Proceedings of 2019 IEEE
International Conference on Mechatronics and Automation
August 4 - 7, Tianjin, China

# A Step-wise Feature Selection Scheme for a Prognostics and Health Management System in Autonomous Ferry Crossing Operation

Xu Cheng, André Listou Ellefsen, Guoyuan Li,
Finn Tore Holmeset, Houxiang Zhang
*Department of Ocean Operations and Civil Engineering*
*Norwegian University of Science and Technology*
*Aalesund, 6009, Norway*
{*xu.cheng, andre.ellefsen, guoyuan.li, fiho, hozh*}*@ntnu.no*

Shengyong Chen
*School of Computer Science and Technology*
*Tianjin University of Technology*
*Tianjin, 300384, China*
*csy@tjut.edu.cn*

*Abstract*— **Developing a reliable algorithm to detect faults automatically within critical components in autonomous ferries is essential for safe and cost-beneficial maritime operations. Autonomous ferries are equipped with hundreds of sensors. Thus, in order to support the algorithm, the input data should be subjected to a feature selection process. This paper introduces a novel step-wise feature selection scheme for prognostics and health management (PHM) system in autonomous ferries. The scheme mainly consists of two steps. The first step is the Pearson correlation analysis to reduce the redundant information among sensors. In order to study the importance of the selected features obtained by correlation analysis and removal of irrelevant features, the second step is sensitivity analysis (SA) based feature selection. The proposed scheme is evaluated on real-operational marine diesel engine data. In the experiments, both fault classification and fault detection demonstrate the feasibility of the proposed approach.**

## I. INTRODUCTION

Today, the demand of autonomous vessels is increasing due to complex maritime operations. Ships, as an important carrier of human marine activities, need to become more intelligent to cope with harsh environmental conditions. Ship intelligence has become a key aspect for the next generation maritime and offshore industries [1].

By the end of this decade, inland semi-autonomous ferries, which were considered as a futuristic fantasy a few years ago, will almost certainly be in commercial use on the west coast of Norway [2], [3]. These ferries are going to navigate a short distance across a river or a fjord by themselves. In the ideal case, the crew members will perform duties other than maintenance, operation, and navigation. Thus, Thus, as the ferries have few or no maintenance personnel on board, a PHM system including automatic fault detection and related residual life prediction (RUL) is crucial in order to schedule maintenance operations to the next appropriate port of call [4]. The PHM system monitors and detects potential faults and predicts future operational trends, aiming to achieve the ideal maintenance policy for the vessel. With early warnings, the system can reduce downtime and avoid component damage due to unexpected failures.

Recently, more and more attentions have been paid to deep learning (DL)-based PHM systems [5], [6]. The core idea of DL-based PHM system is to develop a specific DL algorithm to detect unforeseen degradation trends in the sensor data. It is significant to develop an effective DL-based PHM system for autonomous ferry, however, there are still several challenges: 1) there are hundreds of logging points in a control system and many are coming from sensors on critical components. This makes it hard to select degradation relevant sensors for specific faults; 2) the accuracy of the data obtained from the sensors have a significant impact on the performance of the algorithm. Hence, the reliability of the sensors is crucial for the PHM system; 3) sensor data has great redundancy, and how to effectively remove redundancy is a challenge; 4) there are many uncertainties within the sensor data which are affected by a variety of factors, such as environmental conditions, operational loads, human factors, and so on.

To build a compact PHM system, feature selection methods are helpful to select the most relevant input features in order to support the DL algorithm based PHM system. Liu et al. proposed an entropy-based sensor selection approach for the aircraft engine [7]. A novel optimal feature selection approach was presented which can be a benefit for fault detection and diagnosis in smart buildings [8]. A novel intelligent fault diagnosis method was proposed for the rolling bearing based on the adaptive feature selection technique [9]. Kang et al. suggested a hybrid feature selection method to reduce diagnostic performance deterioration in data-driven PHM system [10]. Chebel-Morello et al. focused on the feature selection for fault detection with aims at reducing the influence of irrelevant features [11]. Although the feature selection methods from the literature are helpful for establishing a DL-based PHM system in autonomous

