Proceedings of the 7th European Conference of the Prognostics and Health Management Society 2022 - ISBN – 978-1-936263-36-3

Wrong Injection Detection in a Small Diesel Engine, a Machine Learning Approach

Piero Danti[1], Giovanni Vichi[2], and Ryota Minamino[3]

[1,2,3] *Yanmar R&D Europe, Florence, 50125, Italy*
piero_danti@yanmar.com
giovanni_vichi@yanmar.com
ryota_minamino@yanmar.com

[1] *Università degli Studi di Firenze, Florence, 50121, Italy*
piero_danti@yanmar.com

Abstract

In the last ten years, Machine Learning (ML) and Artificial Intelligence (AI) have overwhelmed every engineering research branch finding a broad variety of applications; anomaly detection and anomaly classification are two of the topics that have benefited mostly by data-driven methods' insights. On the other side, in the small diesel engine domain, the current trend is to lean on traditional anomaly detection/classification procedures and do not foster the use of AI. The goal of this work is to detect anomalies in the in-cylinders injectors of a small diesel engine as soon as a wrong quantity of fuel is inputted into one or more cylinders by means of ML approaches. Part of the analysis aim to understand which measurements are the most relevant for the detection and to compare different techniques to select the most suitable one. Furthermore, a condition-based methodology for maintenance is proposed. After a brief review of the state-of-the-art, the case-study scenario is presented grouping sensors accordingly to their degree of accessibility; then, the implemented techniques are explained and results are discussed.

1. Introduction

Although McCulloch and Pitts proposed the first computational model of a neuron in the 1943 and Arthur Samuel wrote the first computer learning program in 1952, only in the latest years AI has encountered an impressive growth due to the availability of hardware (HW) with an increased computational power and the capability of storing an huge amount of data. In addition, the decrease of sensors and acquisition systems cost and the development of Internet of Things (IoT) infrustructures have led industries to a digital transition.

Maintenance, intended as the activity of keeping a machinery or equipment in good condition by making repairs, correcting problems and periodic adjustments, has glimpsed the opportunity of a fast improvement moving from a classical Time-Based Maintenance (TBM) to a more modern Condition-Based Maintenance (CBM).

Many field have demonstrated interest in ML and AI in order to improve their maintenance strategies: Rogers et al. (Rogers, Guo, & Rasmussen, 2019) offer a review of Fault Detection and Diagnosis (FDD) methods for residential air conditioning systems, Datta et al. (Datta & Sarkar, 2016) report different pipeline fault detection methods, Meng et al. (Meng & Li, 2019) investigate Prognostics and Health Management (PHM) methods of lithium-ion batteries, Maciejewski et al. (Maciejewski, Treml, & Flauzino, 2020) deal with fault detection and diagnosis methods for induction motors, Li et al. (Li, Delpha, Diallo, & Migan-Dubois, 2020) study the application of Artificial Neural Networks (ANN) to to photovoltaic FDD, Liu et al. (Liu, Yang, Zio, & Chen, 2018) face another blooming subject for AI that is fault detection in rotating machinery while Kumar (Kumar, 2018) takes into account fault detection in a more specific context (bearings and gears), Shi et al. (Shi & O'Brien, 2019) give a comprehensive overview of automated FDD in buildings while Mirnaghi et al. (Mirnaghi & Haghighat, 2020) focus on large-scale HVAC, Gururajapathy et al. review fault location and detection in power distribution systems and, in the end, Habibi et al (Habibi, Howard, & Simani, 2019). are interested in fault detection techniques for wind turbine power generation.

On the other hand, a small number of works has demonstrated interest in detecting and diagnosing faults in Internal Combustion Engines (ICEs) and diesel engines by means of AI and, among these, the majority deals with marine engines: Xu et al. (Xu et al., 2017) propose a new belief rule-based (BRB) expert system for fault diagnosis, Wang et al. (S. Wang, Wang, & Wang, 2020) assemble a novel scheme for fault diagnosis based on k-means, Principal Component Analysis (PCA) and ANN, Cai et al. design a novel FDD method by combining Rule-Based (RB) algorithm and Bayesian networks (BNs) or Back Propagation Neural Networks (BPNNs), Wei et al. (Wei, Liu, Chen, & Ye, 2020) describe a new unsupervised ML algorithm based on One-Class Support Vector Machine (OCSVM), Affinity Propagation (AP) and Gaussian Mixture Model (GMM) for fault diagnosis, Xu et al. (Xu et al., 2020) define wear fault diagnostic model by using a multi-model fusion system based on Evidential Reasoning (ER), Wang (R. Wang, 2021) use an innovative hybrid fault monitoring scheme integrating the manifold learning and the isolation forest to monitor the state of marine diesel engine.

