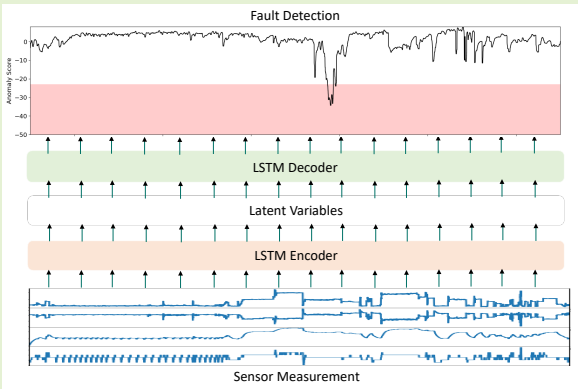
Fault Detection with LSTM-based Variational Autoencoder for Maritime Components

Peihua Han, *Student Member, IEEE*, André Listou Ellefsen, Guoyuan Li, *Senior Member, IEEE*, Finn Tore Holmeset, and Houxiang Zhang, *Senior Member, IEEE*

***Abstract*— Maintenance routines on ships today follow either a reactive maintenance (RM) or preventive maintenance (PvM) approach. RM can be regarded as post-failure repair, which might create large costs. PvM uses predetermined maintenance intervals, which often involves unnecessary maintenance. Recently, prognostics and health management (PHM) has emerged as a potential way to develop an ideal maintenance policy. PHM aims to provide optimal maintenance schedule through the use of sensor measurement for fault detection and fault prognostics, among which fault detection is the first and fundamental action. In this paper, a long-short term memory based variational autoencoder (LSTM-VAE) is proposed for fault detection of maritime components onboard. It is a semi-supervised approach that requires only fault-free data for training. Therefore, it is widely applicable in the maritime industry since operational data in normal conditions already exists. Real-world operation data collected from a diesel engine on the research vessel (RV) Gunnerus is used to validate the method. Results show that the LSTM-VAE can detect the fault accurately.**



***Index Terms*— fault detection; anomaly detection; ship autonomy; condition monitoring; prognostic and health management;**

I. Introduction

MAINTENANCE is the key to ensuring the safe and efficient operation of marine vessels. Currently, reactive maintenance and preventive maintenance are two main approaches used onboard [1]. These approaches are either cost-intensive or labor-intensive. Recently, attention has shifted to prognostics and health management (PHM), which has the greatest promise for managing maintenance operations to archive zero-downtime performance [2]. PHM systems aim to perform fault detection, fault isolation, fault identification, and remaining useful life prediction using available sensor measurements. In this way, an ideal maintenance schedule can be developed by continuously monitoring the status of the components and the evolution of their failures, which will considerably enhance operational availability and reliability as well as system safety.

Fault detection or anomaly detection is the fundamental part of any PHM system. It focuses on identifying when the current execution differs from typical successful experiences. This difference is usually caused by incipient or abrupt faults. Model-based and data-driven methods are two paradigms depending on whether a physical model is used. In the data-driven methods, the semi-supervised anomaly detection method uses only normal data for training [3]. Therefore, it is widely applicable for maritime components since recording anomalous data is costly or even dangerous in comparison to normal data [4]. Researchers often use a one-class support vector machine or isolation forest trained with non-anomalous executions. These methods have difficulty using high-dimensional sensor data, therefore, significant engineering effort is usually involved to produce low-dimensional representation. Another solution is reconstruction-based detection. A dimension reduction technique such as principal component analysis (PCA) or autoencoders (AE) is often used to compressed the data into its low-dimensional representation. Then the original data is reconstructed using the low-dimensional representation. The idea is that unforeseen patterns of anomalous data cannot be reconstructed well compared to foreseen non-anomalous data [5]. In particular, the variational autoencoder (VAE) is favored in this context because it can model the low-dimensional embedding in a probabilistic manner.

Manuscript received xx xx, xxxx; accepted xx xx, xxxx. Date of publication xx xx, xxxx; date of current version xx xx, xxxx. *(Corresponding author: Houxiang Zhang.)*

This work was supported by a grant from the Research Council of Norway through the Knowledge-Building Project for industry “Digital Twins For Vessel Life Cycle Service” (Project nr. 280703) and a grant from the Research Council of Norway through the IKTPLUSS Project “Remote Control Centre for Autonomous Ship Support” (Project nr: 309323).

Peihua Han, André Listou Ellefsen, Guoyuan Li, Finn Tore Holmeset, and Houxiang Zhang are with the Department of Ocean Operations and Civil Engineering, Norweigian University of Science and Technology (NTNU), Aalesund, Norway. (e-mail: peihua.han@ntnu.no, andre.ellefsen@ntnu.no, guoyuan.li@ntnu.no, fiho@ntnu.no, hozh@ntnu.no).

The data streams will be treated as i.i.d. in time if the above method is directly applied to time series data. Many components are subjected to rapid variations in operational loads, depending on both the task of operation and environmental conditions. Therefore, measurements that are normal in one operation condition might be anomalous in another. The temporal dependencies must be included to produce correct predictions. To address this problem, a sliding time-window is often used [6]. However, it does not represent dependencies between nearby windows and the window size is difficult to decide. Another method is to normalize the data streams based on its corresponding operation conditions [7]. This method requires prior knowledge about the operating conditions or otherwise, a clustering must be performed to approximate the operating conditions. To model the temporal dependency, a natural way is to use recurrent neural networks (RNN). We make use of an LSTM network [8], a variant of RNN, to introduce the temporal dependency into the VAE model. The LSTM is chosen since it has the ability to track long-term dependencies.

In this paper, we introduce LSTM-VAE for anomaly detection for maritime components. This model uses only the normal sequences for training. The structure of our proposed LSTM-VAE is different from the seq2seq model in [9] but similar to [5]. This structure allows us to consider long term dependencies and perform online predictions naturally. The encoder projects the multi-sensor measurement values into a latent space representation at each time step, and the decoder uses the latent space representation to reconstruct the measurement. The temporal dependencies of each time step are processed by the LSTM implemented in the encoder and decoder. The log reconstruction probability, which is the log-likelihood of the current observation given the expected distribution, is used as the anomaly score. The effectiveness of the proposed LSTM-VAE is shown through a case study involves real-world operation data collected in a maritime diesel engine. The major contributions of this paper are listed as follows:

- An semi-supervised anomaly detection method is proposed for maritime components. This method uses only normal data for training.
- The LSTM network is implemented in VAE and therefore the temporal dependencies are considered naturally.
- The comparison with the baseline method shows that the proposed method provides better performance.