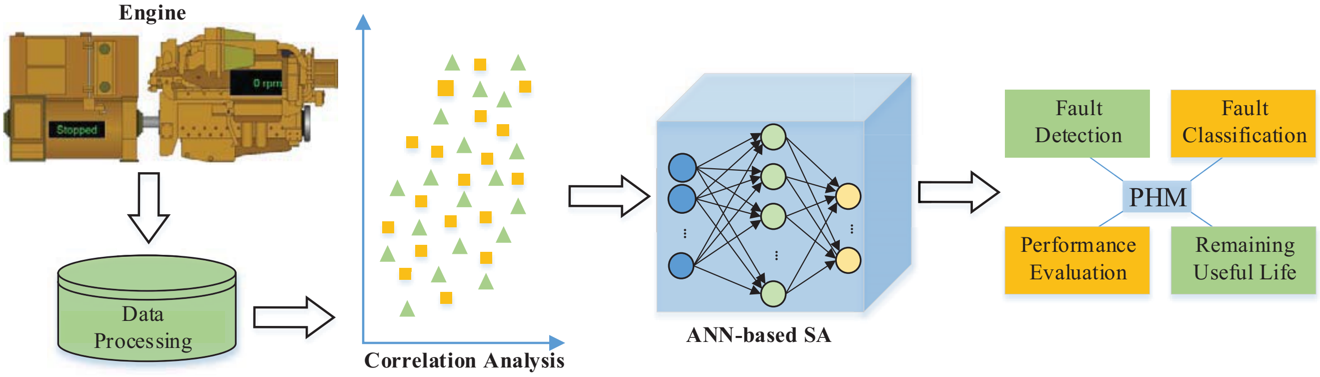


Fig. 1. Scheme of step-wise feature selection

ferries, few of them focused on how to reduce redundancy and uncertainties in the data set.

To address the above-mentioned challenges, a step-wise feature selection scheme is proposed for targeting faults of a marine diesel engine. The first step is the Pearson correlation analysis, aiming at reducing redundant information among the input features. The second step is sensitivity analysis (SA)-based feature selection. The purpose of the SA-based feature selection is to study the uncertainty and identify the importance of the selected features by Pearson correlation analysis. Our on-going project aims to develop an intelligent PHM system for planning and executing real-time support for autonomous or semi-autonomous ship operations. The on-going project mainly consists of two parts. The first part is the feature selection process to support both diagnostics and prognostics algorithms. The second part is the development of both automatic fault detection algorithms and RUL prediction algorithms. Nevertheless, in this paper, we are only focusing on the feature selection process in our on-going project. The main contributions of this paper are as follow: first, it proposes a novel step-wise scheme for the feature selection process for fault detection and fault classification of a marine diesel engine; second, the proposed scheme is evaluated on real-operational marine diesel engine data. The experiments demonstrate the feasibility of the proposed approach.

The paper is organized as follows: the proposed scheme is introduced in Section II. The experiments are discussed in Section III. Section IV concludes the paper.

## II. Step-wise feature selection scheme

This section will present the essential background on the proposed step-wise feature selection scheme. First, the overall structure of the proposed scheme is introduced. Then, the data set and corresponding data collection process are elaborated. Finally, the key components of the proposed step-wise feature selection scheme is explained in detail.

### A. Overall structure

The complete step-wise feature selection scheme is shown in Fig. 1. The first step is data pre-processing. The purpose of pre-processing is to clean the noisy sensor data as well as remove the low variance variables, which can ensure the precision of the analysis result. The second step is the Pearson correlation analysis. The autonomous ferry is equipped with thousands of sensors to measure its various conditions. The purpose of correlation analysis is to reduce the redundant information provided by the sensors. The last step is the SA, which is often used to figure out the intrinsic relationship between input and output [12]. To identify the importance and reduce the uncertainty of each parameter, the variance-based Sobol method is used. To make the conventional Sobol method be applied to the sensor data directly, Artificial Neural Network (ANN)-based meta-model is adopted. The selected features will be tested on different applications to illustrate their effectiveness.

### B. Data set description and purification

The data is collected from the hybrid power lab, founded by the Department of Ocean Operations and Civil Engineering at the Norwegian University of Science and Technology in Aalesund. A small marine diesel engine, which is equipped with a generator, a marine battery system, a marine DC switchboard with necessary power converters, and a marine automation system is used to simulate the whole process of a ferry crossing operation.

