Heliyon 11 (2025) e41251

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Condition monitoring and multi-fault classification of hydraulic systems using multivariate functional data analysis [☆]

Cevahir Yildirim [a,*], Alba M. Franco-Pereira [b,c], Rosa E. Lillo [a,d]

[a] uc3m - Santander Big Data Institute (IBiDat), Spain
[b] Interdisciplinary Mathematics Institute (IMI), UCM, Spain
[c] Department of Statistics, UCM, Spain
[d] Department of Statistics, uc3m, Spain

ARTICLE INFO

Keywords:
Functional data analysis (FDA)
Multivariate functional principal component analysis (MFPCA)
Fault diagnosis and classification
Condition monitoring
Hydraulic systems

ABSTRACT

Condition monitoring and fault classification in engineering systems is a critical challenge within the scope of Prognostics and Health Management (PHM). The fault diagnosis of complex nonlinear systems, such as hydraulic systems, has become increasingly important due to advancements in big data analytics, machine learning (ML), Industry 4.0, and Internet of Things (IoT) applications. Multi-sensor data provides opportunities to predict component conditions; however, environments characterized by multiple sensors and diverse fault states across various components complicate the fault classification process. To address these challenges, this study introduces a novel multivariate Functional Data Analysis (FDA) framework based on Multivariate Functional Principal Component Analysis (MFPCA) for classifying failure conditions in hydraulic systems. The proposed method systematically tackles condition-based diagnostics and addresses fundamental issues in multi-fault classification. Experimental results demonstrate that this approach achieves high classification accuracy using raw multi-sensor data, establishing multivariate FDA as a powerful tool for fault diagnosis in complex systems.

1. Introduction

Fault classification and condition monitoring of engineering systems is one of the major areas of interest in “Prognostics and Health Management (PHM)” studies. Hydraulic systems are among the most commonly used engineering systems in various industries. Implementing condition monitoring for hydraulic systems provides multiple benefits, including increased productivity, reduced maintenance costs, minimized downtime, and enhanced reliability and safety in a variety of operational contexts [32]. The frequent occurrence of malfunctions and the significant costs associated with shutdowns make it impractical to predict machine damage using manual condition monitoring systems [3]. Moreover, fault detection, real-time condition monitoring, and predictive maintenance of hydraulic systems have become increasingly important in recent years [15]. In general, there are two approaches to condition monitoring: the model-based approach, which requires a comprehensive understanding of the system’s physical and mathematical

[☆] This research is part of the I+D+i projects PDC2022-133359, PID2022-137243OB-I00, PID2022-137050NB-I00 and TED2021-131264B-100 funded by Ministerio de Ciencia, Innovación y Universidades/AEI/10.13039/501100011033 and European Union NextGenerationEU/PRTR. This initiative has also been partially carried out within the framework of the Recovery, Transformation and Resilience Plan funds, financed by the European Union (Next Generation) through the grant ANTICIPA and the ENIA 2022 Chairs for the creation of university-industry chairs in AI-AImpulsa: UC3M-Universia.

[*] Corresponding author.
E-mail addresses: cevahir.yildirim@alumnos.uc3m.es (C. Yildirim), albfranc@ucm.es (A.M. Franco-Pereira), rosaelvira.lillo@uc3m.es (R.E. Lillo).

https://doi.org/10.1016/j.heliyon.2024.e41251
Received 7 October 2024; Received in revised form 3 December 2024; Accepted 13 December 2024

Available online 17 December 2024

behavior, but obtaining such detailed information can be challenging for complex systems. The second approach is the statistical approach, which relies on the analysis of past failures observed during monitoring and requires a significant amount of historical data. Real-time fault diagnosis and condition monitoring of hydraulic systems are not new areas of investigation. Vibration analysis stands out as a widely adopted and effective method for monitoring the condition of systems, including rotating components such as hydraulic pumps, bearings, electric motors, and others [39,35]. However, other measures, such as temperature, flow rates, power, and efficiency metrics, can also be important for understanding equipment condition. Many theoretical and practical studies have been conducted, and different Machine Learning (ML) and Internet of Things (IoT) approaches have been applied over the years for PHM. The basic idea is to explore the degradation patterns of different hydraulic components and investigate the correlation between degradation patterns and equipment health conditions under varying operational conditions [24]. By correctly identifying and addressing potential failures, damages can be mitigated, maintenance costs minimized, and productivity increased [17]. As a result, PHM is a growing field that is vital for the application of predictive maintenance in complex production systems.

Today, in-depth research is carried out on single components and single failure types in hydraulic systems. However, simultaneous failures of different degrees of severity are common in complex hydraulic systems and are rarely investigated in detail. In the related literature, there are several approaches to monitor the condition of hydraulic systems using multivariate statistics, such as artificial neural networks (ANN), support vector machines (SVM), decision trees, and semantic-statistical methods. In their study, Helwig et al. [15] examined complex hydraulic systems, investigating the relationship between features extracted from the sensor data and various failure types. Using Linear Discriminant Analysis (LDA), they projected high-dimensional features into a low-dimensional discriminant space to facilitate classification of complex system fault conditions. Using multivariate statistical analysis, they identified key features associated with a failure scenario, drawing on known failure characteristics from experimental data. Chawathe [6] introduces an intelligent system designed to precisely evaluate the condition of hydraulic systems by analyzing continuous real-time streams of sensor data. The rise of big data technology, data mining methods, advanced machine learning algorithms, and industrial IoT platforms has made fault diagnosis and condition monitoring increasingly fascinating. Lei et al. [38] focus on exploring research opportunities within this domain. Xu [37], Lei [23], Zhao et al. [40], and Peng [29] have investigated various fault detection methods using traditional machine learning models to detect faults in hydraulic systems. When taking into consideration the nonlinear characteristics of signals captured in multi-sensor environments, Wang et al. [36] introduced an innovative diagnostic approach for hydraulic systems. Rooted in multi-source information fusion and fractal dimension analysis, their method achieves fault pattern recognition through the combination of the k-nearest neighbor technique and fuzzy clustering approach.

The techniques mentioned above are commonly known as shallow machine learning (SML) methods. Typically, SML methods require initial preprocessing of system monitoring data, followed by the extraction and selection of statistical features that capture system characteristics. Then, fault diagnosis is performed by creating a diagnostic model based on the determined features [17]. Condition monitoring and fault diagnostics of hydraulic systems present many challenges. The biggest issue is understanding which sensor data is correlated with which equipment failure in the system. Feature extraction techniques have been developed to overcome this problem and are used in the preprocessing step in many studies. To determine the correlation between sensors and system components, the first four moments of the data distribution—namely, mean, standard deviation, skewness, and kurtosis—are usually extracted [3,6,15,33,37]. Then, the most-correlated sensor is considered. However, considering only the information from one sensor and ignoring the data from all other sensors may lead to the loss of information from the other sensors. Furthermore, manual feature extraction relies heavily on expert knowledge, making the selection of the most sensitive features in various diagnostic scenarios subjective and time-consuming, especially in cases where expertise is limited [5]. Additionally, since feature extraction schemes are tailored to specific monitoring systems, they need to be redesigned when applied to a new system. Another challenge arises from the varying sampling rates of sensor data collected from the system, which, combined with complex coupling interactions between components, makes it difficult to collect consistent data. Huang et al. [17] and Kim and Jeong [21] address this problem by using a series of deep learning models to analyze sensor data from hydraulic systems. Another challenge is the performance of classifying system faults, as many components may fail, and it is crucial to identify which component is associated with the fault and when it will fail. Many studies have been conducted to classify fault conditions of hydraulic systems, and different metrics such as classification accuracy, precision, recall, and the harmonic mean of precision and recall (called the F1 score) are used to evaluate classification performance [37].

In summary, multiple fault types from different system components need to be classified with minimal loss of information using multi-sensor data, which may have different sampling rates. In this study, a new Functional Data Analysis (FDA) approach is proposed to overcome the above-mentioned challenges in condition monitoring and fault detection of hydraulic systems, as well as effectively classify fault conditions. More precisely, this work aims to contribute in four aspects:

1. A novel Multivariate Functional Data Analysis (FDA) approach based on Multivariate Functional Principal Component Analysis (MFPCA) is proposed for fault classification and condition monitoring of hydraulic systems.
2. The natural structure of the FDA approach allows working with sensors that have different sampling frequencies.
3. Instead of using feature extraction techniques for the most relevant sensor selection, direct use of raw multi-sensor data in a multivariate domain is achieved via MFPCA.
4. Fault classification is done effectively, and the classification performance is competitive compared to other studies in the literature.

The subsequent sections of this article are structured as follows. In Section 2, preliminary information is provided. In Section 3, the experimental dataset is explained. In Section 4, the structure and the notation of the proposed MFPCA-based fault classification approach are explained. In Section 5, the results of the proposed method are presented, and the results are interpreted by comparing them with other studies in the literature. The conclusions are presented in Section 6.

