infrastructures

MDPI

Technical Note

Clustering and Vectorizing Acoustic Emission Events of Large Infrastructures' Normal Operation

Theocharis Tsenis *,†,‡ and Vassilios Kappatos ‡

Department C: Non-land Transport, Environmental and Economic Issues, Sector of Hellenic Institute of Transportation (HIT), Centre for Research and Technology Hellas (CERTH), 57001 Thessaloniki, Greece; vkappatos@certh.gr

* Correspondence: ttsenis@certh.gr

† Current address: Centre for Research and Technology Hellas (CERTH), 6th km Thermi–Charilaou, 57001 Thessaloniki, Greece.

‡ These authors contributed equally to this work.

Abstract: The detection of acoustic emission events from various failing mechanisms, such as plastic deformations, is a critical element in the monitoring and timely detection of structural failures in infrastructures. This study focuses on the detection of such failures in metal gates at rivers' lifting dams aiming to increase the reliability of river transport compared to the current situation, thereby, increasing the resilience of transport corridors. During our study, we used lifting dams in both France and Italy where river transport is thriving. A methodology was developed, processing corresponding acoustic emission recordings originating from lifting dams' metal gates, using advanced denoising—preprocessing, various decompositions, and spectral embeddings associated with various latest nonlinear processing clustering techniques—thus providing a detailed cluster label morphology and profile of water gates' normal operating area. Latest machine learning outlier detection algorithms, like One-Class Support Vector Machine, Variational Auto-Encoder, and others, were incorporated, producing a vector of confidence on upcoming out-of-the-normal gate operation and failure prediction, achieving detection contrast enhancement on out-of-the-normal operation points up to 400%.

Keywords: acoustic emission; infrastructure; support vector machine; variational auto-encoder; spectral clustering

check for updates

Received: 9 December 2024
Revised: 31 January 2025
Accepted: 4 February 2025
Published: 11 February 2025

Citation: Tsenis, T.; Kappatos, V. Clustering and Vectorizing Acoustic Emission Events of Large Infrastructures' Normal Operation. *Infrastructures* **2025**, *10*, 38. https://doi.org/10.3390/infrastructures10020038

Correction Statement: This article has been republished with a minor change. The change does not affect the scientific content of the article and further details are available within the backmatter of the website version of this article.

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Acoustic emission (AE) technology is a non-destructive testing method used for monitoring the structural integrity of materials and components. It works by detecting the transient elastic waves produced by the rapid release of energy from localized sources within a material, such as cracks, dislocations, or other abnormal structural changes. These emissions are typically high-frequency sounds that are inaudible to the human ear but can be detected using sensitive piezoelectric sensors.

AE techniques have demonstrated remarkable suitability for detecting defects across a variety of materials, like composites and metals, due to their sensitivity and adaptability. Studies such as those by Tsenis et al. [1] and Spyrou et al. [2] illustrate the effectiveness of AE in identifying defects within composite specimens used in the transportation domain. These works highlight the capability of AE to detect internal flaws and provide precise insights into structural integrity, emphasizing its relevance in environments where composite materials are prevalent [1,2]. There are dozens and dozens of other papers emphasizing the usage of AE in failure detection in a very large set of materials.

Moreover, AE has proven to be a powerful tool for failure detection in large infrastructures, including maritime and offshore structures. For instance, Kappatos et al. [3] discussed the application of AE for structural health monitoring (SHM) in ships and offshore structures, showcasing its ability to continuously monitor and detect critical anomalies before catastrophic failures occur. Additionally, studies focusing on AE source localization through neural nets, such as those by Kappatos et al. [4], emphasize advancements in using neural networks and feature selection for robust AE source localization in ship hulls, demonstrating the method's effectiveness in addressing the complexities of large-scale infrastructure monitoring [3–5].

These findings collectively underline the versatility of AE techniques, not only in detecting defects in various materials like composites and many others, but also in ensuring the overall structural integrity of large-scale infrastructures, thus preventing failures and enhancing operational safety. Focusing on improving the climate resilience and sustainability of transport infrastructure, particularly at inland waterways, AE could be employed for inspecting waterlock gates. Here is a brief overview of such applications:

- Detection of Cracks and Leaks: AE sensors are placed on or around the waterlock gates to listen for sounds generated by the formation or propagation of cracks due to stress, fatigue, or material degradation. This could be particularly useful in detecting microcracks before they become significant structural issues.
- Material Integrity: Over time, materials like steel or concrete in waterlock gates can degrade due to environmental conditions (e.g., water corrosion, freeze–thaw cycles). AE monitoring can help in assessing the health of these materials by identifying when and where degradation processes are accelerating.
- Operational Safety: By continuously monitoring with AE during the operation of the gates, any sudden changes in the structural integrity can be detected in real-time, allowing for immediate response to prevent potential failures or safety hazards.
- Preventive Maintenance: Data from AE recordings can inform maintenance schedules, allowing for repairs or replacements to be made before major issues arise, thus extending the service life of the infrastructure.