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. The total duration of the ferry crossing is 22 minutes and 40 seconds. First, the ferry is safe to leave the shore at a constant speed. Then, the ferry increases its speed until a suitable speed is reached. The speed remains unchanged before it decreases safely. In the end, the ferry breaks before docking. The raw data set collected from a marine diesel engine can contain hundreds of data-points from various sensors and controllers. Table I describes 47 sensors which are copied from the marine automation logging system.

TABLE I

DESCRIPTIONS OF 47 SENSORS INCLUDED IN THE MARINE AUTOMATION LOGGING SYSTEM.

| Index | Sensor | Unit |
|---|---|---|
| 1 | Cooling Water Into Engine | °C |
| 2 | Cooling Water Temp | °C |
| 3 | Exhaust Temp | °C |
| 4 | Cooling Water From Engine | °C |
| 5 | Lub Oil Pressure | Bar |
| 6 | Lub Oil Temp | °C |
| 7 | Fuel Oil Temp | °C |
| 8 | Engine Speed | RPM |
| 9 | Boost Pressure | Bar |
| 10 | Cooling Water Flow | Liter/min |
| 11 | 1C1 DC Voltage Actual | V |
| 12 | 1C1 Total Current | A |
| 13 | 1C1 Power | KW |
| 14 | 1C1 Active Current | A |
| 15 | 1C1 Reactive Current | A |
| 16 | 1C1 Generator Voltage | V |
| 17 | 2C1 Active Current | A |
| 18 | 2C1 Source Current | A |
| 19 | 2C1 Source Voltage | V |
| 20 | 2C1 DC Link Voltage | V |
| 21 | 3C1 DC-Link Voltage | V |
| 22 | 3C1 Total Current | A |
| 23 | 3C1 Power | KW |
| 24 | 3C1 Line Voltage | V |
| 25 | 3C1 Active Current | A |
| 26 | 3C1 Reactive Current | A |
| 27 | 3C1 Shore Voltage | V |
| 28 | 4C1 DC Link Voltage | V |
| 29 | 4C1 Line Voltage | V |
| 30 | 5C1 Motor Speed | RPM |
| 31 | 5C1 Motor Torque | Nm |
| 32 | 5C1 DC Link Voltage | V |
| 33 | ESS Bus Voltage | V |
| 34 | ESS Bus Current | A |
| 35 | ESS Maximum Cell Voltage | mV |
| 36 | ESS Minimum Cell Voltage | mV |
| 37 | ESS Pack Heartbeat | — |
| 38 | ESS Average Cell Voltage | mV |
| 39 | Cooling Water Loss | W |
| 40 | Engine Power In | KW |
| 41 | Engine Total Efficiency | % |
| 42 | Cooling Water Loss Accumulator | W |
| 43 | 1C1 Energy Accumulator | W |
| 44 | 3C1 Energy Accumulator | W |
| 45 | Radiator Fan Controller | PID |
| 46 | Fuel Meter | Liter |
| 47 | 3C1 Base Current Ref | A |

Ship sensor data is inevitably affected by a variety of factors, such as sensor errors, human factors, and so on. Therefore, to clean the sensor data is of high importance. In this paper, data purification mainly focuses on three aspects: the handling of missing values, the processing of outliners, and the reduction of data noise. If the number of missing data is less than five, a linear function would be used to fill them. Otherwise, missing data is predicted by a model which is trained by the continuous data. The outliners are removed directly in this paper. The moving average is employed to reduce the noise. Furthermore, the data is normalized by the so-called min-max normalization.

### C. Correlation analysis

The common method used for correlation analysis is the Pearson correlation analysis [13]. The definition of two random variables $X$ and $Y$ is as follows:

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \tag{1}$$

where $Cov(X,Y)$ is the covariance of $X$ and $Y$ and $Var(X)$ and $Var(Y)$ represent the variance of $X$ and $Y$, respectively.

The advantage of Pearson correlation analysis is simple and computational cheap calculations. However, its obvious disadvantage is that it can only detect linear relationships between two variables. In this paper, Pearson coefficient is applied to explore the relationship between input parameters rather than the relationship between input and output variables. Thus, the Pearson coefficient aims to remove redundant information among input variables. The reason for this is that if two input parameters exhibit a high linear relationship, we can assume that these two variable contain redundant information. As for the complex non-linear relationship between input and output, the SA based feature selection method can be employed.

### D. SA based feature selection

The variance-based Sobol method identifies the importance of the input variable with respect to output variable [14]. Thus, SA can also be used for feature selection.

