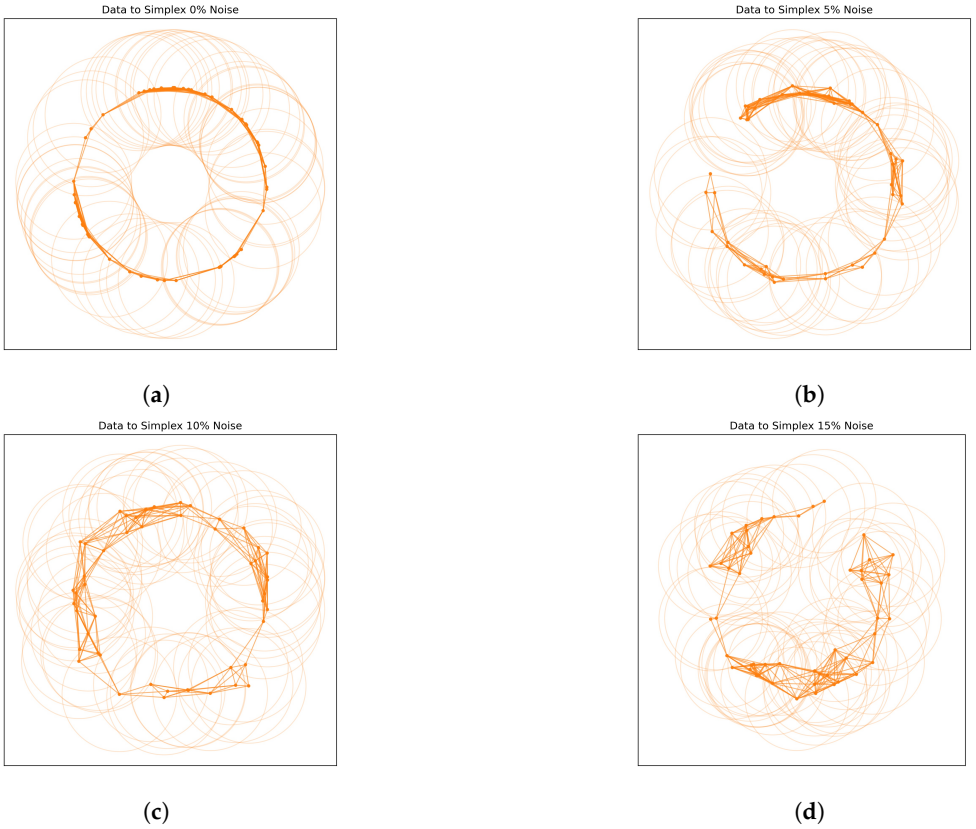


Figure 4. Example of different noise level with same first Betti number 1. (**a**) Noise Level 0, (**b**) Noise Level 5, (**c**) Noise Level 10, (**d**) Noise Level 15.

2.3. Differential Geometry

2.3.1. Ricci Curvature

In Riemannian geometry, the curvature measures how much a smooth object deviates from Euclidean [30], which has been found to encode and express local and global features of smooth manifolds with metric tensors [31]. *Ricci curvature* controls the average dispersion of geodesics around the tangent directions and the growth of the volume of spheres and distance balls, which the following equation can compute [32]:

$$Vol_{\alpha}(\epsilon) = d\alpha\epsilon^{n-1}(1 - \frac{Ric(v)}{3}\epsilon^2 + o(\epsilon^2)) \tag{1}$$

where n is the dimension of the Riemannian manifold, and $Vol_{\alpha}(\epsilon)$ is the $(n-1)$-volume generated within an n-solid angle $d\alpha$ by geodesics of length ϵ in the direction of v. Thus, *Ricci curvature* can represent both divergences of geodesics and volume growth. *Ricci curvature* can be easily converted to the classical Gauss curvature for $n = 2$ case.

2.3.2. Ricci Flow

In 1981, Hamilton presented the concept of the *Ricci flow* [33], which is a differential equation for the Riemannian metric that functions similarly to the diffusion of heat and heat equation. It smoothes out the metric's unevenness [34]. The *Ricci flow* has demonstrated its efficacy in addressing geometric issues [35].

Suppose we have a manifold M with a Riemannian metric g_{ij} and corresponding Ricci curvature R_{ij}. The Hamilton's *Ricci flow* is a second-order partial differential equation on symmetric tensors as:

$$\frac{\partial}{\partial t}g_{ij} = -2R_{ij} \tag{2}$$

The solution to the Ricci flow involves finding a sequence of metrics $g_{ij}(t)$ defined on a smooth manifold M, which satisfies Equation 2. One of the fundamental characteristics of the Ricci flow is that the curvature evolves in accordance with a nonlinear form of the heat equation, leading to a smoothing of the irregular curvature.

3. Materials and Methods

3.1. Materials

3.1.1. Dataset 1

The first benchmark structure under analysis is the ASCE Phase II SHM benchmark experimental structure developed by the IASC-ASCE SHM Task Group [36]. It is a 4-story, 2-bay by 2-bay steel-frame scale-model structure whose footprint dimensions are 2.5 m × 2.5 m, and the height is 3.6 m. Figure 5 shows the benchmark structure. This benchmark model was constructed and tested in the Earthquake Engineering Research Laboratory at the University of British Columbia (UBC). Ambient and forced vibrations were conducted in the tests. Ambient excitations include winds, pedestrian effects, and traffic loads. Forced vibration was conducted using an impact hammer and electrodynamic shaker. The test data were published in 2003 [37]. Fifteen accelerometers were positioned on the structure, with three being placed on each of its four floors and one roof. The placement included one at the east face, one at the west face, and one near the central column. Only twelve accelerometers were used for this study since we ignored accelerometers on the ground. Seven damage cases of the first benchmark structure were simulated under ambient excitation. Each scenario consisted of 20 recorded samples, with each sample featuring 1500 timesteps (a timestep of 0.05 s). Structural benchmark cases are described in Table 1 and Figure 6. More details about this benchmark structure, regarding the structural properties (mass and stiffness assignment, connections, etc.) and dynamic tests, can be found in [36].

Table 1. Dataset 1: First benchmark structural cases.

Case	Description
1	Health
2	No brace on the east side
3	No brace on the SE corner per floor in one bay
4	No brace on the SE corner of the first and fourth floors in one bay
5	No brace on the SE corner of the first floor in one bay
6	No brace on the N side of the second floor
7	No brace for the structure

Figure 5. ASCE experimental SHM benchmark structure: experimental model (excerpt from [38]).

