Postprint of: Kowalski J., Krawczyk B., Woźniak M., Fault diagnosis of marine 4-stroke diesel engines using a one-vs-one extreme learning ensemble, Engineering Applications of Artificial Intelligence, Vol. 57 (2017), pp. 134-141,
DOI: 10.1016/j.engappai.2016.10.015

© 2017. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

Fault diagnosis of marine 4-stroke diesel engines using a one-vs-one extreme learning ensemble

[a] *Gdynia Maritime University, Department of Engineering Sciences, Morska Street 81-87, 81-225 Gdynia, Poland*
[b] *Department of Systems and Computer Networks, Faculty of Electronics, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland*

A B S T R A C T

This paper proposes a novel approach for intelligent fault diagnosis for stroke Diesel marine engines, which are commonly used in on-road and marine transportation. The safety and reliability of a ship's work rely strongly on the performance of such an engine; therefore, early detection of any type of failure that affects the engine is of crucial importance. Automatic diagnostic systems are of special importance because they can operate continuously in real time, thereby providing efficient monitoring of the engine's performance. We introduce a fully automatic machine learning-based system for engine fault detection. For this purpose, we monitor various signals that are emitted by the engine, and we use them as an input for a pattern classification algorithm. This action is realized by an ensemble of Extreme Learning Machines that work in a decomposition mode. Because we address 14 different faults and a correct operation mode, we must handle a 15-class problem. We tackle this task by binarization in one-vs-one mode, where each Extreme Learning Machine is trained on a pair of classes. Next, Error-Correcting Output Codes are used to reconstruct the original multi-class task. The results from experiments that were conducted on a real-life dataset demonstrate that the proposed approach delivers superior classification accuracy and a low response time in comparison with a number of state-of-the-art methods and thus is a suitable choice for a real-life implementation on board a ship.

Keywords:
Marine engine
Fault diagnosis
Fault detection
Diesel engine
Machine learning Ensemble
learning Extreme learning machines
Multi-class decomposition

1. Introduction

In maritime environments, 4-stroke Diesel engines are commonly used in on-road and marine transportation. For this reason, these types of engines are a major source of toxic emissions into the atmosphere. The harmful compounds that are emitted from Diesel engines are carbon and nitrogen oxides and unburned hydrocarbons. The deterioration of the technical state of Diesel engines will decrease their efficiency. As a result, there is higher fuel consumption and increased emissions of harmful compounds into the atmosphere. For this reason, in addition to research on improving the structures of Diesel engines, it is important to perform research on diagnostic methods for these devices. Technical diagnostics of Diesel engines is especially crucial in marine engine maintenance and operations. Based on a simple calculation, it can be estimated that a ships engine with 20 MW output power consumes almost 100 tons of fuel and emits into the atmosphere approximately 6 tons of nitric oxide (NOx) per day. The technical diagnosis of marine engines is an important and relatively difficult task for the marine engine operators. The reason is that the diagnostic signals are changed not only from changing the technical condition of the engine but from the changes in the load and/or speed of the engine.

The presented conditions tend to create automated diagnostic tools that are aimed at assisting the detection of marine equipment malfunctions. The simplest solutions for supporting the diagnostic decisions are systems of automatic signaling when the permissible operating engine parameters exceed their boundaries. Such systems allow for the prevention of sudden damage that could lead to a stoppage of the engine, but only during the engines relatively heavy load operation. The presented method is the most popular during on-board operations and maintenance. Some extensions of the mentioned method are diagnostic systems that have been proposed by ship engine manufacturers, such as the CoCoS Engine Diagnostic System of MAN or DICARE of Caterpillar (Woodyard, 2009). Both systems allow the monitoring of engine parameters, along with reporting and simple trend analysis. In each of the mentioned diagnostic systems, the decision is made by the engine room operator based on his/her knowledge and experience. It should be noted that insufficient or incorrect diagnosis of the engines technical conditions could lead to environmental risk and stoppage of the ship. In such cases, there is, in addition, no possibility of controlling the level of toxic emissions into the atmosphere and reducing the fuel consumption. A solution to this problem is the installation of additional sensors in the functional systems of the engine. One of such solutions is a diagnostic system that is based on fast deterioration of the engine crankshaft (Yang et al., 2001; Renaudin et al., 2010; Dereszewski, 2014) analysis of the boost pressure (Wu and Huang, 2011), combustion pressure (Pawletko, 2015) or acoustic emission (Pontoppidan et al., 2005). To prevent negative effects on the environment, the International Maritime Organization introduced Annex VI to the Marpol 73/78 Convention. This Annex forces ship owners to limit NO_x emissions from ship engines. According to the mentioned regulation, every on-board engine that is above 130 kW that is introduced to operation is obligated to have a valid certificate that confirms the acceptable emissions of NO_x. If ship engines are subjected some alterations during the operation period, they will have to extend the certificate. Prolonging the certificate consists of checking sets of parameters and the structural parts of the engine that influence the NO_x emission. Any changes in the design or adjustment of the engine beyond the framework established during the certification entails the need for direct measurement of the NO_x emission. Therefore, ship owners are encouraged to install systems for the direct measurement of the composition of the exhaust gases on-board. Installation of the systems to exhaust gas analysis on-board can utilize the results of the measurements for the diagnosis of marine Diesel engines. For this reason, the aim of the presented work is diagnostic signal identification in exhaust gas identification of marine 4-stroke Diesel engines. Achieving the objective requires conducting an active experiment that consists of registration of the influence of selected marine engine failures on the composition of the exhaust gas. The obtained results were used for the selection of the diagnostic signals that allow for the technical diagnosis of the engine. To verify the results, a set of computer experiments was conducted that involved the classification of the results by a neural network ensemble (Woźniak, 2014) that utilized the Extreme Learning Machine (Ding et al., 2015) principles and worked in the decomposition mode. We propose to use a one-vs-one approach, where the discussed multi-class problem is divided into a number of simpler pairwise tasks, and each base classifier is trained on a simplified problem. Error-Correcting Output Codes (Dietterich and Bakiri, 1995) are then used to reconstruct the original multi-class decision from a set of binary outputs. In this way, we can achieve improved recognition accuracy by exploiting local specializations of the classifiers in the ensemble.