More specifically, searching works about FDD in small engines (nominal power lower than 155 kW_e, according to YANMAR categorization) has led to unfruitful results; moreover, no paper has shown interest in detecting anomalies concerning the fuel injection into engine's cylinders.

When a diesel engine is calibrated, the quantity of fuel injected in each cylinder is tuned in order to guarantee a certain level of emissions and the desired performance of the system. Due to aging, these values may get uncalibrated and need to be adjusted; usually only a periodical TBM is done and no continuous checks are performed unless the phenomena's evidences are tangible. In this work, a procedure to detect anomalies on the cylinders' injector as soon as a wrong quantity of fuel is inputted in one or more cylinders has been developed; the key questions of this activity are:

- Is it possible to detect a wrong injection anomaly by means of data-driven methods? What kind of accuracy can we expect?
- Is it necessary to install new sensors and which are the most important measurements?

The key findings are:

- Wrong injection can be detected both as general anomaly (at least one cylinder has a wrong injection) and as specific cylinder anomaly with different levels of accuracy.
- Standard sensors acquired by the Engine Control Unit (ECU) are sufficient to train a well performing model but, in case of more stringent requirements, it is highly suggested to measure the cylinder exhaust temperatures.

The paper is structured as follows: after a description of the case-study scenario (Section 2), the ML techniques selected are explained (Section 3), results are discussed (Section 4) and conclusions are drawn (Section 5).

2. Case-study scenario

Data from a four cylinders YANMAR small diesel engine have been exploited. Each sample presents a certain number of features (inputs to the model) and two labels (outputs of the model) referring respectively to a general anomaly (at least 1 out of 4 cylinders has a wrong injection) and to a localized anomaly (the 1_{st} cylinder presents a wrong injection). Data collection phase was not performed during this activity but was carried out previously; indeed the experimental setup, the sensors description and the engine features are extensively reported by Becciani et al. in (Becciani et al., 2019). For a better understanding Figure 1 shows the matrix of tests performed during data collection; in particular it shows the normalized injected fuel quantity at three engine load levels by varying the engine speed from 1000 rpm to 3000 rpm by steps of 250 rpm.

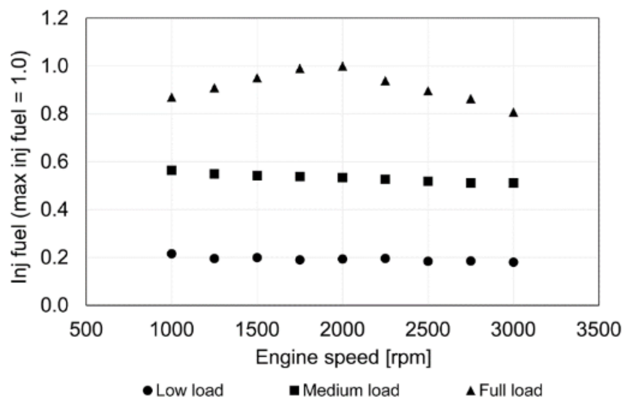

Figure 1. Injected fuel quantity at three load levels by varying the engine speed.

2.1. Steps definition

Measurements have been split in three groups based on the level of accessibility. In the 1_{st} step, signals directly acquired from the ECU (available from default engine sensors setup) have been considered, then, in the 2_{nd} step, additional sensors installed ad-hoc in the test-bench (but with the possibility to be added at sustainable costs on real engines) have been analysed. The 3_{rd} step involves also sensors for combustion analysis (unlikely to be added on real engines due to the unfeasible costs but interesting from a research point of view).

In particular the sensors groups are defined as follows:

- Step 1 (ECU data):
  - Rail pressure;
  - Intake manifold temperature;
  - Exhaust manifold temperature;
  - Intake manifold pressure;
  - Exhaust manifold pressure.
- Step 2 (test-bench data):
  - Air mass flow;
  - Air-Fuel ratio;
  - Oil pressure and temperature;

    - Fuel inlet pressure;
    - Exhaust downstream turbine temperature;
    - Exhaust downstream turbine pressure;
    - Exhaust temperature on cylinders.
- Step 3 (combustion analyser data):
    - Indicated Mean Effective Pressure (IMEP) of cylinders;
    - Max pressure of cylinders;
    - Max pressure angle of cylinders;
    - Burning of 50% of the fuel dose (MBF50) of cylinders.

Actually, in each step additional sensors would have been available but, due to the nature of the data collection phase, some measures had to be neglected.

