From Predictive to Energy-Based Maintenance Paradigm: Achieving Cleaner Production through Functional Productiveness

Abstract: The introduction of Energy-Based Maintenance (EBM) practice in Sustainable Manufacturing (SM) is attracting significant attention from academicians, especially considering imposed European initiatives (e.g., Green Deal). Although traditional Predictive Maintenance (PdM) practice uses Machine Learning (ML) tools, it still relies on secondary (waste) energy indicators of a *p-f* curve (e.g., vibration, temperature). We introduce the notion of Functional-Productiveness (FP) for setting FP thresholds in detecting "Quasi-fault" events considering hydraulic power. Discretised hydraulic power signal with Recursive Feature Elimination (RFE) is used for feature extraction. Support Vector Machine (SVM), Random Forest (RF), Partial Least Square Discriminant Analysis (PLS-DA) and t-Distributed Stochastic Network Embedding (t-SNE) algorithms are used for Feature Selection (FS) process. The extracted features show latent degradation of a hydraulic control system of a Rubber Mixing Machine (RMM) performed by binary classification {None, Quasi-Fault} with SVM, RF, PLS-DA and Logistic Regression (LR). The results show that latent degradation led to a 20% drop in hydraulic power compared to the initial state, while the existing diagnostic practice of Lubricant Condition Monitoring (LCM) fails to provide such insights. Consequently, the study suggests that traditional monitoring practices that rely on static *p-f* indicators are becoming obsolete, leading to unnecessary energy waste and power loss.

Keywords: energy-based maintenance, hydraulic system, t-SNE, support vector machine, random forest, partial least square discriminant analysis, logistic regression

1 Introduction

1.1 Background and rationale

Understanding Cleaner Production (CP) as a: „*...preventive approach to managing the environmental impacts of business processes and products*“ (Srinivas, 2022), it is unavoidable to perceive that *preventive* can be considered formulaic in understanding the degradational consequences of poor maintenance policies and practices. Rhetorically, UNEP's definition of CP as „*...continuous application of an integrated preventive environmental strategy to processes, products, and services to increase overall efficiency and reduce risks to humans and the environment"* (UNEP, 2006) fails to address what are exactly the roles of preventive practices and questions how industrial maintenance is perceived within proposed definitions. As asset-intensive companies set their agendas in accomplishing high-end market demands, responding to disruptive market demands and EU-imposed normative regulations and requirements (e.g., Green Deal, SDGs – Sustainability Development Goals) with the help of Industry 4.0 features (e.g., Internet of Things (IoT), Cloud Computing) became much more manageable (Orošnjak, 2022). However, although existing maintenance research shifted towards „Maintenance 4.0“ with the introduction of Industry 4.0 (I4.0), it is a question, what are the actual maintenance constructs that altered maintenance from PdM (Predictive Maintenance) to M4.0? Other than adding I4.0 nomenclatures and associated features with buzzwords „smart maintenance“, „IoT maintenance“, etc., one cannot, with sophisticated exactitude, claim that such practices revolutionalised maintenance in a scientific sense.

The ongoing PdM concerns are mainly oriented towards DL (Deep Learning) philosophy (Li, 2018), whether for diagnostic or prognostic purposes. Detecting anomalies with, for instance, vibroacoustic signals (indicators) is still mostly researched within the PdM (Tang et al., 2022) of mechanical structures. Using computer vision with IC (infrared cameras) for detecting anomalies (Ullah et al., 2017) is an example of using thermography to predict faults in electrical power equipment. In the hydraulic sphere, the LCM (Lubricant Condition Monitoring) is mostly used since it provides insight into the operational state of the system's internal structures (Karanović et al., 2018). Included CM indicators can be used to assess potential anomalies produced by non-random stress events; however, existing research still builds on static and subjective (pre)set thresholds using „energy waste“ indicators (red line in Fig. 1) of the *p-f* curve. With the introduction of Sustainable manufacturing, we assume that the lack of maintenance advancement in the industrial „4th Wave“ is hidden in the dimension of sustainability. From the technical aspect, the idea is to shift from time and waste energy to primary (functional) energy (green line in Fig. 1), for diagnosis/prognosis purposes. Hence, we suggest shifting to *Energy-Based Maintenance* (EBM) practice.

The EBM is a type of maintenance strategy focusing on equipment and systems' power consumption to identify problems and schedule maintenance (Orošnjak et al., 2021). The goal of EBM is to optimise energy usage, reduce waste, and improve the efficiency and reliability of the equipment. There are several aspects of EBM: (1) Condition Monitoring Practice (CMP) – using sensors or other monitoring devices to measure power consumption and identify trends and changes that may indicate a problem; (2) Predictive Practice (PdP) (diagnosis and/or prognosis) – extracting features to prognose a potential failure or diagnose existing causes of degradation; (3) Corrective Practice (CrP) – identifying and fixing problems that are causing excessive energy consumption, such as worn parts or incorrect settings; (4) Optimisation Maintenance Practice (OMP) – conducting modelling and optimisation of maintenance-associated tasks and activities to minimise energy as a cost function (Xia et al., 2022). In addition, the benefit of using energy as an optimisation factor is that it can be transformed into monetary

value (Xie et al., 2019). Although state-of-the-art literature on maintenance involving sustainable philosophy experiences two streams: Sustainable Maintenance (SM) (Pires et al., 2016) and EBM philosophy (Orošnjak, 2021), different ideas drive the two. Namely, the SM practice is more oriented towards philosophical and social aspects of maintenance, while EBM is more oriented towards the technical and economical aspects.

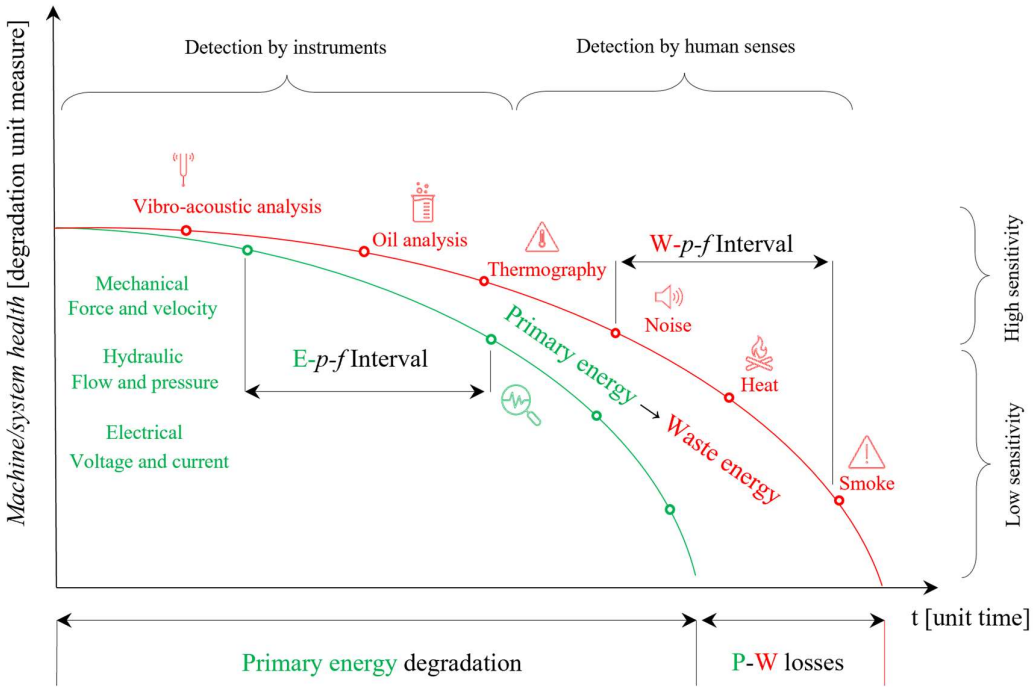


Fig. 1. Conceptualisation of the *p-f* curve with secondary energy indicators (red line) and primary energy indicators (green line)

1.2 Literature review

Existing research on PdM that includes waste-energy indicators usually relies on diagnostics, including contamination (e.g., oil analysis) (Huang et al., 2022), vibration signal (Liu et al., 2021), electrical power intake (Yin et al., 2021), etc. For instance, Jiang et al. (2021) decompose vibration signals by Empirical Wavelet Transform (EWT) to form features in the time and frequency domain and use correlation coefficients to detect changes in signal behaviour. Yin et al. (2021) propose a Genetic Algorithm-based wavelet neural network (GAWNN) for locating abnormal energy consumption patterns. They propose Fuzzy logic evaluation to define abnormal energy consumption by an expert judgement for conducting the correlation between abnormal machine components and energy consumption. Since sampling rates of different sensors are usually inconsistent, efforts have been made to address the problem of the multi-sampling rate problem. Huang et al. (2021) propose a multi-sensor data fusion method with CNN (Convolutional Neural Networks) to detect faults through automatically learned features from raw data. In a similar paper, Huang et al. (2022) propose CNN using classification with LDA (Linear Discriminant Analysis) for comparison and Pearson's correlation for analysing the correlation between features and faults. Under the same problem, Xiangkai Ma et al. (2021) propose multi-rate fusion using CNN architecture for classification for fault diagnosis. Although all of the research combine different data and deep learning, they still rely on secondary energy emission indicators, and ommitt hydraulic power signal.

Moreover, there is an existing problem in anomaly detection in hydraulic systems considering detecting abnormal behaviour patterns in signal processing. The usual practice is to build a classification model with or without prior information about the operational state of the system, using supervised or unsupervised learning, respectively. The problem concerns defining fault or failure thresholds for delineating the operational states as „healthy“ or „non-healthy“. Usually, these thresholds are set subjectively without explanation of these pre-set thresholds (Tang et al., 2021) and with explanations behind the probability assumptions of the behavioural process. It raises the question about Gaussian (or linear) probability assumptions of faults, especially since they (almost) always show non-Gaussian and non-linear characteristics (Pei et al., 2021). Secondly, it also questions whether such subjectively imposed abnormal behaviours by simulating faulty conditions and failures happen in practice. Or what is the benefit of simulating such faults in practical (industrial) encirclement?

On one side, the existing research regarding EBM is currently in the early infancy stage, mostly relying on simulations and cost-benefit analysis (Orošnjak et al., 2021). Only a portion of studies address practical industrial applications (Yin et al., 2021) and compare EBM to other practices (Orošnjak, 2021); however, the results show poor performance compared to traditional practices. The ongoing research reflects a top-down approach, suggesting that extracted primary energy/power data are mostly used for optimisation purposes (Xia et al., 2018) on a strategic level (Zhou and Yi, 2021) and maintenance decision-making (MDM), thus lacking operational and technical applications of the proposed solutions. On the other side, the PdM research still relies on energy-waste indicators of a *p-f* curve, neglecting the deviation of primary (functional) power for gaining insights about the degradation of system performance.

1.3 Study aims and objectives

Considering that labelling (supervised classes) usually includes pre-set 95% control thresholds (González-Muñiz et al., 2022) or static-fixed thresholds in determining the system’s or unit’s failed state (Emec et al., 2016), we aim to resolve this problem of functionality by using the concept of Functional Productiveness (FP) (Orošnjak, 2021). The underlying reason for introducing FP is to avoid labelling data under controlled experimental (testbed) conditions, especially to avoid artificially „simulated faults“. On the other hand, the industrial environment works in an extreme operating regime and, therefore, is much more difficult to classify or label features and fault conditions. Therefore, we experiment in a real-working environment instead of controlled (testbed) conditions using dynamic FP thresholds for binary classification.