As aforementioned, the ANN-based SA will be used. Assuming the ANN model form is $f(\boldsymbol{X}) = f(x_1,...,x_M)$, where $\boldsymbol{X} = (x_1,...,x_M)$ represents the model input which contains $M$ independent parameters. Based on the theory of Sobol [15], [16], [17], the output of the model can be decomposed by the two higher orders, which is shown as follows:

$$f(\boldsymbol{X}) = f_0 + \sum_{i=1}^{M} f_i(x_i) + \sum_{1\leq i\leq j\leq M} f_{ij}(x_i,x_j). \tag{2}$$

where $f_0$ represents the mean value. $f_i(x_i)$ is the effect caused by each input, and $f_{ij}(x_i,x_j)$ stands for the interaction between $i$-th input and $j$-th input. Assuming the $f(\boldsymbol{X})$ is square integrable. By squaring Eq.(2) and integrating over the input space, the following equation is obtained:

$$\int f^2(\boldsymbol{X})d\boldsymbol{X} - f_0^2 = \sum_{i=1}^{M} \int f_i^2(x_i) + \sum_{1\leq i\leq j\leq M} \int f_{ij}^2(x_i,x_j). \tag{3}$$

The total variance and partial variance can be represented by:

$$V = \int f^2(\boldsymbol{X}) d\boldsymbol{X} - f_0^2$$
$$V_i = \sum_{i=1}^{M} \int f_i^2(x_i) \tag{4}$$

The total-order sensitivity index for the $i$-th variable $x_i$ can be defined by:

$$S_{Ti} = 1 - \frac{V_{\sim i}}{V} \tag{5}$$

where $\sim i$ means all input are used except $i$-th input. In this study, the total-order sensitivity index is applied to rank the input parameters.

### E. PHM applications

The selected features will be applied to the PHM systems for autonomous ferries, such as the detection and classification of a fault, the performance evaluation and the estimation of remaining useful life [5]. In this paper, the selected features will be applied to the purpose of both fault detection and fault classification. Fault detection aims to predict a specific fault, occurring at an unknown moment during operation. By utilizing the selected features, a reliable unsupervised fault detection algorithm can be established. Fault classification, on the other hand, aims to establish a model that can separate normal operation conditions from fault conditions after a supervised training procedure is conducted. In other words, the exact moment where the fault occurs is already known.

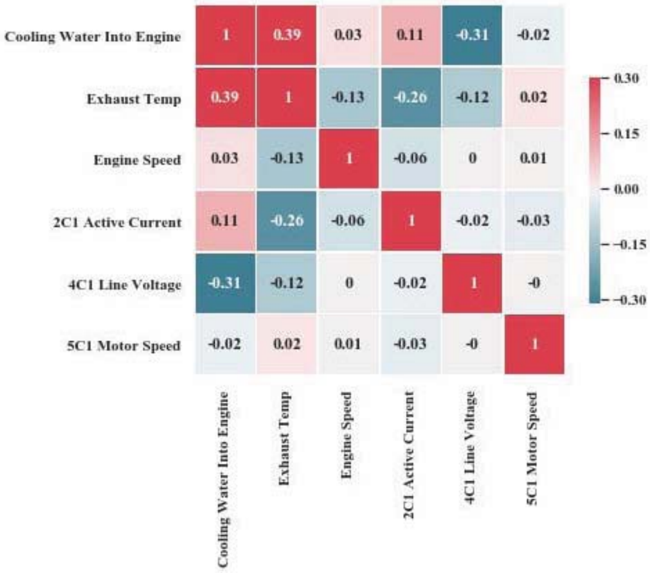
## III. Experiments

This section is devoted to the validation of the proposed approach using real-operational data from a marine diesel engine. All experiments are conducted in a computer equipped with 2.60 GHz i7-6700K CPU and 16 GB RAM.

### A. Experimental setup

In this paper, the fault introduced is a malfunction in the water cooling system. The engine is equipped with both a primary and a secondary cooling system. The purpose of the secondary cooling system is to cool the primary cooling system. The primary cooling is controlled internally in the engine by a bi-metal thermostatic valve, opening at 78°C, fully open at 90°C. The secondary cooling is cooled by a frequency operated fan circulating air through a heat exchanger. The fault introduced is a malfunction of the fan, resulting in loss of cooling efficiency in the secondary cooling system. A fault alarm is triggered in the marine automation system when the cooling water temperature increases 85°C.

