

Fault detection in axial piston hydraulic pumps: integrating principal component analysis with silhouette-based cluster evaluation

Fabian H. Diaz Palencia¹, Carlos Borrás², Cecilia E. García Cena³

Abstract—This paper presents an approach integrating principal component and silhouette analysis with clustering algorithms for fault detection in hydraulic systems. The methodology was validated through a study in which vibration and pressure signals were collected under normal and fault conditions. These signals were then processed through filtering and normalization, followed by dimensionality reduction using principal component analysis. The resulting lower-dimensional feature vectors retained the critical characteristics of both normal and faulty conditions and were subsequently fed into a clustering algorithm. The quality of the resulting clusters was evaluated using silhouette analysis, which offers a reliable means of assessing cluster quality and visualising the outcomes of fault classification. The study demonstrates the effectiveness of this method in accurately representing the patterns of normal and malfunctioning hydraulic pump conditions, ultimately leading to successful diagnostic results.

Keywords: Principal Component Analysis; Silhouette Analysis; Failure Detection; hydraulic piston pump.

Resumen — Este artículo presenta un enfoque que integra el análisis de componentes principales y el análisis de siluetas con algoritmos de agrupamiento para la detección de fallos en sistemas hidráulicos. La metodología se validó a través de un estudio en el que se recopiló señales de vibración y presión en condiciones normales y de fallo. Estas señales fueron procesadas mediante filtrado y normalización, seguidos de una reducción de la dimensionalidad con el análisis de componentes principales. Los vectores de características de menor dimensión resultantes conservaron las características críticas tanto de las condiciones normales como de las defectuosas y posteriormente se introdujeron en un algoritmo de agrupación. La calidad de los conglomerados resultantes se evaluó con el análisis de siluetas, que ofrece un método fiable para evaluar la calidad de los conglomerados y visualizar los resultados de la clasificación de fallos. El estudio demuestra la eficacia de este método a la hora de representar con precisión los patrones de las condiciones normales y defectuosas de las bombas hidráulicas, lo que en última instancia conduce a resultados de diagnóstico satisfactorios.

Palabras clave: Análisis de Componentes Principales; Análisis de Silueta; Detección de fallas; Bomba Hidráulica de Pistones.

1. Mechanical Engineering School, Industrial University of Santander, Colombia. ORCID number <https://orcid.org/0000-0001-6000-0625>

2. Mechanical Engineering School, Industrial University of Santander, Colombia. ORCID number <https://orcid.org/0000-0002-1014-2817>

3. ETSIDI-Centre for Automation and Robotics from Universidad Politécnica de Madrid. Spain. C. Ronda de Valencia 3, 28012. Madrid. Spain. ORCID number <https://orcid.org/0000-0002-1067-0564>

Manuscript Received: 17/12/2024

Revised: 16/01/2025

Accepted: 05/02/2025

DOI: <https://doi.org/10.29019/enfoqueute.1120>

Section Editor: Marcelo Mosquera

I. INTRODUCTION

FAULT monitoring and diagnosis in dynamic systems is a key challenge in industrial engineering, particularly in the analysis of sensor data. The axial piston hydraulic pump is a critical component within hydrostatic systems [1]; it is used as equipment to transmit power in various applications, and its good performance depends on the success and efficiency of the operations in which it is involved within industrial processes. For this reason, companies invest significant efforts and capital in maintenance issues to take advantage of or prolong to the maximum its operation; early detection of failures in the components that integrate it has been a constant research task in recent years. This task consists of determining the type of failure, and for this purpose, it is possible to distinguish three main approaches [2] [3] [4].

- Diagnostic systems for hydraulic systems based on the developed model of the diagnosed system.
- Diagnostic systems for hydraulic systems based on signal analysis.
- Diagnostic systems for hydraulic systems based on knowledge or so-called intelligent fault identification.

Fault detection using knowledge-based methods is a heuristic process. System characteristic values are used to extract features under normal and erratic conditions; once the features are extracted under both conditions, they are compared and the change detection methods are applied. Artificial neural networks, fuzzy logic, principal component analysis and neuro-fuzzy methods can be considered knowledge-based [5]. This paper focuses on applying Principal Component Analysis (PCA) for anomaly and fault detection in time series data, followed by a detailed analysis using t^2 . In addition, silhouette plot analysis is included to assess the quality of the clusters generated from the PCA scores. The aim is to provide a robust methodology for detecting system failures from vibration data or other sensor measurements.

II. BACKGROUND

Several methods have been employed for fault detection and diagnosis in hydraulic systems. Table I compares various techniques used for this purpose. Each process is described in terms of its basic functionality, specific applications in the context of hydraulic systems, key advantages, and inherent limitations. On the opposing side., Table II shows some techniques for detection or diagnosis. of faults in hydraulic systems based on the methods previously described.

A. Principal Components Analysis as a Methodology for Fault Detection

The Principal Components Analysis (PCA) method, introduced by Pearson in 1901 and Hotelling in 1933, is a statistical tool designed to reduce the dimensionality of a dataset containing multiple interrelated variables while preserving as much variation as possible. This is achieved by transforming the original variables into a new set of uncorrelated variables known as principal components. These components are ordered, with the first retaining most of the variation present in the original variables [6].

