An unsupervised machine learning technique to identify knock from a knock signal time-frequency analysis

Authors: Benjamín Pla, Joaquín De la Morena, Pau Bares, Alexandra Aramburu

Affiliation: CMT-Motores Térmicos, Universitat Politècnica de València
Camino de Vera s/n, CP46022, Valencia, Spain

Corresponding author: Alexandra Aramburu
Email address: alaror@mot.upv.es
Telephone: +34963877007 (ext. 76524)

An unsupervised machine learning technique to identify knock from a knock signal time-frequency analysis

Abstract

Knock is a characteristic phenomenon of spark ignition engines that can result in lower fuel efficiency and higher combustion temperatures. When it is not adequately identified and controlled, knock might lead to intensive engine damage. Knock detection methods are based on recognising the large in-cylinder pressure oscillations that resonate in the combustion chamber as a consequence of the rapid auto-ignition of the end-gas. However, in-cylinder pressure sensors may be invasive and expensive. In this work, a method for knock classification is proposed based on the knock sensor measurement, which is a low-cost and easy-to-implement vibration sensor. Unfortunately, the vibration of the engine block contains information from several phenomena and only intensive knock can be identified by classical methods.

To counteract the amount of noise, Machine Learning (ML) techniques are applied to the knock sensor signal. A training database composed of several operating conditions is used to extract the main characteristics of knock. The signal is analysed in the time-frequency domain by using the short-time Fourier transform (STFT), the features of the signal are obtained by singular value decomposition (SVD), and finally, support vector machine (SVM) techniques are used to identify outliers. Experimental data from a spark ignited engine at 2000 rpm and 3000 rpm with spark advance steps was used to train and validate the method. Results are compared to state-of-the-art knock recognition methods based on pressure and knock sensors.

Keywords: knock detection, machine learning, vibration sensor, outlier detection, unsupervised learning

1. Introduction

The automotive industry is performing intense modifications to drastically reduce pollutant emissions to acceptable levels, compatible with the Paris Agreement climate objectives, by reducing the greenhouse gases to ensure that the earth's temperature does not increase more than 2^{o}C [1].

Although electrification is being proposed as an alternative to substitute internal combustion engines, several studies have alerted that because of current limitations in energy production, a combination of electric vehicles (EV) and internal combustion (IC) engines will be necessary to reduce the total emissions produced by transportation sector [2]. Furthermore, the industrial and economic costs of changing the car fleet push the predictions to assume a significant amount of cars with internal combustion engines in the following decades [3].

Regarding IC engines, Spark Ignited (SI) engines represent a mature technology with low pollutant emissions, such as oxides of nitrogen, and because of their high specific power (compared with compression ignited engines) are good candidates to be used in hybrid systems. Hybrid vehicles have attracted attention due to their potential in mechanical efficiency and braking energy recovery [4]. Moreover, new SI combustion modes based on a lean combustion [5] or based on new fuels [6] have demonstrated the potential of being compatible with GHG emission objectives.

Despite this, the maximum compression ratio and the efficiency, and hence the CO2 emissions of SI engines, are limited by knock [7, 8]. Knock is an undesired phenomenon that consists in the autoignition of the end-gas due to the pressure and temperature increase in the combustion chamber caused by the SI combustion flame evolution. The main consequence of knock is the excitation of the combustion resonance due to the rapid and uncontrolled combustion of the end-gas. Knock identification becomes crucial not only for efficiency but also to ensure a safe operation and avoid dangerous operating conditions [9].

In research and calibration, the in-cylinder pressure sensors represent a key tool to detect knock: normally, the maximum amplitude pressure oscillation (MAPO) is used to determine the knock occurrence [10, 11]. Yet, the threshold required to identify knock should be high enough to differentiate knock from other phenomena leading to the misclassification of low-knocking conditions. Additional metrics include the integral of modulus of pressure gradient (IMPG) or the integral of modulus of pressures oscillations (IMPO) as mentioned in Zhen et al. [12]. Other criteria and knock identification methods have been studied in the past, not taking into account only the pressure oscillation but other parameters such us the heat release rate [13]. More recently, Bares et al. [14] proposed a new definition comparing the excitation of the cylinder resonance caused by the combustion and the autoignition. Whereas Kim [15] developed a classification model using machine learning algorithms trained with in-cylinder pressure data. Both using the previous metrics, such as MAPO, for comparing the results obtained.

However, the prohibitive cost of in-cylinder pressure sensors with an accurate performance at high frequency precludes its implementation in commercial vehicles. Current SI engines use a specific sensor (knock sensor) which analyses the block vibration to determine the knock occurrence. Yet, in-cylinder pressure sensor is used to calibrate the system and determine the maximum allowable vibration, which is stored in the electronic control unit (ECU) as a threshold for the knock sensor signal.Knock index metrics using the block vibration are also defined [16, 17]. Most of them being similar to the ones used with the in-cylinder pressure, particularly, when using the filtered signal. Other studies proposed ion current signal to detect knock as an alternative to pressure and knock sensors [18].

The knock sensor measures the block vibration and, henceforth, is affected by the resonant modes in the combustion chamber, but also by other nonlinear phenomena, such as piston slap or valve opening [19]. Several methods have been proposed to improve knocking detectability focused on the knock sensor signal: Napolitano et al. [20] proposed a Wiebe model to predict the combustion phasing in order to discern the end-gas autoignition from normal combustion. Siano and Panza [21] published a nonlinear ARX model to identify knocking conditions with a black box environment while Petrucci et al. [22] applied machine learning techniques combined with a 1D-model to feed the system.

Physical models such as the one proposed by Napolitano et al. [20] are restricted to model precision, while black-box models such as the ones proposed in [21, 22] might suffer from over-fitting and meaningful errors in non-tested operating conditions. In this context, this paper presents a machine learning technique based on the knock sensor signal able to achieve classification results similar to the ones obtained by direct methods such as the in-cylinder pressure sensor approach. The method is able to provide information at each cycle without requiring a threshold to be provided at each operating point making it suitable for real-time applications. The algorithm consists of three phases: Firstly, employing single value decomposition (SVD), the main patterns of the knock spectrogram are obtained. Secondly, the most important patterns for knock recognition are defined, and finally, support vec-

tor machine (SVM) techniques are used to discern outliers and hence non-usual combustion metrics. The method has been tested with experimental data with various engine speed, spark advance and EGR conditions.

