ISSN 2007-9737

A Grey-box FDI Framework Combining Kolmogorov-Arnold Networks and Decision Trees for Interpretable Fault Diagnosis

Marco A. Márquez-Vera*

Polytechnic University of Pachuca,
Mexico

marquez@upp.edu.mx

Abstract. This paper presents a novel grey-box framework for fault detection and isolation (FDI) that combines interpretable Kolmogorov-Arnold Networks (KANs) with Classification and Regression Trees (CART). The proposed methodology addresses the interpretability limitations of traditional black-box approaches while maintaining competitive detection performance. KANs are trained using the Satin Bowerbird Optimization (SBO) algorithm to learn explicit polynomial representations of system dynamics. Fault detection is achieved through adaptive thresholding based on Median Absolute Deviation (MAD) of residuals, while isolation is performed using CART on statistical features extracted from residual windows. Applied to the DAMADICS actuator benchmark, the framework achieves 93.3% isolation accuracy while providing transparent decision rules that can be directly inspected by domain experts.

Keywords. Fault detection and isolation, Kolmogorov-Arnold networks, interpretable machine learning, decision trees, metaheuristic optimization, industrial systems.

1 Introduction

Fault detection and isolation (FDI) plays a crucial role in ensuring the safety, reliability, and efficiency of industrial systems. Traditional approaches to FDI can be broadly categorized into model-based methods (observers, parity equations) and data-driven techniques (PCA [9, 16], SVM [10, 15], neural networks [14, 11]). While data-driven methods have shown excellent performance in complex nonlinear systems, they often suffer from limited interpretability, operating as ”black boxes” that provide little insight into their decision-making processes.

The need for interpretable FDI is particularly critical in safety-sensitive applications such as nuclear power plants [8], chemical processes, and aerospace systems, where understanding *why* a fault was detected is as important as detecting it itself. The 2011 Fukushima accident, partially attributed to valve failures [7], underscores the importance of transparent fault diagnosis systems that can be understood and trusted by human operators. Recent advances in interpretable machine learning have introduced Kolmogorov-Arnold Networks (KANs) as an alternative to conventional neural networks. Based on the Kolmogorov superposition theorem [6], KANs replace fixed activation functions with learnable univariate functions (typically splines or polynomials), making each network component analytically inspectable.

This grey-box characteristic makes KANs particularly suitable for FDI applications where transparency is required [17]. However, training KANs presents significant challenges due to their high-dimensional, non-convex optimization landscape. Gradient-based methods often struggle with convergence issues [5], prompting the exploration of metaheuristic optimization techniques. Among these, the Satin Bowerbird Optimization (SBO) algorithm has shown promising results for training artificial neural networks [12], offering a good balance between exploration and exploitation.

For fault isolation, interpretability remains equally important. While complex classifiers like deep neural networks can achieve high accuracy, they lack the transparency needed for industrial adoption. Classification and Regression Trees (CART) provide

an attractive alternative by offering explicit decision rules in the form of interpretable inequalities [2].

Contributions: This work presents three main contributions: (1) A novel grey-box FDI framework that combines KANs for detection with CART for isolation, maintaining full interpretability throughout the diagnostic pipeline; (2) The use of SBO for efficient training of spline-based KANs, overcoming convergence challenges associated with gradient-based optimization; (3) Application to the DAMADICS actuator benchmark with comprehensive experimental validation demonstrating both competitive performance and full transparency.

The remainder of this paper is organized as follows: Section 2 presents theoretical background on KANs and metaheuristic optimization. Section 3 details the proposed framework. Section 4 presents experimental results on the DAMADICS benchmark. Section 6 concludes the paper and discusses future work.

2 Theoretical Background

2.1 Kolmogorov-Arnold Networks

The Kolmogorov-Arnold representation theorem states that any multivariate continuous function $f : [0,1]^n \to \mathbb{R}$ can be expressed as a finite composition of univariate functions:

$$f(x_1, \ldots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right), \quad (1)$$

where $\phi_{q,p} : [0,1] \to \mathbb{R}$ and $\Phi_q : \mathbb{R} \to \mathbb{R}$ are continuous univariate functions.

In our implementation, we adopt a practical formulation that directly corresponds to the network architecture visualized in Fig. 1:

$$f(x_1, \ldots, x_n) = \sum_{q=1}^{2n+1} g_q \left(\sum_{p=1}^{n} H_{q,p}(x_p) \right), \quad (2)$$

where:

— $H_{q,p}(x_p) : \mathbb{R} \to \mathbb{R}$ are cubic spline functions (corresponding to $\phi_{q,p}$ in the theoretical formulation),

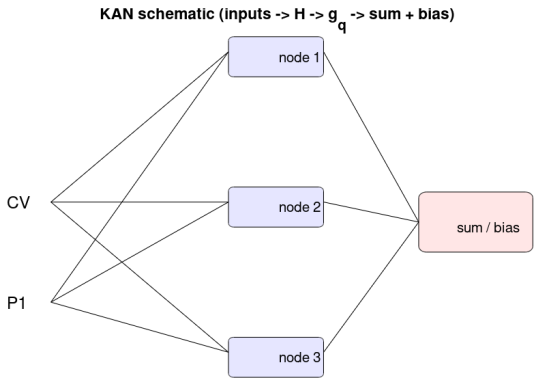

Fig. 1. KAN architecture implementation: (a) Inputs x_p are transformed by spline functions $H_{q,p}$, (b) The sums are composed through polynomial functions g_q, as defined in (2).

