

Received January 20, 2020, accepted February 5, 2020, date of publication February 18, 2020, date of current version February 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2974942

# Health Indicator for Low-Speed Axial Bearings Using Variational Autoencoders

MARTIN HEMMER<sup>1,2</sup>, ANDREAS KLAUSEN<sup>1</sup>, HUYNH VAN KHANG<sup>1</sup>,  
KJELL G. ROBBERSMYR<sup>1</sup>, (Senior Member, IEEE), AND TOR I. WAAG<sup>3</sup>

<sup>1</sup>Department of Engineering Sciences, University of Agder, 4877 Grimstad, Norway

<sup>2</sup>MHWirth AS, 4639 Kristiansand, Norway

<sup>3</sup>NORCE Norwegian Research Centre AS, 4877 Grimstad, Norway

Corresponding author: Martin Hemmer (martin.hemmer@uia.no)

This work was supported by the Norwegian Research Council, SFI Offshore Mechatronics, under Project 237896.

**ABSTRACT** This paper proposes a method for calculating a health indicator (HI) for low-speed axial rolling element bearing (REB) health assessment by utilizing the latent representation obtained by variational inference using Variational Autoencoders (VAEs), trained on each speed reference in the dataset. Further, versatility is added by conditioning on the speed, extending the VAE to a conditional VAE (CVAE), thereby incorporating all speeds in a single model. Within the framework, the coefficients of autoregressive (AR) models are used as features. The dimensionality reduction inherent in the proposed method lowers the need of expert knowledge to design good condition indicators. Moreover, the suggested methodology allows for setting the probability of false alarms when encoding new data points to the latent variable space using the trained model. The effectiveness of the proposed method is validated based on two different datasets: from a workshop test of an offshore drilling machine and from an in-house test rig for axial bearings. In both datasets, the HI is exceeding the warning and alarm levels with a probability of false alarm (PFA) of  $10^{-6}$ , and the method is most effective at lower shaft speeds.

**INDEX TERMS** Bearing fault detection, condition monitoring, unsupervised learning, variational autoencoder, conditional variational autoencoder, generative models.

## I. INTRODUCTION

Rolling element bearings (REBs) are widely used in heavy industrial machinery such as offshore drilling machines, wind turbines, and paper mills. A defect in such bearings might result in a catastrophic failure in the industrial system. Therefore, condition monitoring (CM) for REBs is important to avoid unplanned downtime and production loss in heavy industry. The majority of bearing condition monitoring techniques focus on detecting the presence and development of localized damage in bearing raceways or rolling elements [1]–[3]. CM of low-speed machinery, with a shaft speed below 10 Hz [4], is more challenging. The energy associated with faults is then smaller, resulting in a low signal-to-noise ratio (SNR). This requires more sensitive sensors and development of advanced signal processing methods to extract fault signatures. Operating conditions tend to be less stationary at lower speeds [4], thus resampling to angular domain

is necessary in low-speed applications [5], [6]. Health conditions of large bearings at low speed are usually observed via acoustic emission or vibration measurements [5], [7]–[11]. Cyclostationary methods [12]–[14], wavelet denoising and filtering [15], [16], and empirical mode decomposition (EMD) [17] have all been successful in low-speed bearing fault detection. Data-driven fault diagnosis methods based on machine learning have also been intensively developed in recent years [18]. Fault classifiers based on decision trees (DT) [19], [20], support vector machine (SVM) [21], [22], k-nearest neighbor (k-NN) [23], [24], convolutional neural network (CNN) [25]–[27] and deep belief networks (DBN) [28], [29] are well applied to deal with bearing fault detection. All mentioned machine learning based methods require historical failure data for training, which is hard to obtain in industry. In addition, the authors could not identify previous research dealing with faults on axial bearings, where a characteristic fault frequency might not exist or is inconsistent in spectra. This work aims to develop an anomaly detection method without using historical failure data. Tapered axial

The associate editor coordinating the review of this manuscript and approving it for publication was Hong-Mei Zhang<sup>1</sup>.



**FIGURE 1.** Wear on roller end of offshore drilling machine bearing.

roller bearings, e.g. in drilling machines from the offshore industry, have relative sliding motion in the rib-roller contact area. Low speed makes this area particularly susceptible to wear. In [30], wear on the roller ends was observed in a tapered axial bearing, as shown in Fig. 1.

However, no characteristic frequency component associated with defects, i.e. roller frequency, was observed on the axial bearing during tests. This suggests that diagnosis methods based on detection of defect characteristic frequencies alone are ineffective in detecting wear in large and slow axial bearings. Identifying this defect on the axial bearings is currently relying on offline monitoring methods such as lubricant analysis and visual inspection combined with precautionary maintenance actions [25], [30]. This practice requires interruptions of production and may allow failure to progress inconspicuously between inspections. Therefore, development of online, non-intrusive monitoring methods is very important to facilitate condition based maintenance (CBM) for large axial bearings in heavy industry. Since data from a healthy state is easier to obtain than in a damaged state, a procedure of determining whether or not the observed bearing is normal based on prior knowledge of healthy behavior of the machine, would be very useful to avoid using failure data.

References [31], [32] proposed a method for health threshold setting based on healthy operating characteristics, allowing controlling probability of false alarm ( $P_{FA}$ ). A whitening transformation was applied to a set of correlated condition indicators (CIs) with Gaussian or Rayleigh distributions. These CIs were then used to calculate a health indicator (HI) with a known probability density function (PDF) and cumulative distribution function (CDF). The HI is normalized by the inverse CDF evaluated at  $(1 - P_{FA})$ , and optionally multiplied with a warning factor  $w < 1$ . In this case, let  $HI_0$  denote an observation from a healthy machine. The probability of observing  $HI_0$  above the warning factor is then equal to the  $P_{FA}$ , as shown in (1). The consequences of failures and false alarms must be considered when setting the  $P_{FA}$  threshold. Additionally, the number of inferences to be done must be considered. Multiple testing increases the risk of false positive samples simply by chance [33].

$$P(HI_0 > 1) = P_{FA} \quad (1)$$

However, the method in [31] is only effective if CIs are well selected with known probability distributions. The overall goal of conventional approaches is to perform health assessment with a statistical foundation, based on a potentially large set of observed variables. Due to the curse of dimensionality, also known as Hughes' phenomenon [34], a single indicator might "drown" in high dimensional feature space, which reduces accuracy of the model. Thus, a method for dimensionality reduction of the features while maintaining most of the information is required. To implement the HI threshold setting, features are also required to be independent variables. Principal Component Analysis (PCA) can transform a set of variables to linearly uncorrelated features with decreasing contribution to the variance, but it does not account for non-linear dependencies. Machine learning (ML) algorithms can be an alternative solution since they can capture complex dependencies among the observed variables.

Autoencoders are successfully used for dimensionality reduction in fault detection and classification of rotating machinery [35]–[38], but lack a probabilistic latent representation. Generative models are capable of estimating complicated PDFs of given data, and can generate new samples, which follow the same distribution as the training data. In [39], it was shown that sequential training of restricted Boltzmann machines could discover hidden dependencies between observed variables and a sparse representation. However, training such networks typically requires an additional statistic method, e.g. Markov Chain-Monte Carlo (MCMC) methods, resulting in computational burden.

To achieve dimensionality reduction and reduce computational burden, this work uses a combination of a Variational Autoencoder (VAE) [40] and a Generative Adversarial Networks (GAN) [41], which is similar to the Adversarial Autoencoder (AAE) [42]. The VAE performs inference of variational parameters using neural networks in an encoder-decoder structure by minimizing the reconstruction error and the Kullback-Leibler Divergence ( $D_{KL}$ ) [43] between an encoded sample and a Gaussian standard distribution, which is equivalent to maximizing the evidence lower bound (ELBO). This objective can be optimized with gradient descent algorithms through the "reparameterization trick" [40]. These generative models allow imposing a distribution on the latent variables. In [42], the latent distribution in AAEs seems to follow the target distribution closer, which is desirable for the purpose of a HI. However, the adversarial training in GANs and AAEs is often unstable [44]. This problem was also observed in experiments with AAEs while developing the proposed method. VAEs have been used in ball bearing fault classification by using the latent variables for each data point as input to a classifier [45]. The proposed approach instead utilizes the aggregated distribution of healthy conditions in the latent space of a VAE to calculate a HI for new observations.

