energies

Article

A Novel Deep Clustering Method and Indicator for Time Series Soft Partitioning †

Alexandre Eid [1,2,*], Guy Clerc [1], Badr Mansouri [2] and Stella Roux [3]

[1] University Lyon, Université Claude Bernard Lyon 1, INSA Lyon, École Centrale de Lyon CNRS, Ampère, UMR5005, 69622 Villeurbanne, France; guy.clerc@univ-lyon1.fr
[2] Safran Electronics & Defense, 91344 Massy, France; badr.mansouri@safrangroup.com
[3] Grenoble INP—Ensimag, UGA, 38400 Saint-Martin-d'Hères, France; stella.roux@grenoble-inp.org
* Correspondence: alexandre.eid@univ-lyon1.fr

† In Proceedings of the 2021 IEEE International Conference on Prognostics and Health Management (ICPHM 2021), Detroit, MI, USA, 7–9 June 2021.

Abstract: The aerospace industry develops prognosis and health management algorithms to ensure better safety on board, particularly for in-flight controls where jamming is dreaded. For that, vibration signals are monitored to predict future defect occurrences. However, time series are not labeled according to severity level, and the user can only assess the system health from the data mining procedure. To that extent, a clustering algorithm using a deep neural network core is developed. Time series are encoded into pictures to be fed into an artificially trained neural network: U-NET. From the segmented output, one-dimensional information on cluster frontiers is extracted and filtered without any parameter selection. Then, a kernel density estimation finally transforms the signal into an empirical density. Ultimately, a Gaussian mixture model extracts the latter independent components. The method empowered us to reveal different degrees of severity faults in the studied data, with their respective likelihoods, without prior knowledge. It was then compared to state-of-the-art machine learning algorithms. However, internal clustering results evaluation for time series is an open question. As the state-of-the-art indexes were not producing relevant results, a new indicator was built to fulfill this task. We applied the whole method to an actuator consisting of an induction machine linked to a ball screw. This study lays the groundwork for future training of diagnosis and prognosis structures in the health management framework.

Keywords: semantic segmentation; time series; clustering; deep learning; kernel density estimation; electromechanical actuator; data labeling; prognosis and health management; aeronautics

check for updates

Citation: Eid, A.; Clerc, G.; Mansouri, B.; Roux, S. A Novel Deep Clustering Method and Indicator for Time Series Soft Partitioning. *Energies* **2021**, *14*, 5530. https://doi.org/10.3390/en14175530

Academic Editor: Ahmed Abu-Siada

Received: 22 July 2021
Accepted: 30 August 2021
Published: 4 September 2021

1. Introduction

According to the Maintenance Cost Technical Group (MCTG) from the International Air Transport Association (IATA), USD 69 billion was spent on maintenance repair and overhaul (MRO) in 2018 for a fleet count of 27,535 aircraft from 54 airlines. The aerospace industry is shifting from hydraulic to electromechanical actuation, which could be prone to seizure. Hence, it is mandatory to prevent actuator jamming at any cost during in-flight control, where several ball-screws are used.

The Prognosis and Health Management (PHM) framework can be used to reduce the amount of money spent on unscheduled MRO operations by developing methods to forecast structural defects. Further, it can better anticipate this hazardous jamming event by detecting any unusual behavior from either signal of vibration or electrical.

Nowadays, the PHM framework has moved from conventional-based health monitoring (HM) to a deep learning approach. Therefore, the previously hand-designed features extracted from raw data are automatically determined by the first layers of a deep neural network (DNN). Additionally, in the extensive review of deep learning by [1], visualizing time series behavior as pictures is an ongoing research topic. Time series dynamics can be extracted by the Gramiam Angular Field (GAF) representation of [2] or the Recurrence Plot (RP) of [3]. These representations can be fed into a DNN for better classification of data. In our work, a similar approach is used for visualizing time series behavior.

In [4], Jain defined *clustering* as "the unsupervised classification of patterns into groups". While this exploratory analysis technique is widely used in spatially organized datasets, it is scarcely used in time series where data are organized by time. Hence, there is a need to create a new method with its quality indicator. Several recent studies have been presented, using clustering or pattern recognition to label data for further use in the training of an artificial diagnosis structure. In [5], every deviation from the nominal state of electrical machines is detected from a clustering algorithm. Plus, the operating conditions can be inferred from the data group structure [6]. The association between a group and a fault severity is clearly stated in [7]. Patterns found in this approach can be assumed to be representative of several fault detection. This same hypothesis is followed here: a group of data represents a certain fault severity. For the fuzzy clustering algorithm, Ref. [8] presents another approach.

Further, finding a general indicator is challenging, as the clustering quality indexes are specific to the fundamental hypothesis made in the research [9]. This article presents a new method and a new quality indicator for clustering and quantifying the results for time series data. First, the method will be outlined, and then it will be applied to time series. The time series are vibration signals coming from an induction machine attached to a ball screw. From it, many precursors of jamming defects are extracted. Finally, our method is tested against several state-of-the-art algorithms in the discussion part, and all the results are quantified with a newly created indicator.

2. Materials and Methods

The clustering method is applied to time series in a broad sense, i.e., indexed signals by a timestamp. More generally, signals whose elements abide by a common precedence relation are considered. Let $\mathcal{T}$ be such a signal. Practically, it represents the values of a given sensor indexed by time, on a monitored system. n realizations of $\mathcal{T}$ for the system can be gathered: $\{\mathcal{T}_1, \mathcal{T}_2, ..., \mathcal{T}_n\}$. In order to assess system health, an ensemble of p statistical and model-based parameters are extracted from each $\mathcal{T}_{i\in[\![1;n]\!]}$. Let $\mathbb{X}_{j\in[\![1;p]\!]} \in \mathcal{M}_{n1}(\mathbb{R})$ be such parameter. Each one of them can be concatenated to form what is called a *feature matrix* $\mathbb{M} \in \mathcal{M}_{np}(\mathbb{R})$.

In the following work, studied parameters are the chosen columns of the *feature matrix* $\mathbb{M}$. Let us choose for $j \in [\![1;p]\!], \mathbb{X}_j = \begin{bmatrix} x_1^j & x_2^j & ... & x_n^j \end{bmatrix}^\top$ the jth descriptor.

A clustering method deals with data classification without prior knowledge. To do so, it has to minimize an intra-class distance and maximize an outer-class one. Doing so, it aggregates objects with similar properties and repels objects with heterogeneous ones. A distance is used to compare its elements $x_{i\in[\![1;n]\!]}$ between each other, allowing the extraction of relevant clustering information from each $\mathbb{X}_j$. It allows the extraction of groups with similar characteristics. For that, the Euclidean distance presented in Equation (1) is chosen.

$$\forall (x,y) \in \mathbb{R}^2, d(x,y) = \sqrt{x^2 - y^2} \tag{1}$$

By applying the distance to all the elements of $\mathbb{X}_i$, a square *distance matrix* $\in \mathcal{M}_n(\mathbb{R})$ can be obtained in Equation (2).

$$\mathcal{G}(\mathbb{X}_j) = \begin{bmatrix} d(x_1^j, x_1^j) & d(x_1^j, x_2^j) & \dots & d(x_1^j, x_n^j) \\ d(x_2^j, x_1^j) & d(x_2^j, x_2^j) & \dots & d(x_2^j, x_n^j) \\ \vdots & \vdots & \vdots & \vdots \\ d(x_n^j, x_1^j) & d(x_n^j, x_2^j) & \dots & d(x_n^j, x_n^j) \end{bmatrix} \tag{2}$$

The nearer the points are to each other, the lesser the value of the distance. Hence, there are areas in the matrix filled with near zeros values if the original data present some clustering behavior. The symmetry of the *distance matrix* enables groups of data to be visually recognizable, and the Euclidean distance creates square shapes of near null values along with the matrix diagonal. Let $\mathcal{G}_{j\in[\![1;p]\!]}$ be the matrix associated with each $\mathbb{X}_j$. Each of them will be processed as one channel RGB image. At this stage, one fundamental hypothesis is made:

Hypothesis 1. *The studied system cannot regenerate itself.*

Note that this hypothesis is commonly made in PHM, due in part to the nature of the data available to the community. The famous C-MAPSS dataset from [10] and its recent enhancement by [11] contain run-to-failure data from an aircraft engine, without maintenance operations between several flight cycles. Furthermore, this hypothesis was already clearly made in [12] (Hypothesis 12). Otherwise, one could face two configurations. First, if the maintenance completely rejuvenates the system, the method still applies after moving the origin of the time series before processing. However, if it is not the case, the residual part of useful life gained after maintenance must be evaluated. The latter is out of the scope of this work. Thus, with Hypothesis 1, every cluster found can be associated with a fault severity degree. As the clustering is done on the time series, each new cluster increases the severity value of the underlying default. The distance matrices defined in Equation (2) are the stepping stone for extracting relevant clustering information. As shown in Section 3, the processed data in our application are very noisy, and the tested traditional signal processing techniques were not entirely effective in improving the signal over noise ratio. Hence, the deep learning approach is chosen and applied with an image processing technique. The matrices are considered to be images whose contrast will be improved.

In this study, the computed Euclidean distances in the symmetric matrices make square shapes appear on their diagonals. Detecting shapes in images can be done in two steps:

- First, select a value for each class, and assign the latter values to every pixel in the image. This process is called semantic segmentation. For example, the value 0 is assigned to the wanted square shapes and 1 to the background.
- Secondly, build a structure to reveal the different square shapes and eliminate the noisy background from images automatically, thus conserving only relevant information for clustering. To that extent, deep neural structures are studied.