2. Experimental setup

Experimental tests were carried out in a turbocharged four-cylinder SI engine, which main specifications are summarized in Table 1. The engine is coupled to a fully instrumented test bench with a dynamometric brake AVL AFA 200/4-8EU for speed and torque control.

Table 1: Engine Characteristics

	Units	Value
Displaced volume	[cc]	1300
Stroke	[mm]	81.2
Bore	[mm]	75
Compression ratio	[-]	10.6:1
Combustion	[-]	SI
Number of cylinders	[-]	4
Fuel type	[-]	Gasoline
RON	[-]	95
Fuel injection system	[-]	GDI

The engine block vibration was measured with a knock sensor (Bosh KS4-R2) installed between cylinders 2 and 3, as depicted in Fig. 1. The frequency range of the sensor is 3kHz to 30kHz. For evaluation and comparison purposes, information from in-cylinder pressure sensors was also acquired through instrumented spark-plugs (AVL ZI33). Knock and pressure sensors were connected to a high-frequency acquisition system and synchronized using a research encoder with a 0.2 cad sampling resolution, obtaining 3600 measurement points for each cycle.

The complete database was assembled with data from 5 test campaigns in which variations of SA were made at various operating conditions, maintaining an air/fuel stoichiometric ratio. Table 2 gives a summary of the operating conditions of each test.

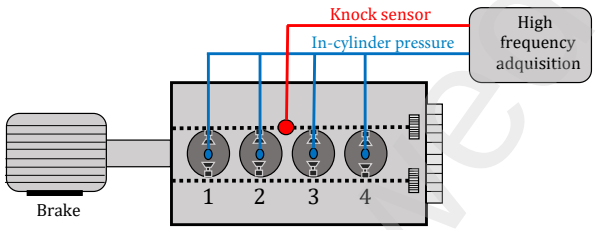

Figure 1: Knock and pressure sensors position

Table 2: Operating conditions of each experimental test

N [rpm]	P_{adm} [bar]	SA range	No of cycles
2000	1	4 - 10	4690
2000	1	7 - 12	5236
3000	0.8	5 - 26.5	6949
3000	1	6 - 16	5692
3000	1	7 - 14	7923
		Total	30490

3. Method description

The method developed in this paper can be divided into three stages:

1. **Signal processing:** The knock sensor signal is acquired time or crank angle based and it is processed with a short time Fourier transform (STFT). Additionally, the signal was normalized taking into account current operating conditions.
2. **Feature extraction:** Once the signal is processed, single value decomposition (SVD) is applied to extract the main modes that represent the spectrogram. From a training dataset the modes that represent the engine behaviour are computed and only two modes are sufficient to determine the knock occurrence. In real time, the ECU only needs to identify the intensity of these two modes.

3. **Knock detection:** Support Vector Machine (SVM) is used to classify each cycle. If the intensity of the representative modes lie on the expected area normal combustion is identified, otherwise an outlier, and hence knock, is detected.

Although the algorithm can be considered an unsupervised technique, as knock is recognized with SVM from an analysis of the signal, it is also true that the selection of the representative SVM modes is made with supervision. An overview of the methodology is presented in the flowchart depicted in Fig. 2, while the following subsections provide a detailed description of the three stages.

3.1. Signal Processing

Short-time Fourier transform is a technique that gives a time-frequency representation from a time-domain signal. It uses a sliding window function over the signal and obtains the energy spectral density of each segment. Different authors have already applied this technique to evaluate the characteristic frequencies in combustion chambers [23, 14, 19].

The STFT was configured with a 128-point Hanning window with 96 overlapping points and a DFT length of 1024 samples. Fig. 3 presents the spectrograms obtained for cycles with different knocking levels for 2000 and 3000 rpm. In cycles with no knock, on the left plots, the primary frequency can be seen around 5 to 10 kHz, which corresponds to the first resonance mode. It also shows some diffuse background noise. For cycles with knock (right plots), more resonance frequencies appear above 10 kHz, which is in accordance with the analysis made by Naber et al. [24]. Moreover, the ones at 10 kHz are prolonged for more than 40 CAD. Below 4 kHz are the components related to the piston movement and the combustion process. The presence of distinct features in the spectrogram of the cycles with knock is essential for the development of the proposed technique.

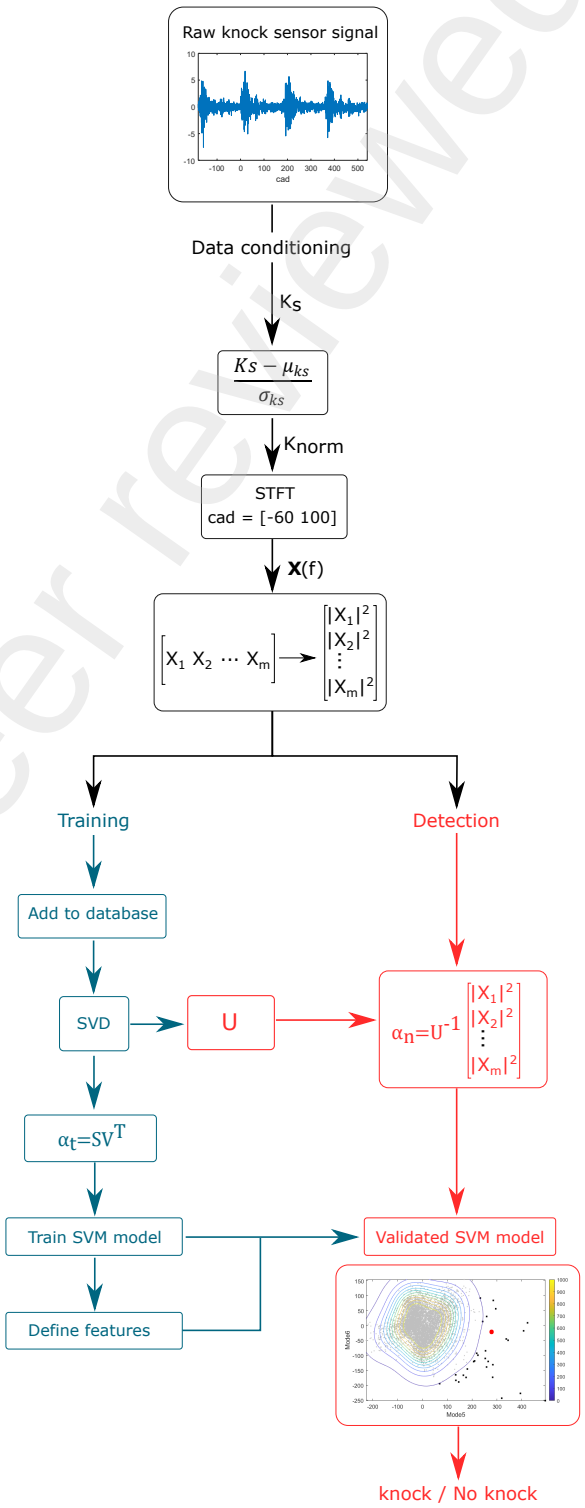

Figure 2: Knock detection method flowchart. (1) Black boxes: signal processing. (2) blue boxes: feature extraction. (3) red boxes: knock detection.

