Fault Detection System for Bearings in Electric Motors using Variational Auto Encoders

A. Menéndez-González, L. Magadán, J. C. Granda, and F. J. Suárez

***Abstract*—Electric motors play a fundamental role in essential industries such as energy, transport and aeronautics, which require efficient maintenance to ensure productivity. Bearings are the most common failure point, making Prognostics and Health Management of this component crucial for Industry 4.0. This paper introduces a Fault Detection System based on Variational Auto Encoders (VAEs) trained exclusively on healthy vibration data from two public datasets. By analysing the resultant Gaussian distributions the system identifies early indicators of faults. This approach overcomes the common challenge of requiring faulty data for training, while also making it applicable to any other dataset. The study reveals an initial degradation stage in the training datasets, a critical oversight in previous studies, providing a more accurate depiction of bearing degradation profiles.**

Link to graphical and video abstracts, and to code: https://latamt.ieeer9.org/index.php/transactions/article/view/9554

***Index Terms*—Electric motors, Prognostics and Health Management, Bearing fault detection, Variational Auto Encoders.**

I. Introduction

MAINTENANCE has always been a major concern for industries. For a machine to be efficient it must meet certain requirements of safety and reliability, but with the emergence of new technologies reactive and preventive maintenance come short. Reactive maintenance can help understand faults, but it can not directly avoid them. Preventive maintenance relays on a schedule, so it cannot fully guarantee the safety between revisions. These policies are also quite expensive, representing between 15% and 60% of production costs [1].

On the other hand, predictive maintenance estimates when a fault will occur. It is based on machine monitoring to assist in determining the optimal moment to perform maintenance tasks. Predictive maintenance techniques usually involve four key steps: extraction of the relevant features from raw data; construction of a Health Index/Indicator (HI) to quantify the state of the machine; division of the machine lifecycle into stages; and prediction of the Remaining Useful Lifetime (RUL). This structured approach allows for extending the operational maintenance cycles and increasing the productivity of the machine by approximately 55%, making predictive maintenance a cornerstone of Industry 4.0 [2], [3].

However, deciding which machine element to focus on is a complex task. Due to digitalisation, electro-mechanical systems have been on the rise, of which 40% contain rotatory components. Among the various faults that can occur in electric motors, around 50% are caused by bearings, followed by the stator and rotor, which account for 30% and 10% respectively. The remaining 10% are due to other factors, such as loosening or gear faults [4].

Traditionally, bearing faults have been diagnosed using data such as vibrations, using the Root Mean Square as the indicator about the machine's operating conditions [5]. With the raise of Industry 4.0, new standards [6] and techniques are constantly being developed, specially in the Deep Learning field [7], [8]. These new models follow a black box approach, sacrificing explainability in search of competitive results, which is unacceptable in industries with tight safety and control measures. Thus, new efforts aim to cover this gap by creating systems that provide verifiable results with interpretable techniques.

Recent studies have explored various innovative approaches for bearing fault detection and RUL prediction. Magadán et al. [9] introduced a Monotonic Smoothed Stacked AutoEncoder (MS2AE) to infer a health index from raw bearing data, providing explainability through correlation matrices and Dynamic Time Warping. Wang et al. [10] proposed a Multiple-Feature Fusion Health Indicator (MFF-HI) and Weighted Temporal Convolution Network (WTCN) for accurate RUL prediction. Guo et al. [11] developed a method combining impulse-driven measures with an improved morphological filter for robust fault detection. Yan et al. [12] enhanced health indicators using a HI-signal-to-noise ratio (HI-SNR) and sum of weighted spectrum amplitudes (SOWSA) for machine performance degradation assessment. Zhou et al. [13] utilized a discrete entropy-based health indicator and LSTM model for bearing health forecasting. Wei et al. [14] proposed a framework with multi-scale attention-based dilated causal convolution (MADCC) and multi-layer temporal convolutional network (MTCN) for RUL prediction. Deng et al. [15] introduced a calibration-based hybrid transfer learning framework to improve data fidelity and model generality. Finally, Li et al. [16] used cumulative summation features (CSF) and discrete wavelet transform (DWT) for online anomaly detection and RUL prediction.

This work presents a novel Fault Detection System for bearings that uses Deep Learning techniques to train models

The associate editor coordinating the review of this manuscript and approving it for publication was Guillermo Valencia-Palomo (*Corresponding author: Alonso Menéndez-González*).

This research was partially funded by the Spanish National Plan of Research, Development, and Innovation under project EDNA (PID2021-124383OB-100), the European Union and the University of Oviedo.

A. Menéndez-González, L. Magadán, J. C. Granda, and F. J. Suárez are with the University of Oviedo (e-mails: alonsomenendezgonzalez@gmail.com, magadanluis@uniovi.es, jcgranda@uniovi.es, and fjsuarez@uniovi.es).

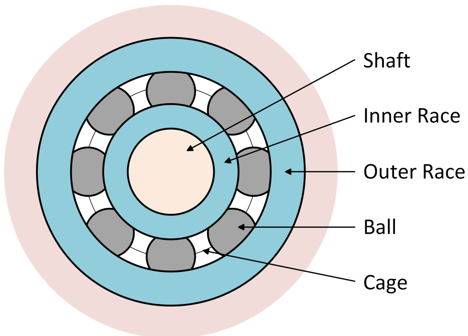

Fig. 1. Ball bearing scheme.