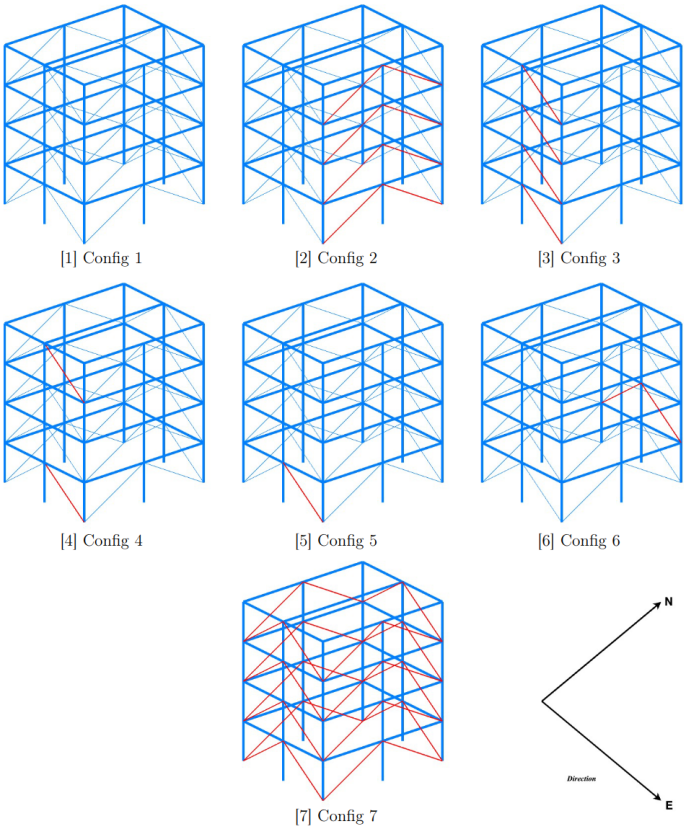


Figure 6. Seven benchmark cases of Structure 1, damage components (red).

3.1.2. Dataset 2

The building shown in Figure 7 is a three-story structure, referred to as the second benchmark building. Its simulated data was published by Los Alamos National Laboratory [39]. The structure features aluminum columns and plates connected using bolted joints, with a rigid base. Each floor of the structure comprises 4 aluminum columns, measuring 17.7 cm in length, 2.5 cm in width, and 0.6 cm in thickness. These columns are attached to top and bottom aluminum plates, which are 30.5 cm by 30.5 cm and 2.5 cm in thickness, creating a 4-degree of freedom system. A center column, which is 15.0 cm in length, 2.5 cm in width, and 2.5 cm in thickness, hangs from the top floor and may cause nonlinear behaviors when it interacts with a bumper located on the second floor. The position of the bumper can be adjusted to vary the level of nonlinearity. For the purposes of this study, a shaker was mounted on a base plate (76.2 $\times$ 30.5 $\times$ 2.5 cm) to provide excitation, and a force transducer was used to measure the force from the shaker at the end of the stingers. The response was measured on the opposite side from the excitation source on each floor using four accelerometers. Each damage case includes 40 recorded samples (with the exception of Case 2, which has 160 samples), and each sample has 2000 timesteps (where each timestep equals 3.1 ms). The description of the structural benchmark can be found in Table 2.

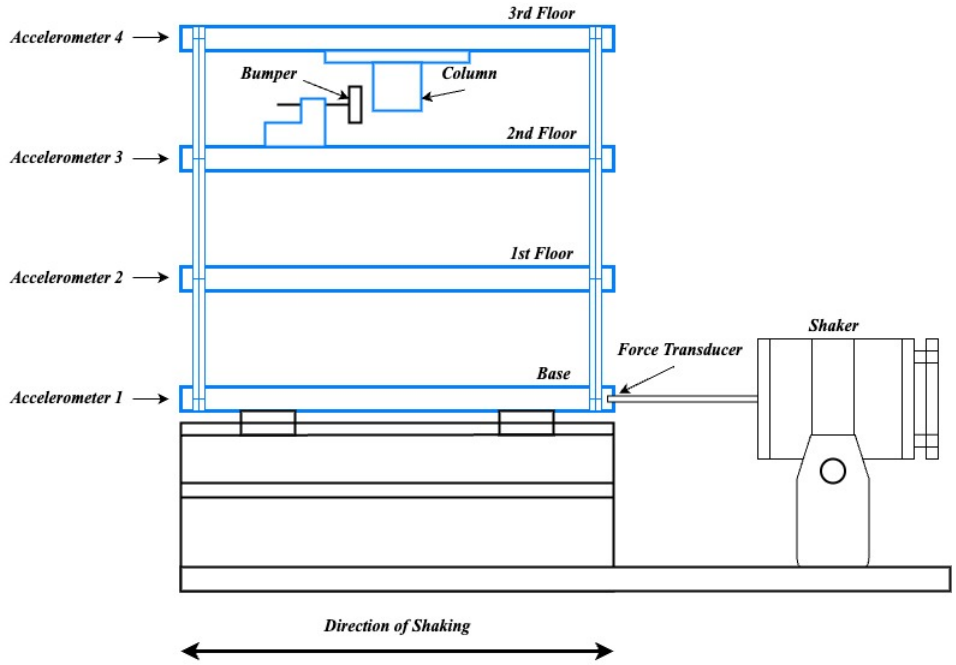


Figure 7. Benchmark Structure 2: the three-story frame structure.

Table 2. Dataset 2: description of structural benchmark cases.

Structural Case	Conditions
1	Mass on the 1st floor
2	Add gap between the bumper and the suspended column
3	Column: 1BD – 50% stiffness reduction
4	Column: 3BD – 50% stiffness reduction
5	Column: 2AD + 2BD – 50% stiffness reduction
6	Column: 2AD + 2BD – 50% stiffness reduction

3.1.3. Dataset 3

The third dataset utilized is a simulation of data obtained from the simulated benchmark structure of Dataset 1. The simulation was generated by using a 12 Degrees of Freedom (DOF) finite element model of the structure for the purpose of evaluating its suitability as a benchmark for SHM/DD problems [40]. This model took into consideration the rigid body movement of each floor, including translation in the $x-$ and $y-$ directions and rotation of θ around the center column. Additional information on the finite element analysis can be found in [40]. In this study, we present binary classification examples that encompass two damage patterns: (i) removal of all braces in the 1st story and (ii) removal of all braces in the 1st and 3rd stories.

