Similarity-based anomaly score for fleet-based condition monitoring

Kilian Hendrickx[1], Wannes Meert[2], Bram Cornelis[3], Konstantinos Gryllias[4], and Jesse Davis[5]

[1,3] *Siemens Digital Industries Software, Interleuvenlaan 68, 3001 Leuven, Belgium*
kilian.hendrickx@siemens.com
bram.cornelis@siemens.com

[1,2,5] *KU Leuven, Department of Computer Science, Celestijnenlaan 200A box 2402, 3001 Leuven, Belgium*
kilian.hendrickx@cs.kuleuven.be
wannes.meert@cs.kuleuven.be
jesse.davis@cs.kuleuven.be

[4] *KU Leuven, Department of Mechanical Engineering, Celestijnenlaan 300, 3001 Leuven, Belgium*
konstantinos.gryllias@kuleuven.be

[4] *Dynamics of Mechanical and Mechatronic Systems, Flanders Make, Belgium*

ABSTRACT

Monitoring machines and early fault detection reduces production downtime, repair costs, and human casualties. Traditionally, machine condition monitoring requires performing a time-consuming manual analysis to find indicators for each potential fault. Moreover, these approaches often only focus on a single machine, while many industrial applications involve monitoring a fleet of machines. In such applications, it is often safe to assume that the majority of machines are in a healthy state. In comparable operating states, the behavior of these healthy machines is similar and any deviating machine is thus likely to be faulty. Previously, we proposed a fleet-based anomaly framework that can assess the health status of each machine in the fleet by detecting these deviations. It groups together similarly behaving machines and assumes that the healthy ones form the largest group and assigns an anomaly score to each machine based on the size of the group it belongs to. In this work, we propose a similarity-based anomaly score that offers multiple benefits over the cluster-based anomaly score. First, this score better represents the severity of a machine fault. Second, it allows to assess the health status of individual machines instead of machine groups. Finally, using similarities provides more nuanced insights in a machine's health status, especially for gradual degrading machines. Experiments show that the similarity-based anomaly score is superior to the cluster-based approach.

1. INTRODUCTION

Monitoring machine performance and reliability is important for both companies and society. Companies risk high repair costs and reduced downtime, while environmental pollution and human casualties are negative for society. Condition monitoring aims to detect such issues in an early phase. These approaches are now able to utilize data from an increasingly large number of monitored industrial assets in order to reduce these risks and costs by detecting unexpected machine failure at an early stage. Many of these monitoring approaches focus on assessing the health status of a single machine (Wong, Jack, & Nandi, 2006). Often, these approaches require historical data sets or handcrafted fault indicators which might be hard to obtain and can be dependent on operational parameters (International Organization for Standardization, 2009; Farrar, Sohn, & Worden, 2001; Randall & Antoni, 2011). Often, this procedure needs to be repeated when a novel fault type is considered. Supervised machine learning is often proposed to automate this procedure, but requires historical annotated data (Darraz et al., 2019; Stetco et al., 2019). These challenges can be overcome by employing an unsupervised anomaly detection approach that considers a fleet consisting of multiple similar operating machines (Monnin, Voisin, Leger, & Iung, 2013; Turrin, Subbiah, Leone, & Cristaldi, 2015). By assuming that the majority of the machine exhibits healthy behavior, deviating parameters or machine measurements (signatures) can indicate a machine fault without the need for historical data (Schmidt & Heyns, 2019). This can for example be applied in industrial applications such as wind farms (Siegel, 2013), production lines, and pairs of aerospace

engines (Jacobs, Edwards, Kadirkamanathan, & Mills, 2018).

In this work, we extend our previously proposed framework for fleet-based condition monitoring (Hendrickx et al., 2020). This framework addresses the challenges of both traditional condition monitoring and supervised machine learning approaches. It uses interpretable machine learning techniques to automatically evaluate assets within a fleet while incorporating domain knowledge if it is available. It consists of four building blocks. In the first block, the user defines a similarity measure to compare machines. This measure can be both data-driven and based on domain knowledge. The second block clusters the machines based on this similarity measure. Next, the third block assesses the health status of a machine by assigning each machine an anomaly score. The higher this score, the more deviating a machine's behavior is considered to be. Finally, each of these blocks is visualized in the fourth block to guide a domain expert to set up and gain trust in the framework.

