Automatic Fault Detection for Marine Diesel Engine Degradation in Autonomous Ferry Crossing Operation

André Listou Ellefsen, Xu Cheng, Finn Tore Holmeset
Vilmar Æsøy and Houxiang Zhang
Department of Ocean Operations and Civil Engineering
as part of the Mechatronics Laboratory
Norwegian University of Science and Technology
Aalesund, 6009, Norway
Email: {andre.ellefsen, xu.cheng, fiho, vilmar.aesoy, hozh}@ntnu.no

Sergey Ushakov
Department of Marine Technology
Norwegian University of Science and Technology
Trondheim, 7491, Norway
Email: sergey.ushakov@ntnu.no

***Abstract*—The maritime industry generally anticipates having semi-autonomous ferries in commercial use on the west coast of Norway by the end of this decade. In order to schedule maintenance operations of critical components in a secure and cost-effective manner, a reliable prognostics and health management system is essential during autonomous operations. Any remaining useful life prediction obtained from such system should depend on an automatic fault detection algorithm. In this study, an unsupervised reconstruction-based fault detection algorithm is used to predict faults automatically in a simulated autonomous ferry crossing operation. The benefits of the algorithm are confirmed on data sets of real-operational data from a marine diesel engine collected from a hybrid power lab. During the ferry crossing operation, the engine is subjected to drastic changes in operational loads. This increases the difficulty of the algorithm to detect faults with high accuracy. Thus, to support the algorithm, three different feature selection processes on the input data is compared. The results suggest that the algorithm achieves the highest prediction accuracy when the input data is subjected to feature selection based on sensitivity analysis.**

***Index Terms*—Automatic fault detection, feature selection, marine diesel engine, prognostics and health management, variational autoencoder**

I. Introduction

Only five years ago, most people considered autonomous and semi-autonomous ships as a futuristic fantasy [1]. Today, however, this assumption has changed dramatically since inland semi-autonomous ferries will most definitely be in commercial use on the west coast of Norway by the end of this decade [2]. These ferries are intended to navigate entirely by themselves a short distance across a river or a fjord. Thus, the crew members will ideally carry out duties other than maintaining, operating, and navigating the vessels. Additionally, securing regulatory permission, support from the industry, and public approval for semi-autonomous ferries requires evidence they are at least as safe as traditional ferries [3].

Ideally, semi-autonomous ferries will transfer real-time diagnostics and prognostics information to a control center onshore to conduct analysis and schedule maintenance operations of critical systems, components, and sub-components. One of the most critical components is the marine diesel engine as it has a leading position in both propulsion and power generation [4]. Even though the navigation mission for the ferry is rather simple in theory, the marine diesel engine will be subjected to changing environmental conditions and various operational loads. Consequently, faults and failures could occur in a totally random pattern [5]. Hence, in a maintenance perspective, a prognostics and health management (PHM) system, which both include automatic fault detection and associated remaining useful life (RUL) predictions, is crucial in autonomous operations. When the RUL is predicted, the maintenance operation can be scheduled to the next appropriate port of call for the ferry [6]. Nevertheless, the RUL prediction is the available time prior to operational failure after a fault is detected within the engine. Thus, any RUL prediction should depend on an intelligent and reliable fault detection algorithm.

During the last two years, the growth of intelligent fault detection algorithms has increased drastically. Usually, the algorithms have dependent on a supervised classifier [7], [8]. In other words, the algorithms demand fault labels in the training procedure. However, due to a general lack of fault labels for critical components in the maritime industry [9], an appropriate fault detection algorithm should not depend on a supervised classifier. An alternative approach is the utilization of unsupervised reconstruction-based fault detection algorithms [10], [11]. Usually, these algorithms train a Variational Autoencoder (VAE), in an unsupervised practice, to reconstruct normal operation data. In this way, the VAE will provide a greater reconstruction error on unexpected patterns in faulty degradation data. Finally, the reconstruction error is used as an anomaly score function (ASF) before an algorithm is applied to detect faults automatically. However, in semi-autonomous ferries, the sensor measurements might differ strongly between different engine operational loads. This increases the difficulty for the VAE to construct an accurate

ASF. Thus, the input data should be subjected to a feature selection process in order to support the VAE in the demanding reconstruction process.

