Application of Machine Learning in Diesel Engines Fault Identification

Denys Pestana-Viana[1(✉)], Ricardo H. R. Gutiérrez[3(✉)], Amaro A. de Lima[1,2(✉)], Fabrício L. e Silva[2(✉)], Luiz Vaz[3(✉)], Thiago de M. Prego[2(✉)], and Ulisses A. Monteiro[3(✉)]

[1] Post graduation Program of Instrumentation and Applied Optics (PPGIO), Federal Center of Technological Education Celso Suckow da Fonseca (CEFET-RJ), Rio de Janeiro, Brazil
denys.cefet@gmail.com

[2] Center for Research in Mechatronics (NUPEM), CEFET-RJ, Nova Iguaçu, Brazil
{amaro.lima,fabricio.silva,thiago.prego}@cefet-rj.br

[3] Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil
{rhramirez,vaz,ulisses}@oceanica.ufrj.br

Abstract. The objective of this work is the fault diagnosis in diesel engines to assist the predictive maintenance, through the analysis of the variation of the pressure curves inside the cylinders and the torsional vibration response of the crankshaft. Hence a fault simulation model based on a zero-dimensional thermodynamic model was developed. The adopted feature vectors were chosen from the thermodynamic model and obtained from processing signals as pressure and temperature inside the cylinder, as well as, torsional vibration of the engines flywheel. These vectors are used as input of the machine learning technique in order to discriminate among several machine conditions, such as normal, pressure reduction in the intake manifold, compression ratio and amount of fuel injected reduction into the cylinders. The machine learning techniques for classification adopted in this work were the multilayer perceptron (MLP) and random forest (RF).

Keywords: Machine learning · Fault identification · Vibration analysis

1 Introduction

In the offshore industry, where the daily operation cost of units rises to exorbitant amounts, unexpected production outages can mean major economic losses. In addition, the unexpected failure of the equipment on board can produce accidents causing damage to the structure, putting at risk the crew, and possibly resulting in environmental impact.

Diesel engines can be used in the offshore industry (support vessels and oil production units) in the main propulsion system, in electric power plants and in the mechanical drive of pumps and compressors. Therefore, the proper

K. L. Cavalca and H. I. Weber (Eds.): IFToMM 2018, MMS 61, pp. 74–89, 2019.
https://doi.org/10.1007/978-3-319-99268-6_6

functioning of the engine components is critical to provide the torque and power for which they were designed. This dependence on diesel engines makes them have a high economic penalty when out of operation, especially when these stops are not programmed.

The artificial neural networks as well as other supervised classifiers are appropriate for machinery applications as they can be trained offline and tested in real-time signals (online), indicating whether there is or not the failure presence in the system and with a reduced time compared to diagnosis through traditional parameter estimation in dynamic model [1], i.e., the artificial intelligence has the advantage of predicting the failure pattern faster than other numerical methods, which are not based on machine learning [2].

This paper proposes an improvement in the method described by [3], through the use of the technique of machine leaning, with the purpose of providing speed in the detection/identification of failures in diesel engines. For this, they will be used classifiers based on artificial neural networks (ANN) and random forest (RF).

Random forest is an ensemble classification method, which means that it combines the decision of a set of classifiers through a voting process, in order to classify an unknown example. An ensemble classifier is generally more effective than any of the individual classifiers that compounds it [4].

The proposed system follows a modular architecture similar to the ones described in [5,6] for a condition-based maintenance system as showed in Fig. 1, which comprises five blocks: the dynamic model block consists in solving the equations of the model based on the severities applied and the fault signals are returned; the database generation block deals with the creation of the normal signals and their association with the fault signals; the AWGN process block is responsible for adding white noise to all 701 signals of the database; the feature extraction block uses techniques to generate features from the data signals that

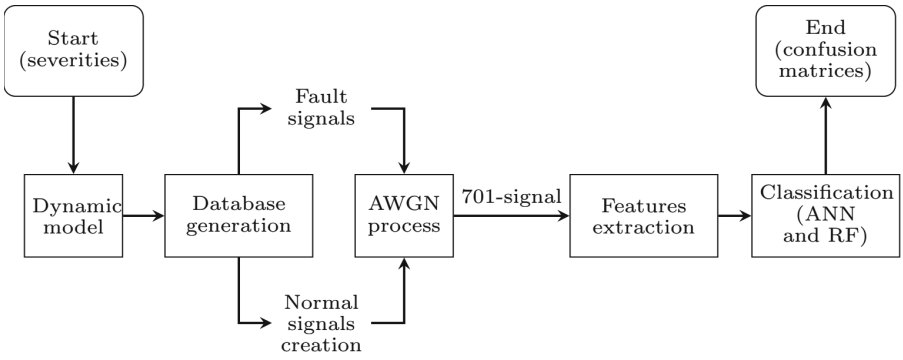

Fig. 1. Block diagram of the proposed system composed by the dynamic model block, the database generation block (split into fault and normal signals creation), the AWGN process block, the feature extraction block, and the classification block (composed by a ANN and RF classifiers).

will be used in the classification block; and the last block, the classification block, is where the machine learning tests occur and the results are displayed through confusion matrices.

2 System Description

The system consists in a diesel engine failures simulation based on the work of [3], which implemented and validated a set of simulation algorithms using the manufacturer's data. Thus creating a program for the simulation of fault routines covering the most common types of failures in diesel engines according to the following cases: normal operation, compression fault, injected fuel mass fault and pressure in the intake manifold fault. The motor chosen as the case study is the series Acteon 6.12TCE with four stroke cycle engine manufactured by MWM Diesel Motors [7].