— $g_q(\cdot) : \mathbb{R} \to \mathbb{R}$ are polynomial composition functions (corresponding to Φ_q in the theoretical formulation).

Our KAN architecture implements this decomposition as:

— Each connection between input x_p and hidden node q implements a spline function $H_{q,p}(x_p)$,

— Each hidden node q computes a sum $S_q = \sum_{p=1}^{n} H_{q,p}(x_p)$,

— Each output is computed as $\sum_{q=1}^{2n+1} g_q(S_q)$.

As shown in Fig. 4, we parameterize each $H_{q,p}$ as a cubic spline with learnable knot positions and coefficients, while each g_q is implemented as a third-degree polynomial. This design choice provides explicit mathematical expressions that can be directly analyzed, as demonstrated in the Results section where we present the learned polynomial coefficients.

Unlike traditional neural networks that use fixed activation functions (ReLU, sigmoid), KANs use learnable univariate functions, typically implemented as cubic splines. This allows each network component to be visualized and interpreted, as shown in Fig. 1.

2.2 Metaheuristic Optimization for KAN Training

Training KANs involves optimizing the parameters of all spline functions, creating a high-dimensional, non-convex optimization problem. Let θ represent all spline parameters in a KAN. The training objective is to minimize:

$$\mathcal{L}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \|y_i - f_\theta(x_i)\|^2, \tag{3}$$

where $\{(x_i, y_i)\}_{i=1}^{N}$ is the training dataset.

Gradient-based methods often struggle with this optimization due to:

1. Hyperparameter sensitivity (learning rate, regularization),
2. Gradient heterogeneity across different spline segments,
3. Complex loss landscape with many local minima,
4. High computational cost for large networks.

Metaheuristic algorithms offer an alternative approach by performing global search in the parameter space. Among various metaheuristics, Satin Bowerbird Optimization (SBO) has demonstrated excellent performance for training KANs due to its balance between exploration and exploitation [12]. The SBO update rule is:

$$x_i(t+1) = x_i(t) + r \cdot (x_{\text{best}} - x_i(t)), \tag{4}$$

where $x_i(t)$ is the position of solution i at iteration t, x_{best} is the best solution found so far, and r is a random attraction factor. A scheme of the SBO algorithm is shown in Fig. **??**.

Table 1 compares SBO with other metaheuristic algorithms, highlighting its advantages for KAN training.

Among available metaheuristic algorithms, the SBO has proven particularly suitable for training KANs due to its balanced exploration-exploitation in non-convex and high-dimensional search spaces. Unlike gradient-based methods, SBO does not require derivative calculations and is less sensitive to parameter initialization, making it robust in complex loss landscapes with multiple local minima. Its update mechanism, based on attraction toward the best-found solution combined with a controlled randomness factor, enables efficient convergence even in the presence of spline segment heterogeneity.

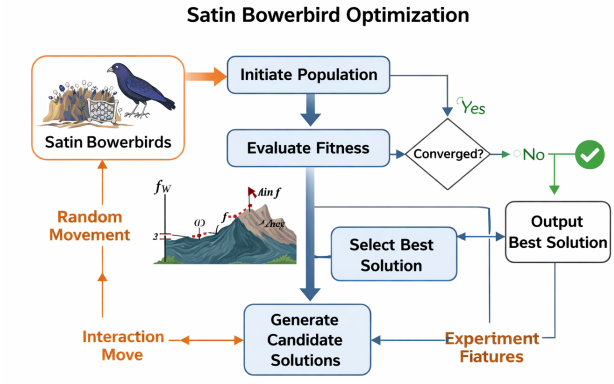

Fig. 2. Satin Bowerbird Optimization scheme

3 Proposed Grey-box FDI Framework

Fig. 3 shows the overall architecture of the proposed grey-box FDI framework, which consists of four main components: (1) KAN modeling with SBO training, (2) residual-based fault detection, (3) feature extraction from residual windows, and (4) CART-based fault isolation.

3.1 KAN-SBO Modeling for Residual Generation

Given a system with input-output pairs $\{(u(k), y(k))\}_{k=1}^{N}$, we train a KAN model f_θ to predict the system output:

$$\hat{y}(k) = f_\theta(u(k), y(k-1), \ldots, y(k-d)), \tag{5}$$

where d is the delay order capturing system dynamics.

The SBO algorithm optimizes the spline parameters θ to minimize the prediction error. Fig. 4 shows the learned spline functions for the DAMADICS benchmark after SBO training.