In large-infrastructure systems, the failure process often follows a complicated pattern that complicates early detection and prediction of failure, particularly using stand-alone traditional methods found in machine learning (ML) and statistics. This pattern typically involves the following:

- A long plateau of good operation: The structure operates without any noticeable signs of failure for an extended period, which could span years or even decades. During this time, small, gradual damage may accumulate at a microscopic level, but it does not manifest in easily detectable ways, making it appear as though the structure is still in optimal condition [6].
- Sudden, catastrophic failure: After the plateau of seemingly stable operation, failure often occurs abruptly and unexpectedly, sometimes in the form of a cascading or avalanche mechanism. This means that small damages rapidly accumulate and propagate, leading to a catastrophic failure within a short time frame. This transition from healthy operation to failure happens too quickly for traditional monitoring techniques to provide adequate warning [7].

The resulting issues faced in failure prediction:

- Lack of Labeled Failure Events: For most of the operational life of large metallic structures, there are no observable 'failure' data. This absence of failure signals over long periods means that datasets used for machine learning models are overwhelmingly biased toward 'good' or normal operation, with very few, if any, samples representing

failure. In traditional ML models, this leads to a severe class imbalance problem, where failure is significantly under-represented or even non-existent in the training data [8,9].

- Nonlinear Damage Accumulation: The damage leading up to failure does not follow a linear progression that can be easily modeled. Small events over time may not be indicative of future failure, as the failure itself may be driven by complex interactions between different damage mechanisms (e.g., cracks, corrosion, fatigue) that accelerate suddenly [10,11].
- Good Operation Plateau: During the long period of normal operation, AE signals may be minimal or exhibit patterns that suggest healthy functioning, even though microscopic damage is accumulating. This plateau complicates the task of detecting when a structure is nearing failure, because early signs of failure may not appear at all or may be so subtle that they are indistinguishable from normal operational noise [12].
- Failure Mechanism: When failure occurs, it often happens quickly, with cascading or avalanche-like propagation of cracks or structural breakdown. This abrupt nature leaves little or no warning time to react or implement preventative measures once the signs become apparent [13].

Moving on, the structure of this paper includes firstly a description of the used methodology, the experimentation setup followed during the AE recording in the visited waterlocks, and finally, the results and a discussion of the conclusions.

2. Our Methodology

In the context of monitoring large metallic infrastructure, AEs are often captured to assess the structural integrity over time. During the normal operation of such infrastructure, AE signals typically exhibit patterns that are consistent with the expected behavior of the structure under normal loads and environmental conditions. These patterns can be visualized as points within a defined 'Normal Operating Space' (NOS) in a multidimensional data space where each dimension represents a characteristic of the AE signal (like amplitude, frequency, and energy). This space is essentially a model of what a 'healthy' operation looks like, established through baseline measurements and historical data. When the infrastructure begins to experience degradation or is nearing a failure state, the nature of AE signals changes. These changes can include an increase in signal amplitude, shifts in frequency content, or altered energy release patterns, among others. Such altered signals would then manifest as outliers relative to the established NOS. Detecting these outliers involves advanced signal processing and anomaly detection techniques, aiming to identify deviations that could signify the initiation or progression of damage mechanisms like crack growth, corrosion, or fatigue. The challenge lies in distinguishing these subtle changes from the normal operational noise or environmental variations, which requires sophisticated algorithms capable of detecting changes in the outlier profile [14].

The process of cropping recorded AE data and coverting them into points in the NOS is called 'pixelization' in this paper, and each point is a 'superpixel'. What follows is a description of the developed 'pixelization' process on waterlocks' AEs, and a description of the 'superpixel' structure. Afterwards, the outlier profiling and change detection methodology are described.

2.1. Normal Operating Space Preprocessing and Pixelization

The first step was to read and group the AE recordings stored in the form of multiple CSVs each named with a timestamp, sequence number, and sensor-channel. A dedicated developed function was used, parsing the CSV directory while excluding non-relevant files or subfolders, creating a structured dataset by splitting file paths and names into meaning-

ful components (e.g., timestamp, sensor channel), grouping files by unique identifiers (e.g., folder names, timestamps, and sensor channels). Through a second dedicated function, these organized groups of recordings were read from median, band-pass filters and then decomposed into heuristically independent components (EMD) [15,16]. The next step was the grouping of the waveforms, consisting of the filtered raw signal and its constituents. Each member of each group was slotified by using all of the following slotification methodologies in parallel:

- Normal Slot Division: Divides the data into fixed-size slots.
- Energy-Based Slot Division: Groups into slots based on energy peaks in the RMS graph of the signal, and uses a threshold for the number of significant peaks to consider within a slot, which are adjusted dynamically based on the distribution of energy peaks
- Rise-Time-Based Slot Division: Groups into slots based on the rise time of the signal, using the time interval between the first threshold crossing and the signal peak.
- Average-Frequency-Based Slot Division: Groups into slots based on the average frequency of the signal.

This differentiation in slotification methodologies proved to be a good mechanism in spotting small deviations from the NOS, attributed to microcracks or other accumulating failing micro-mechanisms. For each slot, taken from raw or decomposed signal, the following features were calculated:

- Spectral Power Band Features: They contain the features of four (adjustable—nth) spectral bands, and for each band extracting the RMS, number of peaks inside the band, and the mean value, variance, and kurtosis of the band's spectral signal.
- HITS Features: They include the peaks inside the time slot and their heights, the number of crosses above a threshold inside the time slot, the number of crosses until the peak value inside the time slot, the rise time, the duration, the average frequency, the initial frequency and the reverberation frequency [17].
- Statistical Features: The mean, variance, kurtosis, and skewness over the whole time slot.