Since the topological invariant is very robust to noise [41], the proposed method is anticipated to exhibit the same characteristic. To validate this, a quantitative analysis has been carried out to examine the impact of noise on classification accuracy. The noise level is quantified following the procedures outlined in [40]:

1. The finite element models give a structure model in terms of active DOF q, which is related to physical DOF by $x = Tq$;
2. The motion equation is $M\ddot{q} + C_d\dot{q} + Kq = T^T f$, where f is a vector of forces applied to the physical DOF, and M and K are mass and stiffness matrices;
3. A total of 16 accelerometers, two each in the $x-$ and $y-$ directions per floor;
4. Hence, we can return noisy sensor measures $\ddot{y} = Cq + Df + v$, where v is a sensor noise vector, the elements of which are Gaussian pulse process with RMS $NoiseLevel$%

of the largest RMS of the acceleration responses (typically one of the roof accelerations), and C is based on the solution of standard $K\Theta = M\Theta A$ eigenvalue problems.

3.2. Methods

In this section, we propose a *Simplicial Complex* Enhanced Manifold Embedding (SCEME) that can cluster different damage scenarios of benchmark structures. SCEME includes two parts. The first part uses topological invariants to describe temporal data. The second part is the dimension reduction of spatial features, where curvature-augmented manifold learning is applied to obtain the corresponding embedding. A detailed methodology is introduced in the following subsections.

3.2.1. Euler Characteristic

In this paper, we use *Euler characteristic (EC)* as the topological invariant for temporal data, which means if two objects are topologically equivalent, they have the same *Euler Characteristic*. The *Euler characteristic* is also called *Euler numbers*, which is used to describe the topological structures and shapes no matter what types of rigid motion have been applied.

The original definition of *EC* is based on the number of vertexes (V), the number of edges (E), and the number of faces (F) as shown in Equation (3).

$$\chi = V - E + F \tag{3}$$

which is a simple equation that only focuses on polyhedra but fails to prove that it is a topological invariant.

An alternative definition of *EC* is based on homology. For any topological spaces, suppose we have the n^{th} *Betti number* b_n as the rank of the n^{th} *singular homology group*. The *Euler characteristic* can then be defined as follows:

$$\chi = \Sigma_0^n (-1)^n b_n \tag{4}$$

If the *Betti numbers* exhibit finiteness and reach zero starting from a specific index n_0, we will have a well-defined *EC* [42].

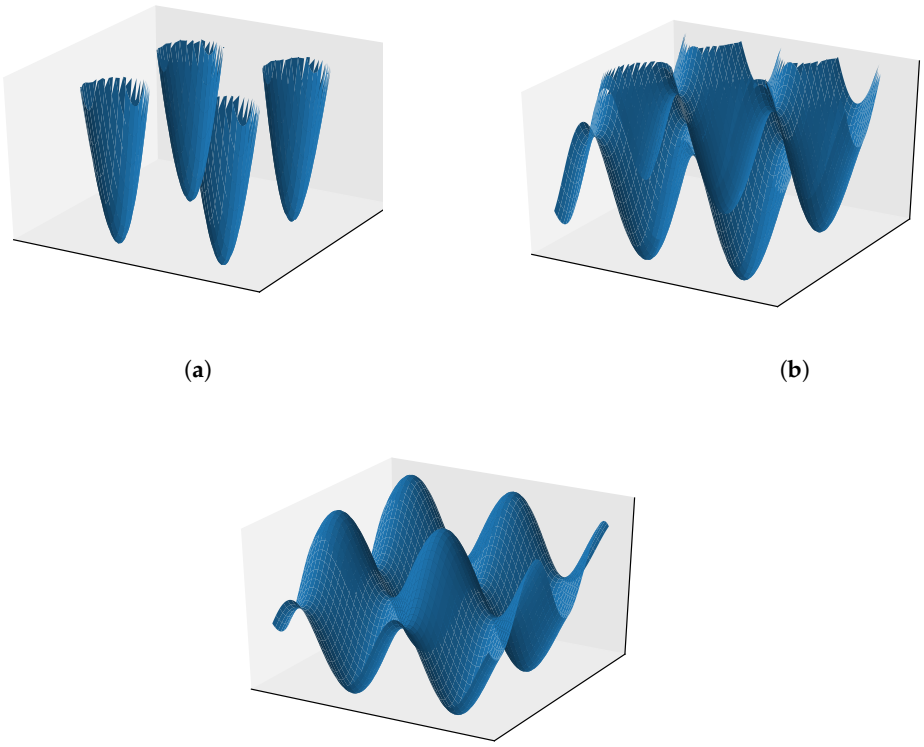
3.2.2. Manifold Filtration

The calculation of the corresponding *EC* of datasets in the temporal domain D_t requires the establishment of a suitable *simplicial complex*. To achieve this, a cubical complex is utilized due to its favorable computational performance, as noted in a study by Ziou et al. [43]. For a given n-dimensional manifold M and a continuous function $f : M \to R$, a sublevel set M_{k_i} is defined as a collection of points $x \in M : f(x) \leq k_i$, with $k_i \in R$ representing the *filtration* threshold. According to Poincaré [44], nested sublevel sets can be generated by incrementally increasing the filtration thresholds:

$$M_{k_1} \subseteq M_{k_2} \subseteq \ldots \subseteq M_{k_n} \subseteq M \tag{5}$$

where k_i denotes the respective thresholds of filtration. The topological features of these sequenced sublevel sets can be quantified by computing the *EC* at each filtration threshold $\chi_0, \chi_1, \ldots, \chi_{n-1}$. By obtaining an ordered pair-value k_i, χ_i, it is possible to outline the topological feature of the manifold and the corresponding function. In this work, 20 different filtration values $k_0 \leq k_1 \leq \cdots \leq k_{19}$ were selected to correspond to the 20 sublevel sets. It should be noticed that there is not much difference in performance when the number of filtrations is greater than 15 in the study. The *EC* was calculated by determining the sum of all n-dimensional topological bases, also known as the total *Betti number* of each base (b_0, b_1, $\ldots$, b_{n-1} for a *n-manifold*). An example of the 3-manifold filtration procedures can be seen in Figure 8. At the first filtration value k_0, the sublevel set comprises 4 connected components, reflecting 4 local minima in the function f, resulting in $\chi_0 = 4$. When the filtration increases

to k_1, only 1 connected component remains, but 2 holes appear, suggesting the existence of local maxima in the function f, yielding an *EC* value of $\chi_1 = 1 - 2 = -1$. At the last filtration k_{19}, the original manifold is restored, with $\chi_2 = 1$. Therefore, the resulting *EC Curve* is $[4, -1, 1]$.