The remaining of the paper is organized as follows: In section II, the network architecture, training procedure and HI calculation are described. Section III details the data

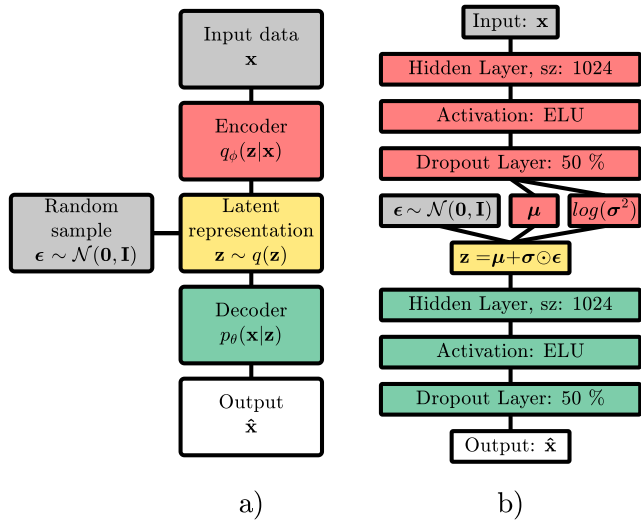


FIGURE 2. a) VAE architecture overview, b) Detailed architecture.

acquisition and pre-processing. Results from two different datasets are presented in section IV. Section V provides conclusions and discussions.

## II. METHODOLOGY

This section presents the approach for calculating a bearing health indicator, utilizing the latent variables in a VAE. The calculation of a HI limits the selection of CIs to those following known distributions as described in [31] for Gaussian and Rayleigh distributions. It also requires the user to pre-select suitable CIs based on domain knowledge. The proposed method performs unsupervised dimensionality reduction from a set of input features, while simultaneously imposing a Gaussian distribution on the latent variables. This section provides a review of the network components, the loss functions and training algorithm. The model was implemented in Python using TensorFlow r1.12 [46].

### A. NETWORK ARCHITECTURE AND LOSSES

The network architecture is shown in Fig. 2. An encoder (red) and a decoder (green) are connected by the latent representation (yellow).

Let  $\mathbf{x}$  be the feature input vector and  $\mathbf{z}$  be the latent variable vector. The encoder consists of a fully connected layer of size 1024 with weights, biases, an exponential linear unit (ELU) activation function, and 50 % dropout. In this work, the coefficients of an autoregressive model are used as features. The output includes two vectors, containing the parameters of the latent representation for each data point. Let  $J$  be the dimension of the latent space. The latent variables are constrained to have a Gaussian distribution with diagonal covariance matrix, so the encoder outputs a vector containing the means,  $\boldsymbol{\mu}$ , and log of the variances,  $\log(\sigma^2)$ , each of length  $J$ . Note that these parameters are for the individual data points, not the aggregated latent distribution  $q(\mathbf{z})$ . Utilizing the reparameterization trick from [40], samples from a white

noise vector  $\boldsymbol{\epsilon}$  are used to obtain a random sample  $\mathbf{z}$  from the latent representation while still allowing gradients to flow through the network. The decoder has the same architecture as the encoder, with a fully connected hidden layer, ELU activation, and 50 % dropout. Weights and biases are denoted  $\boldsymbol{\phi}$ . The desired output is a reconstruction of the input, like a normal autoencoder. Combining these parts of the network results in the VAE. Originally, the VAE was developed as a generative model for producing reconstructions similar to the input by sampling from a given prior distribution  $p(\mathbf{z})$ . The connection between data and  $p(\mathbf{z})$  is in general not known and must be approximated. Let the training data distribution be  $\mathbf{x} \sim p_d(\mathbf{x})$ , and VAE output  $\hat{\mathbf{x}} \sim p(\mathbf{x})$ . Further,  $q_\phi(\mathbf{z}|\mathbf{x})$  and  $p_\theta(\mathbf{x}|\mathbf{z})$  are the encoding and decoding distributions of the encoder and decoder networks. Subscript  $\phi$  and  $\theta$  are the encoder and decoder variables. Thus, the aggregated posterior distribution of the latent variable,  $\mathbf{z} \sim q(\mathbf{z})$ , is defined as in (2). To be utilized in the HI calculation,  $q(\mathbf{z})$  must approximate the desired prior  $p(\mathbf{z})$ .

$$q(\mathbf{z}) = \int_{\mathbf{x}} q_\phi(\mathbf{z}|\mathbf{x})p_d(\mathbf{x})d\mathbf{x} \quad (2)$$

To ensure that the latent representation contains useful information about the input data, the encoder and decoder are trained to minimize the reconstruction loss function  $\mathcal{L}_R$ , as in (3).  $\mathcal{L}_R$  is the mean square error between each feature  $x_{i,j}$  and its reconstruction  $\hat{x}_{i,j}$  over a minibatch,  $\mathbf{x}_M$ , of size  $M$ . The number of features per datapoint is denoted  $N$ .

$$\mathcal{L}_R(\boldsymbol{\phi}, \boldsymbol{\theta}; \mathbf{x}_M, \boldsymbol{\epsilon}) = \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^N (x_{i,j} - \hat{x}_{i,j})^2 \quad (3)$$

This encourages similar input data to cluster in latent space, while dissimilar data are separated. Note that the square error is summed over a datapoint and averaged over the minibatch. This gives more weight to reconstruction error, which helps avoid mode collapse, i.e. the latent vector converges to a Gaussian that does not carry information.

While reducing  $\mathcal{L}_R$  provides a good reconstruction, the aggregated latent distributions will not take a Gaussian distribution. To make the latent distribution approximate the desired prior, KL divergence is introduced as a regularization on the encoder variables  $\boldsymbol{\phi}$ . Given the assumption of diagonal covariance matrix, and Gaussian prior, the KL divergence for a data point can be calculated in a closed form. The combined KL loss over a minibatch is then calculated as in (4).

$$\begin{aligned} \mathcal{L}_{KL}(\boldsymbol{\phi}; \mathbf{x}_M) &= \frac{1}{2M} \sum_{i=1}^M \sum_{j=1}^J \left( 1 + \log(\sigma_{i,j}^2) - \mu_{i,j}^2 - \sigma_{i,j}^2 \right) \end{aligned} \quad (4)$$

The objective function to be minimized is the sum of  $\mathcal{L}_R$  and  $\mathcal{L}_{KL}$ , as given in (5).

$$\mathcal{L}_{VAE}(\boldsymbol{\phi}, \boldsymbol{\theta}; \mathbf{x}_M, \boldsymbol{\epsilon}) = \mathcal{L}_R + \mathcal{L}_{KL} \quad (5)$$

Pseudo-code for the training procedure is given in Algorithm 1. Training was repeated 5 times with different random

**Algorithm 1** Training Algorithm

---

```

 $\phi, \theta \leftarrow x$  Initialize parameters
repeat
  Shuffle training dataset
  repeat
     $\mathbf{x}_M \leftarrow$  Get minibatch from the training dataset
     $\mathbf{g} \leftarrow \nabla \mathcal{L}_{VAE}(\phi, \theta; \mathbf{x}_M, \epsilon)$  Calculate gradients
     $\theta, \phi \leftarrow$  Update encoder/decoder parameters
  until Epoch is completed
until Total number of epochs is completed
return  $\phi, \theta$ 

```

---

seeds for weight initialization and shuffling. Hyperparameters used in the experiments are given in Table 1. Model weight updates are performed using the Adam [47] optimizer with cosine decay of the learning rate, as this has been shown to improve Adam performance [48]. Initial learning rate was set to  $10^{-4}$ , which decays to  $10^{-6}$  over the training epochs. Experiments with higher learning rate values caused unstable convergence and divergence of training loss. A dropout rate of 50% was implemented to reduce overfitting [49]. The ELU activation function used in the encoder and decoder hidden layer is shown to outperform other activation functions both in CNNs and autoencoders [50]. As suggested in [40], minibatch size is set to 100. Early stopping was not implemented but could speed up the training process as the training loss converged well before the number of epochs.