Following this was another developed function as a post-processing utility that organizes the features extracted by the previously mentioned function into structured groupings, 'superpixels', and prepares them for further analysis. It groups slots features into 'superpixels' according to time relevance and sensor type. So, eventually, the original recordings per sensor channel are reduced into a group of 'superpixels' with dimensionality varying depending on the decomposing components, power bands, and others, usually in the range of 1024–1300. Figure 1 displays the whole process starting from the raw AE recordings, including slotification and feature extraction per slot, and the alignment of slots' features according to time and channel, producing a multidimensional space. For the reference set of AE recordings, this multidimensional set of points is the NOS.

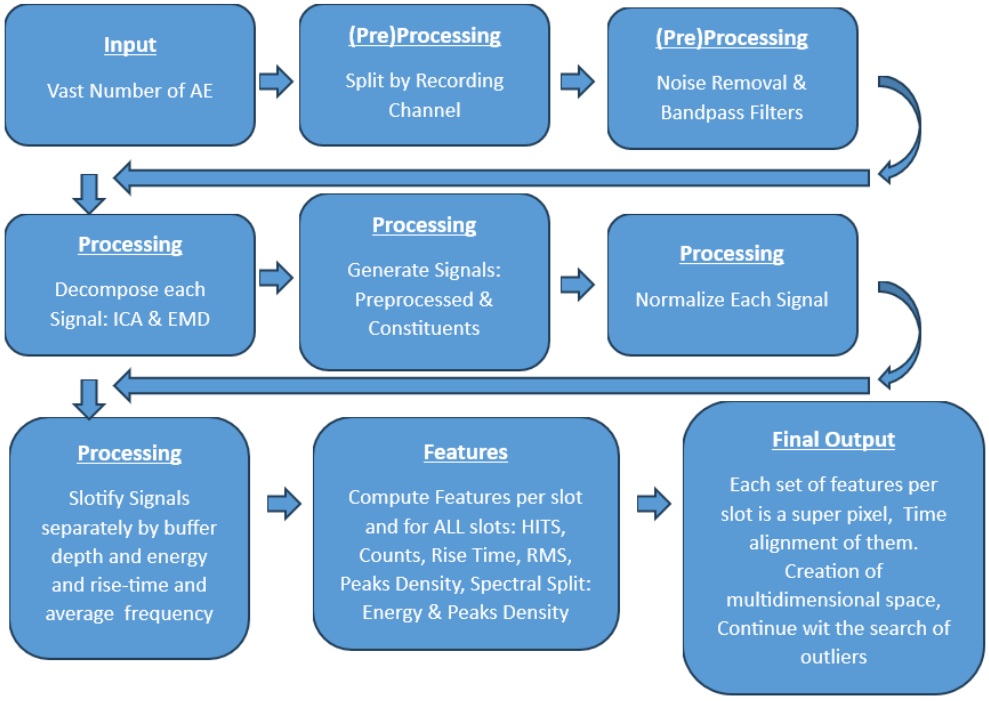

Figure 1. Our methodology layout from accepting the AE signals up to feature extraction and representation in a multidimensional space.

2.2. Mining Failures in Outliers' Profiles

Assuming an initial set of 'superpixels' corresponding to the waterlock gates' NOS, attributed as a reference set, a comparison using as a basis this reference set for identifying future outlying patterns, as opposed to each newly recorded inspection set of AE 'superpixels', lessens the issue. The first sign of abnormal behavior brought on by different malfunctioning processes is the density of observed outliers relative to the reference set. For technicians, the persistence of such outlier densities is a great indication that predictive maintenance should be carried out soon. To show that this comparison with a reference set is a sensitive enough methodology to detect them, we employed in our tests two complementary sets of AE recordings from simulated failures. Since the simulated failures are microcracks, which are precursors to big metal crack failures, we used this outlier identification mechanism as a sensitivity validation methodology. The following outlier methods over the set of 'superpixels' were used in parallel:

- Statistical Mahalanobis Distance Outlier Detection: In this method, although we calculate the Mahalanobis distance and the end from the new set of Mahalanobis distances per 'superpixel', two mean deviation values on the kurtosis and skewness are produced as the final output [18].
- One-Class Support Vector Machine (SVM) Outlier Detection: In this method, although every 'superpixel' produced by the AE recording under investigation is evaluated based on the reference support vector to predict if it is an outlier or not, an average is produced on all 'superpixels' [19].
- Variational Auto-Encoder (VAE) Outlier Detection, where a VAE was trained on the NOS 'superpixels' and the latent statistical space produced by the failing set of 'superpixels' fed to the trained VAE produced a reconstruction statistical space utilizing Mahalanobis distances to detect deviations. Again, in this method, every 'superpixel' is transformed in a statistical latent space with point dimension 2, as created by the reference set of 'superpixels', and the derived features are the mean, std, and

skewness. Over each dimension and for each feature, a z_score was calculated so an array with depth the number of superpixels and length 12 was produced. An append to this vector was to add the Mahalanobis distances of latent space features (mean, std, skewness), thereby adding an additional 6 values to the output vector. Finally, to conclude on the outlier analysis implemented on this output vector, a DBCAN evaluation was added (single value). DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a popular clustering algorithm used in data mining and machine learning. Unlike other clustering methods, such as k-means, DBSCAN does not require the number of clusters to be specified in advance. Instead, it groups together points that are closely packed together (points with many nearby neighbors), marking points in low-density regions as outliers or noise. The method makes an autogroup over the reference superpixels and the new under-inspection 'superpixels'. The number of unconnected 'superpixels' between the two sets is compared and a percentage of difference is produced, which is the number added at the final output outlying vector of this methodology. The dimension of the final output vector is 19 [20].

