Сучасний стан наукових досліджень та технологій в промисловості. 2026. № 2 (36) *ISSN 2522-9818 (print)*
Innovative technologies and scientific solutions for industries. 2026. No. 2 (36) *ISSN 2524-2296 (online)*

UDC 004.85:681.518:621.311.243 DOI: https://doi.org/10.30837/2522-9818.2026.2.153

Pavlo Radiuk, Anatoliy Sachenko, Oleksandr Melnychenko, Ruslan Brukhanskyi, Antonina Kashtalian

WEDD-RST: AN INTERPRETIVE APPROACH TO DETECTING DEFECTS IN PHOTOVOLTAIC PANELS IN CYBER-PHYSICAL SYSTEMS

The **subject** of this study focuses on the interpretability of fault-detection frameworks in solar cyber-physical systems. The rapid expansion of renewable energy networks generates massive volumes of continuous sensor data, requiring sophisticated monitoring strategies to prevent safety hazards such as photovoltaic fires. Nevertheless, dominant deep learning models function as opaque black boxes, providing no transparent reasoning for critical safety decisions. The **goal** of this work is to enhance the interpretability of anomaly detection in cyber-physical environments by formulating a rule-based production knowledge base straight from continuous sensor readings. Key **tasks** include developing an adaptive discretisation method, identifying minimal feature subsets, and reconciling contradictory patterns using an integral class support score. The applied **methods** fuses rough set theory (RST) with an innovative weighted entropy-density discretization (WEDD) algorithm. This combined pipeline fine-tunes thresholds using a dual metric of information entropy and local probability density, utilizing kernel density estimation to position cut points within natural data valleys. Deterministic rules are derived from the lower approximation, whereas probabilistic rules are generated from the boundary region. The **results** illustrate the substantial effectiveness of the suggested approach. Tested on a simulated SCADA dataset designed for fire hazard identification, the framework attains an overall accuracy of 96.2% and a macro F1-score of 0.960. Significantly, it delivers 100% accuracy for deterministic rules and a 98.0% recall rate for the vital Fire Hazard category. Comparative evaluations on the Iris and Wine datasets yield competitive results, achieving 93.3% and 87.6% accuracy, respectively, when compared with Decision Tree and Naive Bayes models. The generated knowledge base is a compact JSON artifact, making it well-suited for edge devices with limited computational resources. In **conclusion**, this research establishes the WEDD-RST approach as a rigorous approach for transforming raw sensor data into fully auditable IF-THEN rules with explicit confidence scores, offering a highly reliable solution for automated safety monitoring in cyber-physical environments.

Keywords: production knowledge base; rough set theory; discretization; information granulation; reduct; classification; cyber-physical systems; photovoltaic monitoring; SCADA.

Introduction

The rapid spread of cyber-physical systems (CPS) in industrial settings, along with the shift away from traditional distributed measurement networks [1] – particularly in renewable energy systems – has generated massive streams of continuous sensor data that require intelligent analysis for real-time decision-making [2]. Photovoltaic (PV) power plants are a critical operational area where thermal anomalies in panels, such as overheated bypass diodes and isolated hot spots, can cause catastrophic fires if not detected and classified quickly [3, 4]. Modern photovoltaic plant monitoring systems utilize Supervisory Control and Data Acquisition (SCADA) platforms that integrate various types of sensors, including thermal imagery from unmanned aerial vehicle (UAV) systems, bypass diode temperature monitors, and fire alarm detectors [5, 6]. Although artificial intelligence boasts extraordinary accuracy in certain data-driven fields, such as biometrics [7], the main challenge in these cyber-physical contexts remains the transformation of raw, continuous sensor readings into actionable insights that human operators can understand, verify, and rely on. Although deep learning approaches demonstrate outstanding detection results across various disciplines – from the YOLO family of architectures in computer vision [8] and the construction of adaptive UAV routes [9] to complex malware identification algorithms [10] – they operate as opaque "black boxes" that cannot be explained. In fields where safety is critical, such as fire risk assessment, this lack of transparency raises significant regulatory, practical, and ethical issues [11]. Personnel responding to a fire alarm must understand the system's logic for classifying hazards; receiving only a probability score from an opaque neural network is insufficient, as a clear understanding of the situation is necessary for correct human psychomotor responses and rapid decision-making in high-pressure situations [12]. Generating production rules based on sensor input data is a fundamental problem in inductive learning, a branch of machine learning based on symbolic computation [13, 14]. Given a feature matrix B (where rows denote measurement instances and columns denote sensor attributes) along with a class label vector Y

ISSN 2522-9818 (print) Сучасний стан наукових досліджень та технологій в промисловості. 2026. № 2 (36)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2026. No. 2 (36)

(denoting operating conditions), the primary goal is to formulate logical rules structured as

$$\textbf{IF}\ \left(\text{Feature}_1 \in \text{Range}_1\right)\ \textbf{AND}\ \left(\text{Feature}_2 \in \text{Range}_2\right)\ \ldots\ \textbf{THEN}\ \text{Class} = k,$$

are accompanied by corresponding measures of confidence (probability) and support (statistical significance).

Converting tabular datasets into logical structures requires solving a series of interrelated tasks: partitioning continuous variables into meaningful intervals, identifying outliers, resolving conflicting data models, and developing inference procedures for processing new input data. Traditional solutions to this dilemma include decision tree methods such as ID3 [14], C4.5 [15], and CART [16], as well as sequential covering methods (e.g., CN2, PRISM, RIPPER) and approaches based on the theory of approximate sets (RST) [17].

However, each has certain drawbacks:

Decision tree models construct rules based on locally optimal splits at each node, failing to ensure a globally optimal set of rules. As a result, the resulting rules may be overly complex or fail to account for important interactions between features [18]. On the other hand, sequential coverage methods generate rules explicitly but remain vulnerable to the order in which classes are processed, which can lead to overlaps or contradictions in rule sets.

Traditional RST methods provide a robust mathematical foundation for managing uncertainty using lower and upper bounds [13, 19]. However, they typically rely on basic equal-width or equal-frequency discretization, thereby ignoring the internal structure of the dataset. Entropy-based discretization, successfully used in [20], although effective in class separation, ignores the distribution of attribute values. This leads to the segmentation of points that break up organic data clusters, creating rules that are highly sensitive to noise.

The goal of this study is to improve the interpretability of fault detection in cyber-physical systems by creating rule-based knowledge bases (PKBs) directly from continuous sensor observations. This is achieved through a hybrid approach that combines advanced discretization with the theory of approximate sets (RST), effectively bridging the gap between artificial intelligence and interpretable industrial automation. The main scientific achievements of this work are as follows:

1. Weighted Entropy-Density Discretization (WEDD): a new multi-criteria discretization algorithm formulated to simultaneously minimize information entropy with respect to class labels and local probability density at boundary positions. This dual-objective strategy positions decision points in natural "valleys" that separate data clusters, creating rules that are highly robust to statistical outliers.

2. An optimized heuristic for identifying minimal subsets of features (reducts), which integrates the information capacity weights obtained during the discretization phase. This addition reduces the complexity of the rules by approximately 40–60% without compromising classification performance.

3. A conflict resolution framework based on an integral class support score that aggregates confidence in correctness, support metrics, and interval reliability weights. This configuration enables accurate classification even when multiple conflicting rules are triggered simultaneously. The remainder of this article is structured as follows: Section 2 reviews the relevant literature. Section 3 presents a comprehensive mathematical foundation for the PKB approach. Section 4 presents the experimental results. Section 5 discusses the results and their broader implications in detail. Finally, Section 6 provides concluding remarks on the study.

Literature Review and Problem Statement

Modern photovoltaic monitoring systems utilize "edge-cloud" architectures that integrate thermal images captured by unmanned aerial vehicles with SCADA systems to detect fire risks [4, 21]. Deep learning paradigms, particularly architectures from the YOLO family [8], are frequently used to detect thermal anomalies in photovoltaic panels [1]. Thakfan et al. [5] conducted an extensive study of artificial intelligence (AI)-based fault detection and diagnosis protocols for building energy networks. They highlighted the inevitable trade-off between prediction accuracy and model transparency, an obstacle that continually drives the development of interpretable diagnostic systems for decentralized energy networks [22].

RST provides a robust mathematical framework for processing fuzzy, uncertain, and partial datasets without the need for external parameters such as probability distributions or fuzzy membership functions [17].

Сучасний стан наукових досліджень та технологій в промисловості. 2026. № 2 (36) *ISSN 2522-9818 (print)*
Innovative technologies and scientific solutions for industries. 2026. No. 2 (36) *ISSN 2524-2296 (online)*

The basis of this theory is the principle of indistinguishability of identical elements, which divides the set of instances into separate equivalence classes according to their specific attributes [13].

Fundamental mathematical elements, including lower and upper approximations and the boundary region, facilitate the derivation of both deterministic and probabilistic conclusions from data sets [19]. Luo et al. [23] clearly defined the discrimination matrix and the discrimination function, laying the theoretical foundation for identifying minimal subsets of features (reducts) that preserve full classification capability.