By modifying the operational parameters (failure parameters) of the engine, the fault signals are created, thus obtaining a model that allows simulating the various conditions of failure in the diesel engine [3,8]. Changing the variables of the functions described in Eqs. (1)–(3) it is possible to emulate the operational condition or a fault situation.

$$P_i(\theta) = f(P_a, P_r, T_p, r_i, m_{a_i}, m_{c_i}, \theta_{inj_i}), \tag{1}$$

$$T_i(\theta) = f(P_a, P_r, T_p, r_i, m_{a_i}, m_{c_i}, \theta_{inj_i}), \text{ and} \tag{2}$$

$$V(t) = f(P_a, P_r, T_p, \{r\}, \{m_a\}, \{m_c\}, \{\theta_{inj}\}, \{K\}), \tag{3}$$

where $P_i(\theta)$ is the instantaneous pressure curve in the interior of each cylinder i as a function of the angle of the crankshaft θ; $T_i(\theta)$ is the instantaneous temperature curve inside each cylinder; $V(t)$ is the instantaneous torsional vibration response measured at the flywheel as a function of the time t; P_a is the pressure in the intake manifold; P_r is the common rail pressure; T_p is the temperature on the cylinder walls; r_i is the compression ratio; m_{a_i} is the mass of air admitted to the cylinders; m_{c_i} is the mass of fuel injected into the cylinders; θ_{inj_i} is the angle of injection start in each cylinder and K is the stiffness of the cranks.

All the simulated signals are generated by solving Eqs. (4)–(6) using as input parameters the severity conditions that will emulate the engine operating condition.

$$P_i(\theta) = f(P_a, \cdots, \theta_{inj_i}, \Delta P_a, \Delta P_r, \Delta T_p, \Delta r_i, \Delta m_{a_i}, \Delta m_{c_i}, \Delta\theta_{inj_i}), \tag{4}$$

$$T_i(\theta) = f(P_a, \cdots, \theta_{inj_i}, \Delta P_a, \Delta P_r, \Delta T_p, \Delta r_i, \Delta m_{a_i}, \Delta m_{c_i}, \Delta\theta_{inj_i}), \text{ and} \tag{5}$$

$$V(t) = f(P_a, \cdots, \{K\}, \Delta P_a, \Delta P_r, \Delta T_p, \{\Delta r\}, \cdots, \{\Delta K\}). \tag{6}$$

The input variables for the severities allocation $\Delta[\cdot]$ is represented in Eq. (7).

$$\Delta p(\%) = \frac{p_n - p_f}{p_n} \cdot 100, \tag{7}$$

where $\Delta p(\%)$ is the percentage variation of the considered failure parameter p (severity inserted for the failures emulation); p_n is the parameter in normal operating condition and p_f is the parameter in fault condition.

The previously presented equations were based on the thermodynamic and dynamic models respectively. They were validated with experimental data provided by the motor performance information (pressures, torque, vibration and torsion). From the calibrated model, it is possible to make the appropriate changes that allow to simulate diesel engine operating faults. Four operating situations were emulated using the system described:

1. **Normal (without faults)**: it is a no fault situation (without the presence of operating situations 2, 3, 4, described below), i.e., representing as accurately as possible the operational condition presented by the performance curves of the manufacturer, or for simulation purposes, with no severity ($\Delta[\cdot]$) related to the input variables of Eqs. (4)–(6);
2. **Pressure reduction in the intake manifold**: this situation occurs when there is a change in the ambient conditions (for naturally aspirated engines), improper operation of the turbocharger, deposits in the intake valves, variation of the opening and closing angles of the intake valves, or for simulation purposes, with severity inserted into the input variable ΔP_a of Eqs. (4)–(6);
3. **Compression ratio reduction in the cylinders**: this situation occurs when there is variation of the cylinder head gasket, change in the dead volume of the combustion chamber due to deposits or corrosion and kinematic variation of the components, or for simulation purposes, with severity inserted into the input variable Δr of Eqs. (4)–(6);
4. **Reduction of amount of fuel injected into the cylinders**: this situation occurs when there is corrosion of the segment rings or cylinder walls, improper sealing of the intake valves due to corrosion, spring failure and deposits, or problems with the injection pump, or for simulation purposes, with severity inserted into the input variable Δm_c of Eqs. (4)–(6).

The outputs from the simulation fault algorithm are the variables $P_i(\theta)$, $T_i(\theta)$ and $V(t)$. $P_i(\theta)$ and $V(t)$ are used in the feature extraction stage, thereafter to be used for training the machine learning classifiers, and $T_i(\theta)$ is used to the diagnosis phase using the thermodynamic model, which will be employed as a comparative basis to evaluate the classifier performance.

The 6-channel pressure simulated signals $P_i(\theta)$ were obtained from each cylinder wall and converted to discrete-time domain generating the signals labelled as $s_{p_1}(n)$, $s_{p_2}(n)$, $s_{p_3}(n)$, $s_{p_4}(n)$, $s_{p_5}(n)$ and $s_{p_6}(n)$, respectively to each cylinder wall. A torsional simulated signal $V(t)$ was obtained from engine flywheel and also converted to discrete-time domain generating the signal labelled as $s_v(n)$. The discrete signals $s_{p_i}(n)$ and $s_v(n)$ compound a 7-channel data signal. All the signals were simulated using sampling frequency of 15 kHz for 1.008 s, making a total of 15120 samples for each channel signal.