The anomaly score proposed in our previous work was purely based on cluster sizes, which has three significant shortcomings. First, its value can change abruptly: a slight deviation can cause a machine's anomaly score to change from being very low to very high or vice versa. Second, the score does not accurately represent the anomalousness of a machine. A machine with the highest anomaly score is not necessarily the one that is exhibiting the most deviating behavior. Finally, the anomaly score is assigned to a group of machines with all machines in the group receiving the same score. Consequently, it is difficult to assess the health status of an individual machine making it difficult to provide insights into a specific machine's performance.

The contribution of this paper is a similarity-based anomaly score for fleet-based condition monitoring that addresses the aforementioned shortcomings. Instead of basing our anomaly score on the output of the clustering, we make use of the similarities among machines within the fleet. While cluster sizes are independent of how faulty a machine is, these similarities better represent a machine's health status. Moreover, they do not abruptly change in case of gradual degradation. Finally, using similarities enables assigning an individualized anomaly score to each machine.

This paper is organized as follows. In Section 2, we summarize the fleet condition monitoring framework of (Hendrickx et al., 2020). In Section 3, we propose our similarity-based anomaly score. Next, we illustrate the differences and advantages of this anomaly score in Section 4. Finally, Section 5 offers some general insights and conclusions.

2. SUMMARY FRAMEWORK

In (Hendrickx et al., 2020), we proposed a framework for fleet-based condition monitoring which detects machines that
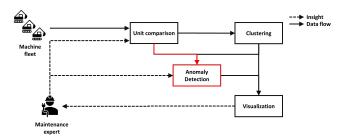

Figure 1. Overview of the fleet-based condition monitoring framework, consisting of 4 interacting building blocks. This paper's contribution is situated in the anomaly-detection block which is highlighted in red.

have deviating signatures from the fleet's general behavior and identifies them as faulty. This process makes the following assumptions: First, the majority of the machines are assumed to be healthy and to exhibit similar signatures. This allows detecting faulty machines if their signature deviates from the majority of the machines. If instead the majority of the machines in the fleet is faulty, our framework would detect that the healthy machines are anomalous. However, if this distribution is expected, a user could set up the framework to identify anomalies as healthy. In some use cases, this assumption might imply that we assume machines are operating in similar operational and environmental conditions. Otherwise, differences in signatures could be caused by differences in utilization instead of health status. However, thoughtful preprocessing still allows applying the framework in some of these cases. For example, if the fleet consists of subgroups of comparable machines, the fleet-based analysis can be performed on each of these groups separately. Additionally, one can exploit domain knowledge to remove the differences caused by operational conditions using techniques such as angular resampling (Lu, Wang, He, Liu, & Liu, 2016). Alternatively, one could identify relevant data samples in a short-term historical database containing recent measurements of other machines.

The framework is implemented in four interacting building blocks: defining a similarity metric between machines, clustering together similar machines in the fleet, performing anomaly detection to identify deviating machines, and visualizing the process to assist the user. Figure 1 shows the framework's setup, indicating the interactions with and between its blocks. We summarize each of these blocks below, a full description can be found in (Hendrickx et al., 2020).

2.1. Machine comparison

The framework expresses the difference of any machine pair (X, Y) through a user-defined and application-specific distance measure $s(X, Y)$[1]. A variety of measures is allowed,

---

[1]For our framework, a measure not satisfying the triangular inequality suffices. We opt to use the term distance for readability.

including those originating from machine learning as well as based on domain knowledge. This enables tailoring the framework towards specific use cases and incorporating domain knowledge if it is available. We demonstrated this in our previous work by considering multiple similarity metrics, ranging from data-driven pattern recognition to expert-based motor current signature analysis.