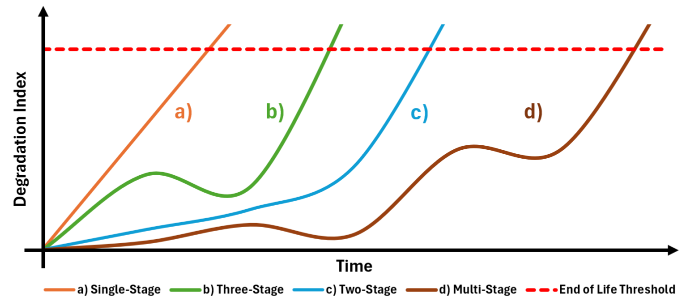

Fig. 2. Common degradation index profiles.

on limited data. The system is trained exclusively on healthy vibration data, making it applicable to any dataset and suitable for newly commissioned equipment. Vibration samples are summarised into Gaussian distributions to extract features and obtain the Health Index. This enables the application of various statistical tests and properties to assess the machine's health status, automating the division of the machine lifecycle. This novel approach overcomes the common challenge of requiring faulty data for training, enhancing the reliability and applicability of the system in real industrial settings.

The rest of the paper is structured as follows. Section II describes the problem at hand, contextualising the predictive maintenance paradigm for bearings. The proposed fault detection system, along with the statistical techniques to be used are described in Section III. Section IV explains the results obtained from applying the solution to two public datasets and compares them to similar works in the literature. Finally, the main conclusions drawn are outlined in Section V, along with future work.

II. Problem Statement

Prognostics and Health Management is a mature paradigm. Over time, new degradation indicators of machines have been identified, but developing a comprehensive measure of machine health is challenging due to the variety of models and operating conditions of these machines. Consequently, efforts have been made to create adaptable methods that can be applied across different machines to proactively avoid breakdowns. Bearings, which are one of the primary causes of faults in electric motors, and well known datasets are used in this work.

A. Bearings

Bearings are machine elements that limit motion and reduce friction between moving parts. In the case of ball bearings, the separation between the inner and outer race is done through balls, or roller elements, as depicted in Fig. 1. Overall, the races and the roller elements are the most common origins of faults.

Although faults are preceded by a degradation stage, there are no visible signs of degradation until the machine breaks down, requiring indirect evidence such as vibration signals, noise, oil debris, temperature or electrical current to detect the degradation [17].

Typically, after data is gathered, it is fused in order to construct a significant indicator that determines the degradation status of the machine. Due to its interpretability, the most common approach is to build a Health Index or Health Indicator, which is a number between 0 and 1 where the former indicates the End of Life (EoL) of the component. Alternatively, a Degradation Index can be used, ranging from healthy (0) to a threshold value indicating the EoL. The most common patterns of degradation, as illustrated in Fig. 2, are the following:

a **Single-stage**: Linear or exponential trend provoked by a continuous and irreversible degradation.

b **Three-stage**: Pattern usually seen in bearings, where manufacturing defects provoke impacts, until they are smoothed and subsequently exhibit characteristics of a healthier stage. Ultimately, the continuous operation generates new defects and also continuous degradation.

c **Two-stage**: Made up of a healthy stage that may show a slow degradation, followed by an unhealthy stage where the heavier degradation process takes place. Detecting this tendency change is the key to predict the RUL of the component.

d **Multi-stage**: In complex systems such as wind turbine generators [18], the degradation process may undergo multiple stages, provoked by different, concurrent faults, causing an alternating degradation.

B. Datasets

To create the best possible Health Index, it is important to have as much high-quality data as possible. However, there are two main reasons why few datasets are available. First, real run-to-failure data is scarce as breakdowns entail undesirable costs and delays. Second, procedures to fully emulate realistic operating conditions are limited, making it difficult to compare experiments with induced faults to real machines. Nonetheless, there exist datasets obtained from experiments that approximate realistic conditions and have been widely used in the literature, such as the IMS [19] and XJTU-SY [20] datasets.

The IMS dataset consists of three run-to-failure experiments, each involving four bearings installed on a shaft powered by an electric motor. These bearings endure a radial load of 26.7 kN while rotating at a constant speed of 2000 revolutions per minute. Vibration data is collected from each experiment using accelerometers at a frequency of 20 480 Hz and captured in 1-second snapshots every 10 minutes. Additionally, the type of fault that occurred is recorded.

The XJTU-SY dataset utilises a smaller testbed consisting of an electric motor, a pair of bearings, and support elements that allow for load adjustments on the tested bearing. Vibration data is collected at a sampling rate of 25 600 Hz using two accelerometers, with each 1-second sample separated by 1 minute. A total of 15 experiments are conducted under three different operating conditions: the first condition uses a shaft frequency of 35 Hz and a 12 kN load; the second condition increases the shaft frequency to 37.5 Hz while reducing the load to 10 kN; and the third condition operates at 40 Hz with a 10 kN load. The XJTU-SY experiments are generally quite short, especially in comparison with the IMS dataset. Only four of the 15 experiments, specifically 2-1, 2-3, 3-1 and 3-4, last longer than eight hours and are thus more generally considered as realistic.

C. Methods

While historically, white-box approaches were followed to obtain the Health Index due to the limited computing power and data available, recent efforts are progressing by using Deep Learning techniques. Generally, these approaches have prioritized robust results over explainability and the concept of HI. However, due to the reliability and safety requirements of the systems being maintained, there is a renewed focus on developing explainable and verifiable procedures for Prognostics and Health Management.

One example is by Magadán et al. [9], a modularised Fault Detection System containing a Monotonic Smoothed Stacked Auto Encoder, whose latent space represents the Health Index and is only trained on healthy samples. By first using a Q-Q diagram to verify the normality of the HIs, the 95 percentile of the healthy HIs is computed to obtain the Health Stage division, which is a threshold representing the change from a healthy state to a faulty state.

The Fault Detection System presented by Chauhan et al. [21] also uses the Health Index approach, but with a novel indicator measuring the sparsity impact. This system utilises single-valued neutrosophic cross-entropy along with the artificial hummingbird algorithm for feature mode decomposition, achieving results that have been proven through case studies from industries.