This paper investigates automatic fault detection for marine diesel engine degradation in a simulated autonomous ferry crossing operation. The unsupervised reconstruction-based fault detection algorithm proposed in [11], is also used in this study to predict faults automatically. The VAE is the selected reconstruction model. Two data sets of real-operational data from a marine diesel engine are used. The first data set is a simulated ferry crossing during normal operation, while the second data set is the exact same ferry crossing operation except a fault is introduced at an unknown time step. First, the VAE is trained on the normal operation data. Then, the VAE estimates an ASF by computing a reconstruction error at each time step in the second data set, namely, the faulty degradation data. In the end, the algorithm detects a fault automatically by predicting the time step with the highest acceleration in the ASF. In order to examine the need for a feature selection process to support the VAE reconstruction process, both the normal operation data and the faulty degradation data are used to create three different input dimension scenarios: all input features, feature selection based on human domain knowledge (HDK), and feature selection based on sensitivity analysis (SA). In all three scenarios, an individual reconstruction model is used due to different input dimensions.

Our on-going project intends to develop an intelligent PHM system to provide real-time decision automation for autonomous maritime operations. Currently, the project mainly consists of two parts. The first part is the development of a step-wise feature selection approach to support both diagnostics and prognostics algorithms. The second part, on the other hand, is devoted to the development of both automatic fault detection algorithms and RUL prediction algorithms. Nevertheless, in this paper, we are only focusing on automatic fault detection. This study's principal contributions are as follows:

- Three input dimension scenarios on real-operational marine diesel engine degradation data are compared.
- Feature selection processes drastically improve the accuracy of unsupervised reconstruction-based fault detection algorithms.

The overall organization of the paper is as follows. Section II introduces the essential background on the VAE and unsupervised reconstruction models. The experimental procedure, results, and discussions are elaborated in section III. Section IV concludes and finishes the paper and presents objectives for future work.

II. Background

This section introduces the essential background on the VAE and the unsupervised reconstruction models.

A. Variational autoencoder

The VAE was developed by Kingma and Welling in 2013 and models the underlying probability distribution utilizing Bayesian inference [12]. The VAE includes an encoder function $z = q_{\theta_e}(z|x)$ and a decoder function $r = p_{\theta_d}(x|z)$. Thus, compared to the traditional autoencoder [13], the VAE improves generalization since the latent variables z are stochastic in nature. The VAE objective function is to maximize the variational lower bound J_{VAE} [14]:

$$J_{VAE}(\theta_e, \theta_d) = -D_{KL}\big(q_{\theta_e}(z|x)\,||\,p_{\theta_d}(z)\big) + E_{q_{\theta_e}(z|x)}[\log\, p_{\theta_d}(x|z)] \tag{1}$$

where D_{KL} is the Kullback-Leibeler divergence. The first expression is referred to the latent loss and measures how close z match the encoder function. The second expression is the reconstruction log-likelihood and referred to the generative loss. Nevertheless, the reconstruction error needs a Monte Carlo estimate of the expectation [12]. Since this estimate is not easily differentiable, a reparameterization scheme of z is used to collect the gradients of the decoder in order to use the back-propagation algorithm [15]. First, the reparameterization scheme applies a deterministic variable such that $z = \boldsymbol{\mu} + \boldsymbol{\sigma}\varepsilon, \varepsilon \sim \mathcal{N}(0,1)$ [12]. In this way, the encoder produces vectors of both means $\boldsymbol{\mu}$ and standard deviations $\boldsymbol{\sigma}$ rather than vectors of real values. Finally, these vectors are applied as the latent vector in the decoder. A Gaussian reconstruction distribution is normally utilized in the decoder for real-valued input data. The VAE can be stacked with many hidden layers in both the encoder and decoder depending on the dimensionality of the input data. It should be noted that unsupervised pre-training should be considered for very deep VAE structures.