The remainder of this paper is organized as follows: a introduction to model-based and data-driven fault detection is given in Section II. the proposed LSTM-VAE is introduced in Section III. The experiments are discussed in Section IV. Section V concludes the paper.

II. RELATED WORK

In the literature, fault detection methods are usually divided into two categories: model-based and data-driven, depending on whether physical models are involved.

A. Model-based fault detection

In the model-based fault detection method, faults are usually categorized into additive modes and multiplicative modes. It can be developed by monitoring the consistency between the measured outputs of the real system and the model outputs [10]. The difference can be represented as residuals and then the residuals are evaluated for fault detection. Existing techniques include the observer-based methods, parity space methods, and parameter estimation approaches. The observer-based methods adopt an observer such as Kalman filter [11], extended Kalman filter [12], particle filter [13] and etc. to estimate the system output. Then the difference between estimated and measured output is utilized to construct the residual. The residual is usually evaluated through statistical testing to determine a fault. For the parity space methods, the residual signals regarding the faults are characterized by the completely decoupling with the system initial states and presented in form of algebraic equations [10]. Odendaal and Jones [14] applied it for actuator fault detection of aircraft. Zhong et al. [15] extended the parity space approach for linear discrete time-varying system and a numerical example is given to demonstrate the application of the proposed method. As for parameter estimation approaches, the idea is to compare the normal parameters in the fault-free case with the parameters estimated by using parameter identification methods. Wang et al. [16] incorporated transient modeling and parameter identification for rotating machine fault detection. Nevertheless, the aforementioned model-based methods require a physical model, which may be unavailable for some maritime components, thus limiting the use of such methods.

B. Data-driven fault detection

Data-driven fault detection is usually related to anomaly detection, novelty detection, and outlier detection in machine learning. The technique can operate in three different modes: supervised, semi-supervised, unsupervised, depending on the available labels [3].

The fundamental idea of supervised methods is to build a binary or multi-class classifier. Existing methods include feature-based machine learning approaches and deep learning approaches. This method has been utilized for rolling bearing fault diagnostics [17], power distribution system fault detection [18], thruster failure detection [19]. Despite of their effectiveness, supervised methods require the availability of labeled instances for normal as well as faulty classes, which might not be available in most cases. Unsupervised approaches assume no labels are available in the training data set. These approaches are usually cluster-based methods and assume that normal data instances belong to large and dense clusters, while anomalies either belong to small or sparse clusters [20], [21].

Semi-supervised methods lies between supervised and unsupervised methods. It requires only the data in the normal class, therefore, it is more widely applicable than supervised methods and it is expected to be more effective than the unsupervised methods. One-class SVM has been used for this purpose and simulation has shown that this method provides satisfactory performance [22]. Chen et al. [23] used PCA for fault detection

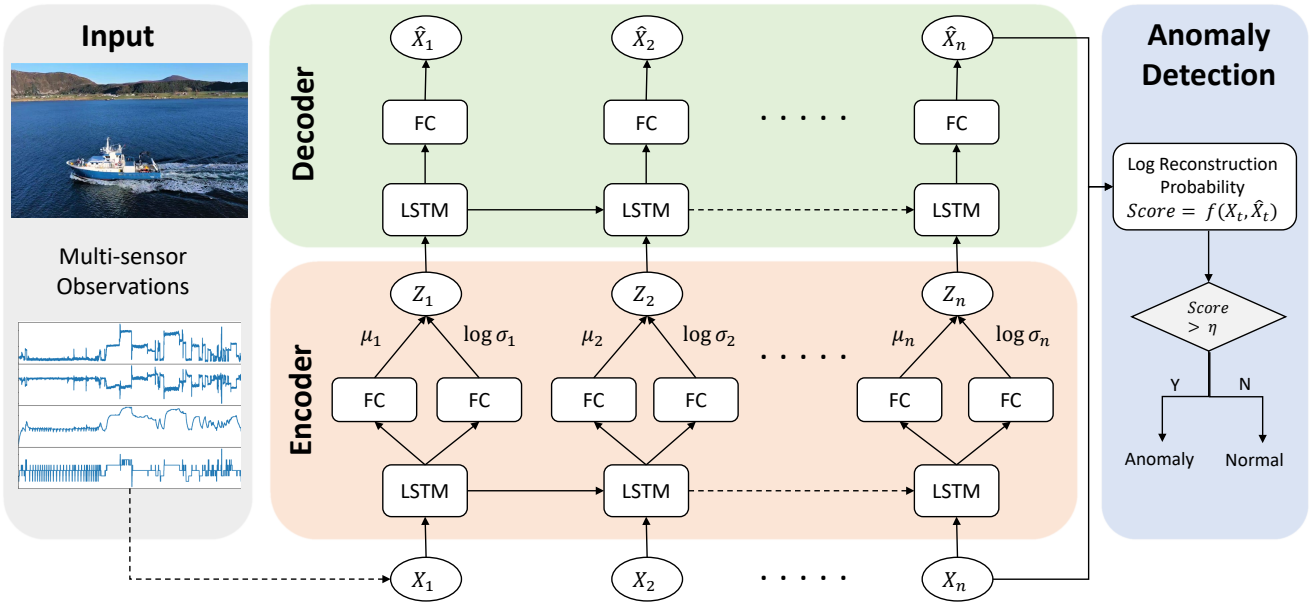


Fig. 1. Illustration of the LSTM-VAE anomaly detector unrolled in time. Note that FC is fully connected layer. The FC in encoder have Relu activation while FC in decoder have identity activation. LSTM uses tanh activation. η is the fault detection threshold.

in high-speed trains. The anomaly score is constructed in the principal component subspace and its effectiveness is demonstrated by practical experiments. Ellefsen et al. [4], [7] proposed to use VAE for anomaly detection for maritime diesel engine and the reconstruction error is used as the anomaly score. To introduce temporal dependencies, RNN or LSTM has been used in the encoder and decoder of AE and VAE. Solch et al. [24] used stochastic RNN to perform anomaly detection for robotics arms. Malhotra et al. [9] proposed a seq2seq AE model for multi-sensor anomaly detection, whose encoder and decoder are parameterized by LSTM to introduce temporal dependency. Pereira and Silveira [25] added attention module for the seq2seq VAE model for time series anomaly detection and applied it solar PV generation. Park et al. [5] combined VAE with LSTM for robot-assisted feeding system and introduced a state-based threshold. For maritime components operated on a vessel, extensive amount of normal operation data has already existed and therefore semi-supervised approaches are more applicable in the maritime domains.

