Electrical and mechanical data fusion for hydraulic valve leakage diagnosis

OPEN ACCESS
IOP Publishing
Measurement Science and Technology
Meas. Sci. Technol. **34** (2023) 044011 (11pp)
https://doi.org/10.1088/1361-6501/acb376

Electrical and mechanical data fusion for hydraulic valve leakage diagnosis

Fabio Conti[1,*], **Federica Madeo**[2], **Antonio Boiano**[2] and **Marco Tarabini**[1]

[1] Department of Mechanical Engineering, Politecnico di Milano, Milano, Italy
[2] School of Industrial and Information Engineering, Politecnico di Milano, Milano, Italy

E-mail: fabio1.conti@polimi.it

Received 29 September 2022, revised 31 December 2022
Accepted for publication 16 January 2023
Published 27 January 2023

Abstract

This paper describes a method for the identification of valves' failure, with the final aim of creating a predictive maintenance architecture. After revising the scientific literature, we selected the electric current, the acoustic emission and the vibration signals as the most promising monitoring techniques. The processes of feature extraction and data fusion have been optimized to detect early symptoms of a failure. Performances of five different machine learning algorithms have been compared. Results, obtained in a specific case study, evidenced that a data fusion process based on vibration and current data, paired with a random forest model allowed a prediction accuracy and a Jaccard index close to 99%.

Keywords: data fusion, vibration measurement, current measurement, predictive maintenance, industry 4.0, fault detection

(Some figures may appear in colour only in the online journal)

1. Introduction

Industrial machineries' faults limit the production efficiency, decrease the goods' quality and may lead to stop the production [1]. It is therefore important to predict incipient faults by identifying physical phenomena that are related to the behavior of the machinery under inspection [2]. These phenomena are usually complex and the earlier the effect of the fault, the harder is the identification. According to the Industry 4.0 paradigms, all the data that can be extracted from the machinery must be used to assess the condition status of the monitored components [3–5]. In this context, data coming from different sources can be used to improve the detection of fault and schedule maintenance accordingly [6].

* Author to whom any correspondence should be addressed.

1.1. Valves monitoring overview

Control valves are important components in all processes involving fluids (i.e. pure gas, mixture of gases, water, oil). Valve failure represents a serious issue in many industrial fields: nuclear power plants, transport of corrosive fluid, controlling of water, oil and gas pipelines [7]. The industrial problem that originated our work is related to aluminum extrusion, in which the desired ram speed is controlled by the pumping of oil in the extrusion cylinder; the quality of the product depends on the combination between the ram speed and the aluminum temperature [8]. Regardless of the task, valves are opened and closed frequently, resulting in different kinds of wear and therefore in oil leakage [9].

Hydraulic blocks used in the die-cast aluminum process are usually equipped with 4/3-way flow directional control valves and are usually connected to hydraulic circuits for primary and auxiliary actuations. A simplified scheme of the hydraulic circuit is shown in figure 1.

A review of the existing literature evidenced that the most common valve failures are related to the leakage generated by the spool wear and to the solenoid burnout [10–13].

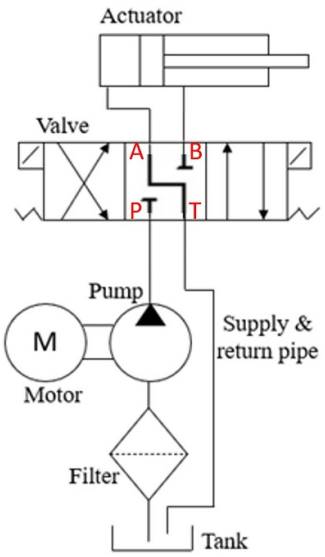

Figure 1. Simplified hydraulic circuit for the main cylinder actuation, a 4/3-way flow directional control valve is used to change the piston position.

1.2. Condition monitoring techniques

Different methods have been proposed to monitor the valves' condition; the main ones are acoustic emission (AE), vibration monitoring, thermography, current analysis, corrosion monitoring, shock pulse method and gas detection [14–18]. Among these, AE, vibration, thermography and current analysis are the ones most commonly used in industrial environments.

AE signals (with bandwidths often exceeding 50 kHz) acquired in presence and absence of leakage show that the amplitude of the dominant leak-related frequency components increase with the leakage rate [19]. Seong and colleagues investigated the correlation of AE features with the occurrence of leakage, the failure entity and the failure mode [20]. Two failure modes were investigated: the disk wear and the obstruction due to a foreign object for different failure sizes (mm) at different inlet pressures. Experiments showed that the peak frequency patterns in the healthy valve cases were random and variable with low amplitude while specific peaks were present in case of failure. Moreover, the characteristic frequencies were independent of the size of the failure, but they were strongly dependent on the types of the failure modes, and that the PSD amplitudes and the AE root mean square (RMS) had a strong relationship with the size of the failure.

AE signals are good indicators of valve wear condition [21] and their variation allow identifying different leakage conditions. However, the equipment needed for the identification of the fault has seldom been used as continuous monitoring system in industrial plants, where hundreds of valves require monitoring.

Vibration analysis is usually more applicable to a larger number of components thanks to the lower sensor cost, albeit the limited bandwidth with respect to AE. Ji *et al* [15] proposed to identify the presence of six different valve wear conditions using six accelerometers (three on the valve spool and three on the valve internal body). The signals were acquired with a sampling frequency of 6 kHz and the analyzed features were the spectral peaks at frequencies of 248.9 Hz, 494.8 Hz, 743.6 Hz and 992.5 Hz, that corresponds to the fluid pulsation frequency and its 2nd, 3rd and 4th harmonics. Tests were performed in laboratory condition but are not hardly applicable in an industrial environment with multiple valves. Vibration analysis was also used by Thompson and colleagues [14] to detect gas leakage. The vibration was analyzed in the 0–20 kHz frequency band; also in this case, the spectral amplitudes were used to detect the presence of leakage. As a general rule, vibration signals can be used to predict the leakage, although spurious components produced by other industrial facilities could interfere with the signal generated by the fault.

Thermography has been used to monitor machinery and structures by assessing the thermal state of the plant and of the hydraulic fluid [22]. Infra-red cameras were used to locate the hot spots in a fluid power system [23]. When leakage occurs, an increase of temperature was detected since a high-pressure fluid was flowing through a small orifice. Takahashi applied this method for internal leakage detection [16]. This heat generation could be detected by measuring different critical points on the solenoid-operated directional control valve. A rarer case is the solenoid burnout fault, which can be detected by monitoring the solenoid temperature. When a solenoid burnout happens, the temperature increases in short time periods [24]. The temperature increase is usually detectable when the leakage reaches critical conditions for the plant and cannot be used to identify early symptoms of the fault, which can be mistakenly associated to daily valve temperature variations.