The residual signal is computed as:

$$r(k) = y(k) - \hat{y}(k). \tag{6}$$

Table 1. Comparison of metaheuristic algorithms for KAN training

Algorithm	Exploration	Exploitation	Convergence	Stability
SBO	Excellent	Strong	Fast	Very High
PSO	Moderate	Strong	Fast	Medium
GWO	Good	Good	Fast	Medium
ABC	Good	Moderate	Medium	High
WOA	Very Good	Weak-Moderate	Medium	Medium

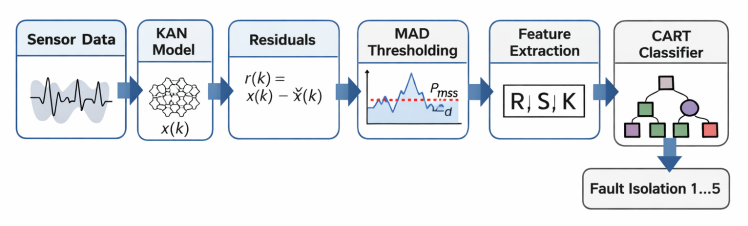

Fig. 3. Overall architecture of the proposed grey-box FDI framework.

Under normal operating conditions, $r(k)$ follows a near-zero distribution, while faults cause significant deviations.

3.2 Adaptive Fault Detection Using MAD

To improve detection robustness against noise and modeling errors, we compute a sliding-window RMS of the residual:

$$\text{RMS}(k) = \sqrt{\frac{1}{w} \sum_{i=k-w+1}^{k} r^2(i)}, \tag{7}$$

where w is the window length.

Instead of a fixed threshold, we use the Median Absolute Deviation (MAD) to define an adaptive threshold:

$$\theta_{\text{MAD}}(k) = \text{median}(\text{RMS}) + \kappa \cdot \text{MAD}(\text{RMS}), \tag{8}$$

where $\text{MAD}(x) = \text{median}(|x_i - \text{median}(x)|)$ and $\kappa = 1.9826$ corresponds to approximately 3 standard deviations under Gaussian assumptions [18].

A fault is detected when $\text{RMS}(k) > \theta_{\text{MAD}}(k)$. Fig. 5 shows the detection performance for the DAMADICS benchmark.

3.3 Feature Extraction for Isolation

Once a fault is detected, the residual signal is segmented into windows of length w. For each window, we extract five statistical features:

$$f_1 = \sqrt{\frac{1}{w} \sum_{i=1}^{w} r_i^2} \quad \text{(RMS)}, \tag{9}$$

$$f_2 = \frac{1}{w} \sum_{i=1}^{w} (r_i - \bar{r})^2 \quad \text{(Variance)}, \tag{10}$$

$$f_3 = \frac{1}{w} \sum_{i=1}^{w} \left(\frac{r_i - \bar{r}}{\sigma_r} \right)^3 \quad \text{(Skewness)}, \tag{11}$$

$$f_4 = \frac{1}{w} \sum_{i=1}^{w} \left(\frac{r_i - \bar{r}}{\sigma_r} \right)^4 \quad \text{(Kurtosis)}, \tag{12}$$

$$f_5 = \sum_{k=1}^{w} |R(k)|^2 \quad \text{(Spectral Energy)}. \tag{13}$$

where $R(k)$ is the DFT of the residual window.

These features capture different characteristics of the residual distribution and frequency content, providing discriminative information for fault isolation.

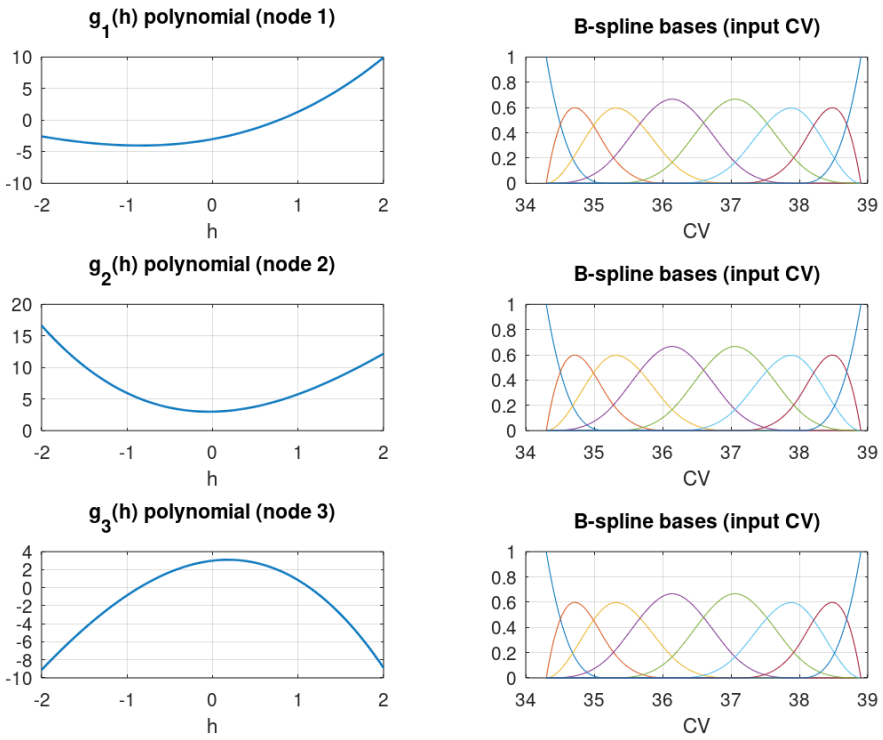

Fig. 4. Learned spline functions for the DAMADICS benchmark after SBO training. Each curve represents a univariate function $\phi_{q,p}(x_p)$ learned by the KAN