The proposed settings allowed the latent variable distribution to converge to a Gaussian distribution while avoiding mode collapse, which suggests that the network configuration is suitable for this particular application. Further experiments to optimize layer size and latent dimension has not been performed. For another dataset with a different AR input size it may be necessary to change the hidden layer size if the latent distribution does not converge to a Gaussian distribution or suffer from mode collapse.

Reconstruction loss constrains the network to ensure that useful information is captured when forming Gaussian latent distributions. Limiting the latent dimension forces the network to infer underlying Gaussian variables that best describes the observed data, at the cost of overall reconstruction performance. Improved reconstruction of input data could be achieved by increasing the latent dimension as well as adding more hidden layers and increasing their size. It is important to note that optimal reconstruction in itself is not the main purpose for the network. Increasing the number of latent variables capability could, however, reduce the HI sensitivity to faults, as the HI is calculated as the norm of the latent vector.

**B. CONDITIONAL VARIATIONAL AUTOENCODER**

With the described approach, it is required to train a separate VAE for each speed. For machines with multiple operating conditions, this is impractical. Therefore, a conditional VAE

**TABLE 1.** Training Parameters.

Parameter	Value
Optimizer	Adam
Learning rate schedule	Cosine decay
Initial learning rate	1e-4
Final learning rate	1e-6
Minibatch size ( $M$ )	100
Hidden layer size	1024
Dropout	50 %
Activation	ELU
Latent dimension ( $J$ )	3
Epochs, Dataset 1	30 000
Epochs, Dataset 2	15 000

(CVAE) is trained for each dataset. CVAEs utilize the same network structure and loss function  $\mathcal{L}_{VAE}$  as VAEs, but can be conditioned on additional information, such as speed. For each datapoint, the speed information is a categorical variable, one-hot encoded into a conditioning vector  $c$ . For example, the speed of 100 rpm in dataset 1 is encoded to  $c_{100} = [0, 1, 0, 0, 0]$  while the speed of 60 rpm in dataset 2 is encoded to  $c_{60} = [0, 1]$ . As the model order is different for rpms,  $\mathbf{x}$  is zero-padded to the largest model order  $p$ . VAE training datasets consist of data from a single speed, while the CVAE uses data from all speeds. Except for these differences, VAEs and CVAEs follow an identical training procedure.

**C. HEALTH INDICATOR**

A methodology for threshold setting given CIs with Rayleigh or Gaussian distributions is proposed in [31]. In this work, a Gaussian distribution is chosen for the latent variables. To verify that  $q(\mathbf{z})$  approximates the standard normal distribution  $\mathcal{N}(\mathbf{0}, \mathbf{I})$ , the Kullback-Leibler Divergence ( $D_{KL}$ ) was calculated as given in (6) for the aggregated posterior.  $\Sigma$  is the covariance matrix of  $\mathbf{z}$ ,  $\boldsymbol{\mu}$  is a vector containing the mean values of  $\mathbf{z}$  and  $J$  is the number of latent variables.

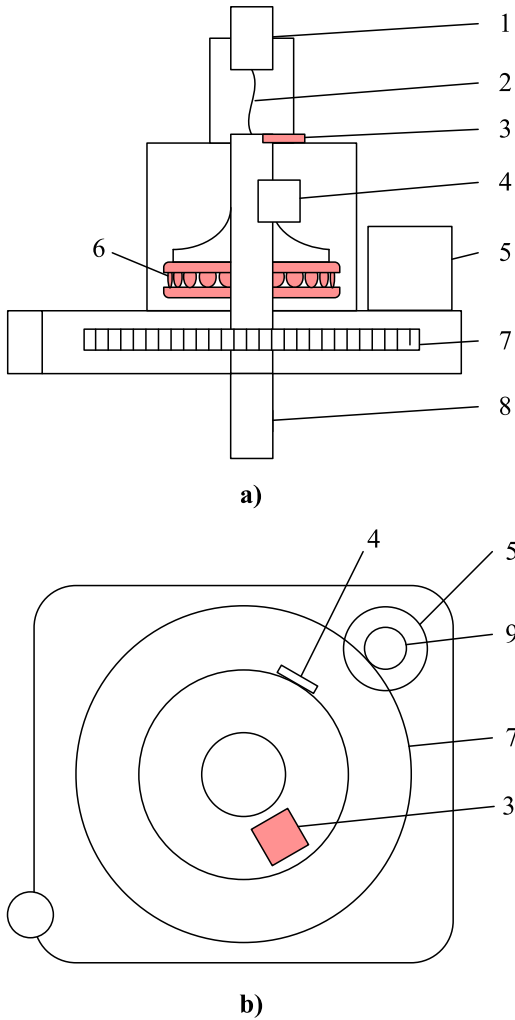
$$D_{KL} = \frac{1}{2} \left( \text{tr}(\Sigma) + \boldsymbol{\mu}^T \boldsymbol{\mu} - J + \log_e \left( \frac{1}{\det \Sigma} \right) \right) \quad (6)$$

The norm of  $J$  Gaussian variables follows a  $\chi$  distribution with  $\nu$  degrees of freedom. Let  $F(\cdot)$  denote the CDF of a  $\chi$ -distribution. The HI is normalized with a factor that is a function of the  $P_{FA}$ . The HI is calculated as shown in (7).

$$\text{HI} = \left( \sum_{j=1}^J |z_j|^2 \right)^{\frac{1}{2}} \left( F^{-1}(1 - P_{FA}) \right)^{-1} \quad (7)$$

**III. EXPERIMENTAL SETUP**

The proposed algorithm is tested on data from two experiments: Vibration data from a workshop test of an offshore drilling machine, and acoustic emission (AE) data from an in-house test rig for axial bearings. A further description of the experimental setup is given in the following sections.

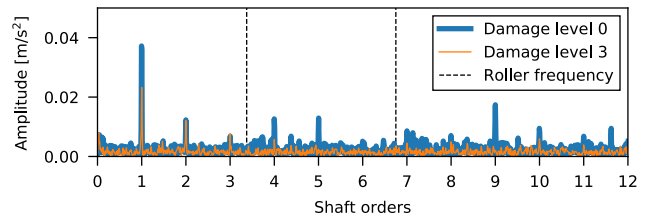


**FIGURE 3.** Conceptual drawing of an offshore drilling machine. a) Side view b) Top view. 1: Encoder 2: Flexible coupling 3: Axial sensor placement 4: Radial sensor placement, 5: Motor, 6: Axial bearing, 7: Gear, 8: Shaft, 9: Pinion.