2.2. Clustering

A clustering algorithm uses the chosen distance measure to group together machines that exhibit similar behavior. The assumption that the machines are in a similar operational condition with a majority in a healthy state implies that they will form the largest cluster. Other clusters thus contain deviating machines, likely grouped by similar fault states. One key challenge in clustering is determining the number of expected clusters. In the condition monitoring setting, this cannot be set upfront. If all machines are in a healthy state, only a single cluster is expected. When one or more faulty machines are present, then two or more clusters would be expected. When one or more faulty machines are present, then multiple clusters would be expected as machines subject to different faults are likely to be clustered in different clusters.

We proposed to use hierarchical clustering for two reasons. First, it is relatively simple and can be visualized using a dendrogram, which is a tree-like structure that is easy to interpret and allows users to obtain better insights into the framework's predictions and the machines' health statuses. The dendrogram visualizes the distance between two subclusters by the height at which they are merged. Since a subcluster can consist of multiple elements, this similarity is defined by a *linkage* function. Popular choices are the minimum (single linkage) or maximum (complete linkage) distance between the subclusters' elements. This pairwise distance also influences the cophenetic distance $t(X, Y)$, the height at which two members X and Y are joined or, more formally, the distance between the two largest possible clusters containing X and Y separately (Sokal & Rohlf, 1962). Second, it does not require setting the number of desired clusters upfront, which alleviates this key clustering challenge. In our previous work, we proposed a partitioning strategy to obtain this number based on the structure of the clustering. This uses the correlation between the pairwise distances $s(X, Y)$ and cophenetic distances $t(X, Y)$: the cophenetic correlation to partition the hierarchical clustering. The higher this correlation, the better the clustering preserves the original pairwise distances (Lessig, 1972). Moreover, a high cophenetic correlation indicates the presence of multiple clusters, as intuitively explained by the example in Figure 2. If one or more faulty machines are present (blue dots), the larger distances between the healthy and faulty machines dominate the cophenetic correlation, which becomes close to 1 and only slightly decreases with an increasing number of faulty machines. Our
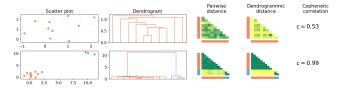

Figure 2. Upper: 10 data points representing a cluster of healthy machines, shown in a dendrogram obtained using single linkage. A low cophenetic correlation is obtained due to the clear difference between the pairwise and dendrogrammic distances, shown in the triangular matrices with a green (low) to yellow (high) color scale.
Lower: a cluster representing 3 faulty machines is added (blue). The cophenetic correlation drastically increases, as the difference in pairwise and dendrogrammic distances becomes relatively small.

partitioning procedure thus recursively partitions the hierarchical clustering such that the cophenetic correlation of each cluster partition is at least thr_{cc}.

2.3. Cluster-based anomaly score

A machine's anomaly score expresses how anomalous a machine is compared to the fleet. Healthy machines are expected to have a low anomaly score, while a faulty machine should have a high score. In our previous work, we proposed a machine's anomaly score to be the (normalized) size of its corresponding cluster. The main contribution of this paper, a similarity-based anomaly score, is presented in Section 3 and is shown to be superior to the cluster-based approach in Section 4.

2.4. Visualization

A high level of interpretability helps domain experts to set up the framework and allows the expert to have a deep understanding of the framework's predictions. This enables the expert to gain trust in automatic monitoring, to correctly set up the framework and to deeper analyze a specific deviating machine.

We previously proposed an interpretable visualization of the framework in our previous work. This visualization shows the machine's signatures, pairwise distances, hierarchical clustering, and the anomaly score of each machine. Figure 3 shows an example visualization where one could easily confirm the prediction of machines D1_2 and D2_10 being faulty. Note that this example makes use of the anomaly score described above.