2. Preliminaries

Functional data analysis (FDA) has grown rapidly in recent years, presenting challenges in providing a precise definition due to its wide range of applications and tools. Essentially, FDA can be considered when the variable or unit of interest in a data set is inherently represented as a smooth curve or function. FDA deals with the statistical analysis of collections of curves [22]. The data explained in Section II can be considered multivariate functional data. Therefore, it is necessary to introduce the essential tools of FDA.

Preliminary information and the basics of the useful tools for applying the FDA approach are provided in this section. Smoothing with B-splines, Multivariate Functional Principal Component Analysis (MFPCA) [4], and the Modified Epigraph Index (MEI) are briefly explained and will be used later in the following sections of this paper. Since the sensor data is recorded at the same time intervals but has different scales for each sensor type, the common practice of normalizing the data has been applied before applying MFPCA.

2.1. Basis representation and smoothing

The first step in FDA is reconstructing the sample in a functional from the corresponding discrete observations. The common approach typically involves assuming an expansion of each sample curve in terms of a basis of functions and subsequently fitting the coefficients of the basis using smoothing techniques. [1]. Let $X_i(t), i = 1, \dots, n$ be a sample function generated by a process $X(t)$. In practice, sample functions are observed in a finite set of time points $(t_{i,0}, t_{i,1}, \dots, t_{i,T_i} \in T)$ for all $i = 1, \dots, n$. Then, the sample information is given by the vectors, $x_i = (x_{i,0}, \dots, x_{i,T_i})$, where $x_{i,0}$ is the ith hydraulic system's sensor observation value at time point 0 and x_{i,T_i} is the ith system's sensor observation value at failure time ($i = 1, \dots, n$). In this section, we assume that the sample paths belong to a finite-dimensional space generated by a basis $\phi_1(t), \dots, \phi_B(t)$, allowing them to be expressed as

$$X_i(t) = \sum_{b=1}^{B} c_{ib}\phi_b(t), \quad i = 1, \dots, n, \tag{1}$$

where c_{ib} are the coefficients of the basis functions for the i^{th} individual observation set and the b^{th} basis function with B number of basis functions. The function $X_i(t)$ is a linear combination of $\phi_1(t), \dots, \phi_B(t) : t \in T$, and it is called a functional data [31]. The selection of the basis and its dimension B is crucial and should be tailored according to the characteristics of the curves. We have used a B-Spline basis of order 4 (cubic splines), which generates a space of splines of the same degree, with each segment having its own cubic polynomial function. Cubic B-Splines are the most popular B-Splines used in practice because they guarantee that the first and second derivatives will be continuous. According to Šulejic [41], a nth-degree B-Spline curve (order of $n+1$) is defined by

$$C(t) = \sum_{k=1}^{\infty} B_{k,n}(t)P_k, \tag{2}$$

where P_k are called control points. The B-Spline curve $C(t)$ is constructed from nth-degree basis functions $B_{k,n}(t)$ defined by recurrence on the inferior degree.

We consider the vector known as the knot vector to be defined as $T = (t_0, \dots, t_m)$ where T is a non-decreasing sequence of real numbers with $t_k \leq t_{k+1}$ and $k = 0, \dots, m-1$. Each t_k is called knot. The kth B-Spline basis function of the nth degree is defined by the Cox-de Boor recursion formula [7]:

$$B_{k,0}(t) = \begin{cases} 1, & \text{if} \quad t_k \leq t \leq t_{k+1}, \\ 0, & \text{otherwise}, \end{cases} \tag{3}$$

with

$$B_{k,n}(t) = \frac{t - t_k}{t_{k+n} - t_k} B_{k,n-1}(t) + \frac{t_{k+n+1} - t}{t_{k+n+1} - t_{k+1}} B_{k+1,n-1}(t) \tag{4}$$

Typically, the knots of a B-Spline are equally spaced, and their number is chosen to be sufficiently large to effectively fit the data while avoiding unnecessary computational overhead.

Regarding smoothing, another option would have been to use P-splines. Introduced by Eilers and Marx [8], P-splines offer a flexible and powerful smoothing tool by combining regression on B-splines with discrete roughness penalties, making them suitable for handling non-normal data and various applications, including regression on signals and varying-coefficient models [9]. B-splines and P-splines are both powerful tools in a variety of applications. P-splines may be a better option when there is a lot of noise in the data or when more flexibility in curve smoothing is needed, as they help control smoothness and prevent overfitting, especially in small data sets. Our choice was B-splines because they are known for their numerical stability and fit quality, making them preferable for smooth interpolation. Additionally, the data studied in this paper are not noisy. The adequacy of the selected base and nodes will also be evaluated in detail in Section 5.6.

2.2. Modified epigraph index (MEI)

Since there is no natural order between functions, a fundamental task in functional data analysis is to provide a way to order curves, which makes it possible to define ranks and L-statistics. A depth definition for functional observations is formulated based on the concepts of hypograph and epigraph of a curve [26]. This functional depth offers a criterion for ordering the sample of curves from center outward. Alternative top-down or bottom-up orders can be obtained via the use of two statistical concepts: the Epigraph Index (EI) Hypograph Index (HI) can be used to sort the functions with an order. In this context, the EI can be calculated for each curve as,

$$EI(X(t)) = \frac{1}{N}\sum_{i=1}^{N} I(G(X_i(t)) \subset epi(X(t))), \tag{5}$$

so,

$$EI(X(t)) = \frac{1}{N}\sum_{i=1}^{N} I(X_i(t) \geq X(t)), \text{for each} \quad t \in I, \tag{6}$$

where $G(X_i(t))$ is the graph of $X_i(t)$, and $epi(X(t))$ is the epigraph of $X(t)$.

A modified version of the EI, which is less restrictive than the previously described definition, can be employed for the analysis of irregular (non-smooth) curves that may frequently cross over [11]. The modified epigraph index (MEI) is defined by,

$$MEI(X(t)) = \frac{1}{N}\sum_{i=1}^{N} \hat{\lambda}(X(t) \leq X_i(t)), \tag{7}$$

where $\hat{\lambda}(A)$ is the normalized Lebesgue measure over $I = [a, b]$, that is $\hat{\lambda}(A) = \lambda(A)/(b-a)$. The MEI function $MEI(X(t))$ yields a vector comprising the MEI values for every element within the functional dataset. These indexes have already been used for outlier detection [27], to design homogeneity tests [10] and to conduct clustering functional data [30]. In this work, MEI is used to rank the sample curves.

2.3. MFPCA

Principal Component Analysis (PCA) serves as an effective method for reducing dimensionality in data analysis. It is a multivariate statistical technique that concentrates multiple linearly related variables into a smaller set of uncorrelated variables. Initially proposed by Pearson [28] in a study focusing on optimal linear and plane fitting of spatial data, PCA was refined by Hotelling [16] and subsequently developed into a widely adopted method used in data dimensionality reduction, fault diagnosis, and anomaly detection. PCA provides valuable insight into the structure of variability within the variance-covariance operator for one-dimensional functional data [12]. The main goal of principal component analysis in a multivariate environment is to reduce the dimension of a data set consisting of a large number of correlated variables while retaining as much of the variation present in the data set as possible [13]. This is accomplished by transforming the original variables into a new set of variables known as the principal components. These components are uncorrelated and ordered such that the first few retain the majority of the variation present in all the original variables. It is very well known in statistics that if we are observing a j-dimensional random vector $\mathbf{X} = (X_1, X_2, \dots, X_j)' \in \mathbb{R}^j$ In the first step we look for a linear combination $U_1 = u_{11}X_1 + u_{12}X_2 + \dots + u_{1j}X_j = u_1'X$ of the elements of vector X having maximum variance. The variable U_1 is called the first principal component. Next, we look for a linear combination $U_2 = u_2'X$, uncorrelated with the first principal component having a maximum variance, and so on. Furthermore, at the kth stage, a linear combination $U_k = u_k'X$ called the kth principal component has maximum variance among those that one uncorrelated with the first $k-1$ principal components [19].