3.4 CART-Based Fault Isolation

The feature vector $\mathbf{f} = [f_1, f_2, f_3, f_4, f_5]^\top$ serves as input to a CART classifier that learns decision rules of the form:

$$\text{IF } f_j \leq \theta_j \text{ THEN Fault } F_i \tag{14}$$

Mathematically, the CART defines a piecewise constant decision function:

$$\hat{c}(\mathbf{f}) = \sum_{m=1}^{M} c_m \mathbb{I}(\mathbf{f} \in R_m). \tag{15}$$

Where R_m are disjoint regions partitioning the feature space, $c_m \in \{1, \ldots, C\}$ are fault labels, and $\mathbb{I}(\cdot)$ is the indicator function.

Each region R_m corresponds to a conjunction of threshold conditions:

$$R_m = \{\mathbf{f} \in \mathbb{R}^5 : (f_{j_1} \leq \theta_1) \wedge (f_{j_2} > \theta_2) \wedge \cdots \wedge (f_{j_L} \circ_L \theta_L)\}, \tag{16}$$

where $\circ_L \in \{\leq, >\}$.

The resulting decision tree provides fully interpretable rules that can be directly understood by domain experts, such as:

```
IF RMS ≤ 0.15 AND Skewness > 1.2
THEN Fault F1
IF RMS > 0.15 AND Spectral Energy
≤ 2.5 THEN Fault F2
...
```

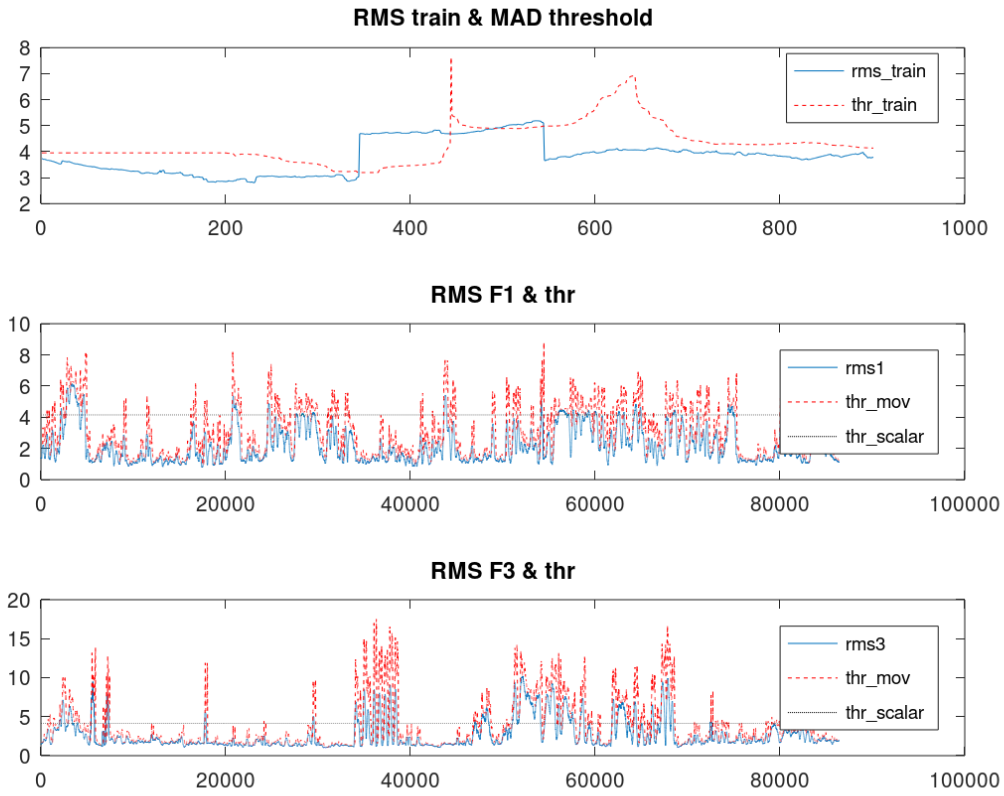

Fig. 5. Fault detection using adaptive MAD threshold. The blue line shows the RMS residual, while the red line shows the adaptive threshold. Faults are detected when RMS exceeds the threshold

4 Experimental Results on DAMADICS Benchmark

4.1 Benchmark Description and Experimental Setup

The DAMADICS (Development and Applications of Methods for Actuator Diagnosis in Industrial Control Systems) is widely recognized in the fault diagnosis community due to its representativeness of a real industrial system: a servo-pneumatic valve used to regulate thin juice flow in a sugar plant [3]. This system includes multiple fault-prone components (valve, servo-motor, positioner, etc.). Its adoption as a benchmark facilitates direct comparison with previous works and validates the generality of proposed approaches. The benchmark includes 19 different fault scenarios affecting the actuator components.