3 Dynamic Model

In order to detect, identify and quantify failures in a diesel engine, a coupled model was proposed by [3]. The proposed model consists on a zero-dimensional

thermodynamic model, from which the pressure profiles inside cylinders are generated, and a torsional dynamic model of the crankshaft is used to obtain the torsional vibration for different operational condition.

For the developed model a 4-stroke diesel engine with common rail injection system was considered, as presented in [3].

The gas mixture inside cylinder pass through several processes: intake, compression, combustion, expansion and exhaust, but as the focus of interest is only the performance, the pressure in the intake and the exhaust processes were considered constant according to the air standard diesel cycle. Meanwhile, the compression, combustion and expansion processes were represented by the universal gas equation [9,10].

Regarding the performance, even that a real diesel engine takes into account the mass flow through the valves during the intake and exhaust processes, it can be neglected once the mass of the control volume is considered constant [9,10].

Although the obtained model refers to a single cylinder, it is possible to obtain the pressure profile for the 6 cylinders if the interval and ignition order were respected [11].

The torsional model for the crankshaft was developed considering the isolated crankshaft chain a system with 11 degrees of freedom, to which an equivalent lumped parameter model was developed [11–13]. The equation of motion for the refereed system is presented in Eq. (8),

$$[\mathbf{J}]\left\{\ddot{\theta}(t)\right\} + [\mathbf{C}]\left\{\dot{\theta}(t)\right\} + [\mathbf{K}]\left\{\theta(t)\right\} = \left\{M(t)\right\}, \tag{8}$$

where [**J**], [**C**] and [**K**], are the inertial matrix, the damping matrix and the torsional stiffness matrix, respectively. The vectors $\{\theta(t)\}$, $\left\{\dot{\theta}(t)\right\}$ and $\left\{\ddot{\theta}(t)\right\}$ are the angular position and its derivatives, and $\{M(t)\}$ is the vector of external torques related to dynamic loads originated by the inertial and the combustion loads.

In the torsional dynamics only the tangential loads are effective for the external torques. The Eq. (9) gives the external torque for a single slide-crank system.

$$M_i(\theta) = F_{ti}(\theta).r, \tag{9}$$

where $F_{ti}(\theta)$ is the effective force for the calculation of the torque related to a single slide-crank system, and r the distance from the crankshaft's center to the point of application of the effective force.

Although the obtained model refers to a single cylinder, it is possible to obtain the torque for the 6 cylinders if the interval and ignition order were respected [11].

4 Database

The database is expected to emulate all operating scenarios under study. In our case, all possible diesel machine faults and system conditions variations, which

correspond to severities levels containing enough information to characterize and discriminate the faults. In this work the developed database covered the following operating conditions:

- **Normal (without faults)**: in this class, no fault is implemented and 51 different instances (realizations) are created from the insertion of 0.1% of maximum severity with normal Gaussian probability distribution covering a range between 0 and 0.1% in the 27 input variables of severity adopted from the dynamic model ΔP_a, ΔP_r, ΔT_p, Δr_i, Δm_{a_i}, Δm_{c_i}, $\Delta \theta_{inj_i}$. The objective of this step is to emulate the real motor in normal operation, where the machine variables drifts with a small range around the optimal functioning.
- **Pressure reduction in the intake manifold**: Several scenarios with severities of $[1, 2, 3, \cdots, 50]\%$ for the variable ΔP_r are considered, making a total of 50 "pressure reduction in the intake manifold" scenarios.
- **Compression ratio reduction in the cylinders**: This condition involves all cylinders to create the scenarios. Several scenarios with severities of $[1, 2, 3, \cdots, 50]\%$ related to variables Δr_i and the cylinders $i = [1, 2, 3, 4, 5, 6]$ are considered, making a total of 50 "compression ratio reduction" different scenarios for each cylinder, respectively, generating a total of 300 "compression ratio reduction" scenarios.
- **Reduction of amount of fuel injected into the cylinders**: This condition involves the all cylinders to create the scenarios. Several scenarios with severities of $[1, 2, 3, \cdots, 50]\%$ related to variables Δm_{c_i} and the cylinders $i = [1, 2, 3, 4, 5, 6]$ are considered, making a total of 50 "reduction of amount of fuel injected" different scenarios addressed for each cylinder, making a total of 300 "reduction of amount of fuel injected into the cylinders" addressed scenarios.

In all scenarios, the motor rotation frequency was set at 2500 RPM. According to [3] the rotation of 2500 RPM was used, since it presented the lowest joint error rate in the estimation of the mean and maximum pressures of the burning cycle, between the experimental data (according to data supplied by the manufacturer) and the simulated data, during the validation stage of the thermodynamic and dynamic models.

The entire database comprises a total of 701 different fault scenarios for 4 distinct operational conditions. 51 of which from the normal class, 50 from "pressure reduction in the intake manifold" class, 300 from "compression ratio reduction in the cylinders" class and 300 from the "reduction of amount of fuel injected into the cylinders" class. This database is named 701-signal database.