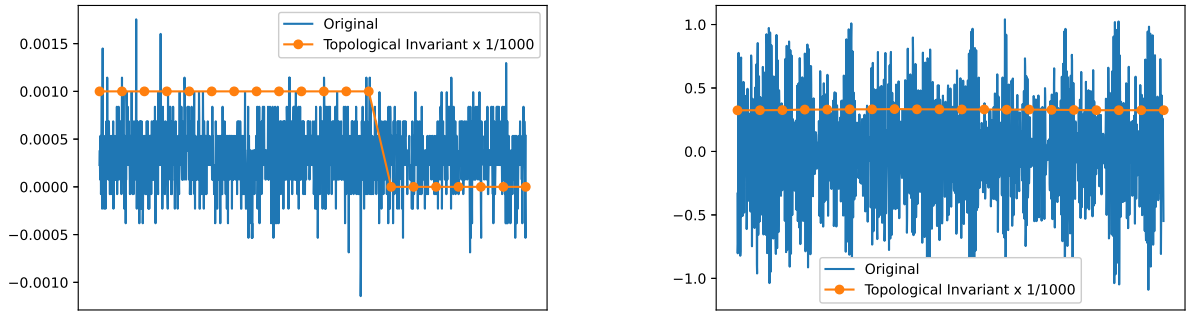


(**a**) (**b**)

(**c**)

Figure 8. An example of *filtration* process: (**a**) k_0; (**b**) k_2; (**c**) k_{19}.

The filtration process of a *simplicial complex* involves removing vertices or edges that have data values below a specified threshold, as depicted in Figure 2. The computation of filtration and topological properties, as represented by the *EC*, are efficient and capable of handling large, high-dimensional datasets. The filtration procedure results in an EC curve, a function that shows the changes in topology through the filtration process, according to [45]. The *EC curve* is a valuable and easily understandable representation of complex data objects, offering an advantage over information gained from persistence diagrams, as noted in [16,18,45]. An illustration of *EC curves* for both datasets is shown in Figure 9.



(**a**) (**b**)

Figure 9. *EC Curves*: (**a**) Dataset 1, Sensor 16; (**b**) Dataset 2, Sensor 4.

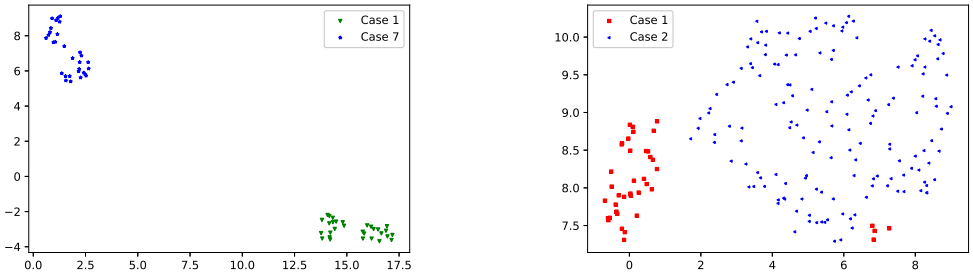
3.2.3. Curvature Enhanced Manifold Learning

Manifold learning, also known as nonlinear dimension reduction, is employed to derive a two-dimensional or three-dimensional representation of the series of structures' *ECs*. The main idea of this paper is that the patterns formed by *ECs* sequences can be embedded in a 2-manifold or 3-manifold, denoted as M^n where $n \leq 3$, and that these patterns exhibit a distinct topology in the low-dimensional space, which leads to the formation of distinguishable clusters. The paper utilizes the CAMEL algorithm to obtain the low-dimensional representation of *ECs* sequences, with four primary assumptions: (a) the data are evenly distributed on the initialized Riemannian manifold, (b) the Riemannian metric is locally constant, (c) the manifold is locally connected, and (d) the final manifold is optimized using *Ricci flow* [22].

4. Results

4.1. Binary Classification

As mentioned above, our algorithm includes two parts: temporal dimension reduction and spatial dimension reduction. Hence, each temporal response is represented by an *EC* curve (20 points for each curve) for both datasets. Therefore, every sample has 320 features for Dataset 1, while every single sample has 100 features for Dataset 2. The purpose of the first example is to evaluate the ability of the proposed study to categorize damage. This demonstration includes three binary classification examples. In the first example, there are 20 samples from Case 1 (healthy state) and 20 samples from Case 7 (all braces removed). The second example, referred to as Dataset 2, includes 40 samples from Case 1 and 120 samples from Case 2 as the last instances. The corresponding result is shown in Figure 10 with 100% accuracy (see Section 4.3 for the definition of accuracy) for Dataset 1 and 98.13% for Dataset 2, respectively.

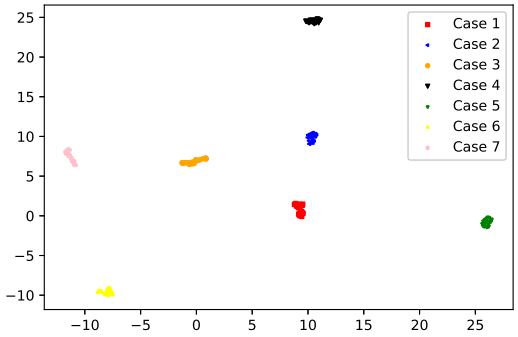


(**a**) (**b**)

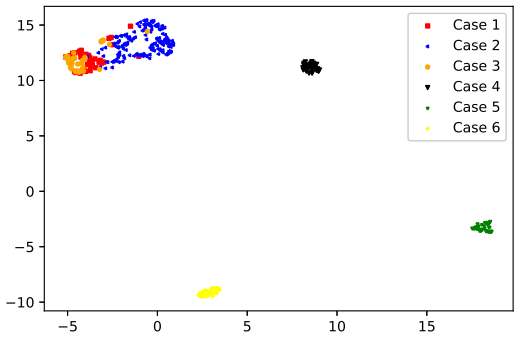
Figure 10. Low-dimensional representation of binary classification for both datasets: (**a**) Dataset 1: Case 1 (Green), Case 7 (Blue); (**b**) Dataset 2: Case 1 (Red), Case 2 (Blue).