The main contributions of this paper are as follows:

- A novel approach for monitoring marine 4-stroke diesel engines based on a set of diagnostic signals.
- Design of effective and intelligent fault-detection system using machine learning techniques.
- Ensemble of binary Extreme Learning Machines in one-vs-one mode utilizing Error Correcting Output Codes.
- Extensive experimental results on a real-life dataset collected by the authors, which prove the high quality of the proposed ensemble-based fault detection in terms of an excellent accuracy and low response times and which allow for the potential to have an on-board implementation in marine vehicles.

The remainder of this manuscript is organized as follows. The next section describes in detail the problem of engine failure description, the proposed diagnostic signals and the considered types of engine failures. Section 3 describes the details of the proposed intelligent fault-detection system based on an ensemble of randomized neural networks, while Section 4 presents the details of the experimental study that was performed and the obtained results along with their thorough analysis. The final section presents the paper's conclusions.

2. Problem of engine failure description

In this section, we present a detailed description of the used engine model, the measured diagnostic signals and the considered types of possible faults to be detected.

2.1. Laboratory research

This study was conducted on a marine, 3-cylinder, 4- stroke, direct injection diesel engine with an inter-cooler system. The engine was loaded with a generator that was electrically connected to the water resistance and supercharged by a VTR 160 Brown-Boveri turbocharger. During the tests, the engine was fueled by diesel oil with known properties and operated at a constant speed, equal to 750 rpm. The fueling system of the engine consists of mechanically controlled Bosh type fuel pumps connected to injectors with multi-hole type nozzles. This type of engine is commonly used as a power generator or main propulsion system with a variable pitch propeller (Carlton, 2012). A total of 56 parameters of the laboratory stand, including the engine load and speed, the parameters of the turbocharger, the systems of cooling, fueling, and lubricating and the air exchange were measured. The composition of exhaust gas was also recorded using an electro-chemical gas analyzer with an infrared carbon dioxide sensor. The pressure, temperature and humidity of the air were also recorded by the laboratory equipment. All of the mentioned results were recorded with a 1-s sampling time. The injection and combustion pressure in all of the cylinders of the engine were also recorded. The scheme of the laboratory stand is presented in Fig. 1, while the most important engine parameters are presented in Table 1.

The laboratory tests consisted of the engine operation with the following faults:

- the throttling of the exhaust gas duct (two adjustments),
- the throttling of the air inlet duct (two adjustments),
- the shift of the fuel pump cam on the camshaft, which causes a delay in the fuel injection,
- the leakage of the air inlet valve,
- the leakage of the exhaust gas valve (two adjustments),
- the decrease in the opening pressure of the fuel injector,
- the increase in the opening pressure of the fuel injector,
- the chocked fuel injector,
- the discalibrated fuel injector,
- the leakage of the fuel injection pump (two adjustments).

The test procedure, the parameters of the measuring devices and the analysis of results are presented, i.e., in Kowalski (2014, 2015a, b).

2.2. Classification problem description

Any fault of the internal combustion engine causes changes in the organization of the combustion process in the engine cylinders and causes changes in the composition of the exhaust gases. Usually, the change in the fuel fraction in the combustible fuel mixture causes a change in the carbon monoxide (CO) and carbon dioxide (CO_2) emissions in the exhaust gas from the engine. The engine faults that are located in the air/exhaust gas exchange system result in changes in the amount of air supplied to the engine cylinders. The resulting effect could be a change in the oxygen (O_2) content of the exhaust gas of the engine. Furthermore, changes in temperature, pressure and time of the combustion in the cylinder results in changes in the (NO_x) content of the exhaust gas. The content of NO_x in the exhaust gas is also dependent on the humidity, temperature and pressure of the charging air. This description is very generalized and simplified; however, it presents the desirability of the use of the mentioned exhaust gas components as carriers of the diagnostic signals. It should be noted that the presented carriers of the diagnostic signals do not allow the

MOST WIEDZY Pobrano z mostwiedzy.pl

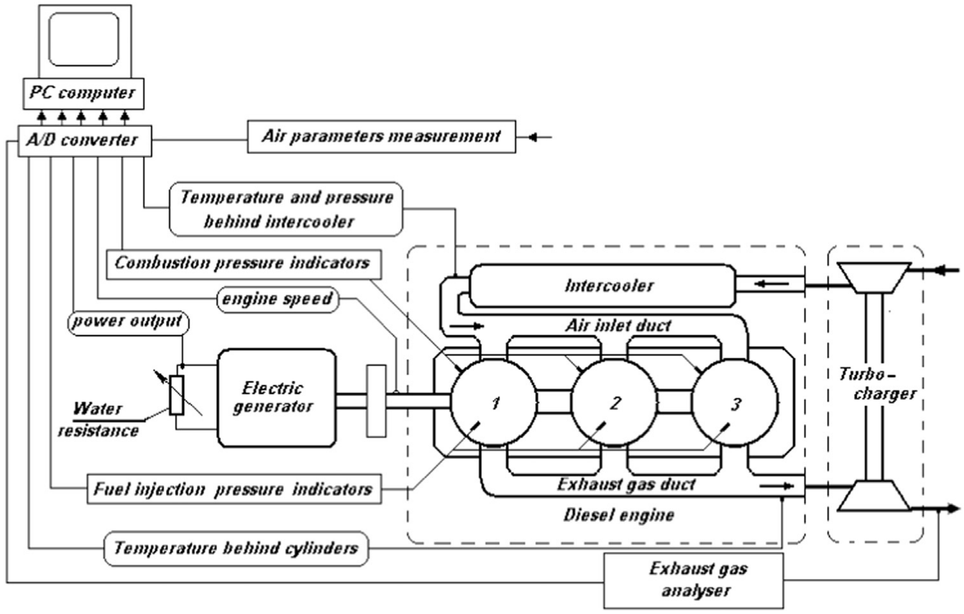

Fig. 1. The scheme of the laboratory stand (Kowalski, 2014).

Table 1
Parameters of the laboratory engine.