III. Methodology

This section first review the variational autoencoder and the long-short term memory. Then the proposed LSTM-based variational autoencoder with the anomaly score in terms of reconstruction probability is introduced. The schematic illustration of the proposed model is shown in Fig. 1.

A. Preliminary: variational autoencoder (VAE)

The VAE is a variant of the AE rooted in Bayesian inference [26]. The VAE replaces an AE's latent representation z of given data x with stochastic variables, as shown in Fig. 2. The encoder $q_\phi(z|x)$ approximates the true posterior and the decoder $p_\theta(x|z)$ represents the likelihood of the complex process of data generation that results in the data x from z.

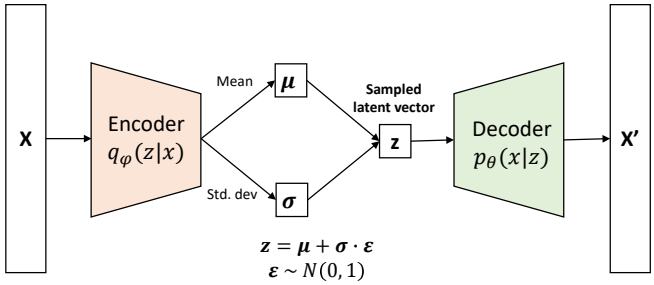


Fig. 2. A simple illustration of a VAE.

The encoder and decoder are modeled in the structure of the neural network which is parametrized by ϕ and θ, respectively. The VAE optimizes the parameters, ϕ and θ, by maximizing the lower bound of the log-likelihood.

$$\begin{aligned}\mathcal{L}_{vae} &= -D_{KL}\big(q_\phi(z|x)||p_\theta(z)\big) + E_{q_\phi(z|x)}[\log p_\theta(x|z)] \\ &\leq \log p(x)\end{aligned} \tag{1}$$

where D_{KL} is the Kullback-Leibler (KL) divergence. Minimizing D_{KL} between the approximated posterior $q_\phi(z|x)$ and the prior $p_\theta(z)$ of the latent variable regularizes the latent space. The common choice of the prior distribution $p_\theta(z)$ is a standard Gaussian distribution $\mathcal{N}(0,1)$ [26].

B. Preliminary: long-short term memory (LSTM)

LSTM is a type of recurrent neural networks (RNN). As opposed to traditional RNN, the LSTM introduces a memory cell that regulates the information flow in and out of the cell. As shown in Fig. 3, the memory cell consists of three non-linear gating units that protect and regulate the cell state. The introduction of these gating units enable easy information flow along the entire chain, therefore, the gradient vanish

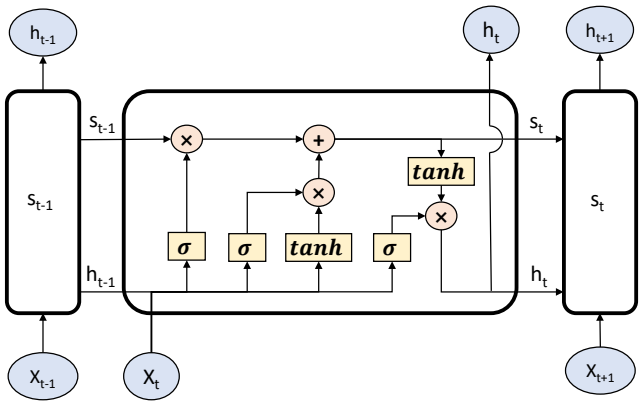


Fig. 3. Schematic illustration of a LSTM cell.

problem can be eliminated and it is able to learn long term dependencies. For each element in the input sequence, the LSTM computes the following function:

$$\begin{aligned} i_t &= \sigma(W_{ii}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi}) \\ f_t &= \sigma(W_{if}x_i + b_{if} + W_{hf}h_{t-1} + b_{hf}) \\ g_t &= \tanh(W_{ig}x_i + b_{ig} + W_{hg}h_{t-1} + b_{hg}) \\ o_t &= \sigma(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho}) \\ c_t &= f_t \odot c_{t-1} + i_t \odot g_t \\ h_t &= o_t \odot \tanh(c_t) \end{aligned} \tag{2}$$

Where h_t is the hidden state at time t, c_t is the cell state at time t, x_t is the input at time t, h_{t-1} and c_{t-1} is the hidden state and cell state at time $t-1$, respectively. i_t, f_t, g_t, o_t are the input, forget, cell and output gates, respectively. σ is the sigmoid function, where $\sigma(x) = 1/(1+e^{-x})$. $\odot$ is the Hadamard product. W and b are the weights and bias in the LSTM cell.

C. Long-short term memory based variational autoencoder

We introduce a long short-term memory-based variational autoencoder (LSTM-VAE). LSTM-VAE is a combination of VAE and LSTM. Specifically, the LSTM as in eq.(2) is used to model the encoder $q_\phi(z|x)$ and decoder $p_\theta(x|z)$ in eq.(1).