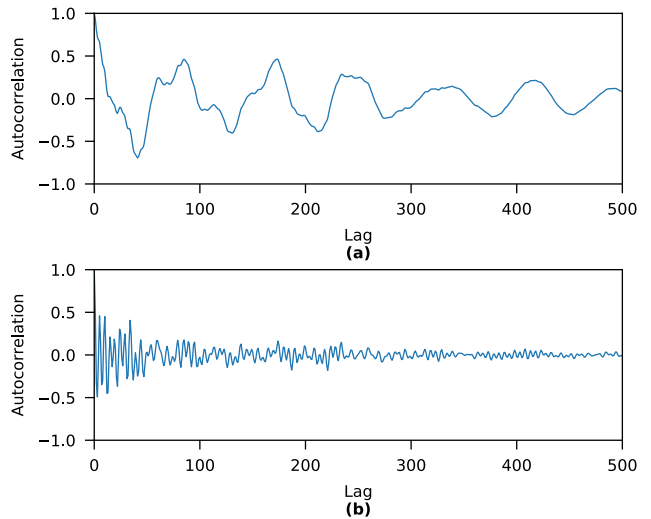
**A. DATASET 1: OFFSHORE DRILLING MACHINE WORKSHOP TEST**

Dataset 1 (DS1) was collected from an offshore drilling machine taken out of operation for maintenance as described in [30]. A schematic drawing of the setup is shown in Fig. 3.

Data is collected from an accelerometer mounted in the axial location. Data was sampled at 102.4 kHz and decimated to 81.92 kHz. The axial bearing showed signs of roller end wear as shown in Fig. 1. Data was first recorded using a healthy bearing, being denoted damage level (DL 0). Then, reassembling the machine with a slightly damaged bearing results in a change of the vibration characteristics and a reduction in root mean square (RMS) [30]. Distinguishing this change from any fault induced change is not possible. Thus, the slightly damaged condition is selected as the baseline condition (DL 1) for training data. Additional damages in the form of indentations from a carbide tip tool were applied to one of the roller end, producing data at DL 2. For data at DL 3, the bearing was further damaged and also run under poor



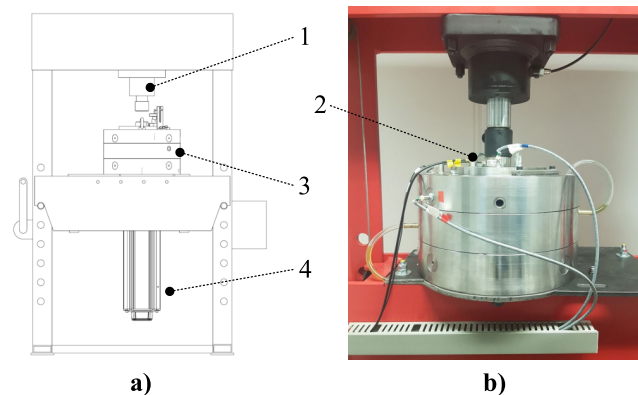
**FIGURE 4.** Envelope spectrum for a healthy bearing (damage level 0) and at damage level 3.



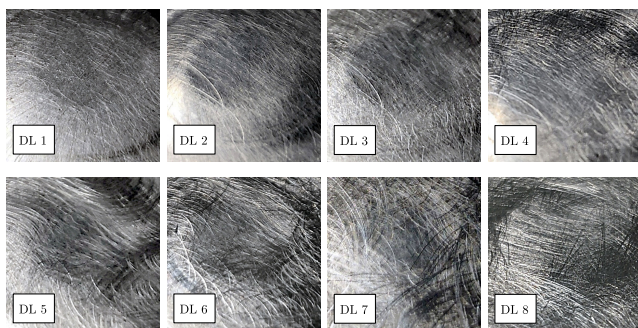
**FIGURE 5.** Autocorrelation function of vibration data captured in Dataset 1. a) Raw data b) After differentiation.

lubricating conditions. Data was recorded at 50, 100, 150, 200 and 250 rpm. At 50 rpm, only data from DL 1 and DL 3 was recorded. The machine was running unloaded, subject to the gravity by its own weight. A quantitative measurement of damage is not available, but a degradation resulting in a measurable change is expected. However, previous analysis of the vibration signal was not successful in detecting any clear indication of the damage [30]. Damage to a roller was expected to cause amplitude modulation at the roller frequency, but as shown in Fig. 4, no peak was observed at either one or two times the roller frequency in the envelope spectrum.

Segments corresponding to approximately 1 revolution are used for calculating the features. To increase the number of data points, an overlap of 75 % is applied. The autocorrelation function (ACF) is examined on a healthy dataset to determine if the signal is stationary or not. If the ACF reduces quickly, the signal is considered stationary [51], otherwise the signal is considered non-stationary. The ACF of a vibration signal acquired at 50 rpm is shown in Fig. 5 a). The ACF is slowly decreasing, and has a cyclic trend, and the signal is therefore considered stationary. To mitigate trends and cyclic signal components, the signal is differentiated once. Effectively, the jerk ( $m/s^3$ ) is calculated with this differentiation, and low-frequency components from shaft and gearbox are mitigated, while high-frequency components are enhanced.



**FIGURE 6.** a) Drawing of in-house axial bearing test rig. b) Assembled bearing test unit. 1: Load cylinder, 2: AE sensor location, 3: Bearing test unit, 4: Motor.



**FIGURE 7.** Damage on roller ends in dataset 2. DL 1: P400, DL 2: P400 Heavy, DL 3: P320, DL 4: P220, DL 5: P220 Heavy, DL 6: P80, DL 7: P8 Heavy, DL 8: P80 Very Heavy.

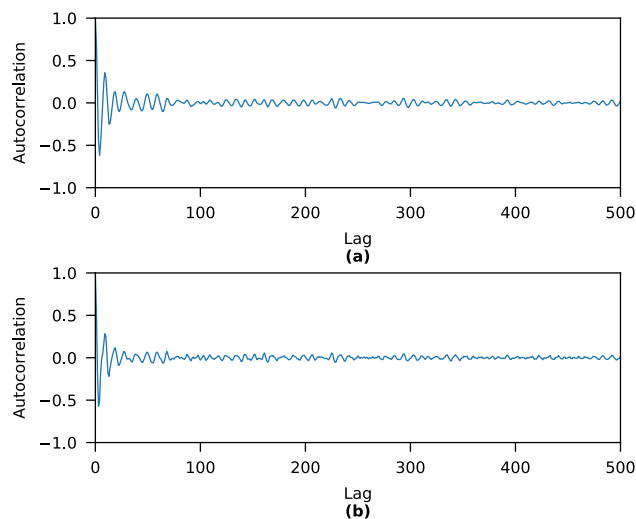
The resulting ACF after differentiation is shown in Fig. 5 b), showing that the ACF is now decreasing fast, and only varies randomly after 100 lags. Given this result, the vibration signal acquired on this test rig is differentiated once to make the signal more stationary.

**B. DATASET 2: AXIAL ROLLER BEARING TEST RIG**

Dataset 2 (DS2) consists of AE data from an in-house test rig, shown in Fig. 6. The test bearing was of type 29230 M from manufacturer ISB, subject to an axial load of 50 kN. Data was recorded at 30 and 60 rpm, in that order. AE data was collected at 1 MHz sampling rate for 10 seconds. Data was then split into constant length segments of 50 000 samples.

To emulate the distributed abrasive wear shown in Fig. 1, the rollers were removed, and roller ends were ground with sandpaper of grit size from ISO/FEPA grit grade P400 (finest), P320, P220 and P80 (coarsest), as shown in Fig. 7. “Heavy” and “Very Heavy” refer to relative degrees of damage using the same sandpaper grade.

The ACFs of the acoustic emission dataset before and after differentiation are shown in Figs. 8. a) and b), respectively. The ACF of the raw signal in Fig. 8 a) decreases rapidly, and differentiating the signal has little effect on the ACF as observed in Fig. 8 b). Therefore, the acoustic emission signal is considered stationary and requires no further differentiation.



**FIGURE 8.** Autocorrelation function of vibration data captured in Dataset 2. a) Raw data b) After differentiation.

**C. FEATURE EXTRACTION AND PREPROCESSING**

The input  $x$  to the autoencoder network is a feature calculated from the vibration and AE data. In the previous work, vibration energy was not significantly increased when the damage level on an axial bearing was escalated [30]. In addition, energies at specific characteristic frequencies do not increase either. However, the bearing condition degradation is expected to produce a change in frequency content of the associated signal. Therefore, features, which are sensitive to changes in the measured signal, are required to be used as input to the autoencoder. An autoregressive (AR) model of order  $p$  can predict the next signal sample based on a linear combination of  $p$  previous samples, assuming that the signal  $s$  is stationary.