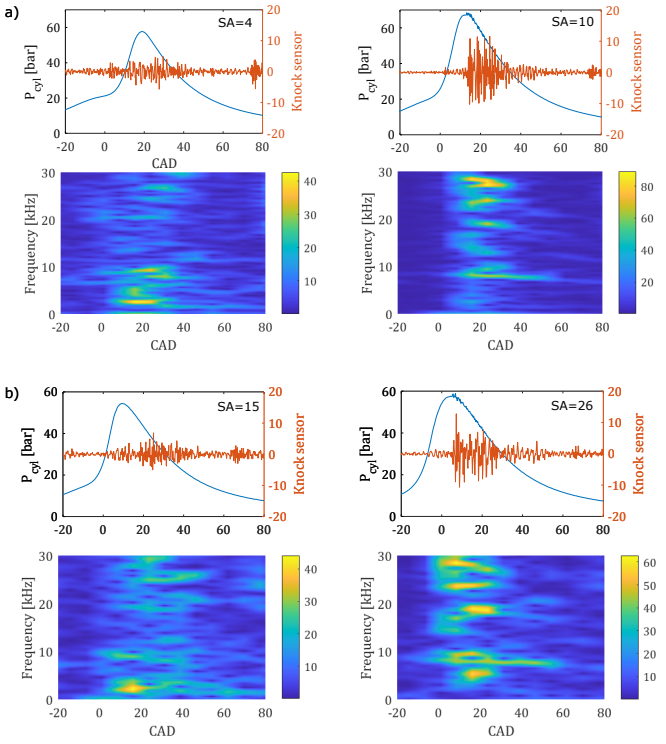

Figure 3: Spectrogram representation of the knock sensor signal for low (left) and high (right) MAPO values. (a) 2000 rpm and $P_{adm} = 1bar$ and (b) 3000 rpm and $P_{adm} = 0.8bar$

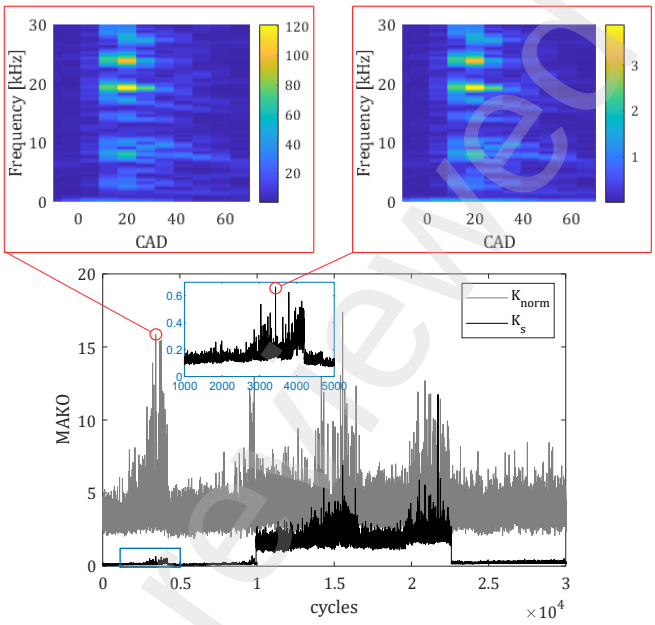

Figure 4: Comparison between original and normalized knock signal. The spectograms above show the STFT magnitude for the same point in both cases.

Additionally, the filtered signal was normalized according to equation 1 where μ_{ks} and σ_{ks} represent the mean and standard deviation from data in each cycle. This normalization was due to the difference in the maximum amplitude achieved by the knock sensor between tests, as can be seen in Fig. 4. Using normalized data prevents having skewed features. It should be noted that, although this procedure changes the STFT magnitude value, it did not affect the spectrogram's main characteristics, as observed in the upper plots of Fig. 4, where spectrograms of the same point before and after normalization are depicted.

$$K_{norm} = \frac{K_s - \mu_{ks}}{\sigma_{ks}} \tag{1}$$

The matrix obtained from the spectrogram in each cycle is reshaped into a column vector. The vector's length will depend on the resolution defined previously to calculate the STFT, yet it will have the same length regardless of the engine velocity (2000 rpm or 3000 rpm). The column vector will be appended to the database if the cycle is used for training the model. On the other hand, if the cycle is for detection, the coefficients associated with each principal mode are obtained for further analysis.

3.2. Feature extraction

The singular value decomposition of X is a factorization of that matrix into three other matrices: U, V and Σ, as in equation 2). U and V are unitary orthogonal matrices hierarchically arranged in terms of their ability to represent the columns and row spaces of X. Whereas Σ is a diagonal matrix with non-negative, decreasing values (σ_i) known as singular values [25].

$$X = U\Sigma V^T \tag{2}$$

The relevance of this tool lies in its ability to extract the main features of any matrix X. This provides two advantages: the first one is related to the fact that matrix U will become a basis of the original data. Hence, any column vector x in the space of the database can be obtained as a linear combination of column vector U and

some coefficients α.

The second advantage is that since Σ is ordered by importance, an approximation of X can be obtained by taking a subset of Σ comprised of the first r singular values (see equation 3). This approximation will be the closest rank-r matrix to X in the least square sense [26]. Meaning that, instead of using the complete database, a reduced version can be used decreasing the computational cost.

$$X = \sum_{i=1}^{m} \sigma_i u_i v_i^T \approx \sum_{i=1}^{r} \sigma_i u_i v_i^T; \quad r < m \tag{3}$$

where, u_i and v_i are the column vectors of matrices U and V, respectively. And m is the number of columns in X.

The experimental data were randomly divided into a training set and two validation sets for the feature selection stage, and for the SVM detection analysis. The training set had a total of 10672 cycles corresponding to 35% of the collected data, and the validation sets had 6606 cycles each one, obtained from the remaining 65%. The number of knocking cycles in each set, according to the reference method, is detailed in Table 3. Also, the knock percentage is indicated, which varies between 0.36% to 0.48% according to the set.