The VAE assumes that data streams are i.i.d. in time. To introduce temporal dependency for this model, we replace the feed-forward network in a VAE to LSTM. Fig. 1 shows the LSTM-VAE structure that is unrolled in time. Given a multivariate input x_t at time t, the encoder LSTM output the hidden state h_t utilizing x_t, h_{t-1}, c_{t-1}. Then h_t is feed into two linear modules to estimate the mean μ_t and log-variance $\log \sigma_z$ of the posterior $p(z_t|x_t)$. A random sample z_t from $p(z_t|x_t)$ feeds into the decoder LSTM and then a final linear module outputs the reconstructed input $\hat{x}_t$. The parameters ϕ for encoder and θ for decoder can be obtained by minimizing the loss function as follows:

$$Loss = \sum_{t=1}^{T} [D_{KL}\big(q_\phi(z_t|x_t)||p_\theta(z_t)\big) + MSE(x_t, \hat{x}_t)] \tag{3}$$

where MSE denotes mean square error, T is the length of the sequences. A standard normal distribution $\mathcal{N}(0,1)$ is used as the prior $p_\theta(z_t)$ of the latent space. Note that eq.(3) is only variation of eq.(1) since a multivariate Gaussian distribution can be assumed for continuous data and therefore maximizing the log-likelihood in eq.(1) equals minimizing the MSE in eq.(3).

D. Online anomaly detection with reconstruction probability

In autoencoders, reconstruction error is usually used as the anomaly score. Since VAE is stochastic in nature, the variability of the latent space can be taken into account. We use the reconstruction probability as the anomaly score for the proposed LSTM-VAE. The reconstruction probability is the Monte Carlo estimate of the log-likelihood $E_{q_\phi(z|x)}[\log p_\theta(x|z)]$ in (1), which can be calculated by a number of samples drawn from the latent variable distribution. Therefore the variability of the latent variable space can be taken into account, which extends its expressive power since normal data and anomaly data might share the same mean value but have different variability [27].

However, the Monte Carlo estimate requires sampling from the latent space and then forward the samples to the decoder to calculate the reconstruction probability. We implemented it in a different way by making use of a batch prediction, i.e., replicate the input by the number of samples and then perform the forward pass through the whole network. Algorithm 1 shows the pseudo-code for the online detection process using reconstruction probability.

Algorithm 1 Online anomaly detection algorithm in terms of reconstruction probability

Input: $x_t \in R^D$, s_{t-1}, n
Output: $p_\theta(x_t|\hat{x}_t)$, s_t
$\phi, \theta \leftarrow$ *load the trained LSTM-VAE model, the* ϕ, θ *is obtained using the loss function in eq.*(3)
$x_t \leftarrow$ *get current multi-sensor data*
$s_{t-1} \leftarrow$ *get the state of LSTM from previous time step*
$x_t \leftarrow Normalize(x_t)$
$\mathbf{x_t} \leftarrow Batch(x_t, n)$
$\hat{\mathbf{x}}_\mathbf{t}, s_t \leftarrow f_{\phi,\theta}(\mathbf{x}_t, s_{t-1})$, refer to eq.(2)
$\mu, \sigma \leftarrow Statistics(\hat{\mathbf{x}}_\mathbf{t})$
$p_\theta(x|\hat{x}) = p(x|\mu, \sigma)$
return $\log p_\theta(x|\hat{x})$, s_t

E. Fault detection threshold

For the fault detection threshold, a constant can be specified. The sensitivity of the detector is then controlled by assigning different constants [5]. One approach to determine such threshold is to maintain a separate validation set with fault and normal data. The threshold can be tuned based on the validation set to get the optimal fault detection accuracy or false alarm rate. However, since no fault data is available in this paper, The fault detection threshold is determined as $\mu + 3\sigma$. The mean μ and standard deviation σ is obtained

Fig. 4. R/V Gunnerus starboard side view.

from anomaly score in the validation set that contains only the normal data. The assumption is that the anomaly score should have a similar range as the validation set when the component is normal. We empirically found that this threshold provides satisfactory results in our case.

IV. Experimental Study

In this section, a maritime diesel engine operated in NTNU's research vessel Gunnerus is used to show the efficacy of the proposed method. The data collection procedure, model training, and the experimental results will be presented.

A. Data collection

The data is collected from a diesel engine operated on Norwegian University of Science and Technology's research vessel Gunnerus, as shown in Fig. 4. Gunnerus is equipped with the latest technology for a variety of research activities within biology, technology, geology, archaeology, oceanography and fisheries research. The diesel electric system of Gunnerus is used to generate electric power which is supplied to the power grid for operating the vessel. We collected the data from an entire month of November 2019. During these periods, the vessel has been sent out for several purposes such as sea trial, maneuvering courses, etc. No specific fault for the engine was found in this period. The time interval when the vessel is in operation is filtered out. In total, we got 10 days that the vessel is in operation and approximately an average of 6 hours for each day. Table I lists the sensor measurement related to the diesel engine from the logging system. The sensor data was collected at a sampling rate of 1 Hz. Fig. 5 presents the sensor measurement from a randomly selected day. It can be found that the measurement varied a lot due to the change in operational conditions.

On 21th, November 2019, we went on board to introduce a fault on this diesel engine. The air filter clogging fault was simulated using a cloth winding tape, as shown in Fig. 6. The left subgraph in Fig. 6 shows the diesel engine onboard and the right subgraph presents that the air filter is clogged with the tape. The outer surface of the air filter of the diesel engine is wrapped with a cloth winding tape. In this way, the heat dissipation and exhaust capacity of the air filter are reduced.

TABLE I
DESCRIPTIONS OF 9 SENSORS INCLUDED IN THE LOGGING SYSTEM.