A comprehensive compilation established the practical application of RST in intelligent decision-making [24], while Błaszczyński et al. [25] extended the methodology to account for multi-criteria decision analysis.

Subsequent research directly linked RST to data mining, establishing decision tables [26], decision rules, and reducts as the primary mechanisms for knowledge discovery. For example, the LERS framework [27] demonstrated effective rule induction using rough sets, paving the way for the development of PKB. Despite these breakthroughs, most RST applications rely on pre-discretized data, often neglecting the quality and methodology of the discretization process.

Discretization, i.e., the transformation of continuous variables into discrete categorical intervals, is an important preprocessing step for symbolic classification models [28]. The study [20] employed a recursive entropy-based discretization technique that relied on the minimum description length (MDLP) principle to determine stopping conditions. This strategy remains the industry standard for supervised discretization. Their approach identifies partition points that minimize the conditional entropy of class labels, thereby isolating threshold values that optimally distinguish between categories.

García et al. [28] compiled a detailed classification of discretization methods, grouping them by supervision level (supervised vs. unsupervised), partitioning logic (top-down vs. bottom-up), and evaluation metrics (entropy, error rate, or statistical measures). Toulabinejad et al. [29] demonstrated that supervised discretization significantly improves the performance of character-based classifiers compared to simple unsupervised binning. Furthermore, Huang et al. [30] investigated binning as a fundamental tool, detailing the theoretical relationships between binning quality and the subsequent accuracy of classification tasks.

A significant drawback of algorithms based solely on entropy is their inability to account for the density distribution of attribute values. As illustrated by Xun et al. [31] in their comparative analysis of entropy frameworks, different discretization metrics can yield radically different results for threshold placement and classification. This insight is the driving force behind the development of the multi-criteria methodology presented in this article, which improves upon traditional entropy metrics by incorporating density analysis.

Kernel density estimation (KDE) is a nonparametric technique for approximating probability density functions from raw data [32, 33]. Silverman's comprehensive text [34] outlines the principles of KDE, including methods for determining bandwidth. When applied to binning, KDE allows for the detection of "natural breaks" in data profiles – specifically, valleys between density peaks that indicate logical separations between distinct subpopulations. The WEDD algorithm uses this feature to precisely position discretization thresholds at the density minima of the dataset, ensuring that the generated intervals correspond to the fundamental structure of the dataset.

In SCADA systems, decision-making frameworks based on Boolean logic cross-reference sensor data, including thermal anomalies, bypass diode temperatures, and fire alarms, to classify operating conditions [35]. Combining fuzzy production rules with existing SCADA logic provides a unique opportunity to enhance the transparency and verifiability of automated fire risk assessments, serving as the primary driving force for the methodology detailed in this study. Guided by these practical operational requirements, the primary objective of this study is to enhance the transparency and reliability of automated fault identification in cyber-physical PV networks by formulating a rule-based PKB that relies exclusively on continuous sensor readings. To achieve this goal, three main tasks are performed: (i) developing an innovative WEDD algorithm designed to segment continuous attributes by weighting class differences with respect to the natural density of the data; (ii) extracting minimal, non-redundant feature sets (reducts) using an information-weighted rough set heuristic aimed at minimizing rule complexity; and (iii) formulating a robust conflict management protocol that uses integral class support metrics to generate practical deterministic and probabilistic IF-THEN rules, ideal for integration into SCADA systems with limited computational resources.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Сучасний стан наукових досліджень та технологій в промисловості. 2026. № 2 (36)
Innovative technologies and scientific solutions for industries. 2026. No. 2 (36)

Methods and Materials

The PKB approach consists of seven steps that transform raw sensor data into a self-contained, interpretable classification system. Figure 1 shows the complete process architecture.

Criterion 1. The information model of the classification task is represented as a decision-making system (DMS) described by a tuple:

$$\text{DMS} = \langle U, A \cup \{d\}, V, f \rangle, \quad (1)$$

where $U = \{x_1, x_2, \ldots, x_n\}$ is a finite set of objects (a space in which each object x_i corresponds to a row of the feature matrix B); $A = \{a_1, a_2, \ldots, a_m\}$ is a set of conditional attributes (features) corresponding to the columns *of B*; *d* is a decision attribute defined by the class label vector *Y*; $V = \bigcup_{a \in A \cup \{d\}} V_a$ is the domain of attribute values; $f : U \times (A \cup \{d\}) \to V$ is an information function that maps objects to the values of their attributes.

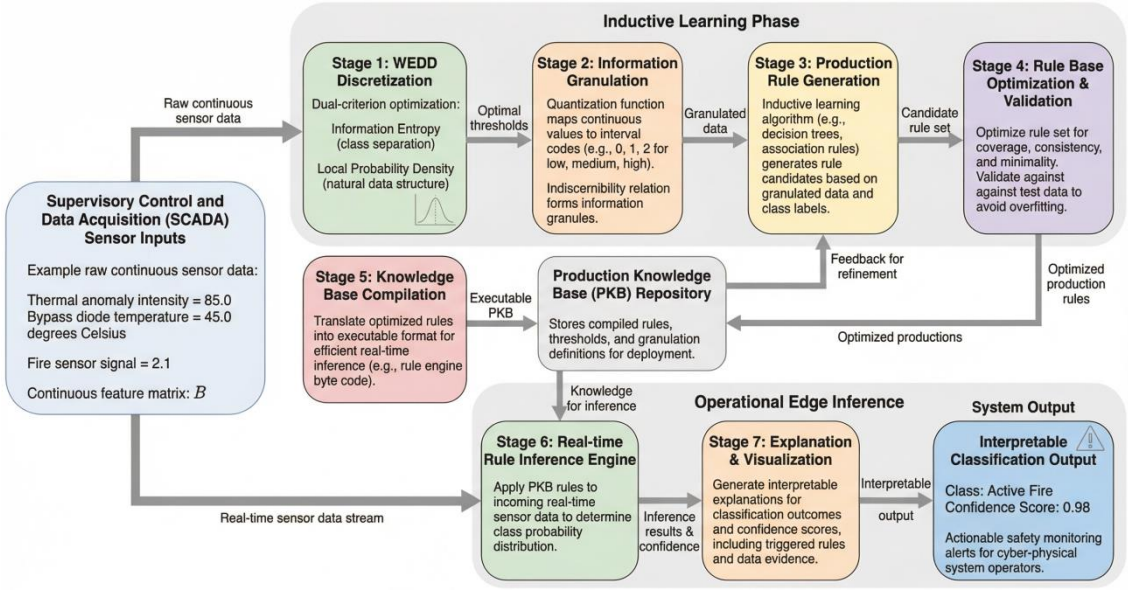

Fig. 1. Schematic architecture of the proposed PKB approach. The system consists of an inductive learning phase (steps 1–5) for data discretization, granularity, and rule generation, combined with an operational inference phase (steps 6–7), which uses the compiled rule base for interpretable real-time defect classification in cyber-physical systems

Fig. 2 shows a visual representation of the DMS formalization, illustrating how the feature matrix B and the class vector Y are mapped to the formal representation of a tuple.

Stage 1. Multi-criteria adaptive discretization (WEDD). Since the feature matrix B contains continuous values, these must be divided into intervals so that symbolic inference rules can be constructed. We propose the WEDD method, which simultaneously optimizes two criteria: information entropy (class separation) and local probability density (the natural structure of the data).

Criterion 2. Information entropy. For attribute $a \in A$ and threshold T for candidates, the classical conditional entropy is calculated as:

$$E(a, T; U) = \frac{|U_{\text{left}}|}{|U|} \text{Ent}(U_{\text{left}}) + \frac{|U_{\text{right}}|}{|U|} \text{Ent}(U_{\text{right}}), \quad (2)$$

where $U_{\text{left}} = \{x \in U : f(x, a) < T\}$ and $U_{\text{right}} = \{x \in U : f(x, a) \geq T\}$ are subsets of objects separated by the threshold T, and $\text{Ent}(S)$ denotes the Shannon entropy [36] of the class distribution in the subset S:

$$\text{Ent}(S) = -\sum_{k=1}^{K} p_k \log_2 p_k, \quad (3)$$

where p_k is the proportion of class k in the total number of classes S and K.

Criterion 3. Density distribution at the local level. KDE [32, 33] is used to estimate the density distribution of attribute values:

$$f_a(v) = \frac{1}{n \cdot h} \sum_{i=1}^{n} K\left(\frac{v - b_{ia}}{h}\right), \quad (4)$$

Сучасний стан наукових досліджень та технологій в промисловості. 2026. № 2 (36) *ISSN 2522-9818 (print)*
Innovative technologies and scientific solutions for industries. 2026. No. 2 (36) *ISSN 2524-2296 (online)*

where b_{ia} is the value of attribute a for object x_i, $K(\cdot)$ is the Gaussian kernel function, and $h = n^{-1/(d+4)}\sigma$ is the bandwidth parameter defined by Scott's rule [34].

The key conclusion is that optimal discretization thresholds must pass through low-density regions, the so-called "valleys" between natural clusters, rather than arbitrarily dividing dense regions.