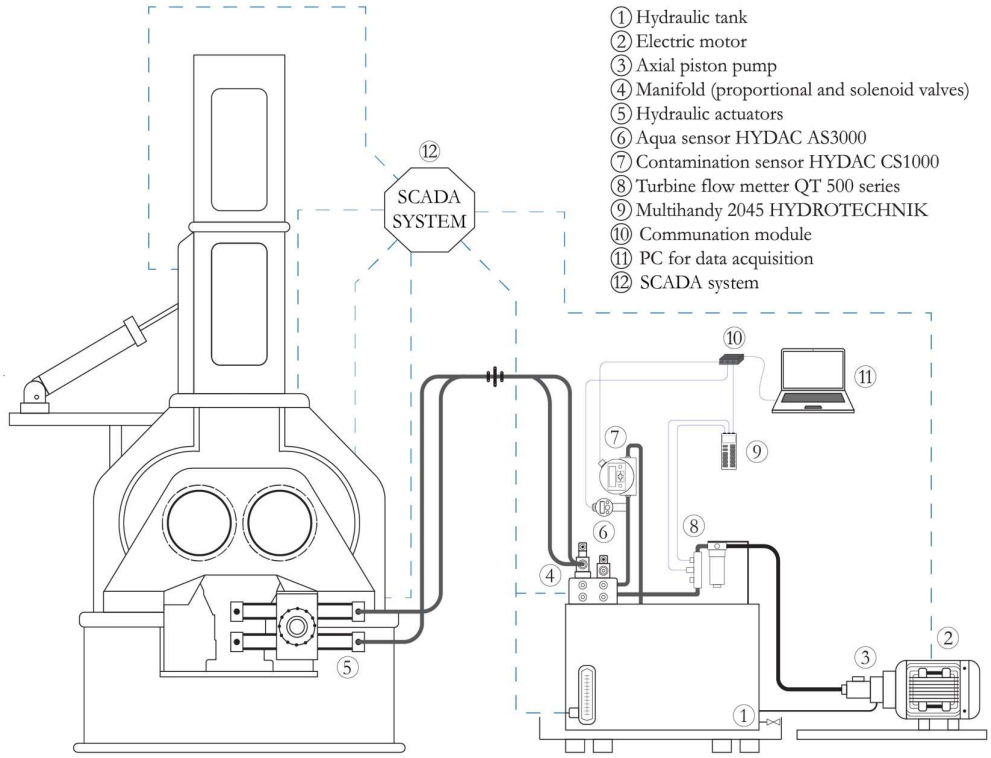
Feature Selection (FS) has become one of the most important topics in feature engineering, and an extensive elaboration on FS is given throughout the paper. The underlying reason for this is that using FS factors like Pearson’s Correlation Score (CS) and XGBoost Feature Importance (XFI) can suffer from redundancy (Ma et al., 2021) or, in the case of inaccurately detecting feature boundaries can easily lead to misclassification (Yu et al., 2020). This paper proposes FS by incorporating RFE (Recursive Feature Elimination) strategy. The models used for RFE-FS include Partial Least Squares Discriminant Analysis (PLS-DA), Support Vector Machine (SVM) and Random Forest (RF). In addition, the t-distributed Stochastic Network Embedding (t-SNE) is used to visualise and gain insight into the global information provided by extracted features. Finally, we compare the results with existing maintenance showing that latent degradation caused 20% of unnecessary energy waste that was not observed by the LCM practice.

The rest of the study is structured as follows. The second chapter explains a study's methodological and experimental design and flow diagram of data processing, including raw data acquisition, feature extraction, selection, and classification algorithms. The third chapter includes results and discusses results obtained from raw data and machine learning modelling. Finally, the study closes by concluding remarks, contributions to the literature and future research directions.

2 Methodology and experimental analysis

2.1 Rubber Mixing Machine Hydraulic Control System

The hydraulic system of a Rubber Mixing Machine (RMM) is used for the experiment. The hydraulic control part (Fig. 2) operates by opening and closing the saddle of the RMM chamber and consists of a tank (1), electric motor (2) and hydraulic pump (3), control and regulation valves (4), and linear position actuators (5). The SCADA (Supervisory Control and Data Acquisition) system monitors the process and provides additional data about the system. The data collection process from RMM was performed from 06.10.2021-12.12.2021. The acquisition of data outside of the SCADA (12) system includes physically mounted contamination sensors: HYDAC Aqua Sensor AS3000 (6), HYDAC CS1220 contamination sensor (7), QT500 Flow Turbine (8), MultiHandy 2045 (9), a communication module for AS3000 and CS1000 (10), and PC for data acquisition (11). The process starts with inserting rubber (bulk) material. When the material reaches the defined temperature (°C) within the chamber, it automatically activates the hydraulic system via SCADA. The pump provides hydraulic power to the actuators (5) that operate the saddle. The actuators are equipped with SCADA sensors that regulate the speed of opening, unloading and closing the saddle (i.e., chamber doors). Existing maintenance practice within the company includes regular monitoring, inspections, and replacements of parts, filters, oil refilling, etc., through the LCM program. The LCM includes physio-chemical and elemental analysis to detect wear or oil-related contamination. For a detailed description of the experiment, methods and procedures for raw data extraction, the reader is referred to **Supplementary material 1.**



147

148 Fig. 2. Experimental installation

149 **2.2 Raw data acquisition and (pre)processing**

150 Data acquisition and pre-processing include three steps (Fig. 3). The first step includes acquiring raw data from
151 the SCADA system, hydraulic power-related data (pressure and flow), LCM (oil-contamination) data, and
152 response *Y* data of each performed cycle. SCADA system's data is automatically generated during the experiment
153 and extracted in csv files. LCM includes physical-chemical and elemental analysis (e.g., oil contamination). The
154 labelling is performed using the FP concept (Orošnjak, 2021). The FP generates dummy variable *Y* as [0, 1], as
155 [None, Quasi-fault] condition. The labelling is performed as a semi-supervised mechanism by setting the FP
156 threshold to 95% normal operation with a penalisation function to remove false-positive points. For a detailed
157 description of FP thresholds, please see appendix **A1.** Response *Y* label by Functional-Productiveness

158 Table 1. Collection of data from SCADA and sensors

No.	Parameter	Unit	Component	Device
1	Hydraulic volumetric flow	l/min	Hydraulic	MultiHandy 2045
2	Hydraulic pressure	bar	Hydraulic	MultiHandy 2045
	Hydraulic power	kW	Hydraulic	MultiHandy 2045
3	Hydraulic oil temperature	°C	Hydraulic	CS/AS/SCADA
4	Chamber temperature	°C	Chamber	SCADA
5	ISO 4406 contamination	ISO Code	Hydraulic	CS1000
6	Water saturation	%water	Hydraulic	AS3000
7	Load	kg	Chamber	SCADA
8	Cycle length	sec	System	SCADA
10	Hydraulic cycle time	sec	Hydraulic	SCADA
11	Hydraulic actuators reaction	sec	Hydraulic	SCADA

12	Energy consumption	kWh	System	SCADA

Data (pre)processing reduces the complexity of data by considering only the data(sets), either extracted or selected, in which the information gained for establishing a model increases the performance of a model. Examples of failures during system operation were jamming the saddle, power-off due to a break in electrical supply, and other cases where total failure/stoppage caused bias in collected data. Because the system is an asset-intensive process and works 24/7, approximately 320 days a year, the stoppages happened regularly due to failures or batch changes of material. Raw data collection includes variables given in Table 1.

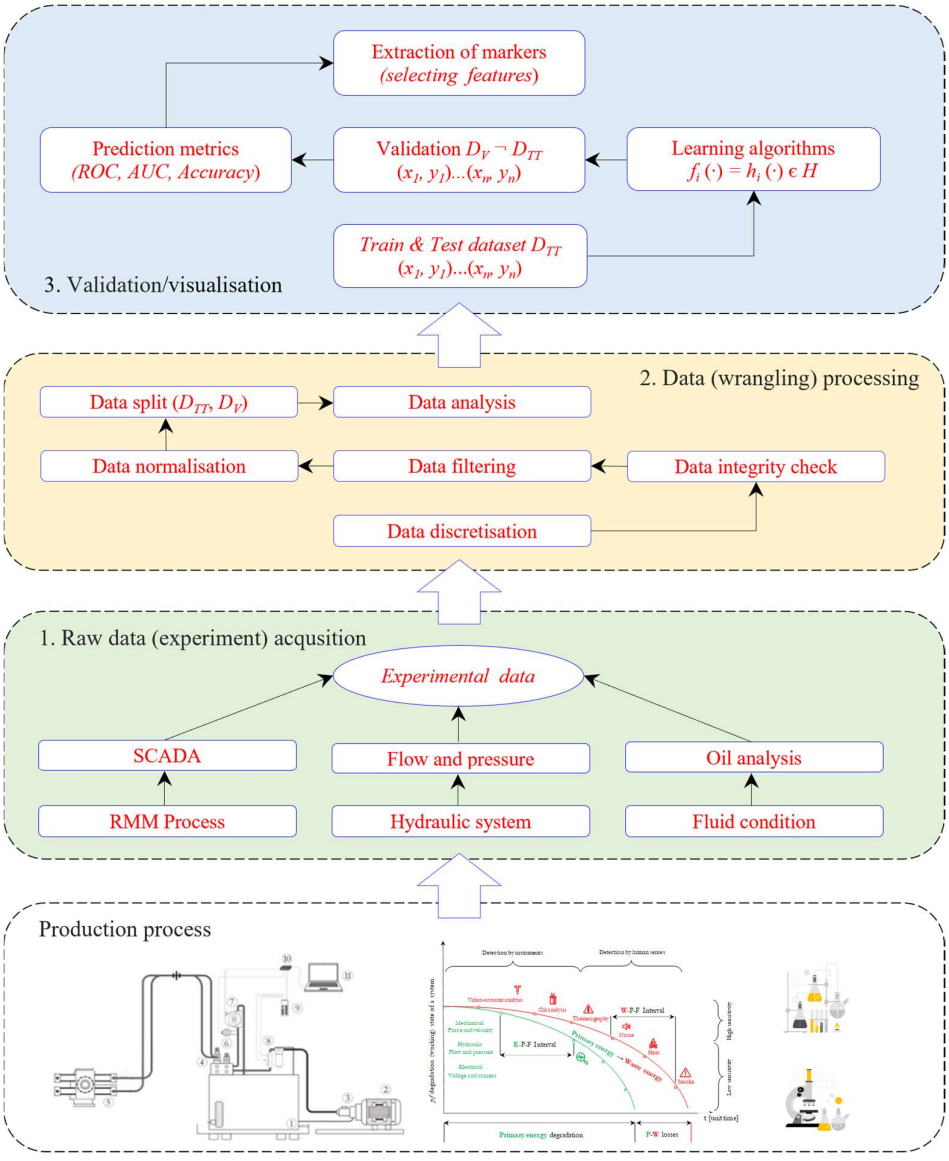


Fig. 3. Data flowchart from the production process to machine learning classifiers and features

After the extracted raw data, *discretisation* is performed on the *HyPower* signal data by sampling, i.e., transferring continuous-time to discrete-time $x(t) \rightarrow x[n]$, such that $x(t) \rightarrow x(nT)$ where $n \in \mathbb{R}$ $\{0, 1, 2 \ldots n\}$ and T is the sampling period of 0.1 sec (f_s = 10 Hz). By binning the signal into three parts for opening, idling, and closing the saddle, we have extracted discrete domain features with statistical formulas given in Table 2.