- Using various embeddings to reduce dimensionality and then clustering them into clusters, with the final outputs been passed again from an outlier detection of One-Class SVM. The embeddings used are 'Locally Linear Embedding'(LLE), 'Locality-Preserving Projection', 'Isomap', 'Spectral Embedding', and 'Multidimensional Scaling' (MDS) [21]. The clustering methodology used is 'Spectral Clustering'(SC). From each set of clusters per methodology, always having the same number of clusters, the silhouette values were extracted and the pairs of normal operation and failure were passed from One-Class SVM outlier detection. In this set of methods, the pair of methods and number of clusters utilized were the following:
  - Number of clusters: 500; number of neighbors: 800; number of components: 25; dimension reduction method: 'LLE'; clustering method: SC.
  - Number of clusters: 500; number of neighbors: 800; number of components: 25; dimension reduction method: 'LPP'; clustering method: SC.
  - Number of clusters: 500; number of neighbors: 800; number of components: 25; dimension reduction method: 'Isomap'; clustering method: SC.
  - Number of clusters: 500; number of neighbors: 800; number of components: 25; dimension reduction method: 'SE'; clustering method: SC.
  - Number of clusters: 500; number of neighbors: 800; number of components: 25; dimension reduction method: 'MD5'; clustering method: SC.

  The results from each method are the output of silhouette values per 'superpixel', and again they are compared with those at the reference set using the One-Class SVM method, with a mean outlying deviation number for each method produced (in percentage) at the end, always in comparison to the reference set. In addition, the silhouette value is a metric used to evaluate the quality of clustering in unsupervised learning. It measures how well each data point fits within its assigned cluster compared to other clusters. The coefficient ranges from -1 to 1 [22].

These methods in various papers have proven their value in their ability to detect outliers. One-Class SVM is a variant of the traditional SVM designed for one-class classification problems. The primary goal is to construct a decision boundary that encapsulates the normal data while isolating outliers. This is achieved by mapping the input data into a high-dimensional feature space using a kernel function, then separating them from the origin with maximal margin. This approach ensures the model's robustness against minor variations within the normal class while identifying anomalous instances that fall outside the boundary.

Several enhancements have been proposed to improve the robustness of One-Class SVM:

- Adaptive Weighted One-Class SVM: By assigning adaptive weights to training instances, this method reduces the influence of outliers in the training set, resulting in more robust anomaly detection (Yang et al. [23]).
- Minimum Within-Class Scatter One-Class SVM (OC-WCSSVM): This method minimizes the within-class scatter of the training data to construct tighter decision boundaries, improving the detection of subtle anomalies (An et al. [24]).

One-Class SVM has been widely applied across various domains for anomaly detection, including the following:

- Fault Detection: One-Class SVM is used to identify faults in systems such as industrial processes and mechanical systems. A robust version of One-Class SVM improves fault detection by removing suspected outliers from the training data before constructing the decision boundary (Xiao et al. [25]).
- Outlier Detection in Data Mining: This is particularly useful in mining datasets with limited labeled instances, as it relies solely on the normal class for training. Enhancements such as Selective Transfer Machine (STM) methodologies enable cross-domain outlier detection using unlabeled datasets (Fujita et al. [26]).
- Anomaly Detection in Sensor Networks: Applications such as wireless sensor networks leverage One-Class SVM to detect deviations caused by environmental or operational anomalies (Amer et al. [27]).

VAEs have proven to be highly effective in detecting outliers due to their ability to model complex data distributions and identify anomalies through reconstruction errors or deviations in latent space representations. Akrami et al. [28] introduced a robust VAE framework tailored for tabular datasets with mixed categorical and continuous features. The model demonstrated resilience against training data outliers caused by random noise or data poisoning, highlighting its utility in anomaly detection applications across network traffic datasets. Similarly, Li et al. [29] integrated a VAE with a genetic algorithm to enhance outlier detection in high-dimensional data. By analyzing subspaces for anomalies, the model achieved higher interpretability and accuracy compared to traditional outlier detection techniques. These studies exemplify the robustness and versatility of VAEs in handling diverse datasets and identifying outliers with high precision. The integration of novel loss functions, such as beta divergence, and hybrid methods, such as combining VAEs with genetic algorithms, highlights the evolving capabilities of VAEs for anomaly detection in various real-world applications.

The results from the aforementioned outlier detection methods, which, it must be stated, are applied in parallel, are formatted into a vector and turned into a percentage of difference/deviation from the referenced NOS points. We call each such resulting vector an Outlier Profiling Vector (OPV). Figure 2 describes the whole procedure.

The dimension of the OVP vector is 27, with OVP[0:1] from the pure Mahalanobis method, OVP[2] from the direct One-Class SVM method, OVP[3:21] from the VAE method, and OVP[22:26] from the embedding reduction and spectral clustering with silhouette values.