Parameter	Value	Unit
Max. electric power	250	kW
Rotational speed	750	rpm
Cylinder number	3	
Cylinder diameter	250	mm
Stroke	300	mm
Compression ratio	12.7	
Injector nozzle		
Holes number	9	
Holes diameter	0.325	mm
Opening pressure	25	MPa

detection of the engine faults that occur in the individual cylinders. The detection of the engine condition requires the identification of the engine cylinder in which the fault occurred. For this reason, the diagnostic signals must also come from the respective engine cylinders. An available and widely used diagnostic signal is the temperature of the exhaust gas from the engine cylinders. Accordingly, the following carriers of the diagnostic signals were selected: the mass fraction of NO_x in the exhaust gas, the corrected-to-standard atmospheric conditions,[1] the mass fractions of CO and CO_2T in the exhaust gas of the engine and the exhaust gas temperature behind all of the cylinders of the engine.

3. Classification model

In this section, we describe the details of the proposed machine learning-based fault-detection system.

3.1. Extreme learning machines

Extreme Learning Machines (ELMs) (Ding et al., 2015) are random-based single-layer feedforward neural networks that are trained in a randomized manner to reduce their computational complexity. Over the past two decades, there have been a number of significant developments in the field of neural network training algorithms (Jain et al., 2014). However, most of these approaches have suffered from the extended computational time that is required for effective execution and the large number of parameters to be set. ELMs are among the emerging trends in neural network learning that aim at alleviating the training complexities with the usage of randomly drawn weights for the neurons in the hidden layer. One must note here that despite the emerging popularity of ELMs-based approaches, this concept can in fact be traced further down in the literature to the proposals of Randomized Neural Networks (Schmidt et al., 1992) and Random Vector Functional Link (Pao et al., 1994). Let us now present the concept of ELMs. We assume that we have n labeled objects in a d-dimensional feature space and a set of M class labels at our disposal. A single-layer feedforward neural network with N hidden neurons could be described by the following equation:

$$\mathbf{y} = \sum_{i=1}^{N} \mathbf{B}_i f(\mathbf{w}_i \cdot \mathbf{x} + b_i), \tag{1}$$

where $f()$ represents the activation function, $\mathbf{x}$ is the considered object, $\mathbf{w}_i$ stands for input weights associated with i-th hidden neuron, b_i is its bias and $\mathbf{B}_i$ are weights assigned to output neurons.

When considering all of n training points, we use this equation in a matrix form:

$$\mathbf{Y} = \mathbf{HB}, \tag{2}$$

where $\mathbf{H}$ is the matrix that stores outputs of the hidden layer for each input object:

$$\mathbf{H} = \begin{pmatrix} f(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & f(\mathbf{w}_2 \cdot \mathbf{x}_1 + b_2) & \cdots & f(\mathbf{w}_N \cdot \mathbf{x}_1 + b_N) \\ f(\mathbf{w}_1 \cdot \mathbf{x}_2 + b_1) & f(\mathbf{w}_2 \cdot \mathbf{x}_2 + b_2) & \cdots & f(\mathbf{w}_N \cdot \mathbf{x}_2 + b_N) \\ \vdots & \vdots & \ddots & \vdots \\ f(\mathbf{w}_1 \cdot \mathbf{x}_n + b_1) & f(\mathbf{w}_2 \cdot \mathbf{x}_n + b_2) & \cdots & f(\mathbf{w}_N \cdot \mathbf{x}_n + b_N) \end{pmatrix}, \tag{3}$$

and $\mathbf{B} = (\mathbf{B}_1, \mathbf{B}_2, \ldots \mathbf{B}_N)^T$ and $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_n)^T$. To calculate weights assigned to outputs $\mathbf{B}$, we must compute the Moore–Penrose generalized inverse of the matrix $\mathbf{H}$, which will be denoted as $\mathbf{H}^{-1}$.

ELMs algorithm consist of three main steps:

1. Randomly generating the bias matrix $\mathbf{b} = (b_1, b_2, \ldots b_N)^T$ and weight matrix $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \ldots \mathbf{w}_N)^T$.
2. Calculating $\mathbf{H}$ using Eq. (3).
3. Calculating output weights $\mathbf{B} = \mathbf{H}^{-1}\mathbf{Y}$.

[1] ISO 8178 standardReciprocating internal combustion engines.

MOST WIEDZY Pobrano z mostwiedzy.pl

One must note that due to the random nature of the ELMs, we might find a high variance in their behavior. To counter this drawback, regularization can be used, which has been reported as having a crucial effect on the quality of the ELMs.

To regularize the ELMs, we use an orthogonal projection to obtain the Moore–Penrose pseudoinverse of $\mathbf{H}$:

$$\mathbf{H}^{-1} \cong (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T \tag{4}$$

where $\mathbf{H}^T$ is a transposed matrix $\mathbf{H}$. This approach allows us to add a ridge parameter $\frac{1}{\lambda}$ to the diagonal of $(\mathbf{H}^T\mathbf{H})$, which is known as the ridge-regression regularization approach (Buteneers et al., 2013). Applying it leads to obtaining a more stable solution.

After regularization, we can calculate the matrix of the output weights in step 3 of the ELMs training according to

$$\mathbf{B} = \left(\frac{\mathbf{I}}{\lambda} + \mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{Y}, \tag{5}$$

where $\mathbf{I}$ is an identity matrix of equal size to $\mathbf{H}$.

A main advantage of ELMs is their low computational complexity, ease of implementation, minimum hardware demands, and high speed of response. These properties make them a highly suitable tool for real-time autonomous fault diagnosis systems.

3.2. Decomposition-based ensemble architecture

ELMs have proven themselves to be one of the most efficient contemporary classifiers. However, we must remember that due to their random initialization, ELMs might be observed as unstable classifiers. Regularization reduces their variance, but there is no guarantee that the outputted model will display the highest possible performance on a given dataset. Therefore, the ensemble learning paradigm (Woźniak, 2014) has started to attract the attention of the ELMs community in recent years (Ayerdi and Graña, 2014; Cao et al., 2015).

The majority of ELMs ensemble architectures proposed in the literature attempt to take advantage of the random weight generation property of these neural networks and utilize it as a method of ensuring diversity among the committee members. Then, a voting procedure is applied to combine the individual outputs into a final compound prediction (Cao et al., 2012).