To illustrate the performance of the proposed approach, two data sets are collected. The first one is the ferry crossing during normal operation, while the second one is the ferry crossing when the fault is introduced. Then, the already developed unsupervised fault detection algorithm [5] is trained on the normal operation data and employed on the faulty degradation data to detect the fault time step automatically. In other words, the algorithm estimates the time step which separates normal operation data from faulty degradation data. Finally, the estimated fault time step is used to create supervised training targets for a Long-Short Term Memory (LSTM)-based classifier. These training targets are labeled with 0 in the normal operation condition and 1 in the fault condition.



|  | Cooling Water Into Engine | Exhaust Temp | Engine Speed | 2C1 Active Current | 4C1 Line Voltage | 5C1 Motor Speed |
|---|---|---|---|---|---|---|
| Cooling Water Into Engine | 1 | 0.39 | 0.03 | 0.11 | -0.31 | -0.02 |
| Exhaust Temp | 0.39 | 1 | -0.13 | -0.26 | -0.12 | 0.02 |
| Engine Speed | 0.03 | -0.13 | 1 | -0.06 | 0 | 0.01 |
| 2C1 Active Current | 0.11 | -0.26 | -0.06 | 1 | -0.02 | -0.03 |
| 4C1 Line Voltage | -0.31 | -0.12 | 0 | -0.02 | 1 | -0 |
| 5C1 Motor Speed | -0.02 | 0.02 | 0.01 | -0.03 | -0 | 1 |

Fig. 2. Correlation analysis of input parameters.

### B. Results of correlation analysis

The raw data sets contain almost 500 features. After the removal of the low variance variables, 47 features, which are described in Table I, are left for further analysis.

19 features are selected by the correlation analysis when the threshold is set to 0.9. The selected variables which will be prepared for analysis of the SA-based feature selection. Fig. 2 shows the Pearson correlation coefficients of six selected parameters. In Fig. 2, it is clear to see all the coefficients in the Pearson matrix are very small. In this case, we can say the redundant information has been removed.

### C. Results of SA-based feature selection

To perform the SA-based feature selection, a five-layer ANN with corresponding hidden nodes 32, 16, and 4, respectively, in the hidden layers is constructed for the Sobol method. To obtain the ANN surrogate model with good performance, 80% of samples are used as the training data set, while the left 20% are used as validation data set. The output of the ANN is the cooling water temperature.

For simplicity, in our analysis, all parameters are assumed independent and the uniform distribution for each parameter.

The variation of each input variable is defined in a certain range within the maximum and minimum value.

In this paper, the threshold for selecting the variables is set to 0.1, that is, those parameters whose sensitivity index is below 0.1 will be removed. From Fig. 3, the 7 variables with the smallest sensitivity index are removed. These seven parameters are: “Engine Speed”, “1C1 Generator Voltage”, “2C1 Active Current”, “2C1 Source Current”, “2C1 Source Voltage”, “5C1 Motor Torque”, and “Engine Total Efficiency”.

It seems reasonable for SA-based feature selection that engine speed, motor torque, generator power, current, and voltage, etc. are not directly related to the cooling water temperature, intuitively. On the other hand, the cooling water into engine, exhaust temp are obviously influential to the cooling water temperature. Finally, the number of input parameters obtained by the proposed scheme is 12.

### *D. Verification of the proposed scheme*

In order to verify the effectiveness, the features selected by our step-wise feature selection scheme are compared to all input features, features selected by human domain knowledge (HDK), Joint Mutual Information (JMI) [18], [19], and Random Forest (RF) [20]. As described above, the number of all input features is 47, the number of features selected by HDK is 22, and the number of features selected by JMI, RF, and SA is 12. The algorithm, which is proposed in our previous paper [5], is used for unsupervised fault detection. Then, a one-layer LSTM with 100 hidden nodes and a softmax classifier is trained on the constructed training targets and used for fault classification, namely, classifying normal operation conditions and fault conditions in the second data set, as described in Section III-A. To perform the classification, the second data set is further divided into 15 shorter sequences. 10 sequences are selected for training and 5 sequences are selected for validation, randomly.