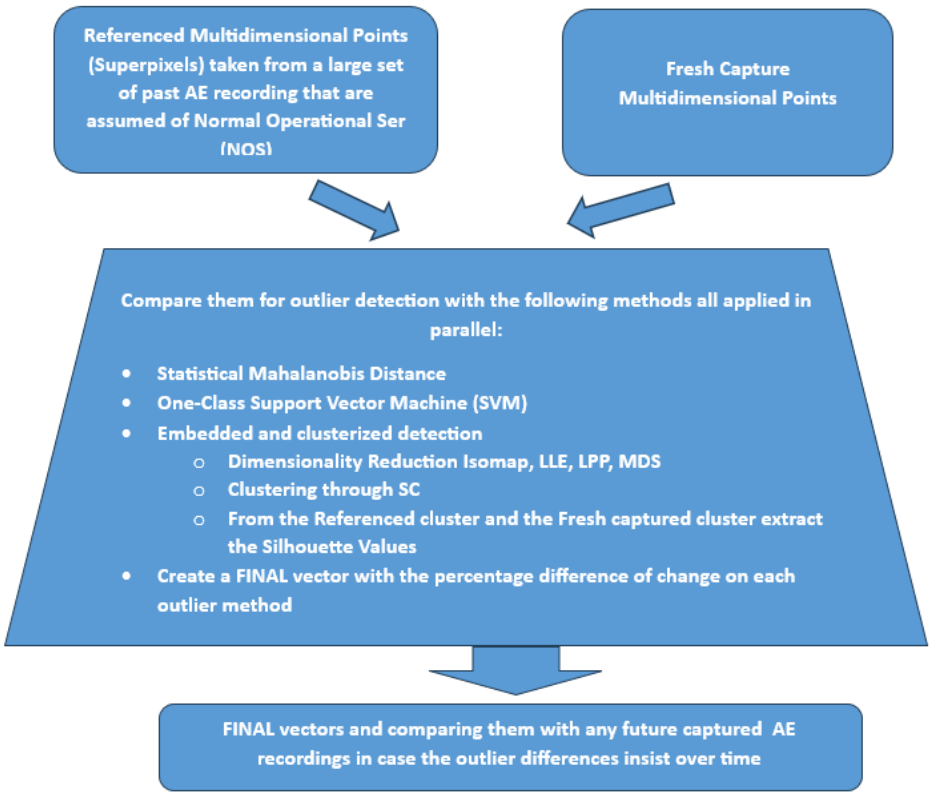

Figure 2. Starting from the referenced and inspected multidimensional space producing the OPV.

3. Our Experimentation and Procedure in Waterlock Gates

The system used for recording the AEs was the micro-SHM by Mistras. This is an industrial-grade recorder with a steel external casing and Wi-Fi connectivity. The system can operate on batteries for up to 3.5 h and features four channels for piezoelectric sensors. The corresponding recording software was installed on a laptop, communicating with the recording unit through WiFi. The piezoelectric sensors used are shown in Figure 3. The mounting of these sensors used a specific procedure for picking the same points on the gates of the various waterlocks which were assumed to be critical in the structural integrity of the gate, and before mounting the sensors, thoroughly cleaning the point and applying paste increased acoustic conductance between the sensors and the metal underneath.



Figure 3. PZT sensors used by the Mistras micro-SHM system, from top to bottom with 60 kHz (R6a), 150 kHz (R15a), and 400 kHz (wideband) useful bandwidth.

After calibration and experimentation, the used sensors for the 4 channels of the micro-SHM AE logger were the R6a with useful bandwidth of 10 kHz–80 kHz for channels 1 and 2 and the R15a with 100 kHz to 170 kHz at channels 2 and 3. When they arrived, the signals all passed from a median digital filter and a bandpass FIR filter over their useful bandwidths. The internal buffer depth of the Micro-SHM was set to 4096 bytes.

Figure 4 displays a setup of the AE logging system with four sensors and the placement of the sensors on the gate. The collection of data came from the waterlocks gates of Conca di Baricetta, Serafini (Italy), and VNF's waterlock gate outside Lyon. On each waterlock, two sets of recordings were taken. In the first set, the metallic gates opened and closed several times in a normal way, and this set was assumed to be the NOS (reference) recording set of the waterlocks' gate operation. In addition, a number of the gates' normal open–close operations were recorded in the second set, but at various points, and a Pencil Lead Break (PLB) test [30] was carried out as a straightforward method of simulating microcracks [31,32].

The third set of recordings was performed in the lab where on top of a 2 × 2 m metallic plate the sensors were placed and PLB tests were performed. As described in the Methodology section, from these recordings, preprocessing and mining codes for features were developed and applied in order finally from each recording set to produce a dataframe with reduced dimensionality and samples containing the important essence of each recording set, the feature vectors per slot of data. We call each row of that dataframe a 'superpixel'. These 'superpixels', which eventually are vectors of data in a hyper-dimensional space, were compared using various outlier techniques, as defined in Section 2.2 on outlier methods subsection, to find their deviation levels. Assuming the first set as the normal one, and the other two vectors' failure deviations from the normal exhibited as outliers, a vector expression of percentage deviation (meaning the comparison—using 100% scale—of a feature in the OVP vector as calculated using the AE reference set, in comparison with the corresponding feature value as calculated for the AE recording set under inspection) was created as a profile of the outlying deviation from the normal/reference NOS.



Figure 4. Placement of sensors (the used AE logger Micro-SHM in the top left corner) at a sample waterlock gate at Conda di Baricetta waterlock.