The electrical current is also widely used for the diagnosis of electrical motors and solenoid-operated valves, where the current absorbed by the solenoid allows detecting the armature's movement. Kulkarni *et al* [25] showed that leakage affects current absorption curve of the solenoid, since unwanted flow may either decrease the differential pressure between front and back opening of the valve. Moreover, if the solenoid is de-energized and the leakage occurs, the valve potentially opens. The solenoid excitation current has a prominent dip during power-up due to the back electromotive force (EMF) generated by plunger movement [17]. Tansel and colleagues [26] monitored the current waveform through a Hall sensor, obtaining a reliable indication of the damage. Ngbede and colleagues [27] proposed a Deep Learning approach for classifying faulty solenoid operated valves. As in the temperature case, it is possible to detect solenoid burnouts by monitoring the valve's current. In a very similar way as for the temperature case, a fast current peak bigger than the nominal current occurs when solenoid burnouts [24].

The existing literature studies evidence that AE, vibration, thermography and current signals can be successfully used to detect the main valves' faults; the works are mainly based on laboratory experiments and use single signals for the discrimination of healthy and faulty equipment conditions. In order to simultaneously use different signals and features to increase the robustness of the classification problem, we decided to

explore the effectiveness of machine learning (ML) techniques for the feature selection and weighting.

1.3. ML for predictive maintenance

Leak detection can be considered as a two-class problem, because the objective is to determine whether the valve is healthy or not [28]. When identifying the leakage level [29] and/or the failure mode [24], the problem becomes a multi-class one: the class zero is the nominal situation, in which leakage has not occurred, while the other classes correspond to different leakage rates and/or failure modes. In order to reduce the task complexity, the problem can also be decomposed in different target layers through a hierarchical ML model: i.e. the failure condition is first identified and then the failure size is recognized [23]. In this context the binary classification is considered and Hyperparameters are chosen accordingly to the task and through multiple runs of optimization [30, 31]. Balanced dataset is another aspect that must be considered, since allows the creation of better prediction models [32]. As example, random forest (RF) aims to minimize the overall error rate, this tends to lower the attention to the data related to the minority cases [33]. Despite there exist some techniques to solve and correct data imbalances [32, 33], the possibility of harvesting equally size of data for different cases must be exploited whenever it is available, as it is in our laboratory conditions.

1.4. Aim of the work

Our study aims at the definition of the possible methods to combine features coming from a set of sensors chosen on the basis of physical principles for the detection of valves' leakage. The paper is structured as follows: section 2 presents the background of the proposed method, that is described in section 3. A case study is presented in section 4 and experimental results are discussed in section 5. Paper conclusions are drawn in Section 6.

2. Background

Due to the different characteristics of hydraulic blocks and valves manufactured by different producers, a data driven approach is the favored one with respect to an analytical one. The analytical identification of a model for the hydraulic block-valve system, indeed, cannot be representative of all the possible electro-actuated systems in an industrial environment. The identification of an analytical relationship between the behavior of the AE and the leakage, on the other hand, can be useful to identify the most generic features to be used in the ML process. The model describing the relationship between the emitted acoustic power and the flow turbulence is based on Lighthill's theory and it is valid in presence of turbulent flow, which typically occurs for Reynolds numbers greater than 1000 [34]. Reynolds number (Re) is defined in equation (1):

$$Re = \frac{\rho v L}{\mu} \tag{1}$$

where ρ is the fluid density, v is the average velocity of fluid through the leak, L is the characteristic length and μ is the fluid dynamic viscosity. According to Lighthill's theory, the sound emission is produced by the net density fluctuation and the net momentum transferred to the turbulence flow region and it must be the result of quadrupole and high order sources. The sound power from turbulence flow is expressed in the Lighthill's equation [35], which follows:

$$P_s = \frac{\rho v^8 D^2}{\alpha^5} \tag{2}$$

where P_s is the sound power, D is the valve size and α is the sound velocity in the fluid. Since $P_s \propto v^8$, the sound power depends on leakage velocity and an increase in leakage rate causes a clear change in AE signal amplitude. The sound power P_s had been derived in terms of the fluid parameters in [36] and it can be expressed as:

$$P_s = C_0 \frac{P_1^4 d^{16}}{\alpha^5 \rho^3 D^{14}} \tag{3}$$

where P_1 is the inlet pressure, d is the leakage orifice diameter and C_0 is the proportionality constant. According to Parseval's theorem, the energy measured in time and frequency domain are equivalent [37]. The sound power of liquid valve leakage P_s is directly proportional to the average signal power AE_{rms}^2:

$$AE_{\text{rms}}^2 = \beta P_s \tag{4}$$

where β is a ratio coefficient whose value is lower than 1, as the signal detected by the AE sensor is a part of the energy of the AE signal. Consequently:

$$AE_{\text{rms}}^2 = C_{0q} \frac{P_1^4 d^{16}}{\alpha^5 \rho^3 D^{14}}. \tag{5}$$

Kaewwaewnoi *et al* [36] used a quadrupole relationship to describe leakage, but a more complete model was explored by El-Shorbagy [38]. A more complex model was proposed to model the control valve noise, according to Curle's theory about the influence of solid boundaries [39]. The addition of a dipole source turns equation (5) into:

$$AE_{\text{rms}}^2 = \frac{(C_{0q} P_1 + C_{0d}) P_1^4 d^{16}}{\alpha^5 \rho^3 D^{14}}. \tag{6}$$

The equations show the dependence between the leakage hole dimensions and the emitted sound power P_s (and consequently with AE_{rms}^2). The substitution of d into equations (5) or (6) allows to identify the direct relationship between leakage and AE and the dominant quantities. Referring to a four-way directional control valve, the orifice flow relationship can be expressed as follows [40]:

$$Q = K_v X \sqrt{\Delta P} \tag{7}$$

where K_v is an overall constant coefficient for a particular valve orifice referred to the spool land circumference, X is the axial opening caused by the displacement of the spool in the valve body from its initial position and ΔP is the pressure drop across the valve. Therefore, equation (7) can be regarded as a resistance equation, in which the control valve orifice acts as a variable non-linear resistor:

$$Q = \frac{1}{R}\Delta P \tag{8}$$

$$R = \frac{1}{K_v X \Delta P^{-1/2}}. \tag{9}$$

The equivalent resistance R varies with valve opening X: the larger the X, the lower the R factor becomes and the larger the flowrate that passes through the orifice, and vice versa. Consequently, the internal leakage of a 4/3-way directional control valve, figure 1, is computed using the following equations [40]:

$$Q_{\text{leak}P-A} = G_{\text{leak}} \mathrm{d}P_{P-A} \tag{10}$$

$$Q_{\text{leak}B-T} = G_{\text{leak}} \mathrm{d}P_{B-T} \tag{11}$$

$$Q_{\text{leak}P-B} = G_{\text{leak}} \mathrm{d}P_{P-B} \tag{12}$$

$$Q_{\text{leak}B-A} = G_{\text{leak}} \mathrm{d}P_{B-A}. \tag{13}$$

Q is the leakage flowrate, G_{leak} is the equivalent conductance of leakage flow and $\mathrm{d}P$ is the pressure differential between two valve ports. In case of liquid valve leakage, the volume flow rate of liquid through a valve was given by [41]:

$$Q = 29.81 c_v d^2 \sqrt{\frac{\Delta P}{S}} \tag{14}$$

where Q is the volume flow rate, c_v is the valve flow coefficient, ΔP is the pressure drop across the valve and S is the specific gravity. The equation (14) may be rearranged as follows:

$$d = \left(\frac{Q}{29.81 c_v}\right)^{1/2} \left(\frac{S}{\Delta P}\right)^{1/4} \tag{15}$$