PCA has proven to be a powerful tool for detecting faults in complex systems, especially in industrial processes and machinery. PCA is well-suited for identifying anomalies in high-dimensional datasets. By projecting new data onto the lower-dimensional space defined by the principal components, deviations from normal operation patterns can be readily identified. This approach allows for the timely detection of subtle changes in system behavior that may indicate the onset of faults. The ability of PCA to handle large datasets, reduce noise, and provide interpretable results has led to its increasing popularity for fault detection (see Fig. 1) across various industries, including manufacturing, chemical processing, and mechanical systems.

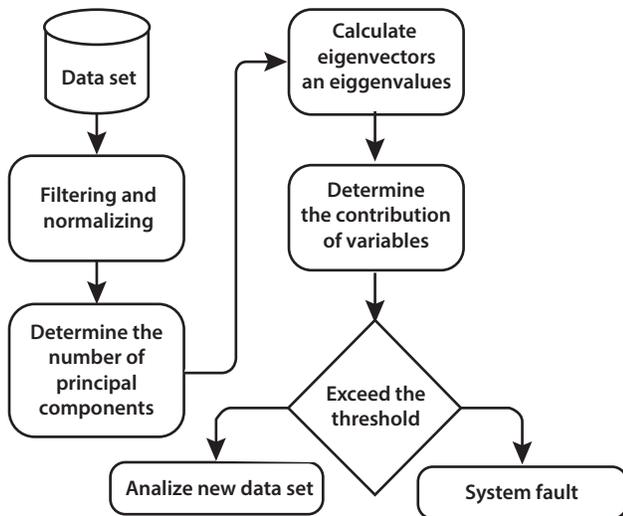


Fig. 1. Typical diagnostic methodology by PCA.

The article [7] mentions that PCA applications in multivariate process failure diagnosis started in the 1990s, and to date, it is still implemented in conjunction with other techniques or methodologies. The idea is to use sensor data and/or variables that describe the operation of devices, processes, or systems and apply PCA to identify the main components that explain the most significant amount of variability in the data to identify patterns or relationships that may indicate the presence of a fault or erratic operation.

L. Siyuan et al., in [8] show a study on the application of PCA for fault detection in Hydraulic Pumps, and in [9] combines

Rough Set (RS) and PCA to diagnose faults based on the energy characteristics of Vibration signals also in hydraulic pumps.

M. Atoui et al., in [10] apply Bayesian Networks (BN) and PCA, BN and 4381 T^2 -SPE for fault detection, validating both methodologies using the Tennessee Eastman Process, showing that both methods produce the same performance at the time of fault detection.

Villegas et al. in [11] describe the PCA application for fault detection and diagnosis in a real plant. The approach includes a PCA model for each system behavior, i.e., models under normal and fault conditions. It demonstrates that detecting and identifying level sensor failures and clogging in a two-communicating tank system is possible. The paper [29] presents a multimode process monitoring technique that integrates density peak clustering and kernel principal component analysis with a multi-strategy zebra optimization algorithm. The proposed method enhances mode identification accuracy and fault detection capabilities in dynamic industrial processes. Experimental validation demonstrates the method's superiority over traditional techniques, achieving high fault detection rates and low false alarm rates across various scenarios, particularly in identifying transition modes.

The study presented in [12] introduces an innovative approach for diagnosing faults in grid-connected photovoltaic systems by combining feature extraction techniques like the Salp Swarm Algorithm with supervised machine learning classifiers. The model's performance is compared against traditional methods such as PCA and Kernel PCA. The findings demonstrate that the model achieves a diagnostic accuracy of over 99 % and greatly enhances computational efficiency compared to PCA and KPCA. This improvement is particularly notable for fault classification in nonlinear systems where PCA and KPCA are less effective.

The research detailed in [13] is centered on creating a fault diagnosis and location system for nuclear plant equipment. It utilizes PCA to reduce the dimensionality of sensor data from 70 to 21 dimensions, resulting in improved classification accuracy during the training of a Residual Network, with a peak accuracy of 98.59 %. In a study referenced as [20], Diaz used PCA for detecting failures related to loss of volumetric efficiency and applied SVM to classify the severity of failure in an axial piston pump. The study yielded results close to 99 %.

Zhao et al. [14] propose a new fault diagnosis method based on PCA. First, they transform the vibration signal to the frequency domain. Then, they use the PCA method to reduce the dimension of the feature matrix. Finally, the reduced feature vector is fed into another model to diagnose faults in a rotating machinery bench.

Cárdenas et al. [15] developed a PCA-based approach to detect and categorize faults in a natural gas engine. Their algorithm analyzed alarm bursts to distinguish normal system behavior from failures. The results they obtained were quite promising and outperformed the existing methods used by operators.