4. Results

Figure 5 is a sample of such clustering and silhouette values produced along with the dimension reduction prior to the clustering.

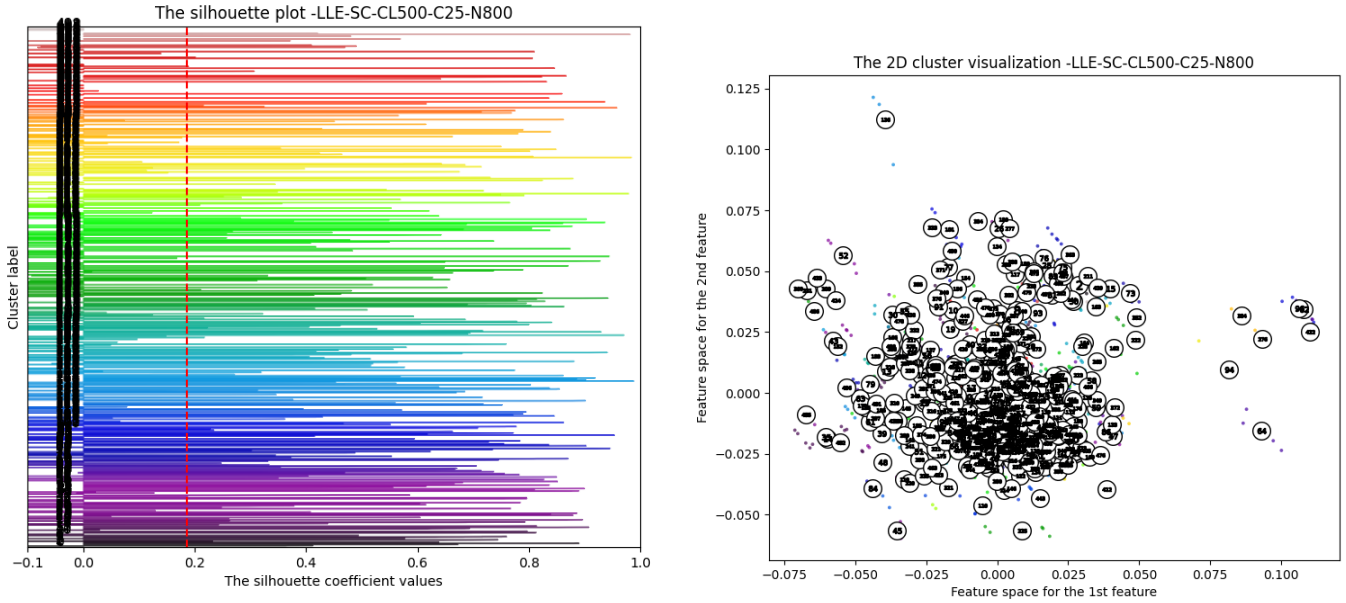

Figure 5. LLE dimension reduction: (**left**) silhouette values and (**right**) clustering (800 clusters) 2D projection, with x-axis the 1st feature (dimension) of the new reduced space due to LLE dimensionality reduction and y-axis the 2nd feature dimension.

Figure 6 displays the comparison of the OPVs, which are the changes/deviations over the referenced NOS set of 'superpixels' directly derived from the referenced AE recordings. The x-axis displays the coordinates of the OVP vector, with OVP[0:1] from the pure Mahalanobis method, OVP[2] from the direct One Class SVM method, OVP[3:21] from the VAE method, and OVP[22:26] from the embedding reduction and spectral clustering with silhouette values.

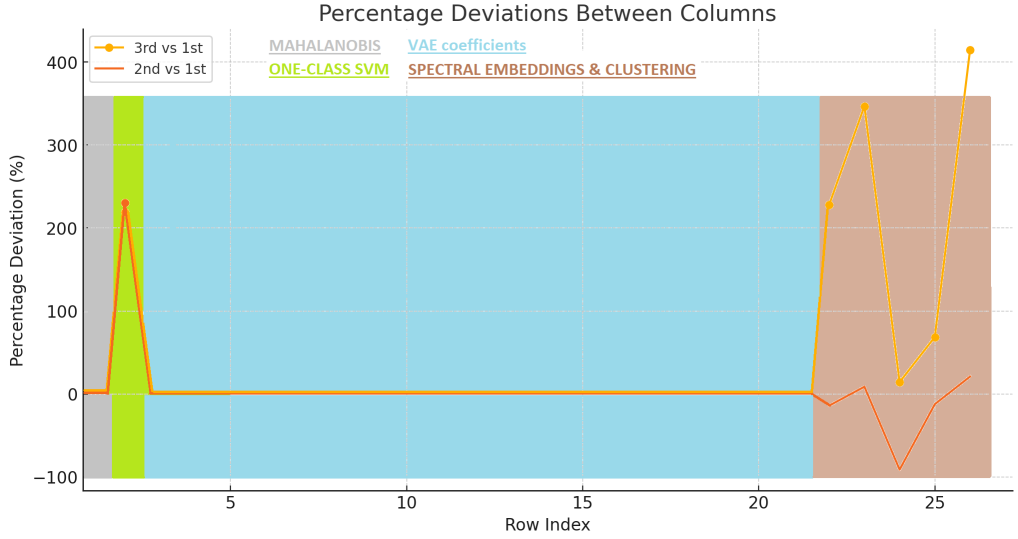

Figure 6. Deviation in the outlier profiles in comparison to the original gate's normal operation, with the horizontal line displaying the 27 indices as described in Section 2.2.