One can discuss the efficacy of such an approach with respect to the ability to both maintain diversity and provide mutually complementary classifiers for the pool. Methods based on varying the input (Jackowski et al., 2014) or output (Galar et al., 2015) spaces were reported to deliver highly efficient performance in varying multi-class scenarios, and thus, they could furnish attractive solutions to training ELM ensembles.

In our design of an intelligent fault diagnosis system, we must be aware of the scale of the problem that we are working with. There are a total of 15 different engine states that must be detected, which leads to 15 different classes to be distinguished. This problem is a complex classification problem, especially because some of the faults could have similar characteristics but have highly differing effects on the engines performance. Therefore, we require a compound and intelligent system that will be able to provide very high classification accuracy in scenarios with many classes and many potentially similar classes.

We propose to address this issue by using the divide-and-conquer approach known as multi-class decomposition (Lorena et al., 2008). This approach works on the basis of creating a set of sub-problems from the original multi-class task, where each problem is characterized by a reduced number of classes. The most popular approach used in this domain is binary decomposition, where the input dataset is transformed into a number of binary problems. In this way, we obtain a significantly simplified classification task, where each base learner is responsible for dichotomization between two classes only. This approach reduces the classifiers complexity and was proven to lead to significantly improved performance when compared with a multi-class version of the same classifier. In binary decomposition, two approaches have received the largest amount of attention: one-vs-one (OVO) and one-vs-all (OVA) (Galar et al., 2011).

In OVO (Liu et al., 2008), a multi-class problem of m classes is divided into $m(m-1)/2$ binary problems. They are formed by exhaustive pairwise selection of all possible two class combinations. Therefore, instances of the pair of classes are used to train each classifier (while ignoring the other classes). This approach leads to simplified classification boundaries and can counter some of the problems that are embedded in the nature of multi-class data, such as labeling noise or overlapping of categories.

In OVA (Rifkin and Klautau, 2004), a multi-class problem of m classes is divided into m binary problems. They are formed by selecting one class as the positive group and aggregating the remaining classes as a single negative category. This approach leads to significantly smaller ensembles than in the OVO case, but it introduces an additional difficulty in the form of an artificial class imbalance. For m classes, each base classifier will have to face a $1{:}\ m-1$ imbalance ratio, which could be harmful to its performance. Recent studies show that the OVA approach is inferior to OVO in many real-life problems (Galar et al., 2011).

Because we address a 15-class problem, the intuitive choice is to choose the OVO strategy. This choice can be justified as follows:

- ELMs can address multi-class classification, but they do not display any robustness to potential class overlapping (similarities in different error characteristics) and could highly benefit from simplification of the learned problem.
- We require the highest possible classification accuracy because even small errors in the ship engines that go undetected or wrongly recognized could lead to significant costs.
- A high number of base classifiers in OVO (105 classifiers for a 15-class task) is not a problem because ELMs are characterized by very low computational complexity and an ensemble of such a size can be efficiently implemented in any portable computational device or embedded as a hardware implementation of the engine itself.
- OVA cannot be efficiently applied because it will impair base classifiers by introducing a 1:14 imbalance ratio into each pairwise problem.

Having selected the decomposition methodology, we require a classifier combination strategy to reconstruct the multi-class problem from the individual binary outputs. Among the plethora of available solutions, we propose to use Error-Correcting Output Codes (ECOC) (Allwein et al., 2000).

An ECOC combiner can be used together with the OVO solution to efficiently choose the final class to be predicted while offering an efficient way to reduce the errors that could occur at the individual classifier level. To predict a label for a new example, the outputs of the classifiers are introduced in a code-word by mapping the positive class to +1 and the negative class to −1. Then, the code-word is compared with each row of the code-matrix (each class is represented by a row), and the most similar row is given as the output. An example of the code-matrices of OVO for a 4-class task is shown in Eq. (6), and the pseudocode of generating such a coding matrix for the OVO strategies is presented in Algorithm 1:

$$M_{ovo} = \begin{pmatrix} +1 & +1 & +1 & 0 & 0 & 0 \\ -1 & 0 & 0 & +1 & +1 & 0 \\ 0 & -1 & 0 & -1 & 0 & +1 \\ 0 & 0 & -1 & 0 & -1 & -1 \end{pmatrix} \tag{6}$$

Algorithm 1. Generating coding matrix for OVO strategies.

MOST WIEDZY Pobrano z mostwiedzy.pl

Require: m – number of classes
$M_{ovo} \leftarrow$ zero matrix with $m(m-1)/2$ columns and m rows
$k_{n-1} \leftarrow 1$
$k_n \leftarrow 0$
for i=1 **to** m **do**
  $k_n \leftarrow k_n + m - i$
  if $k_{n-1} \leq m^*(m-1)/2$ **then**
    for $j = k_{n-1}$ **to** k_n **do**
      $M_{ovo}[i, j] \leftarrow 1$
    end for
    for $k = i + 1$ **to** m **do**
      $M_{ovo}[k, k_{n-1} + k - i - 1] \leftarrow -1$
    end for
  end if
  $k_{n-1} \leftarrow k_n + 1$
end for

There are a number of decoding strategies, and each is characterized by varying the trade-off between the efficacy and the computational time. ECOCs are highly suitable for inclusion in the proposed intelligent fault-detection system due to their high efficiency, increased robustness to classifier-level mistakes and possibility of using decoding strategies with low computational complexity, which allows them to operate in real time. To fulfill these requirements, we propose to use Hamming decoding (Dietterich and Bakiri, 1995).

We present a summary of the proposed intelligent fault-detection system in pseudocode form as Algorithm 2 for the training process and as Algorithm 3 for the classification stage.

Algorithm 2. Training OVO extreme learning ensemble.