TABLE I
METHODOLOGIES FOR FAULT DETECTION

Methodology	Description	Application in Hydraulic Systems	Advantages	Limitations
Support Vector Machines	Classification technique that finds the best margin of separation between classes in a high-dimensional space.	Classification of system status as “normal” or “anomalous” based on sensor data characteristics.	Effective in handling non-linear data; high classification accuracy.	May be sensitive to kernel choice; may not scale well with large data sets.
Artificial Neural Networks	Computational models inspired by the human brain consist of layers of nodes (neurons).	Modeling complex and nonlinear relationships between variables to predict failures based on historical data.	Ability to learn complex relationships; adaptability to different types of data.	Require large amounts of data for training; risk of overfitting.
Convolutional Neural Networks	Deep neural networks are designed to process structured data in the form of images.	Analysis of sensor data or thermal image patterns to identify features that indicate faults.	Efficient in feature extraction; suitable for high-dimensional data.	They require high computational power and training time; they require labeled data.
Recurrent Neural Networks	Networks are designed to handle sequential data and capture temporal dependencies.	Time series analysis of sensor data to detect failure patterns over time.	Capture temporal dependencies; effective for sequential data.	Complexity in training; risk of long-duration gradients and fading.
Decision Trees and Random Forests	Prediction models that divide data into nodes based on binary questions combine multiple trees.	Classification and prediction of failures, combining the output of several trees for more robust decisions.	Easy to interpret; good accuracy; reduce the risk of overfitting.	They can be prone to overfitting without proper adjustment; they may require a lot of training time.
Clustering Algorithms	Group data based on similarities and patterns.	Identification of patterns and anomalies in sensor data that may indicate faults.	No predefined labels are required; suitable for detecting unknown patterns	Sensitive to the number of clusters; may not handle noisy data well.
Principal Component Analysis	Dimensionality reduction technique that transforms data to a new space.	Reduced data complexity to improve the efficiency of other fault detection algorithms.	Simplifies complex data; facilitates visualization of patterns.	May lose relevant information; not always suitable for non-linear data.
Model-Based Methods	They use mathematical models of the system to predict expected behavior and detect anomalies.	Comparison of actual data with the mathematical model to detect deviations that indicate failures.	Accuracy in anomaly detection based on a detailed understanding of the system.	It requires accurate and detailed models, but it can be complex to implement.
Expert and Rule-Based Systems	They use predefined rules and expert knowledge to make decisions and diagnose failures.	Fault diagnosis is achieved by applying specific rules and heuristics that reflect expert knowledge.	Apply domain-specific knowledge; easy to interpret.	Limited by predefined knowledge; cannot adapt to new situations easily.
Anomaly Detection Methods	They focus on identifying significant deviations from normal system behavior.	Using statistical and machine learning techniques, identifying anomalous behaviors that could indicate failures.	Effective in detecting unexpected behavior; can be adapted to different data types.	May generate false positives; require good definition of what constitutes an anomaly.

TABLE II
FAULT DETECTION METHODS AND THEIR CONTRIBUTIONS IN HYDRAULIC SYSTEMS

Article	Fault Type	Analysis Method	Problems Encountered	Contribution
[16]	Fault in the control valve, displacement regulation mechanism, and rotating group.	Kalman Filter with Unknown Input and Residual Evaluation Based on cumulative sum.	Complexity in estimating the swashplate moment and sensitivity to operating conditions.	Improves fault detection sensitivity and speed by considering uncertainties in the swashplate.
[17]	Wear of the valve plate.	Vibration signal analysis using wavelet analysis and Artificial Neural Networks.	No specific problems were mentioned, but spectral analysis and feature selection techniques is emphasized.	Develops a methodology to detect and classify wear faults in valve plates, facilitating condition-based maintenance.
[18]	Weak faults in hydraulic pumps.	Vibration signal fusion using Enhanced Empirical Wavelet Transform and Variance Contribution Rate.	Inadequate segmentation of the spectrum in the Empirical Wavelet Transform.	Introduces an improved method to achieve more accurate spectrum segmentation, enhancing weak fault detection.
[19]	Failures in positive displacement pumps, specifically in three-screw spindle pumps.	Extended Kalman Filter for the estimation of non-measurable signal data.	Difficulty in obtaining accurate mathematical models and the inability to use conventional sensors due to fluid turbulence.	Development and implementation of a prototype that improves accuracy and reliability in fault detection, providing a solution without the need for traditional sensors.
[20]	Abrasive wear in components of a positive displacement multi-piston pump.	K-Nearest Neighbor classifier based on vibration signals, static and dynamic pressure, and working medium flow rate.	Formation of elliptical depressions on the cam plate and microchannels in the valve plate due to the loss of the lubricating layer.	Demonstrates that a basic classifier like K-Nearest Neighbor can achieve high accuracy in detecting wear conditions.

[21]	Wear of the valve plate in a swash-plate axial piston pump.	Causation-based Linear Interpolation and Wavelet-based Adaptive Signal Analysis of Instantaneous Angular Speed Fluctuation waveform.	Interference of sensors position, periodic noise disturbances from bearings, shafts, and other rotating components.	Application of the Instantaneous Angular Speed signal for fault diagnosis of a swash-plate axial piston pump, for detecting wear faults in the valve plate. This approach is validated through experimental results.
[22]	Wear conditions and cavitation phenomena.	Support vector machines, extreme learning machines, deep belief networks and the minimum redundancy maximum relevance algorithm for feature ranking.	The limitation of conducting diagnostics under stationary operating conditions may not fully represent the operational realities of hydraulic systems.	Development of a neural classifier for pump wear state classification with high accuracy rates for pressure signals and the application of deep machine learning techniques to effectively detect and classify multiple faults in hydraulic systems.
[23]	The fault type analyzed is the detection of PDM (Positive Displacement Motor) stalls during coiled tubing operations.	Fuzzy Logic Inference System (FIS) that monitors surface parameters and detects motor stalling using data from coiled tubing unit surface sensors.	The limitation of a relatively small dataset available for development, which contains data from only 4 milling operations, and the reliance on human interaction with the equipment that is not recorded in the data.	The main contribution of the research is the development of a Fuzzy Logic Inference System that supervises surface data, recognizes abnormal situations, and informs the user to help avoid human-induced errors during coiled tubing operations.
[27]	Slight cylinder faults and severe cylinder faults in axial piston pumps.	The physics informed neural network framework is used for predicting pump flow ripple, which serves as a clear indicator of pump health. The study utilized a calibrated pipeline model to obtain simulated pressure ripples.	The time-consuming and expensive nature of solving the inverse problem, poor initial and partial boundary conditions.	Their framework has been validated through numerical and experimental studies, demonstrating high accuracy in predicting flow ripples and identifying fault characteristics, thus enhancing fault diagnosis in hydraulic systems.
[28]	Axial piston pump faults, including slipper wear, loose boot, and valve plate wear.	Domain adversarial transfer fault diagnosis method based on multi-scale attention mechanisms.	Insufficient feature extraction and domain adaptation capability in cross-situation and partially unlabeled samples.	The proposed method effectively improves fault diagnosis accuracy and provides new ideas for further research on axial piston pump fault diagnosis.