Hence, equations (5) and (6) become:

$$AE_{\text{rms}}^2 = C_1 \frac{1}{\alpha^5 \rho^3 D^{14}} \left(\frac{Q}{c_v}\right)^8 \left(\frac{P_1 S}{\Delta P}\right)^4 \tag{16}$$

$$AE_{\text{rms}}^2 = \frac{(C_1 P_1 + C_2) P_1^3}{\alpha^5 \rho^3 D^{14}} \left(\frac{Q}{c_v}\right)^8 \left(\frac{S}{\Delta P}\right)^4. \tag{17}$$

It follows that the AE_{rms} increases with inlet pressure, it is inversely proportional to the size of the valve D and to the flow coefficient c_v. Moreover, the volume flow rate is the dominating factor. Based on the previous information, the AE_{rms} is the first feature candidate. Moreover, the AE signal of valve leakage is stationary random and its mean is close to zero [9]; therefore, the correlation between its RMS and variance is defined as:

$$AE_{\text{rms}}^2 = N\sigma^2. \tag{18}$$

If the AE variance σ^2 is substituted in equations (16) or (17), it turns out that:

$$\frac{\sigma^2}{Q^8} = C_1 \frac{1}{N\alpha^5 \rho^3 D^{14} c_v^8} \left(\frac{P_1 S}{\Delta P}\right)^4 = k \tag{19}$$

where k is constant when the valve and fluid parameters are determined. Therefore, when the valve leakage is small, there is a linear relationship between the standard deviation of the AE signal σ and the leakage rate Q:

$$\sigma = k_1 Q + k_2 \tag{20}$$

where k_1 is related to fluid and valve properties and k_2 denotes various environmental noise in the experiments.

3. Method

The proposed method is based on the fusion of different physical signals coming from the valve to detect incipient faults or variations of the working condition. The method is based on the extraction of features from different signals and on the classification of the valve state using different algorithms.

3.1. Features

The equations presented in the previous section show that the severity of the leakage phenomenon affects the energy of the acoustic signal transferred to the fluid. The fluid energy, in turn, is transmitted to the valve frame and generates different phenomena.

The most evident is the body valve vibration, whose frequency depends on the valve characteristics and on the leakage entity. The possible features that can be extracted from the signal are related to its energy (RMS, Peak) and to frequency domain features (frequencies with dominant spectral components, spectral centroid and spectral Kurtosis). These quantities were selected from the literature studies that are more similar to our application: the magnitude of the dominant frequencies components and the frequency centroid were chosen in [20]. Kurtosis evidenced increasing levels of vibration in [42]. While the RMS usefulness is demonstrated by the formulas of section 2. Frequency-band related RMS is investigated since it is bound to specific harmonic components, while RMS on the whole frequency band could be generated by machineries or vibration sources.

The second group of features is related to the absorbed current, for the detection of the solenoid burnout. In the literature there are no specific studies related to the current signal features allowing to detect the main hydraulic valves' defects. Consequently, features were derived from the analysis of the current signal time history as described in the next section.

3.2. Algorithms

Current and vibration features can be analyzed with RF, support vector machine (SVM) and k-nearest neighbors (k-NN) algorithms. These approaches allow merging current and vibration information at feature level to detect possible defects of hydraulic blocks. Each of the three algorithms' classes have specific tuning parameters, that have to be carefully selected in order to obtain the best classification performances.

The first algorithm to be trained is the RF one, exploiting its features importance parameter. The features importance parameter is computed as the mean of accumulation of the impurity function within each tree of the RF model [43]. Implying that these features are keener to be chosen with respect to others. It can be used to reduce the number of features considered during the training phase of the other algorithms, since it identifies the features with higher correlations to the class variation, reducing the computational load of the algorithms and eventually the feature extraction phase.

The tuning parameters of RF algorithms are the number of trees and the random state parameter. The former parameter can be used to tune the voting classifier population, while the latter controls the randomness of the feature selection used during the training phase. In RF, a sub-optimal greedy algorithm is repeated several times while training using random selections of features and samples. Therefore, the number of trees was selected as the limit above which the benefits in classification performances are negligible.

SVM has three main tuning parameters, that are the Kernel type, the Regularization Factor (C factor) and the Gamma parameter. Different types of kernels exist, such as linear, sigmoid, radial basis function (RBS). RBS was chosen for his high accuracy and training velocity, which could also fit well in low computational power devices. However, linear and sigmoid kernels were also included in our study. The C parameter represents the penalty parameter for the misclassification errors. The gamma parameter defines how far the influence of a single training sample affects the calculation of a plausible line of separation.

k-NN estimates the likelihood that a sample is a member of a class (for instance, faulty valve) rather than another (healthy) based on what group of the training data points that sample is nearest to. The choice is made considering a certain number k of training samples close to the sample that must be evaluated. This algorithm tries to generalize the output if a high value of training samples is used. Conversely, considering a small neighborhood of data (small k) could lead to overfitting (i.e. the model has a good behavior during the training, but has poor performance in real conditions).

Performances of ML models are compared on the basis of three different metrics: the computational time, the confusion matrix and the Jaccard index.

(a) Computational time is evaluated both during the learning phase and during the prediction of the valve state.
(b) The confusion matrix is used to show the model's ability to correctly predict or separate the classes. The confusion matrix provides information about the correctly classified cases (faulty when faulty, healthy when healthy) and about the errors (faulty when healthy or healthy when faulty). Diagonal elements of the confusion matrix indicate the correct predictions, while off-diagonal elements are the ones mislabeled by the classifier. The higher the number of elements on the diagonal, the better, stating more accurate predictions.
(c) The Jaccard index is a measure of similarity between two sample sets. It is computed as the ratio between the size of the intersection and the size of the union of the two labeled sets. The index ranges from 0% to 100% and the higher the percentage, the more similar the two populations.

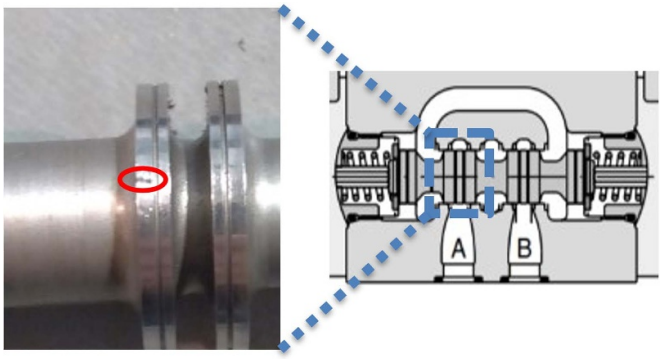

Figure 2. Pictorial view of the simulated defect of the spool.