B. Unsupervised reconstruction models

As similar to [11], the reconstruction models in this study are also configured with three hidden layers and corresponding hidden units ($h1$,$h2$,$h3$) in the encoder and three hidden layers with corresponding hidden units ($h3$,$h2$,$h1$) in the decoder. However, due to different input dimensions, the selection process of the hidden units is based on the following experience-based formula:

$$h1 = \mathbb{Z}\big(n \cdot 1.2\big) \qquad h2 = \mathbb{Z}\left(\frac{h1}{2}\right) \qquad h3 = \mathbb{Z}\left(\frac{h2}{2}\right)$$

where n is the number of input features in the specific scenario. Consider $\boldsymbol{x}_t = [x_1 \dots x_n]_t$ as the input vector of measurements at time step t. In order to train the reconstruction models in an unsupervised practice, $\boldsymbol{x}_t$ is also utilized as the target $\boldsymbol{y}_t$ for reconstruction at each t. To measure error calculations, each reconstruction model uses a fully connected output layer where the mean squared error (MSE) is the chosen loss function:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} ||\hat{y}_i - y_i||^2 \tag{2}$$

where n is the number of input features, $\hat{y}_i$ is the i_{th} predicted measurement and y_i is the i_{th} target measurement.

Fig. 1. The small marine diesel engine included in the hybrid power lab at the Department of Ocean Operations and Civil Engineering at the Norwegian University of Science and Technology in Aalesund.

III. Experimental study

In the ensuing experimental study, Microsoft Windows 10 is the operating system, Java 8 is the programming language, "deeplearning4j" (DL4J) version 1.0.0-beta3 [16] is the deep learning library and NVIDIA GeForce GTX 1060 6 GB is the graphics processing unit used. The reconstruction models are trained and evaluated on real-operational data from a marine diesel engine.

A. Data sets

A hybrid power lab, founded by the Department of Ocean Operations and Civil Engineering at the Norwegian University of Science and Technology in Aalesund, is used to collect the data sets. The lab consists of a small marine diesel engine with a generator, a marine battery system, a marine DC switchboard with necessary power converters, and a marine automation system to control the entire process. The power produced is fed back to the power grid in order to simulate load changes in the system. The marine diesel engine is shown in Figure 1.

During the data collection process, the engine is run by an operating profile that aims to simulate a real-life autonomous ferry crossing on the west coast of Norway. First, the ferry leaves shore in a safe and constant velocity. Then, the ferry increases its velocity until a suitable velocity is reached. This velocity is kept constant before the velocity decreases safely. Finally, the ferry breaks just before it docks. The total duration of the ferry crossing is 22 minutes and 40 seconds and the complete engine operating profile is shown in Figure 2.

The engine operating profile is run both when the normal operation data and the faulty degradation data are collected. Thus, the difference between the two data sets is that a fault is introduced at an unknown time step in the faulty degradation data. Hence, the main goal is to predict the time step where the fault occurs, namely, the fault time step f_t.

The engine has both a primary and a secondary water cooling system, where the secondary cools the primary. The primary cooling is controlled internally in the engine by a bi-metal thermostatic valve, which opens at 78 °C and fully open at 90 °C. The secondary cooling is controlled by a frequency operated fan circulating air through a heat exchanger. The fault introduced is a malfunction of the fan. This results in loss of cooling efficiency in the secondary cooling system. An alarm is triggered in the marine automation system when the cooling water temperature increases 85 °C.

TABLE I
Real-operational data sets collected from the marine diesel engine.

Data set	Time (seconds)	Frequency	Time steps
Normal operation data	1360	2 Hz	2720
Faulty degradation data	1360	2 Hz	2720

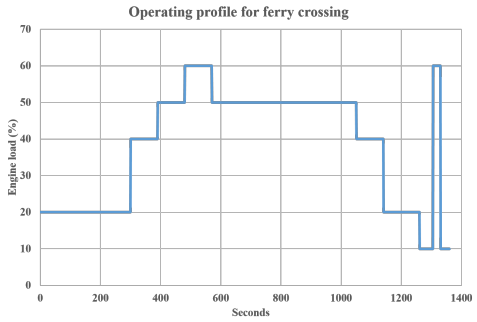

Fig. 2. Operating profile for a simulated autonomous ferry crossing.