Let us suppose we have a sample of curves representing p different variables measured in n individuals. Therefore, we can represent the data with FPCA in a multivariate sense. Now, in the multivariate case, we have $\mathbf{X(t)} = (X_1(t), X_2(t), \dots, X_p(t))'$ with $t \in I$, and assume that $\mathbf{X} \in L_j^2(I)$ is a Hilbert space of square-integrable functions on the interval I, and $j = 1, \dots, p$. For $t, u \in I$, we define the covariance matrix $C(t, u)$ in the jth domain as

$$C_j(t,u) = Cov\big(X_j(t), X_j(u)\big), \quad j = 1, \dots, p \quad t, u \in I. \tag{8}$$

As noted in [31], a suitable inner product is the basis of all approaches for principal component analysis. For functions $X = (X_1, \dots, X_p)'$ with $X_p \in L_j^2(I)$ and $t \in I$, and If the elements vary considerably in domain, range or variation a weight vector $\omega, \dots, \omega_j$ can be supplied and the MFPCA is based on the weighted scalar product can be expressed as [14]

$$< X, Y >_{\omega} = \sum_{j=1}^{p} \omega_j \int_{I_j} X_j(t) Y_j(t) dt, \quad j = 1, \dots, p \quad t \in I, \tag{9}$$

where the covariance operator of $X(t)$ is a positive auto-adjoint compact operator defined by

$$C\big(X\big)_j(t) = \sum_{j=1}^{p} \int_{I_j} C_j(t,u)X_j(u)du, \quad j=1,\dots,p \quad t \in I, \tag{10}$$

then, the spectral representation of C provides the following Karhunen–Loeve orthogonal expansion [20,25] which is the orthogonal decomposition of the process

$$X_j(t) = \mu_j(t) + \sum_{k=1}^{\infty} f_{jk}(t)\xi_{jk}, \quad j=0,\dots,p \quad t \in I \tag{11}$$

where $f_{jk}(t)$ are the orthonormal family of eigenfunctions of the covariance operator C associated with its decreasing sequence of non null eigenvalues λ_k that is

$$C_j\big(f_j(t)\big) = \int_{I_j} C_j(t,u)f_j(u)du = \lambda_{jk}f_{jk}(t), \quad t \in I. \tag{12}$$

Similarly, f_{jk} is the kth principal weight function for the jth domain. Considering that the total variance of $X_j(t)$ is expressed as

$$V_j = \int_{I_j} C_j(t,t)dt = \sum_{k=1}^{\infty} \lambda_{jk}, \quad j=1,\dots,p, \tag{13}$$

where the ratio λ_{jk}/V_j represents the variation explained by the kth functional principal component. Therefore the process admits the following multivariate functional principal component reconstruction in terms of the first q principal components so that the total variance explained by them is expected to be one.

$$X_{ij}^{q}(t) = \mu_j(t) + \sum_{k=1}^{q} f_{jk}(t)\xi_{ijk}, \quad t \in I, \tag{14}$$

where ξ_{ijk} is defined as the family of uncorrelated zero-mean random variables, which can be defined in the multivariate sense by

$$\xi_{ijk} = \int_{I} f_{ijk}\big(X_{ij}(t) - \mu_j(t)\big)dt, \quad t \in I. \tag{15}$$

The random variable ξ_{ijk} is called the kth multivariate functional principal component and has the maximum variance λ_k out of all the generalized linear combinations of $X_{ij}(t)$ which is the function of the ith individual at jth domain.

3. Hydraulic system experimental dataset

Experimental data were obtained from a hydraulic testing system [15] capable of inducing reversible changes in the state or condition of various system components. Using this real test rig, sensor data were collected and buffered in real-time on a Programmable Logic Controller (PLC) (Beckhoff CX5020) before being transferred to a PC via EtherCAT for storage and subsequent analysis. Fault characterization measurements were configured using a custom-developed graphical user interface (GUI) built in LabVIEW, with the execution of these measurements handled by the PLC. All sensors employed in the system were equipped with standard industrial 20 mA current loop interfaces and connected to the data acquisition system. The test system underwent several hundred working cycles, during which various fault conditions including different fault types and severity levels were simulated in all possible combinations. These simulations accounted for the varying time scales of the anticipated effects, ensuring a comprehensive dataset. However, states related to transitions in oil temperature were excluded from the training data to minimize external variability.

The system under consideration consists of both a working cycle and a cooling and filtration cycle. It includes two interconnected cycles, as shown in Fig. 1. The primary cycle includes components such as a main pump (MP1), four accumulators (A1–A4), a filter (F1), and a set of valves. The secondary cycle, responsible for cooling and filtration, consists of a hydraulic pump (SP1), a filter (F2), a cooler (C1), and a valve connecting the pump to the cooling-filtration units. A variety of sensors (see Table 1) are placed in the system to track process measures such as volume flows, temperatures, and pressures at different points during the repeated testing load cycles (60 seconds). In total, 17 sensors are observed, including six pressure sensors (PS1–PS6), two flow rate sensors (FS1, FS2), four temperature sensors (TS1–TS4), one motor power sensor (EPS1), one vibration sensor (VS1), and three virtual sensors. The Cooling Efficiency Sensor (CE) is formulated using TS3, TS4, and T$_{amb}$; the Cooling Power Sensor (CP) is formulated with TS3, TS4, FS2, EPS1, FS1, PS2; and the System Efficiency Sensor (SE) is obtained from the values of EPS1, FS1, and FS2. These sensors collect data as the system goes through predefined fixed operating cycles with varying conditions of the components. The sampling rate of each sensor varies between 1 Hz and 100 Hz, with specific rates given in Table 1.

The failure behavior of the four main components (cooler, valve, pump, and accumulator) varies during the sixty-second working cycles. A total of 2205 simulation cycles are completed with different fault conditions, and the corresponding samples collected for each component of the system are given in Table 2. The “Fault Classification” studied in the given system is performed for four components: Cooler (C1), Valve (V10), Main Pump (MP1), and Accumulators (A1–A4) (see Fig. 1). In addition, 1449 data samples were obtained under stable conditions, while 756 samples were recorded during periods when static conditions were not yet fully

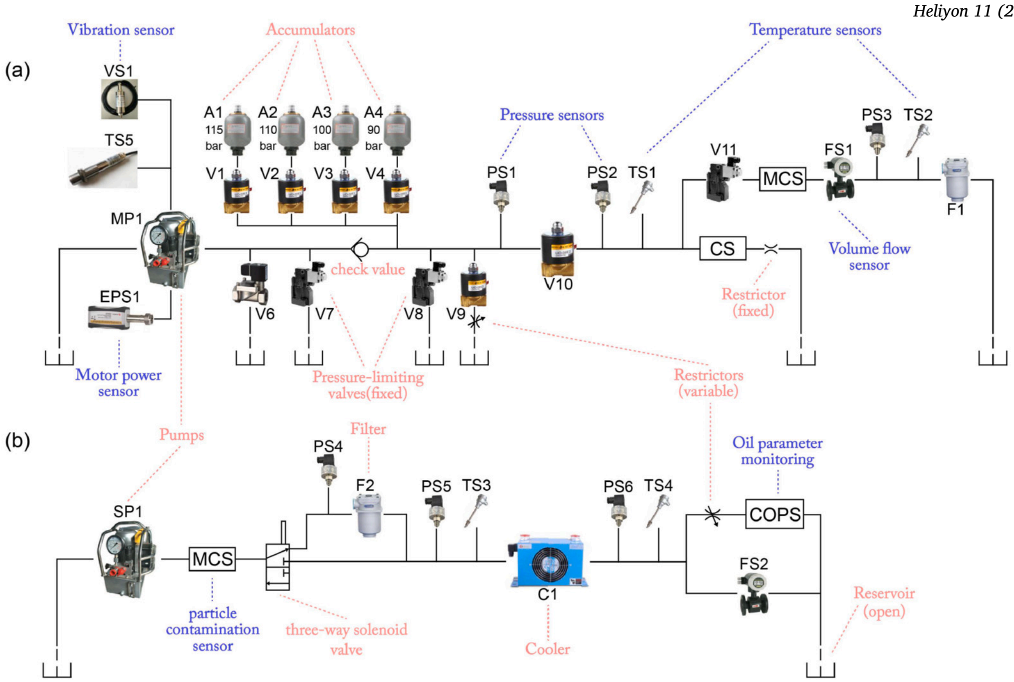

Fig. 1. Hydraulic System Structure: a-working circuit b-cooling and filtration circuit. (The figure is adapted from the data set [15]).

Table 1
Hydraulic System Sensor Details.

Sensor Name	(Sampling Rate)	Sensor Type (Units)
PS1	(100 Hz)	Pressure (bar)
PS2	(100 Hz)	Pressure (bar)
PS3	(100 Hz)	Pressure (bar)
PS4	(100 Hz)	Pressure (bar)
PS5	(100 Hz)	Pressure (bar)
PS6	(100 Hz)	Pressure (bar)
EPS1	(100 Hz)	Motor Power (W)
FS1	(10 Hz)	Volume Flow (l/min)
FS2	(10 Hz)	Volume Flow (l/min)
TS1	(1 Hz)	Temperature (°C)
TS2	(1 Hz)	Temperature (°C)
TS3	(1 Hz)	Temperature (°C)
TS4	(1 Hz)	Temperature (°C)
VS1	(1 Hz)	Vibration (mm/s)
CE	(1 Hz)	Cooling Efficiency [virtual] (%)
CP	(1 Hz)	Cooling Power [virtual] (kW)
SE	(1 Hz)	System Efficiency [virtual] (%)

established. While some studies [37] only take into consideration the stable working cycles, in order to utilize all the available information, we have considered both conditions (2205 samples).