3.3. Case study

The case-study for the validation of the proposed data fusion approach is a hydraulic manifold used to control an aluminum extrusion press. Two 4/3-way flow directional control valves model D1VW by Parker (Ohio, USA) were used in the experiment. One to be considered healthy, while the other one was damaged through an incision on the spool's valve. The leak diameter produced was 0.1 mm hole, figure 2, representative of a condition that is surely faulty in real operations.

3.4. Experiments

Experiments were performed with different circuit operating pressures at the same flow rate. Operating pressures were 5, 12 and 20 MPa. The flow rate was 10 l min^{-1}. This small flow rate aims to simulate the resistance imposed by the presence of a working hydraulic press at the valve output. All the six combinations of operating pressure and flow rate were tested. A total number of 5850 working cycles (14 s each) were acquired, 975 working cycles for each condition to obtain an equally distributed dataset. For each working cycle, case 'a' is the actuation of the spool valve in the damaged direction, while case 'b' is related to the actuation in the opposite direction.

Tests were performed in different days to limit the effect of uncontrolled external disturbances (spurious vibration, temperature increase). Faulty and healthy states were

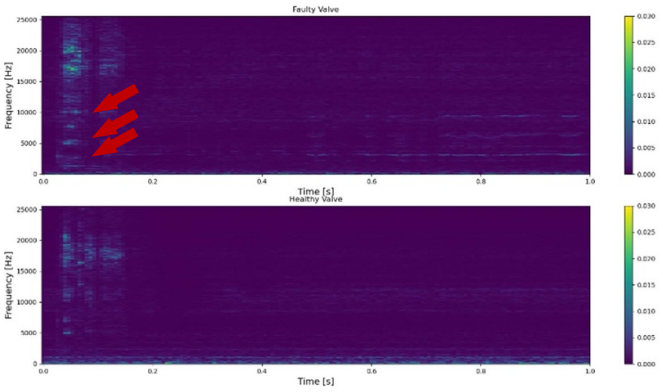

Figure 3. Short-time Fourier Transform of the healthy and faulty valves (10 l min^{-1}, 5 MPa). Specific harmonic components that are present in the Faulty case are highlighted with red arrows.

acquired on the same days to reproduce the same experimental conditions.

The valve vibration was measured by a single-axis accelerometer manufactured by Brüel & Kjær (Nærum, Denmark), model DeltaTron Type 4397. The vibration pick-up was positioned on the valve body with sensing axes aligned with the spool axis.

In order to select the features to be extracted from the vibration signal, a time-frequency analysis was performed (figure 3), with the aim of identifying the frequency components that, at specific times, allow distinguishing between the healthy and faulty states.

In the proposed case-study, the larger differences between the two states occur 0.02 s after the spool switch. Signals' spectra (figure 4) were analyzed in order to identify the frequency bands presenting larger differences between the healthy and faulty states (0–10 kHz and 20–25 kHz). The frequency band 15–20 kHz, although characterized by large spectral components, has not been considered, since the variation between the healthy and faulty state was limited. The vibration features included in our analyses were therefore the Kurtosis and the spectral peaks around the frequencies of 2.6, 5.2, 9.2 and 22.5 kHz. The spectral peak is related with the signal RMS at a specific frequency, and was chosen instead of the narrow-band RMS for the easiness of implementation in the final (real-time) application.

The currents absorbed by the two solenoids were measured by high-side, unipolar, current shunt monitors manufactured by Texas Instrument (Dallas, Texas) model INA169. Current signals are shown in figure 5. $I1$ and $I2$ are the currents feeding the solenoids (being $I1$ the one of the solenoids closer to the damage and recorded on actuation 'a').

Current features were derived from the time domain, at a time close to the spool switch inside the valve body, as indicated in [21]. The push pin in fault condition stops for 0.006 s at the middle of its run, and a local minimum point is observed. Figure 5 shows that, at 0.032 s, the pushpin suddenly stops, to resume its execution at 0.038 s. This is observed by the current sensor; the current suddenly increases for 0.006 s because

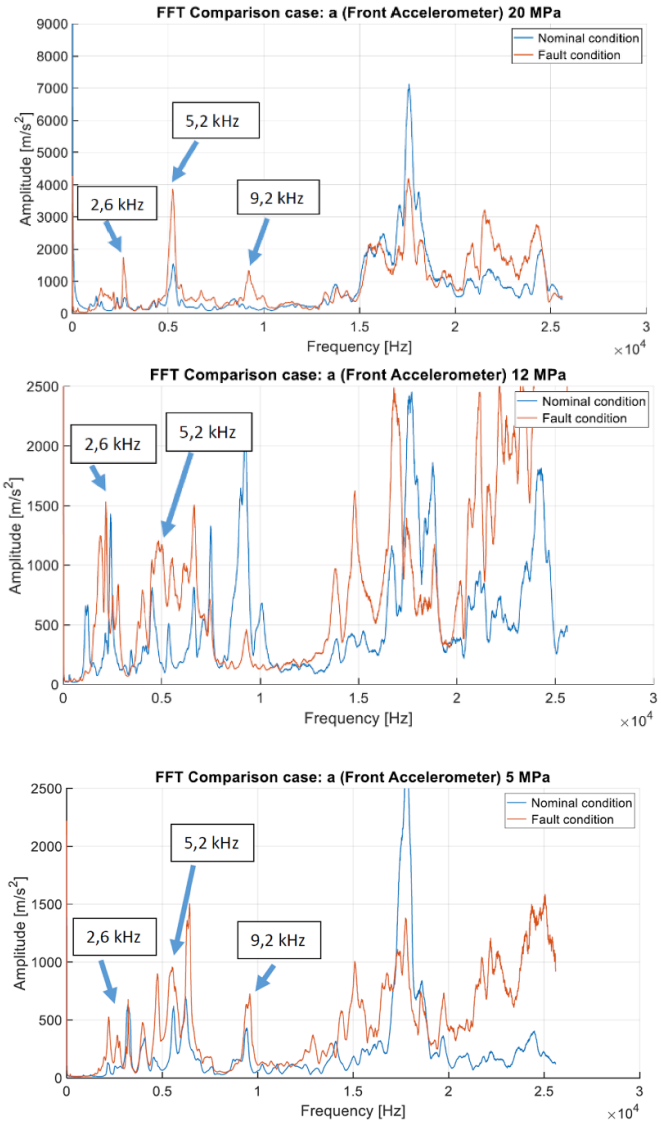

Figure 4. FFT of the vibration signal at 20, 12 and 5 MPa. The selected harmonic components are highlighted with arrows. The choice of the 22.5 kHz harmonic is representative of the whole 20−25 kHz frequency band.

no back EMF is generated, to then decrease again until the pushpin reaches the end of the stroke. As a feature selection analysis, the focus was mainly related to extracting the pushpin behavior just described. Therefore, seven current features were selected:

(a) the current amplitudes for $I1$ and $I2$ (in the 0.032 s–0.038 s time interval) when the current derivative is null (2 features).
(b) the maxima and minima values for both the currents $I1$ and $I2$ (4 features).
(c) the difference between the two currents in the highlighted interval of time (1 feature).