Table I summarizes the two data sets collected. As seen in Figure 2, the engine load changes drastically throughout the ferry crossing operating profile. Thus, the sensor measurements differ strongly between the different engine loads. This affects the ability of the VAE to reconstruct an ASF with high degradation relevance. Therefore, in this study, both the normal operation data and the faulty degradation data are further used to create three different input dimension scenarios: all input features, feature selection based on HDK, and feature selection based on SA.

1) All input features: The raw data sets collected in this study includes 47 input features in total, e.g., operational loads, temperature, pressure, flow, and engine speed measurements. This scenario utilizes all 47 input features, and hence, neglects the degradation relevance for each input feature regarding the specific fault used in this study. Thus, in this scenario, the difficulty for the VAE to reconstruct an accurate ASF increases.

2) Feature selection based on human domain knowledge: In this scenario, valuable HDK is used to select degradation relevant input features concerning the specific fault. The goal of this selection process is to reduce the amount of noise in the reconstructed ASF, and hence, support the algorithm to predict f_t with higher accuracy. This selection process results in 22 input features.

3) Feature selection based on sensitivity analysis: This scenario is based on the first part of our on-going project. The step-wise feature selection approach is based on variance-based sensitivity analysis. In order to remove redundant information among the input features and reduce the computational complexity of variance-based sensitivity analysis, a Pearson

TABLE II
JOINT HYPER-PARAMETERS.

Hyper-parameter	Method/value
Optimization algorithm	Stochastic gradient descent
l_r method	Adaptive moment estimation [19]
l_r	$1 \cdot 10^{-4}$
$l2$ regularization	$1 \cdot 10^{-4}$
Weight initialization	Xavier [20]
Activation function	Rectified linear unit [21]

correlation analysis is conducted. Additionally, a surrogate model is adopted since conventional variance-based sensitivity analysis cannot be applied to the data sets directly [17], [18]. This selection process results in 12 input features.

B. Data normalization

Each input measurement x_n in the normal operation data is normalized with zero mean and unit variance normalization:

$$\hat{x}_n = \frac{x_n - \mu}{\sigma} \tag{3}$$

where μ and σ is the mean and the corresponding standard deviation of the normal operation data, respectively. Then, the normalization statistics obtained from the normal operation data are applied to the faulty degradation data.

C. Hyper-parameter configuration and training

The three reconstruction models are configured with joint hyper-parameters, as similar to [11]. Joint hyper-parameters are used in order to create reliable comparisons between the three input dimension scenarios. The selected hyper-parameters are summarized in Table II. An early stopping (ES) approach is used during the training process of each reconstruction model in order to reconstruct the normal operation data as accurately as possible. The total reconstruction error of all time steps in the normal operation data E_{nod} is monitored by the ES approach for each epoch:

$$E_{nod} = \sum_{t=1}^{T_{nod}} \left(\frac{1}{n} \sum_{i=1}^{n} ||\hat{y}_i - y_i||^2 \right)_t \tag{4}$$

where T_{nod} is the total number of time steps in the normal operation data and the second term is the MSE in Eq. 2. The training process is terminated if the number of epochs with no reduction on E_{nod} is greater than four. Finally, the reconstruction model, obtained from the epoch with the lowest E_{nod}, is used for validation on the faulty degradation data.

D. Fault prediction

Ellefsen et al. [11] used an unsupervised reconstruction-based fault detection algorithm for maritime components. Their proposed algorithm is also used in this work in order to predict f_t. First, the raw ASF is estimated by computing the MSE, Eq. 2, at each time step in the faulty degradation data. Normally, the raw ASF includes high amounts of noise. Thus, the algorithm generates three sliding windows of length w in order to smooth the ASF:

$$w = \frac{T_{fdd}}{p} \tag{5}$$

TABLE III
THE TRUE FAULT TIME STEP f_t COMPARED TO THE PREDICTED FAULT TIME STEP $\hat{f}_t$ ON THE FAULTY DEGRADATION DATA FOR EACH SCENARIO.