4.2. Multi-Labels Classification

Demonstration 2 includes all cases from both benchmark structures in order to demonstrate the overall performance of the proposed study. The first example shows the low dimensional representations from Dataset 1, while the second example shows the corresponding low dimensional representation of Dataset 2 as presented in Figure 11. It is shown that the results have a perfect separation for Dataset 1, which indicates almost perfect classification results. Cases 1–3 of Table 2 in Dataset 2 are not well separated (Cases 4–6 of Table 2 are). This is because the damage extent in Cases 1–3 is very small, and their dynamic responses are very similar in the physical space. Thus, the low-dimensional embedding shows the very close distance of these three cases. This is interesting as the distance in the low-dimensional space may be used as a metric to quantify the damage levels. A detailed investigation is beyond the current scope of the paper (automatic classification, not quantification) and is listed as future work.



(**a**)



(**b**)

Figure 11. Low-dimensional representation of all classes classification for both datasets: (**a**) Dataset 1: low-dimensional representation of all cases; (**b**) Dataset 2: low-dimensional representation of all cases.

4.3. Evaluation and Computational Cost

Since the proposed study is fully unsupervised learning, the traditional cross-validation method cannot be easily applied to evaluate the corresponding performance of the model [46]. Alternatively, the F1 score is commonly used to indicate the corresponding performance [47]. The F1-score includes both precision and the recall of a classifier into a single metric, which is defined by the following equation.

$$F1 = \frac{P \times R}{P + R} \times 100\% \tag{6}$$

where P is the precision, and R is the recall of the classification model. In this study, the precision P is calculated by Equation 7.

$$P = \frac{\#TruePositive}{\#TruePositive + \#FalsePositive} \tag{7}$$

The definition of recall R is in Equation 8.

$$R = \frac{\#TruePositive}{\#TruePositive + \#FalseNegative} \tag{8}$$

Similar to most state-of-the-art methods on unsupervised manifold learning [21,48], we used the k-nearest neighbor (KNN) classifier trained on the low dimensional representation to evaluate the corresponding precision and recall of the obtained embeddings.

For both demonstrations, the 10 runs averaged F1 scores and the corresponding averaged computational time are shown in Table 3, while the highest value of traditional neural network models can only reach 98.93% for limited cases with longer duration of 300 s [37].

Table 3. F1 Score (%) and computational time by a single core Intel(R) Xeon(R) CPU @ 2.20 GHz.

Dataset	F1-Score (%)	Computational Time
1	100.00%	32.61 s
2	94.21%	21.55 s

5. Discussion

In this section, the effect of signal length (time step) of the simplicial simplex and noise level are discussed to show the robustness of the proposed method.

5.1. Signal Length of the Simplicial Simplex

The signal length in the time domain is very important for spatiotemporal data analysis problems. Some techniques require as many as 30,000 steps [37] or a minimum of 1500 steps [15] to capture enough features for SHM. Thus, an algorithm that is capable of classification with only a short duration is preferred. This significantly enhances the robustness of the application as the classification results will be insensitive to the selection of time window and transit change in service.

A parametric analysis using Dataset 1 was performed. The proposed method was applied with different signal lengths (i.e., from 100–1900 time steps) for each category. The embedding results are shown in Figure 12. It is very clear that the proposed technique can separate all cases when we use 100 steps (or 5 s) for each case in the temporal domain. Some cases were divided into multiple sub-classes, and the reason is that local features instead of global features play a dominant role in a shorter duration. Quantitative metrics using F1 score for different signal lengths are shown in Figure 12. It is shown that the proposed method shows stabilized results once the signal length is beyond a certain length (see Figure 13).