One drawback of such architectures is the need for a tremendous amount of training data. In industry, and specifically in aeronautics, relevant training data are scarce and costly to obtain. Consequently, Ref. [13] has developed an architecture, U-NET, for cell images segmentation specialized in small data samples to alleviate this limitation. Note that the dataset dimension should be compared with usual ensembles used for training state-of-the-art neural network architectures. To give an order of magnitude, one could cite Microsoft's COCO dataset [14] and its 3×10^6 images. The network U-NET is comprised of two main parts, as seen in Figure 1. The left part of the configuration compresses input and extracts deep features from it. The most profound features of all are processed through a bottleneck set of layers. Its result is fed through the bottom of the right part to be decompressed. Finally, depending on the number of classes to be segmented, either a *sigmoid* or a *softmax* layer is used to transform each processed input pixel value into the likelihood of belonging to a class. Each block has two convolution layers for the descending part, associated with a *ReLU* activation ending with *MaxPooling*. For the ascending part, the symmetrical structure is used made of *ConvTranspose2d*, an inverse convolution operation to up-sample its input, as developed in Pytorch [15].

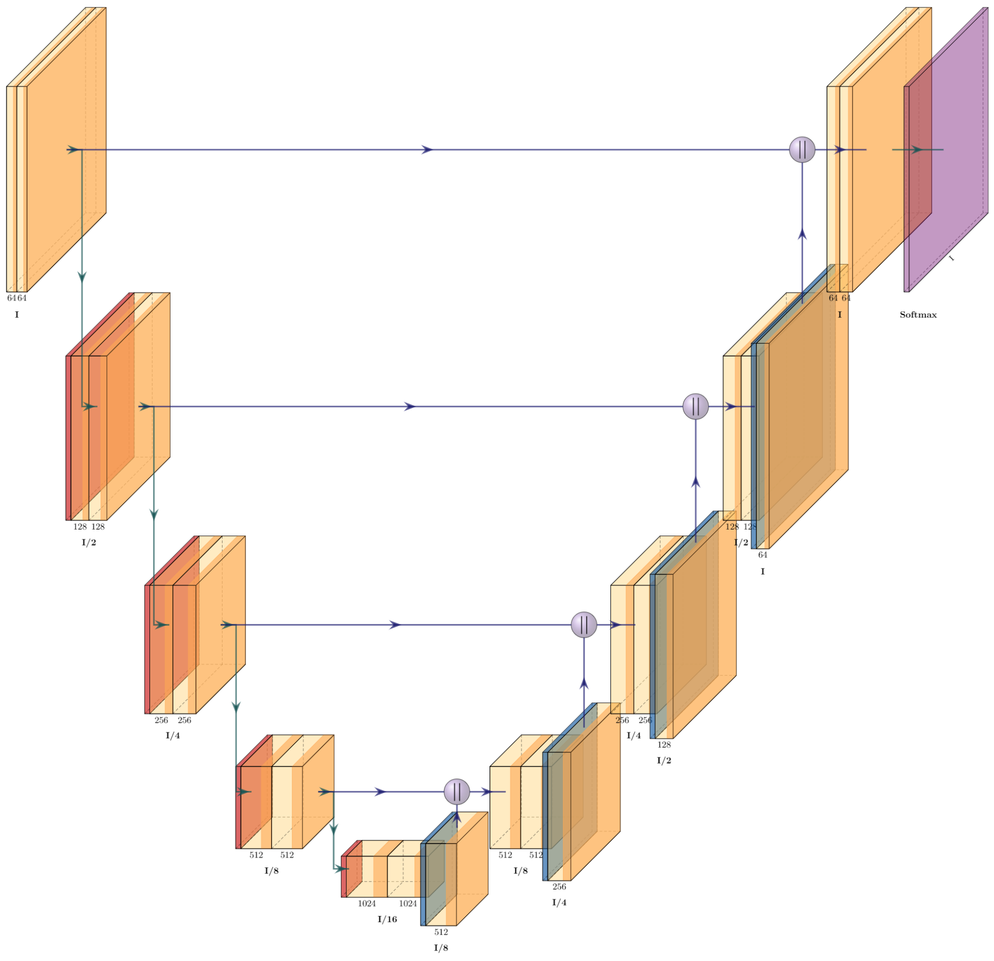

Figure 1. U-NET U-shaped architecture represented using algorithms from [16].

U-NET's goal is to segment every distance matrix $\mathcal{G}_{j\in[\![1;p]\!]}$. As stated earlier, the network has to learn to recognize square patterns alongside each matrix diagonal. Consequently, it has to be trained to do so.

Like many industrial environments, aeronautics lacks an appropriate amount of high sampled training data. An artificial dataset is generated to alleviate this problem. Four types of training patterns are produced, all sharing the same underlying structure: a set of squares with black ones along the diagonal.

The images from Figure 2 are randomly created by calculating the distance matrices of constant continuous piecewise functions (CCPF), linear continuous piecewise functions (LCPF), and adding Gaussian noise. The first row of Figure 2 represents input training data, whereas the second row represents its associated mask. At the end of each forward pass, the neural network compares its output to the mask and backpropagates the gradient accordingly. The best training procedure is empirically determined. First, U-NET is trained with clean square images (CCPF). Once the loss function is converged, the training of U-NET is reinforced by fuzzy data (LCPF). The operation is then repeated, and Gaussian noise is added to each input to further increase the neural network generalization capacity. Finally, a new dataset is generated with a field-specific bias to help the convergence of the training to the study goal. This bias comes from another fundamental hypothesis made in this work:

Hypothesis 2. *The aging process is a compounding effect and can only accelerate through time.*

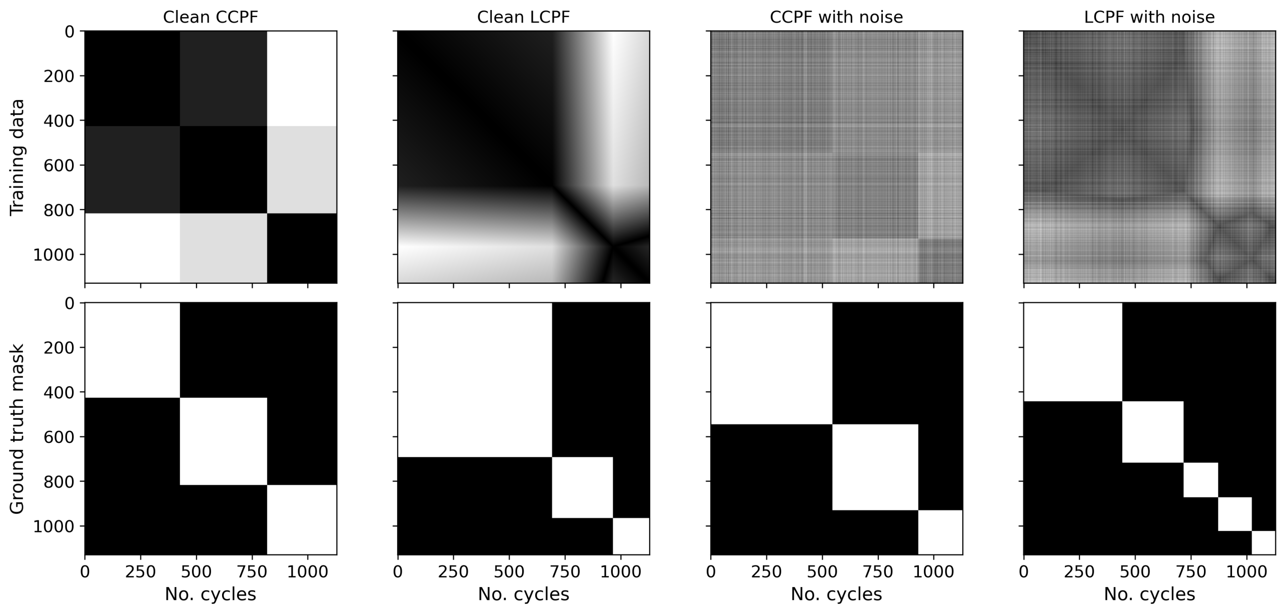

Figure 2. Distance matrices represented as images with several configurations, used for U-NET training with their associated ground truth.

Similarly to Hypothesis 1, Hypothesis 2 is commonly made in literature. In [11], several degradation models are applied to run-to-failure data. The abnormal one tends to represent the behavior of several known damage propagation models, e.g., Arrhenius or Coffin–Manson laws. Consequently, the last batch of training data is generated with one constraint: squares alongside the matrix diagonal must have decreasing dimensions. The latter part of the training reinforcement procedure is subsequently done.

Once U-NET is trained, it can produce segmented images from distance matrices. The segmented output is a binary image, as all of its pixels take either the value 0 or 1. However, as shown in Section 3, the neural network is still very noisy. Therefore, further signal processing procedures should be applied to extract a piece of relevant clustering information from it. The first attempt at data clustering can be directly extracted by looping through the output matrices anti-diagonally. It is a first candidate of a cluster frontier signal. The main idea is to recognize the area in each picture, where square shapes terminate. Figure 3 visually illustrates the algorithm used to extract a first frontier cluster candidate. The "anti-diagonal" term used previously refers to the direction of each red arrow.