Overall, 17 features were selected, 7 related to current and 10 related to acceleration.

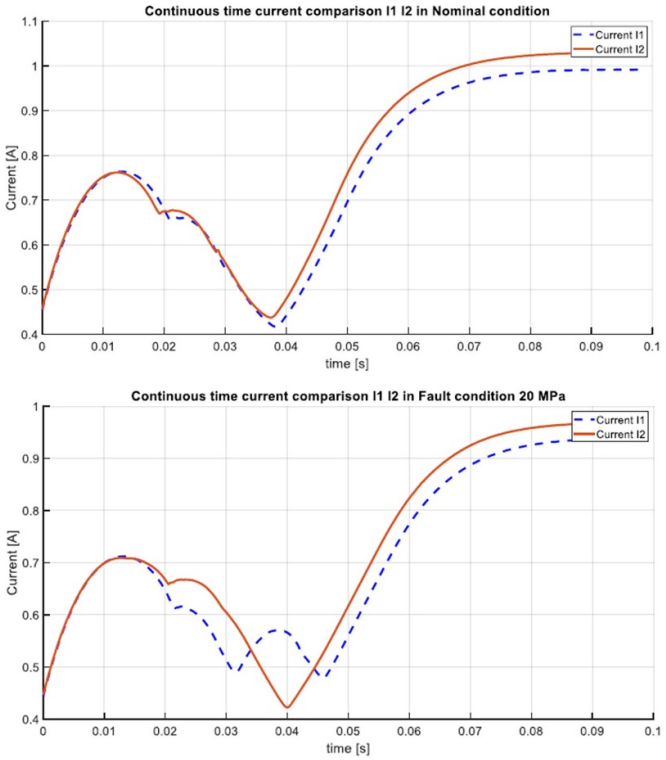

Figure 5. Time history of the currents feeding the solenoids in nominal condition (upper plot) and in faulty condition (lower plot).

4. Results

The relative importance of the 17 different features (figure 6) is one of the results of the RF model. The best features for the classification are the RMS of the vibration in the 22.5 kHz band and the difference of currents of the two solenoids.

In order to limit the total number of features, we selected the ones above the threshold of 0.05, i.e. the difference of the currents absorbed by the two solenoids, the current derivative for $I1$, vibrations RMS at 2.6 kHz, 5.2 kHz, 9.2 kHz, 22.5 kHz.

ML algorithms described in section 3 were trained to create predictive models. Data shown in the followings are related to ML models trained with 75% of the dataset (4524 samples) and tested on the remaining 1326. Each condition is therefore represented by 754 samples for the training step and 221 samples for the prediction step. An optimizer was executed to test out different C and gamma parameters and identify the combinations which perform better for SVM algorithms. At the end of its computation, the results which maximize the SVM performances are $C = 10$ and $gamma = 0.00001$. RF is characterized by 16 trees with a maximum depth of 2. In this case Gini index is selected as split criterium. k-NN is characterized by $k = 7$, obtained with a trial-and-error optimization.

The chosen algorithms present the computational complexity summarized in table 1. n indicates the number of the training sample, p the number of features, d depth of tree, n_{trees} the number of trees, $n_{\text{sv}} = 603$ the total number of supports vectors through the six cases while k is the number of k-nearest samples to be used by k-NN algorithm. k-NN Training Time is comparatively lower than SVM and RF algorithms, given $n = 4521$, $p = 7$, $n_{\text{trees}} = 16$ and $k = 7$. Prediction Time varies according to the parameters used to tune the algorithms, in this case RF has lower computational time than SVM and k-NN.

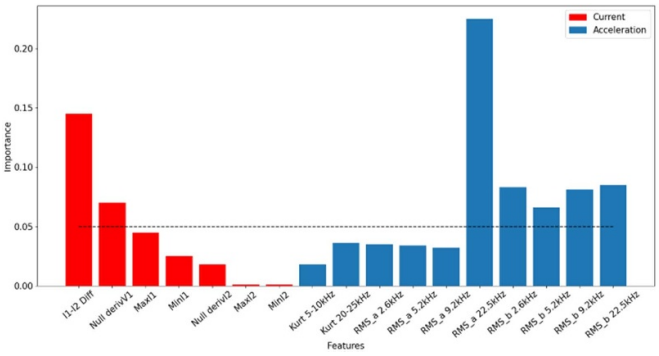

Figure 6. Features importance ranking for random forest algorithm. Features importance is computed as the mean of accumulation of the impurity decrease within each tree.

Table 1. Model Time-complexity for selected algorithms.

Algorithm	Training	Prediction
Random forest	$\mathcal{O}(n\log(n)pn_{\text{trees}}) \sim 1\times10^{6}$	$\mathcal{O}(n_{\text{trees}}d) \sim 1\times10^{0}$
SVM (Kernels)	$\mathcal{O}(n^2) \sim 1\times10^{7}$	$\mathcal{O}(n_{\text{sv}}p) \sim 1\times10^{3}$
k-NN	$\mathcal{O}(np) \sim 1\times10^{5}$	$\mathcal{O}(npk) \sim 1\times10^{5}$

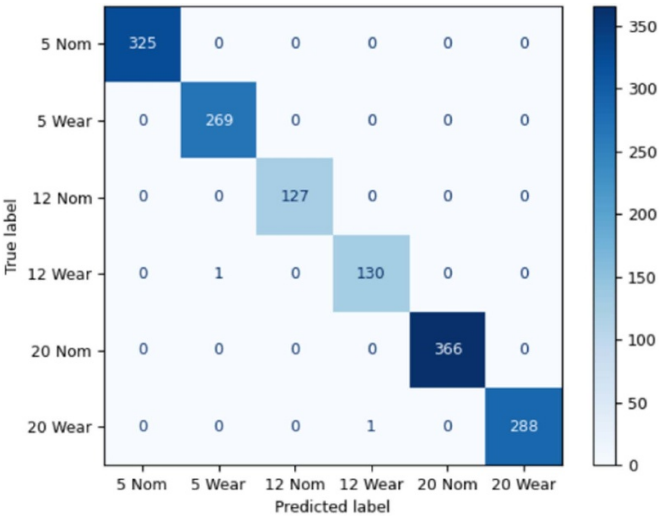

Figure 7. Confusion matrix plot of the random forest algorithm trained with current and acceleration features with feature importance greater than 0.05.