Table 2. Time-discrete domain features

Time-domain feature	Feature notation[1]	The formula for feature generation
Mean value	N_Mean_*X*S	$MEAN = \frac{1}{n}\sum_{i=1}^{n} N_i$
Standard deviation	N_StDev_*X*S	$StDev = \frac{\sqrt{(x_i - \bar{x})^2}}{n-1}$
Root Mean Square	N_RMS_ *X*S	$RMS = \sqrt{\frac{1}{n}\sum_i x_i^2}$
Minimum	N_Min_*X*S	$\min(N_i)$
Maximum	N_Max_ *X*S	$\max(N_i)$
Quartiles (Q1,Q3)	N_nQ_ *X*S	$nQ_{1;3} = x_{(k)} + a(x_{(k+1)} - x_{(k)})$
Interquartile Range	N_IQR_ *X*S	$\mathrm{IQR} = \mathrm{N\_3Q\_}XS - \mathrm{N\_1Q\_}XS$
Peak-to-peak	N_P-P_ *X*S	$P\_P = \lvert \mathrm{N\_MIN\_}XS - \mathrm{N\_MAX\_}XS \rvert$
Skewness	N_Skew_ *X*S	$Skew = \tilde{\mu}_3 = \frac{\sum_i^n (x_i - \bar{x})^3}{(n-1)\cdot\sigma^3}$
Kurtosis	N_Kurt_ *X*S	$Kurt = \tilde{\mu}_4 = \frac{\sqrt{n(n-1)}}{n-2}\frac{\sum_i^n (x_i - \bar{x})^3}{(n-1)\cdot\sigma^3}$

[1]*X*S – OS for Opening Saddle regime; IS for Idle Saddle Regime; CS for Closing Saddle Regime.

Data integrity includes maintaining data accuracy and consistency during the experiment with (pre)defined notations, file types, storage, etc. The data integrity step can be replaced with data quality, which is essential in assuring rigidity and avoiding bias in our data collection and stewardship. *Data filtering* includes the inspection of data and variables (and features) with high multicollinearity used later in modelling. Thus, filtering here reduces the cost of modelling and assuring data redundancy. After the filtering and extracting features, we performed *data normalisation* by standardisation and split dataset D into a training-testing dataset D_{TT} (70%) and validation dataset D_V (30%). The problem of overfitting can still reside and question the model's applicability in practice. Therefore, to reduce the noise and remove redundant features, respecting *Occam's razor* law, we use the RFE-FS strategy. The RFE-FS is a wrapper method. The idea of the RFE-FS is to reduce unnecessary and redundant features and avoid overfitting. Also, the feature subset's cross-validation technique can help with this problem. However, although cross-validation highlights significant features by reducing noise in data, the feature subsets can be biased towards the specific algorithm (Remeseiro and Bolon-Canedo, 2019). The problem also exhibits in, for instance, Decision Trees and Multi-layer Perceptron (MLP) networks with regularisation, which can turn off the irrelevant features (Duch et al., 2004). Also, discarding important features that do not improve the classification by the proposed algorithm can lead to information loss because some features are more informative if combined with other features (Zenglin Xu et al., 2010). Finally, RFE-FS is used with a combination of different algorithms: SVM (Chang and Lin, 2008), PLS-DA (Ruiz-Perez et al., 2020), and Random Forest (RF) (Chen et al., 2020).

2.2.1 Support Vector Machine (SVM)

Usually, distance-based and entropy-based methods with SVM wrappers are used in industrial applications for fault diagnosis (Hui et al., 2017; Liu et al., 2014; Lu et al., 2015). The SVM's hyperplane function that separates two labelled datasets given instance pairs as training vectors $\boldsymbol{x}$ and labels y such that $\boldsymbol{x} \in \mathbb{R}^n$ and a label $y \in \mathbb{R}^m$, $y \in [0, 1]$, where i is instance and j feature such that as $\boldsymbol{x}_i^j$. The SVM attempts to separate data by finding an optimal hyperplane as a function $f(\boldsymbol{x}) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b$, where $\boldsymbol{w}$ is the feature weight (or w_i for each vector) and b is the bias. The loss function solves the optimisation problem:

$$\min_{\boldsymbol{\omega},b} \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{l}\xi(\boldsymbol{w}, b; \boldsymbol{x}; y), \quad (1)$$

$$\text{subject to } y_i(\boldsymbol{w}^T\boldsymbol{x} + b) \geq 1 - \xi, \xi \geq 0.$$

where $\xi(\boldsymbol{w}, b; \boldsymbol{x}; y_i)$ is a loss function, and $C \geq 0$ is a penalty parameter of the training error. The loss functions:

$$\max(1 - y_0(\boldsymbol{\omega}^T\boldsymbol{\phi}(\boldsymbol{x}) + b), 0), \text{ and } \max(1 - y_1(\boldsymbol{\omega}^T\boldsymbol{\phi}(\boldsymbol{x}) + b), 0)^2, \quad (2)$$

are represented as the *l1* and *l2* loss functions of SVM, respectively, where ϕ is a function mapping data into high-dimensional space. Considering that datasets, when projected on a scatter plot, do not reflect linear separation, a kernel trick can be used. The purpose of the kernel, such as $K(\boldsymbol{x}_0, \boldsymbol{x}_1) = \phi(\boldsymbol{x}_0)^T\phi(\boldsymbol{x}_1)$, is to transform low-dimensional input space into „better" high-dimensional space for separation, where hyperplane takes n-1 of the feature vectors (e.g., 3D plot uses 2D hyperplane for separation). From the SVM $\phi(\boldsymbol{x})=\boldsymbol{x}$ with the linear kernel $K(\boldsymbol{x}_0, \boldsymbol{x}_1) = \boldsymbol{x}_0^T\boldsymbol{x}_1$, where $\boldsymbol{x}_0, \boldsymbol{x}_1$ represents the binary classes [0, 1] = [None, Quasi-fault].

2.2.2 Partial Least Squares Discriminant Analysis (PLS-DA)

The PLS-DA is extremely popular in classification (Perk et al., 2011) and feature selection (Yan et al., 2017) in various applications outside industrial systems. The model, however, is often misused since it is prone to overfitting (Ruiz-Perez et al., 2020); thus, cross-validation is used to avoid misinterpretation (Westerhuis et al., 2008). The PLS-DA takes the relation between features and constructs a new set of features corresponding to the projection (loading) into lower dimensional space vectors, i.e., Latent Variables (LVs) or Principal Components (PCs). Unlike unconstrained PCA (Principal Component Analysis), which constructs new features by linear transformations that best explain variance within the data, the PLS-DA is constrained, meaning it projects LVs concerning the class label y_0 or y_1.

Explaining PLS-DA analytically, what it does is that it transforms dataset X into lower dimension matrix A, where X is $n \times m$ matrix, and A is transformed X matrix into a lower dimension of $m \times d$-dimensional vectors C, with error matrix E, such that $C = XA + E$. The transformed C contains rows corresponding to the transformed vectors, while the E matrix contains information for the next PC. The difference between PCA's PCs is that in PC1 (or LV1), the PCA preserves the most variance of the original dataset X, while PLS-DA preserves in PC1 (LV1) as much variance but to target label or class. Eigenvectors of the covariance matrix C give PCs as:

$$C = \frac{1}{n-1}X^T C_n X\ , \quad (3)$$

where C_n is the $n \times n$ centre matrix. The loadings ($L_1 \ldots L_n$) are given, for eigenvectors $e_1 \ldots e_n$ and eigenvalues $\lambda_1 \ldots \lambda_n$ of C such that:

$$L_i = \sqrt{\lambda_i}e_i\ , \text{ for } i = 1,\ldots n, \quad (4)$$

while for PLS-DA, the C is formulated as:

$$C = \frac{1}{(n-1)^2}X^T C_n y y^T C_n X\ , \quad (5)$$

and through an iterative process, we get loading vectors a_i after k iterations, such that:

$$\max_{(a_k, b_k)} cov(X_k a_k, y_k b_k), \quad (6)$$

where b_k is the loading of each label y_k, $X_1 = X$, and X_k and y_k are error matrices after the transformation of previous k-1 components. The misinterpretation lies in believing that the model should be "better" with higher explained

variance, which in the case of PLS-DA, can cause bias. Overall, it can be considered that PLS-DA is a supervised version of PCA (Ruiz-Perez et al., 2020), such that PLS is just ordinary PLS regression with a special dummy *y*-variable where PC1 is used to best separate the classes, while PC1 in PCA is used that contains the most variance in a given dataset. For a detailed mathematical explanation, the reader is referred to (Ruiz-Perez et al., 2020).

2.2.3 Random Forest (RF)

Random Forest (RF) can be considered as a combination of ML methods bagging (bootstrap aggregating) and random selection (Kaur and Malhotra, 2008). The prior makes a prediction based on a majority vote of selected trees, while the latter searches each node for the best *K* split within a given feature subset. Although they are ensembles of decision trees (DTs), unlike DTs, they tend to grow deep and result in overfitting. The practical application of RF can be explained in short as follows. Consider a decision tree with *M* leaves that divides the feature space into *M* regions R_m, $1 \leq m \leq M$. For each tree, the function $f(x)$ is defined as:

$$f(x) = \sum_{m=1}^{M} c_m \prod (x, R_m), \tag{6}$$

where *M* is the number of regions in feature spaces, R_m is the region appropriate to *m*, while c_m is the constant suitable to *m*:

$$\prod (x, R_m) = \begin{cases} 1, & if\ x \epsilon R_m \\ 0, & otherwise \end{cases} . \tag{7}$$

The last classification is set by a majority vote, including all trees. In the RF algorithm, solving is done by the goodness of split *s* at node *t*. Dividing and splitting the tree into nodes concerning the greatest value, i.e., least impurity of a split, is done by breakdown criteria:

$$\Phi(s,t) = \Delta i(s,t) = i(t) - P_R i(t_R) - P_L i(t_L), \tag{8}$$

where *s* is split at node *t*, t_R is split at node *t* proportional to the number of objects, Δi is a function with t_R that has probability P_R, the t_L has probability P_L that is defined as decreasing impurity with objects in t_L, while $\Phi(s, t)$ is the goodness of split, $i(t)$ is a function with t_R. Although the RF is done via the *varImp()* function, the detailed mathematical formulation can be found in Breiman (2001). Also, detailed description of RFE-FS results is given in **Supplementary Material 2**.

3 Results & Discussion

3.1 Raw data and feature extraction

Since different anomalies are detected during an RMM, the sequence window of performing a hydraulic cycle is split into three segments: *OS* as the opening saddle segment, *IS* as the idle saddle segment, and *CS* as the closing saddle segment. The graphs in Fig. 4 depict the hydraulic power signal for performing one hydraulic cycle, which lasts approximately 20 seconds.