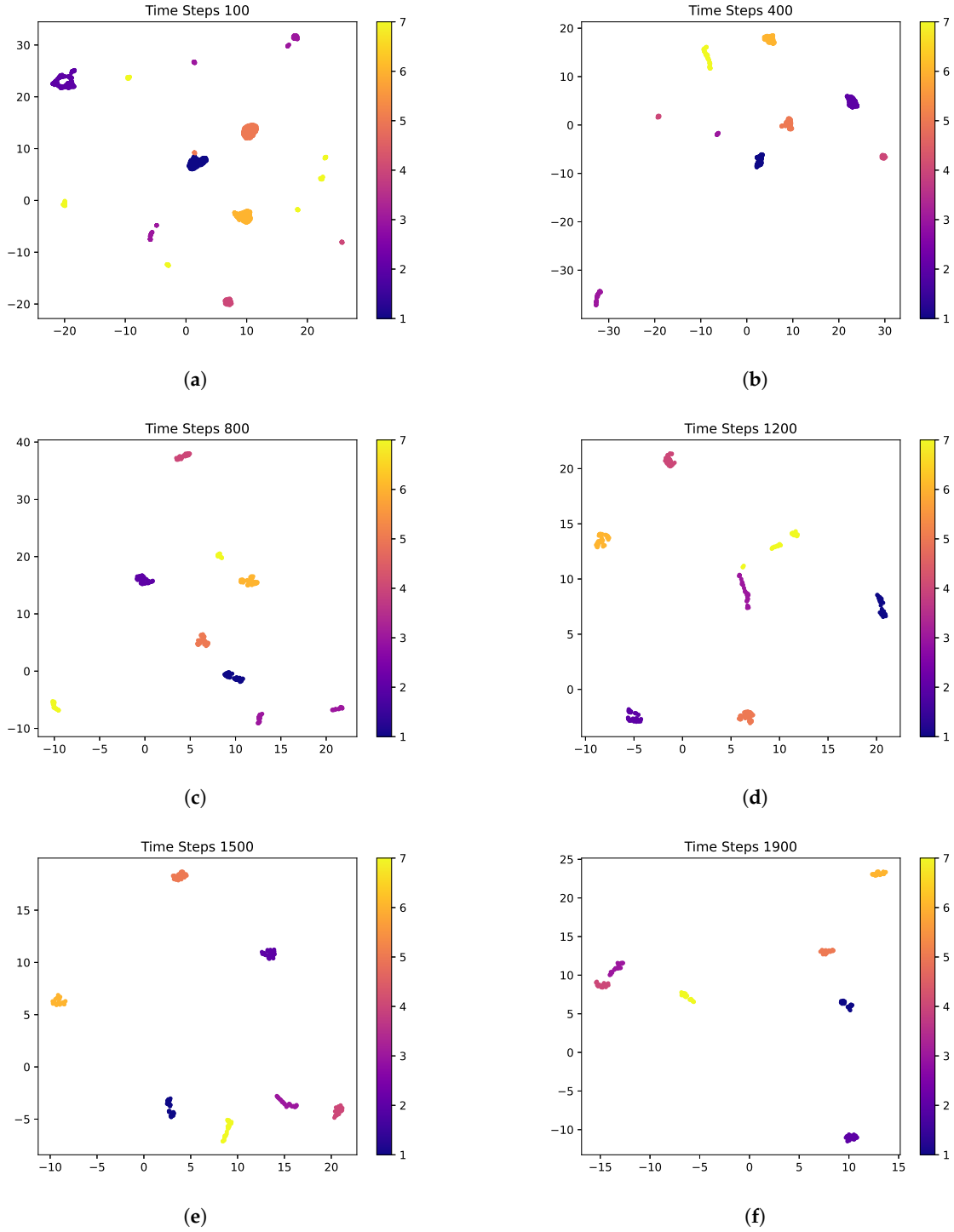
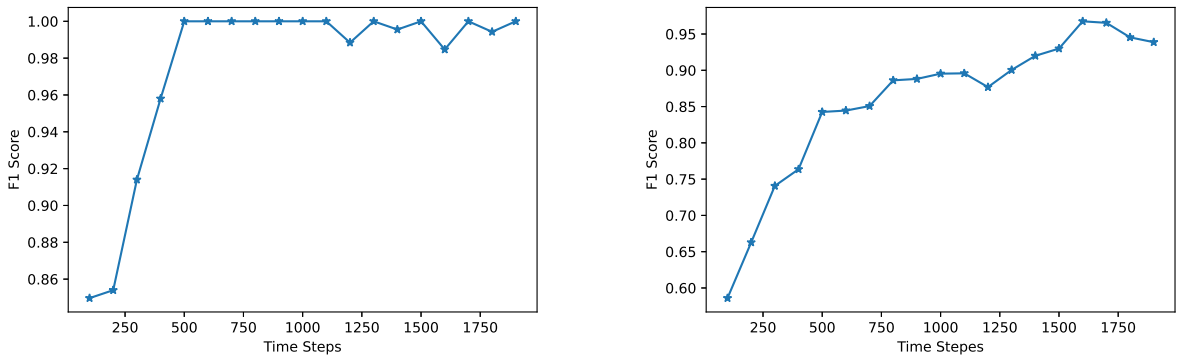


Figure 12. Examples of embeddings in different time lengths ((**a**) 100; (**b**) 400; (**c**) 800; (**d**) 1200; (**e**) 1500; (**f**) 1900) used for a single data point (Dataset 1).

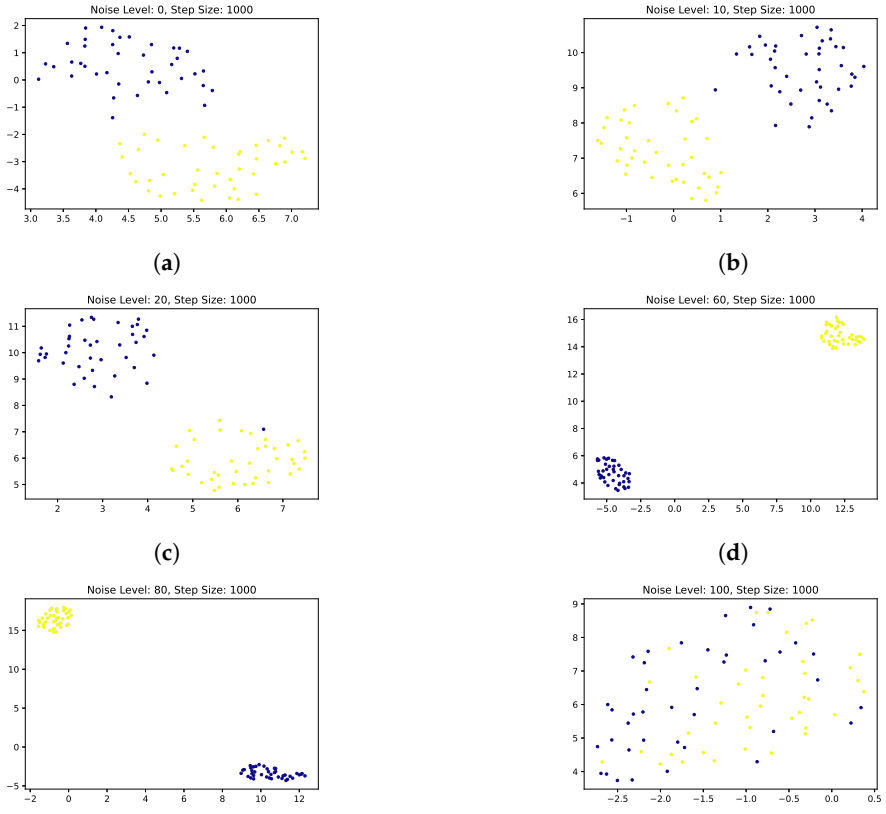


(**a**) (**b**)

Figure 13. Five-runs accuracy performance (F1 score) vs. time steps used for a single point: (**a**) Dataset 1; (**b**) Dataset 2.

5.2. Effect of Noise Levels

In this paper, we checked how accuracy changes at noise levels from 0 to 100, as defined in Step 4. We have some embeddings shown in Figure 14, and the relation of accuracy changes at different noise levels is shown in Figure 15. The results indicate the robustness of the algorithm in the presence of noise. The classification accuracy did not show a significant decrease until the 90 noise level was reached.



(**a**) (**b**)

(**c**) (**d**)

(**e**) (**f**)

Figure 14. Examples of embeddings in different noise levels ((**a**) 0; (**b**) 10; (**c**) 20; (**d**) 600; (**e**) 80; (**f**) 100); 1000 time steps used for single data point (Dataset 1).