4. Proposed method structure and notation

The proposed method provides a systematic machine learning approach that learns from multivariate sensor data to classify multiple fault types of an engineering system. The proposed model structure, as shown in Fig. 2, includes smoothing for multivariate functional data, obtaining derivatives of functional data, calculating modified epigraph indexes (MEIs), and finding correlations between MEIs and component failure conditions. MFPCA is performed by using these correlations as weights, and classification is carried out via Random Forest (RF) after the data is divided into training and test sets. The MFPCA scores of the original function and the 1st and 2nd derivative functions are used as independent variables in the classification step.

The matrix in Table 3 represents the functions and vectors for each hydraulic system and sensor observation set for the given dataset (see Section 3). It shows that the available data can be modeled as an n-dimensional sample of a multivariate functional dataset of dimension J. Mathematically, the available data can be represented in a matrix where each cell (i, j) collects the evolution of sensor j in system i, $i = 1, \dots, n$, and $j = 1, \dots, J$. We denote those functions as $X_{ij}(t)$. Now, in practice, the function $X_{ij}(t)$ is observed at discrete instants of time $(t_{i,1}, \dots, t_{i,60})$ for 60 second of working cycle defined in the experimental test set [15].

Table 2
Fault Conditions for each Hydraulic System Components.

Component	Value	Condition	Samples
Cooler (C1)	3%	total failure	732
	20%	reduced efficiency	732
	100%	full efficiency	741
Valve (V10)	73%	total failure	360
	80%	severe lag	360
	90%	small lag	360
	100%	optimal behavior	1125
Pump (MP1)	0	no leakage	1221
	1	weak leakage	492
	2	severe leakage	492
Accumulators (A1-A4)	90 bar	total failure	808
	100 bar	severely red. pressure	399
	115 bar	slightly red. pressure	399
	130 bar	full efficiency	599

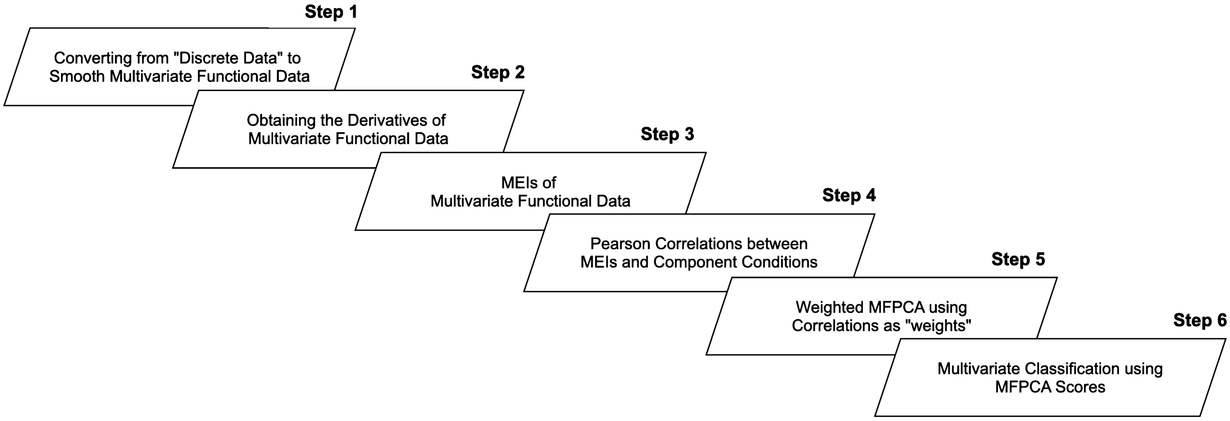
Step 1
Converting from "Discrete Data" to Smooth Multivariate Functional Data
Step 2
Obtaining the Derivatives of Multivariate Functional Data
Step 3
MEIs of Multivariate Functional Data
Step 4
Pearson Correlations between MEIs and Component Conditions
Step 5
Weighted MFPCA using Correlations as "weights"
Step 6
Multivariate Classification using MFPCA Scores

Fig. 2. The six approach for classifying the failures observed in the system using information from all the sensors.

Table 3
Functions and vectors for each hydraulic system observation set.

	sensor 1	…	**sensor J**
Hyd. Sys.1	$X_{11}(t)$ $X_{11}(t_{1,1})$, … , $X_{11}(t_{1,60})$	…	$X_{1j}(t)$ $X_{1j}(t_{1,1})$, … , $X_{1j}(t_{1,60})$
⋮	⋮	⋱	⋮
Hyd. Sys.n	$X_{n1}(t)$ $X_{n1}(t_{n,1})$, … , $X_{n1}(t_{n,60})$	…	$X_{nj}(t)$ $X_{nj}(t_{n,1})$, … , $X_{nj}(t_{n,60})$

5. Analysis, results and discussions

The steps of the proposed FDA approach (see Fig. 2) are applied to the dataset in this section. Each step is presented under a separate subheading. Multi-fault classification is performed, and the results are interpreted. Graphs and tables are provided to explain the proposed technique. A comparison of the results is made with other studies in the literature using the same dataset. The advantages of the proposed method are highlighted, addressing the challenges of the multi-fault classification problem discussed in Section 1.

5.1. Converting from “discrete data” to “smooth multivariate functional data”

The initial step of the proposed approach is to perform smoothing on multivariate sensor observations to obtain smooth multivariate functional data. While smoothing is a pre-processing step to transform discrete data into smooth functional curves before applying MFPCA, it also addresses the issue of varying sampling frequencies across different sensors, which is a common problem in multi-sensor environments with sensors operating at different frequencies (see Table 1). Similar to many financial and meteorological datasets, the hydraulic system data in this study is recorded at discrete time points. Let x_k denote the observed value of process $X(t)$ at the kth time point T_k, where T is a compact set such that $t_k \in T$ for $k = 1, \dots, K$. Thus, our data consists of K pairs (x_k, t_k). These

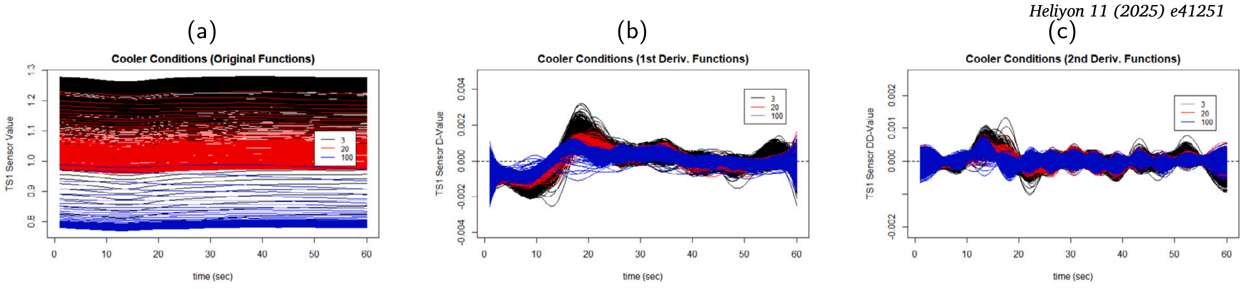

Fig. 3. Sensor “TS1” functions and “Cooler” fault conditions (different colors represent the different fault conditions from Table 2.), (a) original functions. (b) first derivative functions. (c) second derivative functions.

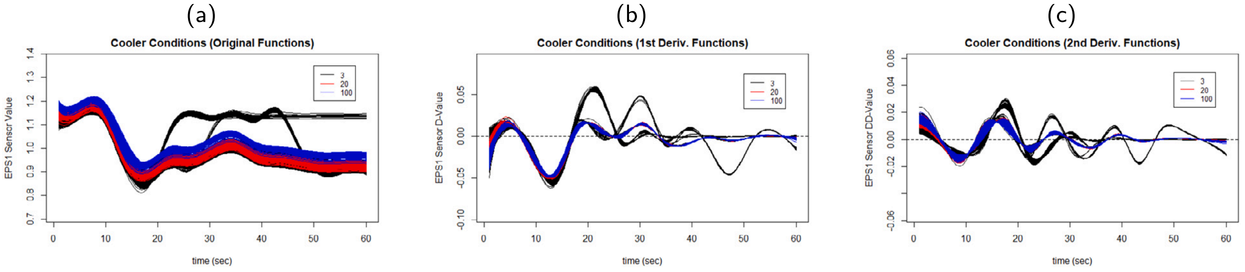

Fig. 4. Sensor “EPS1” functions and “Cooler” fault conditions (different colors represent the different fault conditions from Table 2.), (a) original functions. (b) first derivative functions. (c) second derivative functions.

discrete data points are smoothed into continuous functions as represented by equations (1), as explained in Section 2. This procedure is repeated for each sensor domain. In this paper, cubic B-splines are used to reconstruct the sensor curves, and the number of B-spline basis functions is determined through cross-validation. Consequently, the final format of the equation of smooth sensor curves can be expressed as

$$X_{ij}^{*} = \sum_{b=1}^{B} c_{ijb}\phi_b(t), \quad i = 1,\dots,n \quad j = 1,\dots,J, \tag{16}$$

where c_{ijb} are the coefficients of the basis functions $\phi_b(t)$ for i^{th} hydraulic system, j^{th} sensor and b^{th} basis function and B is number of basis functions. Cubic B-Splines are used to smooth functions (4) and smooth curves for each observation set and each sensor is obtained using equation (2).