Performances of the RF model are summarized in the confusion matrix of figure 7. Only two samples out of 1507 are misclassified during the testing phase. Therefore, the average classification report shows a Prediction Accuracy of more than 99.8% and a Jaccard index greater than 99%.

The confusion matrix of the SVM-rbf model is shown in figure 8. The algorithm generates more false positives than false negatives with a Prediction Accuracy of 95% and a Jaccard index of 91%.

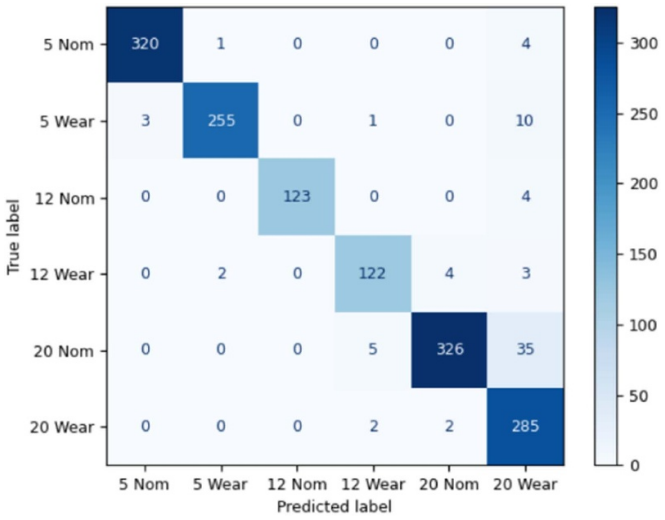

Figure 8. Confusion matrix plot of the SVM-rbf algorithm, trained with current and acceleration features with feature importance greater than 0.05.

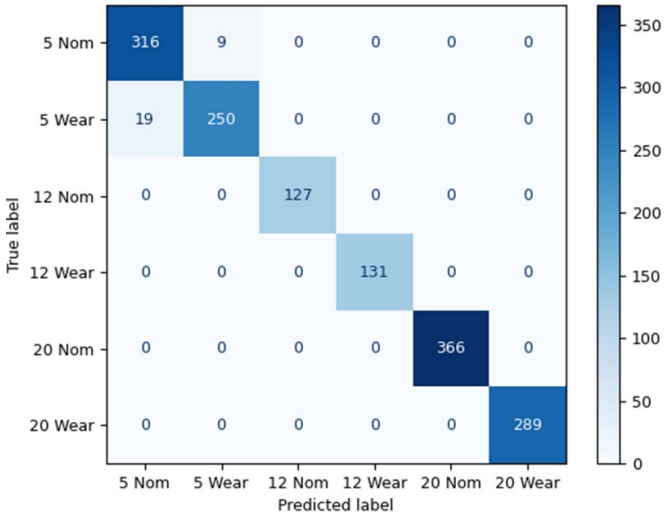

Figure 10. Confusion matrix plot of the SVM-linear algorithm, trained with current and acceleration features with feature importance greater than 0.05.

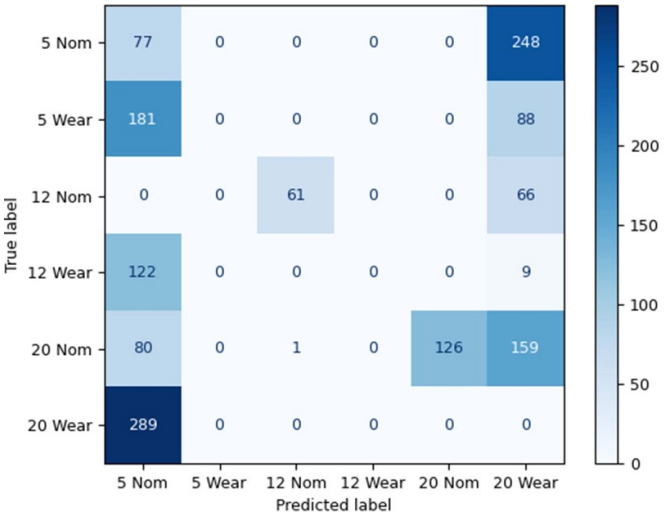

Figure 9. Confusion matrix plot of the SVM-sigmoid algorithm, trained with current and acceleration features with feature importance greater than 0.05.

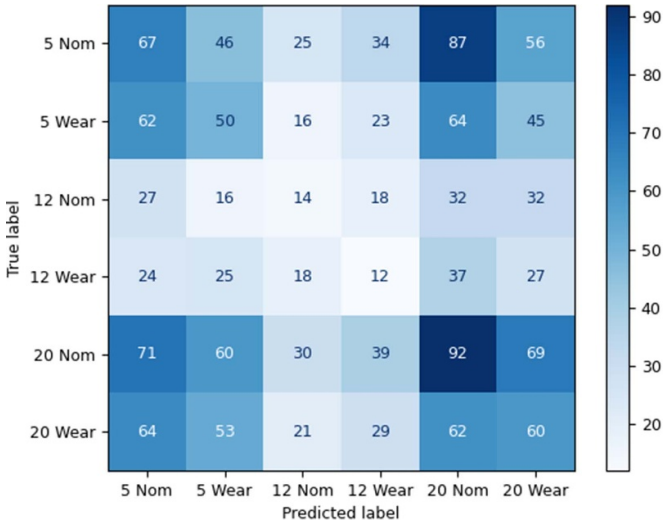

Figure 11. Confusion matrix plot of the k-NN algorithm, trained with current and acceleration features with feature importance greater than 0.05.

The confusion matrix of the SVM-sigmoid model is shown in figure 9. The algorithm is not able to correctly classify the classes and the Jaccard index is 14%.

The confusion matrix of the SVM-linear model is shown in figure 10. The algorithm generates more false negatives than false positives with a prediction accuracy of 98% and a Jaccard index of 96%.

The confusion matrix of the k-NN model is shown in figure 11. The algorithm has a prediction accuracy of 18% and a Jaccard index of 10%.

The time required to train the models was comparable among all the five proposed algorithms (less than 1 s).

The effect of the dataset size on the method performances has been investigated by comparing the accuracy as in [15], RF and SVM-linear were chosen given the achieved performances. RF Accuracy was larger than SVM accuracy for any dataset size, as shown in figure 12.

As can be seen in figure 13 it was trained a RF algorithm with only the accelerometer features, in this case the prediction accuracy is 98% and the Jaccard index 96%, with a total of 30 misclassified samples out of 1507.