Thus, changes in the AR model parameters should reflect that the vibration signal has changed. The AR coefficients may have arbitrary distributions, which makes it challenging to quantify a change. It is therefore easier to threshold in latent space, where the distribution of healthy latent vectors approximates a Gaussian distribution.

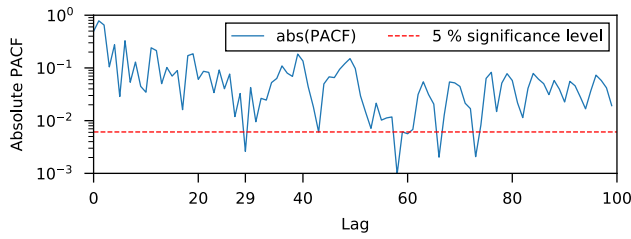
The AR model is depicted as

$$s_i = v_i + \sum_{j=1}^p a_j s_{i-j} \tag{8}$$

where  $s_i$  is the signal at  $i$ 'th time step,  $v$  is the model residual and  $a_j$  is the  $j$ 'th model parameter. The Yule-Walker equations [52], [53] are solved for an input signal  $s$  to obtain the AR model parameters.

The order  $p$  is determined by calculating the partial autocorrelation function (PACF) [54] for an increasing number of lags. The model order  $p$  of a time series with  $N$  samples is considered sufficient where PACF at lag  $p$  is zero with a 5 % significance level [51], as given in (9) [55].

$$PACF \leq \frac{\pm 1.96}{\sqrt{N}} \tag{9}$$



**FIGURE 9.** Absolute value of the partial autocorrelation function of a differentiated vibration data record acquired using test rig 1, model order 29 selected.

**TABLE 2.** Statistics for calculated model order  $p$ .

	DS1-50	DS1-100	DS1-150	DS1-200	DS1-250	DS2-30	DS2-60
Mean ( $p$ )	32	12	20	16	15	14	25
STD	9	9	13	9	10	10	11
Median	29	5	23	9	9	7	17

The smallest lag  $p$ , which results in a PACF below the 5% significance level, is determined for each healthy segment. As an example, the PACF of a differentiated vibration signal acquired at 50 rpm using test rig 1 is shown in Fig. 9.

At lag 29, the PACF is beneath the 5% significance level. This procedure is repeated for all signal segments, and statistics between all segments within each speed range are calculated afterwards and shown in Table 2. As seen in the table, the mean value is selected as model order  $p$ . Standard deviation (STD) and median are also given for each dataset and speed.

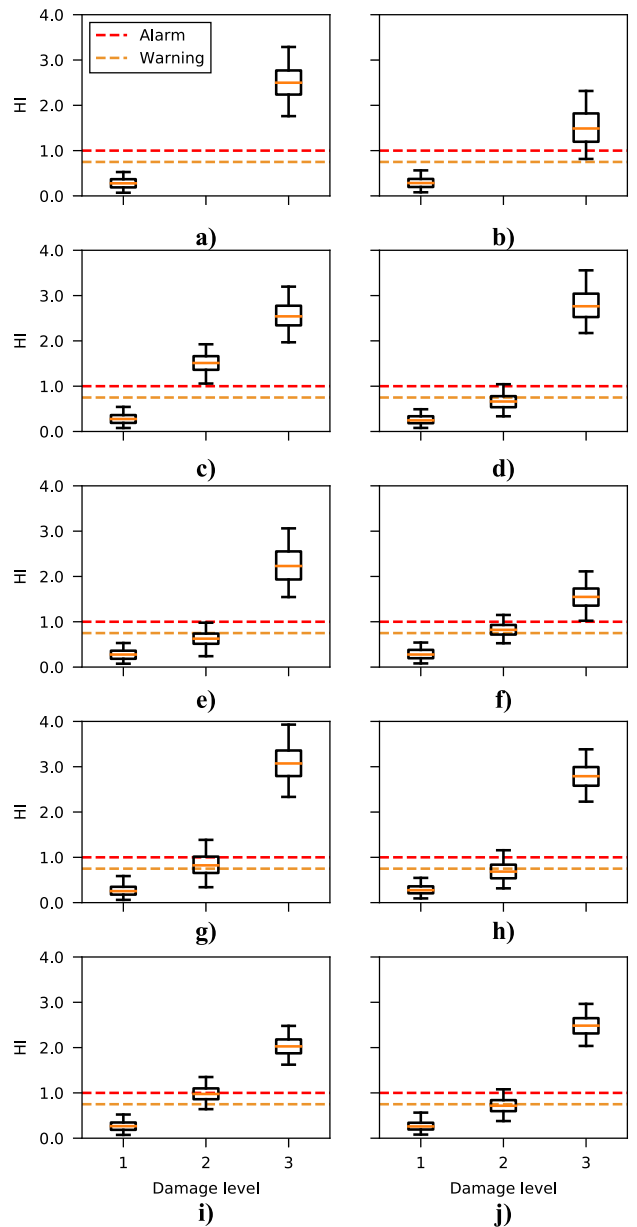
All input data was afterwards standardized using the mean and standard deviation of the remaining training data. Outliers in training data are removed if one AR coefficient differed from the mean value by more than five standard deviations. Baseline data (DL0) was shuffled and split in training (50%), validation (25%) and test (25%) subsets. The remaining DLs were used for testing only. Table 3 shows sample rate, number of samples in the raw data, and size of the datasets at each DL.

#### IV. RESULTS

This section presents the results of the experiments, evaluating the calculated HI using both VAE and CVAE. The validity of the required assumptions of a Gaussian-distributed latent variable is also discussed. The presented results are the aggregate of the 5 models trained with different random initialization.

##### A. HEALTH INDICATOR EVALUATION

In the first dataset, DS1, an increase in HI with damage level is observed at all speeds. The alarm level ( $HI = 1$ ) is calculated with  $P_{FA} = 10^{-6}$ . Boxplots of the calculated HI from VAEs and a CVAE are shown in Fig. 10. Whiskers are set to 2.5th and 97.5th percentile. In the following discussion, the median (orange line inside boxes) is considered as the HI value. In dataset 1, HI at DL2 exceeds the warning level



**FIGURE 10.** HI for dataset 1. Left column: Standard VAE. Right column: Conditional VAE. a-b) 50 rpm, c-d) 100 rpm, e-f) 150 rpm, g-h) 200 rpm, i-j) 250 rpm. Whiskers are set to 2.5th and 97.5th percentile.

0.75 in all speeds except at 150 rpm for VAE ( $HI = 0.63$ ) and 200 rpm for CVAE ( $HI = 0.69$ ). Data for DL2 was not recorded at 50 rpm. At DL2, the HI exceeds alarm value of 1 at all speeds. Results from VAE and CVAE differ more as damage level increases, but the overall results are well aligned with an increase in HI with damage level at all rpms.

The HI calculated for dataset 2 with VAE and CVAE is shown in Fig. 11. At 30 rpm, the HI is above the alarm level from DL2. However, there is no monotonic increase in HI level with damage level. Still, this result should be considered as a clear indication of anomalous behaviour. HI for 60 rpm follow a similar trend, but the HI values are lower, exceeding the warning level in DL3-5 only. As in DS1, the HI values calculated using the standard VAE and CVAE are very similar.

TABLE 3. Dataset details.