Table 3: Percentage of knocking cycles in the training and validation sets

Set	Total cycles	Knock cycles	Knock [%]
Training	10672	43	0.40
Validation 1	6606	32	0.48
Validation 2	6606	24	0.36
Test	6606	26	0.39
TOTAL	30490	125	0.41

A training input matrix $X \in \mathbb{R}^{n \times m}$ was used to obtain the main modes U from the SVD factorization. This matrix is comprised by $x_j (j = 1 : m)$ column vectors, where m denotes the number of cycles ($m = 10672$). Each column vector contains n elements from the reshaped STFT magnitude. The modes obtained represent the main features of the STFT database and will be used to obtain the coefficients in the detection stage. In this stage, the coefficients of X in the U basis (α_t) are obtained with equation 4.

$$\alpha_t = \Sigma_r V_r^T \tag{4}$$

These coefficients reveal the importance of a particular mode for a given cycle. It is worth noting that these modes will be related mainly to the predominant frequencies captured by the knock sensor along a complete cycle, particularly during the combustion stage. For instance, Fig. 5 shows some modes obtained from the training database. Mode 1 represents the mean STFT from the whole set. Mode 2 represent mostly variations around 10 Hz, whereas the other chosen modes capture higher resonance frequencies. Therefore the SVM model will be trained for different mode combinations to find the one that best detects the knocking cycles.

The total number of modes are defined by the amount of columns in X, i.e., the number of cycles in the set. However, the first modes have more relevance than the last ones, this can be seen in Fig. 6 that depicts the singular value associated with each mode. Because there is almost a order of magnitude between mode 1 and mode 20, meaning the 20th mode value is merely a 10% of the first mode, only a combination of the 20 first modes will be considered.

The number of modes to be considered for knock identification is a trade-off between computational cost and precision. In this case combinations of 2, 3 and 4 modes have been tested.

3.3. Knock identification

One of the major challenges in this work is the presence of imbalanced data. This occurs when class distribution is unequal in the training set. For instance, knocking is an event that appears sporadically under some specific oper-

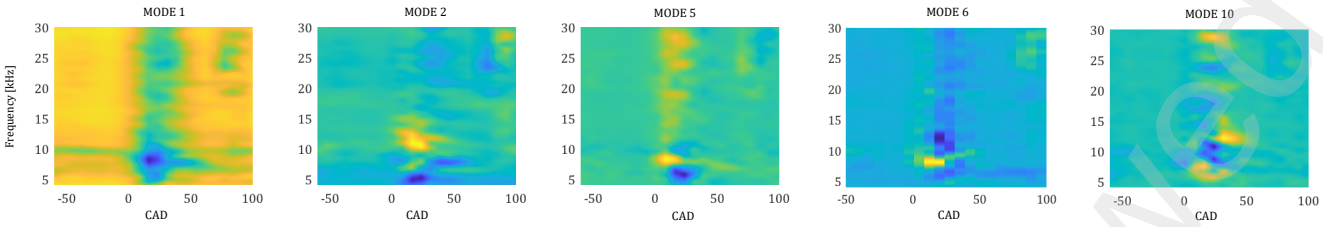

Figure 5: Modes obtained from the left singular vector U

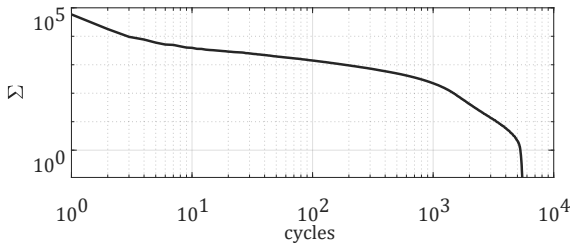

Figure 6: Singular values from the factorization of the training database

ating conditions, creating a biased data distribution as it is illustrated in the probability density function (pdf) in Fig. 7. Most classification algorithms require a similar amount of observations for each class to get optimal results. Hence, poor predictive performance may occur when a minority class exists. In this work, the problem is addressed using one-class support vector machine (OC-SVM).

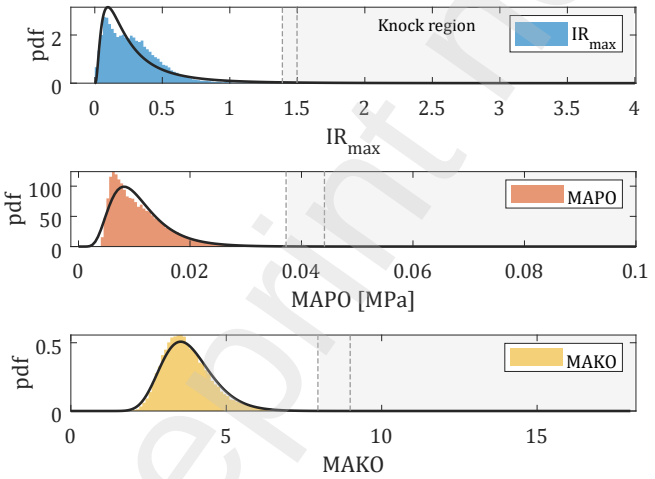

Figure 7: Data distribution and the corresponding log-normal pdf for different knock detection techniques. The dashed lines indicate the minimum and maximum classification threshold used in each case.

SVM is a technique that classifies data by finding the margin that best separates features into different domains. If the training data is not linear, it is mapped into a higher dimension space by means of a kernel function. The instances closest to the separating hyperplane are called support vectors. In particular, one-class SVM is an unsupervised model that, unlike regular SVM, does not require target labels for the training process. Instead, it aims to determine the behaviour of data from a single class and tag the elements that deviate from the established profile as outliers. For more detailed information, refer to Schölkopf et al. [27].

This study used a radial basis function (RBF) as the kernel function [28]. The principal hyperparameters related to the model are the outlier fraction (OL_f), ν and γ. OL_f expresses the expected proportion of outliers in the training data. ν is a regularization parameter from the OC-SVM objective function that defines a lower bound to the number of support vectors and an upper bound to the number of outliers the model allows. A large value of ν leads to more support vectors and, therefore, a flexible decision boundary. Finally, γ is a kernel coefficient related to a single training point's range of influence. A small γ is related to a wide variance in the kernel function, meaning that two points can be considered similar even if they are not close together.

4. Results and discussion

In this work, three other knock detection methods will be considered to compare and validate the results obtained.