Scenario	n	f_t	p	w	$\hat{f}_t$
All input features	47	1979	60	45	2529
			70	39	2540
			80	34	2544
			90	30	2548
			100	27	2549
HDK	22	1979	60	45	2012
			70	39	1852
			80	34	1861
			90	30	1863
			100	27	1867
SA	12	1979	60	45	1994
			70	39	2004
			80	34	2000
			90	30	1863
			100	27	1867

where T_{fdd} is the total number of time steps in the faulty degradation data and p is a tune-able and application-dependent parameter. Next, the three sliding windows slide across the raw ASF for each time step, where a distance equivalent to w is applied between each sliding window. Then, in order to remove a certain amount of noise in the raw ASF, the average reconstruction error is calculated in the three windows. Thus, since p decides the length of w, it also decides the amount of smoothing performed on the ASF. Next, the velocity between windows 1 and 2 and between 2 and 3 are calculated. Finally, the acceleration between the two velocities is estimated. A comparison between the raw ASF and the smooth ASF for each scenario is shown in Figure 3.

According to [11], the maximum increase in sensor measurements deviations compared to typical sensor measurements in normal operation data is a clear symptom of a fault. Therefore, the maximum acceleration is used as the fault indicator since this point indicates increasing velocity, and hence, an accelerated increase in the ASF. The increasing velocity indicates that one or several feature measurements have begun to diverge from the normal operation data quickly. Thus, the algorithm detects the maximum acceleration and the corresponding fault time step $\hat{f}_t$. Please see [11], for a more comprehensive explanation of the algorithm.

E. Experimental results and discussions

Table III shows the predicted fault time step $\hat{f}_t$ for each scenario. Five different p values are used to examine the robustness of each reconstruction model. The lowest p value, however, is determined to 60 since any lower value might smooth the ASF too much, and hence, ignoring important degradation patterns. The true fault time step f_t in the faulty degradation data is determined based on the first time the cooling water temperature increases 85 °C. It should be noted that both f_t and $\hat{f}_t$ can be divided by two in order to be consistent with Figure 2.

As seen in Table III, the scenario utilizing all input features performs late $\hat{f}_t$ predictions for all p values. This scenario neglects the relevance of degradation for each input feature concerning the specific fault used in this study. Thus, as seen