3. SIMILARITY-BASED ANOMALY SCORES

An anomaly score based on similarities instead of cluster sizes offers several benefits. First, the number of machines in each cluster is independent of how deviating (and thus faulty) a machine is. For example, the cluster-based score decreases
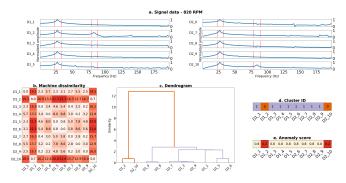
Figure 3. Visualization of the framework's four building blocks. In this example, machines D1_2 and D2_10 are faulty. This can be observed in their current signal frequency signatures (a), pairwise distances (b), the dendrogram (c) and anomaly score (d).

if more machines face the same fault. Second, when machine transitions to another cluster, its cluster-based score can abruptly and drastically change. Using similarities is particularly beneficial for edge cases that would transition due to only a slight change in the machine's health status. In such case, the similarity-based anomaly score changes more gradually. Third, the similarity-based anomaly score can be calculated on a machine-level. This individualized score provides a nuanced prediction and is beneficial to detect early degradation. For the cluster-based anomaly score, slight deviations would remain hidden until severe enough to transition the machine outside the majority cluster.

We propose two similarity-based anomaly scores, based on the pairwise distance $s(X, Y)$ or its approximating cophenetic distance $t(X, Y)$ which have three distinguishing properties:

1. The scores reflect the severity of a machine fault. When comparing two faulty machines, the score of the most degraded one should be the highest.

2. A significant change in anomaly score should only be caused by an underlying change in a machine's health status. Small variations due to noise should not cause a large difference in anomaly score.

3. The health status should be assessed for each machine individually. This is facilitated by anomaly scores calculated per machine.

3.1. Pairwise similarity-based anomaly score

Pairwise comparisons offer detailed insights into machine deviations within the fleet. In similar operating conditions, all healthy machines are expected to have similar signatures and low pairwise distance. However, a faulty machine's signature will be different, resulting in a high distance to healthy machines.

To assess the health status of a machine X, we compare this

with a representative healthy machine. Since most of these cover the majority of the fleet, we ensure this comparison by selecting the median of X's pairwise comparisons. Using non-robust metrics such as average or extrema would cause outliers to influence this comparison. Next, this value is normalized by the median distance within the largest cluster.

Formally, the anomaly score of a machine X in fleet F is defined in Equation 1, with P the set of all machine pairs within the largest cluster and $s(X, Y)$ the distance of a machine pair (X, Y) (block 1, machine similarity):

$$score(X) = \frac{median_{Y_i \in F \setminus X}[s(X, Y_i)]}{median_{(Y1_i, Y2_i) \in P}[s(Y1_i, Y2_i)]}. \quad (1)$$

3.2. Approximate distance-based anomaly score

The distances used in the previous anomaly score are suitable for most use cases of the fleet-based condition monitoring framework. However, one could easily replace $s(X, Y)$ in Equation 1 with another function. This could be beneficial when studying a large number of machines where calculating all pairwise distances may be computational too expensive, and approximating these pairwise distances is required.

One such strategy is to replace these pairwise distances by their cophenetic distances $t(X, Y)$, as these can be approximated more efficiently. Techniques such as HappieClust use heuristics to minimize the number of required pairwise comparisons in hierarchical clustering (Iam-On & Boongoen, 2013). Other alternatives to limit the number of pairwise comparisons are for example approximate nearest neighbor approaches. These use hashing techniques to group items likely to be similar (Cai, 2019). Instead of comparing every machine pair, one could limit the comparisons to the different groups.

3.3. Rescaling

Typically, an anomaly score is defined as a real value between 0 and 1, with 0 indicating completely normal behavior and 1 completely abnormal behavior. However, our proposed anomaly score can be larger than 1 as the median distance of a faulty machine can be larger than the normalization factor of the median distance within the largest cluster. Therefore, we use an exponential squashing function $score_{rescaled}(x)$ to rescale this anomaly score, as defined in Equation 2 (Vercruyssen et al., 2018).

$$score_{rescaled,\alpha}(x) = 2^{-\frac{score(x)^2}{\alpha^2}}. \quad (2)$$

In this equation, α corresponds to the value resulting in an anomaly score of 0.5: $score_{rescaled,\alpha}(\alpha) = 0.5$. In other words, a hypothetical machine obtains an anomaly score of 0.5 when deviating to the fleet α times the median deviation within the healthy cluster. The higher this parameter, the

more severe a deviation should be before resulting in a score of 1. However, lower values cause higher anomaly scores for limited deviations. In general, the optimal value of α depends on the exact use-case and can be used to fine-tune the sensitivity of the fleet-based condition monitoring framework.