Sample plots are created in Fig. 3 and Fig. 4 to represent smooth curves of a system component (Cooler). Each curve represents a 60-second set of observations, while different colors in the figures indicate different failure conditions of that component. Although it is possible to create graphs for all 17 sensors (see Table 1) and four system components (see Table 2) in the dataset. Fig. 3 is generated from the TS1 sensor with a sampling frequency of 1 Hz and Fig. 4 is generated from the EPS1 sensor with a sampling frequency of 100 Hz. Conditions colors are given only for Cooler in this graphs.

5.2. Obtaining the derivatives of multivariate functional data

The second step is to obtain the derivative functions of the original smooth observation functions. The use of derivative functions in FDA is particularly useful for gaining additional insight into the problem, especially when dealing with classification tasks. Fig. 3(b) and Fig. 4(b) show the first derivatives of the original observations, and Fig. 3(c) and Fig. 3(c) indicate the second derivative functions.

5.3. Modified epigraph indexes (MEIs) of the multivariate functional data

In many real-world problems, such as hydraulic systems, individual observations represent real functions of time, and they are observed at discrete moments within a given interval. The dataset described in Section 3 contains multiple observations, indexed by $i = 1,\dots,n$, in multivariate domains, indexed by $j = 1,\dots,p$, defined over a compact time interval $I = [a,b]$, representing the functional data $X_{ij}(t),\dots,X_{nj}(t)$. Each curve represents the evolution over time of a specific process of interest for an individual observation. In this context, the MEI defined in Section 2 is used to detect outliers and also to rank curves with a resulting index vector. The next step is to sort the curves and create an index that will allows us to understand and interpret the correlation between the sensors and the machine component conditions. MEI (7) is used to generate the index vector for each sensor observation set. This involves computing the Spearman’s correlation coefficient, adapted to functional data, using the ordering introduced by the MEIs, similar to the approach outlined in [34].

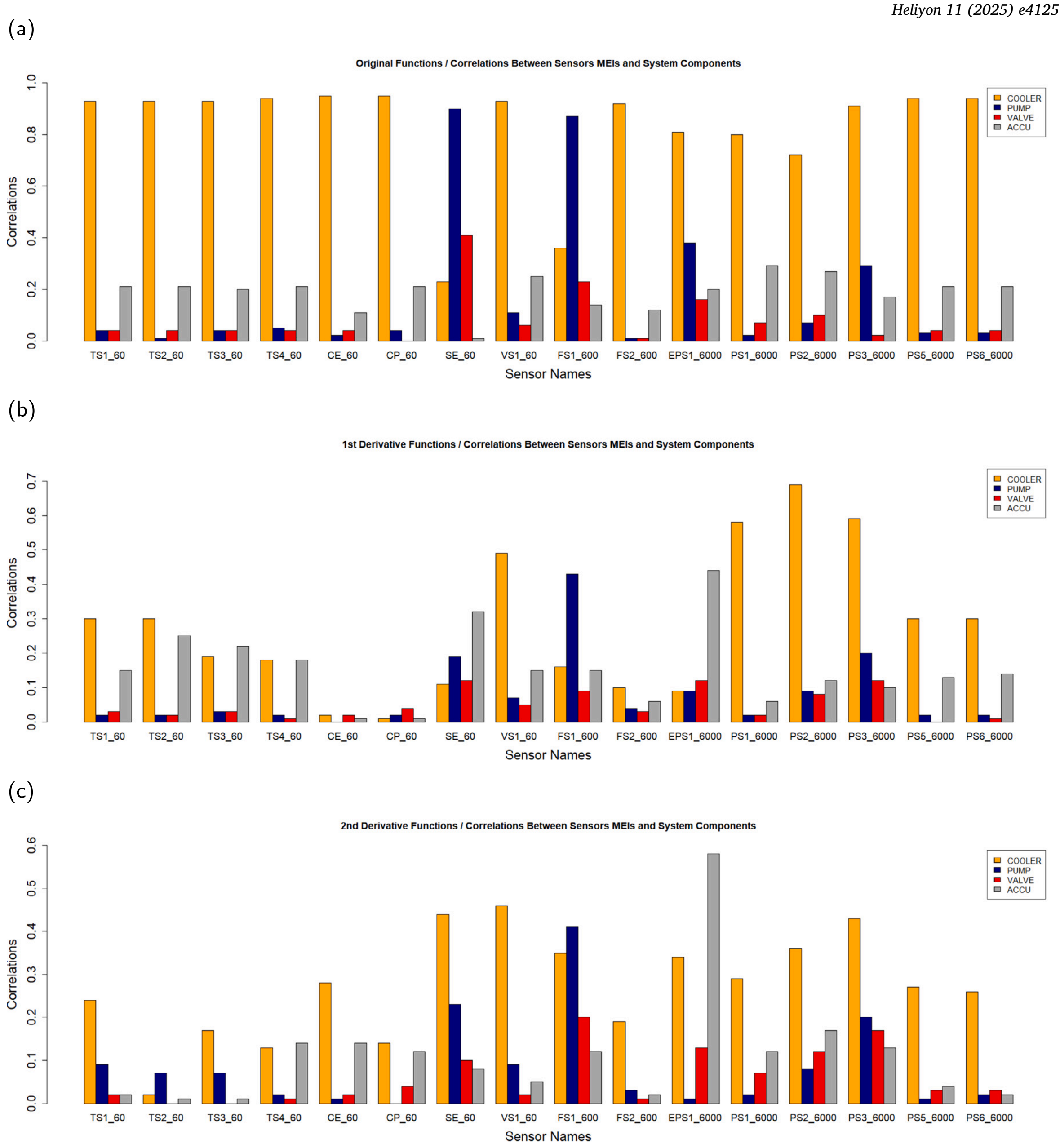

Fig. 5. Correlations between MEI index vector generated for each sensor data and system components. Colors indicate different hydraulic system components and it can be seen that in addition to the original sensor functions, 1st and 2nd derivative functions are also correlated for some sensor/component pairs. (a) Original Function MEIs. (b) 1st Deriv. Function MEIs. (c) 2nd Deriv. Function MEIs.

5.4. Pearson correlations between MEIs and component conditions

The following step is to use MEIs to understand the correlation between the curves and the fault conditions. Through the use of the Pearson correlation coefficient between the MEI vectors and the component conditions. This computation is done for the original functions (Fig. 5(a)), the first derivative functions (Fig. 5(b)), and the second derivative functions (Fig. 5(c)). While most studies in the literature rely on feature extraction techniques to identify the single sensor most relevant to component conditions, the proposed FDA approach allows the use of all multivariate sensor data. The correlation values between the MEI and the component conditions obtained in this step are then used as weighting factors in the next step, MFPCA. Higher correlations result in higher weights, while lower correlations are represented by lower weights. As a result, no information is lost, as all sensors are considered, but any of them contributes proportionally to their correlation.