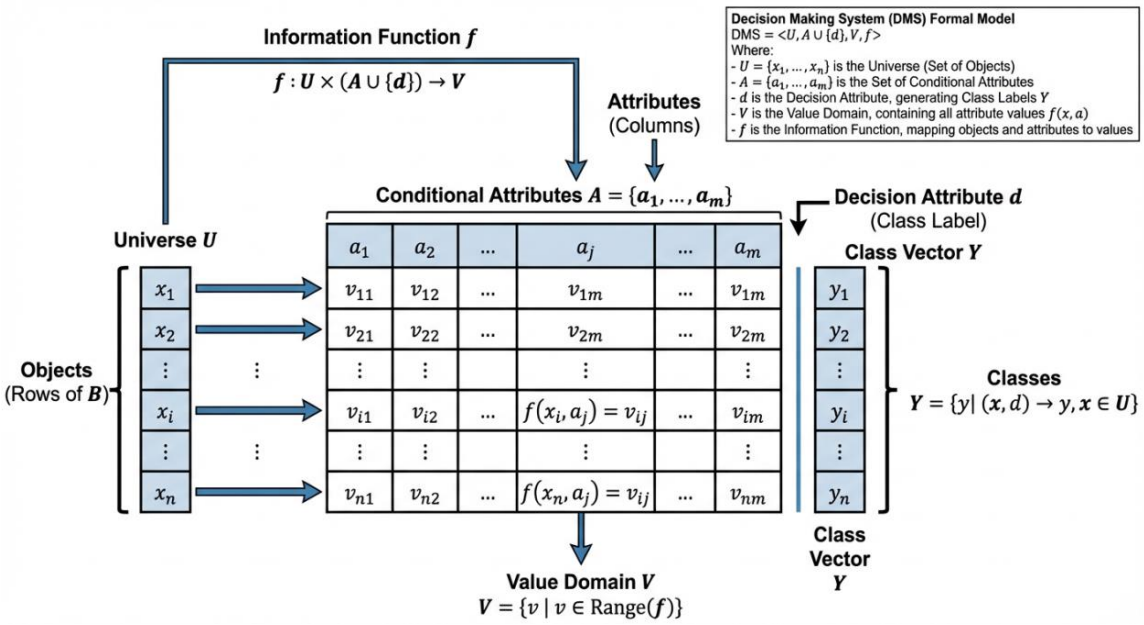

Fig. 2. The formal structure of a DMS. The feature matrix B provides the conditional attributes A, the class vector Y provides the decision attribute d, and the information function f maps objects to their attribute values in the domain V

Multi-criteria objective function. An integral quality criterion $Q(a,T)$ is introduced for threshold optimization, which must be minimized:

$$Q(a,T) = \alpha \cdot \overline{E}(a,T) + (1-\alpha) \cdot \overline{f_a}(T), \qquad (5)$$

where $\overline{E}(a,T)$ is the minimum-maximum normalized entropy, $\overline{f_a}(T)$ is the minimum-maximum normalized density at point T, and $\alpha \in [0,1]$ is a weighting coefficient that controls the trade-off between class separation and structure preservation.

In the experiments conducted, $\alpha = 0.6$ emphasizes class separation while simultaneously incorporating density information.

Recursive discretization algorithm. The WEDD algorithm works as follows.

Algorithm 1. WEDD discretization.

Input: Special column $\mathbf{b}_a$, class vector Y, weight α

1. Sort the values of b_{ia} in ascending order
2. Create threshold values for candidate $\mathcal{T}$ as the midpoints between adjacent distinct values
3. Compute the KDE profile of $f_a(v)$ using a Gaussian kernel with Scott's bandwidth
4. **For** each candidate $T \in \mathcal{T}$, **perform**
5. Compute the conditional entropy $E(a,T;U)$ using formula (2)
6. Calculate the density $f_a(T)$ using equation (4)
7. **Complete for**
8. Normalize:

$$\overline{E}(a,T) = \frac{E - E_{\min}}{E_{\max} - E_{\min}}, \quad \overline{f_a}(T) = \frac{f_a - f_{\min}}{f_{\max} - f_{\min}}$$

9. Compute $Q(a,T) = \alpha\overline{E} + (1-\alpha)\overline{f_a}$ for all T
10. Select $T^* = \arg\min_T \{Q(a,T)\}$
11. Apply the MDLP stopping criterion
12. **If** the MDLP criterion is not satisfied, **then**
13. Recursion on the subintervals of U_{left} and U_{right}
14. **terminate if**
15. **return** $T_a^* = \{\tau_1, \tau_2, \ldots, \tau_{k_a}\}$

Output: Set of optimal thresholds T_a^*

ISSN 2522-9818 (print) Сучасний стан наукових досліджень та технологій в промисловості. 2026. № 2 (36)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2026. No. 2 (36)

Stage 2. Feature space transformation and information granularity. After discretization, continuous values in B are mapped to discrete codes using a quantization function.

Definition 2 (quantization function). For each attribute $a_j \in A$ with optimal threshold values $T_j^* = \{\tau_{j,1}, \ldots, \tau_{j,k_j}\}$, the quantization function $\phi_j : V_{a_j} \to \mathbb{Z}^+$ maps continuous values to interval codes:

$$b_{ij}^* = \phi_j(b_{ij}) = \begin{cases} 0, & \text{if } b_{ij} < \tau_{j,1}; \\ m, & \text{if } \tau_{j,m} \le b_{ij} < \tau_{j,m+1}; \\ k_j, & \text{if } b_{ij} \ge \tau_{j,k_j}. \end{cases} \quad (6)$$

The result is a discrete matrix $B^* = \{b_{ij}^*\}$, in which each element is an integer denoting the qualitative state of the attribute (e.g., 0 = "low", 1 = "medium", 2 = "high").

Definition 3 (Indistinguishability Relation). Based on B^*, the indistinguishability relation $\mathrm{IND}(A)$ defines the condition of logical identity between two objects:

$$\mathrm{IND}(A) = \{(x_i, x_j) \in U \times U : \forall a_k \in A, b_{ik}^* = b_{jk}^*\}. \quad (7)$$

This relation divides the space into equivalence classes (information granules) $U/\mathrm{IND}(A) = \{E_1, E_2, \ldots, E_L\}$, where $L \le n$. Each granule E_p groups objects that have identical discretized attribute signatures $S_p = \langle b_{p1}^*, b_{p2}^*, \ldots, b_{pm}^* \rangle$. A granule is classified as:

– deterministic (pure): if all objects in E_p belong to a single class $(|\partial(E_p)| = 1$, where $\partial(E_p)$ is a set of distinct class values within $E_p)$;

– inconsistent: if the objects in E_p belong to several classes $(|\partial(E_p)| > 1)$, indicating an overlap of classes in the discrete feature space.

Stage 3. Formulation of deterministic production rules. Definition 4 (lower bound). For each decision class $X_j = \{x \in U : f(x,d) = y_j\}$, the lower bound is defined as [17]:

$$\underline{A}X_j = \{x \in U : [x]_A \subseteq X_j\}, \quad (8)$$

where $[x]_A$ denotes the equivalence class containing x.

An object x belongs to the lower approximation of the class X_j if and only if its entire equivalence class consists exclusively of objects of that class. This guarantees deterministic classification with 100% certainty.

Each deterministic granule $E_p \subseteq \underline{A}X_j$ generates a production rule:

$$R_p : \text{IF } (a_1 \in I_{1,v_1}) \wedge \ldots \wedge (a_m \in I_{m,v_m}) \Rightarrow d = y_j, \quad (9)$$

where I_{k,v_k} is the range of values of the k-th attribute corresponding to the discrete code v_k.

The support of the rule R_p quantifies its statistical significance:

$$\mathrm{Supp}(R_p) = |E_p \cap X_j| / |U|. \quad (10)$$

The overall quality of the approximation is measured by Pawlak's classification quality index [13]:

$$\gamma_A(d) = \sum_{j=1}^{K} |\underline{A}X_j| \Big/ |U|, \quad (11)$$

which represents the proportion of objects that can be unambiguously classified using deterministic rules.

Stage 4. Synthesis of probabilistic rules from boundary regions. Conflicting granules, i.e., those where $|\partial(E_p)| > 1$, form a boundary region $BN_A(X_j)$. For each such granule, a probabilistic rule is constructed:

$$R_p^{\text{prob}} : \text{IF conditions} \Rightarrow d = y_j \text{ with Conf}(R_p) = |E_p \cap X_j| / |E_p|. \quad (12)$$

The confidence $\mathrm{Conf}(R_p) < 1.0$ reflects the degree of class ambiguity within the granule. Rules with a confidence level below a threshold value may be flagged for human review in safety-critical applications. Fig. 3 illustrates the relationship between the lower approximation, the boundary region, and the upper approximation in the context of fuzzy set classification.