5 Feature Extraction

The technique for feature extraction consists in estimating the mean and maximum pressure values from the 6 pressure cylinder signals $s_{p_1}(n)$, $s_{p_2}(n)$, $s_{p_3}(n)$, $s_{p_4}(n)$, $s_{p_5}(n)$ and $s_{p_6}(n)$, and obtaining spectral information from the torsional vibration signal $s_v(n)$. The adopted measures associated to the faults in order to discriminate them are:

– **Maximum pressure inside the cylinders estimation**: it is the maximum value of each discrete pressure curve associated to each cylinder $s_{p_1}(n)$, $s_{p_2}(n)$, $s_{p_3}(n)$, $s_{p_4}(n)$, $s_{p_5}(n)$ and $s_{p_6}(n)$, generating M_{p_1}, M_{p_2}, M_{p_3}, M_{p_4}, M_{p_5} and M_{p_6}, respectively. Equation (10) summarizes this subset of features.

$$M_{p_i} = max[s_{p_i}(n)], \tag{10}$$

where M_{p_i} is the maximum value from pressure curves $s_{p_i}(n)$ for each cylinder i.

– **Mean pressure inside the cylinders estimation**: it is the first-order expected value (mean) of each discrete pressure curve associated to each cylinder $s_{p_1}(n)$, $s_{p_2}(n)$, $s_{p_3}(n)$, $s_{p_4}(n)$, $s_{p_5}(n)$ and $s_{p_6}(n)$, generating μ_{p_1}, μ_{p_2}, μ_{p_3}, μ_{p_4}, μ_{p_5} and μ_{p_6}, respectively. Equation (11) summarizes this subset of features.

$$E[s_{p_i}(n)] = \mu_{p_i} = \frac{1}{N}\sum_{n=1}^{N} s_{p_i}(n), \tag{11}$$

where, $E[\cdot]$ is the expected value, μ, associated to each cylinder-pressure curve; N is the number of samples of $s_{p_i}(n)$.

– **Spectral analysis**: It is the technique to estimate the torsional frequency spectrum similarly to the one described in [5]. It consists in calculating a N_{DFT}-point DFT of $s_v(n)$ according to Eq. (12), generating $S_v(k)$, which will be used to calculate $F(k)$, $A(k)$, and $P(k)$ representing frequency (Hz), amplitude (N.m) and phase (degrees) amounts, respectively. The Eqs. (13)–(15) summarize this subset of features.

$$S_v(k) = \frac{1}{N_{\text{DFT}}}\sum_{k=0}^{N_{\text{DFT}}-1} s_v(n) W_N^{kn}, \tag{12}$$

where $S_v(k)$ is the N_{DFT}-point DFT of $s_v(n)$ with $W_N^{kn} = e^{-\frac{j2\pi}{N_{\text{DFT}}}}$ and j representing the complex number.

$$F(k) = \frac{kF_s}{N_{\text{DFT}}}, \tag{13}$$

where $F(k)$ is the harmonic frequency of torsional spectrum $S_v(k)$; k is the frequency bin associated to a frequency (Hz) and F_s is the data acquisition sampling frequency.

$$A(k) = |S_v(k)|, \tag{14}$$

where $A(k)$ is the amplitude (N.m) of torsional spectrum $S_v(k)$.

$$P(k) = \frac{360}{2\pi} arg[S_v(k)], \tag{15}$$

where $P(k)$ is the phase (degree) of torsional spectrum $S_v(k)$ and $arg[\cdot]$ represents the complex argument of the spectrum, i.e., that is the phase of displacement between the real and imaginary part of the complex variable.

5.1 Feature Vector

Compared to the technique presented in [5,6], the proposed one differs by applying the feature extraction in torsional vibration and addressing mean and maximum values in features vector.

In order to generate the feature vector, the first step is to estimate the maximum values of pressure curves for each cylinder according to Eq. (10) providing the features subset:

$$M_{p_i} = [M_{p_1}, M_{p_2}, M_{p_3}, \cdots, M_{p_6}]\,. \tag{16}$$

The second step is to obtain the mean values from the cylinders pressures curves, according to Eq. (11), where the features subset is:

$$\mu_{p_i} = [\mu_{p_1}, \mu_{p_2}, \mu_{p_3}, \cdots, \mu_{p_6}]\,. \tag{17}$$

The third step is to calculate the spectral values from the torsional vibration curves $S_v(k)$ according to Eqs. (12)–(15), and its first 24 harmonics, i.e., the first 24 half orders of the engine. Considering the rotation fixed at 2500 RPM the first half order is given by $\frac{1}{2}\frac{2500}{60}$ Hz. Consequently, 24-half order frequency vector is $f = [21.083, 41.667, \cdots, 500]$, where its elements are the frequencies of the first 24 half orders frequencies. The relation between the spectral bin and frequency is directly obtained from $k = [f]\frac{N_{\text{DFT}}}{F_s}$. The subset related to the spectral values is then given by:

$$F(k) = [F(k_1), F(k_2), F(k_3), \cdots, F(k_{24})]\,, \tag{18}$$

where p_j is the pressure curve associated to each cylinder, with $j = 1, 2, \cdots, 6$.

$$A(k) = [A(k_1), A(k_2), A(k_3), \cdots, A(k_{24})]\,, \text{ and} \tag{19}$$

$$P(k) = [P(k_1), P(k_2), P(k_3), \cdots, P(k_{24})]\,, \tag{20}$$

where k_j is the frequency bin associated to each element of 24-half order vector, with $j = 1, 2, \cdots, 24$.