Another example is Liang et al. [8], which presents a Fault Diagnosis System named IVTN-SA that aims to overcome the difficulties of environments with different operating conditions. This system is made up of a Subdomain Adaptation module and an Improved Vision Transformer Network which uses Deformable Convolutions to perform feature extraction over vibration signals, thus fusing global and local information to perform the diagnosis.

Finally, a different approach is presented by Xu et al. [7], whose objective is to create a reliable fault identification framework based on multi-sensor data, achieved with a Collaborative Fusion Convolutional Neural Network. This system is made up of different CNNs extracting features from different mechanical signals, whose correlations are explored by means of a Central Fusion Module.

As noted, not all of the aforementioned Fault Detection Systems explicitly use a Health Indicator. However, to effectively manage the various available data, features must still be combined and fused to make accurate predictions. It is essential to discover new procedures that can enhance the quality of the Health Index while being applicable to any operating condition and ensuring explainability, as this is crucial for optimising the maintenance of electric motors.

III. Proposed Solution

Overall, to answer the needs of electric motors maintenance a Fault Detection System should meet three requirements:

1) **Flexible:** Applicable to different operating conditions.
2) **Efficient:** Can function even in scenarios where data is scarce.
3) **Explainable:** Does not follow a black box approach, providing verifiable results achieved by explainable techniques.

Thus, this work proposes a Fault Detection System that meets these requirements while aiming to improve the quality of the HI. The proposed system is described in detail below, along with the method used to obtain the HI. Subsequently, the techniques employed to assess the degradation status based on the HI will be discussed.

A. Fault Detection System

The system presented is a procedure that can be applied to any vibration dataset, made up of four interconnected components as depicted in Fig. 3. For future comparison with other approaches in the literature, the IMS dataset and XJTU-SY 2-1, 2-3, 3-1 and 3-4 datasets have been selected, which are considered more realistic due to their longer durations. The details of each component are presented below:

1) **Vibration data:** For each dataset, a set of samples is considered as healthy. In the case of the selected datasets, the first 300 samples are used by similar works, which were checked to be normal by using a Q-Q diagram. The remaining samples are considered as unlabelled, due to their degradation state being unknown. Since no other preprocessing is performed on the data, this straightforward approach ensures that the system can be easily applied to various datasets, enhancing its flexibility.
2) **Health Index Construction:** A Variational Auto Encoder is trained only on Healthy samples, learning how to represent each sample by a Gaussian distribution. Afterwards, it is used to predict the means and variances of all samples.
3) **Healthy Distribution Construction:** Once all healthy distributions are generated, a single General Healthy Distribution (GHD) is obtained from them, or alternatively, an indicator to be used on the next step. The techniques to be applied will be explained in the following subsection.

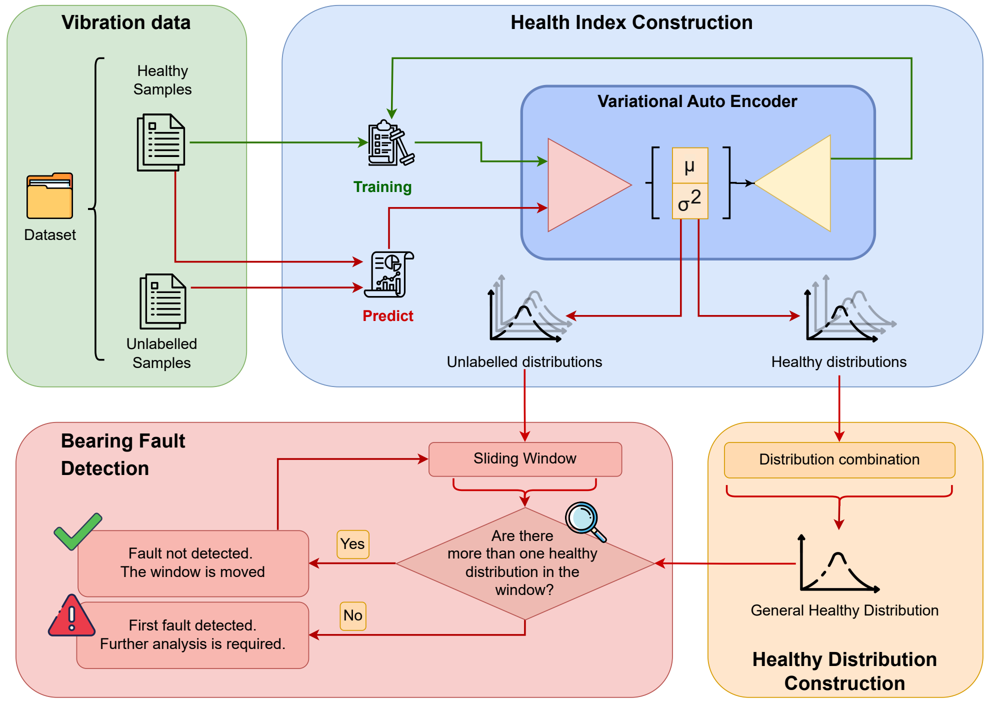

Fig. 3. Schema of the VAE-Based fault detection system.

4) **Bearing Fault Detection:** Unlabelled distributions, whose degradation state is unknown, are analysed by using a sliding window of size five, comparing them with the information gathered in the previous step. If at least four of them are detected as unhealthy, then a new fault is predicted.

It must be noted that a new VAE is trained per dataset, as each machine is different, as well as their operating conditions. The VAE's objective is to capture as much information in two values: Mean and Variance of a Gaussian distribution, which will act as the Health Index for that sample. For both IMS and XJTU-SY, each sample is of size 20 480, so summarising them in a size 2 latent space will inevitably carry an information loss. However, this is not as counterproductive as it may seem, as the aim is to detect when a sample pertains to a healthy status or a faulty status. The Hyper-parameters and architecture used for the VAE, which were empirically obtained, are shown in Table I. Notably, by considering only healthy samples for training, this approach and parameters remain highly applicable in real-world settings where obtaining faulty samples is a major challenge, thus enhancing the technique's usability and flexibility across different datasets.