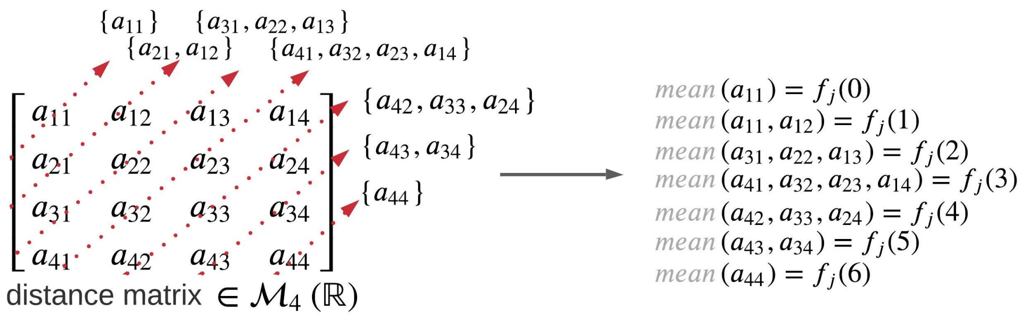
$\{a_{11}\}$
$\{a_{21}, a_{12}\}$
$\{a_{31}, a_{22}, a_{13}\}$
$\{a_{41}, a_{32}, a_{23}, a_{14}\}$
$\{a_{42}, a_{33}, a_{24}\}$
$\{a_{43}, a_{34}\}$
$\{a_{44}\}$
$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$
distance matrix $\in \mathcal{M}_4(\mathbb{R})$
$mean(a_{11}) = f_j(0)$
$mean(a_{11}, a_{12}) = f_j(1)$
$mean(a_{31}, a_{22}, a_{13}) = f_j(2)$
$mean(a_{41}, a_{32}, a_{23}, a_{14}) = f_j(3)$
$mean(a_{42}, a_{33}, a_{24}) = f_j(4)$
$mean(a_{43}, a_{34}) = f_j(5)$
$mean(a_{44}) = f_j(6)$

Figure 3. Visual explanation of cluster frontier extraction—phase 1.

As detailed in Algorithm 1, the matrix indices are fetched and stored into a structure. Note that looping through the matrix in such a way produces an object of size $2n - 1$ bags of items. Indeed, being of size n, the matrix upper left part is covered in n iterations, including the maximum anti-diagonal size. It then remains at $n - 1$ iterations to get through the lower right part of the matrix. The mean is calculated for each ensemble of matrix values as stated in line 16, thus creating a resulting signal of size $2n - 1$.

Algorithm 1 Cluster frontier extraction—phase 1

```
1: function GET INDICES OF MATRIX($n$)                          ▷ $n$ is matrix size
2:     declare indexArray of $[1; 2n-1]$ arrays          ▷ Initialize array of size $2n-1$
3:     for $i \leftarrow 1$ to $n$ do              ▷ get indices for upper left part of the matrix
4:         for $j \leftarrow 1$ to $i$ do
5:             indexArray$[i]$.append($[i-j+1; j]$)
6:     for $i \leftarrow n-1$ to 1 do                   ▷ get indices for other matrix part
7:         for $j \leftarrow i$ to 1 do
8:             indexArray$[2n-i]$.append($[j+1; i-j+2]$)
9: return indexArray
10:
11: function CALCULATE INTERMEDIATE SIGNAL($M$)                ▷ Matrix $M$ to process
12:     declare resArray of $[1; 2n-1]$ floats
13:     $n \leftarrow$ size(M)                                  ▷ M is a square matrix
14:     indexArray $\leftarrow$ GET INDICES OF MATRIX($n$)
15:     for $i \leftarrow 1$ to $2n-1$ do
16:         resArray$[i]$ $\leftarrow$ mean(M[indexArray$[i]$])
17: return resArray
```

The mean is chosen here because of the output pixel values. At the neural network output, pixels belonging to square shapes are encoded to 0, whereas pixels depicted in the image background are encoded to 1. By using the mean, a cluster frontier appears when its value is shifted toward unity. On the opposite, the mean tending toward a null value indicates the inside of a cluster.

The first phase of the cluster frontier extraction algorithm produces a signal with values lying between 0 and 1. The ones representing the potential cluster frontiers, local extrema, have to be found to detect and further refine the clustering boundaries. Each peak in the created function represents a likely square contour in the segmented image. However, standard filtering methods require the non-trivial setting of parameters, such as the sliding window. The latter represents the number of points to which the filtering algorithm is applied simultaneously. Hence, it controls the sensitivity of the algorithm to local and global dynamics. An iterative procedure is created to avoid such shortcomings. It consists of looping through the signal to be filtered with an increased window size for each loop. Figure 4 represents the algorithm visually.

The idea of the procedure is to split the input signal into several sets of values of increasing size throughout the loop. The minimum window is of size two, and it reaches a size n at the last iteration to cover the signal as a whole. Since cluster boundaries are detected for values approaching 1, the maximum is taken on each local window. An array of size $2n-1$ is initialized to 0 before the loop beginning to store each maximum found at its respective index in the input signal. Every time a maximum is locally found, it is added to the null array. Consequently, during initial iterations, the algorithm amplifies the local maxima, whereas in the end, it tends to amplify the more global ones. For further details on how to realize this operation, Algorithm 2 is given.

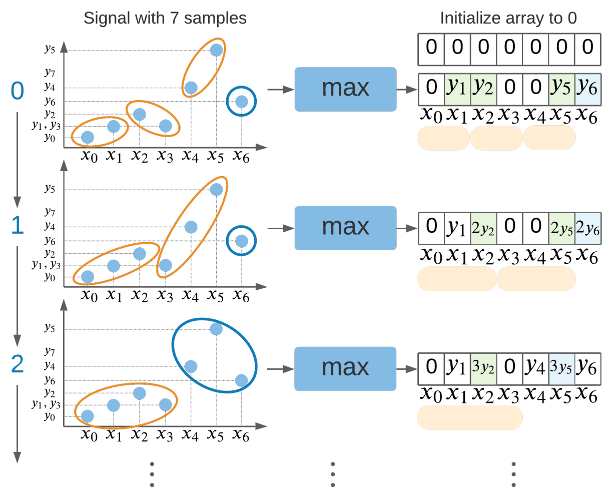
Signal with 7 samples
Initialize array to 0
0
1
2
max
max
max
y_5 y_7 y_4 y_6 y_2 y_1, y_3 y_0
x_0 x_1 x_2 x_3 x_4 x_5 x_6
0 0 0 0 0 0 0
0 y_1 y_2 0 0 y_5 y_6
0 y_1 $2y_2$ 0 0 $2y_5$ $2y_6$
0 y_1 $3y_2$ 0 y_4 $3y_5$ y_6

Figure 4. Visual explanation of cluster frontier extraction—phase 2: batch peak detection.

Algorithm 2 Cluster frontier extraction—phase 2: Batch Peak Detection

```
1:  function BATCH PEAK DETECTION(signalArray)
2:      $n \leftarrow$ size(signalArray)
3:      declare resArray of $n$ floats
4:      for $i \leftarrow 1$ to $n$ do          ▷ fill arrays with zeros and index for initialization
5:          resArray$[i] \leftarrow 0$
6:      declare noWindowList of $n$ integers
7:      declare remainingSizeList of $n$ integers
8:      for $i \leftarrow 2$ to $n$ do          ▷ create array with all the windows and remaining sizes
9:          noWindowList$[i] \leftarrow i$
10:         remainingSizeList$[i] \leftarrow n \,\%\, i$          ▷ % represents the modulus operator
11:     for $i \leftarrow 0$ to $n-1$ do
12:         noWindow $\leftarrow n \,//\, i$
13:         remainElmt $\leftarrow$ remainingSizeList$[i]$
14:         if remainElmt $\neq 0$ then          ▷ suppose that language allows array broadcasting
15:             tmpArray $\leftarrow$ signalArray$[: -$remainElmt$]$.reshape(noWindow, $i$)
16:         else
17:             tmpArray $\leftarrow$ signalArray.reshape(noWindow, $i$)
18:         declare indexStatValueList of unknown size of objects
19:         declare indexMaxArray of unknown size of floats
20:         for $j \leftarrow 0$ to noWindow do          ▷ get an array
21:             indexMaxArray $\leftarrow$ arg(tmpArray$[i,:]$ == max(tmpArray$[i,:]$))
22:             for $k \leftarrow 0$ to $size$(indexMaxArray) do
23:                 indexStatValueList.append($[j$, indexMaxArray$[k]$)          ▷ track maxima and
    their corresponding indexes (like chained list)
24:         declare idxMaskArray of size $m \times l$ integers
25:         for $j \leftarrow 1$ to $m \times l$ do          ▷ local initialization
26:             idxMaskArray$[j] \leftarrow 0$
27:         idxMaskArray.reshape($m, l$)
28:         for $k, p \leftarrow$ [1;noWindow] $\times$ $[1; i]$ do          ▷ create mask
29:             idxMaskArray$[k, p] \leftarrow 1$
30:         apply idxMaskArray to tmpArray
31:         unfold tmpArray of size (noWindow, $i$) to $(1, n)$
32:         for $j \leftarrow [1; m]$ do          ▷ Place maximum value at its corresponding index
33:             resArray$+=$ tmpArray
        return resArray
```

Finally, since the resulting array is still of size $2n - 1$, the abscissa scale of the signal is divided by a factor of two. This compensates the dilation generated by the method explained in Algorithm 1. It is assumed that the uncertainty of two cycles is tolerated for the health monitoring framework. The resulting signals are noted as $f_{j \in [\![1;p]\!]} \in \mathcal{M}_{n1}(\mathbb{R})$. Note that the combination of Algorithms 1 and 2 does not require the choice of any threshold or hyperparameter value.