Index	Sensor	Unit
1	Boost Pressure	bar
2	Engine Speed	RPM
3	Engine Exhaust Gas Temperature 1	°C
4	Engine Exhaust Gas Temperature 2	°C
5	Fuel Rate	liter/min
6	Lube Oil Pressure	bar
7	Lube oil Temperature	°C
8	Engine Power	kW
9	Cooling Water Temperature	°C

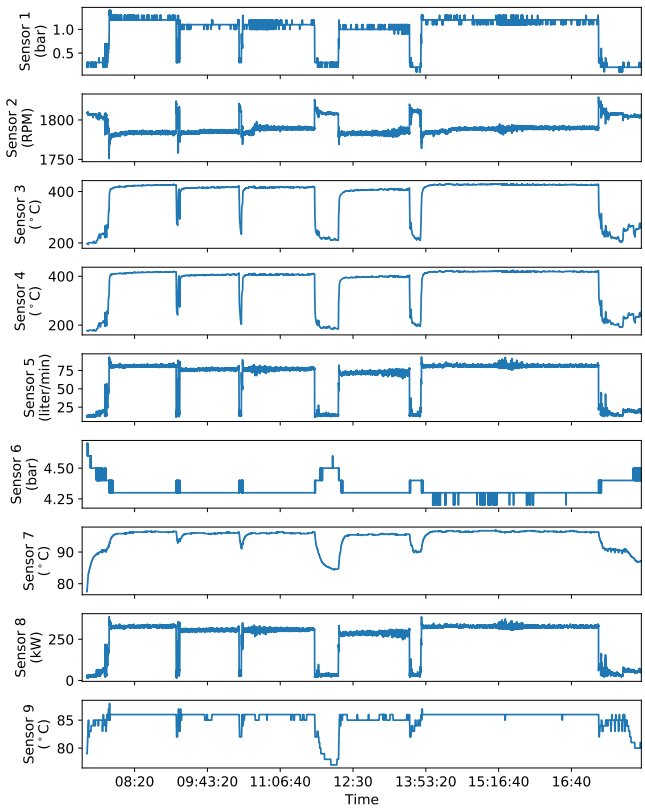


Fig. 5. Sensor measurement in the diesel engine on 8th, November, 2019.

Note that the fault introduced can be categorized as abrupt fault.

The 10 days of collected data is divided into the following three parts: (1) a test set containing 2 days of data: the day when the fault was introduced and a random normal operation day; (2) a validation set containing data for 1 of the remaining 8 days; (3) a training set containing the data for the remaining 7 days. The training set and validation set are used in the training phase of the model. The test set is used to show the efficacy of the model. Fig. 7 shows the path of the vessel operated in these 10 days. The vessel is operated around the fjord of Trondheim. The blue line indicates the training instances, the green line is the validation instances while the red line denotes the test instances. Since the training, verification, and test data come from different paths of the ship, resulting in different operating conditions, data leakage is unlikely to happen. Furthermore, the test result can reveal

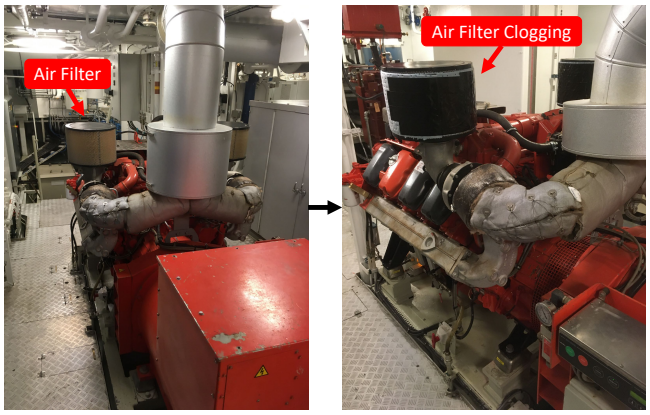


Fig. 6. Diesel engine operated in the NTNU's research vessel. The air filter is manually clogged for a period of time.

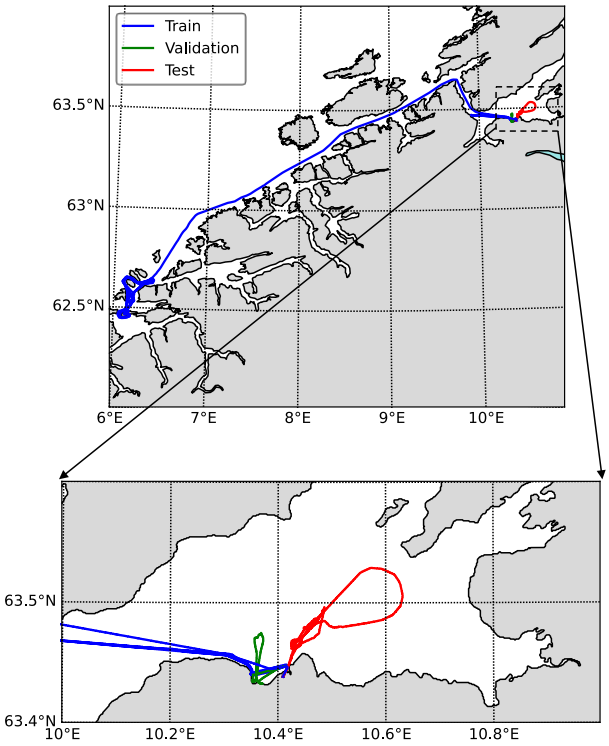


Fig. 7. Path taken by the R/V Gunnerus. The engine data on this path is used in this study.

that the proposed model can detect the fault regardless of the operation conditions.

B. Data pre-processing

Each sensor measurement in the training data sets is scaled with standard (z-score) normalization:

$$\overline{x}_n = \frac{x_n - \mu}{\sigma} \tag{4}$$

Where μ and σ is the mean and standard deviation, respectively. n refers to the sensor index. The normalization statistics obtained from training set is then applied to the validation set and test set, respectively.

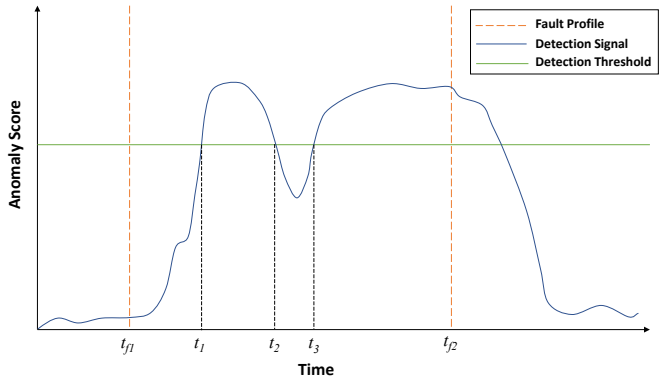


Fig. 8. Illustration of the evaluation metrics.