As can be seen, the direct One-Class SVM and the embeddings with the spectral clustering produced large deviations up to 400%, augmenting the contrast on outliers

greatly. It is observable that OVP is a unification of the latest techniques in outlier detection methods. All these methods are merely observation outlier methods from different points of view. It is also understandable that finding an outlier in a multidimensional space as produced from a physical procedure where numerous failing mechanisms are present (failing waterlock gate) is quite difficult as each mechanism can create different outlying orientations in the multidimensional space during system operation. This difficulty is the reason for creating a complex vector of outlier profiles (OVP), based on different methods, as each one has its limitations over the outlying direction. This fact is displayed at the OVP graph where OVP[3:21] produces small values. Of course, the reader must keep in mind that this is an initial approach and several future papers will follow with refinements and new models added.

5. Conclusions and Discussion

From the results over the specific AE recordings that we recorded, the direct One-Class SVM and the embedding reduction along with spectral clustering are more suitable, although further investigation on the reasons for that along with possible changes could be addressed in future studies. Nonetheless, these methods provide very good augmentation on enhancing the contrast in outlier detection. With this overall methodology, even small simulated microcracks can be spotted even if they are 'buried' inside large amounts of AE recordings. This fact gives the predictiveness of failure as a result of micro-failure accumulation. The methodology is generic, so other failing mechanisms except from metallic microcracks that can produce AEs can be detected, like problems in the gate's hydraulic system, foreign objects bouncing against the gate, and precipitation buildup at the bottom of the waterlock trench. The general and holistic approach we used aims at the ability to add and process a large number of sensors, not only AEs but also force cells, accelerometers, thermocouplings, images, and other. We aim at a universal approach to failure prediction using multi-sensor schemes over large to very large infrastructures.

The next step is to refine and verify the approach in other large structures with other different abnormal operations, under a multi-sensor scenario. Further refinement of the approach will include utilizing more deviation features, always aiming at predictive maintenance over very large infrastructures utilizing a sea of sensors.

Author Contributions: Conceptualization, T.T.; methodology, T.T.; software, T.T.; validation, T.T. and V.K.; formal analysis, T.T. and V.K.; investigation, T.T. and V.K.; resources, T.T. and V.K.; data curation, T.T. and V.K.; writing—original draft preparation, T.T. and V.K.; writing—review and editing, T.T. and V.K.; visualization, T.T. and V.K.; supervision, V.K.; project administration, T.T. and V.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 101021746.

Data Availability Statement: The datasets presented in this article are not readily available because the data are part of an ongoing study. Requests to access the datasets should be directed to authors.

Acknowledgments: This publication has been produced within the project CRISTAL (Critical Infrastructure Simulation of the Multi-Order Interconnected Layers).

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

References

1. Tsenis, T.; Spyrou, E.D.; Kappatos, V. An initial acousto-ultrasonic experimental investigation of defects in composite specimens used in the transportation domain. *Vibroeng. Procedia* **2020**, *34*, 66–71. [CrossRef]
2. Spyrou, E.D.; Tsenis, T.; Kappatos, V. Acousto-ultrasonic analysis of defects in composite specimens used in transportation domain. *J. Meas. Eng.* **2021**, *9*, 117–127. [CrossRef]