At this stage of the method, the p time series are transformed into p distance matrices $\mathcal{G}_j$, and the neural network has segmented each matrix. From each output picture, p first candidates to cluster boundaries signal are obtained, and every one of them is filtered through the batch peak detection Algorithm 2. Hence, p new time series $\mathcal{F}_j$ are created. These could be concatenated into a matrix $\mathcal{F}$ as represented in Equation (3).

$$\mathcal{F} = \begin{bmatrix} f_{11} & f_{12} & f_{13} & \cdots & f_{1p} \\ f_{21} & f_{22} & f_{23} & \cdots & f_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f_{n1} & f_{n2} & f_{n3} & \cdots & f_{np} \end{bmatrix} \in \mathcal{M}_{np}(\mathbb{R}) \tag{3}$$

The developed method assumes that each descriptor contributes to the clustering result. Therefore, this information has to be fused to obtain a global final frontier signal.

The principle of kernel density estimation (KDE) [17] is used to do so. By considering each column of $\mathcal{F}$ to realize a random process, the KDE can be applied to the entire matrix. Consequently, the initial p descriptors are themselves considered random variables. Once these hypotheses are made, the distribution of each $f_{.j}$ provides insight into their respective theoretical probability density. With assumptions made about kernel functions, as presented by [18] (Theorem 6.7), the KDE constructs an empirical density that converges to its theoretical result. By considering the p temporal segments as realizations of several independent and identically distributed random processes, it is coherent to sum all columns of $\mathcal{F}$ defined in Equation (3) to obtain a vector as follows:

$$(y_i)_{1 \le i \le n} = \sum_{j=1}^{p} f_{ij} \tag{4}$$

From [18] (Chapter 6, Equation (6.1)), the empirical density can be expressed in Equation (5).

$$\hat{p}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - y_i}{h}\right) \tag{5}$$

K is a kernel function in Equation (5), applied to the centered variable of the sample time series and scaled by a factor h. Even if the Gaussian kernel defined in Equation (6) is not a computationally optimized one, it is chosen for its theoretical properties.

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp^{-\frac{1}{2}x^2} \tag{6}$$

The best property in this context is the ability to chose a simple bandwidth parameter from the state of the art. As detailed in [19] (Equation (3.31)), the bandwidth h of a Gaussian kernel can be expressed as $h = 0.9An^{-1/5}$, with the constant $A = \min(\sigma, \text{inter-quartile range}/1.34)$.

Finally, gathering all the steps in this Section 2, a function estimating the empirical likelihood of cluster frontiers is created for the studied dataset.

3. Results

In this second part, the clustering method developed in Section 2 is applied to monitoring vibration signals generated by an induction motor driving a ball screw. After presenting the dataset used, several signals are selected to test the previous theoretical part. Finally, our method is compared to other state-of-the-art clustering methods.

3.1. Dataset and Studied System

The dataset comes from four different electromechanical actuators, each made of an asynchronous electrical machine and a ball screw. A controlled test bench is constructed to monitor the vibration behavior of each actuator. The goal of the procedure is to detect any precursor of a jamming defect in each ball screw of the different actuators. That is why an accelerometer is placed on every monitored structure, as shown in Figure 5. This configuration allows the collection of vibration signals coming from the longitudinal axis of the actuation device.

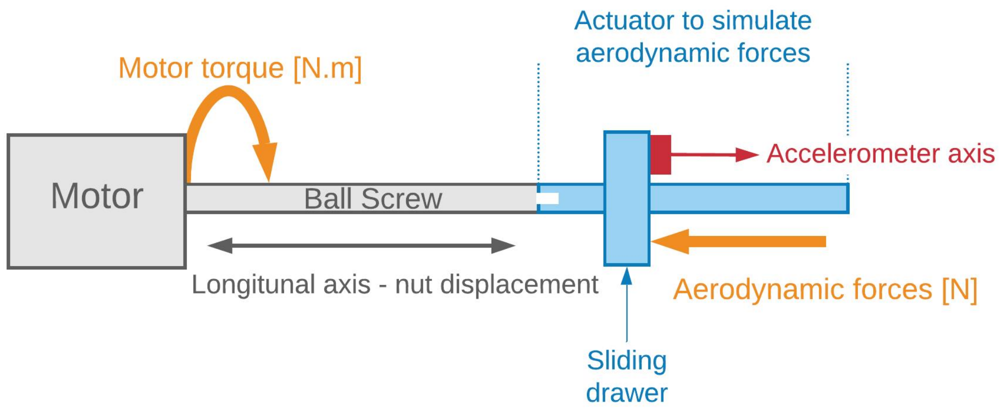

Figure 5. Schematics of the instrumented actuator bench.

Every actuator is stressed through 1130 cycles, specially designed to emulate a realistic environment. With this protocol, each one reaches the end of life at the last cycle. The signal measured is sampled at 1 kHz, standardized, and normalized between 0 and 1.

3.2. Time Series Encoding

As stated in Section 2, p statistical and model-based descriptors are computed from the raw data measured on the monitored actuator. Every descriptor $\mathbb{X}_j$ has 1130 samples. Among all p features, four are chosen. They are referred to as $\{\mathbb{X}_1, \mathbb{X}_2, \mathbb{X}_3, \mathbb{X}_4\}$. Each is standardized and normalized between 0 and 1. They are represented in Figure 6. The four descriptors are selected to evaluate the method on signals that exhibit various dynamics but are still coherent for extracting a health monitoring information extraction. Obviously, from Figure 6, the features are rather noisy and not strictly monotonic. However, a general ascending or descending trend is recognizable.

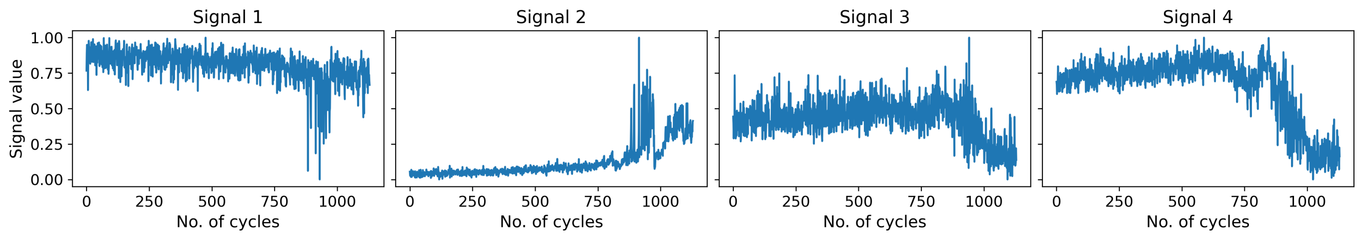

Figure 6. $\{\mathbb{X}_1, \mathbb{X}_2, \mathbb{X}_3, \mathbb{X}_4\}$ collected on our actuator.

Using the Euclidean distance, four distance matrices $\{\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \mathcal{G}_4\}$ are calculated from the previous four statistical features. They are presented in Figure 7.

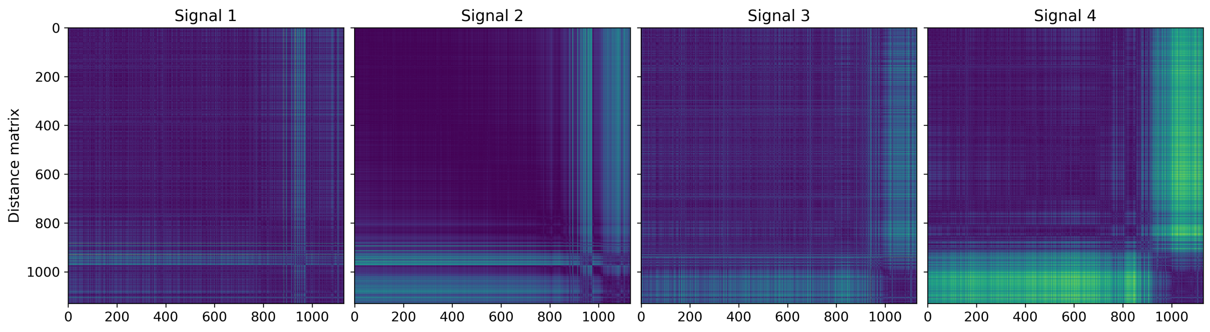

Figure 7. Distance matrices $\{\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \mathcal{G}_4\}$ extracted from $\{\mathbb{X}_1, \mathbb{X}_2, \mathbb{X}_3, \mathbb{X}_4\}$.

In Figure 7, each pixel corresponds to one value of the distance matrix. Note that pictures are colorized to better highlight symmetries, but they are in grayscale with original values between 0 and 1. Here, the darker the pixel, the lower the represented distance. Therefore, as stated in Section 2, potential clusters are represented as dark symmetrical shapes along with each picture diagonal. Indeed, those are the areas where the distance is minimal between points. Outside of it, the distance increases; hence a potential cluster limit can be found. Moreover, horizontal and vertical stripes in the images come from the noise of input data. To further illustrate the concept, at least two probable clusters in signal four of Figure 7 can be seen.