5. Discussion

5.1. Limitations

The main limitation of this study is the reduced set of configurations in which the method has been tested. Our study would surely benefit from both additional tests with different damages of the spool and from testing of the proposed algorithms

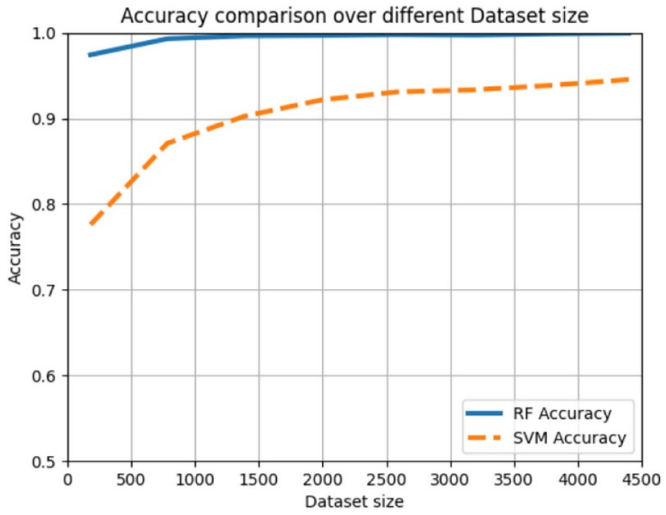

Figure 12. Accuracy comparison of the RF and SVM model trained with different dataset size. Models trained with acceleration and current features.

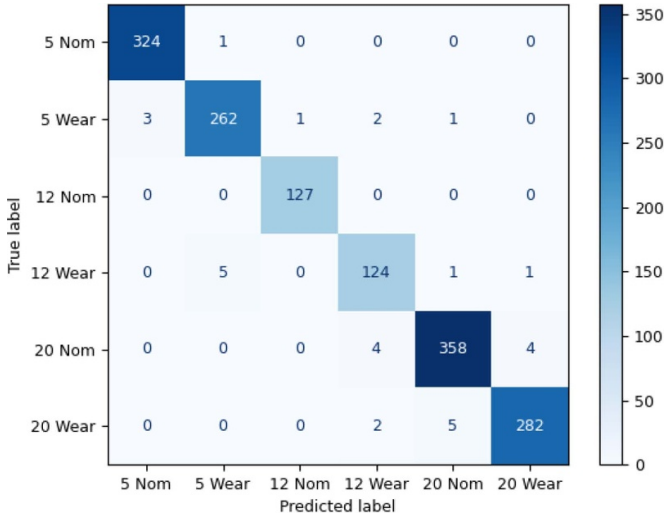

Figure 13. Confusion matrix plot of the random forest algorithm trained with acceleration features with feature importance greater than 0.05.

during the aluminum extrusion process on a real plant. The possibility of additional laboratory tests was prevented by the fact that the hydraulic block had such an high oil capacity that its testing required stopping all the other machineries of our department. Consequently, once the tests described in this work were completed, it was not possible to perform additional experiments. Further experiments are ongoing in a real industrial environment, but the time required to obtain wear damage, at the moment, is unpredictable.

Another limitation arises from the applicability of results to similar case-studies. In general, the relative importance of the different features and the numerical values obtained by the algorithms can be different from case to case, but the proposed approach based on RF features importance ranking allows selecting the features in laboratory tests, before the real application in industrial environment.

5.2. Features extraction

The vibration features to be selected depend on the valve mechanical characteristics, and in general the signal analysis is required in order to identify the features for the fault identification. In case of valve leakage, a portion of the fluid is forced to flow through the small orifice, generating structural vibration [18]. In the presented case-study four bands were selected, representing the valve response to the fluid-induced turbulence [15]. Parallel investigations performed with a laser Doppler vibrometer evidenced that the ultrasonic signals in the frequency band from 20 to 25 kHz allows distinguishing between the healthy and faulty conditions. The higher sampling rate, however, may be a problem for industrial hardware.

Features extracted from the current signals were particularly meaningful: after 0.1 s from the valve activation, the reverse electromotive force produced by the push pin's movement allows identifying the leakage condition. The selected features provided a good separation between the valve in a fault condition and the valve in nominal condition.

The size of the damage (0.1 mm) may be larger than the early damage of the valve and current or vibration signals alone could not be sufficient to evidence the spool wear. In these cases, only the sensor fusion can provide information about damages in the order of hundreds of nanometres.

5.3. Comparison on different data fusion approaches

The RF model ensured superior classification performances with respect to SVMs and k-NN. The RF complexity is also smaller than the SVM complexity of at least one order of magnitude. On the other hand, k-NN has a time complexity that is lower for training time but higher for prediction time. This difference was not noticeable during training time, due to the small size of the dataset, but it should be considered for bigger datasets in industrial environments. Moreover, RF is less sensitive to the dataset size. RF evidenced a higher false negative error, while the SVM-rbf is more balanced in the predicted errors with a tendency to false positive errors. SVM-rbf and SVM-linear have similar results, but SVM-linear has a tendency to obtain false negative errors. Even with higher prediction accuracy this property is less desirable in an industrial environment since it may lead to underestimate the health status of the valve and potential safety hazards. As shown in figure 10, training the RF model with only accelerometer's features provides lower Prediction Accuracy, also in this case features from different sensors provide better results.

6. Conclusions

Results presented in this paper evidenced that the data fusion approach allows the identification of the valve leakage phenomenon. Current and vibration data were selected after the literature review. The proposed method was tested in a specific case-study, that evidenced excellent capability in detecting incipient spool faults. Among the tested methods, the Random Forest model provided better performances; further studies

are needed to validate the method in a relevant industrial environment.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Fabio Conti https://orcid.org/0000-0003-4759-9076
Antonio Boiano https://orcid.org/0000-0002-5552-3680