2.2. Targets definition

As mentioned in the beginning of this section, the available dataset is constituted by two labels:

1. label *Anomaly* refers to a general anomaly, when at least 1 out of the 4 cylinders presents a wrong-injection quantity;
2. label *Cylinder1* refers to a particular anomaly located in the first cylinder. A wrong quantity of fuel has been injected in the first cylinder, no matter what is happening in the others cylinders.

The interest of YANMAR, reported in Table 1, is to detect an *Anomaly* using measurements from step 1 or measurements from Step 1 together with measurements from Step 2. On the other hand, it is interesting to evaluate the detection of *Cylinder1* using measurements from step 1 and 2 or measurements from step 1 and 2 together with measurements from step 3.

Table 1. Targets of the activity with relevant steps.

Sensors from steps	Referred as	Anomaly	Cylinder1
1	Step 1	X	
1 + 2	Step 2	X	X
1 + 2 + 3	Step 3		X

As matter of simplicity, when the authors refer to a model trained with the features belonging to a particular step, it is implicit that also the features from the previous steps have been considered; e.g. a model built using the 2_{nd} step features means that also 1_{st} step features have been used.

2.3. Dataset presentation

The dataset has a matricial shape of $447 \times (n + 2)$, where 447 are the examples populating the dataset, n is the number of features accordingly to the relevant step and 2 are the available labels (as explained in Section 2.2). When dealing with the *Anomaly* label, 1/3 of the examples are normal while 2/3 are anomalous. When considering the *Cylinder1* label, the acquisitions are quite balanced: 55.2% of the examples are anomalous while the 44.8% are normal. These considerations lead to face two classification problems since there are many examples of the anomalous behaviour.

The whole dataset has been split keeping the 80%-20% proportion in train-set and test-set; in order to tune hyper-parameters and to perform a fair comparison all algorithms described in Section 3 have been trained and validated using a k-fold cross-validation routine: a procedure that divides a limited dataset into k non-overlapping folds (Hastie, Tibshirani, & Friedman, 2009). In this work k has been set to 3.

3. Algorithms selection

As mentioned in Section 2.3, main properties of the dataset are a small amount of data, a small-medium quantity of features and two boolean target variables: therefore, the most suitable approach is to face a supervised classification problem by means of classical ML techniques. Some of the state-of-the-art statistical techniques to approach a 2-classes classification problem, described in this section, have been selected accordingly to (Hastie et al., 2009) and following authors' experience.

3.1. Linear Discriminant Analysis (LDA)

LDA is a discriminant approach that attempts to model differences among samples assigned to certain groups. The aim of the method is to maximize the ratio of the between-group variance and the within-group variance. When the value of this ratio is at its maximum, then the samples within each group have the smallest possible scatter and the groups are separated from one another the most (Stanimirova, Daszykowski, & Walczak, 2013). The two assumptions that must be fulfilled are the Gaussian distribution and the equal group covariances for the two classes; below (equation 1) the mathematical formulation where x is the vector of features and y is the target variable (0 stays for normal behaviour, 1 for anomaly):

$$\begin{cases} P\left(x \mid y=0\right) \sim \mathcal{N}(\mu_0,\, \Sigma_0^2) \\ P\left(x \mid y=1\right) \sim \mathcal{N}(\mu_1,\, \Sigma_1^2) \\ \Sigma_0 = \Sigma_1 = \Sigma \end{cases} \tag{1}$$

The algorithm finds a linear decision boundary separating the two classes described by equation 2:

$$\begin{aligned} 2(\Sigma^{-1}(\mu_0 - \mu_1))^T x + (\mu_0 - \mu_1)^T \Sigma^{-1} (\mu_0 - \mu_1) + \\ +2 \ln \left(\frac{P(y=1 \mid x)}{P(y=0 \mid x)} \right) = 0 \end{aligned} \tag{2}$$

3.2. Quadratic Discriminant Analysis (QDA)

QDA is a discriminant approach equivalent to LDA with the exception of the assumption of equal co-variances for the two classes (equation 3) (Ghojogh & Crowley, 2019); these hypotheses lead to a quadratic decision boundary (equation 4).

$$\begin{cases} P\left(x \mid y=0\right) \sim \mathcal{N}(\mu_0,\, \Sigma_0^2) \\ P\left(x \mid y=1\right) \sim \mathcal{N}(\mu_1,\, \Sigma_1^2) \\ \Sigma_0 \neq \Sigma_1 \end{cases} \tag{3}$$

$$\begin{split} x^T(\Sigma_0^{-1} - \Sigma_1^{-1})x + 2(\Sigma_1^{-1}\mu_1 - \Sigma_0^{-1}\mu_0)^T x + \\ +(\mu_1^T \Sigma_1^- 1\mu_1 - \mu_0^T \Sigma_0^- 1\mu_0)^T + \\ +2\ln\left(\frac{|\Sigma_0|}{|\Sigma_1|}\right) + 2\ln\left(\frac{P(y=1 \mid x)}{P(y=0 \mid x)}\right) = 0 \end{split} \tag{4}$$