3.3. Semantic Segmentation

U-NET is trained with an artificial dataset containing 3000 images. A glimpse of its contents can be seen in Figure 2. The data are randomly split for the learning phase and shuffled into 2250 training pictures and 750 testing ones. As presented in Section 2, four types of artificial signals are equally present in the data set: signals with successive constant stages, signals with successive linear stages, and an adaptation of those signals with a white noise of ten percent. A study is carried out to determine the best learning hyper-parameters. For the learning rate, the adaptive scheduler *ReduceLROnPlateau* from package *optim.lr_scheduler* [15] with an initial value of 1×10^{-4} is selected. After monitoring the learning phase of U-NET with *tensorboard* [20], 200 epochs are sufficient to reach a stable accuracy. Because of a limited GPU memory bandwidth, a batch size, and a hyperparameter of 4 images for training is the maximum allowed. The input is an image of size 1130×1130; consequently, it rapidly saturates the VRAM during training. Finally the loss function used is *BCEWithLogitsLoss* from the *torch.nn* package [15]. Note that the learning phase is done only once.

The results of Figure 8 are obtained after 38 h of training on HPC resources with a bi-Intel Xeon Silver 4215R, a bi-NVIDIA Quadro RTX 6000 with 24 GB of GDDR6 each and 192 GB of RAM. U-NET outputs binary images. Purple corresponds to null values and yellow corresponds to ones.

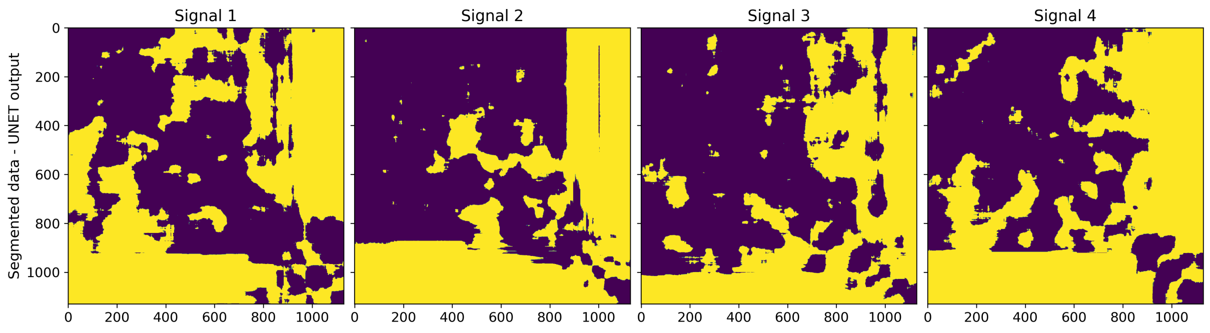

Figure 8. Segmented $\{\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \mathcal{G}_4\}$ after U-NET inference phase.

From Figure 8, it is clear that U-NET's training phase can be improved. Despite the noisy results, the latter signal processing algorithms are designed to alleviate this problem.

3.4. Cluster Frontier Extraction—First Phase

By applying Algorithm 2 of Section 2 to each $\mathcal{G}_i$ of $\{\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \mathcal{G}_4\}$, the signals in Figure 9 are extracted.

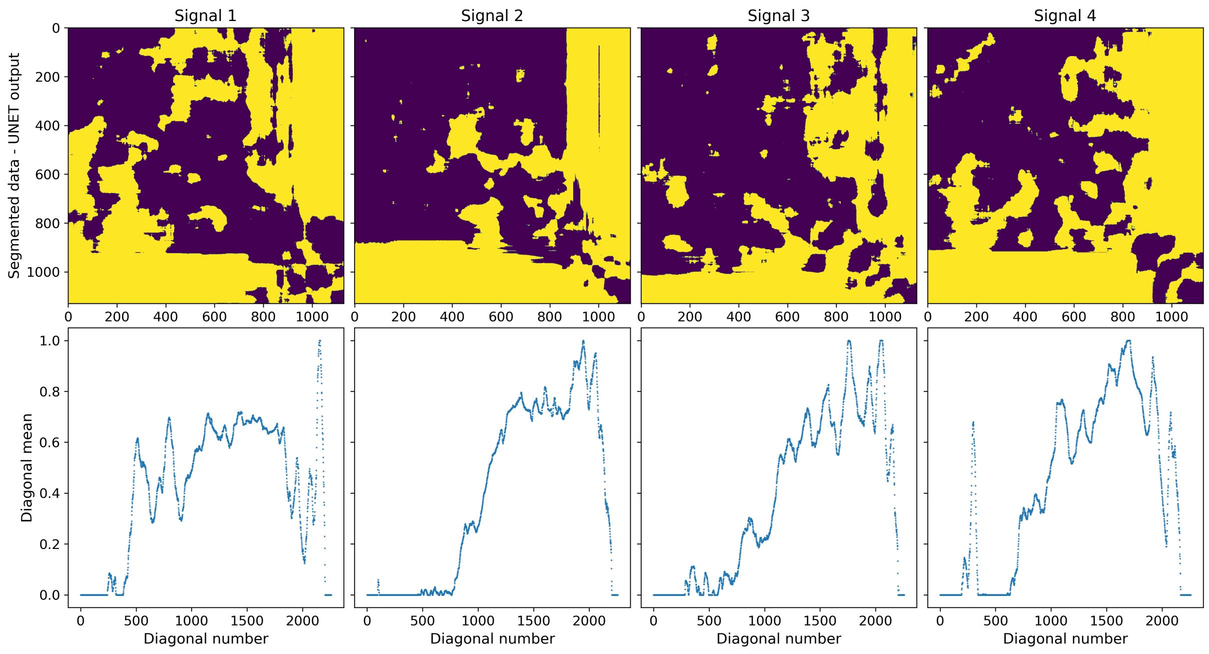

Figure 9. Mean of the values (**second row**) of the segmented distance matrices (**first row**) on each anti-diagonal.

The segmented images being of size 1130 × 1130, the temporary frontier signal is of size 2 × 1130 − 1, thus of size 2259. As stated in Section 2, the local maxima of these signals have to be found. If an anti-diagonal contains only yellow pixels, the signal will reach 1.

3.5. Cluster Frontier Extraction—Second Phase

The cluster frontier extraction, the *batch peak detection* Algorithm 2 of Section 2 is applied to each signal obtained at the previous phase. This current phase consists of local magnifying maxima in the signal to obtain the first discrete potential cluster boundaries. The dimension of the *batch peak detection* results is then shrunk by a factor of 2 to obtain the final signal of Figure 10. It now contains only 1130 values: one for each actuator cycle.

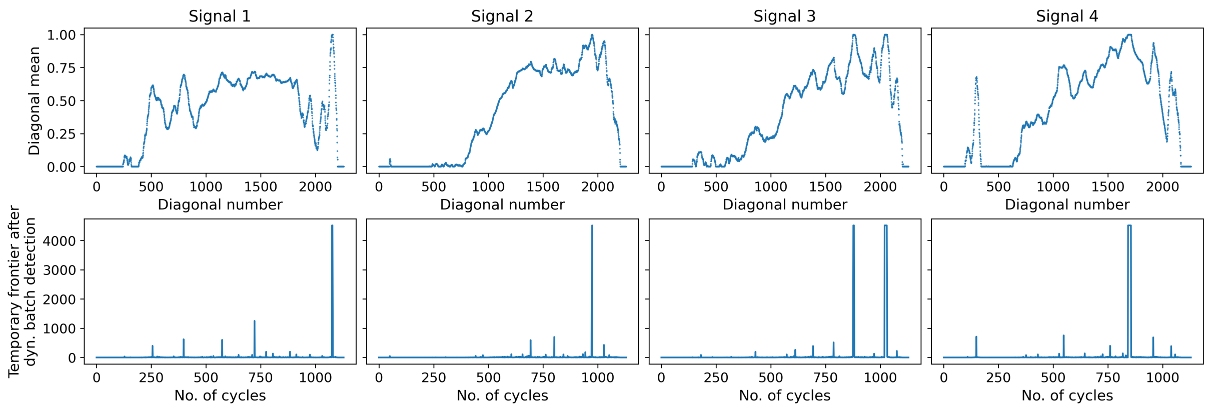

Figure 10. Frontier extraction (**second row**) from signal computed in Section 3.4 (**first row**) on each anti-diagonal.

From Figure 10, cluster frontiers are mainly concentrated after the 750th cycle. This behavior is consistent with motifs from Figure 7, where the discontinuity in signal essentially occurs at the bottom right of each image.

3.6. Cluster Frontier Extraction—Third Phase

In order to obtain a continuous signal out of the frontier extraction from phase two, a Gaussian kernel density estimation of *scipy.stats.gaussian_kde* with Silverman [19] bandwidth method is used. Then, the four frontier extraction signals from Figure 10 are summed, and a Gaussian kernel density estimation on this global frontier extraction is computed. The Figure 11 is obtained.

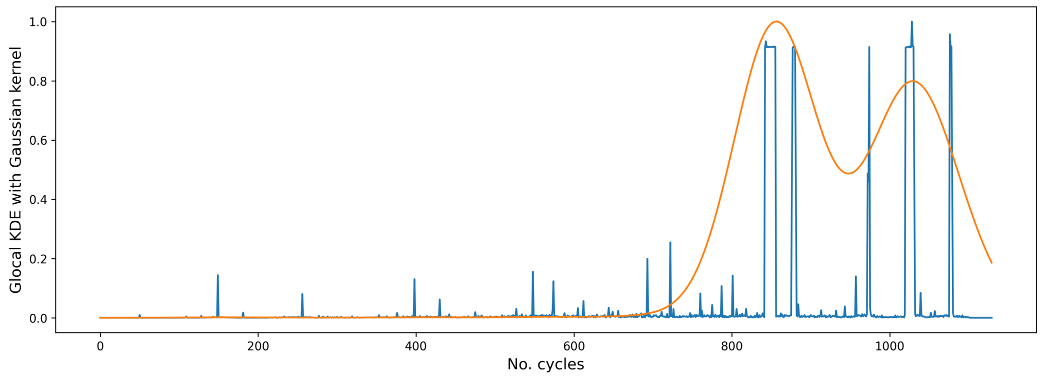

Figure 11. Gaussian kernel density estimation computed with the frontier extraction results of all four statistical parameters.