The first method is the maximum amplitude pressure oscillation (MAPO) index, which is one of the most extended for knock detection; it is calculated as in equation 5. A threshold value, defined according to the operating point, separates cycles with knock from the ones with normal combustion.

$$MAPO = max|p_{bp}| \tag{5}$$

Where, p_{bp} is the filtered pressure signal using a band-pass filter between 4 kHz and 20 kHz.

The second method was proposed in Pla et al. [29]; it considers the intensity of the in-cylinder pressure oscillation and its angular evolution to obtain a new definition of knock. It deduces a minimum auto-ignition resonance indicator (Ir_{min}) and compares it with the resonance indicator (Ir) developed in Guardiola et al. [30], as in equation 6. It gives a more complex analysis when compared with MAPO, but it can detect low knocking cycles and reduce false negative cases. Thus, the dataset analysed with this method will be considered the ground truth dataset. Henceforth, it will be referred to as IR method.

$$if \quad max(I_r) \geq Ir_{min}(\alpha_{maxIr}) \rightarrow \quad knock \tag{6}$$

Both previous methods are based on in-cylinder pressure sensors. However, to be able to compare with another knock sensor-based technique, the third method is proposed. It uses the knock sensor signal in a procedure analogous to MAPO [16]; thus, it will be referred to as MAKO. A high-pass filter is applied to the acquired data with a cut-off frequency of 10kHz before obtaining the maximum amplitude. A threshold also needs to be defined.

The experimental data were labelled according to the IR method. A total of 125 knocking cycles were detected on the complete database. This number was used to define the thresholds for MAPO and MAKO methods so that all of them detect the same amount of knocking cycles. Results obtained from these techniques and the one proposed in this paper will be compared and evaluated against results from the IR method. A scheme of this methodology is presented in Fig. 8.

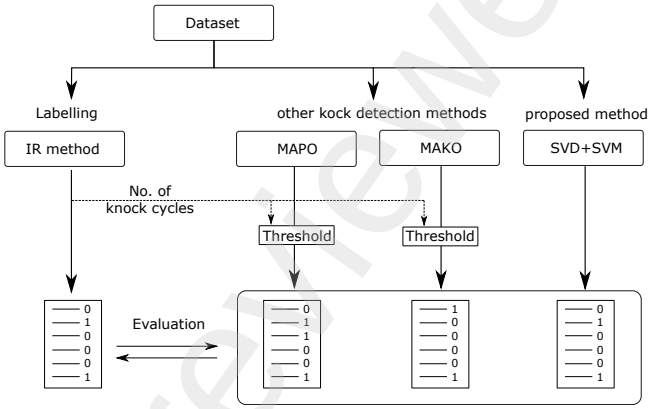

Figure 8: Labelling and performance evaluation approach

To properly assess the effectiveness of the method, metrics that will adequately value the minority class are essential. Standard classification metrics such as accuracy do not represent the model's performance properly when dealing with imbalanced data [31, 32]. Accordingly, the evaluation criteria considered in this paper include:

- Precision: measures the number of correct positive predictions. It is defined as the ratio of correct positive predictions out of all positive predictions made.

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

- Recall (Sensitivity): measures the fraction of actual positive class instances effectively predicted, providing an indication of missed positive predictions.

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

- F_1-score: defined as the harmonic mean between precision and recall.

$$F1 = 2\frac{Precision \cdot Recall}{Precision + Recall} \tag{9}$$

These metrics are derived from the concept of the confusion matrix, which gives a summary of the prediction results. For binary classification, the confusion matrix has four components: true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN).

4.1. Mode selection

First, all the possible combinations with 2, 3 and 4 modes were tested with the training set. The best three combinations according to the F1-score were evaluated with the validation sets. For illustration purposes, Fig. 9 displays a matrix with the F1 score for each pair of modes in the case of 2-mode combinations. It can be seen that combinations including modes 1, 5 and 8 give higher results.

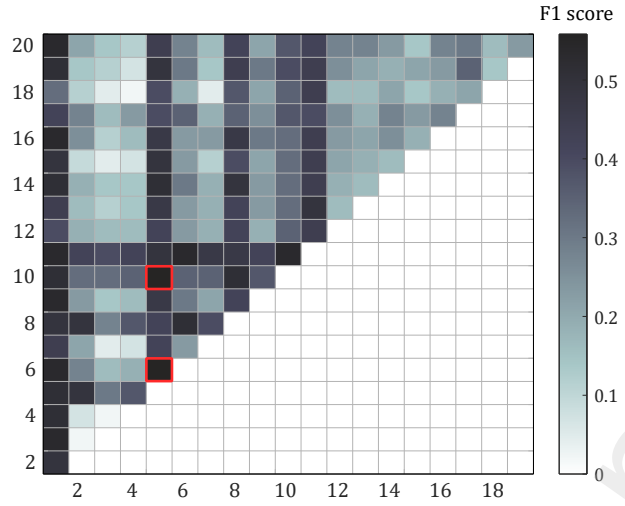

Figure 9: F1 score for 2-mode combinations. The red squares indicate the modes with the higher F1 score attained during the training stage.

This is corroborated in Table 4 where the results from the different mode combinations with 2, 3 and 4 elements are presented, highlighting the ones with higher mean F1-score according to the validation tests.

One of the reasons for using validation tests is to confirm that the best combination of modes selected during the training stage is also suitable for other sets. Using more than two modes does not reduce the error but turns in over-fitting. Consequently, a combination with two modes (5-6) is selected as it will also require less computational cost.

4.2. Knock detection

Table 4: F1 score for different mode combinations.

		F1 score	
k	**Modes**	**Train**	**All**
2	5 - 6	0.56	**0.65**
	5 - 10	0.56	**0.50**
	1-6	0.53	**0.55**
3	1-5-6	0.63	**0.61**
	1-5-10	0.60	**0.58**
	2-5-6	0.60	**0.63**
4	1-5-6-9	0.65	**0.60**
	1-2-5-9	0.65	**0.58**
	1-2-5-6	0.63	**0.64**

Once the mode combinations are decided, the OS-SVM model's hyperparameters are calibrated, namely the outlier fraction, ν and γ. For OL_f and ν, a 3-fold cross-validation was made. On the one hand, a variation of OL_f between 0.001 and 0.006 is made. The influence of this hyperparameter will depend on the knock percentage of the training set, which will change in every fold. Nevertheless, in all cases, it was around $OL_f = 0.004$. An uncertainty of ± 0.001 does not affect the F1-score significantly; however, higher values decrease the efficiency of the method. On the other hand, ν is varied from 0.01 to 0.8 in each fold. Fig. 10 shows the mean F1-score obtained for the training and validation sets; in both cases, F1 score increases up to $\nu = 0.2$ where the model becomes insensitive to variations of this hyperparameter. Because higher ν values increase the number of support vectors, making the model prone to overfitting, ν was set to the lowest possible value in the area of flat F1 ($\nu = 0.2$). Finally, γ was selected automatically by the Matlab software using a heuristic procedure. For the selected mode combination $\gamma = 1.4$.