III. METHODOLOGY

To develop this work, a test bench (see Fig. 2) equipped with a Siemens 40 [HP] 1200 rpm electric motor and an Eaton 54 series axial piston pump was used to induce the failure conditions (see Fig. 3). The load was generated by means of a manifold consisting of two crossed relief valves, with which it was possible to maintain the same load conditions during the experiment (see Fig. 4).



Fig. 2. Test bench



Fig. 3. Positions taken for vibration measurement.

Two datasets were collected, one for normal and another for fault conditions test data are also separated to perform failure predictions under unknown conditions. At a pre-established load of 700 psi, maintaining the measurement variables such as operating regime, external noises, and hydraulic oil temperature in a constant range. To perform the proposed study, we proceeded to capture signals of preload pressure, discharge

pressure, and mechanical vibrations in four points at the pump, in fault and non-fault conditions must be gathered (Table III).

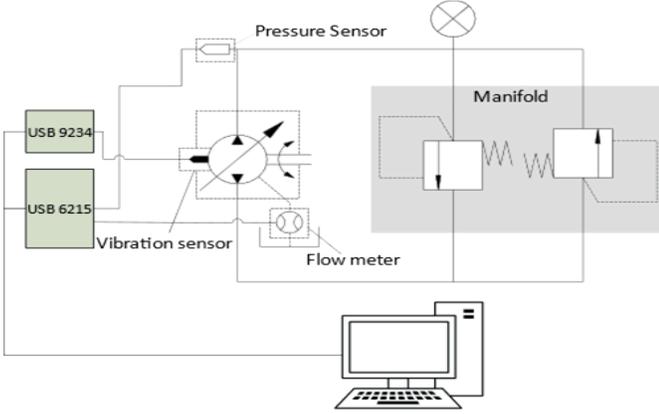


Fig. 4. Experiment's schematic.

Instrumentation and sensors:

- WIKA pressure transducer model ECO-1 to measure discharge pressure.
- Diagnostic systems for hydraulic systems based on signal analysis.
- Diagnostic systems for hydraulic systems based on knowledge or so-called intelligent fault identification.
- NI USB-6215 card
- NI USB-9234 card
- Laptop for data analysis.

TABLE III
PROCESS DATA CORRESPONDING TO NORMAL
AND FAULT OPERATING CONDITIONS

Case	Vib 1	Vib 2	Vib 3	Vib 4	Precharge pressure Volt	Discharge pressure Volt
1	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}
2	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}	X_{26}
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
n	X_{n1}	X_{n2}	X_{n3}	X_{n4}	X_{n5}	X_{n6}

The experiment uses the components mentioned above and LabVIEW software for signal acquisition. The sampling frequency was 100 kHz for pressure data and 50 kHz for vibration data. Intentionally wearing out the valve plate induces one kind of fault condition: loss of volumetric efficiency. The data set includes data obtained when the pump worked in both fault and normal conditions and comprises 2^{16} observations from the six variables.

The Fast Fourier Transform (FFT) was applied to time series data to explore signals' spectral characteristics. The FFT decomposes vibration signals into their frequency components,

enabling visualization of how fault conditions alter the data's spectral characteristics. The amplitude spectrum is calculated for each signal of each variable (sensor), highlighting frequency differences between normal and fault conditions.

In the Fig. 5, we can see the difference in the vibration spectrum in fault and not fault for the four accelerometer positions.

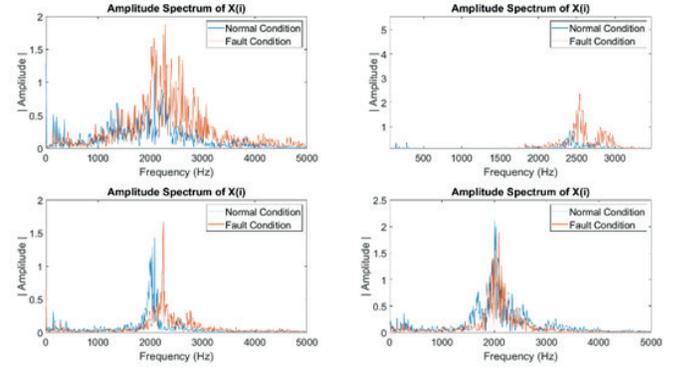


Fig. 5. Vibration Spectrum for normal and fault condition.