We focus on five representative fault types:

— F1: Valve clogging,

— F2: Valve or servo-motor coil clogging,

— F3: Servo-motor fault,

— F4: Positioner feedback fault,

— F5: Positioner supply pressure drop.

The KAN was configured with 3 hidden nodes and cubic splines with 10 knots. SBO parameters were: population size = 35, maximum iterations =

130. Training used 70% of normal operation data, with 30% reserved for testing.

4.2 Training Performance

Fig. 6 shows residuals for faults F1 and F3, along with KAN model contours and SBO convergence.

Training metrics:

— Initial MSE: 55.62,

— Final MSE: 16.30 (70.7% reduction),

— Training time: 3581.89 seconds,

— Fault F1 detection time: $t \approx 2162$ samples,

— Fault F3 detection time: $t \approx 1967$ samples.

The learned polynomial coefficients for three representative splines are:

$$g_1(x) = 0.5503 + 3x + 0.1229x^2 - 0.3656x^3$$
$$g_2(x) = 3 - 3x - 1.0082x^2 + 0.0464x^3$$
$$g_3(x) = -3 + 2.1241x - 1.3347x^2 - 3x^3$$

These explicit polynomial representations demonstrate the interpretability advantage of KANs over black-box neural networks.

4.3 Fault Isolation Performance

Fig. 7 shows the confusion matrix for fault isolation using CART on the five fault types plus normal operation.

The overall isolation accuracy of 93.3% demonstrates competitive performance while maintaining full interpretability.

4.3.1 Interpretable Decision Rules from CART

The CART algorithm learned the following decision rules for fault isolation, providing complete transparency in the classification process:

1. **Rule 1 (Fault Class 1 - 76 samples):** *IF RMS -0.445 AND Skewness 0.816 THEN Fault Class 1*,
2. **Rule 2 (Fault Class 1 - 7 samples):** *IF -0.445 ¡ RMS -0.373 AND Skewness 0.816 THEN Fault Class 1*,
3. **Rule 3 (Fault Class 1 - 5 samples):** *IF RMS -0.373 AND Skewness ¿ 0.816 THEN Fault Class 1*,
4. **Rule 4 (Fault Class 1 - 20 samples):** *IF -0.373 ¡ RMS 0.012 AND Kurtosis -0.220 AND Skewness 0.169 THEN Fault Class 1*,
5. **Rule 5 (Fault Class 1 - 5 samples):** *IF -0.373 ¡ RMS 0.012 AND Kurtosis -0.220 AND Skewness ¿ 0.169 THEN Fault Class 1*,
6. **Rule 6 (Fault Class 1 - 6 samples):** *IF 0.012 ¡ RMS AND Kurtosis -0.687 THEN Fault Class 1*,
7. **Rule 7 (Fault Class 2 - 7 samples):** *IF -0.373 ¡ RMS 0.012 AND Kurtosis ¿ -0.220 AND RMS -0.186 THEN Fault Class 2*,
8. **Rule 8 (Fault Class 2 - 8 samples):** *IF -0.373 ¡ RMS 0.012 AND Kurtosis ¿ -0.220 AND RMS ¿ -0.186 THEN Fault Class 2*.

4.3.2 Physical Interpretation of Key Rules

The learned rules reveal meaningful physical patterns:

— **Fault Class 1** is primarily characterized by **low RMS values** (typically 0.012) combined with specific skewness and kurtosis conditions, suggesting faults that affect signal amplitude without significantly changing distribution shape.