Stage 5. Minimizing the feature space using a reduction search. To determine the minimum set of features required for classification, a discriminability matrix M [23] is constructed. For each pair of granules (E_i, E_j) from different classes:

$$c_{ij} = \{a_k \in A : b_{ik}^* \ne b_{jk}^*\}, \quad \text{when } y_i \ne y_j. \quad (13)$$

The discrimination function is a combination of all disjunctions from non-empty cells:

$$f_{\mathrm{DS}}(a_1, \ldots, a_m) = \bigwedge \{ \bigvee (c_{ij}) : c_{ij} \ne \varnothing \}. \quad (14)$$

Сучасний стан наукових досліджень та технологій в промисловості. 2026. № 2 (36) *ISSN 2522-9818 (print)*
Innovative technologies and scientific solutions for industries. 2026. No. 2 (36) *ISSN 2524-2296 (online)*

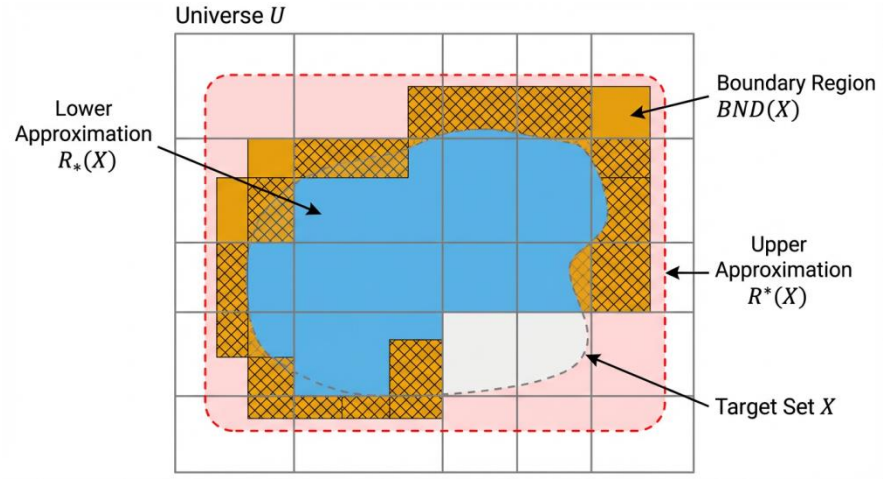

Fig. 3. RST approximation. Deterministic granules (entirely within the target class X_j) form the lower approximation (solid green); granules that partially overlap the class boundaries form the boundary region (striped yellow); their combination forms the upper approximation. The lower approximation guarantees 100% confidence in the classification

Johnson's modified algorithm with information weights. Finding all reducibles is NP-hard [23]. A weighted heuristic based on Johnson's greedy algorithm is proposed, where the priority of selecting each attribute includes its information capacity weight from the discretization stage:

$$W(a_k) = \text{Count}\left(a_k \in c_{ij}\right) \cdot \left(1 - \overline{E}(a_k)\right), \qquad (15)$$

where $\overline{E}(a_k)$ is the normalized average conditional entropy of attribute a_k, calculated over its discretization points.

Attributes with lower entropy (better class separation) receive higher weights, which shifts the greedy search toward features that are both discriminative and informative. The core of the attribute set is the intersection of all reducts:

$$\text{CORE}(A) = \left\{a_k \in A : \exists c_{ij} = \{a_k\}\right\}. \qquad (16)$$

The principal attributes are represented as single elements in the discriminability matrix and are indispensable for classification.

Stage 6. Compiling the PKB. The PKB is compiled as a standalone data structure that includes:

– discretization thresholds $\left\{T_j^*\right\}$ for all attributes;

– the selected reduction $\text{RED} \subseteq A$;

– deterministic lower approximation rules (confidence = 1.0);

– probabilistic rules from the boundary region (confidence < 1.0);

– metadata: support, reliability, coverage for each rule.

This structure allows the system to be deployed without external dependencies. The PKB is a fully-fledged, verifiable artifact that can be serialized in JSON format and embedded in peripheral devices or SCADA controllers.

Stage 7. Weighted inference mechanism. The inference procedure for classifying a new object $X_{\text{new}} = \langle b_1, \ldots, b_m \rangle$ proceeds as follows:

Step 1. Discretization. Apply the saved quantization functions:

$$x_j^* = \phi_j(b_j), \quad \forall j \in \{1, \ldots, m\}. \qquad (17)$$

Step 2. Rule activation. Search the knowledge base for all rules whose conditions match the discretized signature, using only the attributes selected in the reduct RED.

Step 3. Conflict resolution. When multiple rules are activated for different classes, calculate the class's integral support score:

$$S(y_k) = \sum_{\substack{R_i \in R_{\text{active}} \\ \text{Class} = y_k}} \text{Conf}(R_i) \cdot \text{Supp}(R_i) \cdot w_i, \qquad (18)$$

where w_i is the interval reliability weight obtained from the density profile in the discretization zone. The final classification is:

$$y^* = \arg\max_{y_k} S(y_k). \qquad (19)$$

Step 4. Matching with error. If no rule matches perfectly, the Hamming distance-based matching procedure finds the closest rule. If no acceptable match is found, the default class (the most frequent in Y) is assigned.

ISSN 2522-9818 (print) Сучасний стан наукових досліджень та технологій в промисловості. 2026. № 2 (36)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2026. No. 2 (36)

Case study

Experimental Setup

SCADA simulation dataset. To validate the PKB approach in the context of monitoring fire hazards in photovoltaic systems, a simulated SCADA dataset was created with the following characteristics:

– size – $n = 1,000$ samples (measurements from arrays of photovoltaic modules);

– characteristics – 3 continuous sensor measurements:

a) thermal anomaly intensity (X_i) – range $[0, 100]$, threshold 50;

b) bypass diode temperature (X_{1i}) – range $[20, 120]°C$, threshold 70°C;

c) fire sensor signal (X_{2i}) – range $[0, 10]$, threshold 5;

– classes – 5 operational states based on Boolean logic combinations (Table 1); – noise – Gaussian noise added to class-dependent distributions ($\mu = 0$, $\sigma = 0.5$, random seed = 42).

Benchmark Datasets. Two publicly available benchmark datasets from the UCI Machine Learning Repository [37] were used:

– Iris [38]: 150 samples, 4 features, 3 classes (setosa, versicolor, virginica). – Wine [39]: 178 samples, 13 features, 3 classes (grape varieties). For the SCADA dataset, the full dataset was used for both training and validation to evaluate PKB's ability to model the full feature space. For the benchmark datasets, 5-fold stratified cross-validation [40] was used with a fixed random seed of 42 to ensure reproducibility. The PKB approach was compared with two baseline models: a decision tree (a CART implementation using scikit-learn [37]) and Gaussian naive Bayes.

Table 1. *SCADA operational states and their Boolean logical definitions, which correspond to the fire hazard truth table implemented in the PV SCADA monitoring system*

State	X_i	X_{1i}	X_{2i}	Count
Normal	Low	Low	Low	300
Safe bypass	High	High	Low	250
Fire hazard	High	Low	Low	150
Defective diode	Low	High	Low	150
Active fire	Any	Any	High	150

WEDD discretization results

The WEDD algorithm with $\alpha = 0.6$ gave the following discretization thresholds (Table 2)..

Table 2. *WEDD discretization thresholds compared to the class boundaries in the real-world scenario. The algorithm successfully restores thresholds close to the true values, with cutoff points located in the density valleys*

Characteristics	Truth	WEDD Cuts	Bins
Thermal Intensity	50.0	42.8, 47.5, 54.7	4
Diode temperature (ºC)	70.0	58.5, 70.0, 76.0	4
Fire alarm	5.0	5.4	2

Fig. 4 shows a visualization of the WEDD discretization results, displaying class-specific histograms, KDE density curves, and selected cutoff points for each feature.

Key observations in Fig. 4 include:

a) the temperature limit of the bypass diode at 70.0°C accurately reproduces the threshold of the real-world scenario;

b) the intensity of the thermal anomaly is subject to three thresholds that encompass the true boundary at 50.0 on both sides, creating a transition zone that captures the overlap between classes;

c) the fire sensor signal requires only a single threshold at 5.4, located in the density valley between the fire and non-fire populations.