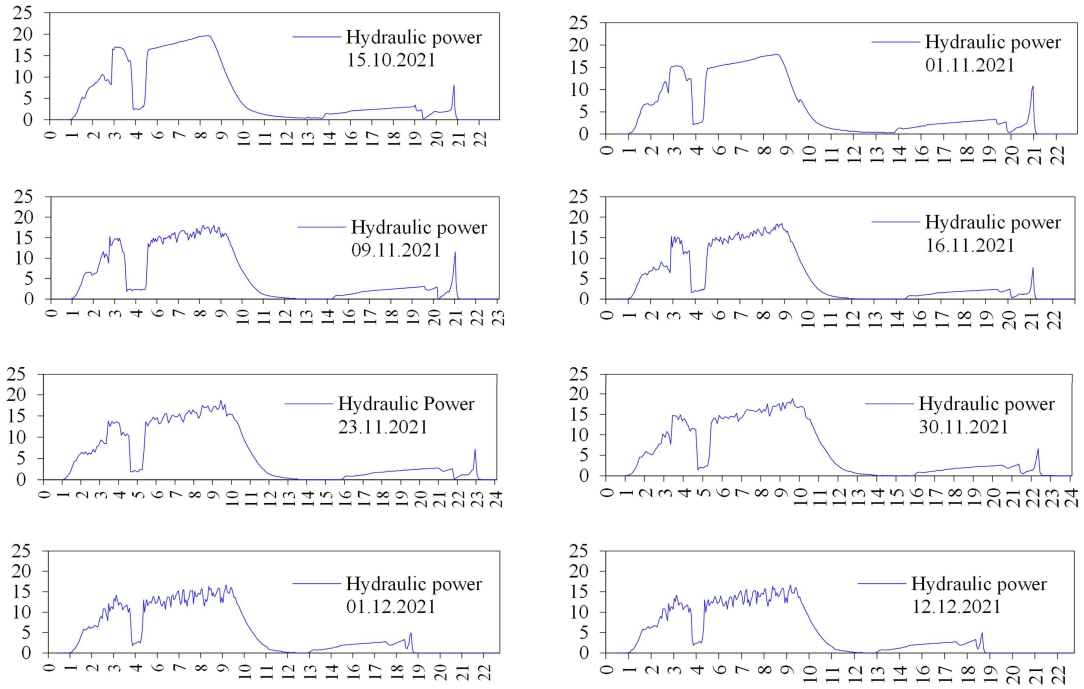


Fig. 4. Hydraulic power [kW] readings (*y*-axis) and time [sec] (*x*-axis) for performing the hydraulic cycle

Data acquired from SCADA include hydraulic cycle time *HS_cycle_time* and time between the rubber mixing process and hydraulic cycle as *HS_idle_time* (Fig. 5).

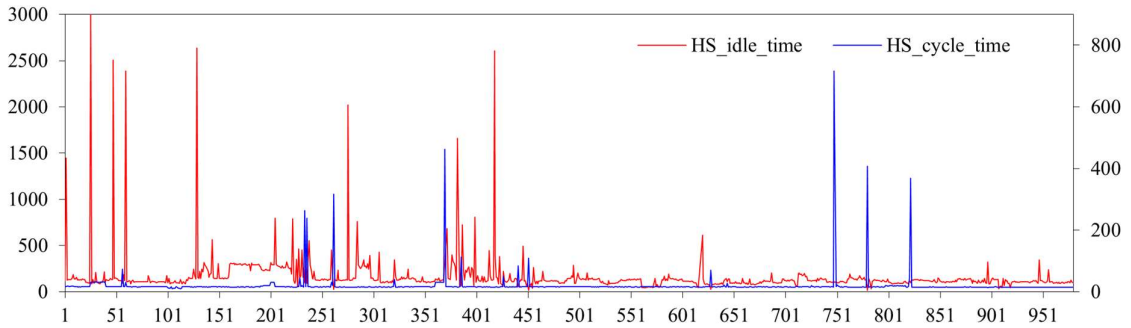


Fig. 5. Measurement of hydraulic idle time [sec] between cycles (primary *y*-axis) and hydraulic cycle time [sec] (secondary *y*-axis) of performed cycles (*x*-axis)

Within the monitoring period, some cycles were stopped before starting again since *HS_idle_time* had a discrepancy between cycles (Fig. 5red line). The time to perform a hydraulic cycle (Fig. 5 blue line) was over 3 minutes in a few instances due to the jamming of the saddle. The movement speeds *T1* and *T2* included measurements of fast and slow saddle opening, the same as *T4 and T5* for fast and slow saddle closing, respectively. The *T3* includes an idle time of the saddle for unloading the rubber material.

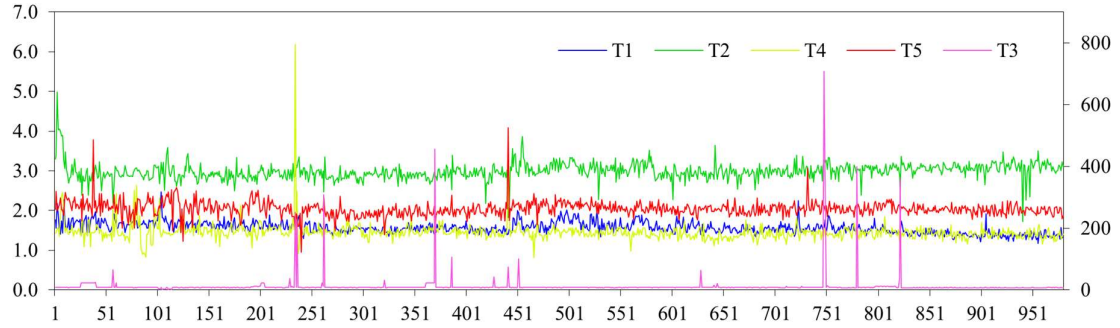


Fig. 6. Actuator speed [sec] for opening (T1 and T2), closing (T4 and T5) (primary *y*-axis), and idle saddle position (T3) of measured cycles (secondary *y*-axis)

The *T3* (Fig. 6 purple line) signal shows that in a few instances, the saddle was jammed at the 750th cycle (>660 seconds). The obtained signals also depict several deviations in cylinder speed of extension (*T1* and *T2*) and retraction (*T4* and *T5*). The LCM data includes the measurement of water saturation via AS3000 (Fig. 7) and particle contamination via the CS1000 sensor (Fig. 8).

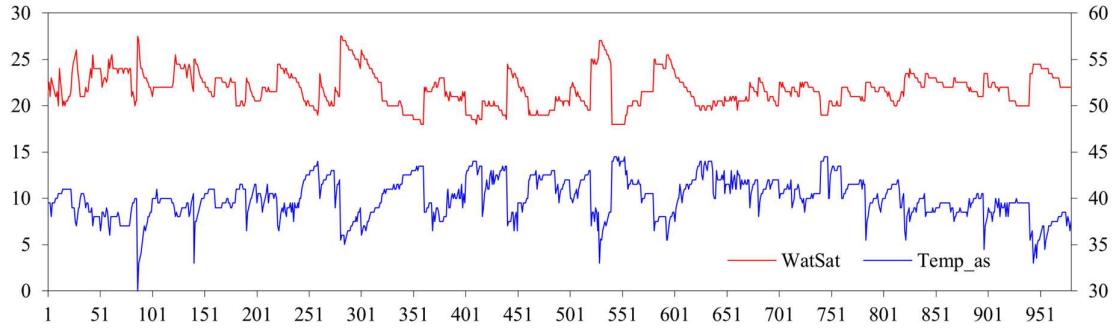


Fig. 7. Water saturation level (y_1-axis) and temperature (y_2-axis) during experiment monitoring cycles (x-axis)

The obtained results from inline monitoring of water saturation do not show significant discrepancies concerning water-related contamination of hydraulic oil. The rule of thumb for an alarming limit of possible water-induced contamination is at 200 ppm, which is by measurement of AS3000 sensor at ~50% saturation (Day and Bauer, 2007), which in this case did not show the separation of water molecules from entrained into free form.

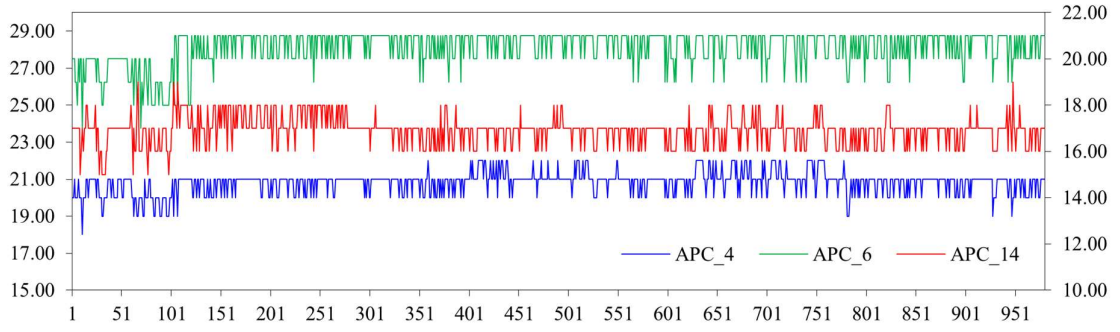


Fig. 8. APC readings according to ISO4406 code for particles ≥4 µm (y_1-axis) and for particles ≥6 µm and ≥14 µm (y_2-axis) during experiment monitoring cycles (x-axis)

The readings from APC (Automatic Particle Counter) show relative contamination considering recommended cleanliness levels by ISO4406 are at 18/15/13 for the specific system (MPFiltri, 2020). The readings throughout the experiment show 21/20/17, suggesting contamination either with particles (e.g., dirt or wear). However, the problem with APCs is that they can show biased results since they also include water and air droplets as readings. Even so, elevated levels suggested further laboratory oil analysis. The laboratory analysis included physio-chemical analysis (Table 3) – density (ASTM D1298), viscosity (ASTM D445), viscosity index (ASTM D2270), flame (ASTM D97) and flow point (ASTM D92), TAN (ASTM D664), Zinc (ASTM D4927), and analysis using Wavelength Dispersive X-Ray Fluorescence (WDXRF) spectrophotometry (Table 4).

Table 3. Laboratory analysis of physical and chemical characteristics of hydraulic fluid

TBS [h]	CO	DEN	V40	V100	VI	FP	FLP	WC	TAN
New oil	BY	863.50	45.78	6.98	109	230	-32	13	0.42
392.0	DY	880.70	53.30	7.53	103	218	-39	24	0.43
427.4	DY	880.40	53.35	7.57	103	216	-38	19	0.45
270.4	DY	880.00	53.63	7.57	103	216	-39	17	0.53
252.8	DY	880.00	53.33	7.55	104	215	-39	16	0.48
221.6	DY	880.23	53.32	7.55	104	215	-39	25	0.46

TBS = Time between samples; CO = color and odor; DEN = density of oil [kg/m^3]; V40 = Viscosity at 40°C [mm^2/s]; V100 = Viscosity at 100°C [mm^2/s]; VI = Viscosity index [-]; FP = Flame point [°C]; FLP = Flow point [°C]; WC = Water content [ppm]; TAN = Total acid number [mgKOH/g].