While the proposed detection scheme can detect the presence of a fault, it cannot specify which bearing is faulty when multiple bearings are monitored. This limitation arises because vibrations from a failing bearing can propagate through the machine's structure, affecting readings from other accelerometers, especially when bearings are located on the same axis. Additionally, real industrial environments are filled with various sources of noise that can also affect the readings, making it challenging to isolate the exact source of vibrations. To address this, further diagnostic steps can be taken after fault detection, such as using additional sensors placed closer to each bearing to better isolate the source of unwanted vibrations. Despite this limitation, the system's early fault detection capability remains valuable, as it allows for timely maintenance actions, reducing the risk of severe failures and associated costs.

TABLE I
VARIATIONAL AUTO ENCODER HYPER-PARAMETERS

Architecture	IS-3500-700-200-10-2-10-200-700-3500-IS
Activation	ReLu
Learning Rate	1e-5
Epochs	5 for IMS-1, 100 for IMS-2 and IMS-3, 50 for XJTU-SY
Batch Size	8 (2 for IMS-1)

B. Health Stage Division Techniques

The encoder of the VAE summarizes each sample in two values that act as the HI: mean and variance of a Gaussian distribution. As it can be noted, this HI does not follow the definition given in Section II, which means that it must be interpreted in order to understand the degradation status of the bearing.

Nonetheless, due to the HI being a normal distribution, several methodologies and properties can be used to compare the healthy and the unlabelled distributions, which are derived from encoding healthy and unlabelled samples. In this subsection five techniques will be tested, which can be classified into two different approaches.

The first approach will compute a General Healthy Distribution (GHD). For that, it is assumed that the output (300 means and variances) follows a normal distribution, computing its average. This GHD will be compared to the unlabelled distributions with the following techniques:

- **Kolmogorov-Smirnov Test:** Also known as KS Test, measures the fitness between two distributions, obtaining a p-value. This allows for a direct comparison between the GHD and any unlabelled distribution, by means of a 95% confidence level.
- **Kullback-Leibler Divergence:** It measures the distance between two distributions, where values close to 0 indicate a high degree of similarity.The comparisons will always be done using the GHD as the reference, obtaining a threshold value by first computing the distance between the GHD and healthy distributions.

The second approach is more direct, not using the GHD and focusing on finding a threshold from the healthy distributions:

- **Kullback-Leibler Divergence over Prior:** The VAE internally employs the KL divergence in its loss function, assuming that the distributions have a mean of 0 and a variance of 1. If the samples encountered during testing significantly differ from those observed during training, the KL divergence of the resulting distribution from the prior will approach zero. Leveraging this fact, a threshold can be established.
- **Entropy:** This metric quantifies the uncertainty of a distribution. It can be calculated using only the variance and is applicable even in higher-dimensional latent spaces which would result in Multivariate Normal Distributions. Distributions with high entropy can be deemed unhealthy based on a predetermined threshold.
- **Variance:** This value is provided by the VAE encoder directly. Similar to entropy, if a sample is significantly different from those encountered during training, the variance will increase. By first analysing this metric for healthy distributions a threshold can be computed.

Most of the techniques explained require a threshold value to classify unlabelled distributions. This is done by first computing the metrics for healthy distributions and then obtaining the 95th percentile. The threshold determination process was ultimately empirical due to the limited number of realistic datasets available. This limitation arises because real run-to-failure data is scarce, and procedures to fully emulate realistic operating conditions are challenging. However, this empirical approach allows for flexibility and adaptability across different datasets. With more datasets, a more robust threshold could be developed, enhancing the system's reliability.

In industrial environments, early detection is crucial because the cost of a false positive is minimal compared to the cost of a developed fault or a false negative. Therefore, the threshold must balance robustness and sensitivity: if the threshold is too high, the system may fail to detect early signs of faults; if it is too low, it may result in frequent false alarms. This approach aims to find a compromise that ensures timely fault detection without overwhelming the system with false positives. Specifically, the system uses a sliding window of size five, predicting a fault if at least four out of five consecutive samples exceed the threshold. This method provides a practical balance, ensuring both sensitivity to early faults and robustness against false alarms.

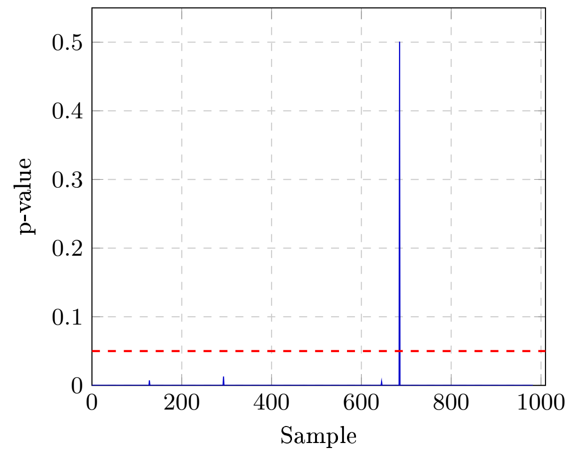

Fig. 4. KS Test for IMS-2

IV. Results

This Section will show the most interesting results. First, the outcomes per each of the five techniques will be described and explained. Subsequently, a comparative with other similar approaches in the literature will be made in order to verify the results.