Сучасний стан наукових досліджень та технологій в промисловості. 2026. № 2 (36) *ISSN 2522-9818 (print)*
Innovative technologies and scientific solutions for industries. 2026. No. 2 (36) *ISSN 2524-2296 (online)*

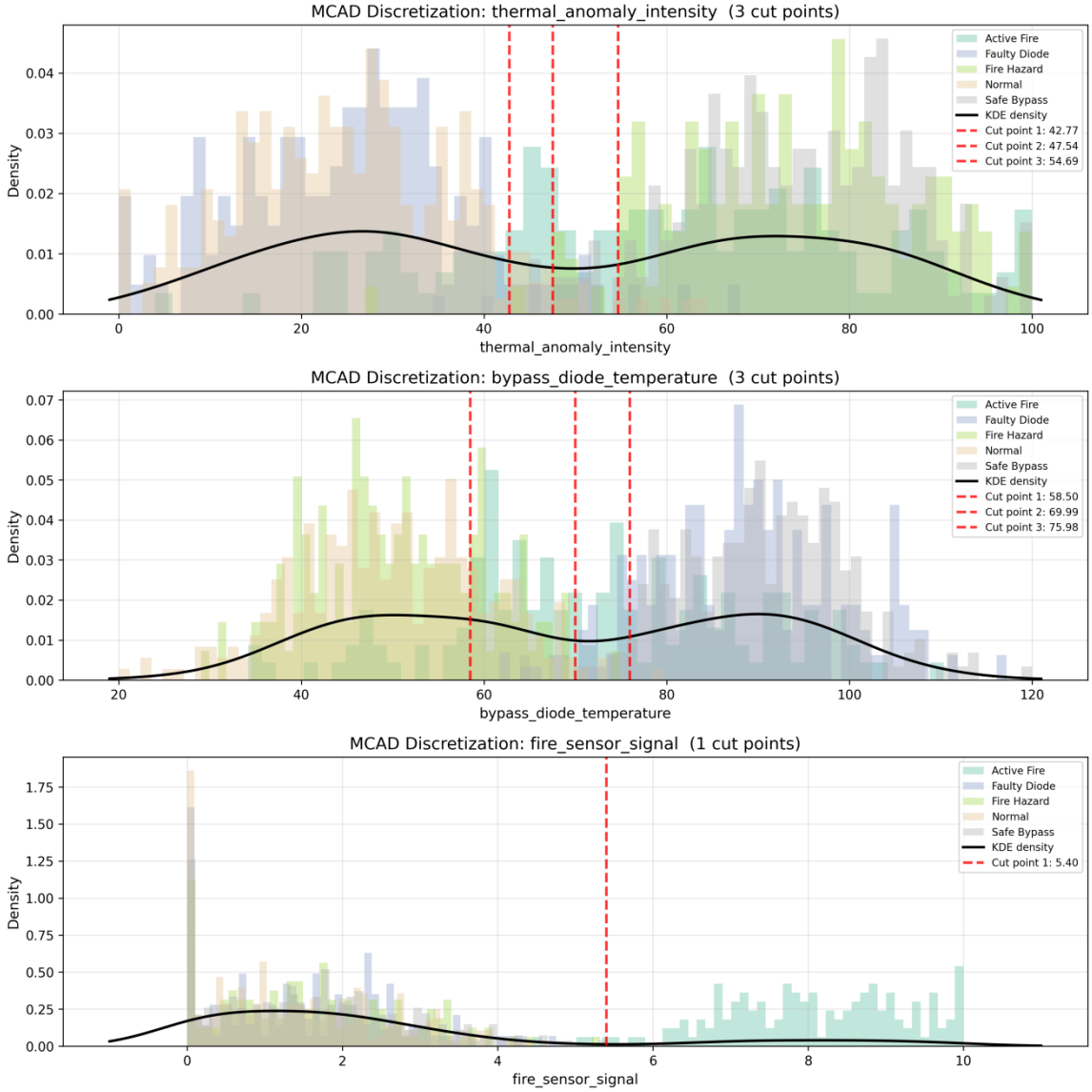

Fig. 4. WEDD discretization results for three SCADA sensor characteristics. Each panel shows: class-dependent histograms (colored bars), Gaussian density curves (KDE) (solid lines), and selected cutoff points (vertical dashed lines). The bypass diode temperature threshold at 70.0 degrees Celsius almost exactly matches the actual situation. The fire sensor signal requires only one cutoff point at 5.4, which closely corresponds to the actual threshold of 5.0. The multi-criteria objective places the cutoff points in the density valleys while simultaneously maximizing class separation

Granulation results and rule extraction

The discretized feature space (combinations) yielded 32 information granules, achieving a compression ratio of 31× from 1,000 objects. Table 3 presents the summary results of the granulation process.

The 19 deterministic rules are distributed as follows: 16 rules for an active fire (covering 98.0% of the class), and one rule each for a faulty diode, a normal state, and a safe bypass. It is noteworthy that the fire hazard class has no deterministic rules: it lies entirely in the boundary region, reflecting the inherent difficulty of distinguishing fire hazard conditions (high thermal anomaly + low diode temperature) from neighboring states. Thirteen probabilistic rules derived from conflicting granules complement the coverage with confidence levels ranging from 50.0% (P12, an equal distribution between Fire Hazard and Normal Mode) to 99.6% (P5, close to a deterministic Normal classification).

ISSN 2522-9818 (print) Сучасний стан наукових досліджень та технологій в промисловості. 2026. № 2 (36)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2026. No. 2 (36)

Table 3. *The granularity of information results from IND(a) equivalence class construction*

Indicators	Values
Total number of objects $(\|U\|)$	1,000
Total number of granules (L)	32
Compression ratio $(L/\|U\|)$	0.032
Deterministic granules	19 (59.4%)
Contradictory granules	13 (40.6%)
$\gamma_A(d)$ (Classification quality)	0.150
Objects in the lower approximation	150 (15.0%)
Objects in the boundary region	850 (85.0%)
Extracted deterministic rules	19
Extracted probabilistic rules	13
Total number of production rules	32

Results of the reductive analysis

Analysis of the discriminant matrix revealed 351 non-empty entries out of 496 pairs of granules with different majority classes. All three attributes appear as non-zero elements in the discriminant matrix (Table 4), since all three attributes have non-zero entries $\mathrm{CORE}(A) = \mathrm{RED} = A$ (a complete set of attributes). It is not possible to reduce the number of features; that is, each sensor measures a distinct physical phenomenon that is indispensable for classification. This confirms the well-thought-out design of the SCADA sensor suite.

Table 4. *Analysis of reducts: feature weights and discriminative singletons*

Characteristics	Weight	Singletons	Count
Heat intensity	0.328	15	88.5
Diode temperature	0.314	17	20.4
Fire alarm	0.280	16	4.5

Validation results

The PKB extraction mechanism yielded the results shown in Table 5.

Table 5. *Validation metrics for the PKB extraction mechanism on the SCADA dataset* $(n = 1000)$*: the system achieves an overall accuracy of 96.2% with perfectly deterministic rule performance*

Class	Accuracy	Recall	F1	Count
Active fire	1.000	0.980	0.990	150
Defective diode	0.958	0.913	0.935	150
Fire hazard	0.907	0.980	0.942	150
Normal	0.983	0.957	0.970	300
Safe bypass	0.953	0.976	0.964	250
Average macro	0.960	0.961	0.960	1000
Weighted average	0.963	0.962	0.962	1000

The "Fire Hazard" critical safety class achieves a recall rate of 98.0% (147 out of 150 correctly identified), despite relying entirely on probabilistic rules. The 90.7% accuracy indicates some false alarms in the "Normal" class, which is an acceptable trade-off in safety-critical applications where missed hazards have catastrophic consequences. Figure 5 shows the confusion matrix for verifying PKB's conclusions on the SCADA dataset.

An analysis by rule type reveals a clear dichotomy: deterministic rules achieve 100% accuracy $(150/150)$, while probabilistic rules achieve 95.5% accuracy $(812/850)$, all 38 misclassifications stem from probabilistic rules in granules with a high level of conflict.

Сучасний стан наукових досліджень та технологій в промисловості. 2026. № 2 (36) *ISSN 2522-9818 (print)*
Innovative technologies and scientific solutions for industries. 2026. No. 2 (36) *ISSN 2524-2296 (online)*

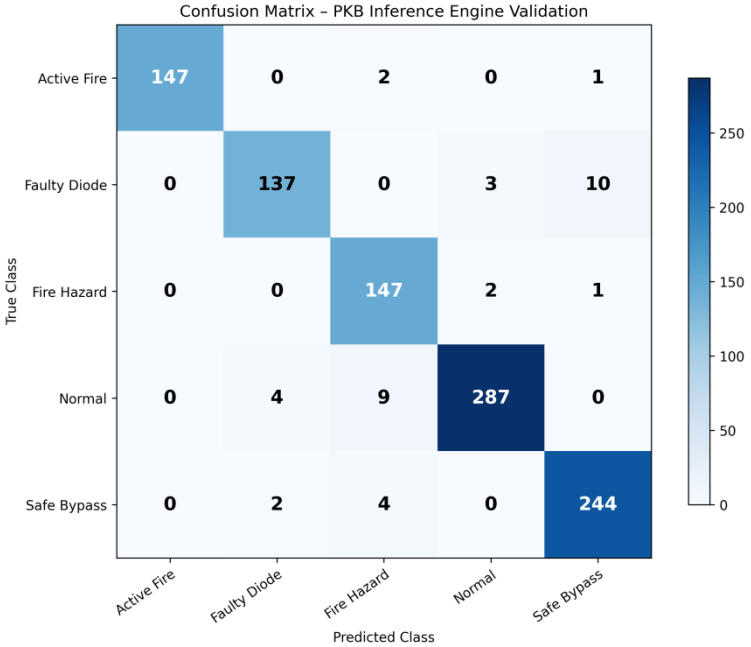

Fig. 5. A confusion matrix for verifying PKB's conclusions on the SCADA dataset. The dominance of the diagonal confirms high classification accuracy across all five operating states. The main sources of errors are misclassifications between a faulty diode and a safe bypass (10 cases) and between a normal state and a fire hazard (9 cases); both types of errors occur in border regions where sensor signals overlap

Comparative testing results

Table 6 and Figure 6 present comparative results for the Iris and Wine test datasets.