3.3. Ensemble Discriminant Analysis (EDA)

Both LDA and QDA are discriminant algorithms that assign a class predicting the conditional distribution of each class. Analysing equations 2 and 4, it is trivial to note that the component deciding which class to assign is the decision function $\delta(x)$ (equation 5).

$$\delta(x) = \ln\left(\frac{P(y=1 \mid x)}{P(y=0 \mid x)}\right) \tag{5}$$

$\delta(x)$ can be seen as the confidence of the discriminant algorithm in predicting the class. When $\delta(x) \approx 0$ it means that, according to the algorithm, $P(y=0 \mid x) \approx P(y=1 \mid x)$ and the confidence is low; when $\delta(x) \gg 0$ the algorithm is confident to predict anomaly, otherwise when $\delta(x) \ll 0$ the algorithm is confident to predict a normal behaviour.

With these assumptions, the authors theorized a new technique coupling LDA and QDA. Below the procedure is explained:

Algorithm 1 EDA algorithm

- train the QDA model on the train-set
- predict by means of QDA both class $\hat{y}$ and decision function δ
- check δ for wrong prediction in train-set
- find δ^+_{max} and δ^-_{min} to identify a boundary of indecision
- train an LDA model on the train examples included in the QDA boundary of indecision
- use the LDA model to predict only the example of test in the QDA boundary of indecision
- take as final predictions of the QDA boundary of indecision outcomes with the higher decision function value (between the original QDA and the LDA)

3.4. k-Nearest Neighbours (kNN)

The intuition underlying kNN Classification is quite straightforward, examples are classified based on the class of their k nearest neighbours. More information can be retrieved in (Cunningham & Delany, 2007).

3.5. Support Vector Machines (SVM)

SVM is a classification technique based on the statistical learning theory proposed by Vapnik (Vapnik, 1999). The objective of the SVM algorithm is to find a hyper-plane in an n-dimensional space, where n is the number of features, that distinctly classifies the data points. To separate the two classes of data points, there are many possible hyper-planes that could be chosen; the SVM algorithm finds a plane that has the maximum margin between the data points of both classes. Maximizing the margin distance provides some robustness to the future classifications. Further information are well explained in (James, Witten, Hastie, & Tibshirani, 2013).

3.6. Classification And Regression Tree (CART)

As the name suggests, CART are suitable both for classification and regression problems. The difference lies in the target variable; this work aims to classify between healthy and anomalous status, indeed the dataset presents a 2-classes target variable. Classification and regression trees are prediction models constructed by recursively partitioning a dataset and fitting a simple model to each partition. Their name derives from the usual practice of describing the partitioning process by a decision tree. For a deeper description and a brief review the reader can refer to (Loh, 2011).

3.7. Artificial Neural Networks (ANN)

Since first years of 21th century, ANN have emerged as an important tool for classification. The vast research activities performed in the last 20 years in neural classification have established that neural networks are a promising alternative to various conventional classification methods. The advantage of neural networks lies in the following theoretical aspects. First, neural networks are data-driven self-adaptive methods in that they can adjust themselves to the data without any explicit specification of functional or distributional form for the underlying model. Second, they are universal functional approximators, indeed ANN can approximate any function with arbitrary accuracy (Zhang, 2000).

3.8. eXtreme Gradient Boosting (XGB)

In (Chen & Guestrin, 2016b), Chen and al. proposes the most used algorithms to solve classification problems in Kaggle competitions. Gradient boosting is a ML technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically CART. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function. The extreme in the name refer to the fact that is designed to push the extreme of the computation limits of machines to provide a scalable, portable and accurate library.

4. Results

All algorithms described in Section 3 have been trained (both for predicting *Anomaly* and *Cylinder1*, as summed up in Table 1) on the 80% of the whole dataset using a 3-fold cross-validation routine and then they have been tested on the remaining 20%.

Table 2 reports the algorithms' accuracy for the *Anomaly* case and Table 3 for the *Cylinder1* case; both tables are ordered by the descending accuracy value of the test-set. Accuracy has been used as evaluation metric in view of the fact that the anomalous/healthy samples are balanced within the dataset as explained in Section 2.3.