A. *KS Test*

The Kolmogorov-Smirnov Test is the technique used to compare the GHD with the unlabelled distributions. Results show that this technique is not sensible enough, as samples that are close to the failure point do not reach the 95% confidence level that was set as a threshold, so it would seem that this technique is not adequate for this Fault Detection System. An example of the outcomes that were achieved through this technique is illustrated in Fig 4, where sample #684 exceeds the threshold, but the subsequent samples do not, thereby failing to meet the conditions set by the sliding window. Nonetheless, removing the sliding window altogether does not yield better results, as for most of the datasets the threshold is not reached by any sample.

B. *KL Divergence: GHD*

The first approach for the KL Distance uses the GHD as a reference to classify unlabelled distributions. The outcomes for most of the datasets do not yield the expected results, caused by the system failing to detect a significant difference between healthy and faulty distributions. However, the case of IMS-1 perfectly exemplifies the expected behaviour of this system. Fig. 5 shows three different stages. The green stage corresponds to the first 300 healthy distributions. The yellow area corresponds to the unlabelled distributions contrasted against the threshold using the sliding window to discard noise. The read area starts with a great change at sample #2004 which triggers the alarm.

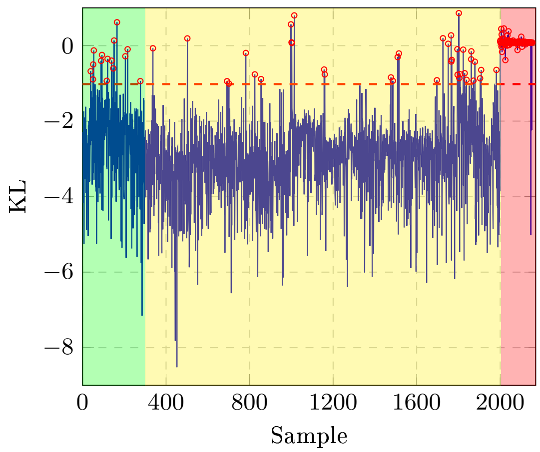

Fig. 5. KL divergence for IMS-1 using GHD.

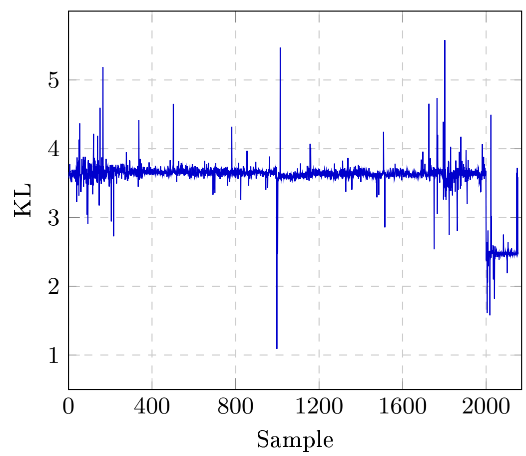

Fig. 6. KL divergence for IMS-1 second approach.

C. KL Divergence: Over Prior

The second approach for the KL Distance leverages the fact that unknown distributions will be close to the $\mathcal{N}(0,1)$ distribution. As the system approaches the fault state, the encoded distributions become more similar to this normal distribution. Consequently, the KL distance, which measures the divergence between the encoded distribution and $\mathcal{N}(0,1)$, should decrease as the fault nears. Therefore, a lower limit can be set as a threshold to detect the proximity of a fault.

This is exactly what can be observed in Fig. 6, where the tendency change at sample #2000 of IMS-1 dataset occurs in the opposite way as seen in Fig. 5, indicating a heavier degradation process. However, results are inconclusive for other datasets, as the KL shows unexpected behaviours, such as decreasing on healthy samples and increasing again on samples close to the end of the experiment. The remaining two techniques will be able to shed light to these contradictions.

D. Entropy

This property of normal distributions, which measures uncertainty, is expected to increase as the machine approaches a fault. This is because the healthy samples seen during training

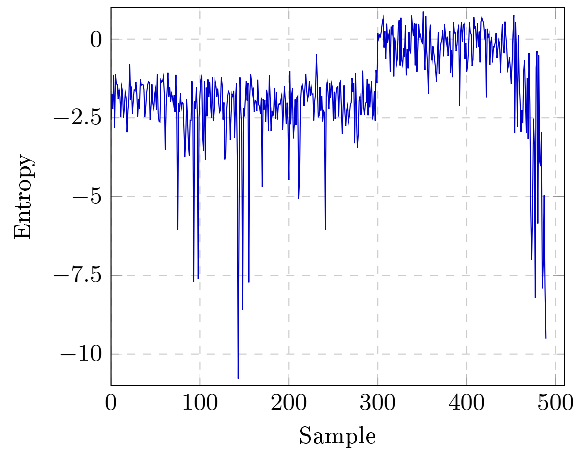

Fig. 7. Entropy for XJTU-SY 2-1

and the unlabelled samples nearing the end of the experiment are significantly different.

However, the results show the opposite trend. For instance, in the case of XJTU-SY 2-1, illustrated in Fig 7, there is a slight increase at sample #300 due to overfitting, followed by an abrupt decrease towards the final phase of the experiment. This descent occurs because the VAE identified unlabelled samples similar to those seen during training, implying that the training set included samples that were not healthy.

To understand how faulty samples could be present at the start of the experiment, it is essential to recall the common degradation profiles of bearings shown in Fig. 2. These results indicate early-stage degradation due to the smoothing of manufacturing defects, a behavior observed in bearings that follow a three-stage profile. Fig. 7 further supports this by showing an abnormal entropy decrease around sample #142, which is indicative of the initial degradation phase.

This finding challenges the prevailing assumptions in the literature, which typically consider the first 300 samples to be healthy. However, results indicate that these initial samples include some that are not healthy and belong to an initial degradation phase. This suggests a three-stage degradation profile, with early-stage degradation occurring within the first 300 samples. This new understanding highlights a critical oversight in previous studies and provides a more accurate depiction of bearing degradation profiles.