Fig. 11 shows the classification results from the OC-SVM. Points detected as knocking cycles are indicated with squares and can be compared with those detected

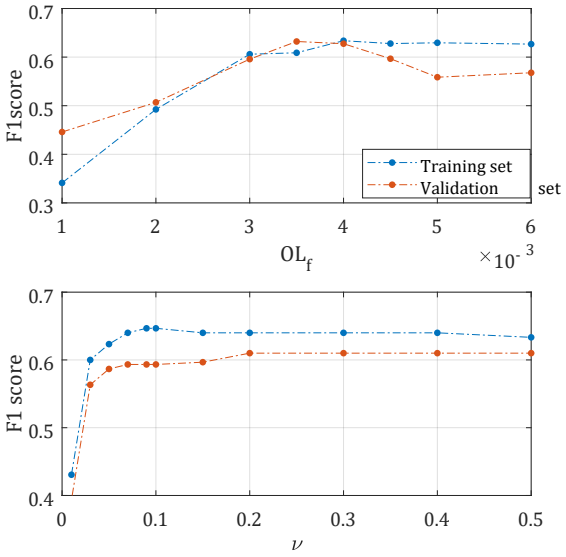

Figure 10: Mean F1 score values for hyperparameters calibration. **Upper: Outlier fraction. Lower:** ν

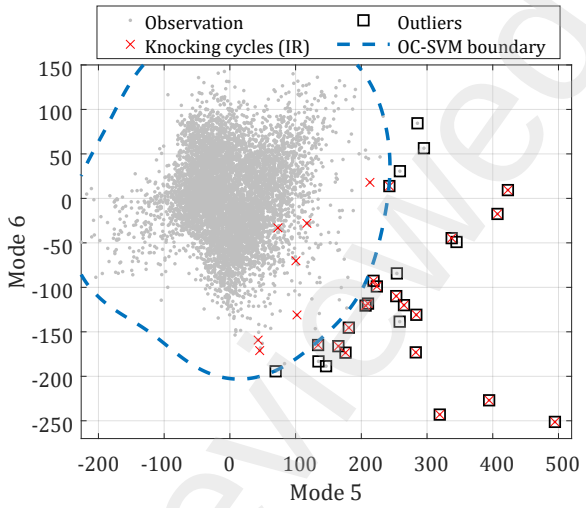

Figure 11: Outlier detection results for test set. The axis represent the coefficients of each observation associated to mode 5 and 6. The dashed line represents the decision boundary used by the OC-SVM model.

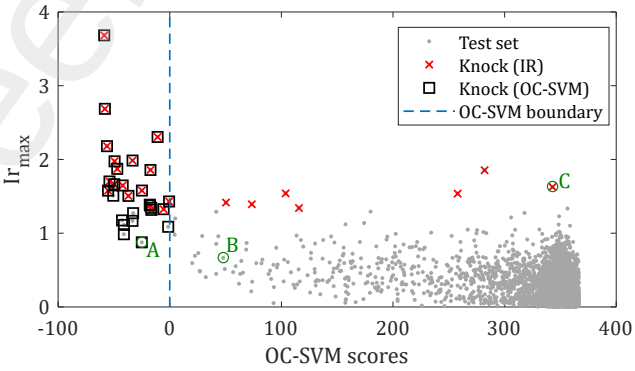

Figure 12: Maximum resonance index compared with OC-SVM scores.

with the IR method (red crosses). The dashed line is the model's decision boundary obtained from the training stage. For the selected mode combination, a condensed cloud of points exists around the graph's origin ($\alpha_5 = 0$ and $\alpha_6 = 0$) representing the majority class, i.e., the normal combustion cycles. In contrast, most of the cycles labelled as knock are separated. Most outliers are located in the fourth quadrant, corresponding to positive α_5 and negative α_6.

OC-SVM designate a score for each observation, indicating the signed distance to the decision boundary. This value represents the likelihood of belonging to a specific class. Negative scores indicate that the point is an outlier, as illustrated in Fig. 12. It can be seen that most of the misclassified points are low-knocking cycles according to the Ir index. The three points (A, B and C) marked in green are chosen for further analysis, Table 5 contains the information regarding the different indicators. Observing the "Label" category in this table, only point A is classified as knock. However, all three points present evidence of knock, as shown by the HRR and p_{bp} analysis in the upper plots from Fig. 13.

Point A is classified correctly by OC-SVM but not recognised by the IR method. Point B is recognised by neither method, yet it lies close to the decision boundary with a low positive score. Furthermore, the spectrogram of points A and B are similar, showing a high resonance near 12 kHz (see lower plots in Fig. 13). Then, although point B was not detected, the score can be a useful indicator to further evaluate the cycle's knocking condition through other characteristics. Finally, point C is not identified by the proposed method despite being classified as knock by the IR method. However, the spectrogram shows no particular frequency characteristic, indicating that the resonance mode of this cycle was not detected, which may

Table 5: Data for points A, B and C analysis

Point	α_5	α_6	OC-SVM score	IR [bar]	MAPO [bar]	MAKO	Label
A	134.8	-183.2	-25	0.87	0.35	6.8	Knock
B	136.1	-129.4	47.9	0.67	0.21	6.6	No knock
C	73	-33.12	343	1.62	0.49	4.9	No knock