Require: m – number of classes,
$\mathcal{TR} = \{\mathcal{TR}_1, \mathcal{TR}_2, \ldots, \mathcal{TR}_m\}$ – training set, where $\mathcal{TR}_i$ stands for its subset groups examples from class i only,
$\Pi = \Psi_{1,2}, \Psi_{1,3}, \ldots, \Psi_{1,m}, \Psi_{2,3}, \ldots \Psi_{m-1,m}$ – set of binary base classifiers, where $\Psi_{i,j}$ returns 1 when class i is recognized and -1 when class j is returned
$\Pi \leftarrow \emptyset$
for i=1 **to** $m - 1$ **do**
  for $j = i + 1$ **to** m **do**
    Randomly generate bias matrix **b** and weight matrix **W** for $\mathcal{TR}_i \cup \mathcal{TR}_j$
    Calculate output matrix **H** using Eq. (3) for $\mathcal{TR}_i \cup \mathcal{TR}_j$
    Calculate $\mathbf{H}^{-1}$ using Eq. (4)
    Calculate matrix of output weights **B** using Eq. (5)
    $\Pi \leftarrow$ add $\Psi_{i,j}$
  end for
end for

Algorithm 3. Classification stage using Hamming decoding for OVO strategies.

Require: x – observation
$D \leftarrow$ zero matrix with $m(m-1)/2$ columns and m rows
$k_{n-1} \leftarrow 1$
$k_n \leftarrow 0$
for i=1 **to** m **do**
  $k_n \leftarrow k_n + m - i$
  if $k_{n-1} \leq m^*(m-1)/2$ **then**
    for $j = k_{n-1}$ **to** k_n **do**
      if $\Psi_{i,j} = 1$ **then**
        $D[i, j] \leftarrow 1$
      end if
    end for
    for $k = i + 1$ **to** m **do**
      if $\Psi_{k,kn_1+k-i-1} = 1$ **then**
        $D[k, kn - 1 + k - i - 1] \leftarrow -1$
      end if
    end for
  end if
  $k_{n-1} \leftarrow k_n + 1$
end for
recognized class
$\leftarrow \arg\max_{i \in \{1,\ldots,m\}} \left(\sum_{j=1}^{m(m-1)/2} |M_{ovo}(i, j) - D(i, j)| \right)$

4. Experimental study

To evaluate the actual potential of the proposed diagnostic system, we conducted a detailed experimental study on a real-life dataset of engine fault observations collected by the authors.

We designed the experiments to obtain answers to the following questions:

- Is it possible to obtain a high accuracy rate of marine engine diagnosis using the proposed set of features, which originated from engine measurements?
- Does the OVO-based ensemble of ELMs allow us to efficiently tackle this complex multi-class problem and offer a significant improvement over a number of state-of-the-art solutions in machine learning-based fault diagnosis?
- Is the proposed system suitable for real-life implementation on boats and is it characterized by a complexity that allows it to operate under online and real-time conditions?

The following subsections will present details that regard the used data and methods as well as the obtained results, along with a discussion.

4.1. Data

For the purpose of this experiment, we have collected a real-life dataset. It consisted of 798 observations (separate engine readings) described by 15 features, according to the framework described in Section 2. We address a 15-class problem, where one of the classes is the correct state of the engine and the 14 remaining classes represent various possible failures. The distribution of objects among these classes was roughly balanced.

4.2. Set-up

To place the quality of the proposed OVO ELMs system into context, we propose to compare it to a number of state-of-the-art machine learning classifiers. These include single-model solutions based on decision trees, neural networks and kernel classifiers as well as multiple classifier systems. The details of algorithms together with their description and range of parameters evaluated are presented in Table 2.

The following experimental framework was used during the computational experiments:

- For the performance metric in this multi-class problem, we use the average accuracy; it assigns identical weights to all of the classes, thus assuming their equal importance to the problem. It is computed as follows:

$$\text{AvgAcc} = \frac{1}{m} \sum_{i=1}^{m} \text{TPR}_i, \tag{7}$$

where TPR_i stands for true positive rate on i-th class.

- We use a 5×2 CV combined F-test (Alpaydin, 1999) for simulta-

MOST WIEDZY Pobrano z mostwiedzy.pl

Table 2
Details of classifiers used in the experimental study.

Abbr.	Description	Hyperparameters
C5.0	Efficient decision tree induction algorithm	Confidence $\in\{0.10, 0.25, 0.35\}$
		min. instances per leaf=2 Pruned=TRUE
k-NN	k nearest neighbors classifier	$k \in \{1, 3, \ldots, 9, 11\}$
		Distance=Euclidean
RIPPER	Repeated Incremental Pruning to Produce Error Reduction	Folds $\in\{2, 3, 4, 5\}$
	rule-based classifier	Optimizations $\in\{2, 3, 4, 5\}$
ELM	Extreme Learning Machine	No. of hidden neurons $\in\{10, 15, \ldots, 45, 50\}$
	randomized neural network described in Section 3.1	Activation function=sigmoid
		$\lambda \in \{2, 4, \ldots, 8, 10\}$
SVM-OVA	Support Vector Machine using RBF kernel	$\mu \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$
	with OVA decomposition and ECOC combiner	$\log_{10}(\gamma) \in \{-7, -6, \ldots, 3, 4\}$
SVM-OVO	Support Vector Machine using RBF kernel	$\mu \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$
	with OVO decomposition and ECOC combiner	$\log_{10}(\gamma) \in \{-7, -6, \ldots, 3, 4\}$
C5.0-OVO	C5.0 with OVO decomposition and ECOC combiner	Same as in C5.0
k-NN-OVO	k-nn with OVO decomposition and ECOC combiner	Same as in k-NN
RIPPER-OVO	RIPPER with OVO decomposition and ECOC combiner	Same as in RIPPER
BAGG	Bagging with C5.0 as base classifier	No. of base classifiers $\in\{10, 20, \ldots, 90, 100\}$
BOST	AdaBoost.M2 with C5.0 as base classifier	No. of iterations $\in\{5, 10, \ldots, 45, 50\}$
RANF	Random Forest ensemble	No. of base classifiers $\in\{40, 80, \ldots, 260, 300\}$
ROTF	Rotation Forest ensemble	Feature extraction=PCA No. of base classifiers $\in\{10, 20, \ldots, 90, 100\}$
VELM	Voting ensemble of ELMs	No. of base classifiers $\in\{10, 20, \ldots, 90, 100\}$
ELM-OVO	Proposed ELMs ensemble with OVO decomposition and ECOC combiner	Same as in ELM

neous training/testing and pairwise statistical analysis. It repeats five times a two-fold cross-validation. The combined F-test is conducted by a comparison of all versus all. As a test score, the probability of rejecting the null hypothesis is adopted, i.e., that the classifiers have the same error rates. As an alternative hypothesis, it is conjectured that the tested classifiers have different error rates. A small difference in the error rate implies that the different algorithms construct two similar classifiers with similar error rates; thus, the hypothesis should not be rejected. For a large difference, the classifiers have different error rates, and the hypothesis should be rejected.