The final step is to combine all measures in a feature vector, V_f, using steps described from Eqs. (16)–(20) concatenating the 3 spectral variables, maximum and mean values, which can be expressed by:

$$\begin{aligned} V_f = \{&M_{p_1},\, M_{p_2},\, M_{p_3},\, \cdots,\, M_{p_6}, \mu_{p_1},\, \mu_{p_2},\, \mu_{p_3},\, \cdots,\, \mu_{p_6},\\ &F(k_1),\, F(k_2),\, F(k_3),\, \cdots,\, F(k_{24}), A(k_1),\, A(k_2),\, A(k_3),\, \cdots,\, A(k_{24}),\\ &P(k_1),\, P(k_2),\, P(k_3),\, \cdots,\, P(k_{24})\}. \end{aligned} \tag{21}$$

The feature vector V_f achieving a 84 dimensionality vector, which will be used for the training and the test steps of the machine learning algorithms discussed in the next section. The respective values of the distribution of 701-signal database of the several subsets of the vector V_f, which is composed of M_{p_i}, μ_{p_i}, $F(k)$, $A(k)$ and $P(k)$, according to Eqs. (16)–(20), where are organized in box plot and shown in Fig. 2.

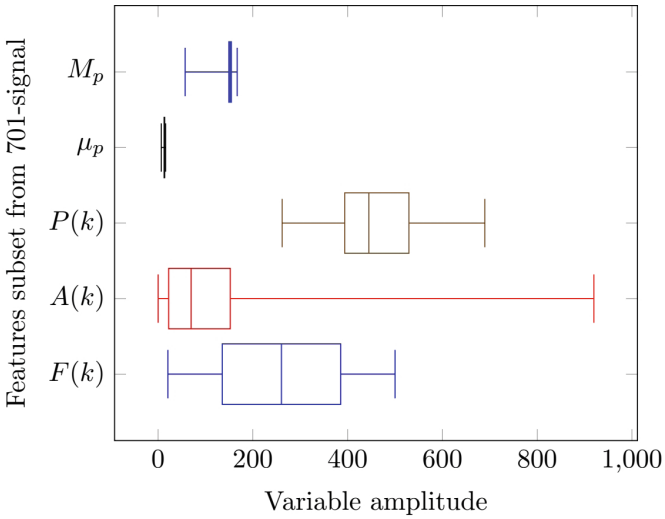

Fig. 2. Box plot of 701-signal database for $F(k)$, $A(k)$, $P(k)$, μ_{p_i} and M_{p_i}, respectively. The plot shows the median, 25% quartile, 75% quartile and the lower and upper range (whiskers), which are the max and min for each distribution, respectively.

6 Experimental Results

The fault classification experiment consists in adopting similar procedure presented in [5,6] to evaluate the system ability of fault discrimination by adding the maximum and average pressure and spectral measures, not only for normal, but also to other faults. These tasks were performed with the classifiers: random forest (RF) with the feature vector as input, the number of trees of 84 obtained empirically as made in [5] and the output with the size of the number of fault classes; and multilayer perceptron (MLP) with input layer with the same size of input feature vector, one hidden layer with the amount of neurons approximately equals to the input layer size also obtained empirically, and the output layer with the number of neurons equals to the number of classes to be discriminated.

The original 701-signal database was divided into 2 disjoint sets with approximately 80% and 20% of the signals for training and test, respectively and into 3 disjoint sets with approximately 70%, 10%, and 20% of the signals for training, validation and test, respectively to ANN classifier. Each of the sets is chosen to represent the data with maximum variability from fault severity intensity aspect. The validation set is employed to avoid the ANN to become excessively specialized on the training signals thus losing its generalization capability [5].

In order to avoid data biased performance in classification, the k-fold cross-validation technique was applied in all classification tests, dividing the base 701-signal in k = 5 folds with the purpose of circularly changing the test subset maintaining the proportion of 80 and 20% for the training and test sets, respectively.

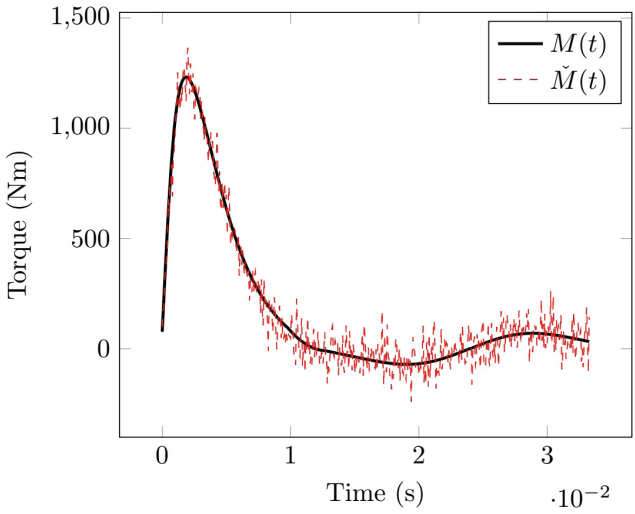

Fig. 3. Process of applying 15 dB SNR of additive noise in torque variable. In graph above, the dashed line is the signal with noise addition and the straight line is the original signal without noise. The new variable torque with AWGN ($\{\check{M}(t)\}$) is $\{\check{M}(t)\} = \{M(t)\} + \nu(t)$, where $\nu(t)$ is white Gaussian noise with 15 dB SNR. The process is the same for the remaining variables (pressure and torsional vibration).

All tables related to the fault classification experiment present the classification performance X/Y, also called confusion matrix, for the test data from the 701-signal database, where X represents the recognized and Y is the total number of signals for the target class. The total accuracy classification performance of confusion matrix is represented by $W \pm \sigma$, where W is the total accuracy and σ is its standard deviation during k-fold validation.