Each local maximum of the density function is likely to be a cluster frontier. Hence, two cluster frontiers are detected in Figure 11: a first one at cycle 856, and a second at cycle 1029. It means that the data are likely to have three clusters. The first cluster is from cycles 0 to 855, which corresponds to the first degree of severity where the system is healthy. The second cluster, from cycle 856 to cycle 1028, is where the system is in the second stage of fault severity. The last cluster, starting at cycle 1029, is the third and worst stage of fault severity.

The two independent components of the density estimation of Figure 11 are computed in the first column of Figure 12. The mean and the 95% respective confidence interval of each independent component are represented in second column of Figure 12.

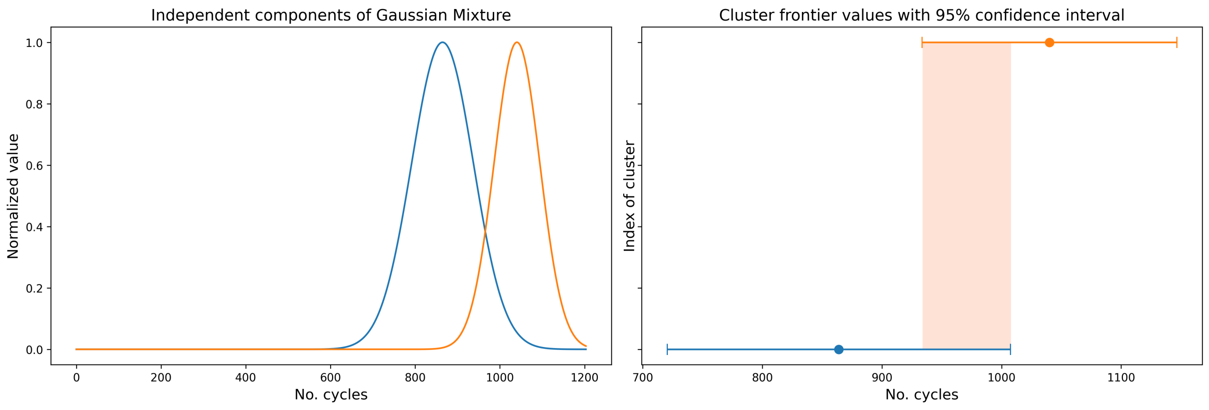

Figure 12. Cluster frontiers with 95% confidence interval.

3.7. Clustering Results

In order to represent the clustering results on the four descriptors, each local maximum of the density function is considered to be a cluster frontier. Figure 13 is obtained.

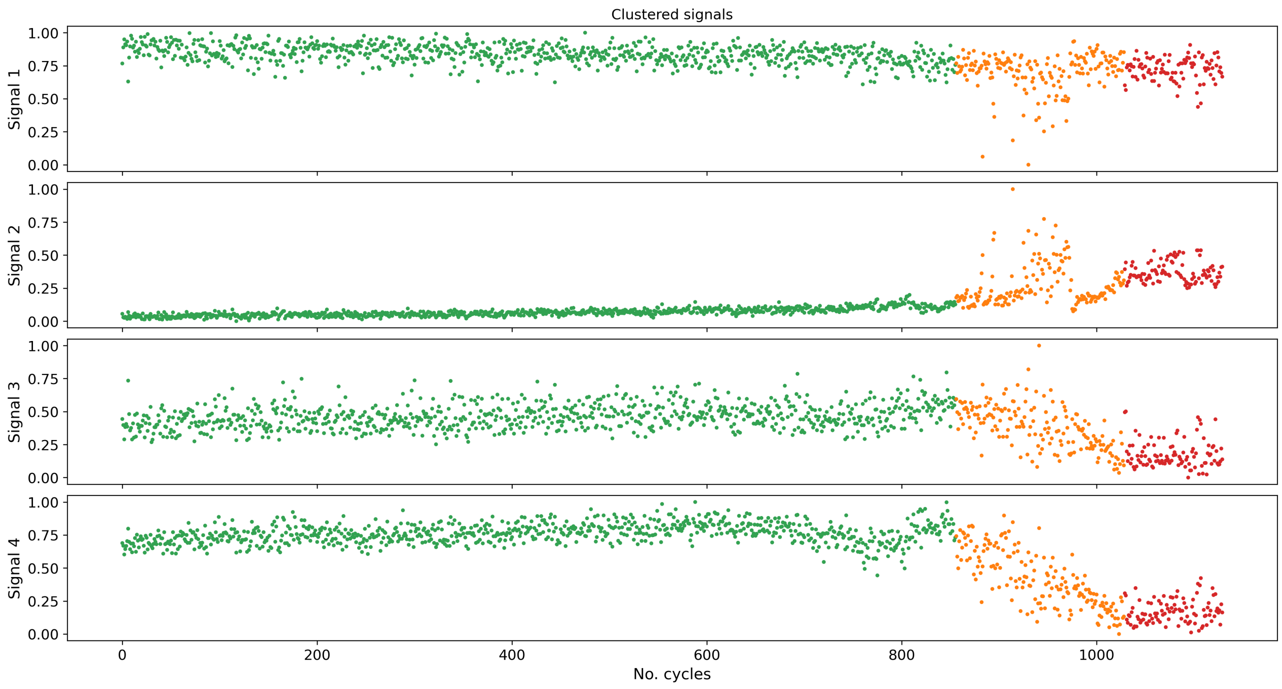

Figure 13. Clustering results of initial signals from Figure 6.

In Figure 13, each cluster has a different color corresponding to the health stage of the actuator. The first cluster is green, the second cluster is orange, and the last cluster is red. One can notice that the clusters obtained are coherent with the shapes of the signals.

4. Discussion

In this subsection, the clustering method is compared to other state-of-the-art algorithms. The latter are summarized in Table 1. They are taken from the Python library *scikit-learn* [21] except *Kmeans*, which comes from the *tslearn* [22] package.

Table 1. Clustering methods considered in our study.

Clustering Method	Category	Metric	Function Name	Parameters

K-Means	Partitioned	metric = 'euclidean'	KMeans	max_iter = 200
				n_jobs = −1
				init = 'k-means++'
				n_init = 10

		metric = 'dtw'		

K-medoids	Paritional	metric = 'euclidean'	KMedoids	max_iter = 200
				init = 'k-medoids++'

Mean Shift	Partitional	/	MeanShift	max_iter = 200
				n_jobs = −1

OPTICS	Density-based	metric = 'euclidean'	OPTICS	n_jobs = −1
				min_samples = 3

		metric = 'minkowski'		
Agglomerative	Hierarchical	metric = 'euclidean'	Agglomerative Clustering	min_cluster_size = 50
		metric = 'manhattan'		

Gaussian Mixture Model	Model-based	cov = 'spherical'	GaussianMixture	max_iter = 1000
				n_init = 100

		cov = 'diag'		
		cov = 'full'		
		cov = 'tied'		

These methods were selected according to two criteria. The first is the ease of implementation to our benchmark scope, as they are present in standard machine learning libraries in Python. The second is the representativeness of the diversity of the clustering algorithms. Indeed, according to [23] (Figure 6), they can be separated into six different approaches. At least three of them have been studied: partitional, hierarchical, and model-based. Each method in Table 1 was fed a data structure with five columns and 1130 lines. Four of the columns were the concatenation of $\mathbb{X}_{i\in[\![1;4]\!]}$, and the fifth one was an index array ranging from 0 to 1129 representing time. The same $\mathbb{X}_i$ of Figure 6 were used, allowing each algorithm to be aware of the time precedence relation between points. Each clustering method fits the data and then predicts the clustering labels. The labels are then stored in a matrix to represent the clustering results.

Since the clustering methods use the four $\mathbb{X}_i$ to produce one clustering result, the clustering outcome can only be represented for one signal. Figure 14 plots the clustered first signal. The number of groups in data were chosen according to a new quality indicator.

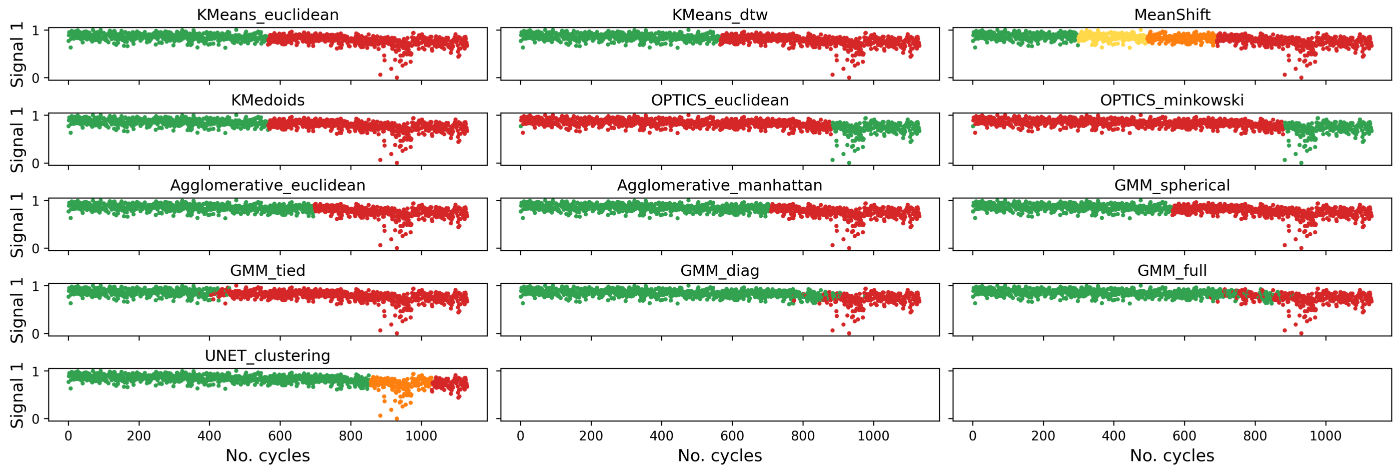

Figure 14. Clustering results with state-of-the-art algorithms for $\mathbb{X}_1$.