C. Hyperparameters for training

Both the encoder LSTM and decoder LSTM consist of two-layer LSTM with hidden units number $\{12, 8\}$. The size of the latent space is selected as 4.

The proposed LSTM-VAE is trained using the back-propagation algorithm. Since one day of data usually contains over 25,000 time steps, it is difficult to train LSTM with such time length. The training data in each day was cut into segments with 1,000 data-point using stride 10 and then the batch learning can be applied. The mini-batch size is set to be 512 and Adam [28] is used as the optimizer with a learning rate of 1×10^{-3}. The l_2 regularization term with coefficient 1×10^{-3} was used. The training process stops if the loss for the validation set with no decrease for 100 epochs. The hyperparameters are selected by trial and error and these values provide satisfactory results for the training of the proposed model.

D. Evaluation metrics

Two performance evaluation metrics, *time to detect* and *detection stability factor* [29], are used in this paper to evaluate the performance of proposed method. *Time to detect* (TTD) is defined as the period of time from the beginning of a fault injection to the moment of the first detection signal occurs. *Detection stability factor* (DSF) is the level of stability of the detection signal measured as a percentage of the sum of duration of fault detection signals to the total time elapsed after fault injection. Fig. 8 presents an simple schematic illustration of the fault detection process, where t_{f1} is the moment when the fault is introduced while t_{f2} is the time that the fault is ended. These two evaluation metrics can be calculated through Eq.(5). The *time to detect* attempts to measure how quickly the fault detection algorithms respond to faults while the *detection stability factor* measures the stability of fault detection.

$$\begin{aligned} TTD &= t_1 - t_{f1} \\ DSF &= \frac{(t_2 - t_1) + (t_{f2} - t_3)}{t_{f2} - t_{f1}} \end{aligned} \tag{5}$$

E. Baseline methods

To evaluate the performance of the proposed method, we implemented 6 baseline methods:

TABLE II
TRAINING, VALIDATION, TESTING LOSS OF DIFFERENT METHODS.

Method	Training	Validation	Testing
iForest	0.437	0.454	0.471
iForest (MRN)	0.402	0.410	0.450
AE	0.007	0.014	0.022
AE (MRN)	0.114	0.134	0.329
VAE	0.358	0.436	0.504
VAE (MRN)	0.772	0.820	1.195
LSTM-VAE	0.486	0.496	0.503

- iForest: An isolation forest based detector with standard normalization.
- iForest (MRN): An isolation forest based detector with multi-regime normalization (MRN).
- AE: An autoencoder based detector with standard normalization.
- AE (MRN): An autoencoder based detector with multi-regime normalization.
- VAE: A variational autoencoder based detector with standard normalization.
- VAE (MRN): A variational autoencoder based detector with multi-regime normalization.

For a fair comparison, the hidden units size and the latent space size for both AE and VAE are set as the same as LSTM-VAE. Multi-regime normalization [30] refers to normalize the data based on its operation conditions. Since no prior knowledge about the number of operating conditions or the way to divide the data into different operation conditions is available, we perform multi-regime normalization through three steps: (1) Calculate the coefficient of variation (CV) for different sensors and select four sensors with the highest CV; (2) Perform K-Means clustering based on these four sensors; (3) Normalize all the sensors measurements based on the cluster it belongs. The first step is to select the relevant sensor measurements to the operation condition. The operation condition is approximated by the cluster.

F. Experimental results and discussions

1) Training, validation, testing loss: Table II shows the loss for different methods. The testing loss is calculated from the day where no fault is introduced to the engine. The loss of isolation forest based anomaly detector is the anomaly score defined in [31]. The loss of AE is the reconstruction error while the loss of VAE and LSTM-VAE is the KL-divergence plus the reconstruction error. It is shown that the training, validation, testing loss are in the similar range for these methods. This indicates that the models is well trained and the training, validation, testing data is from the same distribution.

2) Qualitative results: Fig. 9 shows the qualitative comparison of our proposed LSTM-VAE with the baselines method. The fault detection algorithms are run on the test set to produce the anomaly score. The left subgraphs are from the day where a fault is introduced while the right subgraphs are from the normal operation test day. The period where the air filter clogging fault is introduced is marked with red background. The anomaly score is filtered by a median filter with kernel size 79 to remove the random spike.

The anomaly score presented in Fig. 9 is reconstruction error for AE, VAE and the anomaly score defined in [31] for iForest. For LSTM-VAE, both the reconstruction error and log reconstruction probability are provided. We perform sampling 100 times and then compute the mean reconstruction error for VAE and LSTM-VAE. From the left first three subgraphs, the fault can be only detected when the multi-regime normalization is used. When standard normalization is used, there is no distinguishable increase in anomaly score at the fault time step for AE and VAE. Even though the anomaly score is noticeable at fault time step for iForest, the score in normal operation is relatively high. The results emphasize the necessity of taking temporal dependencies or operation conditions into account to successfully detect the fault for maritime systems. The right four subgraphs show the anomaly score for one normal operation day. The scores are therefore relatively low.

The left fourth and fifth subgraphs in Fig. 9 show that the LSTM-VAE provides a similar result to iForest, AE, VAE when the multi-regime normalization is used. A clear increase in reconstruction error can be found at the fault time step. The LSTM-VAE applies directly to the standard normalized data, which makes the method easy to scale to a complex system. Generally, lots of sensors are equipped in a maritime system and complex operation conditions are involved. Performing multi-regime normalization is unrealistic in most scenarios. Even for this diesel engine, we spent lots of effort to decide the relevant sensor and the number of clusters. LSTM-VAE naturally includes the temporal dependencies and there is potential to easily scale the model in maritime systems.

From the left fourth and fifth subgraphs in Fig. 9, it is shown that the reconstruction probability provides a more noticeable change than reconstruction error. It indicates that it is beneficial to include the variety in the latent space. The reconstruction probability is expected to more expressive than the reconstruction error.