E. Variance

While this value is provided directly by the encoder and should behave similarly to the entropy, it yields more consistent results, being able to find a fault point for each dataset. Despite this fact, the detection is done quite early for most of them, which could be an indicative of overfitting, so in the next subsection the results will be contrasted against important works of the literature. Fig. 8 shows the results in IMS-1 once again, where the degradation status change can be seen at sample #2004, being generally a less noisy representation compared to the other techniques that allows for a clean threshold.

TABLE II
SUMMARY OF THE RESULTS

Dataset	IMS 1	IMS 2	IMS 3	XJTU 2-1	XJTU 2-3	XJTU 3-1	XJTU 3-4
VAE + KS Test (No window)	-	684	774	-	-	-	-
VAE + KL Divergence (GHD)	2004	-	-	-	-	-	-
VAE + Variance	2004	335	310	303	321	303	307
MS2AE + Dynamic Time Warping [9]	1875	536	5967	451	301	2347	1416
MFF-HI + WTCN [10]	-	934	-	-	-	2348	1418
Impulse-Driven measures + Morphological Filter [11]	1917	632	-	458	-	2407	1446
HI-SNR + SOWSA [12]	1833	533	5952	-	-	-	-
Discrete Entropy-Based HI + LSTM [13]	1971	530	-	-	-	-	-
MADCC + MTCN [14]	-	-	-	455	327	2344	1418
Calibration-Based Hybrid Transfer Learning Network [15]	-	-	-	452	302	2346	1417
CSF + DWT [16]	-	-	-	450	313	2376	1430

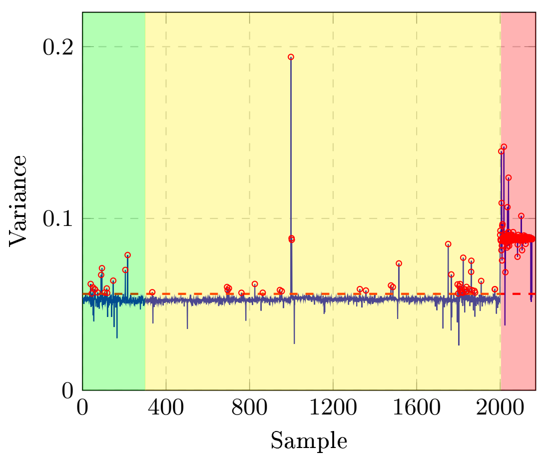

Fig. 8. Variance for IMS-1

F. Compared Results

The datasets used are not labelled, with the only known detail being that the last sample corresponds to the fault itself. Therefore, the best way to verify the results is by comparing them with those from significant works in related literature. Table II presents a summary of the results obtained with the presented Fault Detection System and those found in the literature. This table indicates, for each dataset, the sample number at which a fault is detected to occur. While early detection is the objective, detecting too early may be deemed as unreliable. The following remarks can be extracted for each dataset:

- **IMS-1:** The results are close to those of other works. Even if the detection is done later, graphs like Fig. 8 show variance disturbances at around sample #1800 corresponding to other works' predictions. A further analysis is required to determine whether the samples correspond to an incipient fault or, on the other hand, they are just noise.
- **IMS-2:** The variance leads to a very early detection, probably due to overfitting, but at the same time, closer inspection of its graph does not indicate any sign for the fault to start at sample #530, as other works suggest, while also revealing a peak that relates to the detection done by [10].
- **IMS-3**: While the prediction using the variance seems to be affected by overfitting, there were no signs that indicate that there would be faulty samples at #5900, as other works suggest.
- **XJTU-SY 2-1:** As literature indicates that the fault is probably at sample #450, an analysis of the variance graph shows a great decline from this sample onwards, which can be related to the bearing having a three-stage degradation profile.
- **XJTU-SY 2-3:** For this dataset, the results for the variance coincide with the ones of other works, so this early detection is probably not caused by overfitting.
- **XJTU-SY 3-1:** Once again, when using the variance the alarm is triggered too early, but there are no indications of a fault at sample #2300, as other works suggest. An explanation could be that the model was unable to capture the necessary information to differentiate between healthy and faulty samples by just using a size-2 latent space in this experiment.
- **XJTU-SY 3-4:** Similar to the previous dataset, but this time there are a few peaks before sample #1400 that could allow an earlier prediction than those of the literature, although it is also possible that these peaks are not related to the actual fault.

Overall, although most of the results did not align with those reported in the literature, the datasets IMS-1 and XJTU-SY 2-3 demonstrated the expected behavior of this Fault Detection System. An additional verification supporting this claim is presented in Fig. 9, which plots the entire latent space of the VAE for the IMS-1 dataset, classifying each sample using the Variance criteria, as seen in Fig. 8. The scatterplot clearly distinguishes between samples labeled as healthy and those labeled as unhealthy, highlighting the system's effectiveness in identifying faults.

In contrast, the fault detection points for datasets IMS-2, IMS-3, XJTU-SY 2-1, XJTU-SY 3-1, and XJTU-SY 3-4 differed significantly from those reported in other studies. For instance, IMS-2 and IMS-3 exhibited signs of overfitting, leading to very early detections that did not correspond to the expected fault points. Similarly, XJTU-SY 3-1 and XJTU-SY

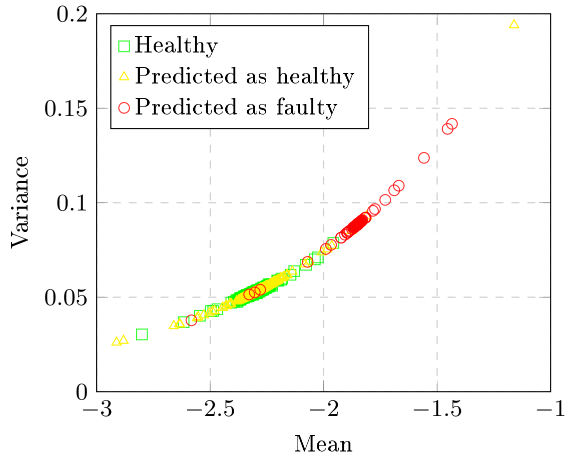

Fig. 9. Scatterplot of the latent space for IMS-1.