- We fix the significance level $\alpha = 0.05$ for the statistical testing.
- All of the classifiers hyperparameters were optimized using the internal 3-fold CV on the training set.

All of the experiments were conducted in the R environment[2] on a machine that was equipped with a four core Intel Core i7-4700MQ Haswell @ 2.40 GHz processor and 8.00 GB of RAM.

[2] http://www.r-project.org/

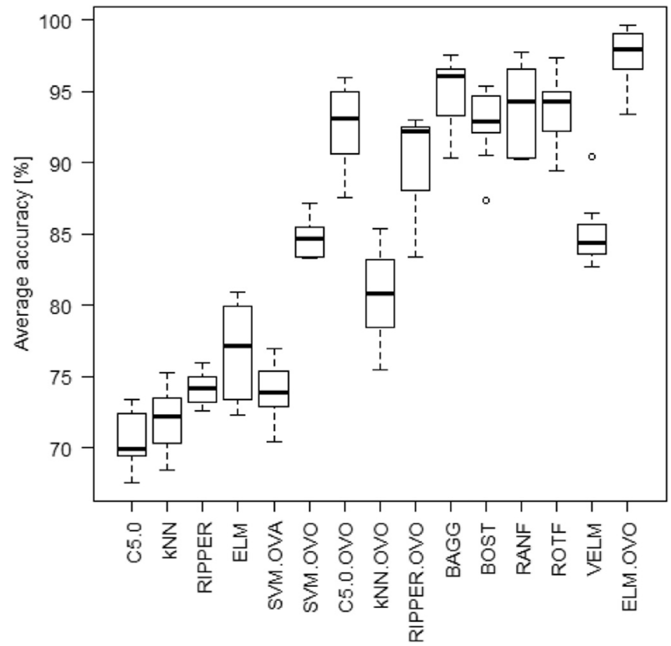

Fig. 2. Boxplot of average accuracy (%) metric obtained examined classifiers used in the proposed intelligent marine engine fault detection system.

4.3. Results and discussion

The performances of the fault diagnosis system with different base classifiers are given in the form of a boxplot in Fig. 2. The results of the pairwise combined 5×2 CV F-test and the obtained p-values are depicted in Table 3, while Table 4 presents the average training and testing times for the examined methods.

Let us now take a look into the obtained results.

When analyzing the performance of single model classifiers (C5.0, k-NN, RIPPER and ELM), it can be observed that they are not suitable to be used as a basis for a marine engine fault-detection system. All of them achieve an accuracy of below 80%, which cannot be accepted in such a demanding industrial application. Assuming that more than every fifth possible fault will be ignored and misdiagnosed imposes high potential costs (in terms of human safety, engine damage and time needed to localize the actual fault), which are prohibitive from a practical point of view. The poor performance of these methods can be explained by a high number of classes to be considered, which leads to a highly complex decision boundary and a possible occurrence of class overlapping that furthers hampers the recognition process.

The lack of SVMs in the previous group can be explained by the fact that they cannot be *de facto* treated as single models. Because they are binary in nature, when applying them to multi-class problems, one must conduct a decomposition. In this study, we have compared the effectiveness of the SVMs in the OVA and OVO scenarios. The OVA approach returns a similar performance to the single models, which proves our previous claims that this model is not suitable for the problem under consideration. Because of the high number of classes, the introduced imbalance ratio results in each base model being biased toward the negative class, which in turn results in reduced classification accuracy.

When comparing the OVO solutions using SVM, C5.0, k-NN and RIPPER, we can observe a significant gain in the classification performance compared with using their multi-class equivalents. Among these methods, k-NN is characterized by the smallest improvement, despite being the best-performing single model. This finding shows that the minimal-distance approach does not gain very much from the multi-class decomposition as in the remaining algorithms. C5.0 and RIPPER are the two best-performing classifiers in OVO, with

MOST WIEDZY Pobrano z mostwiedzy.pl

Table 3
Results of the combined 5×2 CV F-test for comparison between the proposed OVO-ELM and reference methods. Symbol '=' stands for classifiers without significant differences, '>' for situation in which the proposed method is superior and '<' vice versa.

Hypothesis	p-value	Hypothesis	p-value
ELM-OVO vs C5.0	>0.000001	ELM-OVO vs C5.0-OVO	>0.010950
ELM-OVO vs k-NN	>0.000001	ELM-OVO vs k-NN-OVO	>0.000961
ELM-OVO vs RIPPER	>0.000001	ELM-OVO vs RIPPER-OVO	>0.008526
ELM-OVO vs ELM	>0.000001	ELM-OVO vs BAGG	>0.036057
ELM-OVO vs SVM-OVA	>0.000001	ELM-OVO vs BOST	>0.021196
ELM-OVO vs SVM-OVO	>0.002456	ELM-OVO vs RANF	>0.027545
ELM-OVO vs VELM	>0.001992	ELM-OVO vs ROTF	>0.025293

Table 4
Averaged training and testing (per sample) times (s) for different examined classifiers used in the intelligent fault detection system.

Classifier	Training	Testing	Classifier	Training	Testing
C5.0	7.46	0.05	C5.0-OVO	27.98	0.54
k-NN	0.00	34.53	k-NN-OVO	0.00	67.43
RIPPER	12.94	0.07	RIPPER-OVO	40.54	0.41
ELM	27.23	0.18	BAGG	104.56	0.37
SVM-OVA	380.34	1.87	BOST	378.92	1.19
SVM-OVO	436.12	2.02	RANF	118.32	0.42
VELM	2081.45	3.65	ROTF	284.32	1.03
ELM-OVO	314.98	1.01			

C5.0 achieving slightly better recognition rates.