On the Iris dataset (4 features, 3 classes), PKB achieves an accuracy of 93.3%, which is 2.0 percentage points lower than the best baseline (a decision tree at 95.3%). On the Wine dataset (13 features, 3 classes), PKB achieves an accuracy of 87.6% thanks to greedy feature selection based on reduction, which reduces the feature space from 13 to 3–6 features per fold. Although Naive Bayes outperforms Wine (97.2%) due to the nearly Gaussian distribution of features, which satisfies its assumption of independence, PKB significantly outperforms it in terms of interpretability.

Table 6. *Comparison of tests using 5-fold stratified cross-validation. The mean standard deviation is shown. The best results are highlighted in* ***bold***

Data set	Method	Accuracy	Macro F1
Iris	PKB	0.933 ± 0.030	0.933 ± 0.030
	Decision tree	**0.953 ± 0.034**	**0.953 ± 0.034**
	Naive Bayes	0.947 ± 0.040	0.947 ± 0.040
Wine	PKB	0.876 ± 0.078	0.873 ± 0.083
	Decision tree	0.893 ± 0.038	0.896 ± 0.036
	Naive Bayes	**0.972 ± 0.025**	**0.973 ± 0.024**

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Сучасний стан наукових досліджень та технологій в промисловості. 2026. № 2 (36)
Innovative technologies and scientific solutions for industries. 2026. No. 2 (36)

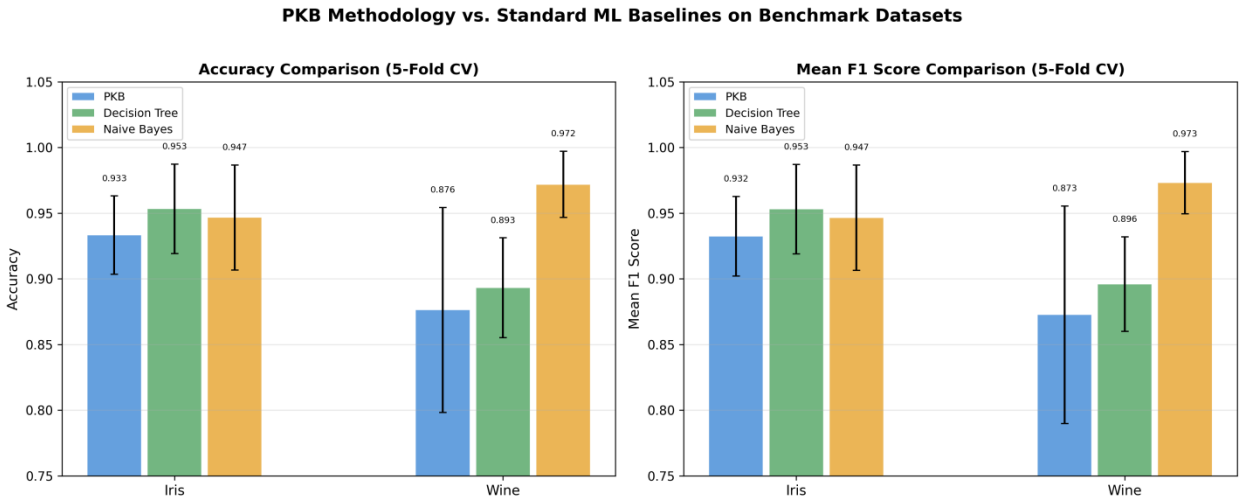

Fig. 6. A comparison of the PKB approach with the baseline "Decision Tree" and "Naive Bayes" models on the Iris and Wine datasets using 5-fold stratified cross-validation. PKB achieves competitive accuracy (left) and F1 score (right), while offering the unique advantage of interpretable IF-THEN production rules with explicit confidence scores. The error bars represent ±1 standard deviation across all folds

Application context:
detection of fire hazards using photovoltaic systems

Figures 7 and 8 show the improved detection accuracy and system performance, respectively.

The PKB approach integrates with the SCADA system's Boolean logic layer (Fig. 9) by replacing binary threshold values with interpretable production rules that provide classifications based on confidence levels, enabling operators to understand the logic behind each fire hazard assessment.

Fig. 10 illustrates the system's ability to distinguish between different types of defects based on thermal profile characteristics.

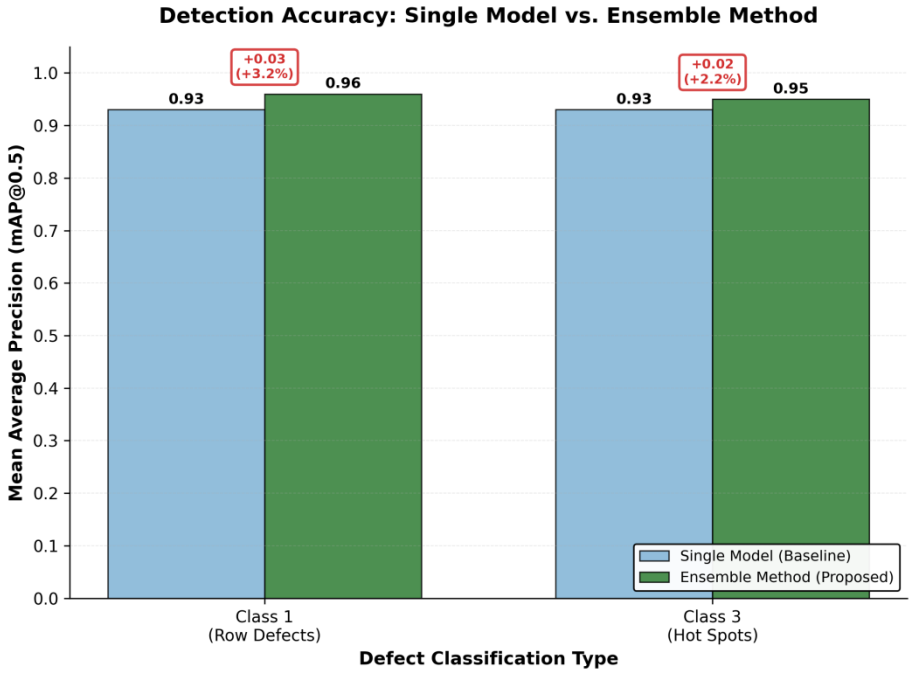

Fig. 7. Detection of improved accuracy in CPS PV monitoring. A combined method that integrates thermal palettes M2 (detection of large defects) and M3 (detection of hot spots) achieves a mAP@0.5 of 0.96 for Class 1 defects (+3%), 0.90 for Class 2, and 0.95 for Class 3 (+2%). The PKB approach complements this detection level by providing an interpretable classification of operating states based on detected thermal anomalies

Сучасний стан наукових досліджень та технологій в промисловості. 2026. № 2 (36) *ISSN 2522-9818 (print)*
Innovative technologies and scientific solutions for industries. 2026. No. 2 (36) *ISSN 2524-2296 (online)*

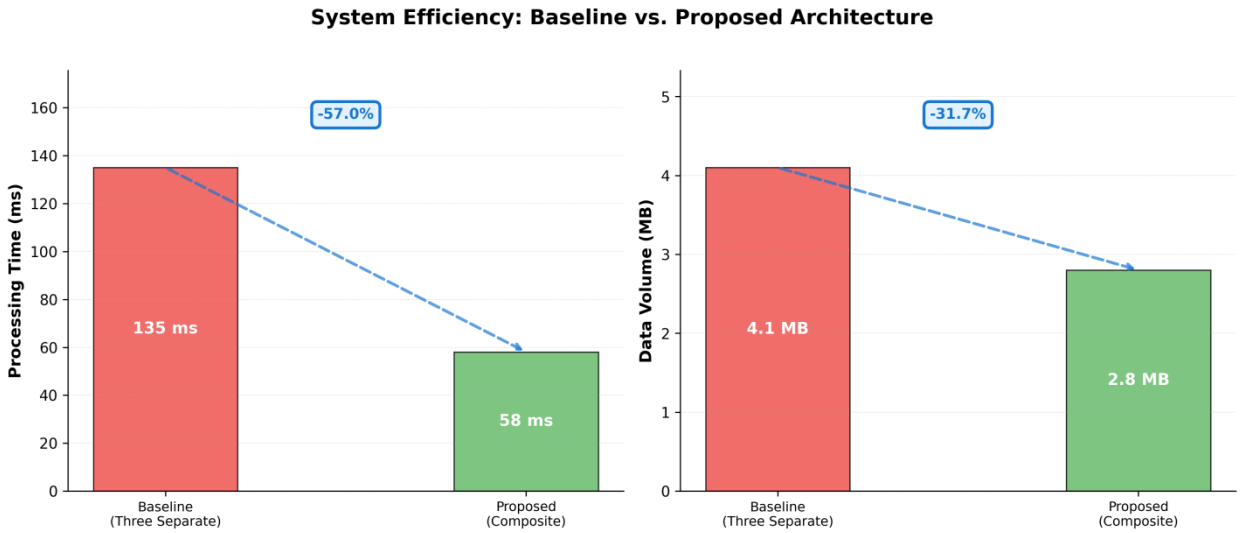

Fig. 8. Improvements in system performance in a CPS architecture with edge cloud computing. Left: 57.0% reduction in processing time thanks to edge computing. Right: 31.7% reduction in data transfer volume. In addition to these performance improvements, the PKB output mechanism adds an interpreted fire hazard classification, enabling proactive safety assessment using SCADA Boolean logic