3-4 triggered alarms too early, without clear indications of faults at the expected samples.

These discrepancies highlight potential areas for improvement in the system's robustness. The early detections in some datasets suggest that the model may be sensitive to noise or overfitting, which could affect its reliability. However, these differences also provide valuable insights. The entropy results indicate that the datasets may contain initial degradation profiles that were previously overlooked, challenging the widely accepted assumption from prior literature that these datasets did not present an initial degradation profile. By developing procedures to identify and remove these abnormalities during preprocessing, the system's accuracy and robustness could be significantly enhanced. Implementing such procedures would likely yield even more promising results across all techniques tested in this work.

V. Conclusion

The objective of this work was to use neural networks with a probabilistic approach to early detect faults in the bearings of electric motors. A Fault Detection System was designed in which Variational Auto Encoders were trained on healthy samples, allowing unhealthy samples to be represented by distinct Gaussian distributions.

By means of different statistical tests and properties, the distributions obtained from the IMS and XJTU-SY datasets were analysed, in search of early fault indicators. Altough not all of the results were in line with those of the literature, it was possible to extract useful information from them. The main conclusions drawn from these experiments are presented below:

- **VAE summarises reality:** The variance results obtained for IMS-1 and XJTU-SY 2-3 demonstrates what should be expected from the proposed Detection System, clearly showing the vibration abnormalities prior to the fault, confirming the results of other works in literature. A deeper analysis regarding the origins of these abnormalities may not only improve the explainability of this work, but also help to identify their root causes, affecting any future work on this field.
- **Three-stage profiles:** Entropy results show that the VAE is able to recognise degradation samples causing entropy reductions at the end of the experiments. This can only mean that datasets such as XJTU-SY 2-1 and 3-1 correspond to bearings with a Three-stage degradation profile, that is, manufacturing defects were present during training.
- **Inaccurate training approach:** Given the previous conclusion, then it can be deducted that the incoherent results of the applied techniques were caused by the initial degradation stage that was present during training. Correct identification of the problematic samples will not only improve the proposed Fault Detection System, but also help new approaches take into account these degradation profiles that are so often overlooked.

Although the Fault Detection System faced certain challenges, the application of different techniques found evidence of a three-stage degradation profile that will undoubtedly impact how new research will approach the Prognostics and Health Assesment of bearings.

Overall, thanks to the results of this work, new possibilities are opened for more detailed exploration of the IMS and XJTU-SY datasets, in order to advance towards the final objective of building a Fault Detection System that provokes a qualitative leap in the prediction of the Remaining Useful Lifetime of electric motors.

Future work will focus on enhancing the understanding of abnormal results, improving the robustness of the training process via the development of a preprocessing step able to filter samples belonging to an initial degradation process. The resulting methodology will be contrasted against both traditional algorithms, such as Random Forests and Support Vector Machines, and new techniques such as Tabular Prior-data Fitted Network (TabPFN). Additionally, the integration of new methodologies and architectures will be considered to further refine the system's capabilities.

References

[1] T. Zonta, C. A. da Costa, R. da Rosa Righi, M. J. de Lima, E. S. da Trindade, and G. P. Li, "Predictive maintenance in the industry 4.0: A systematic literature review," *Computers & Industrial Engineering*, vol. 150, p. 106889, 2020, https://doi.org/10.1016/j.cie.2020.106889.

[2] L. Magadán, F. J. Suárez, J. C. Granda, F. J. delaCalle, and D. F. García, "A robust health prognostics technique for failure diagnosis and the remaining useful lifetime predictions of bearings in electric motors," *Applied Sciences*, 2023, https://doi.org/10.3390/app13042220.

[3] A. Gholaminejad, F. S. Bidgoli, J. Poshtan, and M. Poshtan, "A novel kurtogram-based health index for induction motor fault diagnosis," *2019 International Aegean Conference on Electrical Machines and Power Electronics (ACEMP) & 2019 International Conference on Optimization of Electrical and Electronic Equipment (OPTIM)*, pp. 85–92, 2019, https://doi.org/10.1109/acemp-optim44294.2019.9007198.

[4] A. J. Bazurto, E. C. Quispe, and R. C. Mendoza, "Causes and failures classification of industrial electric motor," *2016 IEEE ANDESCON*, pp. 1–4, 2016, https://doi.org/10.1109/andescon.2016.7836190.

[5] I. O. for Standardization, "Mechanical vibration - evaluation of machine vibration by measurements on non-rotating parts," 1995, https://doi.org/10.3403/bs7854.

[6] International Organization for Standardization, "Mechanical vibration – measurement and evaluation of machine vibration – part 1: General guidelines," International Organization for Standardization, Geneva, CH, Standard, 2016, https://doi.org/10.3403/30328957.

[7] Y. Xu, K. Feng, X. Yan, R. Yan, Q. Ni, B. Sun, Z. Lei, Y. Zhang, and Z. Liu, "Cfcnn: A novel convolutional fusion framework for collaborative fault identification of rotating machinery," *Information Fusion*, vol. 95, pp. 1–16, 2023, https://doi.org/10.1016/j.inffus.2023.02.012.