In the 701-signal base different noise levels were applied with 60, 30, 15 and 0 dB of SNR using additive Gaussian white noise (AWGN) in the original signals, in order to evaluate their influence on classification performance. The Fig. 3 shows the variables torque $\{M(t)\}$ without noise addition as in Eq. (8), and torque with 15 dB SNR AWGN ($\{\check{M}(t)\}$).

The performance of the machine learning was compared with the approach using the Levenberg-Marquardt least squares (LMLS) technique [3] in order to assess the efficiency of the proposed system.

6.1 Classification Results Using 60 dB SNR AWGN

For the tests below, all base faults were included by testing the performance for all fault scenarios of the 701-signal database. Different classes divisions were used to investigate the performance of the classifier for different configurations in the output layer, in addition to obtaining the results by expanding the classes

Table 1. Confusion matrix with 4-class pattern recognition using feature vector V_f with 60 dB noise and random forest.

Class	Target			
	c1	c2	c3	c4
c1	**10/10**	0/9	0/61	0/60
c2	0/10	**9/9**	1/61	0/60
c3	0/10	0/9	**60/61**	0/60
c4	0/10	0/9	0/61	**60/60**
Total	**99.3**	± **0.23**	(%)	

Table 2. Confusion matrix with 4-class pattern recognition using feature vector V_f with 60 dB noise and ANN.

Class	Target			
	c1	c2	c3	c4
c1	**10/10**	0/9	4/61	2/60
c2	0/10	**9/9**	0/61	0/60
c3	0/10	0/9	**57/61**	0/60
c4	0/10	0/9	0/61	**58/60**
Total	**95.7**	± **0.97**	(%)	

Table 3. Confusion matrix with 8-class pattern recognition using feature vector V_f with 60 dB noise and random forest.

Class	Target							
	c1	c2	c3	c4	c5	c6	c7	c8
c1	**5/5**	0/18	0/27	0/14	1/20	0/16	0/28	0/12
c2	0/5	**18/18**	0/27	0/14	0/20	0/16	0/28	0/12
c3	0/5	0/18	**27/27**	0/14	0/20	0/16	0/28	0/12
c4	0/5	0/18	0/27	**14/14**	0/20	0/16	0/28	0/12
c5	0/5	0/18	0/27	0/14	**19/20**	0/16	0/28	0/12
c6	0/5	0/18	0/27	0/14	0/20	**16/16**	0/28	0/12
c7	0/5	0/18	0/27	0/14	0/20	0/16	**28/28**	0/12
c8	0/5	0/18	0/27	0/14	0/20	0/16	0/28	**12/12**
Total (%)	**99.3 ± 0.27**							

Table 4. Confusion matrix with 8-class pattern recognition using feature vector V_f with 60 dB noise and ANN.

Class	Target							
	c1	c2	c3	c4	c5	c6	c7	c8
c1	**5/5**	0/18	0/27	0/14	0/20	1/16	0/28	0/12
c2	0/5	**18/18**	0/27	0/14	0/20	0/16	0/28	0/12
c3	0/5	0/18	**27/27**	0/14	0/20	0/16	0/28	0/12
c4	0/5	0/18	0/27	**14/14**	0/20	0/16	0/28	0/12
c5	0/5	0/18	0/27	0/14	**20/20**	0/16	0/28	0/12
c6	0/5	0/18	0/27	0/14	0/20	**15/16**	0/28	0/12
c7	0/5	0/18	0/27	0/14	0/20	0/16	**28/28**	0/12
c8	0/5	0/18	0/27	0/14	0/20	0/16	0/28	**12/12**
Total (%)	**99.3 ± 0.85**							

Table 5. Confusion matrix with 4-class severity pattern recognition using feature vector V_f with 60 dB noise and RF.

Class	Target			
	S1	S2	S3	S4
S1	**13/13**	0/20	0/49	0/58
S2	0/13	**19/20**	1/49	0/58
S3	0/13	1/20	**48/49**	1/58
S4	0/13	0/20	0/49	**57/58**
Total	**97.9**	**± 0.85**	**(%)**	

Table 6. Confusion matrix with 4-class severity pattern recognition using feature vector V_f with 60 dB noise and ANN.

Class	Target			
	S1	S2	S3	S4
S1	**0/13**	0/20	0/49	0/58
S2	13/13	**19/20**	2/49	0/58
S3	0/13	1/20	**46/49**	2/58
S4	0/13	0/20	1/49	**56/58**
Total	**86.4**	**± 2.7**	**(%)**	

of failures. The tests were defined by the number of classes to be predicted by the classifier and in each test it was adopted the following class nomenclature:

- 4-class test: normal (c1), pressure reduction in the intake manifold (c2), compression ratio reduction in the cylinders (c3) and reduction of amount of fuel injected into the cylinders (c4);
- 8-class test: normal (c1), cylinder failure: cylinder 1 (c2), cylinder 2 (c3), cylinder 3 (c4), cylinder 4 (c5), cylinder 5 (c6) and cylinder 6 (c7) and failure on all cylinders (c8);
- 4-class severity test (S4): percentage severity level (S): $1 \leq S \leq 10$ (S1), $10 < S \leq 20$ (S2), $20 < S \leq 30$ (S3) and $30 < S \leq 50$ (S4).