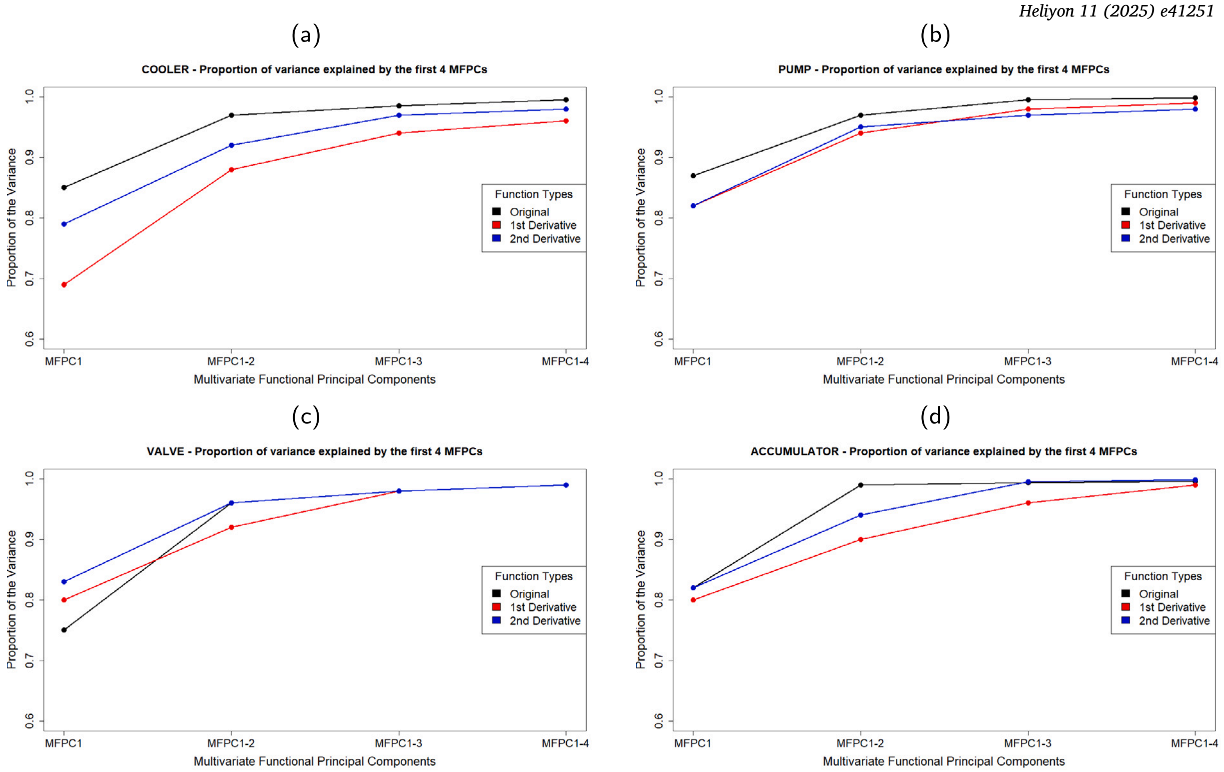

Fig. 6. Cummulative variance explained by the first four MFPCs where the colors represent the original, 1st derivative and the 2nd derivative functions, and the graph (a) for COOLER, (b) for PUMP, (c) for VALVE and (d) for ACCUMULATOR.

5.5. Weighted MFPCA using correlations as weight factors

The failure behavior of an hydraulic system is a stochastic process affected by uncertainties arising from physical degradation dynamics coming from different system components. The degradation pattern can be considered as complex and unknown. To overcome this complexity, MFPCA is applied to understand and explain the data in the multivariate context for the training data set. The multivariate and functional case of PCA called MFPCA has already been propsed by Ramsay and Silverman [31] and Berrendero et al. [4]. The principal components of MFPCA have the same interpretation as those in the functional univariate case. The truncation of (11) at the first q terms provides a reduced dimensional space (see (14)) where classical tools (clustering, regression, . . .) from multivariate analysis can be used to describe the whole process [18]. Estimation of kth principal component score of ith observation set and jth sensor is done following equation (15) in multivariate sense and the weight function $\widehat{f}_k^*(t)$ is the eigenfunctions of covariance operator $\widehat{C}^*$. The solution of the second order eigenequation for each sensor is calculated in equation (12) where $\widehat{C}_j^*(t,u)$ is jth sensor's covariance function is calculated by equation (8). The total explained variance is given by $\sum_{k=1}^{q} \widehat{\lambda}_k^*$ (13) and the proportion of the variance explained by the kth principal component can be calculated as $\lambda_k / \sum_{k=1}^{q} \lambda_k$ [2]. The first four MFPCs for original, first derivative and second derivative functions, can be seen in Fig. 6. Each row in Fig. 6 represents different system components and MFPCA with different weights, obtained from MEIs as explained in Section 5.4. It can be seen that a very high proportion of the variance is already explained by the first three MFPCAs. In calculating the proportion of variance explained, it can be seen that sometimes the original functions perform best and sometimes it is the derived functions. The first three MFPC scores and their relationship with fault conditions (different colors) are given in Fig. 7 for all four component of the hydraulic system.

5.6. Results and discussions

When it comes to interpreting and understanding the first three MFPC scores, the 3D score plot provides a graphical representation of how MFPCA describes and classifies fault conditions in a multivariate context. By examining the first three MFPC scores in a 3D plot, Fig. 7 illustrates how different fault types (represented by various colors) are classified using MFPC scores. Fig. 7(a) demonstrates how "Cooler" fault conditions can be classified using the first three principal component scores of the original function, while Fig. 7(b) shows the classification of "Valve" conditions. Figs. 7(c) and 7(d) display the classifications for "Pump" and "Accumulator" faults, respectively. Additionally, this figure highlights the importance of using derivative functions. In some cases, while MFPCs based on the original function may not be effective for classification, MFPCs derived from the second derivative function can be more useful, as seen in the valve fault in Fig. 7(b).

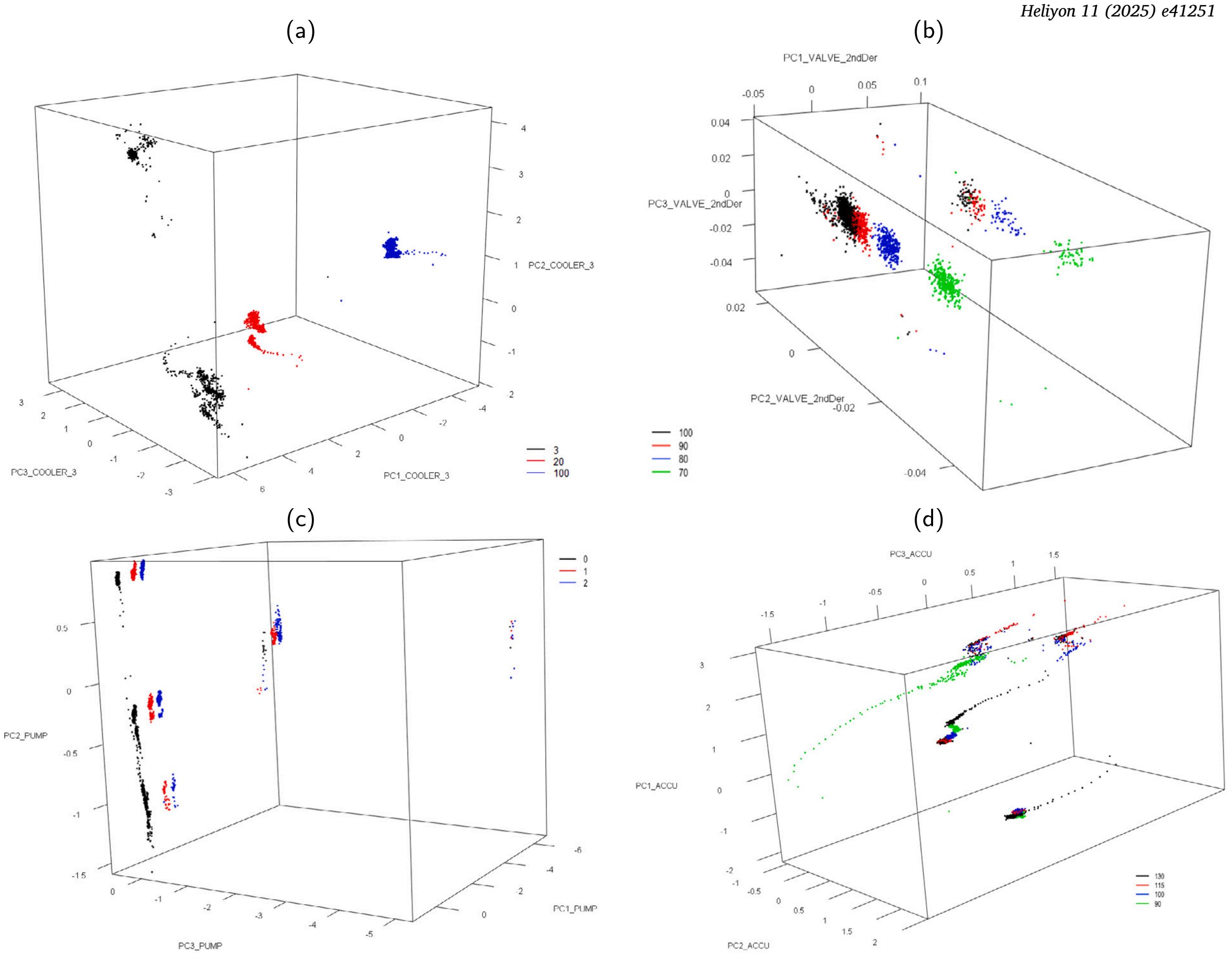

Fig. 7. 3D MFPCA Score Plots for each system component conditions (different colors). a-Cooler. b-Valve. c-Pump. d-Accumulator.