4. EXPERIMENTS

We study the three main anomaly score properties for the cluster-based and similarity-based scores, using both artificial and experimental data sets. First, we use an artificial data set to show that the similarity-based anomaly score is representative of a machine's health status. Second, a similar data set is used to demonstrate the benefits of individualized anomaly scores. Finally, we study an experimental data set of electrical drive trains validating the similarity-based anomaly score in a gradually changing condition. In this data set, we introduced a voltage unbalance fault whose severity changes with the drive train's speed.

4.1. Property 1: Health status representation

An anomaly score can help an operator deciding on how urgent maintenance is. However, this requires the score to be related to the machine's health status: when comparing two machines, a higher anomaly score should indicate a more severe machine fault. In this scenario, we study this by evaluating a data set with different machine fault severities.

Dataset In this use-case, we study a data set simulating a fleet of 20 machines, consisting of 17 healthy machines, one having a minor fault (ID 16) and two having a severe but similar issue (ID $19-20$). A machine i's signature is represented by a two-dimensional data point $(D1_i, D2_i)$. The values of $D1$ and $D2$ are randomly drawn from a Gaussian distribution $R \sim \mathcal{N}(0, 0.05)$. Exceptions are the $D2$ values of the faulty machines, being set to $D2_{16} = 1,25; D2_{19} = D2_{20} = 2,5$.

Evaluation In this experiment, the distance $s(X, Y)$ of a machine pair (X, Y) is calculated by their data point's Euclidean distance. Next, these distances are used to cluster the fleet through hierarchical clustering with *complete* linkage. Finally, the hierarchical cluster is partitioned with thr_{cc} set to 0.9. Figure 4b visualizes this clustering in a dendrogram. The cluster partitioning results in three cluster partitions: a majority cluster of size 17 containing the healthy machines, a cluster of size 1 with machine 16, and a cluster of size 2 having machines 19 and 20.

The cluster-based anomaly scores, shown in Figure 4c, clearly do not represent a machine's health status. The score of machine 16 is higher than those of machines 19 & 20, while the former is only affected by a minor fault. The cluster-based anomaly score assumes a fault is more severe
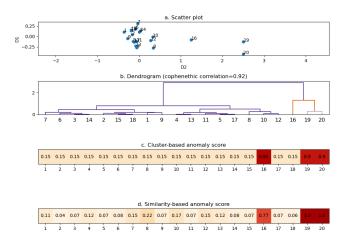

Figure 4. The cluster-based anomaly score assumes a fault is more severe if fewer machines are affected by it, which is invalid in many cases. Deviations, used by the similarity-based approach, are a better health indicator.

if it affects fewer machines, which is invalid in many cases.

We obtain the similarity-based anomaly scores in a two-step approach. First, we calculate the score according to Formula 1. In this case, machines $1-15$, 17 and 18 form the largest cluster P. The anomaly score of any machine X is thus the median distance from X to the aforementioned machines, divided by the median distance between any machine pair in this largest cluster P. Second, we rescale the obtained valued according to Formula 2, using $\alpha = 3$. This results in an anomaly score of 0.5 for a hypothetical machine, deviating to the fleet of three times the median deviation within the healthy cluster. The resulting anomaly scores, shown in Figure 4d, are representative for the machines' health statuses: minor faulty machine 16 is given a lower anomaly score compared to machines 19 and 20.

Using the similarity-based anomaly score thus helps a user interested in the true machine health status. Moreover, this user can obtain additional insights based on hierarchical clustering. The anomaly scores indicate a severe fault for machines 19 and 20, while the dendrogram also suggests their signatures are similar. An operator could thus infer that both machines suffer a similar fault.