More complex models, XGB and ANN, show unstable results in the majority of the proposed tests due to the low amount of training examples available, indeed they are prone to overfitting. On the other hand, CART and kNN appear not feasible to catch engine's non-linearities. Unexpectedly, also SVM seems weak in predicting wrong injections. Focusing on the 2_{nd} step, the most interesting from a business point of view, best performing algorithms are discriminant analysis techniques and in particular QDA and EDA, the new method proposed by the authors.

Considering test-set performance, best algorithms' result for each case can be summarized as follows:

- *Anomaly*, step 1: kNN with a 88% of accuracy;
- *Anomaly*, step 2: EDA and QDA with a 93% of accuracy;
- *Cylinder1*, step 2: EDA and QDA with a 88% of accuracy;
- *Cylinder1*, step 3: ANN with a 91% of accuracy.

Discriminant analysis algorithms demonstrated to perform well in terms of robustness and accuracy; moreover, the decision function values generated by EDA and QDA (equation 5) for the misclassified examples always have a value close to zero. As explained in 3.3, this means that the algorithm claims to be uncertain about its prediction; otherwise, when the examples are well classified, decision function values present high values in module and consequently the algorithm predictions have a high confidence index.

Previous consideration opens the doors to a new perspective, indeed, considering both accuracy and decision function metrics, a new CBM approach can be designed to substitute the classical TBM adopted by engines' manufacturers.
For sake of simplicity, the authors define t the current time, x_t the vector of features at time t, $\delta(x_t)$ the decision function generated by the discriminant ML algorithm for the features x_t and T_m the injectors maintenance date (used in classical TBM) indicated in the engine's data-sheet; clearly, when a maintenance intervention is performed, the current time t is reset.
The above mentioned CBM approach can be detailed in the following enumeration:

1. for $t < T_m$:
   - when the ML algorithm detects an anomaly with high index of confidence $\delta(x_t)$, a maintenance intervention is needed;
   - when the ML algorithm detects an anomaly with low index of confidence $\delta(x_t)$, the engine can continue to operate normally;
   - when the ML algorithm predicts a normal behaviour no action is needed and the value of $\delta(x_t)$ is not considered.
2. for $t = T_m$ the scheduled TBM intervention is not performed and this casuistry is incorporated in the next point;
3. for $t \geq T_m$:
   - when the ML algorithm detects an anomaly, no matter the value of $\delta(x_t)$, a maintenance intervention is needed;
   - when the ML algorithm foresees a normal behaviour with low index of confidence $\delta(x_t)$, a maintenance intervention is needed;
   - when the ML algorithm foresees a normal behaviour with high index of confidence $\delta(x_t)$, the engine can continue to operate normally.

The engine's manufacturer has the duty to set a coherent threshold for the confidence index $\delta(x_t)$ in order to define the concepts of high and low confidence, minimizing the risks and maximizing costumer's profit.
The authors define a general rule to set the threshold on $\delta(x_t)$ as follows:

- in critical applications, where it is highly hazardous to incur in a faulty status, $\delta(x_t) \geq 0.85$ is considered a high index of confidence while $\delta(x_t) < 0.85$ a low index of confidence.
- in non-critical applications, where an anomaly is tolerable, $\delta(x_t) \geq 4.60$ is considered an high index of confidence while $\delta(x_t) < 4.60$ a low index of confidence.

As last consideration, a further analysis has been carried out to understand which measurements are more important to detect a wrong injection. Under a common agreement with the YANMAR engine experts, a feature importance analysis of the step 2 of the *Anomaly* case has been performed since it has been considered the most promising case-study from an application point of view.

Table 2. Accuracy values for algorithms predicting *Anomaly*.

	Anomaly							
step	**1**							
alg.	**kNN**	**XGB**	**ANN**	**LDA**	**SVM**	**CART**	**EDA**	**QDA**
acc. train	1.000	0.985	0.855	0.750	0.979	0.958	0.763	0.760
acc. val	0.775	0.812	0.781	0.753	0.747	0.736	0.758	0.756
acc. test	0.877	0.843	0.798	0.787	0.787	0.764	0.742	0.742
step	**2**							
alg.	**EDA**	**QDA**	**SVM**	**LDA**	**ANN**	**XGB**	**kNN**	**CART**
acc. train	0.969	0.938	0.919	0.868	0.987	0.961	1.000	0.878
acc. val	0.947	0.927	0.868	0.868	0.865	0.843	0.784	0.775
acc. test	0.933	0.933	0.876	0.865	0.820	0.798	0.764	0.764

Table 3. Accuracy values for algorithms predicting *Cylinder 1*.