Table 4. Elemental analysis of particles within hydraulic fluid using WDXRF spectrophotometry

TBS [h]	Zn	Fe	Pb	Cu	Si	Sn	Cr	Al	Ag	Ni	Mn	Cd
New oil	0.037	2	0	0	27	0	2	0	0	3	0	0
392.0	0.034	5	0	0	24	0	1	0	0	0	0	0
427.4	0.034	4	0	0	29	0	1	0	0	0	0	0
270.4	0.034	4	0	0	18	0	1	0	0	0	0	0
252.8	0.034	3	0	0	20	0	2	0	0	0	0	0
221.6	0.034	5	0	0	27	0	2	0	0	0	0	0

TBS = Time between samples; Zn = Zinc; Fe = Iron; Pb = Lead; Cu = Copper; Si = Silicon; Sn = Tin; Cr = Chromium; Al = Aluminium; Ag = Silver; Ni = Nickel; Mn = Manganese; Cd = Cadmium.

Based on the results obtained, the evidence suggest changes in oil viscosity, flow and flame point. Considering that system was refilled to refresh the oil, which can be noticed with TAN change after 4th sample, the conclusions cannot be obtained solely from the laboratory analysis. In addition, the elemental analysis using WDXRF does not show signs of wear due to appropriate filtering and maintenance, suggesting that high APC readings are due to rubber material that contaminated the fluid. In sum, non-destructive LCM results fail to provide conclusions regarding the degradation within the system.

3.2 Feature extraction through RFE

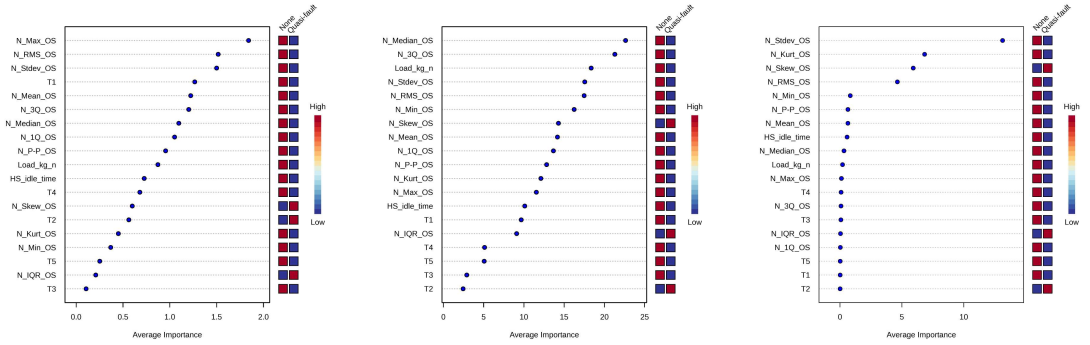
Obtained features by the RFE-FS method using feature weight of PLS-DA, SVM, and RF are depicted in Fig. 9. The performances of models are depicted through ROC (Receiver Operator Characteristic) curve. The ROC curve is a plot of sensitivity (*y*-axis) and 1-specificity (*x*-axis) and is calculated as:

$$Sensitivity = TPR = \frac{TP}{TP+FN}, \tag{7}$$

$$1 - Specificity = FPR = \frac{FP}{FP+TN}. \tag{8}$$

The given equations provide values for estimating AUC (Area under the ROC curve). The AUC values with ROC curves for different features are given with confidence intervals (CI). The accuracy is generated from two (2) to all features (19) by depicting model improvement scores. Additionally, the AUC scores with CI are obtained via

cross-validation (CV), specifically, MCCV (Monte Carlo Cross-Validation). Compared to other CV techniques (e.g., LOOCV – leave one out cross-validation; SCV – stratified cross-validation), the MCCV reduces bias in contributing to measurement accuracy.



(a) (b) (c)

Fig. 9. Multivariate analysis of feature (y-axis) importance measure (x-axis) by models (a) PLS-DA ranking features model; (b) RF-ranking features model; (c) SVM-ranking features model

The feature importance score by PLS-DA shows poor performance, even with all included features (Fig. 10). Moderate results are achieved with PLS-DA or SVM model (Fig. 11) with RF classification. The models using SVM-selected features seem to contribute the most to the accuracy regardless of the model (Fig. 12, Fig. 13). However, although RF classification using PLS-DA and SVM provide the highest scores, after evaluation of the misclassification using the validated dataset the RF shows a reduction in accuracy.

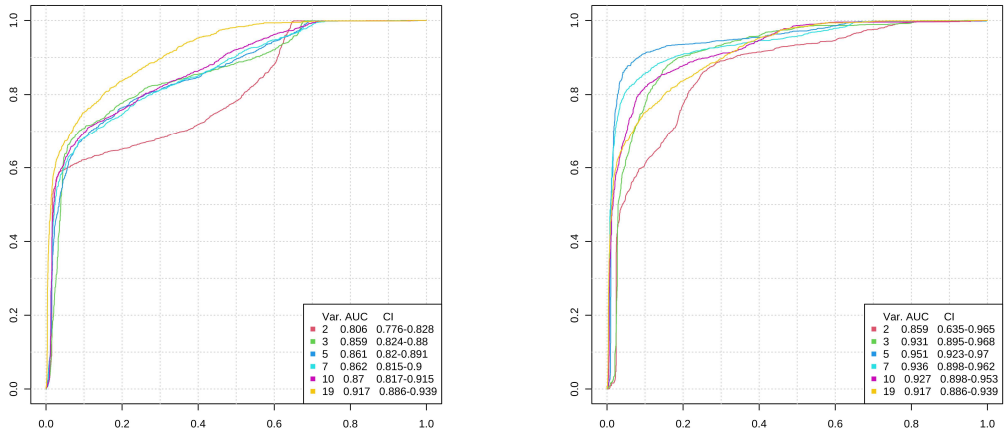
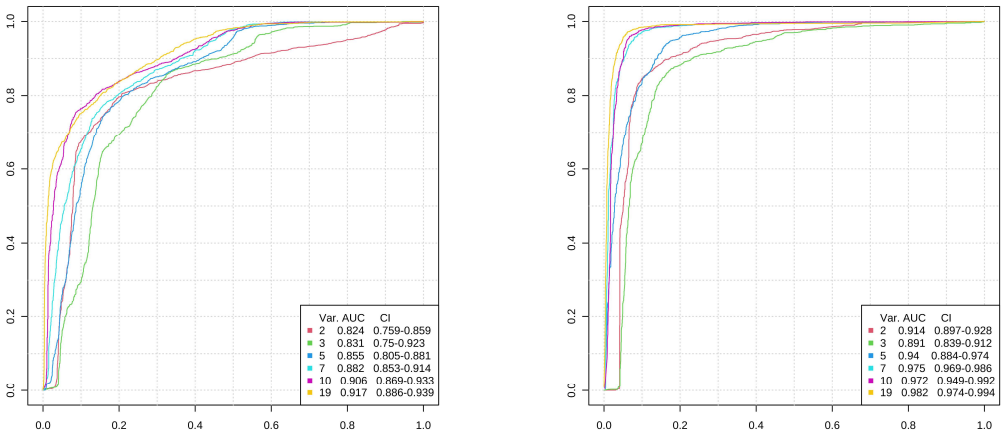
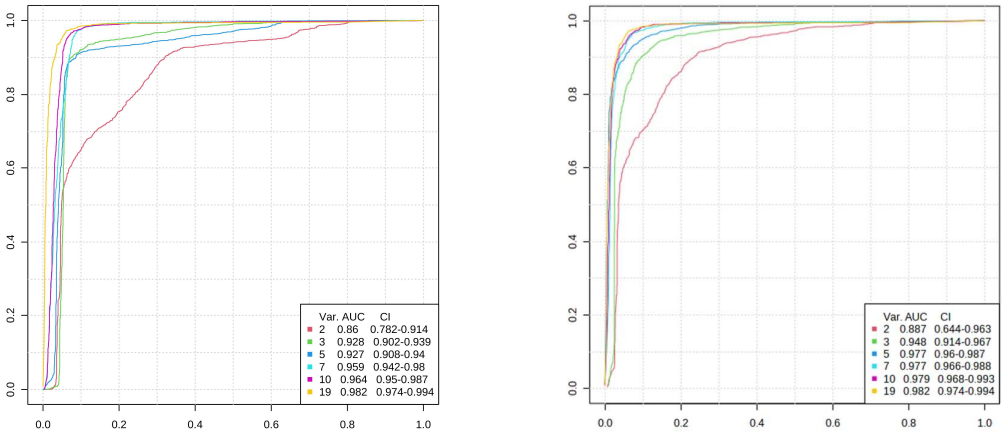


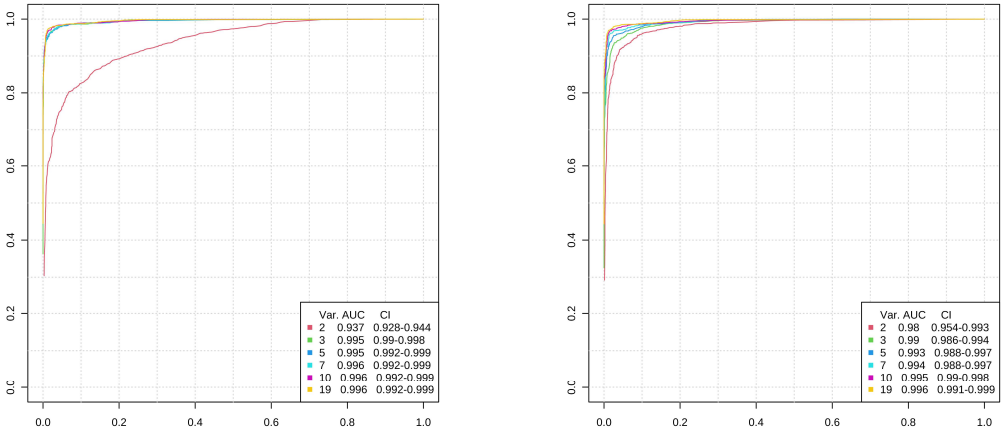
Fig. 10. PLS-DA classification with 2 LVs of PLS (left) and with SVM features (right)



314 Fig. 11. PLS-DA with RF features (left) and SVM with RF features (right)



315 Fig. 12. SVM classification with PLS-DA selected features (left) and with SVM features (right)



316 Fig. 13. RF classification with PLS-DA features (left) and with SVM features (right)

Contributions with more than five features do not significantly increase the accuracy of included models (Fig. 9c). For comparative purposes and to gain insight into model-selected features, we perform data visualisation via t-SNE.