4.2. Property 2: Individualized anomaly scores

Machine wear can cause both sudden failure and degradation. In the latter case, a machine's performance might decrease over time. Individualized scores thus allow an operator to assess each machine's health status and to detect degradation at an early stage. In this experiment, we generate a data set which the hierarchical clustering partitioning strategy considers as a single cluster. Since this results in an identical cluster-based anomaly score for all machines, this experiment highlights the benefits of individualized anomaly scores.
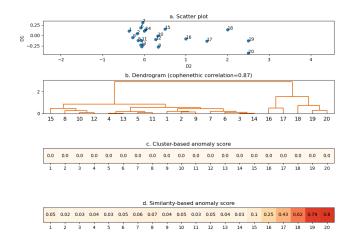
5

Figure 5. The similarity-based anomaly score assesses each machine's health status individually and detects gradual degradation. This is not possible using the cluster-based anomaly score, as all machines within a cluster receive the same score.

Dataset The data set used in this experiment is similar to the data set of Section 4.1; the same fleet of 20 machines is considered. However, gradual degradation is simulated for 6 machines (15-20) an increasing value for $D2 = 0.5*(14-id)$ with id being the machine's ID (Figure 5a).

Evaluation In this experiment, we perform complete link hierarchical clustering using Euclidean distance and partition the obtained clusterings using $thr_{cc} = 0.9$, as shown in Figure 5b.

While some separation between machines 1-14 and 15-20 exists, it is not substantial enough for the partitioning strategy to create multiple clusters. Hence, the cluster-based anomaly score for each machine is 0, as shown in Figure 5c. These scores do not allow an operator to differentiate between machine 1, supposedly in good health condition, and machine 19, which is clearly the most deviating. In contrast, the individualized similarity-based anomaly scores shown in Figure 5d show distinct values for each machine. Moreover, they show clear increases for the machines that are degrading, which yields a more nuanced view of each machine's health status.

Similarly to the previous experiment, the similarity-based anomaly scores are calculated using Formula 1. In this example, the largest cluster P contains all machines, including those that are faulty. Our anomaly score is robust to this, since it makes use of the median distances. As only a minority of the machines are faulty, using the median distance always results in a comparison with a healthy machine. Finally, we rescale the anomaly score by Formula 2 with $\alpha = 3$.

	SCIM	WRSM	DCM
Power	3	3	3
Phase-to-phase voltage (V)	400	400	270
Current (A)	6.2	4.3	11.2
Frequency (Hz)	50	50	-
Pole pairs	2	2	1
Speed	1385	1500	1500

Table 1. Rated parameters of the drivetrain's motors

Figure 6. The experimental fleet of ten drive trains with drive and load side motors connected by a rubber coupling.

4.3. Property 3: gradual changes

In this experiment, we study the anomaly scores' ability to observe gradual changing machine behavior. More specifically, we compare how well the anomaly scores represent a machine's health status while facing varying severeness of degradation.

Experimental setup We study current signals of a fleet of ten electrical three-phase 3 kW drive trains, measured at 25600 Hz. As shown in Figure 6, each drive train consists of two motors, connected by a flexible jaw coupling. One of these motors, a squirrel cage induction motor (SCIM), drives the shaft, the other motor acts as load and is connected to a resistor. For five of the drive trains, this load-side motor is a direct current motor (DCM), the others have wound rotor synchronous motors (WSRM). Their resulting load torque is proportional to speed and corresponding to the rated load at rates speed (Table 1). In our setup, each drive train is operated by an ABB drive controller with internal closed-loop direct-torque control. These controllers are linked such that all drive trains have identical speeds. This speed is measured by a tacho connected to one of the drive trains.

For this experiment, we used a 3 ohm external resistor R_{add} to emulate a voltage unbalance in drivetrain D2_10, inserted between the drive controller and drive-side motor. This is shown in the schematic of Figure 7. We found the severity of the voltage unbalance to be affected by the drivetrain's speed. The higher the speed, the less effect the voltage unbalance has due to the machine's inductance becoming more dominant.
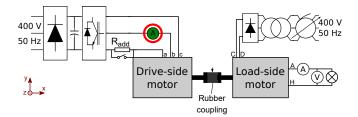
6

Figure 7. Schematic of the introduced voltage unbalance R_{add}. The analyzed current sensor (green) measures phase B.