Multi-class ensembles can deliver very good quality and, in all cases, are similar to that of C5.0 enhanced with OVO. They achieve this result despite working with the native multi-class task. Because all of them are based on input space partitioning, we can deduce that this arrangement leads to the creation of base classifiers with individual domains of competence that can capture the complexities of the faced engine diagnosis task. Interestingly, Bagging (Breiman, 1996) returns the best performance from these four models, while usually Boosting (Schapire and Freund, 2012), Random Forest (Breiman, 2001) or Rotation Forest (Rodriguez et al., October 2006) are believed to be the most efficient methods. This finding can be explained by the relatively small dimensionality of the analyzed problem and the fact that all of the features appear to be relevant. Therefore, both forest models, by utilizing random feature reduction, in fact harm their performance. This finding shows that for the considered task, the entire feature set should be used to properly capture the class properties, while the partitioning of training objects is highly beneficial.

A voting ensemble of ELMs delivers highly unsatisfactory performance, which is much below the standard ensembles or canonical classifiers enhanced with OVO. This finding shows that relying on the role of random initialization as a good way to ensure both the diversity and individual quality fails in the considered engine monitoring problem. In the given case, we can only intuitively guess that either the feature space, parameter or object space randomization does not provide sufficient diversity for the ensembles to work well. However, it is worthwhile to note that the variance in the obtained accuracy is greatly reduced compared to the single ELM.

Finally, the proposed ensemble architecture that utilizes both the ELMs and OVO strategy delivers excellent performance. With error rates of approximately 2%, the proposed system is highly suitable for implementation in real-life scenarios, to provide highly reliable engine state monitoring. This approach took advantage of the decomposition approach because each base ELM was trained on a significantly reduced problem. This input reduction leads to, at the same time, diversification of the base learners and simplification of each individual learning process. The Hamming-based ECOC allowed us to efficiently reconstruct the original multi-class task from the binary outputs. Similar to with VELM, the variance of this model is significantly smaller than in the case of a single ELM. This finding shows that OVO can contribute highly to the ELMs performance when addressing complex multi-class scenarios. The combined F-test (see Table 3) proves that the proposed ELM-OVO solution outperforms all of the other methods in a statistically significant manner.

Because our aim is to develop an online fault diagnosis system, it is required to work in real time. Therefore, in addition to the accuracy, we must also investigate the time complexities of the methods. While from a practical point of view the testing time is of crucial value (because it shows how fast the system can react to changes in the engines state), we also include the training times for completeness. From the results presented in Table 4, it can be observed that the k-NN method cannot be used in this application. Despite having no training phase (this method is a so-called lazy classifier (Aha, 1997)), its average response time per sample is prohibitory. In the case of a rapid accident, a system based on such a learner would report it a few minutes later, which is unacceptable. The remaining classifiers offer acceptable response times. When accounting for the training times, we can see that the VELM ensemble stands out from the others. This finding is due to its sequential structure, where we must train a high number of ELMs one after another.

We can see that the proposed fault diagnosis system based on ELM-OVO offers a very good response time, roughly one second, which is fully suitable for the considered real-life application. Its training time is slightly higher than that of the other ensemble methods because it always needs to train 105 classifiers (one per each pair of classes). However, when comparing the training times of the single ELM and ELM-OVO, it can be observed that OVO leads to a significant reduction in the training time per model, due to the simpler decision space that each of them handles. While the obtained training time is satisfactory, one must remember that it is not a crucial factor in the fault diagnosis system design. In practice, each such ensemble will be trained before being applied to the engine, and thus, we are interested in only the response time. If an adaptive fault diagnosis system design is wanted, then online ELMs (Shao and Er, 2016) can be used because they are characterized by very fast re-training from new examples. This scenario will be developed in our future research, but at the current time, it can be observed as an additional advantage toward using ELMs in this application.

5. Conclusions

The aim of the presented work was the identification of diagnostic signals in the composition of exhaust gas from marine 4-stroke Diesel engines. To achieve the presented aim, an active experiment was performed that consists of measurements conducted during a laboratory engine operation with simulated malfunctions. The cause-and-effect analysis allows us to separate the following diagnostic signals: the NO_x, CO, CO_2 and O_2 fractions in the exhaust gas and the temperatures of the exhaust gas behind each of the engine cylinders. Verification of the selected signals was prepared by using classification tools based on one-vs-one ensembles of ELMs. The obtained results allow us to formulate the following conclusions:

- The composition of the exhaust gas emitted from the marine 4-stroke diesel engine is the carrier of the diagnostic signals of malfunctions located in the engine cylinders and the air-exhaust gas exchange system.
- The values of the presented signals are sufficient for clear identification of the simulated malfunctions.
- The proposed decomposition-based ELM ensemble can tackle efficiently the complex 15-class problem, offering high recognition accuracy and low classification time, which allows for the rapid detection of failures.
- Due to its low complexity, fast response and high accuracy, the proposed ensemble-based intelligent fault diagnostic system is

MOST WIEDZY Pobrano z mostwiedzy.pl

suitable for real-life implementation on board of marine vehicles.

Our future research will investigate this marine engine monitoring problem from the data stream perspective (Woźniak, 2011) and will propose online systems that can update themselves during their exploitation and react to shifts and drifts in the diagnostic signals.

Acknowledgment

Jerzy Kowalski was supported by the Polish National Science Center under the grant no. DEC-2011/01/D/ST8/07142.

Bartosz Krawczyk and Michał Woźniak were supported by the Polish National Science Center under the grant no. DEC-2013/09/B/ST6/02264 and by the statutory fund of the Department of Systems and Computer Sciences, Wrocław University of Science and Technology.

All computational experiments were carried out using computer equipment sponsored by EC under FP7, Coordination and Support Action, Grant Agreement Number 316097, ENGINE - European Research Centre of Network Intelligence for Innovation Enhancement (http://engine.pwr.wroc.pl/).