3) Quantitative results: Table III summarizes the performance of different methods in terms of time to detect (TTD) and detection stability factor (DSF). Different kernel size ω in the median filter is used to smooth the anomaly score. Only the results of the Iforest, AE, and VAE with the multi-regime normalization is shown since these models with standard normalization fail to detect the fault. It is shown that the AE, VAE, and LSTM-VAE performs better than the iForest in our case. The AE, VAE and LSTM-VAE shows a similar performance with TTD around 80 seconds and DSF around 0.78. For the LSTM-VAE, it is shown that using log reconstruction probability as anomaly score provides lower TTD as well as higher DSF than using reconstruction error. With log reconstruction probability, the LSTM-VAE can archive TTD as 60 seconds and DSF as 0.791.

V. CONCLUSION AND FUTURE WORK

In this paper, a long-short term memory based variational autoencoder (LSTM-VAE) is proposed for anomaly detection for maritime systems. The encoder and decoder of VAE are implemented with LSTM to introduce temporal dependencies.

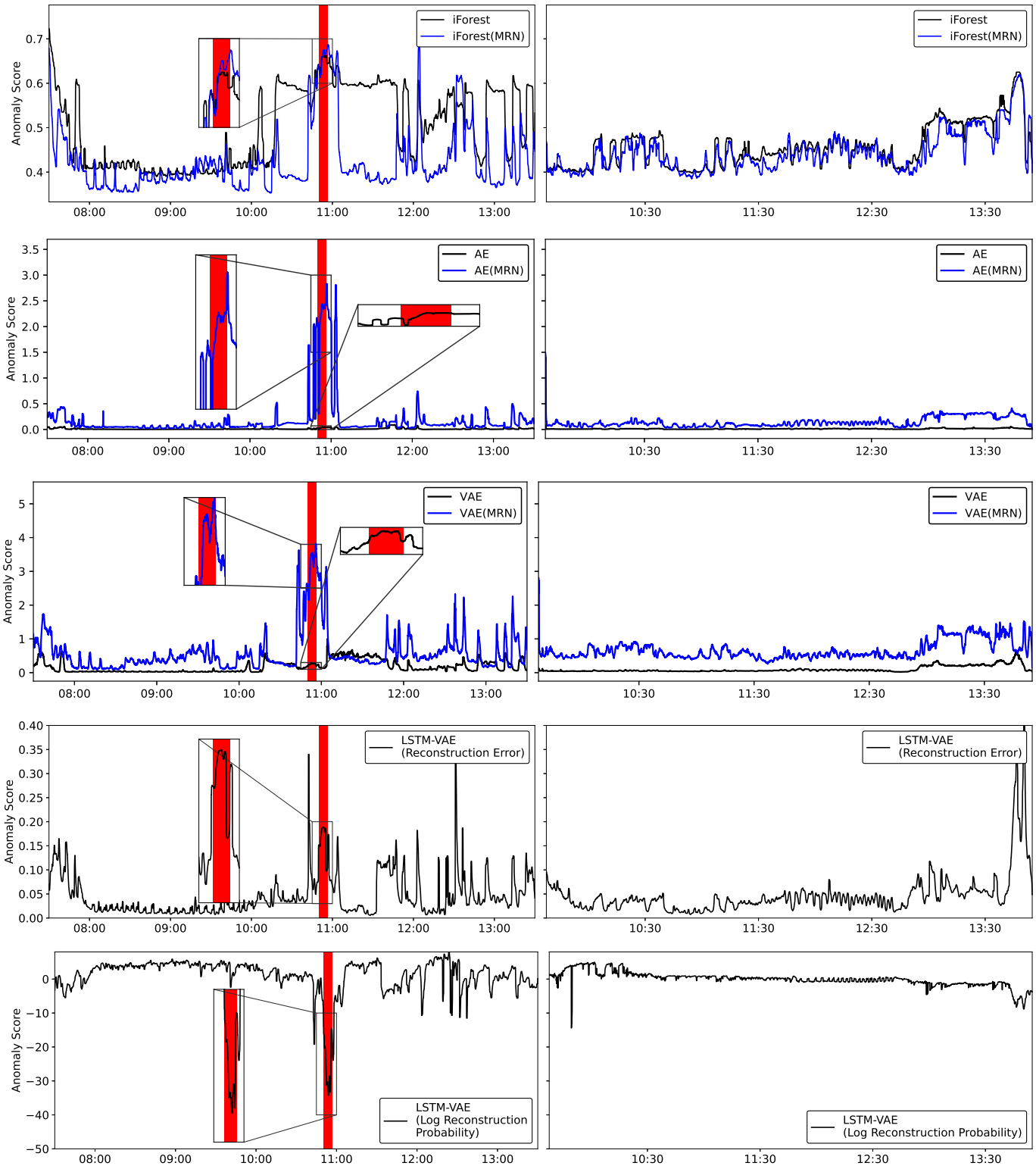


Fig. 9. Visualization of the anomaly scores over time in the test set. The left five sub graphs are the anomaly score from the day where a fault is introduced. The right five sub graphs show the anomaly score from the normal operation test day. The red background on the graphs represents the ground truth of the fault.

This method enables feasible and robust detection without further assumptions on data. In particular, no knowledge of the complex operation conditions is required to preprocess the data since the temporal dependencies are included in this method naturally. The proposed method follows the semi-supervised framework that only the data in normal operation is nessary for training. Since the underlying distribution of multi-dimensional signals is modeled and the signals are

TABLE III
COMPARISON OF DIFFERENT METHODS.

Method	ω	TTD ↓	DSF ↑
iForest (MRN)	39	163	0.553
	59	175	0.564
	79	178	0.565
	99	182	0.565
AE (MRN)	39	77	0.755
	59	78	0.781
	79	78	0.781
	99	79	0.781
VAE (MRN)	39	76	0.757
	59	80	0.749
	79	83	0.783
	99	84	0.785
LSTM-VAE (Reconstruction Error)	39	75	0.782
	59	77	0.782
	79	80	0.784
	99	83	0.782
LSTM-VAE (Log Reconstruction Probability)	39	**60**	0.774
	59	65	0.786
	79	66	**0.791**
	99	72	**0.791**

reconstructed with expected distribution information, the log reconstruction probability can be used as the anomaly score. From the experiment on a maritime diesel engine operating in the real world, we showed that the LSTM-VAE can accurately detect the air filter clogging fault and it outperforms several baseline methods in terms of two temporal metrics: time to detect and detection stability factor.