Data set Using the experimental setup, we generate a data set of the drive trains performing a run-up from 500 to 1500 RPM. This increasing operational speeds causes the severity of the voltage to gradually decrease, making it an ideal use-case for this experiment. We downsample the current signals to 50 data points per period to remove signal noise and study the signal's waveforms. In our previous work, we found these to be indicative of a voltage unbalance. Next, the current signals are split in non-overlapping analysis windows of 0.5 seconds. Due to the small size, the RPM is relatively constant within each window.

Evaluation We analyze non-overlapping 0.5 seconds windows to assess the health status of the fleet over time. In each window, we pairwise compare the machines using a similarity measure derived from Dynamic Time Warping, which we found to be a good fault indicator in our previous work. The single linkage hierarchical clustering is partitioned with thr_{cc} set to 0.9.

Figures 8 and 9 show the relation between each drive train's speed and anomaly score for the cluster-based and similarity-based methods respectively. Both detect the voltage unbalance fault up until around 1250 RPM. However, the cluster-based score remains constant as long the fault can accurately be detected. When the voltage unbalance becomes less prominent, the score abruptly drops to 0. Moreover, minor fluctuations in the signal cause the anomaly score of healthy machines to spike at various speeds. Based on these scores, one could suspect a severe health issue for all machines. In contrast, the similarity-based score gradually decreases from 1 to 0.4 during the same period, aligning with the decreasing severity of the fault. Moreover, the variation of the healthy machines' anomaly scores is limited and offers better insights into the fleet's health conditions. For example, at low speed, one could observe a slight deviation for one particular machine. While we did not introduce a fault for this machine, the similarity-based score suggests it is slightly degraded.

The cluster-based anomaly score only allows a binary health status prediction, as it considers a machine as either healthy or faulty. This is a clear benefit of the similarity-based anomaly score since it makes more nuanced predictions. This is ex-
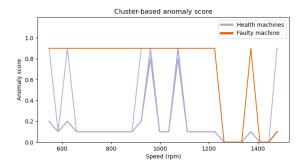

Figure 8. The cluster-based anomaly score has very abrupt transitions. Moreover, minor signal fluctuations cause spikes in the anomaly scores of healthy machines, as a machine is either considered to be healthy or faulty. It is impossible for an operator to assess the severity of a fault.
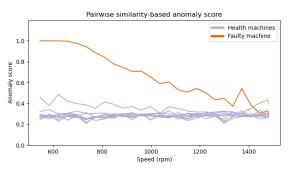

Figure 9. The similarity-based anomaly score gradually decreases as speed increases, following the decreasing severity of the fault. While variation within the healthy machines exists, none of those machines have high anomaly scores. An operator thus obtains a nuanced overview of each machine's health condition.

tremely powerful, as it allows an operator to estimate the severity of a machine failure and take appropriate action. For instance, the operator would not have to stop a production process for a minor failure, which could be very expensive. Moreover, insights into the fault severity for different operational settings help optimize production planning in such a case.

5. CONCLUSION

Our past work showed the benefits of fleet-based condition monitoring. In this approach we assigned an anomaly score to each machine based on the number of similarly behaving machines. In this work, we propose a superior anomaly score which makes use of similarities within the fleet to estimate a machine's health status. It offers multiple benefits to an end-user, by giving a more nuanced and gradually changing prediction of each machine's health condition separately. This allows an operator to estimate the severity of a machine fault, even in cases of early degradation. As the anomaly score reflects the severity of a machine fault, the operator can avoid

unnecessary production shutdowns and schedule predictive maintenance.

REFERENCES

Cai, D. (2019, dec). A Revisit of Hashing Algorithms for Approximate Nearest Neighbor Search. *IEEE Transactions on Knowledge and Data Engineering*, 1–1. doi: 10.1109/TKDE.2019.2953897

Darraz, A., Antoni, J., Mollon, P., Kirie, G., Larcher, A., Darraz, A., . . . Thomas, J.-h. (2019). Rotating machine diagnosis using acoustic imaging and artificial intelligence. In *Proceedings of survishno 2019.* Lyon, France.