In Figure 14, the majority of the methods have detected two clusters, except for *MeanShift* and our clustering method *UNET clustering*. One detects four groups, whereas ours provides three groups in the data, as seen previously in Figure 13. Regarding the shape of signal $\mathbb{X}_1$, three clusters seem to be the optimal number embedded in the data.

To evaluate the accuracy of our method regarding the state of the art, a new clustering indicator is created. Indeed, whereas indicators for spatial clustering are well known, specific existing indicators to infer clustering quality for internal time series clustering are scarce. The *S_Dbw* [24] can be cited to measure the compactness and the good separation of clusters or the exhaustive review of [23] on time series clustering. Additionally, recently, an invariance guided criterion was created by [25]. Nevertheless, these indexes can be separated into two main groups: external and the internal ones [26]. Since class labels for our data are unknown, one must count on information contained in the studied data as stated in [26]. Eleven widely used indicators are compared. In this work, *Silhouette* [27], the Davies–Bouldin separation measure [28], the Calinski Harabasz [29] indicator, the conjunction of an inertia and temporal consistency measure from [12], and finally the SD [30] and *S_dbw* [24] validity indexes are tested. None of them could manage to select the right clustering result in Figure 14. As those are not fitted to the task, the new measure should quantify relevant information according to the physics of the system. On the first hand, as stated in fundamental Hypothesis 1, the studied system cannot be repaired during its life cycle. Once the severity degree has increased, i.e., once a cluster frontier is crossed, the system cannot go back to a previous, healthier stage. In other words, each class must be time continuous. Temporal jumps in clusters are forbidden. Consequently, the indicator must penalize any temporal inconsistency. At the second end, one can see in Figure 14 that good clustering is made when two consecutive groups do not share the same density properties. This shape information also has to be captured.

Equation (7) presents the newly developed quality indicator for internal clustering of time series.

$$\chi = \frac{1}{m} \sum_{j=0}^{m-1} \left| a_{j+1} - a_j \right|, \text{ with} \tag{7}$$

$$a_j = \frac{\sigma_j \left(\max y_j - \min y_j \right)}{\dfrac{1}{n_j - 1} \sum\limits_{\substack{k=0 \\ k \in \mathcal{J}}}^{n_j - 1} t_{k+1} - t_k} \tag{8}$$

where m is the number of clusters found in the data, σ_j is the standard deviation of the jth cluster, $\max y_j - \min y_j$ is the range of the jth cluster, y_j is its elements, t_k is its respective indexes, and n_j is its number of points. Note that $\mathcal{J}$ is the ensemble of the point indexes in cluster j. While the denominator of a_j measures the temporal consistency, the numerator measures the shape of the clusters.

Finally, the quality index χ deals with the variation between two consecutive groups and not a_j in itself. Note that the denominator in a_j can never be null for time series. Indeed, Figure 14 shows, for the new algorithm, that the first and third data partitions exhibit similar dynamics. However, since the system cannot rejuvenate itself, those should be considered two different groups. To recognize the better partitioning result, the value $\mathcal{I}$ defined in Equation (7) has to be maximized.

The ideal number of clusters in Figure 14 is selected according to the results of the Table 2. Note that the new clustering algorithm obtains the global maximum of the array. For ease of reading, the maximum obtained for each algorithm is highlighted in green, whereas the global maximum of the Table 2 is highlighted in dark green. Two main trends can be extracted from these results. There are some indicator values in the neighborhood of the global maximum 0.083: *KMeans_euclidean*, *KMeans_dtw*, *KMedoids*, *Agglomerative_euclidean*, *Agglomerative_manhattan* and *GMM_spherical*. This result is coherent with the presented data, as these algorithms tend to cluster data similarly; thus, all the partitioning algorithms have very similar indicator values. Moreover it seems that the chosen distance does not have a significant impact on the results.

Table 2. Global indicator χ value depending on the clustering method and the number of clusters.

	Number of Clusters						
	2	**3**	**4**	**5**	**6**	**7**	**8**
KMeans_euclidean	0.074	0.051	0.034	0.025	0.026	0.035	0.032
KMeans_dtw	0.074	0.051	0.036	0.025	0.026	0.035	0.031
MeanShift			0.040				
KMedoids	0.074	0.051	0.034	0.026	0.027	0.037	0.032
OPTICS_euclidean	0.008						
OPTICS_minkowski	0.008						
Agglomerative_euclidean	0.073	0.050	0.024	0.020	0.018	0.019	0.017
Agglomerative_manhattan	0.074	0.051	0.038	0.032	0.027	0.030	0.027
GMM_spherical	0.074	0.051	0.034	0.025	0.026	0.034	0.032
GMM_tied	0.068	0.021	0.025	0.021	0.012	0.011	0.009
GMM_diag	0.041	0.027	0.027	0.023	0.013	0.011	0.011
GMM_full	0.049	0.034	0.026	0.023	0.028	0.023	0.021
UNET_clustering		0.083					

To better grasp the idea of the indicator, two tables are separately given. Table 3 represents the contribution of the *shape* measure to χ, and Table 4 represents the temporal consistency one. Practically, Table 3 contains the values of $\sigma_j \left(\max y_j - \min y_j \right)$, and Table 4 contains the values of $\frac{1}{n_j-1}\sum_{k=0}^{n_j-1} t_{k+1}-t_k$ for each clustering results. From that, one can see that the *shape* part of the indicator is coherent with the results found in Figure 14, as different clustering behavior obtains different values. From Equation (7), one could infer that there is no temporal inconsistency. This is proved by the unit value of the denominator. The problem with such methods as *GMM_tied*, *GMM_diag*, *GMM_full*, *OPTICS_euclidean* and *OPTICS_minkowski* is that their clusters are not time continuous. Hence, the value in Table 4 is greater than one. Thus, the numerator of a_j is divided by an amount greater than one, which minimizes its value. This penalization is then propagated in the total calculus of χ.

Table 3. Shape score $\sigma_j\left(\max y_j - \min y_j\right)$ depending on the clustering method and the number of clusters.

	Number of Clusters						
	2	**3**	**4**	**5**	**6**	**7**	**8**
KMeans_euclidean	0.074	0.051	0.034	0.025	0.026	0.035	0.032
KMeans_dtw	0.074	0.051	0.036	0.025	0.026	0.035	0.031
MeanShift			0.040				
KMedoids	0.074	0.051	0.034	0.026	0.027	0.037	0.032
OPTICS_euclidean	0.055						
OPTICS_minkowski	0.055						
Agglomerative_euclidean	0.073	0.050	0.024	0.020	0.018	0.019	0.017
Agglomerative_manhattan	0.074	0.051	0.038	0.032	0.027	0.030	0.027
GMM_spherical	0.074	0.051	0.034	0.025	0.026	0.034	0.032
GMM_tied	0.072	0.037	0.028	0.023	0.018	0.016	0.015
GMM_diag	0.061	0.042	0.030	0.026	0.039	0.035	0.031
GMM_full	0.068	0.036	0.028	0.025	0.040	0.037	0.033
UNET_clustering		0.083					

The same formatting used in Table 2. Shades of green are used to highlight the minimal value for each algorithm and the global minima of the table.

Table 4. Temporal consistency score depending on the clustering method and the number of clusters.

	Number of clusters						
	2	**3**	**4**	**5**	**6**	**7**	**8**
KMeans_euclidean	1.000	1.000	1.000	1.000	1.000	1.000	1.000
KMeans_dtw	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MeanShift			1.000				
KMedoids	1.000	1.000	1.000	1.000	1.000	1.000	1.000
OPTICS_euclidean	2.785						
OPTICS_minkowski	2.785						
Agglomerative_euclidean	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Agglomerative_manhattan	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GMM_spherical	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GMM_tied	1.075	1.310	1.197	1.268	1.919	1.804	1.789
GMM_diag	1.259	1.758	1.903	2.503	2.700	3.002	2.980
GMM_full	1.247	1.472	1.633	1.746	1.609	1.889	1.761
UNET_clustering		1.000					

In Table 4, algorithms that do not produce any temporal inconsistencies have a minimal value of one. Dark green is used to highlight the result of the developed method. For other algorithms, shades of orange to red are used to highlight positive variation.