3.3 t-distributed Stochastic Network Embedding (t-SNE)

The t-SNE is a non-linear dimensionality reduction technique mostly used to visualise high-dimensional data. The t-SNE can be considered the best data dimensionality reduction and visualisation technique (Jiang et al., 2021). The t-SNE works by computing the probability p_{ij} of a given dataset by estimating conditional probability $p_{j|i}$, as:

$$p_{j|i} = \frac{e^{\left(\frac{-\|x_i - x_j\|^2}{2\sigma_i^2}\right)}}{\sum_{k \neq i} e^{\left(\frac{-\|x_i - x_k\|^2}{2\sigma_i^2}\right)}}, \tag{9}$$

where σ_i is the variance of Gaussian that is centred around x_i and $p_{i|i}$ is set to 0 for searching pairwise similarities. It should be emphasised that probability is 1, such that $\sum_j p_{j|i} = 1$ for all i. Hence, defining:

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N}, \tag{10}$$

because p_i and p_i of N-dimensional space is $1/N$, the conditional probability can be written as $p_{j|i} = Np_{ij}$ and $p_{i|j} = Np_{ji}$. Since $p_{ji} = p_{ij,}$ then eq.10 is proven. Using y_i and y_j of $x_i...x_j$, we calculate conditional probability $q_{j|i}$ as:

$$q_{j|i} = \frac{e^{-\|y_i - y_j\|^2}}{\sum_{k \neq i} e^{-\|y_i - y_k\|^2}}, \tag{11}$$

and also for modelling pairwise similarities, the $q_{i|i}$ is set to 0. Hence, the conditional probabilities $p_{j|i,}$ and $q_{j|i}$ will be equal for modelling data points into low-dimensional space from high-dimensional space. The measure of in which $q_{j|i}$ models $p_{j|i}$ is done by KL (Kullback-Leibler) divergence, such that the model minimises the sum of KL divergences over all data points using a gradient as:

$$\frac{\partial C}{\partial y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j)\left(1 + \|y_i - y_j\|^2\right)^{-1}. \tag{12}$$

The cost function is defined as C:

$$C = \sum_i KL(P_i \| Q_i) = \sum_i \sum_j p_{j|i} log \frac{p_{j|i}}{q_{j|i}}, \tag{13}$$

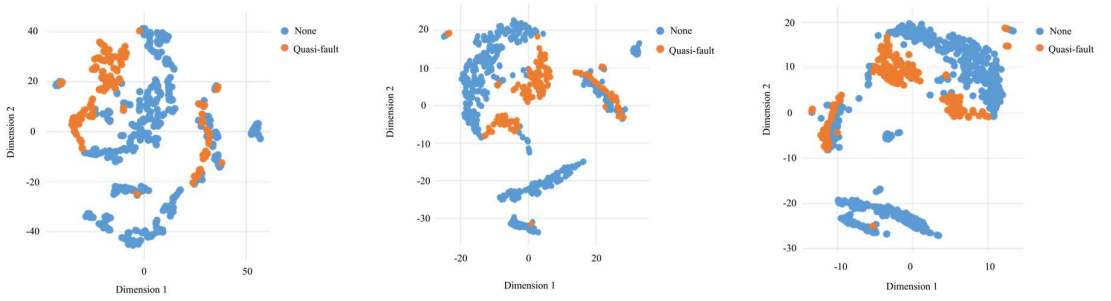
in which P_i is conditional probability over given data points x_i and Q_i is a conditional probability of given y_i. Finally, t-SNE performs a binary search of σ_i getting P_i with a defined perplexity. The perplexity is defined as:

$$Perp(P_i) = 2^{H(P_i)}, \tag{14}$$

and H represents the Shanon entropy of P_i:

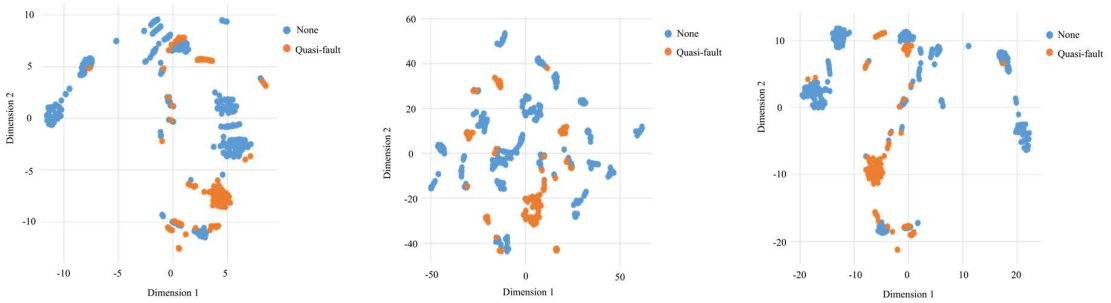
$$H(P_i) = -\sum_j p_{j|i} log_2 p_{j|i}. \tag{15}$$

The perplexity can be represented as a measure of the effective number of neighbours. Usually, perplexity is between 5 and 50 (Maaten and Hinton, 2008). Hence, the expected outcome of t-SNE is separation such that feature-containing information is more clustered with a good potent capability of separability (Tang et al., 2020). To represent the local geometry and global information preserved, we depict t-SNE results of PLS-DA (Fig. 14), RF (Fig. 15) and SVM (Fig. 16).



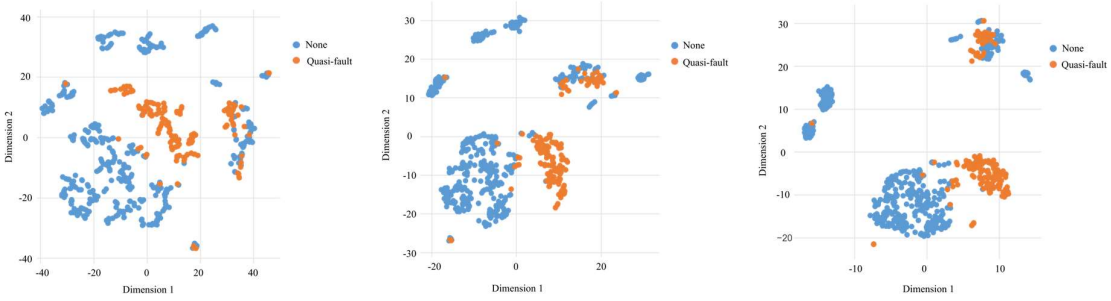
(a) (b) (c)

Fig. 14. t-SNE of PLS-DA with (a) 10 perplexity; (b) 25 perplexity; (c) 50 perplexity



(a) (b) (c)

Fig. 15. t-SNE of RF with (a) 10 perplexity; (b) 25 perplexity; (c) 50 perplexity



(a) (b) (c)

Fig. 16. t-SNE of SVM with (a) 10 perplexity; (b) 25 perplexity; (c) 50 perplexity

In the visualisation results of PLS-DA and RF-extracted features, the t-SNE shows misclassification increased clusters' density and inability to form clusters by the t-SNE. The PLS-DA shows slight signs of discrimination properties (Fig. 14), while RF features provide hard-to-understand information for clustering (Fig. 15). The SVM-selected features, in contrast to previous models, show clustering properties (Fig. 16c), where further multi-class labelling is needed (e.g., for diagnosis and fault detection).

3.4 Classification results using SVM features

From the RFE-FS method, the extracted SVM features are used for classification in the holdout sample. In addition, the current research on binary classification also suggests using LR (Logistic Regression) (Orošnjak, 2022) due to good prediction properties; the LR algorithm is also added. Although the evidence shows that SVM and LR classifiers have excellent prediction properties, the RF outperforms both models by almost 4% (Table 5) and visualisation by AUC also reflects such statements.

Table 5. Classification results of models using holdout datasets

LR classification results using the training dataset				LR classification results using holdout dataset			
	None	Quasi-fault	%		None	Quasi-fault	%
None	352	11	96.96	None	144	8	94.74
Quasi-fault	11	322	96.69	Quasi-fault	9	123	93.18
Overall			**96.83**				**93.96**
SVM classification results using the training dataset				SVM classification results using holdout dataset			
	None	Quasi-fault	%		None	Quasi-fault	%
None	353	16	95.66	None	143	8	94.71
Quasi-fault	10	317	96.94	Quasi-fault	10	123	92.48
Overall			**96.30**	Overall			**93.59**
PLS-DA classification results using the training dataset				PLS-DA classification results using holdout dataset			
	None	Quasi-fault	%		None	Quasi-fault	%
None	313	64	83.03	None	127	27	82.47
Quasi-fault	50	269	84.33	Quasi-fault	26	104	80.00
Overall			**83.68**	Overall			**81.23**
RF classification results using the training dataset				RF classification results using holdout dataset			
	None	Quasi-fault	%		None	Quasi-fault	%
None	359	8	97.82	None	152	5	96.82
Quasi-fault	4	325	98.78	Quasi-fault	1	126	99.21
Overall			**98.30**	Overall			**98.02**

The resulting PLS-DA model shows poor prediction properties in this particular case. By visualising obtained results from a binary classification using hydraulic power, we can see that >95% accuracy can be achieved with the SVM (Fig. 17left) and LR model (Fig. 17right) and >99% accuracy with the RF model (Fig. 18left); however, PLS-DA show poor prediction properties considering AUC obtained results (Fig. 18right).

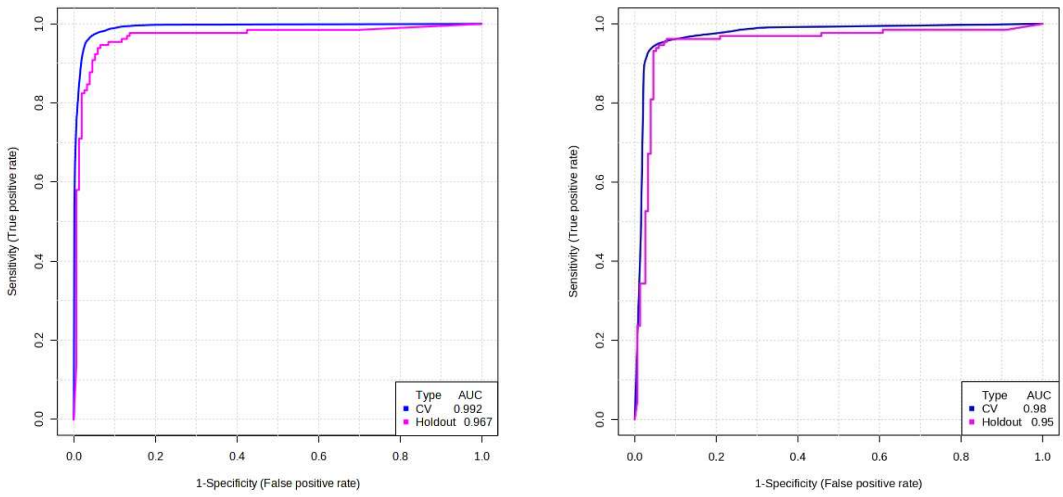


Fig. 17. SVM classifier results (left) and LR classifier results (right)

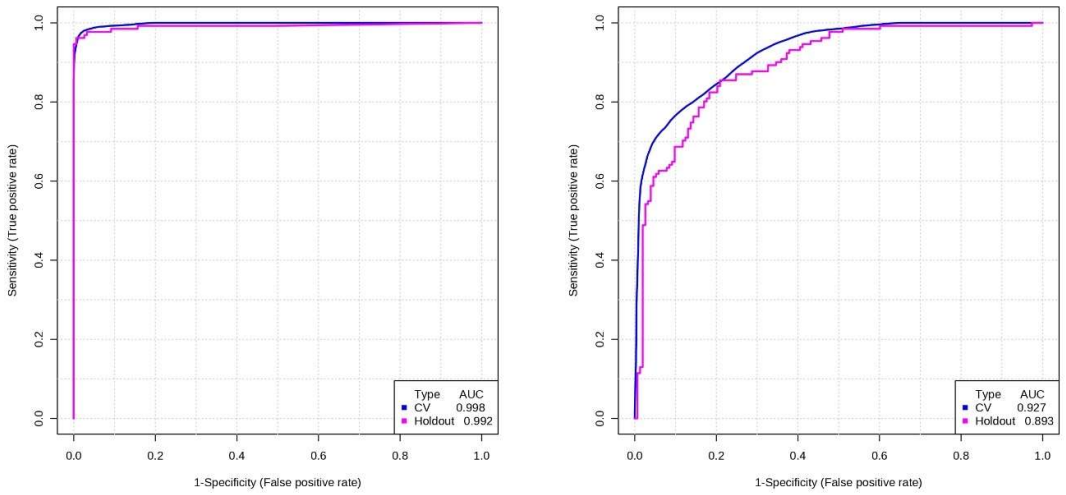


Fig. 18. RF classifier results (left) and PLS-DA classifier results (right)

The final idea is to use extracted features as **FP-markers.** Hence, if we consider the rise of elemental particles in the oil as „markers“ suggesting degradation of, e.g., the hydraulic pump, then we can consider features from a hydraulic power signal as a variable suggesting the loss of output power from the pump – suggesting internal leakage, external leakage, or wear of the pump observed by latent degradational effect.