	DS1-50	DS1-100	DS1-150	DS1-200	DS1-250	DS2-30	DS2-60
Sample rate	81.92 kHz	81.92 kHz	81.92 kHz	81.92 kHz	81.92 kHz	1000 MHz	1000 MHz
Samples	98304	49152	32768	24576	19661	50000	50000
DL0 size	988	568	596	328	312	2099	2098
DL1 size	-	536	480	280	372	420	420
DL2 size	964	496	388	304	280	420	420
DL3-8 size	-	-	-	-	-	420	420
	420	420					

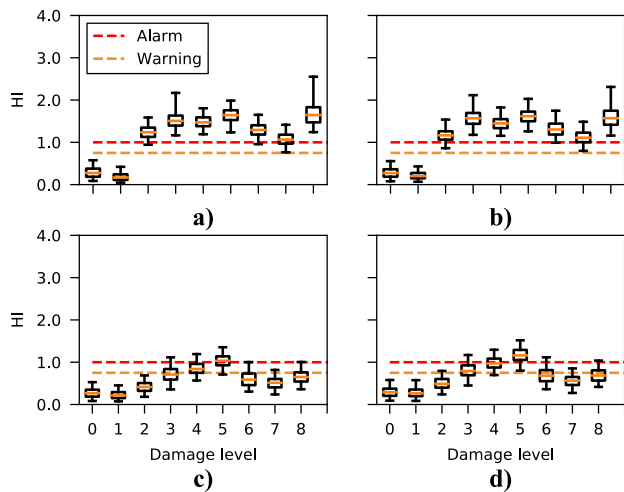


FIGURE 11. HI for dataset 2. Left column: standard VAE. Right column: Conditional VAE. a-b) 30 rpm, c-d) 60 rpm. Whiskers are set to 2.5th and 97.5th percentile.

Compared to dataset 1, there is less consistency in the HI with increasing HI, and larger differences between speeds. The inconsistency between damage levels may be caused by removing the bearing for applying damage. This procedure introduces differences in the mechanical assembly that may affect the results. Also, the damage was applied manually, which gives room for more variations between damage levels. Finally, data for increasing speeds were recorded consecutively. The seeded damage may therefore be smoothed over time during acquisition. This is a possible explanation for the differences between 30 rpm and 60 rpm. If the smoothing effect differs with damage severity, this will also contribute to the HI inconsistency between damage levels. Further, higher speed may generate high energy frequency components, which dominate the AR coefficients but are not associated with the bearing damage.

**B. MODEL PROPERTIES**

A summary of final training, validation and test losses for the VAE are shown in Table 4, including the median values for the 5 models. The ability of latent representations carrying useful information is measured by the reconstruction loss  $\mathcal{L}_R$ . Examining  $\mathcal{L}_R$  in Table 4 reveals that the value is correlated with model order  $p$ , which is expected from the square error summation per datapoint in (3). The reconstructed AR

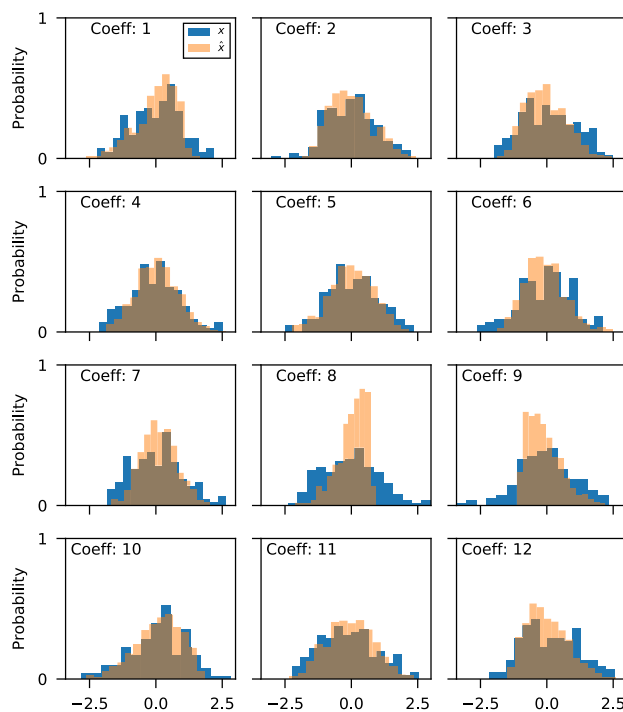


FIGURE 12. Histogram of test data  $x$  and reconstructions  $\hat{x}$  for the 12 coefficients in DS1-100.

coefficients for DS1-100 are shown in Fig. 12. This speed has the lowest number of features ( $p = 12$ ) in the dataset, and also the lowest reconstruction loss. Still, we observe that reconstructions of coefficient 8 and 9 are skewed. It is likely that further tuning of hyperparameters such as hidden layer size, number of hidden layers and latent dimension can improve reconstruction, but a systematic investigation of parameters search was not performed due to the associated computational cost of training.

The statistical properties of the HI assume that  $q(\mathbf{z})$  approximates a multivariate standard Gaussian distribution  $p(\mathbf{z}) \sim \mathcal{N}(0, \mathbf{I})$ . The Gaussian latent space is imposed by  $\mathcal{L}_{KL}$ , which takes values between 1.838 (DS1-250) and 2.106 (DS1-50) in the test dataset. The loss values are more stable than  $\mathcal{L}_R$ , as the latent dimension  $J$  is constant.

However,  $\mathcal{L}_{KL}$  describes the mean KL divergence of each datapoint rather than the aggregated distribution of  $q(\mathbf{z})$ . Therefore, the KL divergence  $D_{KL}$  between the aggregated distribution (after sampling) and  $p(\mathbf{z})$  are calculated as in (6).



TABLE 4. Final loss values after training of the VAEs.

	$\mathcal{L}_R$			$\mathcal{L}_{KL}$			$KL_D$		
	train	val	test	train	val	test	train	val	test
DS1-50	20.668	23.130	22.607	3.484	3.266	2.106	0.033	0.017	0.025
DS1-100	4.432	5.375	5.400	2.765	2.550	1.950	0.033	0.033	0.050
DS1-150	9.814	10.397	11.623	3.205	2.991	2.085	0.038	0.044	0.021
DS1-200	6.987	8.778	9.189	3.001	2.853	2.071	0.028	0.091	0.112
DS1-250	5.328	6.090	6.215	3.178	2.739	1.838	0.020	0.067	0.054
DS2-30	5.255	5.670	5.569	2.844	2.803	2.099	0.050	0.010	0.007
DS2-60	7.418	7.658	7.236	3.586	3.452	1.943	0.041	0.019	0.017

TABLE 5. Final loss values after training of the CVAE.

	$\mathcal{L}_R$			$\mathcal{L}_{KL}$			$KL_D$		
	train	val	test	train	val	test	train	val	test
DS1	14.453	14.080	13.977	2.910	2.825	1.988	0.068	0.007	0.008
DS2	5.302	6.303	8.389	2.702	3.015	2.748	0.038	0.005	0.020

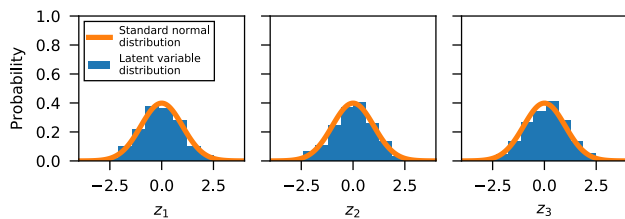


FIGURE 13. Histogram of the aggregated posterior  $\mathbf{z} = [z_1, z_2, z_3]$  for the test data of DS1-100.

The value is bounded to  $D_{KL} \geq 0$ , and a value of zero means that  $q(\mathbf{z})$  and  $p(\mathbf{z})$  are identical distributions. In the test datasets,  $D_{KL}$  takes values between 0.007 (DS2-30) and 0.112 (DS1-100). Fig. 13 shows histograms of each dimension of  $\mathbf{z}$  for DS1-100, which has  $D_{KL} = 0.05$ . A qualitative evaluation confirms that it approximates a Gaussian distribution.