In Fig. 4, it is obvious to see that the validation accuracy for fault classification of the five methods is very high and almost the same. This is due to a binary classification task, which means that it does not require too many related features to get good results. On the application of fault detection, on the other hand, the proposed step-wise feature selection scheme achieves the highest average accuracy. Table II is provided to explain the reason why the proposed scheme achieves the highest fault detection accuracy and to investigate the overlap between the selected features in our proposed scheme compared to the selected features in RF, JMI, and HDK. As seen in Table II, the bold digits indicates overlap with our proposed scheme. There are five overlapping features selected by JMI and only one overlapping feature selected by RF. However, the HDK method utilizes almost all the features which is also selected by our scheme. Thus, HDK and our scheme provides the highest fault detection accuracy since engine speed, redundant measurements on engine loads, and several battery measurements are removed. These features are highly influenced by the engine operational loads and is not relevant for the specific fault used in this study. However, these features are, in fact, selected by both RF and JMI. In Fig. 4 and Table II, the results indicate that the proposed scheme removes both irrelevant and redundant input features concerning the specific fault used in this study. Thus, drastically improving the fault detection accuracy of the algorithm compared to both RF and JMI.

TABLE II

COMPARISON OF SIMILARITY OF SELECTED PARAMETERS.

| **Method** | **Selected parameters**[*] |
|---|---|
| Proposed scheme | **1**,**3**,**5**,**9**,**14**,**20**,**24**,**28**,**30**,**31**,**32**,40 |
| RF | **1**,4,6,7,10,35,36,37,42,43,45,46 |
| JMI | **1**,**3**,4,**5**,6,7,8,**9**,10,12,**14**,18 |
| HDK | **1**,2,**3**,4,**5**,**9**,10,13,**14**,**20**,21,22,23,<br>**24**,25,26,27,**28**,29,**30**,**31**,**32** |

[*] Only the index of parameters is shown. The meaning of selected parameter can be found in Table I.

## IV. CONCLUSIONS

This paper introduces a novel step-wise feature selection scheme for a PHM system for autonomous ferries. The scheme mainly includes two parts: correlation analysis and sensitivity analysis with the aim of reducing redundant information among inputs and selecting the most relevant features for the specific fault used in this study. To verify the feasibility of the proposed approach, two experiments were conducted, namely, fault detection and fault classification. In these experiments, five feature selection methods are compared. In both experiments, the HDK-based method selected 22 features, while 12 features were selected by the proposed scheme, JMI, and RF. The experimental results indicates that the proposed approach can achieve the best accuracy in both fault detection and fault classification.

In future work, we aim to develop more domain-optimized feature selection for ship engine data. We are also working on some solutions of feature selection for ultra-high dimensional data sets at low computational cost. Introducing several other faults in the hybrid power lab will be part of future work.

## ACKNOWLEDGMENT

This work was supported in part by the project “SFI MOVE” and “Digital Twins for Vessel Life Cycle Service” (Project no.: 237929 and 280703) and in part by the National Natural Science Foundation of China (U1509207). The author Xu Cheng is financially supported by the Chinese Scholarship Council. The authors also thank the suggestions and comments from Dr. Shuai Yuan.

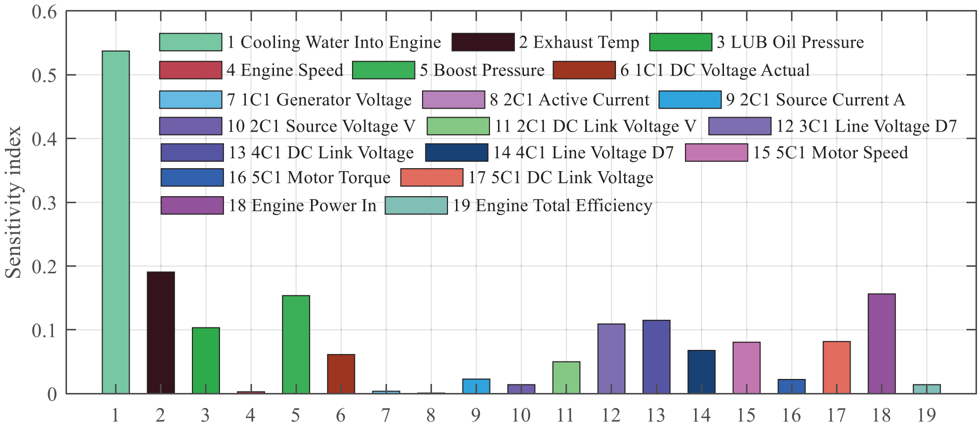


Fig. 3. SA of input parameters.