3.5 Energy waste due to latent degradation

After the experimental period (49 days), we plotted the results of hydraulic power loss. The results show that the power output significantly reduces over time with the same power input—an average of 0.63% power loss/daily compared to the start of the experiment. More importantly, at the end of the experiment, the total degradation reached a 25.9% loss of hydraulic power, i.e., energy waste (Fig. 19).

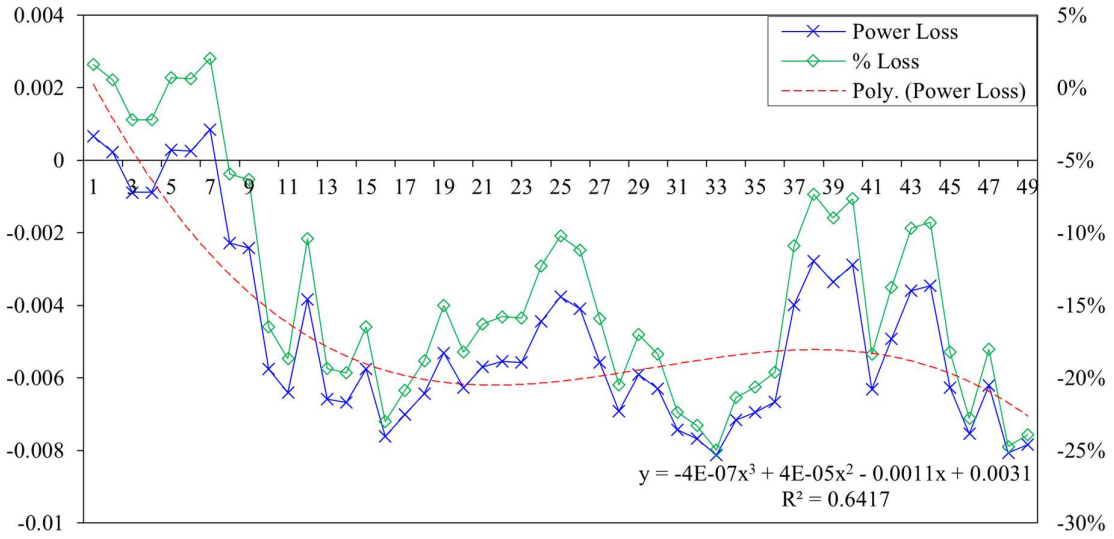


Fig. 19. Power loss (y_1-axis) and %loss (y_2-axis) per day (x-axis) during the experiment

Considering that approximately 358 hydraulic cycles were performed daily, a total of 96.38 kWh of energy are irreversibly wasted due to latent degradation of the hydraulic system, excluding the first week, which is an

approximation of 2.26±0.006 kWh of loss daily. To illustrate the „latent effect“ of degradation, we depict a hydraulic power with (Fig. 20 blue line) and without outliers (Fig. 20 red line) and show how degradation on each monitored cycle at the start and the end of the experiment.

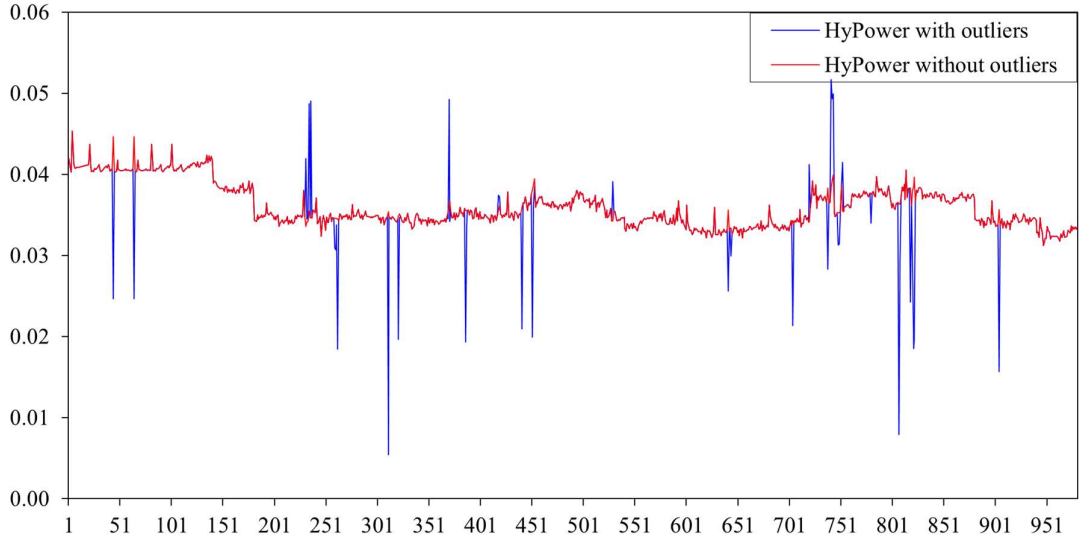


Fig. 20. Monitored hydraulic power loss (*y*-axis) with experimental cycles (*x*-axis)

The power output increase was noticed on the 24th and 36th day, respectively (Fig. 19). These two instances included maintenance activities of adjusting regulation valves, filter replacement and oil refilling. Significant energy waste was achieved by reducing the time for opening and closing the saddle. This can also be observed in Fig. 4 (December), where the hydraulic cycle was reduced from 20 to 18 seconds. The extensive degradation started after the replacement of parts, and potential leakage and wear of the hydraulic pump caused a stoppage of the RMM process. Although the study uses a binary classification problem for gaining insight into the system degradation using FP notation, further diagnosis needs to be conducted for associating „Quasi-faults“ with root causes of degradation.

4 Concluding remarks

For the first time, the study uses hydraulic power as a CM indicator for delineating the healthy from the non-healthy state. Using the RFE-FS strategy, we show that SVM features outperform other models and, in combination with RF for classification, can provide >95% accuracy given unseen data. The extracted features and classification highlighted that observed degradation caused 26% of energy waste. On the other side, applied LCM practice within the company failed to provide such insight, consequently wasting 0.63% of energy/daily. The results obtained from laboratory analysis (physio-chemical properties and WDXRF) fail to provide information about the degradation of the hydraulic system (e.g., wear, leakage). Therefore, through inexpensive sensors and machine learning, we show how cleaner production can be achieved using FP markers (i.e., extracted features) through functional productivity.

The contributions to the literature are seen through introduction of the functional-energy *p-f* curve as a new „maintenance construct“ that shifts the focus from waste to functional energy indicators. Emphasising that altering maintenance practice from PdM to EBM requires switching attention from secondary (waste) to primary

(functional) energy indicators are justified because the same indicator can be used for (1) predictions (diagnostics/prognostics); (2) optimisation (energy efficiency); and (3) operational efficiency, since energy can be transferred into monetary value. To do so, we conceptualise the functional-productiveness as a semi-unsupervised mechanism that labels degradation events as „Quasi-faults“.

We consider the ongoing and future research of the EBM and FP beneficial for upcoming EU-imposed policies and regulations for reducing energy waste. Therefore, we consider expanding and advancing the FP as an unsupervised labelling mechanism for deep learning networks where the faster reaction to changes in signal behaviour can be observed and reacted to without production stoppages and interventions.

References

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Chang, Y.-W., Lin, C.-J., 2008. Feature Ranking Using Linear SVM, in: JMLR: Workshop and Conference Proceedings. WCCI2008, pp. 53–64.

Chen, R.-C., Dewi, C., Huang, S.-W., Caraka, R.E., 2020. Selecting critical features for data classification based on machine learning methods. J. Big Data 7, 52. https://doi.org/10.1186/s40537-020-00327-4

Day, M., Bauer, C., 2007. Water Contamination in Hydraulic and Lube Systems. Mach. Lubr. 9.

Duch, W., Wieczorek, T., Biesiada, J., Blachnik, M., 2004. Comparison of feature ranking methods based on information entropy, in: 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541). IEEE, pp. 1415–1419. https://doi.org/10.1109/IJCNN.2004.1380157

Emec, S., Krüger, J., Seliger, G., 2016. Online Fault-monitoring in Machine Tools Based on Energy Consumption Analysis and Non-invasive Data Acquisition for Improved Resource-efficiency. Procedia CIRP 40, 236–243. https://doi.org/10.1016/j.procir.2016.01.111

González-Muñiz, A., Díaz, I., Cuadrado, A.A., García-Pérez, D., Pérez, D., 2022. Two-step residual-error based approach for anomaly detection in engineering systems using variational autoencoders. Comput. Electr. Eng. 101, 108065. https://doi.org/10.1016/j.compeleceng.2022.108065

Huang, K., Wu, S., Li, F., Yang, C., Gui, W., 2022. Fault Diagnosis of Hydraulic Systems Based on Deep Learning Model With Multirate Data Samples. IEEE Trans. Neural Networks Learn. Syst. 33, 6789–6801. https://doi.org/10.1109/TNNLS.2021.3083401

Huang, K., Wu, S., Li, Y., Yang, C., Gui, W., 2021. A multi-rate sampling data fusion method for fault diagnosis and its industrial applications. J. Process Control 104, 54–61. https://doi.org/10.1016/j.jprocont.2021.06.003

Hui, K.H., Ooi, C.S., Lim, M.H., Leong, M.S., Al-Obaidi, S.M., 2017. An improved wrapper-based feature selection method for machinery fault diagnosis. PLoS One 12, 1–10. https://doi.org/10.1371/journal.pone.0189143

Jiang, W., Li, Z., Zhang, Sheng, Wang, T., Zhang, Shuqing, 2021. Hydraulic Pump Fault Diagnosis Method Based on EWT Decomposition Denoising and Deep Learning on Cloud Platform. Shock Vib. 2021, 1–18. https://doi.org/10.1155/2021/6674351

Karanović, V. V, Jocanović, M.T., Wakiru, J.M., Orošnjak, M.D., 2018. Benefits of lubricant oil analysis for maintenance decision support: a case study. IOP Conf. Ser. Mater. Sci. Eng. 393, 012013. https://doi.org/10.1088/1757-899X/393/1/012013

Kaur, A., Malhotra, R., 2008. Application of random forest in predicting fault-prone classes. Proc. - 2008 Int. Conf. Adv. Comput. Theory Eng. ICACTE 2008 37–43. https://doi.org/10.1109/ICACTE.2008.204

Li, Z., 2018. Deep Learning Driven Approaches for Predictive Maintenance: A Framework of Intelligent Fault Diagnosis and Prognosis in the Industry 4.0 Era. Norwegian University of Science and Technology.