3. Kappatos, V.; Dermatas, E. Structural health monitoring of ship and other offshore structures using acoustic emission testing. In *Theory and Uses of Acoustic Emissions*; Nova Science Publishers: Hauppauge, NY, USA, 2012; pp. 95–106.
4. Kappatos, V.; Dermatas, E. Neural localization of acoustic emission sources in ship hulls. *J. Mar. Sci. Technol.* **2009**, *14*, 248–255. [CrossRef]
5. Kappatos, V.; Dermatas, E. Feature selection for robust localization of acoustic emission sources in ship-hull structures. *Res. Nondestruct. Eval.* **2010**, *21*, 127–140. [CrossRef]
6. Prakash, G.; Narasimhan, S.; Al-Hammoud, R. A two-phase model to predict the remaining useful life of corroded reinforced concrete beams. *J. Civ. Struct. Health Monit.* **2019**, *9*, 183–199. [CrossRef]
7. Hesser, D.F.; Mostafavi, S.; Kocur, G.K.; Markert, B. Identification of Acoustic Emission Sources for Structural Health Monitoring Applications Based on Convolutional Neural Networks and Deep Transfer Learning. *Neurocomputing* **2021**, *442*, 321–330. [CrossRef]
8. Abouhussien, A.A.; Hassan, A.A.A. Assessment of Crack Development in Engineered Cementitious Composites Based on Analysis of Acoustic Emissions. *J. Mater. Civ. Eng.* **2019**, *31*, 04019202. [CrossRef]
9. Haixiang, G.; Yijing, L.; Shang, J.; Mingyun, G.; Yuanyue, H.; Bing, G. Learning from class-imbalanced data: Review of methods and applications. *Expert Syst. Appl.* **2017**, *73*, 220–239. [CrossRef]
10. Kreiser, D.; Jia, S.X.; Han, J.J.; Dhanasekar, M. A nonlinear damage accumulation model for shakedown failure. *Int. J. Fatigue* **2007**, *29*, 1523–1530. [CrossRef]
11. Yuan, H.; Zhang, W.; Kim, J.; Liu, Y. A nonlinear grain-based fatigue damage model for civil infrastructure under variable amplitude loads. *Int. J. Fatigue* **2017**, *104*, 389–396. [CrossRef]
12. Pineau, A.; McDowell, D.L.; Busso, E.P.; Antolovich, S.D. Failure of metals II: Fatigue. *Acta Mater.* **2016**, *107*, 484–507. [CrossRef]
13. Hou, N.; Ding, N.; Qu, S.; Guo, W.; Guo, W.; Liu, L.; Xu, N.; Xu, N.; Tian, L.; Xu, H.; et al. Failure modes, mechanisms and causes of shafts in mechanical equipment. *Eng. Fail. Anal.* **2022**, *136*, 106216. [CrossRef]
14. Gul, M.; Catbas, F.N. Statistical pattern recognition for Structural Health Monitoring using time series modeling: Theory and experimental verifications. *Mech. Syst. Signal Process.* **2016**, *23*, 2192–2204. [CrossRef]
15. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [CrossRef]
16. Kedadouche, M.; Thomas, M.; Tahan, A. Empirical mode decomposition of acoustic emission for early detection of bearing defects. In *Advances in Condition Monitoring of Machinery in Non-Stationary Operations: Proceedings of the Third International Conference on Condition Monitoring of Machinery in Non-Stationary Operations CMMNO*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 367–377.
17. Aggelis, D.G.; Sause, M.G.; Packo, P.; Pullin, R.; Grigg, S.; Kek, T.; Lai, Y.K. Acoustic emission. In *Structural Health Monitoring Damage Detection Systems for Aerospace*; Springer Nature: Berlin, Germany, 2021; pp. 175–217. .
18. Cabana, E.; Lillo, R.E.; Laniado, H. Multivariate outlier detection based on a robust Mahalanobis distance with shrinkage estimators. *Stat. Pap.* **2021**, *62*, 1583–1609. [CrossRef]
19. Schölkopf, B.; Platt, J.C.; Shawe-Taylor, J.; Smola, A.J.; Williamson, R.C. Estimating the support of a high-dimensional distribution. *Neural Comput.* **2001**, *13*, 1443–1471. [CrossRef]
20. Eduardo, S.; Nazábal, A.; Williams, C.K.; Sutton, C. Robust variational autoencoders for outlier detection and repair of mixed-type data. In Proceedings of the International Conference on Artificial Intelligence and Statistics, Online, 26–28 August 2020; pp. 4056–4066.
21. Bengio, Y.; Delalleau, O.; Le Roux, N.; Paiement, J.F.; Vincent, P.; Ouimet, M. *Spectral Dimensionality Reduction*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 519–550.
22. Pavlopoulos, J.; Vardakas, G.; Likas, A. Revisiting Silhouette Aggregation. In *International Conference on Discovery Science*; Springer Nature: Cham, Switzerland, 2024; pp. 354–368.
23. Yang, J.; Deng, T.; Sui, R. An Adaptive Weighted One-Class SVM for Robust Outlier Detection. In Proceedings of the 2015 Chinese Intelligent Systems Conference: Volume 1, Yangzhou, China, 7 December 2015; Springer: Berlin/Heidelberg, Germany, 2016; pp. 475–484.
24. An, W.; Liang, M.; Liu, H. An improved one-class support vector machine classifier for outlier detection. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2015**, *229*, 580–588. [CrossRef]
25. Xiao, Y.; Wang, H.; Xu, W.; Zhou, J. Robust one-class SVM for fault detection. *Hemometrics Intell. Lab. Syst.* **2016**, *151*, 15–25. [CrossRef]
26. Fujita, H.; Matsukawa, T.; Suzuki, E. Detecting outliers with one-class selective transfer machine. *Knowl. Inf. Syst.* **2020**, *62*, 1781–1818. [CrossRef]
27. Amer, M.; Goldstein, M.; Abdennadher, S. Enhancing one-class support vector machines for unsupervised anomaly detection. In Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description, Chicago, IL, USA, 11 August 2013; pp. 8–15.

28. Akrami, H.; Aydore, S.; Leahy, R.M.; Joshi, A.A. Robust Variational Autoencoder for Tabular Data with Beta Divergence. *arXiv* **2020**, arXiv:2006.08204v2.
29. Li, J.; Zhang, J.; Wang, J.; Zhu, Y.; Bah, M.J.; Yang, G.; Gan, Y. Vaga: Towards accurate and interpretable outlier detection based on variational auto-encoder and genetic algorithm for high-dimensional data. In Proceedings of the 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 15–18 December 2021; pp. 5956–5958.
30. Hsu, N.N.; FR, B. Characterization and calibration of acoustic emission sensors. *Inist-CNRS* **1981**, *39*, 60–68.
31. Sause, M.G. *Investigation of Pencil-Lead Breaks as Acoustic Emission Sources*; University of Augsburg, Institute for Physics: Augsburg, Germany, 2011.
32. Prathuru, A.; Faisal, N.; Steel, J.; Jihan, S. Application of pencil lead break (PLB) point source in the detection of interfacial defects in adhesive bonds. *J. Nondestruct. Eval.* **2022**, *41*, 65. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.