Training and test datasets are separated to evaluate classification performance. As mentioned in Section 3, all samples (2205) are taken into acccount, not just the duty cycles under steady conditions (1449 samples). The first three MFPC scores of the original functions, the first derivative functions, and the second derivative functions were considered as independent variables. The data were divided into 70% for training and 30% for testing, and classification was performed using 1-Nearest Neighbor (1-NN), 5-Nearest Neighbor (5-NN), Support Vector Machine (SVM), and Random Forest (RF). The actual and predicted classification results, along with the confusion matrix, are shown in Table 4 for each system component.

Classification performance is evaluated using *Accuracy* based on confusion matrix (M), the common metric used for fault classification studies that indicates how accurate predicted labels are compared to the corresponding actual labels, and is defined as follows [3]:

$$\text{Accuracy} = 100\frac{\sum_{i=1}^{C} \mathrm{M}(i,i)}{\sum_{i=1}^{C}\sum_{j=1}^{C} \mathrm{M}(i,j)} \tag{17}$$

where C is the total number classes, $\sum_{i=1}^{C} \mathrm{M}(i,i)$ represents the correctly predicted samples and $\sum_{i=1}^{C}\sum_{j=1}^{C} \mathrm{M}(i,j)$ is all the samples in the given dataset. The classification accuracy and the comparison with the other studies in the literature are given in Table 5. Although various classifiers are employed, the highest overall accuracy performance of 96.2% is achieved using RF, as shown in the confusion matrix in Table 4. Confusion matrices for 1-NN, 5-NN, and SVM are available in the Supplementary Material (see Tables S1-S3). Additionally, accuracy calculations and comparisons between the different classification techniques are provided in the Supplementary Material (see Table S8).

As mentioned in Section 2.1, smoothing is performed using cubic B-Splines. In problems involving multivariate functional data, the smoothing process can play a critical role in the performance of MFPCA. While the optimal approach depends on the specific dataset, there are several key parameters related to B-spline fitting that need to be carefully chosen. Theoretically, the number of control points and spline functions can be infinite, but in practice, they must be limited at some point. Key parameters to consider include the smoothing parameter, the order of the penalty, the degree of the B-spline basis, and the number of knots. A common and effective approach, which generally works well across most applications, is to use cross-validation to select the smoothing parameter,

Table 4
Confusion Matrix representing the classification performance using RF after MFPCA

[COOLER] Actual

[COOLER] Predicted	3	20	100
3	227	0	0
20	0	206	0
100	0	0	229

[VALVE] Actual

[VALVE] Predicted	73	80	90	100
73	104	0	0	0
80	0	122	0	1
90	0	0	86	13
100	0	0	15	321

[PUMP] Actual

[PUMP] Predicted	0	1	2
0	374	0	0
1	0	148	0
2	0	0	140

[ACCU] Actual

[ACCU] Predicted	90	100	115	130
90	238	14	1	2
100	9	95	8	2
115	3	6	97	3
130	0	2	7	175

Table 5
Classification Accuracy performance comparison with other studies in the literature.

Component	LDA	ANN	SVM (Linear)	SVM (RBF)	ETSC	**MFPCA** (RF)
Cooler	100	100	100	95.7	100	**100**
Valve	100	100	100	100	100	**95.2**
Pump	73.6	80.0	72.4	64.2	96.0	**100**
Accumulator	54.6	59.4	51.6	6.57	84.4	**89.6**
Overall	81.9	82.6	81.0	81.4	95.1	**96.2**

apply a quadratic penalty, use cubic splines, and set one knot for every four or five observations, with a maximum of around twenty knots. For further details see [1], which provides a comparative study on the performance of regression splines, smoothing splines, and P-splines on both simulated and real-life data.

To understand the impact of the correction on MFPCA performance, we performed a sensitivity analysis by adjusting the number of bases from six bases to twenty-four bases by increasing them by three. Results of the sensitivity analysis with varying bases can be found in the Supplementary Material (see Table S9). As it can be observed, no drastic changes are seen in the classification performance, and we can consider that the classification performance is robust against different smoothing.

The fault classification accuracy of the given approach can be considered competitive when compared to some other studies in the literature. A key strength of the proposed methodology lies in its use of raw multivariate sensor data, in contrast to traditional machine learning approaches that often rely on feature extraction techniques or sensor selection methods. In many existing studies, the classification process involves selecting only the most informative or highly correlated sensors, which can result in the loss of potentially valuable information from other sensors. However, in our study, we intentionally avoid this practice and instead utilize all available sensors across all components, without any prior feature reduction or sensor selection. By considering the full spectrum of sensor data, we preserve the richness and diversity of the information, ensuring a more comprehensive understanding of the system's behavior. This approach eliminates the risk of overlooking subtle fault indicators that may be present in less dominant sensors, which are often excluded in traditional methods. The ability to incorporate the complete set of sensor data not only enhances the robustness and accuracy of fault detection and classification but also significantly broadens the applicability of the method across various engineering domains. This inclusive approach makes the proposed method particularly adaptable to complex systems, where different types of sensors are employed, each contributing unique and critical information to the diagnostic process.

The benefits of utilizing derivative functions have been previously discussed, and their impact on accuracy performance is evident. While the classification performance for the cooler and pump appears robust when only sample curves are considered, the accuracy significantly declines for the valve and accumulator. This discrepancy arises because the MFPC scores for the first and second derivative functions significantly enhance the ability to classify fault conditions for these components. Table 6 compares the classification performance of applying RF after MFPCA, between using only the sample curves and including the derivatives. Other classifier results and CMs can be found in the Supplementary Material (see Tables S4-S7). Not surprisingly, classification performance when including derivative functions is much higher than classification accuracy when considering only sample curves (see Supplementary Material,

Table 6
Comparison of the classification accuracy performance when using solely the sample curves and including the derivatives using RF.

	Only Sample Functions	Derivative Functions Included
Cooler	99.85	100.00
Valve	68.90	95.24
Pump	98.56	100.00
Accumulator	77.73	89.60
Overall	86.26	96.21

Table S8). This was expected because, when we check Fig. 7(b), four colors represent four different fault conditions of the valve, but this classification was visible when we checked the weighted MFPC scores of the second derivative functions. Original sample functions' scores were not sufficient to classify the valve conditions. Furthermore, as shown in Fig. 6, the proportion of variance explained by the first MFPC for the valve and accumulator is lower in the original sample functions compared to the derivatives.

6. Conclusion

With the increasing interest in the Internet of Things (IoT), Machine Learning (ML), and Industry 4.0 fields, Prognostic Health Management (PHM) has also gained importance in engineering applications. In this paper, a novel FDA methodology is presented for multi-fault classification and condition monitoring of engineering systems, which is one of the biggest challenges in PHM. A new, step-by-step Multivariate Functional Principal Component Analysis (MFPCA)-based Functional Data Analysis (FDA) approach is provided as a guide to fault diagnosis. The benefits of this novel multivariate FDA approach are demonstrated, and results are compared with the related literature.

A dataset from a hydraulic system test rig is divided into training and testing groups. Multivariate time series data with different sampling rates, coming from the multisensor environment, are first converted into smooth functional data, and then MFPCA is applied. Before MFPCA, Modified Epigraph Index (MEI) vectors are calculated to rank the observations, and these ranks are used to evaluate the correlation between component conditions and sensor functions. These MEI vectors are used as weight factors in MFPCA. Additionally, considering derivative functions in addition to the original multivariate functional data reveals one of the biggest advantages of FDA studies. To understand the impact of smoothing on classification performance, sensitivity analysis has been done, and the proposed method is robust across varying smoothing parameters. It has been shown that the MFPC scores of the original and derivative functions describe the fault conditions of different system components very well. Multi-fault classification has been performed using different classification methods. Moreover, the ability to use all raw sensor data instead of the feature extraction techniques given in the literature is considered another contribution. Considering the applicability of multiple fault classification techniques, using all sensors through functional data analysis but with different weights, instead of application-specific feature extraction techniques, will increase the applicability of the given approach.

One of the limitations of this study arises when working with machines that have different lifetimes, which leads to modeling the sensors with variable-domain functional data. Currently, a methodological and scalable extension of the Multivariate Functional Principal Component Analysis (MFPCA) procedure for this type of data is under development. This extension could potentially be applied to a broader range of engineering problems involving failures and sensors in the future. Another limitation is that although MFPCA is based on nonparametric statistics, allowing it to work with time series functional data independent of data dispersion or noise, certain specific datasets, such as sensor data containing large discrepancies or sensor data that are extremely constant over time, may not be suitable for applying this approach. The methodology developed in this study can be extended to solve similar fault diagnosis problems across various application domains, such as power systems, manufacturing, and others, where fault detection and classification are critical. This includes industries like defense, aerospace, power and energy, electronics, and manufacturing, among others. Additionally, considering the lifespan of sensors as functional data opens the possibility of combining Reliability with FDA in a more comprehensive manner. This would provide more granular information about the evolution of the sensors, offering insights that go beyond feature extraction or reduction to a single sensor.