5. Conclusions

Prior work in time series partitioning for condition monitoring in aircraft systems did not consider the temporal dimension of data, as illustrated by [7]. It used K-Nearest Neighbors or the Support Vector Machine between others to cluster data without considering its time precedence relation. The work of [12] is the most recent attempt, to our knowledge, to partition time series while using and keeping time information. However, the latter required the user to have prior knowledge of the data. By doing so, the risk is for the user to add their own biases into the procedure. Moreover, as developed in [9], it is challenging to find a cluster indicator that could assess the clustering quality algorithm, as it is highly dependent on the hypotheses made in the research and the dataset used.

This study developed a new clustering method for time series with a deep neural network core. The whole procedure consists of three main steps. First, temporal data are encoded into pictures to be fed to the DNN. Then, the segmented images are processed with consistent signal processing algorithms. Finally, the information from all the descriptors is fused to obtain the general data partitioning profile.

The whole algorithm empowers us to find cluster frontiers with a likelihood measure from a set of time series. It was designed specifically to consider the precedence relation of elements and to avoid temporal consistency incoherence. It was applied to an industrial dataset to assess its relevance. Note that the whole algorithm could be used with multivariate time series, as the nature of the data is irrelevant to the results. Hence, the automatic partitioning of time series without prior knowledge in a prognosis and health monitoring context is possible. Although it requires more significant computing resources than its counterpart to train the deep structure, this new approach is a real improvement on traditional techniques since the neural network and the signal processing operations following the image segmentation can cluster very noisy data, and the learning phase is done only once. The better the deep structure training is, the better the clustering results will be.

Furthermore, to quantify the results of this new method against state-of-the-art machine learning, a new indicator was created, alleviating the downsides of the usual indicators for clustering. Whereas the state-of-the-art focuses on measuring the relative shapes of the groups made by the clustering algorithm, ours can evaluate two properties: the temporal consistency of clustering and the form of the cluster. We designed it with our physical assumptions in mind; hence, it exhibits better results than its state-of-the-art counterpart. A variant from the usual indicators can be found in [25], where the clustering result invariance is quantified.

Finally, this whole method is a stepping stone to a more broad framework, aiming to determine the remaining useful life of actuators in a PHM framework. Future work should adapt this method without considering Hypothesis 2 or Hypothesis 1 to use the process in a broader framework.

Author Contributions: Conceptualization, A.E., G.C. and B.M.; methodology, A.E., G.C. and B.M.; software, A.E. and S.R.; validation, A.E.; formal analysis, A.E.; investigation, A.E.; resources, B.M.; data curation, A.E.; writing—original draft preparation, A.E., G.C., B.M. and S.R.; writing—review and editing, A.E., G.C. and B.M.; visualization, A.E. and S.R.; supervision, G.C. and B.M.; project administration, G.C. and B.M.; funding acquisition, B.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the French agency for research and technologye (ANRT) through the CIFRE grant 2018/0649, by Safran Electronics & Defense (Safran Group) and by laboratoire Ampère CNRS—UMR5005 own funds.

Data Availability Statement: 3rd Party Data. Restrictions apply to the availability of these data. Data was obtained from Safran Electronics & Defense and are not publicly available.

Acknowledgments: This work was supported by the PMCS2I from École Centrale de Lyon, France.

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

Abbreviations

The following abbreviations are used in this manuscript:

CCPF	Constant Continuous Piecewise Function
DL	Deep Learning
DNN	Deep Neural Network
GDDR	Graphics Double Data Rate
GPU	Graphics Processing Unit
HPC	High Performance Computing
IATA	International Air Transport Association
MCTG	Maintenance Cost Technical Group
MRO	Maintenance Repair and Overhaul
KDE	Kernel Density Estimation
LCPF	Linear Continuous Piecewise Function
PHM	Prognosis and Health Management
RAM	Random Access Memory
VRAM	Video Random Access Memory

References

1. Fink, O.; Wang, Q.; Svensén, M.; Dersin, P.; Lee, W.J.; Ducoffe, M. Potential, challenges and future directions for deep learning in prognostics and health management applications. *Eng. Appl. Artif. Intell.* **2020**, *92*, 103678. [CrossRef]
2. Wang, Z.; Oates, T. Spatially Encoding Temporal Correlations to Classify Temporal Data Using Convolutional Neural Networks. *arXiv* **2015**, arXiv:1509.07481.
3. Hatami, N.; Gavet, Y.; Debayle, J. Classification of Time-Series Images Using Deep Convolutional Neural Networks. In Proceedings of the Tenth International Conference on Machine Vision (ICMV 2017) 2018, Rome, Italy, 8–12 November 2018; Volume 10696. [CrossRef]
4. Jain, A.K.; Murty, M.N.; Flynn, P.J. Data clustering: A review. *ACM Comput. Surv.* **1999**, *31*, 264–323. [CrossRef]
5. Hendrickx, K.; Meert, W.; Mollet, Y.; Gyselinck, J.; Cornelis, B.; Gryllias, K.; Davis, J. A general anomaly detection framework for fleet-based condition monitoring of machines. *Mech. Syst. Signal Process.* **2020**, *139*, 106585. [CrossRef]
6. Perafán-lópez, J.C.; Sierra-pérez, J. An unsupervised pattern recognition methodology based on factor analysis and a genetic-DBSCAN algorithm to infer operational conditions from strain measurements in structural applications. *Chin. J. Aeronaut.* **2020**, *34*, 165–181. [CrossRef]
7. Zaporowska, A.; Liu, H.; Zakwan, S.; Yifan, Z. A clustering approach to detect faults with multi-component degradations in aircraft fuel systems. *IFAC-PapersOnLine* **2020**, *53*, 113–118. [CrossRef]
8. Chen, J.; Chen, S.; Liu, Z.; Luo, C.; Jing, Z.; Xu, Q. Health Monitoring of Landing Gear Retraction/Extension System Based on Optimized Fuzzy C-Means Algorithm. *IEEE Access* **2020**, *8*, 219611–219621. [CrossRef]
9. Javed, A.; Lee, B.S.; Rizzo, D.M. A benchmark study on time series clustering. *Mach. Learn. Appl.* **2020**, *1*, 100001. [CrossRef]
10. Saxena, A.; Goebel, K. *Turbofan Engine Degradation Simulation Data Set*; NASA Ames Prognostics Data Repository: Washington, DC, USA, 2008.
11. Arias Chao, M.; Kulkarni, C.; Goebel, K.; Fink, O. Aircraft Engine Run-to-Failure Dataset under Real Flight Conditions for Prognostics and Diagnostics. *Data* **2021**, *6*, 5. [CrossRef]
12. Breuneval, R. Surveillance de L'état de Santé des Actionneurs Électromécaniques: Application à L'aéronautique. Ph.D. Thesis, Université de Lyon, Lyon, France, 2017.
13. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* **2015**, arXiv:1505.04597.
14. Lin, T.Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common Objects in Context. *arXiv* **2015**, arXiv:1405.0312.
15. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*; Wallach, H., Larochelle, H., Beygelzimer, A., dAlché Buc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2019; pp. 8024–8035.
16. Iqbal, H. *HarisIqbal88/PlotNeuralNet*; Programmers: _:n2500 Original-Date: 2018-07-24T16:51:34Z. 2020. Available online: https://doi.org/10.5281/zenodo.2526396 (accessed on 18 September 2020).
17. Parzen, E. On Estimation of a Probability Density Function and Mode. *Ann. Math. Stat.* **1962**, *33*, 1065–1076. [CrossRef]
18. Scott, D.W. *Multivariate Density Estimation: Theory, Practice, and Visualization*, 2nd ed.; Wiley Series in Probability and Statistics; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2015. [CrossRef]
19. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; CRC Press: Boca Raton, FL, USA, 1986.
20. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-scale machine learning on heterogeneous systems. *arXiv* **2015**, arXiv:1603.04467.

21. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
22. Tavenard, R.; Faouzi, J.; Vandewiele, G.; Divo, F.; Androz, G.; Holtz, C.; Payne, M.; Yurchak, R.; Rußwurm, M.; Kolar, K.; et al. Tslearn, a machine learning toolkit for time series data. *J. Mach. Learn. Res.* **2020**, *21*, 1–6.
23. Aghabozorgi, S.; Seyed Shirkhorshidi, A.; Ying Wah, T. Time-series clustering—A decade review. *Inf. Syst.* **2015**, *53*, 16–38. [CrossRef]
24. Halkidi, M.; Vazirgiannis, M. Clustering validity assessment: finding the optimal partitioning of a data set. In Proceedings of the 2001 IEEE International Conference on Data Mining, San Jose, CA, USA, 29 November–2 December 2001; pp. 187–194. [CrossRef]
25. Forest, F.; Mourer, A.; Lebbah, M.; Azzag, H.; Lacaille, J. An Invariance-guided Stability Criterion for Time Series Clustering Validation. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 9296–9303. [CrossRef]
26. Liu, Y.; Li, Z.; Xiong, H.; Gao, X.; Wu, J. Understanding of Internal Clustering Validation Measures. In Proceedings of the 2010 IEEE International Conference on Data Mining, Sydney, Australia, 13–17 December 2010; pp. 911–916. [CrossRef]
27. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [CrossRef]
28. Davies, D.L.; Bouldin, D.W. A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *PAMI-1*, 224–227. [CrossRef]
29. Caliński, T.; Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat.* **1974**, *3*, 1–27. [CrossRef]
30. Halkidi, M.; Vazirgiannis, M.; Batistakis, Y. Quality Scheme Assessment in the Clustering Process. In Proceedings of the European Conference on Principles of Data Mining and Knowledge Discovery 2000, Lyon, France, 13–16 September 2000. [CrossRef]