A. Principal Component Analysis (PCA)

PCA reduces data dimensionality and extracts the primary characteristics describing data variations. The technique projects data into a lower-dimensional space while preserving maximum variance [24]. Given a set of observations associated either with control, monitoring or simply as indicators of the process, new variables called principal components are constructed such that considering a data matrix $X \in R^{n \times m}$ (Eq. 1):

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{pmatrix} \quad (1)$$

It is convenient to normalise the data for each variable so that all the variables have the same weight in the computation. Then from this matrix, the covariance matrix can be calculated as follows (Eq. 2):

$$S = \frac{1}{n-1} X^t X \quad (2)$$

Performing singular value decomposition (SVD) (Eq. 3):

$$S = V \Lambda V^2 \quad (3)$$

Where Λ is a diagonal matrix with the eigenvalues of S sorted in descending order ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0$), the columns of the matrix V are the eigenvectors of S . The transformation matrix $P \in \mathbb{R}^{m \times a}$ is generated by choosing the eigenvectors or columns V corresponding to the eigenvalues. For each dataset (normal and fault), the number of principal components necessary to explain at least 90 % of the variance is selected by the percentage of variability explained criterion,

where the number of principal components a is taken so that P_a is close to a user-specified value [25], and the new dataset of smaller dimensions than the original is given by Eq. 4:

$$T = XP \quad (4)$$

Now, the original dataset can be represented in terms of its eigenvectors, which define the direction of the principal components (Eq. 5):

$$\hat{X} = TP^T \quad (5)$$

The difference between the original dataset and the transformed dataset is the residue matrix (Eq. 6):

$$E = X - X^X \quad (6)$$

B. Statistics for monitoring with PCA

- Hotelling (T^2): This statistic is used to detect anomalies in new observations by comparing T^2 values to a determined threshold. Given an observation vector of the process, we can define this states statistic of the form (Eq. 7):

$$T^2 = X^T P \Lambda_a^{-1} P^T X \quad (7)$$

This threshold can be calculated from the sample data following Eq. 8:

$$T_a^2 = \frac{a(n^2 - 1)}{n(n - a)} F_{\alpha}(a, n - a) \quad (8)$$

Where n is the number of samples taken for the calculation of the PCA and $F_{\alpha}(a, n - a)$ is the critical value of the Fisher-distribution) with n and $(n - a)$ degrees of freedom and a level of significance, which specifies the degree of commitment to false alarms. Its most typical values are 0.01 and 0.05.

C. Cluster analysis

According to [25], K-means clustering is an unsupervised non-hierarchical clustering algorithm focusing on similarity. It was applied to group data into two clusters: one for normal conditions and another for fault conditions. Once PCA has reduced the data, we apply the K-means clustering algorithm to group the data points into clusters to identify normal and fault condition data clustering patterns [26]. Given a set of observations $(x_1, x_2, x_3, \dots, x_n)$ and $(S = S_1, S_2, S_3, \dots, S_n)$ is the sum of the distances from the objects to its centroid and minimizing it; m_i is the mean (also called centroid) of points in cluster (Eq. 9).

$$\min \sum_{i=1}^K \sum_{x_j \in S_{\bullet i}} \|x_j - \mu_i\|^2 \quad (9)$$

Finally, we used silhouette analysis to evaluate the quality of the clusters obtained. According to [5] Data i in the cluster C_i , $a(i)$ is the average intra-cluster distance and $b(i)$ is the average inter-cluster distance. The number $s(i)$ is obtained by combining $a(i)$ and $b(i)$ following Eq. 10:

$$s(i) = \frac{b(i) - a(i)}{\max\{b(i), a(i)\}}; -1 \leq s(i) \leq 1 \quad (10)$$

The silhouette coefficient measures the coherence and separation of the points within the clusters. However, instead of calculating the silhouette coefficient in the original data space, we calculate it in the principal component space. This allows us to visualize and evaluate the quality of the clusters in a lower-dimensional space. These plots assess cluster cohesion and separation, indicating the effectiveness of segmentation. The silhouette score measures how well samples are clustered with similar samples to evaluate the quality of clusters produced by clustering algorithms like K-Means [4]. The silhouette score can range from -1 to 1, with higher values indicating well-clustered objects and lower values suggesting that an object might belong to the wrong cluster.

D. Parameter Selection

The number of clusters ($k=2$) was determined for K-means clustering based on our prior knowledge of the system states (normal and fault conditions). For the silhouette analysis, a minimum silhouette score threshold of 0.5 was established to ensure cluster quality. For both analyses, the distance metric employed was ‘‘correlation,’’ which is defined by the Matlab Help Center as ‘‘One minus the sample correlation between points (treated as sequences of values).’’ In this context, each centroid represents the component-wise mean of the points within that specific cluster, following the procedure of centering and normalizing these points to achieve a zero mean and unit standard deviation. This approach ensures that the clustering process effectively captures the underlying relationships between data points while accounting for variations in scale and distribution.

IV. RESULTS

A. Data Preprocessing and Principal Component Analysis

The results of the PCA analysis indicated significant dimensionality reduction. The initial dataset, which comprised six variables, was effectively condensed into a lower-dimensional space. This reduction was achieved using three principal components, which collectively accounted for at least 90 % of the total variance, (see Fig. 6). This finding underscores the efficacy of PCA in simplifying complex datasets while retaining critical information.