Farrar, C. R. C., Sohn, H., & Worden, K. (2001). *Data Normalization : A Key For Structural Health Monitoring* (No. LA-UR-01-4).

Hendrickx, K., Meert, W., Mollet, Y., Gyselinck, J., Cornelis, B., Gryllias, K., & Davis, J. (2020, may). A general anomaly detection framework for fleet-based condition monitoring of machines. *Mechanical Systems and Signal Processing*, *139*. doi: 10.1016/j.ymssp.2019.106585

Iam-On, N., & Boongoen, T. (2013). Pairwise Similarity for Cluster Ensemble Problem: Link-Based and Approximate Approaches. In *Transactions on large-scale data- and knowledge-centered systems ix* (pp. 95–122).

International Organization for Standardization. (2009). *ISO 10816-3:2009 – Mechanical vibration – Evaluation of machine vibration by measurements on non-rotating parts – Part 3: Industrial machines with nominal power above 15 kW and nominal speeds between 120 r/min and 15 000 r/min when measured in situ.* Geneva, Switzerland.

Jacobs, W., Edwards, H., Kadirkamanathan, V., & Mills, A. R. (2018). Inter-engine variation analysis for health monitoring of aerospace gas turbine engines. *European PHM*.

Lessig, V. P. (1972, feb). Comparing Cluster Analyses with Cophenetic Correlation. *Journal of Marketing Research*, *9*(1), 82. doi: 10.2307/3149615

Lu, S., Wang, X., He, Q., Liu, F., & Liu, Y. (2016). Fault diagnosis of motor bearing with speed fluctuation via angular resampling of transient sound signals. *Journal of Sound and Vibration*, *385*, 16–32. doi: 10.1016/j.jsv.2016.09.012

Monnin, M., Voisin, A., Leger, J.-B., & Iung, B. (2013). Fleet-wide health management architecture. In *Annual conference of the prognostics and health management society.*

Randall, R. B., & Antoni, J. (2011, feb). Rolling element bearing diagnostics—A tutorial. *Mechanical Systems and Signal Processing*, *25*(2), 485–520. doi: 10.1016/j.ymssp.2010.07.017

Schmidt, S., & Heyns, P. S. (2019, apr). Localised gear anomaly detection without historical data for reference density estimation. *Mechanical Systems and Signal Processing*, *121*, 615–635. doi: 10.1016/j.ymssp.2018.11.051

Siegel, D. (2013). *Prognostics and Health Assessment of a Multi-Regime System using a Residual Clustering Health Monitoring Approach* (Unpublished doctoral dissertation). University of Cincinnati.

Sokal, R. R., & Rohlf, F. J. (1962, feb). THE COMPARISON OF DENDROGRAMS BY OBJECTIVE METHODS. *TAXON*, *11*(2), 33–40. doi: 10.2307/1217208

Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., . . . Nenadic, G. (2019). Machine learning methods for wind turbine condition monitoring: A review. *Renewable Energy*, *133*, 620–635. doi: 10.1016/j.renene.2018.10.047

Turrin, S., Subbiah, S., Leone, G., & Cristaldi, L. (2015, may). An algorithm for data-driven prognostics based on statistical analysis of condition monitoring data on a fleet level. In *2015 ieee international instrumentation and measurement technology conference (i2mtc) proceedings* (Vol. 2015-July, pp. 629–634). IEEE. doi: 10.1109/I2MTC.2015.7151341

Vercruyssen, V., Meert, W., Verbruggen, G., Maes, K., Baumer, R., & Davis, J. (2018, nov). Semi-Supervised Anomaly Detection with an Application to Water An-

alytics. In *2018 ieee international conference on data mining (icdm)* (Vol. 2018-Novem, pp. 527–536). IEEE. doi: 10.1109/ICDM.2018.00068

Wong, M., Jack, L., & Nandi, A. (2006, apr). Modified self-organising map for automated novelty detection applied to vibration signal monitoring. *Mechanical Systems and Signal Processing*, *20*(3), 593–610. doi: 10.1016/j.ymssp.2005.01.008