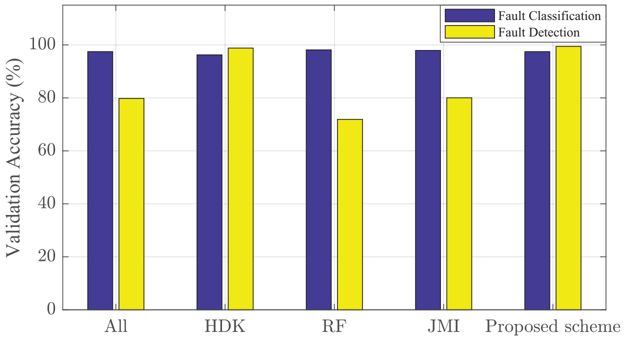


Fig. 4. Accuracy comparison for fault detection and fault classification.

## References

[1] 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.

[2] 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.

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

[4] 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," *Manuscript accepted for publication in IEEE Transactions on Reliability*, 2019.

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

[6] 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.

[7] L. Liu, S. Wang, D. Liu, Y. Zhang, and Y. Peng, "Entropy-based sensor selection for condition monitoring and prognostics of aircraft engine," *Microelectronics Reliability*, vol. 55, no. 9-10, pp. 2092–2096, 2015.

[8] D. Li, Y. Zhou, G. Hu, and C. J. Spanos, "Optimal sensor configuration and feature selection for ahu fault detection and diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1369–1380, 2017.

[9] Z. Wei, Y. Wang, S. He, and J. Bao, "A novel intelligent method for bearing fault diagnosis based on affinity propagation clustering and adaptive feature selection," *Knowledge-Based Systems*, vol. 116, pp. 1–12, 2017.

[10] M. Kang, M. R. Islam, J. Kim, J.-M. Kim, and M. Pecht, "A hybrid feature selection scheme for reducing diagnostic performance deterioration caused by outliers in data-driven diagnostics," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3299–3310, 2016.

[11] B. Chebel-Morello, S. Malinowski, and H. Senoussi, "Feature selection for fault detection systems: application to the tennessee eastman process," *Applied Intelligence*, vol. 44, no. 1, pp. 111–122, 2016.

[12] F. Fernández-Navarro, M. Carbonero-Ruz, D. B. Alonso, and M. Torres-Jiménez, "Global sensitivity estimates for neural network classifiers," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 11, pp. 2592–2604, 2017.

[13] S.-D. Bolboaca and L. Jäntschi, "Pearson versus spearman, kendall's tau correlation analysis on structure-activity relationships of biologic active compounds," *Leonardo Journal of Sciences*, vol. 5, no. 9, pp. 179–200, 2006.

[14] X. Cheng, G. Li, R. Skulstad, P. Major, S. Chen, H. P. Hildre, and H. Zhang, "Data-driven uncertainty and sensitivity analysis for ship motion modeling in offshore operations," *Ocean Engineering*, vol. 179, pp. 261–272, 2019.

[15] A. Saltelli, S. Tarantola, and K.-S. Chan, "A quantitative model-independent method for global sensitivity analysis of model output," *Technometrics*, vol. 41, no. 1, pp. 39–56, 1999.

[16] G. Li, J. Hu, S. W. Wang, P. G. Georgopoulos, J. Schoendorf, and H. Rabitz, "Random sampling-high dimensional model representation (rs-hdmr) and orthogonality of its different order component functions," *The Journal of Physical Chemistry A*, vol. 110, no. 7, pp. 2474–2485, 2006.

[17] A. Saltelli and I. M. Sobol', "Sensitivity analysis for nonlinear mathematical models: numerical experience," *Matematicheskoe Modelirovanie*, vol. 7, no. 11, pp. 16–28, 1995.

[18] G. Brown, A. Pocock, M.-J. Zhao, and M. Luján, "Conditional likelihood maximisation: a unifying framework for information theoretic feature selection," *Journal of machine learning research*, vol. 13, no. Jan, pp. 27–66, 2012.

[19] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature selection: A data perspective," *ACM Computing Surveys (CSUR)*, vol. 50, no. 6, p. 94, 2018.

[20] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.