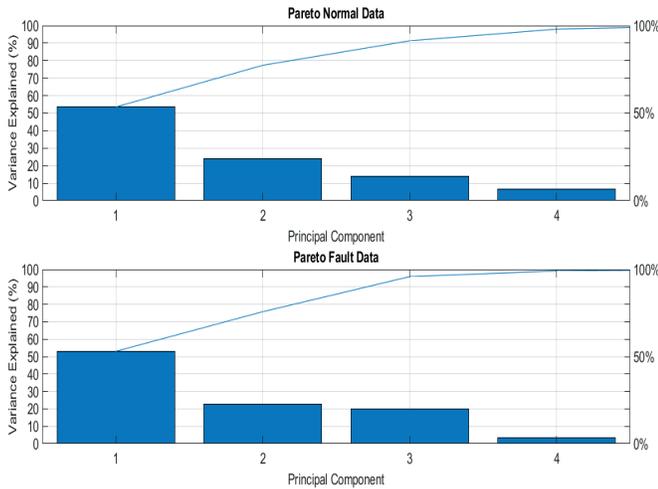


Fig. 6. Pareto

B. CLUSTERING RESULTS

K-means clustering was applied to the data transformed through Principal Component Analysis (PCA), successfully generating two distinct clusters. One cluster represents normal operational conditions, while the other encapsulates fault conditions. This method effectively separates the data points corresponding to normal and fault conditions, thereby enhancing the understanding of the underlying patterns within the dataset. Such differentiation is crucial for monitoring and diagnosing system performance in various applications (see Fig. 7). The Silhouette scores were computed within the principal component space, revealing strong indications of effective cluster separation and coherence. This analysis suggests that the clustering methodology employed is successful in distinguishing between the identified groups while maintaining internal consistency among observations within each cluster (see Fig. 8).

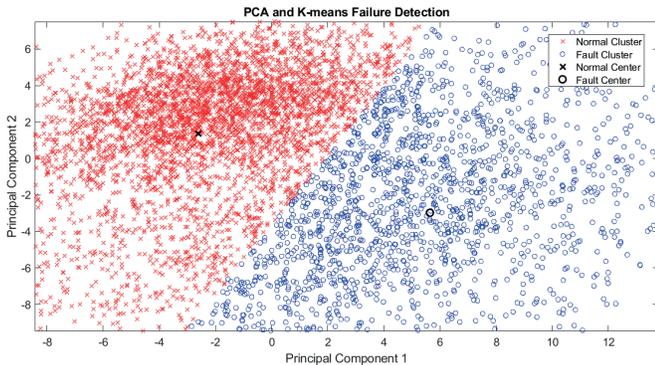


Fig. 7. Clusters from Normal and Fault data by K-means

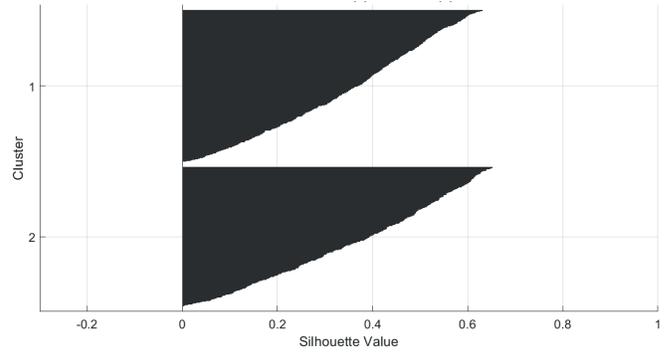


Fig. 8. Clusters from Normal (1) and Fault (2) data

C. ANOMALY DETECTION STATISTICS

The T^2 statistic threshold was determined using the Fisher distribution (see Fig. 9). Observations that exceed this threshold were identified as potential anomalies, with significance levels set at 0.05. Figure 10 depicts the monitoring of the T^2 and how is able to detect the failure at the instant of occurrence (1000).

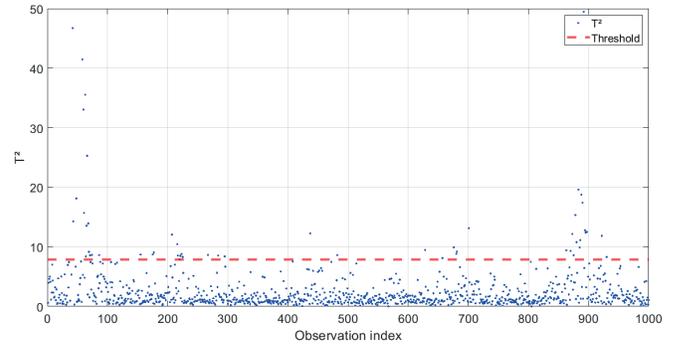


Fig. 9. Threshold

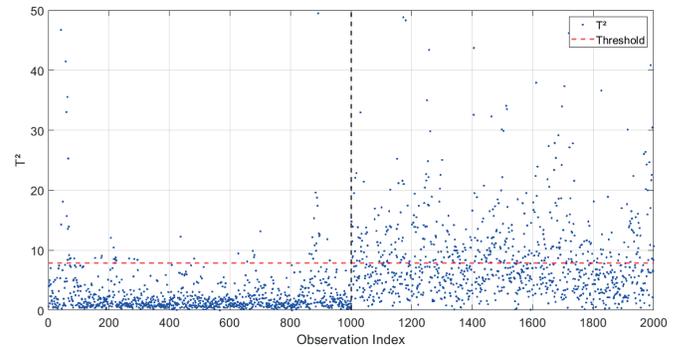


Fig. 10. Fault Detection

V. DISCUSSION

While our methodology has demonstrated promising results in fault detection, certain limitations must be acknowledged. Primarily, the approach tends to assume steady-state operational conditions, which may restrict its applicability in more dynamic environments characterized by rapidly changing system states. Moreover, although Principal Component Analysis (PCA) is effective for dimensionality reduction, it may unintentionally overlook complex nonlinear relationships present within the data. Additionally, the current clustering methodology necessitates complete retraining when encountering new fault types, which could hinder its adaptability and scalability across various hydraulic system configurations.