SCADA Fire Hazard Assessment Logic
Boolean Truth Table for Cyber-Physical System

Xi (UAV Defect)	X1i (Diode Temp.)	X2i (Fire Sensor)	System Status	Action Required
0	0	0	Normal Operation	None
0	0	1	Fire Detected	EMERGENCY
0	1	0	Faulty Diode	Maintenance
0	1	1	Fire Detected	EMERGENCY
1	0	0	Fire Hazard	ALERT
1	0	1	Fire Detected	EMERGENCY
1	1	0	Protected (Safe)	Monitor
1	1	1	Fire Detected	EMERGENCY

Logic Variables:
• Xi: Thermal defect detected by UAV (YOLOv11m-seg)
• X1i: Elevated bypass diode temperature (on-site sensor)
• X2i: Fire sensor activation (on-site sensor)

Decision Rules:
• Defect + No Diode Activation = Fire Hazard (protection failed)
• Defect + Diode Activation = Protected (safety system working)
• Fire Sensor Active = Emergency Response Required

Integration: UAV Edge Computing → Ground Station → SCADA → Alarm System

Fig. 9. The logic for making decisions regarding fire hazards, implemented in the SCADA system.
A Boolean truth table correlates three sensor input signals (X_i, X_{1i}, X_{2i}) to classify operating states.
The PKB approach improves this logic by providing rules weighted by confidence level instead of rigid binary thresholds, allowing for a more detailed assessment of borderline cases

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Сучасний стан наукових досліджень та технологій в промисловості. 2026. № 2 (36)
Innovative technologies and scientific solutions for industries. 2026. No. 2 (36)

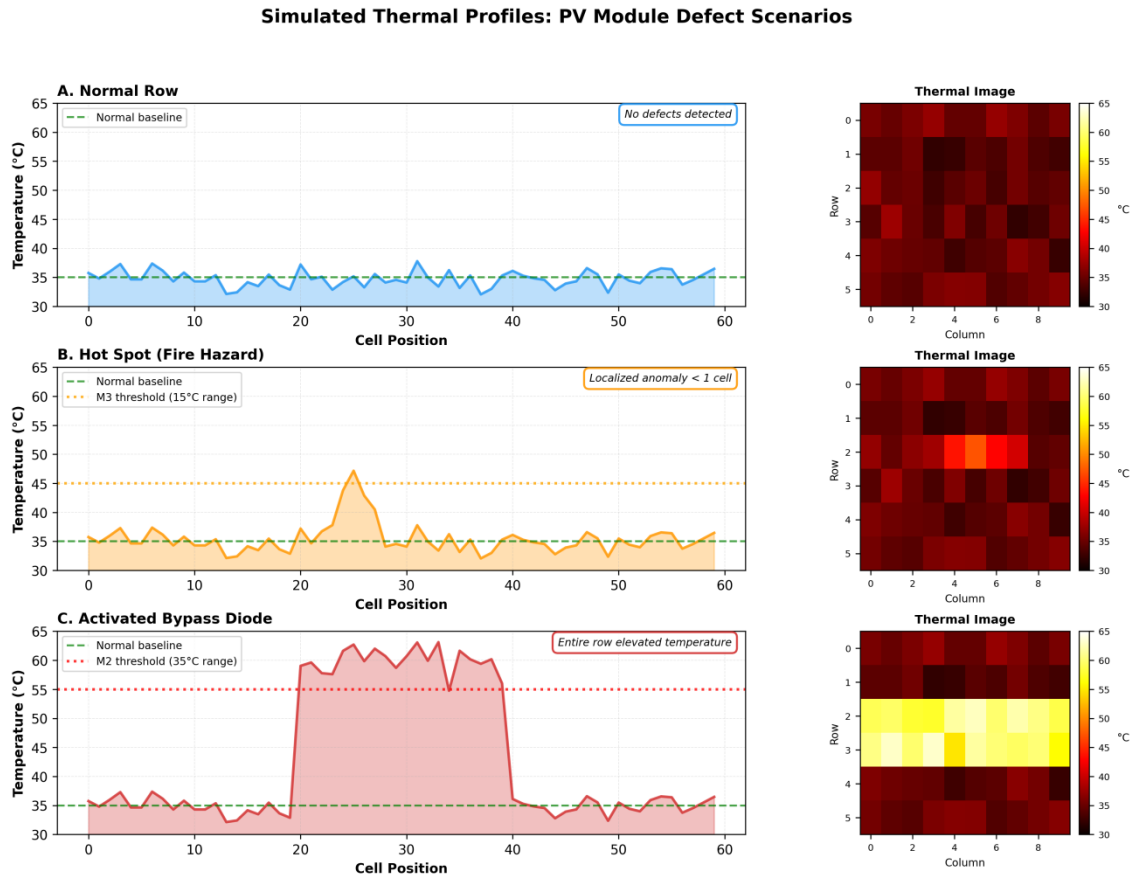

Fig. 10. Thermal profiles for two critical types of defects in photovoltaic modules. Top: a "hot spot" defect, showing a sharp Gaussian temperature peak (detected by the M3 palette). Bottom: bypass diode activation, showing a step-like profile (detected by the M2 palette). When quantified by SCADA sensors, these thermal signatures serve as input features for PKB classification

Discussions

The results of this study confirm the effectiveness of the PKB framework as a reliable alternative to opaque "black-box" classification models in industrial settings where safety is critical. As noted in Section 1, a significant challenge in modern CPS is reconciling the high predictive accuracy of deep learning algorithms, such as YOLO frameworks [3, 8], with the transparency required to build operator confidence. The proposed approach overcomes this obstacle by transforming continuous sensor measurements into explicit symbolic knowledge. A key element of this achievement is the WEDD algorithm. Unlike traditional entropy-based discretization, which isolates class boundaries and often breaks up natural data clusters, the presented technique uses KDE to precisely determine threshold values in density valleys that separate distinct subpopulations. As shown in Fig. 4, this dual-criterion optimization allowed the framework to accurately identify physical boundaries of the real-world scenario, such as the 70.0°C threshold for bypass diode temperature, without manual tuning. This demonstrates that combining information entropy with density parameters preserves the intrinsic semantic architecture of the data, creating a robust foundation for subsequent rule generation.

The implementation of RST introduced a rigorous mathematical framework for measuring uncertainty, a feature often lacking in traditional decision trees or rule extraction algorithms such as CN2 and RIPPER. The granularity metrics, detailed in Table 3, indicate that although only 15% of the synthetic SCADA cases were mapped to the lower bound (deterministic region), the architecture maintained exceptional accuracy within the boundary region using confidence-weighted logic. This differentiation has significant practical value: the system can autonomously trigger responses based on deterministic rules (confidence probability = 1.0) while simultaneously flagging probabilistic rules (confidence probability < 1.0) for manual review by human operators.

Сучасний стан наукових досліджень та технологій в промисловості. 2026. № 2 (36) *ISSN 2522-9818 (print)*
Innovative technologies and scientific solutions for industries. 2026. No. 2 (36) *ISSN 2524-2296 (online)*

Such functionality aligns well with the hybrid decision-making systems highlighted in recent studies [11], which advocate for a balance between machine automation and human oversight. Furthermore, the flawless 100% accuracy recorded for deterministic rules (Table 5) provides empirical confirmation of the fundamental premise of the lower bound approximation: provided the data structure is clear and well-defined, the resulting logical rules are absolutely error-free.

An analytical evaluation of the SCADA dataset confirms that the proposed approach meets stringent fire safety requirements. The confusion matrix in Fig. 5 demonstrates the model's ability to maximize reproducibility under the most demanding operational conditions. Although the "Fire Hazard" category did not yield deterministic rules due to significant attribute overlap, the probabilistic reasoning module achieved 98.0% reproducibility, missing only 2% of actual threats. This achievement comes with a small trade-off, yielding 90.7% accuracy and a moderate level of false alarms. In the field of protecting photovoltaic facilities [4], this approach is highly desirable, as the catastrophic financial and safety consequences of an undetected fire far outweigh the minor logistical costs of investigating false alarms.