	Cylinder 1							
step	**2**							
alg.	**QDA**	**EDA**	**ANN**	**XGB**	**SVM**	**LDA**	**kNN**	**CART**
acc. train	0.885	0.896	0.979	1.000	0.952	0.780	1.000	0.801
acc. val	0.868	0.865	0.809	0.820	0.801	0.733	0.767	0.663
acc. test	0.876	0.876	0.854	0.830	0.787	0.775	0.764	0.742
step	**3**							
alg.	**ANN**	**QDA**	**XGB**	**kNN**	**SVM**	**LDA**	**EDA**	**CART**
acc. train	0.972	0.806	1.000	1.000	0.822	0.787	0.876	0.956
acc. val	0.834	0.748	0.879	0.767	0.767	0.756	0.834	0.739
acc. test	0.910	0.888	0.865	0.831	0.809	0.798	0.778	0.776

Among the best performing algorithms, QDA is the most interpretable. Indeed, analysing the coefficients of the decision boundary (equation 4) quadratic term $(\Sigma_0^{-1} - \Sigma_1^{-1})$, it is straightforward to give a degree of importance to each feature: as physically expected, the cylinders exhaust temperatures are the main drivers to detect a wrong injection (Table 4).

5. Discussion

AI and ML are not extensively used to enhance small diesel engines' performance and, in particular, in-cylinders wrong injection detection or classification usually leans on classical strategies based on physical knowledge; the present work aims to emphasize the exploitability of state-of-the-art ML models to improve commonly used techniques in the ICE domain.
The authors compared various ML algorithms applied to three different groups of features, demonstrating that an high accuracy can be obtained in both the wrong injection classifications investigated: the case of a general *Anomaly* and the case of a particular cylinder anomaly (*Cylinder1*).
Best results have been pursued when detecting an *Anomaly* using as features the measurements collected both from the ECU and from the additional sensors installed in the ad-hoc test bench; in this case the accuracy obtained by the EDA algorithm proposed by the authors reached the value of 95%.
Moreover, also in the less performing tested cases, taking into account the decision function parameter, it is possible to grade an index of confidence that results very low in the misclassified examples. Thus, considering both metrics, accuracy and decision function, the authors proposed a CBM approach to maximize the operative hours of the engine minimizing failures risk.
In the end, a feature importance analysis have been explained detailing the most impacting measurements for detecting an *Anomaly* and, how expected by the domain experts, cylinders exhaust temperatures gained the greatest importance.

As future development, the authors plan to extend the research to other engines in order to validate the discussed results and to apply the *transfer learning* concept. Indeed, the possibility to design a detection methodology based on models trained off-line on data from another engine (with strong similarities) would bring many advantages from a commercial perspective.

Table 4. Analysis of the coefficients of the decision boundary quadratic term $(\Sigma_0^{-1} - \Sigma_1^{-1})$.

	Intake manifold pressure	**Fuel inlet pressure**	**Rail pressure**	**Exhaust temp. on cyl1**	**Exhaust temp. on cyl2**	**Exhaust temp. on cyl3**	**Exhaust temp. on cyl4**	**Oil temp.**
Intake manifold pressure	-8.0	0.1	6.9	43.5	2.6	-60.9	19.5	1.4
Fuel inlet pressure	0.1	3.3	1.2	23.7	-1.1	-18.9	-1.9	0.5
Rail pressure	6.9	1.2	-9.6	-23.0	-0.1	29.3	-11.8	6.3
Exhaust temp. on cyl1	43.5	23.7	-23.0	**409.8**	**-103.6**	**-422.7**	**118.1**	12.4
Exhaust temp. on cyl2	2.6	-1.1	-0.1	**-103.6**	**138.1**	-6.7	-38.4	9.8
Exhaust temp. on cyl3	-60.9	-18.9	29.3	**-422.7**	-6.7	**695.4**	**-253.9**	-19.9
Exhaust temp. on cyl4	19.5	-1.9	-11.8	**118.1**	-38.4	**-253.9**	**168.2**	1.0
Oil temp.	1.4	0.5	6.3	12.4	9.8	-19.9	1.0	-9.9

NOMENCLATURE

ANN	Artificial Neural Networks
AP	Affinity Propagation
BN	Bayesian Network
$BPNN$	Back Propagation Neural Network
BRB	Belief Rule-Based
$CART$	Classification And Regression Tree
CBM	Condition-Based Maintenance
ECU	Engine Control Unit
EDA	Ensemble Discriminant Analysis
ER	Evidential Reasoning
FDD	Fault Detection and Diagnosis
GMM	Gaussian Mixture Model
$HVAC$	Heating, Ventilation, and Air Conditioning
ICE	Internal Combustion Engine
$IMEP$	Indicated Mean Effective Pressure
IoT	Internet of Things
kNN	k-Nearest Neighbours
LDA	Linear Discriminant Analysis
$MBF50$	Burning of 50% of the fuel dose
$OCSVM$	One-Class Support Vector Machine
PCA	Principal Component Analysis
PHM	Prognostics and Health Management
QDA	Quadratic Discriminant Analysis
RB	Rule-Based
SVM	Support Vector Machines
TBM	Time-Based Maintenance
XGB	eXtreme Gradient Boosting