References

[1] Zhang C, Zhao G, Chen H, Guan Y and Li H 2012 Optimization of an aluminum profile extrusion process based on Taguchi's method with S/N analysis *Int. J. Adv. Manuf. Technol.* **60** 589–99
[2] Marcotuli V, Marelli S, Casartelli R, Scaccabarozzi D, Saggin B and Tarabini M 2020 Compensation of temperature effects on an automatic system for diameter measurement *2020 IEEE Int. Work. Metrol. Ind. 4.0 IoT, MetroInd 4.0 IoT 2020—Proc.* pp 283–7
[3] Shao H, Lin J, Zhang L, Galar D and Kumar U 2021 A novel approach of multisensory fusion to collaborative fault diagnosis in maintenance *Inf. Fusion* **74** 65–76
[4] Li X, Cheng J, Shao H, Liu K and Cai B 2022 A fusion CWSMM-based framework for rotating machinery fault diagnosis under strong interference and imbalanced case *IEEE Trans. Ind. Inform.* **18** 5180–9
[5] Safizadeh M S and Latifi S K 2014 Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell *Inf. Fusion* **18** 1–8
[6] Diez-Olivan A, Del Ser J, Galar D and Sierra B 2019 Data fusion and machine learning for industrial prognosis: trends and perspectives towards Industry 4.0 *Inf. Fusion* **50** 92–111
[7] Yan J, Heng-Hu Y, Hong Y, Feng Z, Zhen L, Ping W and Yan Y 2015 Nondestructive detection of valves using acoustic emission technique *Adv. Mater. Sci. Eng.* **2015** 1–9
[8] Zhou J, Li L and Duszczyk J 2004 Computer simulated and experimentally verified isothermal extrusion of 7075 aluminium through continuous ram speed variation *J. Mater. Process. Technol.* **146** 203–12
[9] Ye G Y, Xu K J and Wu W K 2018 Standard deviation based acoustic emission signal analysis for detecting valve internal leakage *Sens. Actuators* A **283** 340–7
[10] Angadi S V and Jackson R L 2022 A critical review on the solenoid valve reliability, performance and remaining useful life including its industrial applications *Eng. Fail. Anal.* **136** 106231
[11] Xie L, Håbrekke S, Liu Y and Lundteigen M A 2019 Operational data-driven prediction for failure rates of equipment in safety instrumented systems: a case study from the oil and gas industry *J. Loss Prevention Process Ind.* **60** 96–105
[12] Zhang H 2018 The characteristic improvement of electromagnetic proportional directional control valve *J. Control Sci. Eng.* **2018** 1–8
[13] Trunzer E, Weis I, Folmer J, Schrufer C, Vogel-Heuser B, Erben S, Unland S and Vermum C 2018 Failure mode classification for control valves for supporting data-driven fault detection *2017 IEEE Int. Conf. on Industrial Engineering and Engineering Management (IEEM)* pp 2346–50
[14] Thompson G and Zolkiewski G 1997 An experimental investigation into the detection of internal leakage of gases through valves by vibration analysis *Proc. Inst. Mech. Eng.* E **211** 195–207
[15] Ji X, Ren Y, Tang H, Shi C and Xiang J 2020 An intelligent fault diagnosis approach based on Dempster-Shafer theory for hydraulic valves *Meas. J. Int. Meas. Conf.* **165** 108129
[16] Takahashi S 1972 United States Patent : 55,367 device for monitoring a fluid pressure system *United States Patent*
[17] Balakrishnan M and Kumar N 2015 Detection of plunger movement in dc solenoids, White Paper *Tex. Instrum. Inc.* pp 1–10
[18] Kaneko S, Nakamura T, Inada F, Kato M, Ishihara K, Nishihara T and Langthjem M A 2014 Vibration induced by pressure waves in piping *Flow-induced Vibrations* 2nd edn (Amsterdam: Elsevier) ch 5, pp 197–275
[19] Sharif M A and Grosvenor R I 1998 Internal valve leakage detection using an acoustic emission measurement system *Trans. Inst. Meas. Control* **20** 233–42
[20] Seong S H, Hur S, Kim J S, Kim J T, Park W M, Lee U C and Lee S K 2005 Development of diagnosis algorithm for the check valve with spectral estimations and neural network models using acoustic signals *Ann. Nucl. Energy* **32** 479–92
[21] Conti F, Conese C, Colombo M, Maggioni L, Moschioni G and Tarabini M 2021 Vibration signals for condition based maintenance of hydraulic valves *2021 IEEE Int. Work. Metrol. Ind. 4.0 IoT* pp 259–63
[22] Keith Mobley R 2002 *An Introduction to Predictive Maintenance* 2nd edn (Woburn, MA: Elsevier) (https://doi.org/10.1016/B978-0-7506-7531-4.X5000-3)
[23] Athanasatos P, Costopoulos T and Skarmea M 2005 Condition monitoring techniques for industrial fluid power systems *1st Int. Conf. Exp. Model. Athens*
[24] Angadi S V, Jackson R L, Yul Choe S, Flowers G T, Suhling J C, Chang Y K, Ham J K and Bae J I 2009 Reliability and life study of hydraulic solenoid valve. Part 2: experimental study *Eng. Fail. Anal.* **16** 944–63
[25] Kulkarni C S, Daigle M and Goebel K 2013 Implementation of prognostic methodologies to cryogenic propellant loading testbed *AUTOTESTCON (Proc.)* pp 314–20
[26] Tansel I N, Perotti J M, Yenilmez A and Chen P 2005 Valve health monitoring with wavelet transformation and neural networks (WT-NN), *2005 ICSC Congr. Comput. Intell. Methods Appl. 2005* pp 1–6
[27] Ngbede U M, Oluwasegun A, Choi M J and Jung J C 2019 Fault state detection in solenoid operated valve based on convolutional neural network using coil current signature
[28] Totten G E and De Negri V J 2012 *Handbook of Hydraulic Fluid Technology* 2nd edn (Boca Raton, FL: CRC Press)
[29] Keith Mobley R 2002 Vibration monitoring and analysis *Maintenance Engineering Handbook* 2nd edn (New York: McGraw Hill)
[30] Gold C and Sollich P 2003 Model selection for support vector machine classification *Neurocomputing* **55** 221–49
[31] Oshiro M O, Perez P S and Baranauskas J A 2012 How many trees in a random forest? *Int. Workshop on Machine Learning and Data Mining in Pattern Recognition* (*Berlin, Germany*, *July 2012*) pp 154–68
[32] Gustavo R C P, Batista E A P A and Monard M C n.d. A study of the behavior of several methods for balancing machine learning training data *ACM SIGKDD Explor. Newsl.* **6** 20–29
[33] Rahman M M and Davis D N 2013 Addressing the class imbalance problem in medical datasets *Int. J. Mach. Learn. Comput.* **3** 224–8
[34] Meland E, Henriksen V, Hennie E and Rasmussen M 2011 Spectral analysis of internally leaking shut-down valves *Meas. J. Int. Meas. Conf.* **44** 1059–72

[35] Morse P M and Ingard K U 1968 *Theoretical Acoustics* (New York: McGraw Hill)
[36] Kaewwaewnoi W, Prateepasen A and Kaewtrakulpong P 2010 Investigation of the relationship between internal fluid leakage through a valve and the acoustic emission generated from the leakage *Meas. J. Int. Meas. Conf.* **43** 274–82
[37] Kaewwaewnoi W 2008 Theoretical and experimental investigation to measure internal leakage of fluid through valves using acoustic emission *PhD. Prog. Rep. King's Mon*
[38] El-Shorbagy K A 1983 An investigation into noise radiation from flow control valves with particular reference to flow rate measurement *Appl. Acoust.* **16** 169–81
[39] Curle N 1955 The influence of solid boundaries upon aerodynamic sound *Proc. R. Soc.* A **231** 505–14
[40] Athanasatos P and Costopoulos T 2008 The effect of internal leakage of a 4/3 way valve on the response of industrial hydraulic systems *NTUA ME Proc. of Hydraulics-Pneumatics Training. TR-HP-0801*
[41] Lyons J L and Askland C L 1975 *Lyons' Encyclopedia of Valves* (Malabar, FL: Krieger Publishing Company)
[42] Yang H, Mathew J and Ma L 2003 Vibration feature extraction techniques for fault diagnosis of rotating machinery - a literature survey *AsiaPacific Vibration Conf.* (*Gold Coast, Australia*, *12–14 November 2003*)
[43] Louppe G 2014 Understanding random forests: from theory to practice (arXiv: 1407.7502)