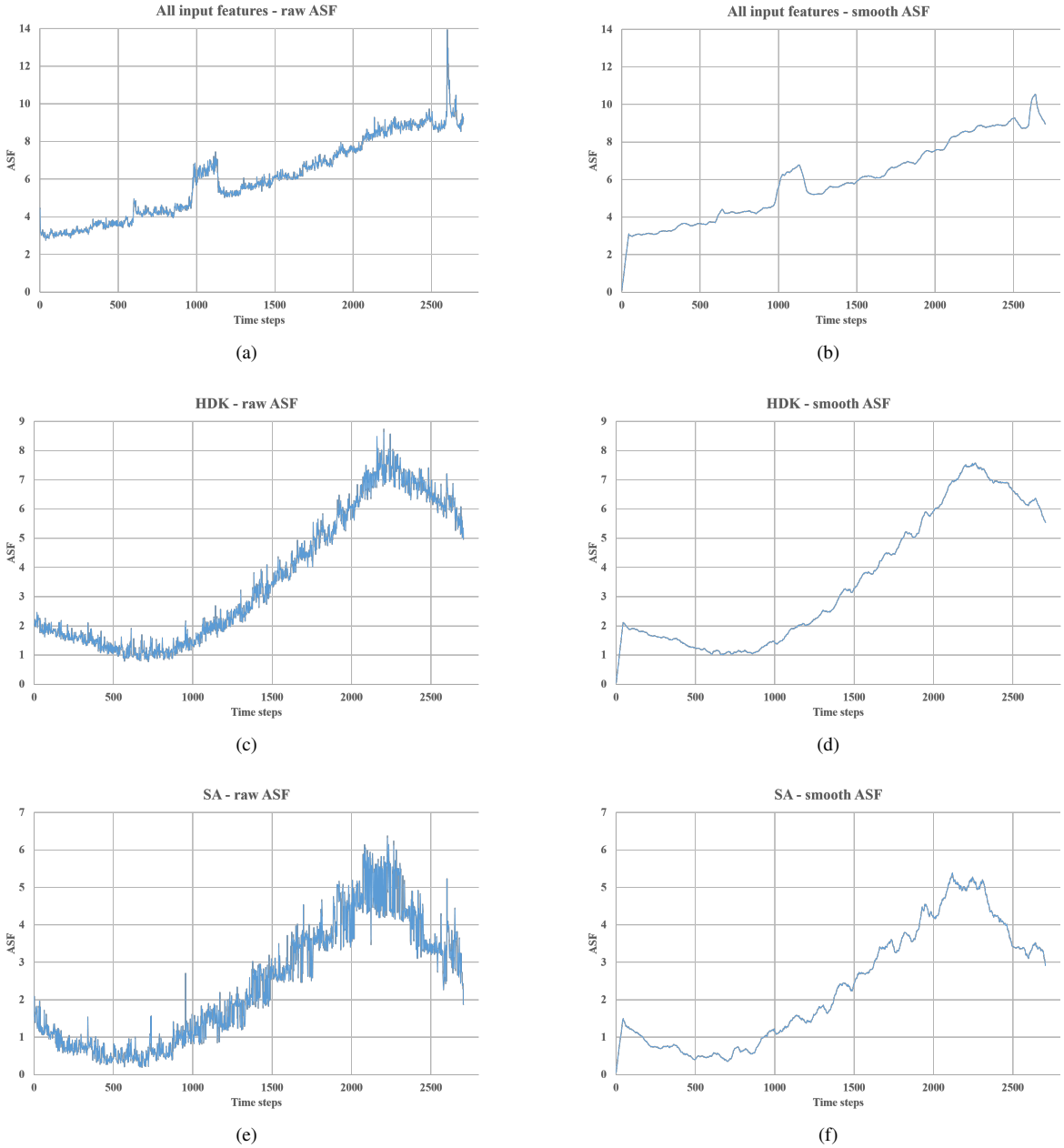

Fig. 3. ASF comparison between the three scenarios. p = 60 in the smooth ASF. (a) All input features - raw ASF. (b) All input features - smooth ASF. (c) HDK - raw ASF. (d) HDK - smooth ASF. (e) SA - raw ASF. (f) SA - smooth ASF.

in Figure 3, a spike occurs in the smooth ASF when the engine load increases rapidly in order for the ferry to break just before it docks. Hence, the algorithm detects the maximum acceleration in the breakpoint in front of the spike. The spike occurs since engine speed, redundant measurements on engine loads, and several battery measurements are included in the data sets. These features increase the difficulty for the VAE to construct an accurate ASF, especially when the engine load is above 50%. However, these features have no degradation relevance for the fault, and hence, they should be disregarded in order to remove noise for the fault detection algorithm. As opposed to the scenario utilizing all input features, the scenarios based on HDK and SA remove both irrelevant and redundant input features concerning the fault. Thus, these scenarios perform accurate $\hat{f}_t$ predictions, especially when p = 60.

The accuracy evaluations on the faulty degradation data in the three scenarios are shown in Table IV. The accuracy is defined as follows:

$$Accuracy\,(\%) = \left(1 - \frac{||\hat{f}_t - f_t||}{2720}\right) \cdot 100 \qquad (6)$$

where 2720 is the total number of time steps in the faulty

TABLE IV
ACCURACY EVALUATION ON THE FAULTY DEGRADATION DATA FOR EACH SCENARIO.

p	Accuracy (%)		
	All input features	HDK	SA
60	79.78	98.79	99.45
70	79.38	95.33	99.08
80	79.23	95.66	99.23
90	79.08	95.74	95.74
100	79.04	95.88	95.88
Avg. Accuracy	79.30	96.28	**97.88**

degradation data. As seen in Table IV, both the HDK and SA scenario perform consistent accuracy above 95% for all p values. Nevertheless, the scenario based on SA performs the highest average accuracy.