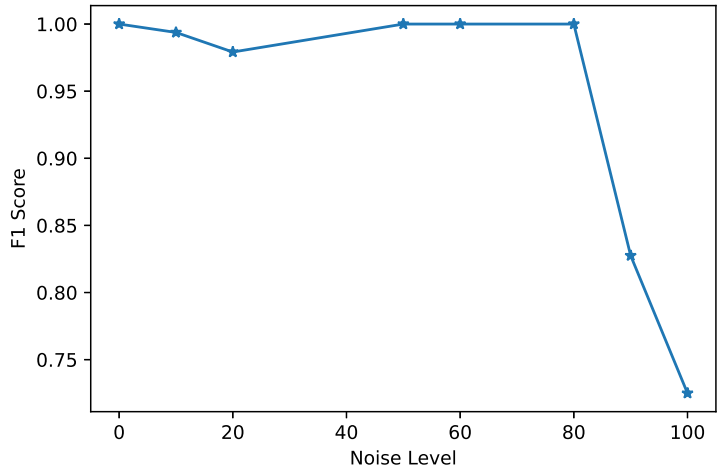


Figure 15. Five-runs averaged accuracy changed with different noise levels (Start).

6. Conclusions

This paper introduces a novel technique called SCEME for spatiotemporal structural health monitoring data. It leverages the topological invariant, Euler characteristic (EC), to describe the behaviors of spatiotemporal systems in the temporal domain and utilizes manifold learning to obtain the low-dimensional representation that obtains the spatial and inter-series behaviors. The proposed method can effectively capture both inter-series and intra-series correlation, particularly for complex super high-dimensional data. As a result, it can classify different damage levels with high accuracy and shorter run time, while being stable with noise. The proposed technique holds the potential to greatly improve the efficiency of structural damage detection using measured structural responses, which are typically spatiotemporal data with large dimensions and uncertainty levels. Currently, the focus is on damage classification, and damage quantification in the low-dimensional space is a future direction of study. Additionally, while this study focuses on diagnosis, future works will explore prognosis in the embedding space and its mapping back to the physical space. It is crucial to keep in mind that the method proposed in this study assumes that the structure will experience similar or equivalent excitations during the diagnostic process. If the loading type changes significantly, the detection results may not be accurate. Thus, future works will also involve testing more complex structures to further validate the robustness of the proposed technique.

Author Contributions: N.X.: Conceptualization, methodology, data curation, writing—original draft, visualization, investigation; Z.Z.: visualization, writing—review and editing, investigation. Y.L.: conceptualization, supervision, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: The research reported in this paper is partially supported by funds from NASA University Leadership Initiative program (Contract No. NNX17AJ86A, Project Officer: Dr. Anupa Bajwa, Principal Investigator: Dr. Yongming Liu) and by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Solar Energy Technology Office (SETO) Award Number DE-EE0009354 (Program Manager: Dr. Tassos Golnas). The support is gratefully acknowledged.

Data Availability Statement: https://github.com/ymlasu/Simplicial-Complex-Enhanced-Manifold-Embedding-of-Spatiotemporal-Data-for-Structure-Health-Monitorin (accessed on: 15 March 2023).

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

References

1. Sony, S.; Laventure, S.; Sadhu, A. A literature review rerof next-generation smart sensing technology in structural health monitoring. *Struct. Control Health Monit.* **2019**, *26*, e2321. [CrossRef]
2. González, M.P.; Zapico, J.L. Seismic damage identification in buildings using neural networks and modal data. *Comput. Struct.* **2008**, *86*, 416–426. [CrossRef]
3. De Lautour, O.R.; Omenzetter, P. Prediction of seismic-induced structural damage using artificial neural networks. *Eng. Struct.* **2009**, *31*, 600–606. [CrossRef]
4. Alves, P.; Couto, E.B.; Francisco, P.M. Board of directors' composition and capital structure. *Res. Int. Bus. Financ.* **2015**, *35*, 1–32. [CrossRef]
5. Bornn, L.; Farrar, C.R.; Park, G.; Farinholt, K. Structural health monitoring with autoregressive support vector machines. *J. Vib. Acoust.* **2009**, *131*, 2. [CrossRef]
6. Chong, D. *Collective Action and the Civil Rights Movement*; University of Chicago Press: Chicago, IL, USA , 2014.
7. Kim, S.M.; Hsu, A.; Lee, Y.H.; Dresselhaus, M.; Palacios, T.; Kim, K.K.; Kong, J. The effect of copper pre-cleaning on graphene synthesis. *Nanotechnology* **2013**, *24*, 365602. [CrossRef]
8. Liang, X.; Mosalam, K.; Muin, S. Simulation-based data-driven damage detection for highway bridge systems. In Proceedings of the 11th National Conference on Earthquake Engineering (NCEE), Los Angeles, CA, USA, 25–29 June 2018.
9. Sajedi, S.O.; Liang, X. Vibration-based semantic damage segmentation for large-scale structural health monitoring. *Comput. Aided Civ. Infrastruct. Eng.* **2020**, *35*, 579–596. [CrossRef]
10. Abdeljaber, O.; Avci, O.; Kiranyaz, S.; Gabbouj, M.; Inman, D.J. Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *J. Sound Vib.* **2017**, *388*, 154–170. [CrossRef]
11. Liang, X.; Sajedi, S.O. Deep Bayesian U-Nets for Efficient, Robust and Reliable Post-Disaster Damage Localization. *arXiv* **2020**, arXiv:2009.11460.
12. Lin, C.; Griffith, C.; Zhu, K.; Mathur, V. *Understanding Vulnerability of Children in Surrey*; The University of British Columbia: Vancouver, BC, Canada, 2018.
13. Sajedi, S.O.; Liang, X. Uncertainty-assisted deep vision structural health monitoring. *Comput. Aided Civ. Infrastruct. Eng.* **2021**, *36*, 126–142. [CrossRef]
14. Farrar, C.R.; Worden, K. *Structural Health Monitoring: A Machine Learning Perspective*; John Wiley & Sons: New York, NY, USA, 2012.
15. Xu, N.; Liu, Y. Fractal-based manifold learning for structure health monitoring. In Proceedings of the AIAA Scitech 2021 Forum, Online, 19–21 January 2021; p. 1167.
16. Cohen-Steiner, D.; Edelsbrunner, H.; Harer, J. Stability of persistence diagrams. In Proceedings of the Twenty-First Annual Symposium on Computational Geometry, Pisa, Italy, 6–8 June 2005; pp. 263–271.
17. Smith, A.; Zavala, V.M. The Euler characteristic: A general topological descriptor for complex data. *Comput. Chem. Eng.* **2021**, *154*, 107463. [CrossRef]
18. Zhang, Z.; Xu, N.; Liu, Y. Robust physics discovery via supervised and unsupervised pattern recognition using the Euler characteristic. *Comput. Methods Appl. Mech. Eng.* **2022**, *397*, 115110. [CrossRef]
19. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [CrossRef]
20. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 11.
21. McInnes, L.; Healy, J.; Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv* **2018**, arXiv:1802.03426.
22. Xu, N.; Liu, Y. CAMEL: Curvature-Augmented Manifold Embedding and Learning. *arXiv* **2023**, arXiv:2303.02561.
23. Stahl, S.; Stenson, C. *Introduction to Topology and Geometry*; John Wiley & Sons: New York, NY, USA, 2014.
24. Mandelbrot, B.B. Self-affine fractals and fractal dimension. *Phys. Scr.* **1985**, *32*, 257. [CrossRef]
25. Seifert, H.; Threlfall, W. *A Textbook of Topology*; Academic Press: Cambridge, MA, USA, 1980.
26. Atkin, R.H. *Combinatorial Connectivities in Social Systems: An Application of Simplicial Complex Structures to the Study of Large Organizations*; Springer: Cham, Switzerland, 1977.
27. Kühnel, W. *Differential Geometry*; American Mathematical Society: Providence, RI, USA, 2015; Volume 77.
28. Matoušek, J. *Topological Methods in Combinatorics and Geometry*; Lecture Notes; Department of Applied Mathematics: Prague, Czech Republic, 1994.
29. Kelleher, A.P.C.; Pantano, A. Introduction to Simplicial Complexes. (cit. on p. 40). Available online: https://www.math.uci.edu/~mathcircle/materials/MCsimplex.pdf (accessed on 6 January 2023).
30. Colding, T.H. Ricci curvature and volume convergence. *Ann. Math.* **1997**, *145*, 477–501. [CrossRef]
31. Bauer, F.; Hua, B.; Jost, J.; Liu, S.; Wang, G. The geometric meaning of curvature: Local and nonlocal aspects of ricci curvature. In *Modern Approaches to Discrete Curvature*; Springer: Cham, Switzerland, 2017; pp. 1–62.
32. Samal, A.; Sreejith, R.; Gu, J.; Liu, S.; Saucan, E.; Jost, J. Comparative analysis of two discretizations of Ricci curvature for complex networks. *Sci. Rep.* **2018**, *8*, 1–16.
33. Hamilton, R.S. Three-manifolds with positive Ricci curvature. *J. Differ. Geom.* **1982**, *17*, 255–306. [CrossRef]
34. Ni, C.C.; Lin, Y.Y.; Luo, F.; Gao, J. Community detection on networks with ricci flow. *Sci. Rep.* **2019**, *9*, 1–12.
35. Perelman, G. The entropy formula for the Ricci flow and its geometric applications. *arXiv* **2002**, arXiv:math/0211159.