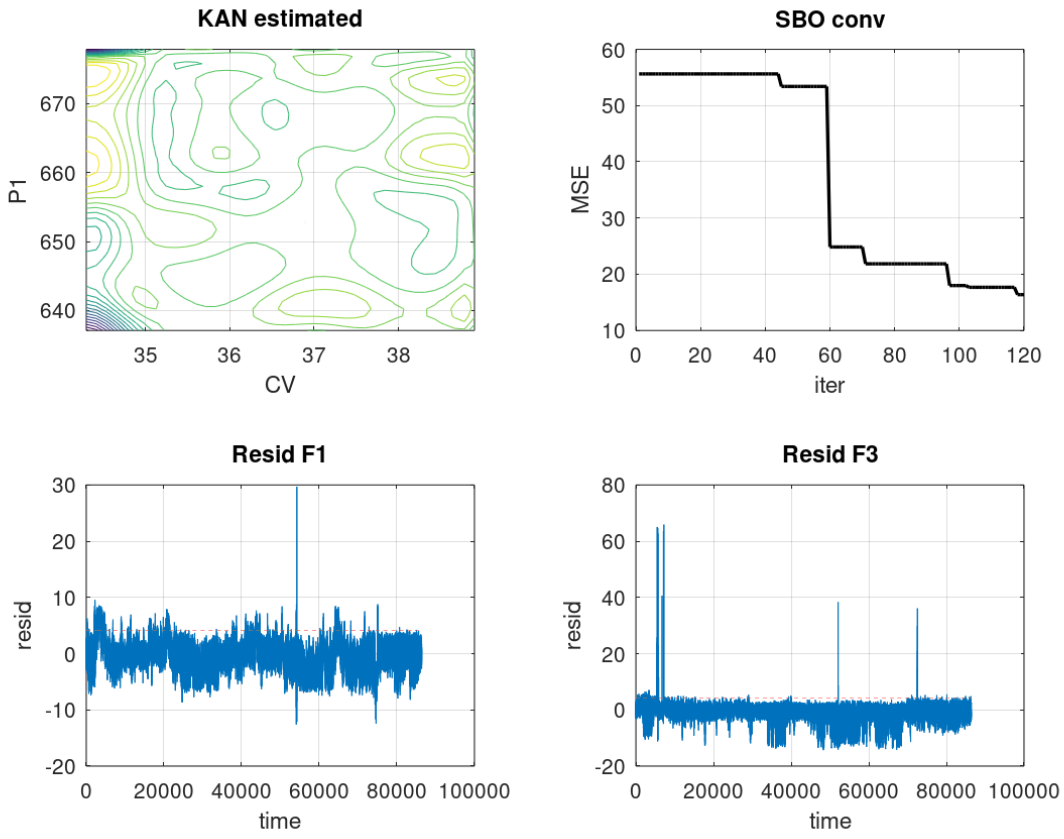

Fig. 6. Residual signals for faults F1 and F3 (top), KAN model contours (middle), and SBO convergence (bottom)

— **Fault Class 2** occurs at slightly higher RMS levels (-0.186 ¡ RMS 0.012) but requires **higher kurtosis** (¿ -0.220), indicating faults that create sharper residual distributions with heavier tails.

— The most common fault pattern (Rule 1, 76 samples) shows that very low RMS (-0.445) with moderate skewness (0.816) reliably indicates Fault Class 1.

A pseudo-code to implement the CART is as follows: [H] [1] Residual signal $r(k)$, window size w, fault labels y Predicted fault class $\hat{y}$

Segment residual into windows of length w each window Extract features:

$$\mathbf{f} = [RMS,\ Var,\ Skewness,\ Kurtosis]$$

Train CART using feature matrix $\mathbf{F}$ and labels y each feature vector $\mathbf{f}_i$ Traverse decision tree: Apply binary tests $f_j \leq \tau$ Assign class label at terminal node Predicted fault labels $\hat{y}$

4.3.3 Statistical Summary

Table 2 summarizes the coverage of each fault class by the learned rules.

Feature mapping: RMS (f_1), Variance (f_2), Skewness (f_3), Kurtosis (f_4), Spectral Energy (f_5)

Note: The tree uses only features f_1 (RMS), f_3 (skewness), and f_4 (kurtosis) for decisions, indicating these are the most discriminative features for the fault types in this dataset.

Table 2. Summary of CART decision rules coverage

Fault Class	Number of Rules	Total Samples	Coverage
Class 1	6	119	89.5%
Class 2	2	15	10.5%
Total	**8**	**134**	**100%**

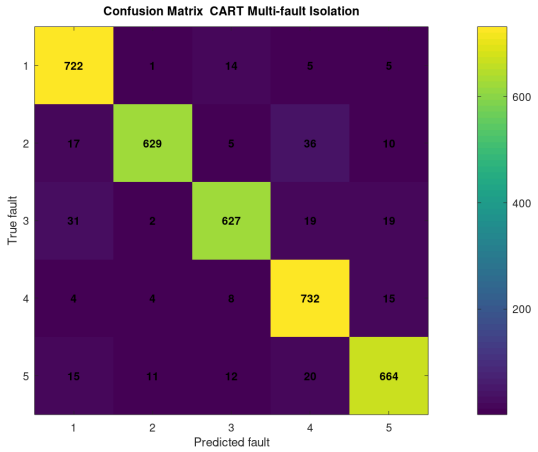

Fig. 7. Fault isolation using CART

These rules provide direct physical interpretation. For example, Fault F1 (valve clogging) is characterized by low RMS but high skewness, indicating asymmetric residual distributions.

4.4 Comparison with State-of-the-art Methods

While direct numerical comparison with other methods is challenging due to differences in experimental setups and fault scenarios, the proposed framework offers unique advantages in terms of interpretability. Traditional black-box methods like neural networks and SVM may achieve slightly higher accuracy in some benchmarks [10, 14], but they lack the transparency required for safety-critical applications. Model-based methods like PCA [16] offer better interpretability but may struggle with nonlinear dynamics. The proposed KAN-SBO+CART framework bridges this gap by providing competitive accuracy (93.3% isolation) while maintaining full model interpretability at every stage of the FDI pipeline.