444 Liu, C., Jiang, D., Yang, W., 2014. Global geometric similarity scheme for feature selection in fault diagnosis.
445 Expert Syst. Appl. 41, 3585–3595. https://doi.org/10.1016/j.eswa.2013.11.037

446 Liu, X., Yang, X., Shao, F., Liu, W., Zhou, F., Hu, C., 2021. Composite Multi-Scale Basic Scale Entropy Based
447 on CEEMDAN and Its Application in Hydraulic Pump Fault Diagnosis. IEEE Access 9, 60564–60576.
448 https://doi.org/10.1109/ACCESS.2021.3074498

449 Lu, L., Yan, J., de Silva, C.W., 2015. Dominant feature selection for the fault diagnosis of rotary machines using
450 modified genetic algorithm and empirical mode decomposition. J. Sound Vib. 344, 464–483.
451 https://doi.org/10.1016/j.jsv.2015.01.037

452 Ma, X., Wang, P., Zhang, B., Sun, M., 2021. A Multirate Sensor Information Fusion Strategy for Multitask Fault
453 Diagnosis Based on Convolutional Neural Network. J. Sensors 2021, 1–17.
454 https://doi.org/10.1155/2021/9952450

455 Maaten, L. van der, Hinton, G., 2008. Visualizing Data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605.

456 MPFiltri, 2020. Fluid Condition and Filtration Handbook: Manual of analysis and comparison of photographs, 1st
457 ed. M PFiltri, Milano, Italy.

458 Orošnjak, M., 2022. The Concept of Functional-Productiveness for Modelling Reliability in Energy-Based
459 Maintenance Domain. University of Novi Sad.

460 Orošnjak, M., 2021. Maintenance Practice Performance Assessment of Hydraulic Machinery: West Balkan Meta-
461 Statistics and Energy-Based Maintenance Paradigm, in: 2021 5th International Conference on System
462 Reliability and Safety (ICSRS). IEEE, pp. 108–114. https://doi.org/10.1109/ICSRS53853.2021.9660739

463 Orošnjak, M., Jocanović, M., Čavić, M., Karanović, V., Penčić, M., 2021. Industrial maintenance 4(.0) Horizon
464 Europe: Consequences of the Iron Curtain and Energy-Based Maintenance. J. Clean. Prod. 314, 128034.
465 https://doi.org/10.1016/j.jclepro.2021.128034

466 Pei, M., Li, H., Yu, H., 2021. A Novel Three-stage Feature Fusion Methodology and its Application in
467 Degradation State Identification for Hydraulic Pumps. Meas. Sci. Rev. 21, 123–135.
468 https://doi.org/10.2478/msr-2021-0018

469 Perk, S., Teymour, F., Cinar, A., 2011. Adaptive agent-based system for process fault diagnosis. Ind. Eng. Chem.
470 Res. 50, 9138–9155. https://doi.org/10.1021/ie102058d

471 Pires, S.P., Sénéchal, O., Loures, E.F.R., Jimenez, J.F., 2016. An approach to the prioritization of sustainable
472 maintenance drivers in the TBL framework. IFAC-PapersOnLine 49, 150–155.
473 https://doi.org/10.1016/j.ifacol.2016.11.026

474 Remeseiro, B., Bolon-Canedo, V., 2019. A review of feature selection methods in medical applications. Comput.
475 Biol. Med. 112, 25–29. https://doi.org/10.1016/j.compbiomed.2019.103375

476 Ruiz-Perez, D., Guan, H., Madhivanan, P., Mathee, K., Narasimhan, G., 2020. So you think you can PLS-DA?
477 BMC Bioinformatics 21, 1–10. https://doi.org/10.1186/s12859-019-3310-7

478 Srinivas, H., 2022. Sustainable Development: Concepts - Cleaner Production [WWW Document]. Glob. Dev.

Res. Cent. URL https://www.gdrc.org/sustdev/concepts/02-c-prod.html# (accessed 12.22.22).

Tang, S., Yuan, S., Zhu, Y., 2020. Deep Learning-Based Intelligent Fault Diagnosis Methods Toward Rotating Machinery. IEEE Access 8, 9335–9346. https://doi.org/10.1109/ACCESS.2019.2963092

Tang, S., Zhu, Y., Yuan, S., 2022. An adaptive deep learning model towards fault diagnosis of hydraulic piston pump using pressure signal. Eng. Fail. Anal. 138, 106300. https://doi.org/10.1016/j.engfailanal.2022.106300

Tang, S., Zhu, Y., Yuan, S., 2021. An improved convolutional neural network with an adaptable learning rate towards multi-signal fault diagnosis of hydraulic piston pump. Adv. Eng. Informatics 50, 101406. https://doi.org/10.1016/j.aei.2021.101406

Ullah, I., Yang, F., Khan, R., Liu, L., Yang, H., Gao, B., Sun, K., 2017. Predictive maintenance of power substation equipment by infrared thermography using a machine-learning approach. Energies 10. https://doi.org/10.3390/en10121987

UNEP, 2006. Environmental Agreements and Cleaner Production [WWW Document]. URL https://www.unep.org/resources/report/environmental-agreements-and-cleaner-production (accessed 12.29.22).

Westerhuis, J.A., Hoefsloot, H.C.J., Smit, S., Vis, D.J., Smilde, A.K., Velzen, E.J.J., Duijnhoven, J.P.M., Dorsten, F.A., 2008. Assessment of PLSDA cross validation. Metabolomics 4, 81–89. https://doi.org/10.1007/s11306-007-0099-6

Xia, T., Si, G., Shi, G., Zhang, K., Xi, L., 2022. Optimal selective maintenance scheduling for series–parallel systems based on energy efficiency optimization. Appl. Energy 314, 118927. https://doi.org/10.1016/j.apenergy.2022.118927

Xia, T., Xi, L., Du, S., Xiao, L., Pan, E., 2018. Energy-Oriented Maintenance Decision-Making for Sustainable Manufacturing Based on Energy Saving Window. J. Manuf. Sci. Eng. 140. https://doi.org/10.1115/1.4038996

Xie, C., Li, Y., Ding, Y., Radcliffe, J., 2019. Evaluating Levelized Cost of Storage (LCOS) Based on Price Arbitrage Operations: with Liquid Air Energy Storage (LAES) as an Example. Energy Procedia 158, 4852–4860. https://doi.org/10.1016/j.egypro.2019.01.708

Yan, Z., Kuang, T.-H., Yao, Y., 2017. Multivariate fault isolation of batch processes via variable selection in partial least squares discriminant analysis. ISA Trans. 70, 389–399. https://doi.org/10.1016/j.isatra.2017.06.014

Yin, S., Yang, H., Xu, K., Zhu, C., Wang, Y., 2021. Location of abnormal energy consumption and optimization of energy efficiency of hydraulic press considering uncertainty. J. Clean. Prod. 294, 126213. https://doi.org/10.1016/j.jclepro.2021.126213

Yu, K., Han, H., Fu, Q., Ma, H., Zeng, J., 2020. Symmetric co-training based unsupervised domain adaptation approach for intelligent fault diagnosis of rolling bearing. Meas. Sci. Technol. 31, 115008. https://doi.org/10.1088/1361-6501/ab9841

Zenglin Xu, King, I., Lyu, M.R.-T., Rong Jin, 2010. Discriminative Semi-Supervised Feature Selection Via Manifold Regularization. IEEE Trans. Neural Networks 21, 1033–1047. https://doi.org/10.1109/TNN.2010.2047114

Zhou, B., Yi, Q., 2021. An energy-oriented maintenance policy under energy and quality constraints for a multielement-dependent degradation batch production system. J. Manuf. Syst. 59, 631–645. https://doi.org/10.1016/j.jmsy.2021.04.015

Appendices

A1. Response *Y* label by Functional-Productiveness

The $X^j_{i,d}$ is a dataset of *i* instances, *j* variables, and *d* matrices (days) [1, 49]. In our *FP* case, we use *j* variables {*HS_cycle_time, T1, T2, T3, T4, T5*, *HyPower_XS*}. The generated *FP* label dataset $Y^j_{i,d}$ matrix consists of instances *i* and variables *j*. Value in $X^j_{i,d} \neq Y^j_{i,d}$ because the value in Y^j_i is dummy $C^{\mathbb{R}}$, $\mathbb{R} \in [0, 1]$, i.e., class label [None, Quasi-fault]. The labels are semi-supervised where each $Y^j_{i,j}$ is inputted as 0 or 1, as:

$$Y_d^j = 0, for\ \forall i, i = 1,2 \dots 20\ if\ X_{i,d}^j \in \left[\overline{X_{d-1}^J} - 2\sigma_{d-1}^j, \overline{X_{d-1}^J} + 2\sigma_{d-1}^j\right] \tag{A.1}$$

and

$$Y_d^j = 1, for\ \exists i, i = 1,2 \dots 20\ if X_{i,d}^j \notin \left[\overline{X_{d-1}^J} - 2\sigma_{d-1}^j, \overline{X_{d-1}^J} + 2\sigma_{d-1}^j\right] \tag{A.2}$$

where $\overline{X_d^J}$ represents the average value of a variable *j* in a *d* dataset *X* as:

$$\overline{X_d^J} = \frac{1}{n}\sum_i^n X_{i,d}^j\,, \tag{A.3}$$

and σ^j_d is the standard deviation of a variable *j* in a *d* dataset:

$$\sigma_d^j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\overline{X_d^J} - X_{i,d}^j\right)^2}, \tag{A.4}$$

where *d-l* is used for setting thresholds, i.e., upper and bottom control limits of a functional productiveness where $l = 1$, restricting to the previous *d* sequence. Secondly, we assume that at the start of the experiment, each *j* variable follows ℵ distribution with parameters (μ, σ^2), such that $X^j_d \sim \aleph(\mu, \sigma^2)$. The „Quasi-fault" event $Y^j_{i,d} = 1$ is for any event that breaks the threshold set in eq.A2.

Considering that outliers can be total failure events, such instances *i* are removed from eq.A1-A4, and after removing the *i = n-1* instances, the new FP threshold is set. The outliers removal is done using *Grubb's* test for ≤ 2 number of outliers. More importantly, since there is a significant chance that labels can be marked as false positives due to random events, we introduce the second penalisation performed by the researcher. The penalisation is performed in eq.A2 when $Y^j_d = 1$ if the $X^j_{i,d}$ value escapes the FP ~ 95% threshold $\left[\overline{X_{d-1}^J} \pm 2\sigma_{d-1}^j\right]$, but does not correlate with changes in system response. For instance, the length of the *HyPower_XS* sequence should correlate with *HS_cycle_time* since the length of the sequence window measures the cycle time. Specifically, *N_Mean_OS* is the feature explaining the average power of the sequence being transferred to the actuators that operate the speeds *T1* and *T2*. The *T3* as the time between saddle positions should correlate with *HyPower_IS and HS_cycle_time*. Finally, the *HyPower_CS* should correlate with *T4* and *T5*, which are the speeds for fast and slow closing of the saddle by actuators. Reasonably, the pump's power (pressure and flow) is in a serial relationship with the actuators' movement force and speed – changes in the hydraulic power from the pump cause changes in the force and speed of the actuator.