36. Dyke, S.J.; Bernal, D.; Beck, J.; Ventura, C. Experimental phase II of the structural health monitoring benchmark problem. In Proceedings of the 16th ASCE Engineering Mechanics Conference, Seattle, WA, USA, 16–18 July 2003.
37. Abdeljaber, O.; Avci, O.; Kiranyaz, M.S.; Boashash, B.; Sodano, H.; Inman, D.J. 1-D CNNs for structural damage detection: Verification on a structural health monitoring benchmark data. *Neurocomputing* **2018**, *275*, 1308–1317. [CrossRef]
38. Omenzetter, P.; de Lautour, O.R. Classification of damage in structural systems using time series analysis and supervised and unsupervised pattern recognition techniques. In Proceedings of the Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems, San Diego, CA, USA, 8–11 March 2010; SPIE: New York, NY, USA, 2010; Volume 7647, pp. 1424–1435.
39. Figueiredo, E.; Flynn, E. Three-story building structure to detect nonlinear effects. In *Report SHMTools Data Description*; Triad National Security, LLC: Columbus, OH, USA, 2009.
40. Johnson, E.A.; Lam, H.F.; Katafygiotis, L.S.; Beck, J.L. Phase I IASC-ASCE structural health monitoring benchmark problem using simulated data. *J. Eng. Mech.* **2004**, *130*, 3–15. [CrossRef]
41. Zhang, P.; Shen, H.; Zhai, H. Machine learning topological invariants with neural networks. *Phys. Rev. Lett.* **2018**, *120*, 066401. [CrossRef] [PubMed]
42. Adler, R.J. Some new random field tools for spatial analysis. *Stoch. Environ. Res. Risk Assess.* **2008**, *22*, 809–822. [CrossRef]
43. Ziou, D.; Allili, M. Generating cubical complexes from image data and computation of the Euler number. *Pattern Recognit.* **2002**, *35*, 2833–2839. [CrossRef]
44. Poincaré, H. *Analysis Situs*; Gauthier-Villars: Paris, France, 1895.
45. Smith, A.D.; Dłotko, P.; Zavala, V.M. Topological data analysis: Concepts, computation, and applications in chemical engineering. *Comput. Chem. Eng.* **2021**, *146*, 107202. [CrossRef]
46. Conneau, A.; Khandelwal, K.; Goyal, N.; Chaudhary, V.; Wenzek, G.; Guzmán, F.; Grave, E.; Ott, M.; Zettlemoyer, L.; Stoyanov, V. Unsupervised cross-lingual representation learning at scale. *arXiv* **2019**, arXiv:1911.02116.
47. DeVries, Z.; Hoda, M.; Rivers, C.S.; Maher, A.; Wai, E.; Moravek, D.; Stratton, A.; Kingwell, S.; Fallah, N.; Paquet, J.; et al. Development of an unsupervised machine learning algorithm for the prognostication of walking ability in spinal cord injury patients. *Spine J.* **2020**, *20*, 213–224. [CrossRef]
48. Tang, J.; Liu, J.; Zhang, M.; Mei, Q. Visualizing large-scale and high-dimensional data. In Proceedings of the 25th International Conference on World Wide Web, Montreal, QC, Canada, 11–15 April 2016; pp. 287–297.