REFERENCES

Becciani, M., Romani, L., Vichi, G., Bianchini, A., Asai, G., Minamino, R., ... Ferrara, G. (2019). Innovative control strategies for the diagnosis of injector performance in an internal combustion engine via turbocharger speed. *Energies*, *12*.

Chen, T., & Guestrin, C. (2016a). *Xgboost.* `https://xgboost.readthedocs.io/en/stable/`.

Chen, T., & Guestrin, C. (2016b). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining.* ACM.

Chollet, F., et al. (2015). *Keras.* `https://keras.io`.

Cunningham, P., & Delany, S. (2007). k-nearest neighbour classifiers. *Mult Classif Syst*, *54*.

Datta, S., & Sarkar, S. (2016). A review on different pipeline fault detection methods. *Journal of Loss Prevention in the Process Industries*, *41*.

Ghojogh, B., & Crowley, M. (2019). Linear and Quadratic Discriminant Analysis: Tutorial. *arXiv:1906.02590 [cs, stat]*.

Habibi, H., Howard, I., & Simani, S. (2019). Reliability improvement of wind turbine power generation using model-based fault detection and fault tolerant control: A review. *Renewable Energy*, *135*.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference and prediction.* Springer.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: with applications* *in r*. Springer.

Kumar, S. (2018). Condition based maintenance of bearings and gears for fault detection – A review. *Materials Today*.

Li, B., Delpha, C., Diallo, D., & Migan-Dubois, A. (2020). Application of Artificial Neural Networks to photovoltaic fault detection and diagnosis: A review. *Renewable and Sustainable Energy Reviews*.

Liu, R., Yang, B., Zio, E., & Chen, X. (2018). Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, *108*.

Loh, W.-Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*.

Maciejewski, N. A. R., Treml, A. E., & Flauzino, R. A. (2020). A Systematic Review of Fault Detection and Diagnosis Methods for Induction Motors. In *2020 FORTEI-International Conference on Electrical Engineering (FORTEI-ICEE)*.

Meng, H., & Li, Y.-F. (2019). A review on prognostics and health management (PHM) methods of lithium-ion batteries. *Renewable and Sustainable Energy Reviews*, *116*, 109405.

Mirnaghi, M. S., & Haghighat, F. (2020). Fault detection and diagnosis of large-scale HVAC systems in buildings using data-driven methods: A comprehensive review. *Energy and Buildings*, *229*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*.

Rogers, A., Guo, F., & Rasmussen, B. (2019). A review of fault detection and diagnosis methods for residential air conditioning systems. *Building and Environment*, *161*.

Shi, Z., & O'Brien, W. (2019). Development and implementation of automated fault detection and diagnostics for building systems: A review. *Automation in Construction*, *104*.

Stanimirova, I., Daszykowski, M., & Walczak, B. (2013). Chapter 8 - robust methods in analysis of multivariate food chemistry data. In F. Marini (Ed.), *Chemometrics in food chemistry* (Vol. 28). Elsevier.

Vapnik, V. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, *10*.

Wang, R. (2021). Research on the fault monitoring method of marine diesel engines based on the manifold learning and isolation forest. *Applied Ocean Research*.

Wang, S., Wang, J., & Wang, R. (2020). A novel scheme for intelligent fault diagnosis of marine diesel engine using the multi-information fusion technology. *IOP Conference Series: Materials Science and Engineering*, *782*, 032022.

Wei, Y., Liu, H., Chen, G., & Ye, J. (2020). Fault Diagnosis of Marine Turbocharger System Based on an Unsupervised Algorithm. *Journal of Electrical Engineering & Technology*, *15*.

Xu, X., Yan, X., Sheng, C., Yuan, C., Xu, D., & Yang, J. (2017). A Belief Rule-Based Expert System for Fault Diagnosis of Marine Diesel Engines. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*.

Xu, X., Zhao, Z., Xu, X., Yang, J., Chang, L., Yan, X., & Wang, G. (2020, feb). Machine learning-based wear fault diagnosis for marine diesel engine by fusing multiple data-driven models. *Knowledge-Based Systems*, *190*.

Zhang, P. (2000). Neural networks for classification: A survey. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, *30*.