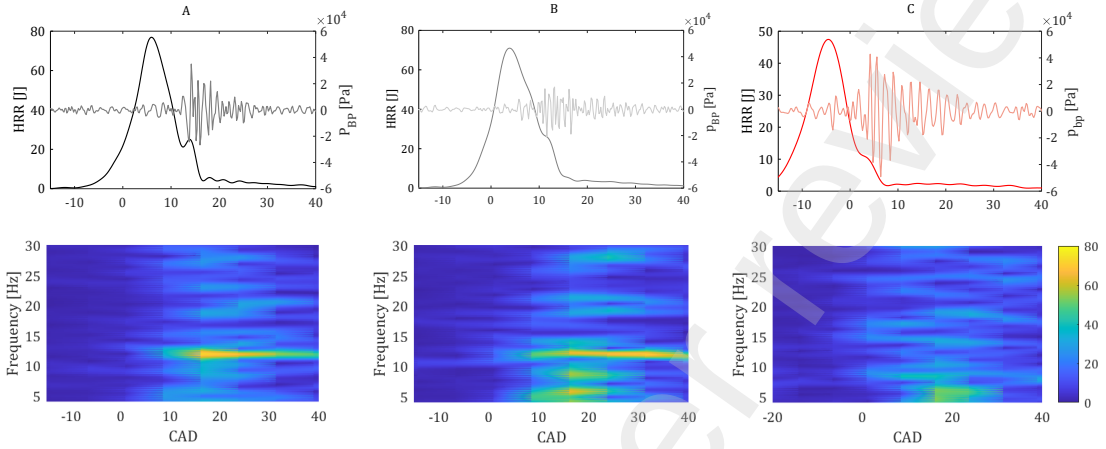

Figure 13: Knocking analysis of points A, B and C. **Upper:** HRR and P_{bp}. **Lower:** STFT.

be due to the position of the sensor being over a vibration node [23].

In Fig. 14 MAPO and MAKO techniques are compared. Regarding MAPO indicator, the proposed method is in good agreement for cycles with a high value of MAPO. However, for lower values, OC-SVM has a better performance. This can be explained by the fact that MAPO tends to underestimate knocking because its ability to detect knock relies on the predefined threshold. As for MAKO, it is to be expected that higher amplitude values coincide with detected knock events, mainly because both methods are based on the same signal. Even so, OC-SVM was able to classify correctly points whose oscillation was not over the threshold. If the threshold is lowered to increase knock detection, both methods, MAPO and MAKO, will incur in type-1 error,i.e. classifying normal cycles as knock.

The performance of the proposed model is analysed in Fig. 15 where a confusion matrix for each method tested is presented. Additionally, table 6 displays the metrics obtained for each case. It is worth recalling that the three

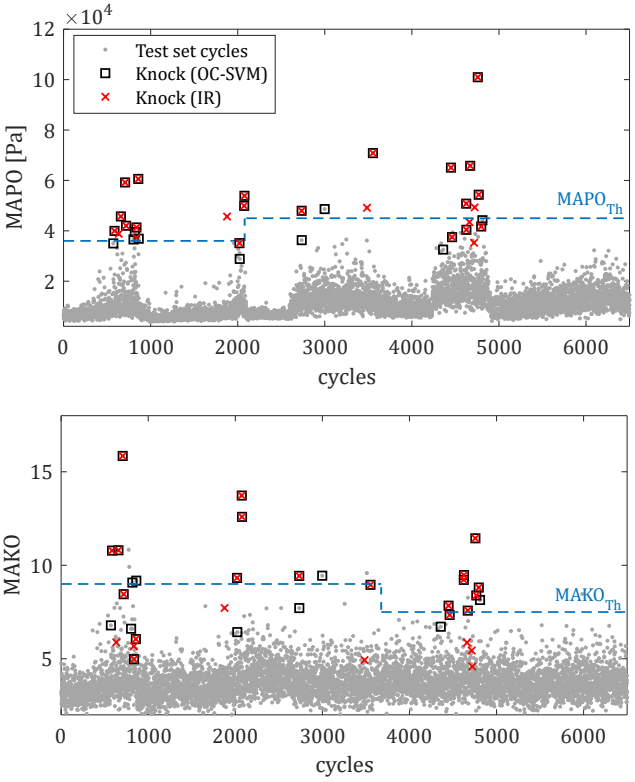

Figure 14: OC-SVM results compared to MAPO (upper plot) and MAKO (lower plot) indicators.

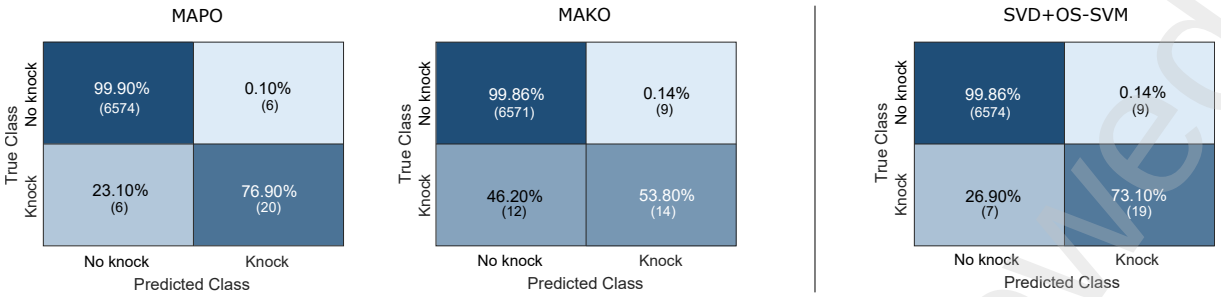

Figure 15: Confusion matrix comparison between MAPO, MAKO and the proposed method

techniques are compared to the IR method, whose results were taken as the ground truth. The new approach is able to detect 73% of knocking cycles accomplishing a performance similar to the one obtained with MAPO, which uses the in-cylinder pressure sensor. Compared to MAKO, the proposed method got better recall and precision results, as it could detect more knocking cycles. Consequently, the false negative cases were reduced by almost 40%.

Table 6: Performance metrics from the knock detection methods employed performance

Method	Recall	Precision	F1 score
OC-SVM	0.73	0.68	0.70
MAPO	0.77	0.77	0.77
MAKO	0.54	0.61	0.57

5. Conclusions

The present work proposes a new approach for knocking cycle detection based on the knock sensor measurements. The method considers the frequency and time evolution of the knock sensor signal through an STFT analysis to extract the main features observed in knocking cycles. These cycles are detected using an unsupervised classification technique (OC-SVM). The method was validated with experimental data at two engine speeds and different SA. Its performance was compared with other commonly used techniques. From the observed results, the following conclusions can be drawn:

- The classification model using as inputs the features extracted from the STFT and a previously established basis U demonstrated a 68% precision and a recall (true positive rate) of 73%. The model managed to detect cycles with both high and low resonance intensity.
- The proposed method has a detection performance similar to that achieved with MAPO. Because it does not require an in-cylinder pressure sensor, it could be tested for its application in onboard control techniques. In addition, it does not require a threshold to be defined at every operation point.
- Regarding MAKO indicator, which is also based on the knock sensor signal, the method improves knock recognition by identifying 20% more knocking events. Consequently, reducing type-2 error.