IV. CONCLUSION AND FUTURE WORK

This paper has examined automatic fault detection for marine diesel engine degradation in a simulated autonomous ferry crossing operation. An unsupervised reconstruction-based fault detection algorithm has been used to predict faults automatically. The VAE is used as the reconstruction model. Two data sets of real-operational data have been collected from a hybrid power lab including a marine diesel engine. The first data set is a simulated ferry crossing during normal operation, while the second data set is the exact same ferry crossing except a fault is introduced at an unknown time step. First, the VAE is trained on the normal operation data. Then, the VAE estimates an ASF by computing a reconstruction error at each time step in the faulty degradation data. In the end, the algorithm detects a fault automatically by predicting the time step with the highest acceleration in the ASF. Although the navigation mission for the ferry is simple, the engine is subjected to drastic changes in operational loads during the simulated ferry crossing operation. This increases the difficulty of the algorithm to detect faults with high accuracy. Thus, to support the algorithm, three different feature selection processes on the input data have been compared.

The algorithm achieved an average accuracy of 97.88% when the input data were subjected to feature selection based on SA. SA removes both irrelevant and redundant input features concerning the specific fault used in this study. Thus, drastically improving the prediction accuracy of the algorithm. However, any feature selection process might remove input features which could be of relevance for other faults with different degradation nature. Hence, introducing several other faults in the hybrid power lab will be part of future work.

ACKNOWLEDGMENT

This work was supported by the Norwegian University of Science and Technology within the Department of Ocean Operations and Civil Engineering under project no. 90329106. The authors would like to thank Digital Twins For Vessel Life Cycle Service (DigiTwin) and the Research Council of Norway, grant no. 280703.

REFERENCES

[1] E. Jokioinen, "Remote and autonomous ships - the next steps: Introduction," *Rolls-Royce, Buckingham Gate, London: The Advanced Autonomous Waterborne Applications (AAWA)*, pp. 4–14, 2016.

[2] O. Levander, "Autonomous ships on the high seas," *IEEE Spectrum*, vol. 54, no. 2, pp. 26–31, February 2017.

[3] R. Jalonen, R. Tuominen, and M. Wahlström, "Remote and autonomous ships - the next steps: Safety and security," *Rolls-Royce, Buckingham Gate, London: The Advanced Autonomous Waterborne Applications (AAWA)*, pp. 56–73, 2016.

[4] J. A. P. Rubio, F. Vera-García, J. H. Grau, J. M. Cámara, and D. A. Hernandez, "Marine diesel engine failure simulator based on thermodynamic model," *Applied Thermal Engineering*, vol. 144, pp. 982–995, 2018.

[5] T. M. Allen, "Us navy analysis of submarine maintenance data and the development of age and reliability profiles," *U.S. Navy SUBMEPP, Kittery, ME, USA, Tech. Rep.*, 2001.

[6] 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*, pp. 1–21, 2019.

[7] 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 Processing*, vol. 130, pp. 377–388, 2017.

[8] H. Liu, J. Zhou, Y. Zheng, W. Jiang, and Y. Zhang, "Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders," *ISA transactions*, vol. 77, pp. 167–178, 2018.

[9] A. S. Zymaris, Ø. Å. Alnes, K. E. Knutsen, and N. M. Kakalis, "Towards a model-based condition assessment of complex marine machinery systems using systems engineering," in *PHM Europe, Bilbao, Spain*, 2016, Conference Proceedings.

[10] 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.

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

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

[13] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[14] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[15] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient backprop," in *Neural networks: Tricks of the trade*. Springer, 2012, pp. 9–48.

[16] "Eclipse deeplearning4j development team, deeplearning4j: Open-source distributed deep learning for the jvm," *Apache Software Foundation License 2.0, http://deeplearning4j.org*, 2019.

[17] X. Cheng, S. Chen, C. Diao, M. Liu, G. Li, and H. Zhang, "Simplifying neural network based model for ship motion prediction: a comparative study of sensitivity analysis," in *ASME 2017 36th International Conference on Ocean, Offshore and Arctic Engineering*. American Society of Mechanical Engineers, 2017.

[18] X. Cheng, G. Li, R. Skulstad, S. Chen, H. P. Hildre, and H. Zhang, "A neural-network-based sensitivity analysis approach for data-driven modeling of ship motion," *IEEE Journal of Oceanic Engineering*, 2019.

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

[20] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, pp. 249–256.

[21] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, vol. 15, 2011, pp. 315–323.