Our current work only considers two temporal metrics to evaluate the proposed method. In future work, more data including different fault types in different operation conditions will be collected, which can enable a comprehensive evaluation of the performance of the proposed method with static metrics such as false positive rate. The next step is to develop fault isolation, fault identification as well as remaining useful life prediction to establish a PHM system.

REFERENCES

[1] K. Knutsen, G. Manno, and B. Vartdal, "Beyond condition monitoring in the maritime industry," *DNV GL Strategic Research & Inovation Position Paper*, 2014.

[2] A. L. Ellefsen, V. Æsøy, S. Ushakov, and H. Zhang, "A comprehensive survey of prognostics and health management based on deep learning for autonomous ships," *IEEE Transactions on Reliability*, vol. 68, no. 2, pp. 720–740, 2019.

[3] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, pp. 1–58, 2009.

[4] A. L. Ellefsen, E. Bjørlykhaug, V. Æsøy, and H. Zhang, "An unsupervised reconstruction-based fault detection algorithm for maritime components," *IEEE Access*, vol. 7, pp. 16 101–16 109, 2019.

[5] D. Park, Y. Hoshi, and C. C. Kemp, "A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1544–1551, 2018.

[6] K. Noda, H. Arie, Y. Suga, and T. Ogata, "Multimodal integration learning of robot behavior using deep neural networks," *Robotics and Autonomous Systems*, vol. 62, no. 6, pp. 721–736, 2014.

[7] A. L. Ellefsen, P. Han, X. Cheng, F. T. Holmeset, V. Æsøy, and H. Zhang, "Online fault detection in autonomous ferries: Using fault-type independent spectral anomaly detection," *IEEE Transactions on Instrumentation and Measurement*, 2020.

[8] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[9] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, "Lstm-based encoder-decoder for multi-sensor anomaly detection," *arXiv preprint arXiv:1607.00148*, 2016.

[10] S. X. Ding, *Model-based fault diagnosis techniques: design schemes, algorithms, and tools*. Springer Science & Business Media, 2008.

[11] S. Zhao and B. Huang, "Iterative residual generator for fault detection with linear time-invariant state–space models," *IEEE Transactions on Automatic Control*, vol. 62, no. 10, pp. 5422–5428, 2017.

[12] G. H. B. Foo, X. Zhang, and D. M. Vilathgamuwa, "A sensor fault detection and isolation method in interior permanent-magnet synchronous motor drives based on an extended kalman filter," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 8, pp. 3485–3495, 2013.

[13] S. Yin and X. Zhu, "Intelligent particle filter and its application to fault detection of nonlinear system," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 6, pp. 3852–3861, 2015.

[14] H. M. Odendaal and T. Jones, "Actuator fault detection and isolation: An optimised parity space approach," *Control Engineering Practice*, vol. 26, pp. 222–232, 2014.

[15] M. Zhong, Y. Song, and S. X. Ding, "Parity space-based fault detection for linear discrete time-varying systems with unknown input," *Automatica*, vol. 59, pp. 120–126, 2015.

[16] S. Wang, W. Huang, and Z. Zhu, "Transient modeling and parameter identification based on wavelet and correlation filtering for rotating machine fault diagnosis," *Mechanical systems and signal processing*, vol. 25, no. 4, pp. 1299–1320, 2011.

[17] Z. Wang, Q. Zhang, J. Xiong, M. Xiao, G. Sun, and J. He, "Fault diagnosis of a rolling bearing using wavelet packet denoising and random forests," *IEEE Sensors Journal*, vol. 17, no. 17, pp. 5581–5588, 2017.

[18] M.-F. Guo, N.-C. Yang, and W.-F. Chen, "Deep-learning-based fault classification using hilbert–huang transform and convolutional neural network in power distribution systems," *IEEE Sensors Journal*, vol. 19, no. 16, pp. 6905–6913, 2019.

[19] P. Han, G. Li, R. Skulstad, S. Skjong, and H. Zhang, "A deep learning approach to detect and isolate thruster failures for dynamically positioned vessels using motion data," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2020.

[20] J. Yu, "A support vector clustering-based probabilistic method for unsupervised fault detection and classification of complex chemical processes using unlabeled data," *AIChE Journal*, vol. 59, no. 2, pp. 407–419, 2013.

[21] S. Rajasegarar, C. Leckie, and M. Palaniswami, "Hyperspherical cluster based distributed anomaly detection in wireless sensor networks," *Journal of Parallel and Distributed Computing*, vol. 74, no. 1, pp. 1833–1847, 2014.

[22] S. Mahadevan and S. L. Shah, "Fault detection and diagnosis in process data using one-class support vector machines," *Journal of process control*, vol. 19, no. 10, pp. 1627–1639, 2009.

[23] H. Chen, B. Jiang, and N. Lu, "A newly robust fault detection and diagnosis method for high-speed trains," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 6, pp. 2198–2208, 2018.

[24] M. Sölch, J. Bayer, M. Ludersdorfer, and P. van der Smagt, "Variational inference for on-line anomaly detection in high-dimensional time series," *arXiv preprint arXiv:1602.07109*, 2016.

[25] J. Pereira and M. Silveira, "Unsupervised anomaly detection in energy time series data using variational recurrent autoencoders with attention," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2018, pp. 1275–1282.

[26] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[27] J. An and S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability," *Special Lecture on IE*, vol. 2, no. 1, 2015.

[28] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[29] T. Kurtoglu, O. J. Mengshoel, and S. Poll, "A framework for systematic benchmarking of monitoring and diagnostic systems," in *2008 International Conference on Prognostics and Health Management*. IEEE, 2008, pp. 1–13.

[30] O. Bektas, J. A. Jones, S. Sankararaman, I. Roychoudhury, and K. Goebel, "A neural network filtering approach for similarity-based remaining useful life estimation," *The International Journal of Advanced Manufacturing Technology*, vol. 101, no. 1-4, pp. 87–103, 2019.

[31] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *2008 Eighth IEEE International Conference on Data Mining*. IEEE, 2008, pp. 413–422.