In general, the method has proven efficient for knocking detection. Although the feature selection process and the training stage have a high computational cost, this aspect is significantly reduced for the detection stage once the model is calibrated.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C. . Schleussner, G. Ganti, J. Rogelj, M. J. Gidden, An emission pathway classification reflecting the paris agreement climate objectives, Communications Earth and Environment 3 (2022).

[2] T. Burton, S. Powers, C. Burns, G. Conway, F. Leach, K. Senecal, A data-driven greenhouse gas emission rate analysis for vehicle comparisons, SAE International Journal of Electrified Vehicles 12 (2022).

[3] J. Gao, J. Huang, X. Li, G. Tian, X. Wang, C. Yang, C. Ma, Challenges of the uk government and industries regarding emission control after ice vehicle bans, Science of the Total Environment 835 (2022).

[4] Y. Wang, A. Biswas, R. Rodriguez, Z. Keshavarz-Motamed, A. Emadi, Hybrid electric vehicle specific engines: State-of-the-art review, Energy Reports 8 (2022) 832–851.

[5] S. Vedharaj, Advanced Ignition System to Extend the Lean Limit Operation of Spark-Ignited (SI) Engines—A Review, Energy, Environment, and Sustainability, 2021.

[6] F. Negüs, M. Grill, M. Bargende, Efficiency potential of si engines with gasoline and methanol: A 0d/1d investigation, in: SAE Technical Papers, 2021.

[7] J. Mason, B. Kaufmann, P. Tartarini, M. Puglia, N. Morselli, G. Veratti, A. Bigi, Compression ratios comparison between engines operating with producer gas, in: European Biomass Conference and Exhibition, ETA-Florence Renewable Energies, 2019, pp. 1927–1931.

[8] J. Rodríguez-Fernández, Á. Ramos, J. Barba, D. Cárdenas, J. Delgado, Improving fuel economy and engine performance through gasoline fuel octane rating, Energies 13 (2020) 3499.

[9] X. Zhen, Y. Wang, S. Xu, Y. Zhu, C. Tao, T. Xu, M. Song, The engine knock analysis - an overview, Applied Energy 92 (2012) 628–636.

[10] K. Uddeen, H. Shi, Q. Tang, G. Magnotti, J. Turner, The effects of compression ratio and combustion initiation location on knock emergence by using multiple pressure sensing devices, International Journal of Engine Research (2022) 14680874221075343.

[11] G. Brecq, J. Bellettre, M. Tazerout, Experimental determination of knock in gas SI engine, Technical Report, SAE Technical Paper, 2001.

[12] X. Zhen, Y. Wang, S. Xu, Y. Zhu, C. Tao, T. Xu, M. Song, The engine knock analysis–an overview, Applied Energy 92 (2012) 628–636.

[13] E. Ollivier, J. Bellettre, M. Tazerout, G. C. Roy, Detection of knock occurrence in a gas si engine from a heat transfer analysis, Energy Conversion and Management 47 (2006) 879–893.

[14] P. Bares, D. Selmanaj, C. Guardiola, C. Onder, A new knock event definition for knock detection and control optimization, Applied Thermal Engineering 131 (2018) 80–88.

[15] J. Kim, In-cylinder pressure based engine knock classification model for high-compression ratio, automotive spark-ignition engines using various signal decomposition methods, Energies 14 (2021) 3117.

[16] M. Rosas, G. Amador, Knock detection method for dual-fuel compression ignition engines based on block vibration analysis, SAE International Journal of Engines 14 (2021) 199.

[17] F. Millo, C. V. Ferraro, Knock in si engines: a comparison between different techniques for detection and control, SAE transactions (1998) 1091–1112.

[18] A. A. Lagana, L. L. Lima, J. F. Justo, B. A. Arruda, M. M. Santos, Identification of combustion and detonation in spark ignition engines using ion current signal, Fuel 227 (2018) 469–477.

[19] K. Prażnowski, J. Mamala, A. Bieniek, M. Graba, Identification and classification of selected internal combustion engine inefficiency based on vehicle structure vibrations, Vibrations in Physical Systems 31 (2020).

[20] P. Napolitano, I. Jimenez, B. Pla, C. Beatrice, Knock recognition based on vibration signal and wiebe function in a heavy-duty spark ignited engine fueled with methane, Fuel 315 (2022) 122957.

[21] D. Siano, M. A. Panza, A nonlinear black-box modeling method for knock detection in spark-ignition engines, in: AIP Conference Proceedings, volume 2191, AIP Publishing LLC, 2019, p. 020137.

[22] L. Petrucci, F. Ricci, F. Mariani, V. Cruccolini, M. Violi, Engine knock evaluation using a machine learning approach, Technical Report, 2020.

[23] D. Scholl, C. Davis, S. Russ, T. Barash, The volume acoustic modes of spark-ignited internal combustion chambers, SAE transactions (1998) 1379–1386.

[24] J. D. Naber, J. R. Blough, D. Frankowski, M. Goble, J. E. Szpytman, Analysis of combustion knock metrics in spark-ignition engines, SAE Transactions (2006) 223–243.

[25] G. H. Golub, C. F. Van Loan, Matrix computations. johns hopkins studies in the mathematical sciences, 1996.

[26] C. Eckart, G. Young, The approximation of one matrix by another of lower rank, Psychometrika 1 (1936) 211–218. doi:10.1007/BF02288367.

[27] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, R. C. Williamson, Estimating the support of a high-dimensional distribution, Neural computation 13 (2001) 1443–1471.

[28] Y. Xiao, H. Wang, L. Zhang, W. Xu, Two methods of selecting gaussian kernel parameters for one-class svm and their application to fault detection, Knowledge-Based Systems 59 (2014)

75–84.

[29] B. Pla, J. De La Morena, P. Bares, I. Jiménez, Knock analysis in the crank angle domain for low-knocking cycles detection, Technical Report, SAE Technical Paper, 2020.

[30] C. Guardiola, B. Pla, P. Bares, A. Barbier, An analysis of the in-cylinder pressure resonance excitation in internal combustion engines, Applied Energy 228 (2018) 1272–1279.

[31] Y. Ma, H. He, Imbalanced learning: foundations, algorithms, and applications (2013).

[32] A. Luque, A. Carrasco, A. Martín, A. de Las Heras, The impact of class imbalance in classification performance metrics based on the binary confusion matrix, Pattern Recognition 91 (2019) 216–231.