CRediT authorship contribution statement

Cevahir Yildirim: Writing – original draft, Methodology, Investigation. **Alba M. Franco-Pereira:** Writing – review & editing, Supervision, Methodology, Funding acquisition. **Rosa E. Lillo:** Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare no conflict of interest

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e41251.

References

[1] Ana M. Aguilera, M.C. Aguilera-Morillo, Comparative study of different b-spline approaches for functional data, Math. Comput. Model. 58 (7–8) (2013) 1568–1579.

[2] M. Carmen Aguilera-Morillo, Ana M. Aguilera, Francisco Jiménez-Molinos, Juan B. Roldán, Stochastic modeling of random access memories reset transitions, Math. Comput. Simul. 159 (2019) 197–209.

[3] Bahman Askari, Raffaele Carli, Graziana Cavone, Mariagrazia Dotoli, Data-driven fault diagnosis in a complex hydraulic system based on early classification, IFAC-PapersOnLine 55 (40) (2022) 187–192.

[4] José Ramón Berrendero, Ana Justel, Marcela Svarc, Principal components for multivariate functional data, Comput. Stat. Data Anal. 55 (9) (2011) 2619–2634.

[5] G.F. Bin, J.J. Gao, X.J. Li, B.S. Dhillon, Early fault diagnosis of rotating machinery based on wavelet packets—empirical mode decomposition feature extraction and neural network, Mech. Syst. Signal Process. 27 (2012) 696–711.

[6] Sudarshan S. Chawathe, Condition monitoring of hydraulic systems by classifying sensor data streams, in: 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), IEEE, 2019, pp. 0898–0904.

[7] Carl De Boor, Carl De Boor, A Practical Guide to Splines, vol. 27, Springer-Verlag, New York, 1978.

[8] Paul HC Eilers, Brian D. Marx, Flexible smoothing with b-splines and penalties, Stat. Sci. 11 (2) (1996) 89–121.

[9] Paul HC Eilers, Brian D. Marx, Practical Smoothing: The Joys of P-Splines, Cambridge University Press, 2021.

[10] Alba M. Franco-Pereira, Rosa E. Lillo, Rank tests for functional data based on the epigraph, the hypograph and associated graphical representations, Adv. Data Anal. Classif. 14 (3) (2020) 651–676.

[11] Alba M. Franco-Pereira, Rosa E. Lillo, Juan Romo, Extremality for functional data, in: Recent Advances in Functional Data Analysis and Related Topics, Springer, 2011, pp. 131–134.

[12] Tomasz Górecki, Mirosław Krzyśko, Functional principal components analysis, in: Data analysis methods and its applications, 2012, pp. 71–87.

[13] Tomasz Górecki, Mirosław Krzyśko, Łukasz Waszak, Waldemar Wołyński, Selected statistical methods of data analysis for multivariate functional data, Stat. Pap. 59 (1) (2018) 153–182.

[14] Clara Happ, Sonja Greven, Multivariate functional principal component analysis for data observed on different (dimensional) domains, J. Am. Stat. Assoc. 113 (522) (2018) 649–659.

[15] Nikolai Helwig, Eliseo Pignanelli, Andreas Schütze, Condition monitoring of a complex hydraulic system using multivariate statistics, in: 2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings, IEEE, 2015, pp. 210–215.

[16] Harold Hotelling, Analysis of a complex of statistical variables into principal components, J. Educ. Psychol. 24 (6) (1933) 417.

[17] Keke Huang, Shujie Wu, Fanbiao Li, Chunhua Yang, Weihua Gui, Fault diagnosis of hydraulic systems based on deep learning model with multirate data samples, IEEE Trans. Neural Netw. Learn. Syst. 33 (11) (2021) 6789–6801.

[18] Jacques Julien, Cristian Preda, Model-based clustering for multivariate functional data, Comput. Stat. Data Anal. 71 (2014) 92–106.

[19] I.T. Jolliffe, Principal Component Analysis, 2nd edn, Springer-Verlag, New York, 2002.

[20] Kari Karhunen, Zur spektraltheorie stochastischer prozesse, Ann. Acad. Sci. Fenn. AI 34 (1946).

[21] Kyutae Kim, Jongpil Jeong, Real-time monitoring for hydraulic states based on convolutional bidirectional lstm with attention mechanism, Sensors 20 (24) (2020) 7099.

[22] Piotr Kokoszka, Matthew Reimherr, Introduction to Functional Data Analysis, CRC Press, 2017.

[23] Yafei Lei, Wanlu Jiang, Anqi Jiang, Yong Zhu, Hongjie Niu, Sheng Zhang, Fault diagnosis method for hydraulic directional valves integrating pca and xgboost, Processes 7 (9) (2019) 589.

[24] Yaguo Lei, Naipeng Li, Liang Guo, Ningbo Li, Tao Yan, Jing Lin, Machinery health prognostics: a systematic review from data acquisition to rul prediction, Mech. Syst. Signal Process. 104 (2018) 799–834.

[25] Michel Loève, * calcul des probabilites-analyse harmonique generale dune fonction aleatoire, C. R. Hebd. Séances Acad. Sci. 220 (12) (1945) 380–382.

[26] Sara López-Pintado, Juan Romo, A half-region depth for functional data, Comput. Stat. Data Anal. 55 (4) (2011) 1679–1695.

[27] B. Martin-Barragan, R.E. Lillo, J. Romo, Functional boxplots based on epigraphs and hypographs, J. Appl. Stat. 43 (6) (2016) 1088–1103.

[28] Karl Pearson, Liii. on lines and planes of closest fit to systems of points in space, Lond. Edinb. Dublin Philos. Mag. J. Sci. 2 (11) (1901) 559–572.

[29] Zhijie Peng, Ke Zhang, Yi Chai, Multiple fault diagnosis for hydraulic systems using nearest-centroid-with-dba and random-forest-based-time-series-classification, in: 2020 39th Chinese Control Conference (CCC), IEEE, 2020, pp. 29–86.

[30] Belén Pulido, Alba M. Franco-Pereira, Rosa E. Lillo, A fast epigraph and hypograph-based approach for clustering functional data, Stat. Comput. 33 (2) (2023) 36.

[31] James O. Ramsay, Bernhard W. Silverman, Functional Data Analysis, 2 edition, Springer-Verlag, New York, 2005.

[32] Tizian Schneider, Nikolai Helwig, Andreas Schütze, Automatic feature extraction and selection for classification of cyclical time series data, Tech. Mess. 84 (3) (2017) 198–206.

[33] Andreas Schütze, Nikolai Helwig, Tizian Schneider, Sensors 4.0–smart sensors and measurement technology enable industry 4.0, J. Sens. Sens. Syst. 7 (1) (2018) 359–371.

[34] Dalia Jazmin Valencia García, Rosa Elvira Lillo Rodríguez, Juan Romo, Spearman coefficient for functions. 2013.

[35] Chuan Wang, Xinxin Chen, Ning Qiu, Yong Zhu, Weidong Shi, Numerical and experimental study on the pressure fluctuation, vibration, and noise of multistage pump with radial diffuser, J. Braz. Soc. Mech. Sci. Eng. 40 (2018) 1–15.

[36] Wei Wang, Yan Li, Yuling Song, Fault diagnosis method of hydraulic system based on multi-source information fusion and fractal dimension, J. Braz. Soc. Mech. Sci. Eng. 43 (2021) 1–13.

[37] Zi Xu, Honggan Yu, Jianfeng Tao, Chengliang Liu, Compound Fault Diagnosis in Hydraulic System with Multi-Output Svm, CSAA/IET International Conference on Aircraft Utility Systems (AUS 2020), vol. 2020, IET, 2020, pp. 84–89.

[38] Yaguo Lei, Jia Feng, Kong Detong, Lin Jing, Xing Saibo, Opportunities and challenges of machinery intelligent fault diagnosis in big data era, J. Mech. Eng. 54 (5) (2018) 94–104.

[39] Shaogan Ye, Junhui Zhang, Bing Xu, Shiqiang Zhu, Jiawei Xiang, Hesheng Tang, Theoretical investigation of the contributions of the excitation forces to the vibration of an axial piston pump, Mech. Syst. Signal Process. 129 (2019) 201–217.

[40] Xiaohang Zhao, Ke Zhang, Yi Chai, A multivariate time series classification based multiple fault diagnosis method for hydraulic systems, in: 2019 Chinese Control Conference (CCC), IEEE, 2019, pp. 6819–6824.

[41] Šulejic Marko, B-Spline and NURBS Curves, Universität, Salzburg, Austria, 2011.