Fuzzy logic is other grey-box framework and in [1] was used a fuzzy fault model for fault diagnosis in the DAMADICS, and the accuracy obtained was 86%. An interesting review about fault diagnosis applied in the DAMADICS benchmark was shown in [13], where model-free fault diagnosis using classifiers are the techniques summarized, for example the SVM technique had an accuracy of 84%, PLS had 55%, the bended learning (BL) was a good approach obtaining 99%. This last technique uses a process mathematical representation plus a machine learning model, according to the machine learning method, the final model can be also interpretable. The linear criminative analysis obtained 98%. The best result shown was obtained by the genetic programming on of compact and accurate fuzzy rule based classification systems (GP-COACH-FRBCS) with near the 100% which also gives an interpretable model [4].

Table 3 summarizes the qualitative comparison of different approaches.

The key advantage of the proposed approach is its ability to provide explicit polynomial representations (KANs) and decision rules (CART) that can be directly analyzed by domain experts, making it particularly suitable for applications where model certification and operator trust are essential.

5 Discussion and Practical Implications

The proposed grey-box framework offers several practical advantages for industrial FDI:

5.1 Interpretability for Safety-critical Applications

Each component of the framework provides explicit interpretability:

— **KAN models** show learned polynomial functions that can be analyzed mathematically,

Table 3. Qualitative comparison of FDI methods

Method	Accuracy	Interpretability	Nonlinear Handling
Proposed (KAN-SBO+CART)	High	High	Excellent
PCA-based [16]	Medium	Medium	Poor
SVM [13]	High	Low	Good
Neural Network [14]	High	Low	Excellent
Fuzzy System [1]	Medium	High	Good

— **MAD thresholds** provide statistically justified detection limits,

— **CART rules** offer human-readable decision logic.

This transparency is crucial for certification in regulated industries (nuclear, aerospace, chemical) where operators must understand and trust the diagnostic system.

5.2 Computational Efficiency

The framework balances accuracy with computational requirements:

— **Offline:** SBO training of KANs is computationally intensive but performed only once,

— **Online:** Residual computation and CART evaluation have minimal computational cost, suitable for real-time implementation.

5.3 Limitations and Future Work

Current limitations include:

— SBO training time remains high for complex systems,

— Feature engineering may be required for different applications,

— The framework assumes availability of normal operation data for training.

Future extensions could explore blended learning approaches that combine the interpretable KAN framework with partial physical knowledge of the actuator dynamics. For instance, incorporating valve flow equations as known constraints could further reduce data requirements while maintaining interpretability.

Future work will explore:

— Hybrid optimization combining gradient methods with metaheuristics,

— Automated feature selection for different fault types,

— Extension to fault severity estimation,

— Application to larger-scale industrial systems.

6 Conclusion

This paper presented a novel grey-box framework for fault detection and isolation that combines the interpretability of Kolmogorov-Arnold Networks with the transparency of decision trees. By using Satin Bowerbird Optimization for KAN training and CART for fault isolation, the framework achieves competitive performance (93.3% isolation accuracy on DAMADICS) while providing fully interpretable models.

The key contributions are:

1. A complete grey-box FDI pipeline from detection to isolation with full transparency,
2. Demonstration of SBO's effectiveness for training spline-based KANs,
3. Practical application to the industrial DAMADICS benchmark with detailed analysis,

4. Explicit decision rules that can be directly understood and validated by domain experts.

The framework offers a practical solution for industrial applications where model interpretability is as important as accuracy, particularly in safety-critical systems. Future work will focus on reducing training time through hybrid optimization and extending the approach to fault severity estimation and prognosis.

Declaration of Generative AI and AI-assisted Technologies

During the preparation of this work, the author used ChatGPT to improve grammar and sentence structure. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