To mitigate these limitations, we could consider the integration of deep learning techniques to enhance the management of complex nonlinear relationships. Developing adaptive clustering parameters would facilitate a more dynamic assessment of different operating conditions. Furthermore, extending the method to accommodate multi-fault classification scenarios would significantly enhance its versatility in hydraulic systems.

Performing a comparative analysis. The proposed method of PCA and silhouette analysis offers advantages in terms of computational efficiency, interpretability of results and ability to operate without labeled data. However, it is limited by its dependence on linear relationships and requires careful parameter tuning. In contrast, deep learning techniques excel at capturing complex patterns and handling nonlinear relationships, but require large data sets and considerable computational resources. Finally, model-based approaches offer a unique perspective due to their physics-based understanding and ability to operate with limited data; however, they face challenges related to complex model development processes and system-specific constraints. These methodologies present a unique balance between computational complexity, data requirements, and diagnostic accuracy, demonstrating the importance of selecting an approach that fits the specific characteristics of the hydraulic system and the objectives of each investigation.

VI. CONCLUSION

The findings illustrate the effectiveness of this methodology in revealing clearer clustering patterns and helping to identify potential anomalies or deviations from standard operational conditions.

The proposed methodology effectively reduces the dimensionality of the data while preserving crucial essential information for fault detection. Silhouette analysis within the principal component space emerges as a valuable tool to assess and visualize the quality of the cluster, which aids in the identification of anomalies. The method shows promise for early fault detection by identifying possible transient states or emerging fault conditions.

Despite the encouraging results, further research is needed to validate the reliability of the method across various operating conditions and fault types. Future studies should incorporate systematic comparisons with other fault detection methods, investigate alternative clustering techniques, and validate the findings using a more diverse and extensive dataset.

Looking ahead, future work will focus on further exploring and refining this methodology across different domains and datasets, as well as investigating other techniques and algorithmic combinations to achieve even more robust results. Fault detection is a continually evolving field, and the use of combined approaches can enhance the accuracy and efficiency of these processes.

ACKNOWLEDGMENT

This research was supported by Vicerrectoría de Investigación y Extensión (VIE) of the Universidad Industrial de Santander, UIS, Colombia. UIS - VIE 1366 Research Funding Project.

REFERENCES

- [1] S. Liu et al., A new test method for simulating wear failure of hydraulic pump slipper pair under high-speed and high-pressure conditions, *Front. Energy Res.*, vol. 10, no. January, pp. 1-14, 2023, <https://doi.org/10.3389/fenrg.2022.1096633>
- [2] M. T. Amin, F. Khan, S. Imtiaz and S. Ahmed, Robust Process Monitoring Methodology for Detection and Diagnosis of Unobservable Faults, *Ind. Eng. Chem. Res.*, vol. 58, no. 41, pp. 19149-19165, 2019. <https://doi.org/10.1021/acs.iecr.9b03406>
- [3] G. Patterson-hine, G. Aaseng, S. Biswas, S.Narasimhan and K. Patipati, "A Review of Diagnostic Techniques for ISHM Applications", *NASA Ames Res. Cent. Honeywell Def. Sp. Electron. Syst.*, vol. 1, 2005.
- [4] J. Watton, *Modelling, Monitoring and Diagnostic Techniques for Fluid Power Systems*. Wales: Springer, 2007.
- [5] V. K. Kandula, *Fault detection in process control plants using principal component analysis* (LSU Master's Theses). 2011.
- [6] I. T. Jolliffe, "Principal Component Analysis, Second Edition", *Encycl. Stat. Behav. Sci.*, vol. 30, no. 3, pp. 487, 2002. <https://doi.org/10.2307/1270093>
- [7] J. Mina and C. Verde, *Detección de fallas usando análisis de componentes principales*, pp. 431-436, 2004. Instituto de Ingeniería, UNAM.
- [8] L. Siyuan et al., "Application of PCA for fault detection in Hydraulic Pumps," *Journal of Hydraulic Engineering*, vol. 15, no. 4, pp. 234-245, 2010.
- [9] L. Siyuan et al., "Rough Set and PCA for fault diagnosis in hydraulic systems", *International Journal of Fluid Power*, vol. 10, no. 2, pp. 50-160, 2012.
- [10] M. A. Atoui et al., "Bayesian Networks and PCA for fault detection in Tennessee Eastman Process," *AI in Manufacturing*, vol. 5, no. 1, pp. 45-58, 2015.
- [11] Villegas et al., "PCA application for fault detection in real plants", *IEEE Transactions on Industrial Applications*, vol. 29, pp. 255-263, 2017.
- [12] A. Hichri, M. Hajji, M. Mansouri, H. Nounou and K. Bouzrara, "Supervised machine learning-based salp swarm algorithm for fault diagnosis of photovoltaic systems", *J. Eng. Appl. Sci.*, vol. 71, no. 1, pp. 1-17, 2024, <https://doi.org/10.1186/s44147-023-00344-z>
- [13] X. Ying Huang, H. Xia, W. zhe Yin and Y. kuo Liu, "Research on fault diagnosis and fault location of nuclear power plant equipment", *Ann. Nucl. Energy*, vol. 205, no. April, p. 110556, 2024, <https://doi.org/10.1016/j.anucene.2024.110556>
- [14] H. Zhao et al., "PCA and Salp Swarm Algorithm for fault detection in photovoltaic systems", *Renewable Energy*, vol. 23, pp. 98-107, 2019.
- [15] Y. Cardenas, G. Carrillo, A. Alviz, A. Alviz, I. Portnoy, J. Fajardo, E. Ocampo and E. Da-Costa, "Application of a Pca-Based Fault Detection and Diagnosis Method in a Power Generation System With a 2 Mw Natural Gas Engine", *EUREKA, Phys. Eng.*, vol. 2022, no. 6, pp. 84-98, 2022, <https://doi.org/10.21303/2461-4262.2022.002701>
- [16] B. Xu, X. Huang, J. Zhang, W. Huang, F. Lyu and H. Xu, "A Fault Detection Method for a Practical Electro-Hydraulic Variable-Displacement Pump with Unknown Swashplate Moment", *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1-11, 2023, <https://doi.org/10.1109/TIM.2023.3265090>
- [17] J. G. Maradey Lázaro and C. Borrás Pinilla, "A methodology for detection of wear in hydraulic axial piston pumps", *Int. J. Interact. Des.*