Tables 1, 2, 3, 4, 5 and 6 present similar framework to the ones described in [5,6], i.e., it performs a classification task using V_f with an ANN of 84-84-4 and 84-84-8 neurons in the input, hidden and output layers for tests with 4 and 8 classes, respectively and 84 trees with 500 splits to the RF classifier.

Comparing Tables 1, 2, 3, 4, 5 and 6, it is clearly noticed that the use of RF improves each individual class performance and consequently the overall system accuracy from up to 99% in the 4 and 8-class tests. Comparing the performance with the ANN classifiers for the same task, there was a decrease in accuracy of 3.6% for both, the 4 and 8-class tests. However, for the severity classification test, there was a difference of 11.5% in performance between RF and ANN classifiers.

6.2 Classification Results Combining All Scenarios in a Summary Table

For the tests below, all database faults were included to evaluate the performance for all fault scenarios of the 701-signal database. In this section a similar configuration as the one adopted in Sect. 6.1 was applied, except for an extra test class. The test were established with the same nomenclature for 4-class, 8-class and 4-class severity as presented in Sect. 6.1 and an additional 14-class nomenclature defined as follows:

Table 7. Classification accuracy results for RF and ANN classifiers with 60, 30, 15 and 0 dB SNR using feature vector V_f.

Noise (dB)	N. class	Accuracy (%)	
		RF	ANN
60	4	99.3 ± 0.2	95.7 ± 1.0
	8	99.3 ± 0.3	99.3 ± 0.9
	14	99.3 ± 0.3	96.4 ± 0.9
	S4	97.9 ± 0.9	86.4 ± 2.7
30	4	98.6 ± 0.3	92.9 ± 1.0
	8	98.6 ± 0.3	97.1 ± 0.8
	14	97.9 ± 0.4	95.8 ± 1.0
	S4	97.9 ± 0.9	86.1 ± 2.3
15	4	95.1 ± 0.3	88.6 ± 1.0
	8	95.8 ± 0.3	90.7 ± 0.8
	14	94.4 ± 0.4	92.1 ± 0.8
	S4	92.3 ± 0.9	76.4 ± 3.2
0	4	83.5 ± 0.9	75.7 ± 1.0
	8	81.4 ± 0.8	79.3 ± 1.3
	14	80.7 ± 0.9	72.9 ± 1.2
	S4	76.4 ± 1.3	60.7 ± 3.7

– 14-class test: normal (c1), pressure reduction in the intake manifold (c2), compression ratio reduction in: cylinder 1 (c3), cylinder 2 (c4), cylinder 3 (c5), cylinder 4 (c6), cylinder 5 (c7) and cylinder 6 (c8) and reduction of amount of fuel injected in: cylinder 1 (c9), cylinder 2 (c10), cylinder 3 (c11), cylinder 4 (c12), cylinder 5 (c13) and cylinder 6 (c14).

The results shown in Table 7 indicate an overall classification performance for RF classifier of 99.3%, which is suitable for the problem of diesel faults classification, making the method proposed effective for the recognition of failure patterns under analysis in this work. The ANN classifier, in general, obtained lower performance when compared to the RF, making the RF more appropriate for the classification tasks addressed in this work.

6.3 Comparative Analysis of LMLS and the Proposed Approaches

The comparative analysis between LMLS technique [3] and the proposed approaches using 4 different classifiers aiming to detect the type of fault (normal, ΔP_a, Δr and Δm_c), the location of fault (normal, cylinder 1 to 6, and all cylinders), the type and location of fault combined, and the fault severity range, is addressed in order to observe the qualitative aspects of both techniques.

It has to be noticed that LMLS technique and the proposed approaches present significant structural differences. The LMLS algorithm requires that the type of fault is known previously to the system application and returns the severity level in numerical terms, while the proposed approaches uses classification techniques to detect the fault type and location, and its severity range.

Table 8 presents the scenarios, where the techniques were analyzed. Cases 1 to 6 were contaminated with only one type of fault each with a certain severity level and measurement noise. Only fault ΔP_a with 25% of severity in cylinder 1 was addressed in case 1. In case 2 only fault Δr in cylinder 1 with 25% of severity was simulated and case 3 only fault Δm_c with 25% of severity and cylinder 1 was considered. The cases 7, 8 and 9 are equivalent to 1, 2 and 3 aggregating noise comprising 15 dB of SNR in fault signals. In the 6 cases the dynamic model coupled with the thermodynamic was used.

Table 8. Summary table with faults and severity levels addressed for each input scenario/case.

Case	Cyl.	SNR(dB)	Simulated fault & severity (%)		
			ΔP_a	Δr	Δm_c
1	1	-	25	-	-
2	1	-	-	25	-
3	1	-	-	-	25
4	1	15	25	-	-
5	1	15	-	25	-
6	1	15	-	-	25

In Table 9 the absolute severity estimation error for cases 1 to 6 were less than 10^{-5}% indicating a high efficiency if estimating severity levels for LMLS technique. However it should be mentioned that algorithm requires previous knowledge of the fault type to be analyzed and takes about 38 h to evaluate a single scenario. The proposed approaches evaluates 4 different classifiers. The first, a 4-class (4-c) experiment recognizes the 3 types of faults under analysis as being of the classes c2, c3 and c4, associated to ΔP_a, Δr and Δm_c, respectively. The 8-class (8-c) classifiers detect appropriately in which cylinder the fault occurred as being class c2, i.e., cylinder 1. The combined fault type and location fault classifier, which is a 14-class (14-c) classifier, was able to detect accurately the classes c2, c3 and c9 related to ΔP_a and cylinder 1, Δr and cylinder 1, and Δm_c and cylinder 1, respectively. And finally the severity range evaluation used a 4-class severity classifier (S4) and could assertively point out the class s3 corresponding to the severity range from 20% to 30%. The computational cost of one assessment of the proposed approach took about 21 ms.