BIOGRAPHIES

Piero Danti is an Electronic and Automation engineer, he received his Master degree at the University of Florence with a master thesis titled "Coordinated collective motion of a sensor network for optimum target tracking using Extended Kalman Filter" developed at the Technical University of Munich (TUM) in Germany. In June 2011, he joins the french firm Altran where he works as RAMS consultant. In March 2012 he starts to work as Control and Software engineer in the Oil and Gas business for the General Electric supplier Promel. After 3 years of activities, both as PLC/HMI programmer and field service engineer, he has the chance to join the european R&D department of the Japanese company YANMAR where he deepens Machine Learning and Artificial intelligence expertise. His main research fields are ML and AI applied to statistical analysis, anomaly detection and time-series forecasting. Currently he is a PhD candidate at the University of Florence dealing with Machine Learning applied to PHM.

Giovanni Vichi received the degree in Doctor of Philosophy in Energy and Innovative Industrial Technology in April 2013 with a thesis work titled "Numerical-experimental methodologies for the development of highly efficient engines for two wheel vehicles". His main research fields are the one-dimensional engine simulation, engine control, and engine measurement technique. As research fellow, he participated in numerous founded projects. Since 2018 is working for Yanmar R&D Europe, now heading the Engine and Powertrain group.

Minamino Ryota received the Master degree in Social Science of Energy at the Graduate School of Energy Science in Kyoto in 2010. His main research fields are the one-dimensional engine simulation, engine control, and engine measurement technique. In 2010 he joins YANMAR and in 2020 start working for Yanmar R&D Europe as senior researcher in the Engine and Powertrain group.

APPENDIX

6. HYPER-PARAMETERS SELECTION

The analyses presented in this work have been carried out using Python and all algorithms have been imported form the scikit-learn library (Pedregosa et al., 2011) (version 0.24.2) with the exception of ANN that has been assembled using Keras (Chollet et al., 2015) (version 2.7.0) and XGB that is provided by a dedicated library (Chen & Guestrin, 2016a) (version 1.3.3). In particular:

- LDA uses default hyper-parameters;
- QDA uses default hyper-parameters;
- EDA has been designed by authors ensembling LDA and QDA with default parameters;
- kNN hyper-parameters are listed in Table 5;
- SVM hyper-parameters are listed in Table 6;
- CART hyper-parameters are listed in Table 7;
- ANN hyper-parameters are listed in Table 8;
- XGB hyper-parameters are listed in Table 9;

Hyper-parameters not reported in the tables have been set to the relative library's default value.

Table 5. Best hyper-parameters configuration for kNN algorithm.

	kNN			
	Anomaly		**Cylinder 1**	
step	**1**	**2**	**2**	**3**
n_neighbors	1	16	1	1
weights	uniform	distance	distance	distance
algorithm	ball_tree	kd_tree	auto	auto
leaf_size	67	67	30	30
p	4	1	1	1

Table 6. Best hyper-parameters configuration for SVM algorithm

	SVM			
	Anomaly		**Cylinder 1**	
step	**1**	**2**	**2**	**3**
C	200	9.73	8.52	500
gamma	5	0.1	0.385	1
kernel	rbf	rbf	rbf	linear

Table 7. Best hyper-parameters configuration for CART algorithm.

	CART			
	Anomaly		**Cylinder 1**	
step	**1**	**2**	**2**	**3**
max_depth	None	None	None	None
min_samples_split	2	18	2	2
min_samples_leaf	2	7	6	2
max_leaf_nodes	44	12	12	45
min_impurity_decrease	0	0	0	0
max_features	5	12	3	19
ccp_alpha	0	0	0	0

Table 8. Best hyper-parameters configuration for ANN algorithm.

	ANN			
	Anomaly		**Cylinder 1**	
step	**1**	**2**	**2**	**3**
hidden_layers	1	1	1	1
hidden_activation	ReLu	ReLu	ReLu	ReLu
hidden_units	16	16	128	64
output_activation	sigmoid	sigmoid	sigmoid	sigmoid
output_units	1	1	1	1
batch_size	64	64	64	128
learning_rate	0.5	0.5	0.05	0.1
optimizer	Adam	Adam	Adam	Adam

Table 9. Best hyper-parameters configuration for XGB algorithm.

	XGB			
	Anomaly		**Cylinder 1**	
step	**1**	**2**	**2**	**3**
n_estimators	50	250	200	150
subsample	0.5	0.5	1	1
colsample_bytree	1	0.5	0.5	0.5
max_depth	7	7	3	7
gamma	0.25	0.5	0.25	0.25
learning_rate	0.25	0.25	0.25	0.25
eval_metric	logloss	logloss	logloss	logloss