Table 5 lists  $\mathcal{L}_R$ ,  $\mathcal{L}_{KL}$  and  $KL_D$  from the network trained as CVAEs, where all speeds in the dataset are used simultaneously in training.  $\mathcal{L}_R$  in the test datasets is higher than the average for the separate speeds in Table 4. This is reasonable, as the same number of neurons in the network must learn to reconstruct data from 5 and 2 rpms in DS1 and DS2 respectively, instead of just one. However, we see that values for  $\mathcal{L}_{KL}$  and  $KL_D$  are similar to the VAE. This indicates that the assumption of a Gaussian latent space is valid for the CVAE as well.

### V. CONCLUSION

This paper proposes a method for unsupervised learning of a Health Indicator (HI), aiming to detect defects in large, slow-rotating axial bearings, by performing variational inference using a variational autoencoder (VAE) and a conditional variational autoencoder (CVAE). Within the framework, coefficients from autoregressive (AR) models were used for both vibration and acoustic emission measurements. The proposed method is proven to be effective using both vibration and acoustic emission (AE) measurements. Using vibration measurements, as opposed to acoustic emission data, allows the proposed method to be cost-effective. In contrast, the previous work of dataset 1 was not able to reveal any degrada-

tion of the bearing using vibration measurements. The HI calculated from AE data in dataset 2 was less consistent with the applied damage. However, the experimental design may have had an impact on the calculated HI, in particular at 60 rpm. In both datasets, the proposed method was able to uncover and quantify a significant change in machine operation through the HI. The possibility to calibrate the HI to a desired level of Probability of False Alarm (PFA) allows the alarm setting to adapt to the criticality of the equipment.

Challenges of detecting defects on axial, large bearings at low speeds were discussed in this study. The effectiveness of the proposed method for axial bearing fault detection at low speeds is validated by data from 2 test rigs. As the proposed method does not rely on detection of fault frequencies, changes in machine operation can be detected regardless of failure mode and fault location. In future studies, the methodology can be extended to include other types of feature input, such as time series data. The effect of the network hyperparameters on reconstruction error, latent variable distribution and HI sensitivity should be investigated along with evaluating generalization performance on other applications. The HI is capable of capturing changes in the condition of the axial bearing, so a logical next step is to incorporate it in prognostics and remaining useful life estimation.

### REFERENCES

- [1] N. Tandon and A. Choudhury, "A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings," *Tribol. Int.*, vol. 32, no. 8, pp. 469–480, Aug. 1999.
- [2] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—A tutorial," *Mech. Syst. Signal Process.*, vol. 25, no. 2, pp. 485–520, 2011.
- [3] M. Cerrada, R.-V. Sánchez, C. Li, F. Pacheco, D. Cabrera, J. V. de Oliveira, and R. E. Vásquez, "A review on data-driven fault severity assessment in rolling bearings," *Mech. Syst. Signal Process.*, vol. 99, pp. 169–196, Jan. 2018.
- [4] Z. Stamboliska, E. Rusiński, and P. Moczko, *Proactive Condition Monitoring of Low-Speed Machines*, 1st ed. Cham, Switzerland: Springer, 2014, doi: 10.1007/978-3-319-10494-2.
- [5] E. Bechhoefer, R. Schlanbusch, and T. I. Waag, "Fault detection on large slow bearings," in *Proc. PHME*, vol. 7, 2016, pp. 1–8.
- [6] Y. Pan, R. Hong, J. Chen, Z. Qin, and Y. Feng, "Incipient fault detection of wind turbine large-size slewing bearing based on circular domain," *Measurement*, vol. 137, pp. 130–142, Apr. 2019.