Table 9. Table of maximum error in identifying the severity for each variable using LMLS technique and Machine Learning Random Forest (MLRF) classification results for 4-c, 8-c, 14-c and S4 classifiers, for different noise levels and 2500 RPM.

Case	Methodology							Time	
	LMLS			Proposed				LMLS (%)	Proposed
	ΔP_a	Δr	Δm_c	4-c	8-c	14-c	S4	(hours)	(ms)
1	$< 10^{-5}$	0	0	c2	c2	c2	s3	38	21
2	0	$< 10^{-5}$	0	c3	c2	c3	s3	38	21
3	0	0	$< 10^{-5}$	c4	c2	c9	s3	38	21
4	$< 10^{-5}$	0	0	c2	c2	c2	s3	38	21
5	0	$< 10^{-5}$	0	c3	c2	c3	s3	38	21
6	0	0	$< 10^{-5}$	c4	c2	c9	s3	38	21

Both techniques performed quite well in the presence of noise for all testes scenarios. The LMLS technique is very accurate in estimating the numerical level of fault severity. However it requires previous information about the severity type to be analyzed and has an excessively computational cost compared to proposed technique. The proposed approaches are based on classification and could accurately recognize the fault classes and severity range analyzed in the test scenarios. Different from LMLS, the proposed approaches do not require previous information about the fault. However they do not provide a numerical severity level and take a very reduced computational time.

The computational time evaluation was performed in a common personal computer, with the following characteristics: CPU core i5, 8 GB RAM, without GPU and using the 2 physical cores for parallel processing.

7 Conclusions

This paper proposed a modification in fault identification technique compared to the one adopted in [3], differing by the utilization of machine learning technique. To evaluate this approach, a new database with 701 fault scenarios was developed. The proposed feature vector applied to the 4-class recognition problem (normal, pressure reduction in the intake manifold, compression ratio reduction and reduction of amount of fuel injected) described in [3,8] have reached the maximum system performance of 99.3%.

Even in a more complex discrimination task, a 14-class problem, where the compression ratio and amount of fuel injected reduction are divided into individual cylinder fault, and new fault categories are added, the proposed feature vector applied to the classification system achieved 99.3% of overall accuracy in low signal-to-noise ratio (60 dB SNR). The addition of higher signal-to-noise ratios, 30, 15 and 0 dB SNR, decreases the overall classification performance. In real systems, the SNR should be taken into account and techniques to keep it as

small as possible should be employed. Otherwise, it could severely impairs the classification performance.

In future work, it is intended to combine the advantages of the proposed approaches and LMLS possibly including a machine learning regression step to estimate the severity level numerically and assess signals corrupted by multiple faults.

References

1. Hoffman, A.J., van der Merwe, N.T.: The application of neural networks to vibrational diagnostics for multiple fault conditions. J. Comput. Stand. Interfaces **24**, 139–149 (2002)
2. Haykin, S.: Neural Networks: A Comprehensive Foundation, Prenctice-Hall, Upper Saddle River (1999)
3. Gutirrez, R.H.R.: Simulação e identificação de falhas de motores diesel, 128p. D.Sc. Thesis-Universidade Federal do Rio de Janeiro (UFRJ) (2016)
4. Yang, B.S., Di, X., Han, T.: Random forests classifier for machine fault diagnosis. J. Mech. Sci. Technol. **22**, 1716–1725 (2008)
5. de Lima, A.A., Prego, T.M., Netto, S.L., da Silva, E.A.B., Gutierrez, R.H.R., Monteiro, U.A., Troyman, A.C.R., Silveira, F.J.C., Vaz, L.: On fault classification in rotating machines using fourier domain features and neural networks. In: Proceedings of the IEEE Latin American Symposium on Circuits and Systems (LASCAS), Cusco, Peru, p. 1–4, March 2013
6. Pestana-Viana, D., Zambrano-Lopez, R., de Lima, A.A., Prego, T.D.M., Netto, S.L., da Silva, E.A.B.: The influence of feature vector on the classification of mechanical faults using neural networks. In: Proceedings of the IEEE Seven Latin American Symposium on Circuits and Systems (LASCAS), Pará, Brazil, p. 115–118, March 2016
7. Mendes, A.S.: Development and validation of a methodology for torsional vibrations analysis in internal combustion engines, 135p. M.Sc. Dissertation–Universidade Estadual de Campinas (UEC) (2005)
8. Monteiro, U.: Thermodynamic simulation of gas turbines for fault diagnosis. D.Sc. Thesis–Universidade Federal do Rio de Janeiro (UFRJ) (2010)
9. Heywood, J.: Internal Combustion Engine Fundamentals. Mechanical Engineering. McGraw-Hill, New York (1988)
10. Stone, R.: Introduction to Internal Combustion Engines, 2nd edn. Macmillan, Basingstoke (1992)
11. Mendes, A.S.: Development and validation of a methodology for torsional vibrations analysis in internal combustion engines (2005)
12. Inman, D.J.: Engineering Vibrations, 4th edn. Pearson, London (2014)
13. Rao, S.S.: Mechanical Vibrations, 5th edn. Pearson, London (2010). Recherche