- Manuf.*, vol. 14, no. 3, pp. 1103-1119, 2020, <https://doi.org/10.1007/s12008-020-00681-w>
- [18] H. Yu, H. Li, and Y. Li, "Vibration signal fusion using improved empirical wavelet transform and variance contribution rate for weak fault detection of hydraulic pumps", *ISA Trans.*, vol. 107, pp. 385-401, 2020, <https://doi.org/10.1016/j.isatra.2020.07.025>
- [19] A. Dabrowska, R. Stetter, H. Sasmito and S. Kleinmann, "Extended Kalman Filter algorithm for advanced diagnosis of positive displacement pumps", *IFAC Proceedings*, vol. 45, no. 20, <https://doi.org/10.3182/20120829-3-MX-2028.00068>
- [20] J. Konieczny and J. Stojek, "Use of the k-nearest neighbour classifier in wear condition classification of a positive displacement pump", *Sensors*, vol. 21, no. 18, 2021, <https://doi.org/10.3390/s21186247>
- [21] J. M. Liu, L. C. Gu, and B. L. Geng, "A practical signal processing approach for fault detection of axial piston pumps using instantaneous angular speed", *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.*, vol. 234, no. 19, pp. 3935-3947, 2020, <https://doi.org/10.1177/0954406220917704>
- [22] J. Konieczny, W. Łatas, and J. Stojek, "Classification of Wear State for a Positive Displacement Pump Using Deep Machine Learning †", *Energies*, vol. 16, no. 3, pp. 1-19, 2023, <https://doi.org/10.3390/en16031408>
- [23] R. A. Galo Fernandes, P. M. Silva Rocha Rizol, A. Nascimento, and J. A. Matelli, "A Fuzzy Inference System for Detection of Positive Displacement Motor (PDM) Stalls during Coiled Tubing Operations", *Appl. Sci.*, vol. 12, no. 19, 2022, <https://doi.org/10.3390/app12199883>
- [24] C. Liu, J. Bai, and F. Wu, "Fault Diagnosis Using Dynamic Principal Component Analysis and GA Feature Selection Modeling for Industrial Processes", *Processes*, vol. 10, no. 12, 2022, <https://doi.org/10.3390/pr10122570>
- [25] Y. Huo, Y. Cao, Z. Wang, Y. Yan, Z. Ge, and Y. Yang, "Traffic anomaly detection method based on improved GRU and EFMS-Kmeans clustering", *Comput. Model. Eng. Sci.*, vol. 126, no. 3, pp. 1053-1091, 2021, <https://doi.org/10.32604/cmescs.2021.013045>
- [26] J. Wu, *Advances in K-means Clustering a Data Mining Thinking*. Springer-Verlag Berlin Heidelberg, 2012.
- [27] C. Dong, J. Tao, H. Sun, Q. Wei, H. Tan, and C. Liu, "Innovative fault diagnosis for axial piston pumps: A physics-informed neural network framework predicting pump flow ripple", *Mech. Syst. Signal Process.*, vol. 225, no. January, p. 112274, 2025, <https://doi.org/10.1016/j.ymssp.2024.112274>
- [28] Z. Dong, H. An, S. Liu, S. Ma, Y. G. H. Pan and Ch. Ai, "Domain adversarial transfer fault diagnosis method of an axial piston pump based on a multi-scale attention mechanism", *Meas. J. Int. Meas. Confed.*, vol. 239, no. July 2024, p. 115455, 2025, <https://doi.org/10.1016/j.measurement.2024.115455>
- [29] D. Ling, T. Jiang, Y. Zheng and Y. Wang, "An adaptive mode identification and fault detection scheme for nonlinear multimode process monitoring using improved DPC-KPCA", *J. Taiwan Inst. Chem. Eng.*, vol. 169, no. January, p. 105915, 2025, <https://doi.org/10.1016/j.jtice.2024.105915>