[8] P. Liang, Z. Yu, B. Wang, X. Xu, and J. Tian, "Fault transfer diagnosis of rolling bearings across multiple working conditions via subdomain adaptation and improved vision transformer network," *Advanced Engineering Informatics*, vol. 57, p. 102075, 2023, https://doi.org/10.1016/j.aei.2023.102075.

[9] L. Magadán, C. Ruiz-Cárcel, J. Granda, F. Suárez, and A. Starr, "Explainable and interpretable bearing fault classification and diagnosis under limited data," *Advanced Engineering Informatics*, vol. 62, p. 102909, Oct. 2024, https://doi.org/10.1016/j.aei.2024.102909.

[10] H. Wang, X. Zhang, X. Guo, T. Lin, and L. Song, "Remaining useful life prediction of bearings based on multiple-feature fusion health indicator and weighted temporal convolution network," *Measurement Science and Technology*, vol. 33, no. 10, p. 104003, jul 2022, https://doi.org/10.1088/1361-6501/ac77d9.

[11] W. Guo, X. Li, and X. Wan, "A novel approach to bearing prognostics based on impulse-driven measures, improved morphological filter and practical health indicator construction," *Reliability Engineering & System Safety*, vol. 238, p. 109451, 2023, https://doi.org/10.1016/j.ress.2023.109451.

[12] T. Yan, D. Wang, J.-Z. Kong, T. Xia, Z. Peng, and L. Xi, "Definition of signal-to-noise ratio of health indicators and its analytic optimization for machine performance degradation assessment," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–16, 2021, https://doi.org/10.1109/TIM.2021.3075779.

[13] Y. Zhou, A. Kumar, C. P. Gandhi, G. Vashishtha, H. Tang, P. Kundu, M. Singh, and J. Xiang, "Discrete entropy-based health indicator and LSTM for the forecasting of bearing health," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 45, no. 2, p. 120, Feb. 2023, https://doi.org/10.1007/s40430-023-04042-y.

[14] H. Wei, Q. Zhang, and Y. Gu, "Remaining useful life prediction of bearings based on self-attention mechanism, multi-scale dilated causal convolution, and temporal convolution network," *Measurement Science and Technology*, vol. 34, no. 4, p. 045107, jan 2023, https://doi.org/10.1088/1361-6501/acb0e9.

[15] Y. Deng, S. Du, D. Wang, Y. Shao, and D. Huang, "A calibration-based hybrid transfer learning framework for rul prediction of rolling bearing across different machines," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–15, 2023, https://doi.org/10.1109/TIM.2023.3260283.

[16] Z. Li, P. Xu, and X.-B. Wang, "Online anomaly detection and remaining useful life prediction of rotating machinery based on cumulative summation features," *Measurement and Control*, vol. 56, no. 3-4, pp. 615–629, 2023, https://doi.org/10.1177/00202940221098048.

[17] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems—reviews, methodology and applications," *Mechanical Systems and Signal Processing*, vol. 42, no. 1, pp. 314–334, 2014, https://doi.org/10.1016/j.ymssp.2013.06.004.

[18] M. Tian, X. Su, C. Chen, and W. An, "A novel method for multistage degradation predicting the remaining useful life of wind turbine generator bearings based on domain adaptation," *Applied Sciences*, vol. 13, no. 22, 2023, https://doi.org/10.3390/app132212332.

[19] H. Qiu, J. Lee, J. Lin, and G. Yu, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," *Journal of Sound and Vibration*, vol. 289, no. 4, pp. 1066–1090, 2006, https://doi.org/10.1016/j.jsv.2005.03.007.

[20] B. Wang, Y. Lei, N. Li, and N. Li, "A hybrid prognostics approach for estimating remaining useful life of rolling element bearings," *IEEE Transactions on Reliability*, vol. 69, no. 1, pp. 401–412, 2020, https://doi.org/10.1109/TR.2018.2882682.

[21] S. Chauhan, G. Vashishtha, R. Kumar, R. Zimroz, M. K. Gupta, and P. Kundu, "An adaptive feature mode decomposition based on a novel health indicator for bearing fault diagnosis," *Measurement*, vol. 226, p. 114191, 2024, https://doi.org/10.1016/j.measurement.2024.114191.

A. Menéndez-González received his B.S. degree in Computer Engineering in Information Technology from the University of Oviedo in 2022 and his M.S. in Artificial Intelligence from Carlos III University of Madrid in 2023. He is currently working towards his Ph.D. in Computer Science from the University of Oviedo. His research work is focused on predictive maintenance of electric motors, applied artificial intelligence and deep learning.

L. Magadán received his B.S. degree in Computer Engineering in Information Technology in 2019, his M.S. degree in Computer Science in 2021, and his Ph.D. in Computer Science from the University of Oviedo, Spain. He is currently working as an assistant proffesor in the Department of Computer Science and Engineering at the University of Oviedo. His main interests are the Industrial Internet of Things and artificial intelligence applied to fault detection and health prognostics of rotating machinery in industrial environments.

J. C. Granda received his M.S. and Ph.D. in Computer Science from the University of Oviedo, Gijón, Spain, in 2004 and 2008. He is an associate professor with the Department of Computer Science and Engineering at the University of Oviedo. In recent years, he has been working on several projects related to applied Wireless Sensor Networks and Industrial Internet of Things. His research interests include quality inspection systems and predictive maintenance. He is also working on several projects related to multimedia systems and multimedia networking.

F. J. Suárez is a full professor in the Department of Computer Science and Engineering at the University of Oviedo, Spain, where he received his Ph.D. in 1998. His current research is focused on Wireless Sensor Networks, Edge/Fog/Cloud Architectures for Industrial Internet of Things and AI Applications for Predictive Maintenance in Industry. In recent years he has lead several projects in those fields in collaboration with industrial companies as head of the "Sensor to Cloud Systems and Services" research team.