References

Aha, D.W., 1997. Lazy Learning. Kluwer Academic Publishers, Norwell, MA, USA.
Allwein, E.L., Schapire, R.E., Singer, Y., 2000. Reducing multiclass to binary: a unifying approach for margin classifiers. J. Mach. Learn. Res. 1, 113–141.
Alpaydin, E., 1999. Combined 5×2 cv f test for comparing supervised classification learning algorithms. Neural Comput. 11 (8), 1885–1892.
Ayerdi, B., Graña, M., 2014. Hybrid extreme rotation forest. Neural Netw. 52, 33–42.
Breiman, Leo, 1996. Bagging predictors. Mach. Learn. 24 (August (2)), 123–140.
Breiman, Leo, 2001. Random forests. Mach. Learn. 45 (October (1)), 5–32.
Buteneers, P., Caluwaerts, K., Dambre, J., Verstraeten, D., Schrauwen, B., 2013. Optimized parameter search for large datasets of the regularization parameter and feature selection for ridge regression. Neural Process. Lett. 38 (3), 403–416.
Cao, J., Kwong, S., Wang, R., Li, X., Li, K., Kong, X., 2015. Class-specific soft voting based multiple extreme learning machines ensemble. Neurocomputing 149, 275–284.
Cao, J., Lin, Z., Huang, G.-B., Liu, N., 2012. Voting based extreme learning machine. Inf. Sci. 185 (1), 66–77.
Carlton, John, 2012. Marine Propellers and Propulsion. Butterworth-Heinemann, Amstredam.
Dereszewski, Mirosław, 2014. Validation of failure detection by crankshaft angular speed analysis under sea waving conditions. Diagnostyka 15.
Dietterich, T.G., Bakiri, G., 1995. Solving multiclass learning problems via error-correcting output codes. J. Artif. Intell. Res. 2 (January), 263–286.
Ding, S., Zhao, H., Zhang, Y., Xu, X., Nie, R., 2015. Extreme learning machine: algorithm, theory and applications. Artif. Intell. Rev. 44 (1), 103–115.
Galar, M., Fernández, A., Barrenechea, E., Bustince, H., Herrera, F., 2011. An overview of ensemble methods for binary classifiers in multi-class problems: experimental study on one-vs-one and one-vs-all schemes. Pattern Recognit. 44 (8), 1761–1776.
Galar, M., Fernández, A., Barrenechea, E., Herrera, F., 2015. DRCW-OVO: distance-based relative competence weighting combination for one-vs-one strategy in multi-class problems. Pattern Recognit. 48 (1), 28–42.
Jackowski, K., Krawczyk, B., Woźniak, M., 2014. Improved adaptive splitting and selection: the hybrid training method of a classifier based on a feature space partitioning. Int. J. Neural Syst. 24 (3).
Jain, L.C., Seera, M., Lim, C.P., Balasubramaniam, P., 2014. A review of online learning in supervised neural networks. Neural Comput. Appl. 25 (3-4), 491–509.
Kowalski, Jerzy, 2014. An experimental study of emission and combustion characteristics of marine diesel engine with fuel pump malfunctions. Appl. Therm. Eng. 65 (1), 469–476.
Kowalski, Jerzy, 2015a. An experimental study of emission and combustion characteristics of marine diesel engine in case of cylinder valves leakage. Polish Maritime Res. 22 (3), 90–98.
Kowalski, Jerzy, 2015b. Concept of the multidimensional diagnostic tool based on exhaust gas composition for marine engines. Appl. Energy 150, 1–8.
Liu, B., Hao, Z., Tsang, E.C.C., 2008. Nesting one-against-one algorithm based on SVMs for pattern classification. IEEE Trans. Neural Netw. 19 (12), 2044–2052.
Lorena, A.C., Carvalho, A.C., Gama, J.M., 2008. A review on the combination of binary classifiers in multiclass problems. Artif. Intell. Rev. 30 (1-4), 19–37.
Pao, Y.-H., Park, G.H., Sobajic, D.J., 1994. Learning and generalization characteristics of the random vector functional-link net. Neurocomputing 6 (2), 163–180.
Pawletko, Rafał, 2015. Analysing applicability of selected methods to smooth indicator diagrams of marine medium-speed engine. Polish Marit. Res. 22 (2), 55–61.
Pontoppidan, Niels Henrik, Sigurdsson, Sigurdur, Larsen, Jan, 2005. Condition monitoring with mean field independent components analysis. Mech. Syst. Signal Process. 19 (6), 1337–1347.
Renaudin, Laurence, Bonnardot, Frédéric, Musy, Olivier, Doray, J.B., Rémond, Didier, 2010. Natural roller bearing fault detection by angular measurement of true instantaneous angular speed. Mech. Syst. Signal Process. 24 (7), 1998–2011.
Rifkin, R., Klautau, A., 2004. In defense of one-vs-all classification. J. Mach. Learn. Res. 5, 101–141.
Rodriguez, Juan J., Kuncheva, Ludmila I., Alonso, Carlos J., 2006. Rotation forest: a new classifier ensemble method. IEEE Trans. Pattern Anal. Mach. Intell. 28 (October (10)), 1619–1630.
Schapire, Robert E., Freund, Yoav, 2012. Boosting: Foundations and Algorithms. The MIT Press, Cambridge MA.
Schmidt, W.F., Kraaijveld, M.A., Duin, R.P.W., 1992. Feedforward neural networks with random weights. In: 11th IAPR International Conference on Pattern Recognition, 1992. Vol. II. Conference B: Pattern Recognition Methodology and Systems, Proceedings, pp. 1–4.
Shao, Z., Er, M.J., 2016. An online sequential learning algorithm for regularized extreme learning machine. Neurocomputing 173, 778–788.
Woodyard, D., 2009. Pounder's Marine Diesel Engines and Gas Turbines. Butterworth-Heinemann, Amsterdam.
Woźniak, M., Graña, M., Corchado, E., 2014. A survey of multiple classifier systems as hybrid systems. Inf. Fusion 16 (1), 3–17.
Woźniak, Michał, 2011. A hybrid decision tree training method using data streams. Knowl. Inf. Syst. 29 (2), 335–347.
Wu, Jian-Da, Huang, Cheng-Kai, 2011. An engine fault diagnosis system using intake manifold pressure signal and Wigner–Ville distribution technique. Expert Syst. Appl. 38 (1), 536–544.
Yang, Jianguo, Pu, Lijun, Wang, Zhihua, Zhou, Yichen, Yan, Xinping, 2001. Fault detection in a diesel engine by analysing the instantaneous angular speed. Mech. Syst. Signal Process. 15 (3), 549–564.

MOST WIEDZY Pobrano z mostwiedzy.pl