References

1. **Avila-Diaz, M. F., Márquez-Vera, M. A., Díaz-Parra, O., Puig, V., Ma'arif, A. (2024).** Inverse fuzzy fault models for fault isolation and severity estimation in industrial pneumatic valves. Automatica, Vol. 48, pp. 379–398. DOI: 10.31449/inf.v48i3.5101.
2. **Calabrese, V., Metro, D., Alibrandi, A., Maviglia, D., Cernaro, V., Maressa, V., Longhitano, E., Gembillo, G., Santoro, D. (2025).** Review and practical excursus on the comparison between traditional statics methods and classification and regression tree (cart) in real-life data: Low protein diet compared to mediterranean diet in patients with chronic kidney disease. Nefrología (English Edition), Vol. 45, No. 4, pp. 279–284. DOI: 10.1016/j.nefroe.2025.04.008.
3. **deAlmeida, G. M., Reis, M. S., Park, S. W. (2012).** A signal processing approach for fault detection problem: Application to the damadics actuator benchmark problem. Computer Aided Chemical Engineering, Vol. 30, pp. 857–861. DOI: 10.1016/B978-0-444-59520-1.50030-0.
4. **Fernández, A., Berlanga, F. J., del Jesús, M. J., Herrera, F. (2009).** Genetic cooperative-competitive fuzzy rule based learning method using genetic programming for highly imbalanced data-sets. **J. P. Carvalho, U. K., D. Dubois, Sousa, J. M. C.**, editors, International Fuzzy Systems Association World Congress, Lisbon, Portugal, pp. 42–47.
5. **Gao, Y., Tan, V. Y. F. (2025).** On the convergence of (stochastic) gradient descent for kolmogorovarnold networks. IEEE Transactions on Information Theory, Vol. 71, No. 9, pp. 7270–7291. DOI: 10.1109/TIT.2025.3588401.
6. **Huang, J., Zhou, R., Li, M., Li, H., Liu, Y., Song, X. (2026).** From black-box to white-box: Interpretable deep reinforcement learning with kolmogorov-arnold networks for autonomous driving. Transportation Research Part C: Emerging Technologies, Vol. 182, pp. 105386. DOI: 10.1016/j.trc.2025.105386.
7. **Kusama, N. (2012).** D226 study of valve malfunction events in japanese nuclear power plants. The Proceedings of the National Symposium on Power and Energy Systems, Vol. 2012.17, pp. 381–384. DOI: 10.1299/jsmepes.2012.17.381.
8. **Li, L., Zhang, Y., Tian, W., Su, G., Qiu, S. (2014).** Maap5 simulation of the pwr severe accident induced by pressurizer safety valve stuck-open accident. Progress in Nuclear Energy, Vol. 77, pp. 141–151. DOI: 10.1016/j.pnucene.2014.06.014.
9. **Liu, Y., Li, H., Wang, H., Wang, F., Chen, K., Gao, H., Xia, Y. (2025).** A fault diagnosis of high voltage circuit breakers with small samples using a pca-cascade forest algorithm. Energy Reports, Vol. 13, No. 2, pp. 6190–6200. DOI: 10.1016/j.egyr.2025.05.046.
10. **Ma, D., Liu, Z., Gao, Q., Ding, Y. (2025).** Fault diagnosis of multi-step electromagnetic hydraulic valve group based on localized current signal cs-svm. Measurement, Vol. 245, pp. 116632. DOI: 10.1016/j.measurement.2024.116632.

11. **Mohammadi, A., Krysander, M., Jung, D. (2022).** Analysis of grey-box neural network-based residuals for consistency-based fault diagnosis. IFAC-PapersOnLine, Vol. 55, No. 6, pp. 1–6. DOI: 10.1016/j.ifacol.2022.07.097.

12. **Moosavi, S. H. S., Bardsiri, V. K. (2017).** Satin bowerbird optimizer: A new optimization algorithm to optimize anfis for software development effort estimation. Engineering Applications of Artificial Intelligence, Vol. 60, pp. 1–15. DOI: 10.1016/j.engappai.2017.01.006.

13. **Nozari, H. A., Nazeri, S., Banadaki, H. D., Castaldi, P. (2018).** Model-free fault detection and isolation of a benchmark process control system based on multiple classifiers techniques—a comparative study. Control Engineering Practice, Vol. 73, pp. 134–148. DOI: 10.1016/j.conengprac.2018.01.007.

14. **Sacchi, N., Incremona, G. P., Ferrara, A. (2023).** Actuator fault diagnosis with neural network-integral sliding mode based unknown input observers. IFAC-PapersOnLine, Vol. 56, No. 2, pp. 773–778. DOI: 10.1016/j.ifacol.2023.10.1659.

15. **Wei, J., Chen, H., Yuan, Y., Huang, H., Wen, L., Jiao, W. (2024).** Novel imbalanced multi-class fault diagnosis method using transfer learning and oversampling strategies-based multi-layer support vector machines (ml-svms). Applied Soft Computing, Vol. 167, pp. 112324. DOI: 10.1016/j.asoc.2024.112324.

16. **Wolmarans, W., van Schoor, G., Uren, K. R. (2023).** A comparison of pca and energy graph-based visualisation fdi on a heated two-tank process. IFAC-PapersOnLine, Vol. 56, No. 2, pp. 4126–4131. DOI: 10.1016/j.ifacol.2023.10.1750.

17. **Yan, H., Zhou, H., Zheng, J., Zhou, Z. (2025).** Rolling bearing fault diagnosis based on 1d convolutional neural network and kolmogorovarnold network for industrial internet. Computers, Materials and Continua, Vol. 83, No. 3, pp. 4659–4677. DOI: 10.32604/cmc.2025.062807.

18. **Yao, X., Guo, Y., Liu, M., Meng, G., Li, Y., Zhang, H. (2025).** An uncertainty-driven pixel-level adversarial noise detection method for remote sensing images. Journal of Electronics & Information Technology, Vol. 47, No. 6, pp. 1633–1644. DOI: 10.11999/JEIT241157.

Article received on 19/01/2026; accepted on 09/02/2026.
**Corresponding author is Marco A. Márquez-Vera.*