- [7] N. Jamaludin and D. Mba, "Monitoring extremely slow rolling element bearings: Part I," *NDT E Int.*, vol. 35, no. 6, pp. 349–358, Sep. 2002.
- [8] N. Jamaludin and D. Mba, "Monitoring extremely slow rolling element bearings: Part II," *NDT E Int.*, vol. 35, no. 6, pp. 359–366, Sep. 2002.
- [9] B. Van Hecke, J. Yoon, and D. He, "Low speed bearing fault diagnosis using acoustic emission sensors," *Appl. Acoust.*, vol. 105, pp. 35–44, Apr. 2016.
- [10] W. Caesarendra, B. Kosasih, A. K. Tieu, H. Zhu, C. A. S. Moodie, and Q. Zhu, "Acoustic emission-based condition monitoring methods: Review and application for low speed slew bearing," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 134–159, May 2016.
- [11] Z. Mo, J. Wang, H. Zhang, X. Zeng, H. Liu, and Q. Miao, "Vibration and acoustics emission based methods in low-speed bearing condition monitoring," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Chongqing)*, Oct. 2018, pp. 871–875.
- [12] D. Abboud, J. Antoni, S. Sieg-Zieba, and M. Eltabach, "Envelope analysis of rotating machine vibrations in variable speed conditions: A comprehensive treatment," *Mech. Syst. Signal Process.*, vol. 84, pp. 200–226, Feb. 2017.
- [13] E. Bechhoefer, R. Schlanbusch, and T. I. Waag, "Techniques for large, slow bearing fault detection," *Int. J. Prognostics Health Manage.*, vol. 7, no. 1, pp. 1–12, 2016.
- [14] S. Kass, A. Raad, and J. Antoni, "Self-running bearing diagnosis based on scalar indicator using fast order frequency spectral coherence," *Measurement*, vol. 138, pp. 467–484, May 2019.
- [15] C. Mishra, A. K. Samantaray, and G. Chakraborty, "Rolling element bearing fault diagnosis under slow speed operation using wavelet de-noising," *Measurement*, vol. 103, pp. 77–86, Jun. 2017.
- [16] H. Nguyen, J. Kim, and J.-M. Kim, "Optimal sub-band analysis based on the envelope power spectrum for effective fault detection in bearing under variable, low speeds," *Sensors*, vol. 18, no. 5, p. 1389, May 2018.
- [17] Q. Xiong, Y. Xu, Y. Peng, W. Zhang, Y. Li, and L. Tang, "Low-speed rolling bearing fault diagnosis based on EMD denoising and parameter estimate with alpha stable distribution," *J. Mech. Sci. Technol.*, vol. 31, no. 4, pp. 1587–1601, Apr. 2017.
- [18] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vol. 108, pp. 33–47, Aug. 2018.
- [19] M. Saimurugan, K. I. Ramachandran, V. Sugumaran, and N. R. Sakthivel, "Multi component fault diagnosis of rotational mechanical system based on decision tree and support vector machine," *Expert Syst. Appl.*, vol. 38, no. 4, pp. 3819–3826, Apr. 2011.
- [20] L. Song, H. Wang, and P. Chen, "Vibration-based intelligent fault diagnosis for roller bearings in low-speed rotating machinery," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 8, pp. 1887–1899, Aug. 2018.
- [21] M. Kang, J. Kim, J.-M. Kim, A. C. C. Tan, E. Y. Kim, and B.-K. Choi, "Reliable fault diagnosis for low-speed bearings using individually trained support vector machines with kernel discriminative feature analysis," *IEEE Trans. Power Electron.*, vol. 30, no. 5, pp. 2786–2797, May 2015.
- [22] A. Soualhi, K. Medjaher, and N. Zerhouni, "Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 1, pp. 52–62, Jan. 2015.
- [23] J. Tian, C. Morillo, M. H. Azarian, and M. Pecht, "Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with K-Nearest neighbor distance analysis," *IEEE Trans. Ind. Electron.*, vol. 63, no. 3, pp. 1793–1803, Mar. 2016.
- [24] P. Nguyen, M. Kang, J. Kim, and J. M. Kim, "Reliable fault diagnosis of low-speed bearing defects using a genetic algorithm," in *Proc. PRICAI Trends Artif. Intell.*, 2014, pp. 248–255.
- [25] M. Hemmer, H. Van Khang, K. Robbersmyr, T. Waag, and T. Meyer, "Fault classification of axial and radial roller bearings using transfer learning through a pretrained convolutional neural network," *Designs*, vol. 2, no. 4, p. 56, Dec. 2018.
- [26] D. Peng, Z. Liu, H. Wang, Y. Qin, and L. Jia, "A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains," *IEEE Access*, vol. 7, pp. 10278–10293, 2019.
- [27] M. M. M. Islam and J.-M. Kim, "Automated bearing fault diagnosis scheme using 2D representation of wavelet packet transform and deep convolutional neural network," *Comput. Ind.*, vol. 106, pp. 142–153, Apr. 2019.
- [28] J. Xie, G. Du, C. Shen, N. Chen, L. Chen, and Z. Zhu, "An end-to-end model based on improved adaptive deep belief network and its application to bearing fault diagnosis," *IEEE Access*, vol. 6, pp. 63584–63596, 2018.
- [29] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1693–1702, Jul. 2017.
- [30] M. Hemmer and T. I. Waag, "A comparison of acoustic emission and vibration measurements for condition monitoring of an offshore drilling machine," in *Proc. Annu. Conf. Prognostics Health Manage. Soc.*, 2017, pp. 278–285.
- [31] E. Bechhoefer and A. P. F. Bernhard, "A generalized process for optimal threshold setting in HUMS," in *Proc. IEEE Aerosp. Conf.*, 2007, pp. 1–9.
- [32] E. Bechhoefer, D. He, and P. Dempsey, "Gear health threshold setting based on a probability of false alarm," in *Proc. Annu. Conf. Prognostics Health Manage. Soc.*, J. R. Celaya, S. Saha, and A. Saxena, Eds., 2011, pp. 275–281.
- [33] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Stat. Soc. B, Methodol.*, vol. 57, no. 1, pp. 289–300, Dec. 2018.
- [34] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.
- [35] S. Tao, T. Zhang, J. Yang, X. Wang, and W. Lu, "Bearing fault diagnosis method based on stacked autoencoder and softmax regression," in *Proc. 34th Chin. Control Conf. (CCC)*, Jul. 2015, pp. 6331–6335.
- [36] H. Shao, H. Jiang, H. Zhao, and F. Wang, "A novel deep autoencoder feature learning method for rotating machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 95, pp. 187–204, Oct. 2017.
- [37] Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, and Z. Zhu, "Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery," *IEEE Access*, vol. 5, pp. 15066–15079, 2017.
- [38] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Process.*, vol. 130, pp. 377–388, Jan. 2017.
- [39] M. A. Ranzato, Y.-L. Boureau, and Y. Lecun, "Sparse feature learning for deep belief networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2008, pp. 1185–1192.
- [40] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*. [Online]. Available: <http://arxiv.org/abs/1312.6114>
- [41] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
- [42] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," 2015, *arXiv:1511.05644*. [Online]. Available: <http://arxiv.org/abs/1511.05644>
- [43] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.
- [44] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," 2017, *arXiv:1701.04862*. [Online]. Available: <http://arxiv.org/abs/1701.04862>
- [45] G. San Martin, E. López Droguet, V. Meruane, and M. das C. Moura, "Deep variational auto-encoders: A promising tool for dimensionality reduction and ball bearing elements fault diagnosis," *Struct. Health Monitor.*, vol. 18, no. 4, pp. 1092–1128, Jul. 2018.
- [46] M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2015, *arXiv:1603.04467*. [Online]. Available: <https://arxiv.org/abs/1603.04467>
- [47] D. P. Kingma and J. Lei Ba, "Adam: A method for stochastic optimization," 2015, *arXiv:1412.6980*. [Online]. Available: <https://arxiv.org/abs/1412.6980>
- [48] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2017, *arXiv:1711.05101*. [Online]. Available: <http://arxiv.org/abs/1711.05101>
- [49] N. Srivastava, G. Hinton, A. Krizhevsky, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [50] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," 2015, *arXiv:1511.07289*. [Online]. Available: <http://arxiv.org/abs/1511.07289>
- [51] K. Gairaa, A. Khellaf, Y. Messlem, and F. Chellali, "Estimation of the daily global solar radiation based on Box–Jenkins and ANN models: A combined approach," *Renew. Sustain. Energy Rev.*, vol. 57, pp. 238–249, May 2016.
- [52] G. U. Yule, "On a method of investigating periodicities disturbed series, with special reference to Wolfer's sunspot numbers," *Philos. Trans. Roy. Soc. London. A, Containing Papers Math. Phys. Character*, vol. 226, nos. 636–646, pp. 267–298, 1927.

- [53] S. G. T. Walker, "On periodicity in series of related terms," *Proc. Roy. Soc. London*, vol. 131, no. 818, pp. 518–532, 1931.
- [54] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ, USA: Wiley, 2015.
- [55] Y. Vardanyan and M. R. Hesamzadeh, "Modeling regime switching in day-ahead market prices using Markov model," in *Proc. IEEE PES Innov. Smart Grid Technol. Conf. Eur. (ISGT-Eur.)*, Oct. 2016, pp. 1–6.



**MARTIN HEMMER** received the B.Sc. degree in mechanical engineering from the Oslo University College, in 2012, and the M.Sc. degree in mechatronics from the University of Agder, in 2014, where he is currently pursuing the Ph.D. degree in mechatronics, as a part of the SFI Offshore Mechatronics project. His project deals with condition monitoring and condition-based maintenance in offshore applications, focusing on large, as well as axial rolling element bearings rotating at low speed. His research interests include the areas of machine learning, signal processing and condition monitoring, and condition-based maintenance of rotating machinery.



**ANDREAS KLAUSEN** received the B.Sc., M.Sc., and Ph.D. degrees in mechatronics from the University of Agder, Norway, in 2013, 2015, and 2019, respectively. He is currently a Postdoctoral Research Fellow in the field of mechatronics with the University of Agder. His Ph.D. project dealt with condition monitoring of rolling element bearings operating during low and variable speed conditions. His research interests include the areas of signal processing, condition monitoring, autonomous systems, and optimization.



**HUYNH VAN KHANG** received the B.Sc. degree from the Ho Chi Minh City University of Technology, Vietnam, in 2002, and the M.Sc. degree from Pusan National University, South Korea, in 2008, and the D.Sc. (Tech.) degree from Aalto University, Finland, in 2012, formerly known as Helsinki University of Technology. He was a Postdoctoral Researcher at the Tampere University of Technology, Finland, from 2012 to 2013. He has been an Associate Professor in electrical power engineering with the University of Agder (UiA), Norway, from 2013 to 2019, where he is currently a Professor with the Department of Engineering Sciences. His research interests are condition-based maintenance, applied machine learning, advanced signal processing, as well as electrical machines and drives.



**KJELL G. ROBBERSMYR** (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in mechanical engineering from the Norwegian University of Science and Technology, Trondheim, Norway, in 1985 and 1992, respectively. He is currently a Professor and the Head of the Top Research Center Mechatronics with the University of Agder, Grimstad, Norway. His main research interests include machine design, rotating machines, condition monitoring, and vehicle crash simulations.



**TOR I. WAAG** received the M.Sc. and Ph.D. degrees in signal processing for laser light scattering from the Norwegian University of Science and Technology (NTNU), Trondheim. His background is in technical physics from NTNU. He is currently a Senior Scientist at the NORCE Norwegian Research Center. His work has been concentrated on the entire chain from sensor data via signal processing to decision support, mainly for the offshore industry in Norway. He is a member of the Society of Petroleum Engineers (SPE) and of the Norwegian Academy of Technological Sciences (NTVA). His recent activity within the SFI Offshore Mechatronics at the University of Agder has been focused on condition-based maintenance, mainly studying big, slow rotating bearings using vibration measurements, and acoustic emission.

• • •