Furthermore, comparative baseline tests on the Iris and Wine datasets (Table 6 and Fig. 6) confirm that the proposed technique performs on par with well-known algorithms. Although Gaussian Naive Bayes outperformed the proposed model on the Wine dataset (97.2% vs. 87.6%) due to differences in the distribution of these metrics, the current approach has the distinct advantage of generating transparent IF-THEN rules. As shown in Section 4.5, these rules remain fully controllable from a regulatory standpoint and are easily interpreted by industry experts. Integrating this architecture into edge-cloud ecosystems offers significant advantages over alternatives that are entirely cloud-dependent or based exclusively on neural networks. As shown in Fig. 8, this structure provides a significant reduction in data transfer volumes and computation latency. By exporting the complete knowledge base – that is, including discretization boundaries, reducers, and extracted rules – into a compact JSON format of approximately 32 KB, the platform ensures exceptional portability. This artifact can be seamlessly integrated into the operational logic of edge devices, ranging from the NVIDIA Jetson AGX Orin to simpler microcontrollers. Here, it serves as an interpreted reasoning layer, linking the visual detection input data of the neural network with the output data of the SCADA system's mechanical actuator. Such a configuration ideally supports the industry-wide trend toward transitioning to an edge intelligence system [6], facilitating decentralized decision-making that maintains functionality even during potential network outages. Despite these advantages, various limitations and potential obstacles to research exist. First, the existing empirical validation relies on synthetic SCADA data with added Gaussian noise. Authentic sensor readings from operational photovoltaic installations often exhibit non-Gaussian distortions, temporal drift, and hardware malfunctions, which can reduce the effectiveness of the density estimation module. The relatively modest classification quality index $\gamma A(d) = 0.150$ indicates that the three-dimensional feature space exhibits significant class overlap. This problem may be difficult to solve without incorporating additional sensor modalities. Furthermore, the feature reduction process relies on an adapted greedy algorithm. Although incorporating information capacity weights improves the traditional Johnson algorithm, it still cannot guarantee the discovery of a globally optimal reduction, which may result in the creation of rule sets that are more extensive than necessary. The observed performance degradation on the high-dimensional Wine dataset indicates that the adopted heuristic may struggle to manage complex feature interdependencies in high-dimensional spaces. Future research initiatives should address these limitations by exploring evolutionary optimization methods for identifying reducibles and testing the approach on various real-world industrial datasets to ensure robustness to environmental changes and sensor degradation.

Conclusion

This study confirms that the PKB approach is a carefully designed and mathematically sound framework for transforming continuous input data from cyber-physical sensors into clear and useful information. By combining the recently introduced WEDD algorithm with the robust mathematical foundations of RST, a comprehensive pipeline was created to generate highly accurate, linguistically intuitive, and logically rigorous production rules. The proposed model underwent comprehensive testing in a seven-stage workflow, confirming its effectiveness in managing the complex demands of industrial monitoring. When applied to

ISSN 2522-9818 (print) Сучасний стан наукових досліджень та технологій в промисловості. 2026. № 2 (36)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2026. No. 2 (36)

a synthetic SCADA dataset on fire hazards, the architecture achieved an impressive overall accuracy of 96.2%, with flawless 100% accuracy in deterministic cases within the lower bound. Most importantly, for safety-critical infrastructure, the model achieved 98.0% recall for fire emergencies, ensuring reliable detection of hazardous conditions even when their signatures overlap with those of standard operations. Comparative benchmarking using the Iris (93.3%) and Wine (87.6%) datasets confirms that, despite being optimized for industrial logic, the model maintains competitive performance in general classification tasks compared to opaque "black-box" alternatives. The main identified limitation is the greedy nature of the feature reduction stage, which can lead to suboptimal subsets in high-dimensional spaces, as well as the framework's dependence on simulated noise models. Looking ahead, three key research directions are proposed for further improving this approach. First, incorporating palette-independent radiometric models will enhance the system's robustness to input data regarding ambient thermal noise. Second, the application of evolutionary algorithms will help optimize the trade-off between the compactness of the rule set and classification coverage. Finally, the development of online learning mechanisms will allow the knowledge base to dynamically adapt to the aging of photovoltaic system components.

Conflict of interest

The authors declare that they have no conflict of interest, including financial, personal, authorship, or any other conflict that could influence the research or the results published in this article.

Funding

This research was supported by the Ministry of Education and Science of Ukraine and funded by the European Union's external aid instrument as part of Ukraine's commitments under the EU Framework Program for Research and Innovation "Horizon 2020". The work was performed as part of the research project "Intelligent Defect Detection System for Green Energy Facilities Using UAVs", state registration number 0124U004665 (2024–2026).

Data availability

Data will be provided upon reasonable request. The base datasets (Iris and Wine) are publicly available. Analysis scripts may be provided by the authors via correspondence upon reasonable request. The proprietary dataset obtained during field validation is not publicly available due to commercial confidentiality.

Use of artificial intelligence

The authors declare the use of artificial intelligence (AI) technologies during the preparation of this manuscript as follows:

– AI model and number: the large language model Gemini 3.1 Pro Preview (from Google LLC) and the language service Grammarly (from Grammarly, Inc.);

– where exactly they were used: in all sections of the manuscript for linguistic and stylistic editing of the text;

– what exactly was done using AI tools: the tools were used exclusively to improve the linguistic quality of the text, namely: checking grammar, spelling, and punctuation, as well as ensuring stylistic consistency and the accuracy of the academic translation;

– how the authors verified the results provided by the AI tools: the authors conducted a thorough analysis and manual review of all suggested corrections to ensure that the entire content of the manuscript accurately and without distortion reflects the essence of the research; the authors retain full responsibility for the scientific integrity and final content of this article;

– how the results provided by the AI tool influenced the study's conclusions: the use of AI had no impact on the scientific conclusions, research results, or methodology; it affected only the linguistic quality of the text.

Acknowledgments

The authors express their sincere gratitude to the Ministry of Education and Science of Ukraine for its support. This research was funded by the European Union's external aid instrument under the "Horizon 2020" research and innovation program. The work was carried out as part of the project "Intelligent Defect Detection System for Green Energy Facilities Using UAVs" (state registration number 0124U004665).

All authors have read and approved the published version of this manuscript.

Сучасний стан наукових досліджень та технологій в промисловості. 2026. № 2 (36) *ISSN 2522-9818 (print)*
Innovative technologies and scientific solutions for industries. 2026. No. 2 (36) *ISSN 2524-2296 (online)*

References

1. Carnì, D. L., Grimaldi, D., Lamonaca, F., Nigro, L., and Sciammarella, F. (2017), "From distributed measurement systems to cyber-physical systems: A design approach", *International Journal of Computing*, Vol. 16, No. 2, pp. 66–73. DOI: https://doi.org/10.47839/ijc.16.2.882
2. Osadchy, S., Demska, N., Oleksandrov, Y., and Nevliudova, V. (2021), "Research of DIKW and 5C architectural models for creation of cyber-physical production systems within the concept of industry 4.0", *Innovative Technologies and Scientific Solutions for Industries*, No. 1(15), pp. 132–140. DOI: https://doi.org/10.30837/itssi.2021.15.132
3. Xie, H., Yuan, B., Hu, C., Gao, Y., Wang, F., Wang, C., Wang, Y., and Chu, P. (2024), "ST-YOLO: A defect detection method for photovoltaic modules based on infrared thermal imaging and machine vision technology", *PLOS ONE*, Vol. 19, No. 12, Article e0310742. DOI: https://doi.org/10.1371/journal.pone.0310742
4. Lysyi, A., Sachenko, A., Radiuk, P., Lysyi, M., Melnychenko, O., Ishchuk, O., and Savenko, O. (2025), "Enhanced fire hazard detection in solar power plants: An integrated UAV, AI, and SCADA-based approach", *Radioelectronic and Computer Systems*, Vol. 2025, No. 2, pp. 99–117. DOI: https://doi.org/10.32620/reks.2025.2.06
5. Thakfan, A., and Bin Salamah, Y. (2024), "Artificial-intelligence-based detection of defects and faults in photovoltaic systems: A survey", *Energies*, Vol. 17, No. 19, Article 4807. DOI: https://doi.org/10.3390/en17194807
6. Svystun, S., Melnychenko, O., Radiuk, P., Savenko, O., Sachenko, A., and Lysyi, A. (2024), "Thermal and RGB images work better together in wind turbine damage detection", *International Journal of Computing*, Vol. 23, No. 4, pp. 526–535. DOI: https://doi.org/10.47839/ijc.23.4.3752
7. Donida Labati, R., Genovese, A., Muñoz, E., Piuri, V., Scotti, F., and Sforza, G. (2016), "Computational intelligence for biometric applications: A survey", *International Journal of Computing*, Vol. 15, No. 1, pp. 42–53. DOI: https://doi.org/10.47839/ijc.15.1.829
8. Terven, J., Córdova-Esparza, D.-M., and Romero-González, J.-A. (2023), "A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS", *Machine Learning and Knowledge Extraction*, Vol. 5, No. 4, pp. 1680–1716. DOI: https://doi.org/10.3390/make5040083
9. Svystun, S., Scislo, L., Pawlik, M., Melnychenko, O., Radiuk, P., Savenko, O., and Sachenko, A. (2025), "DyTAM: Accelerating wind turbine inspections with dynamic UAV trajectory adaptation", *Energies*, Vol. 18, No. 7, Article 1823. DOI: https://doi.org/10.3390/en18071823
10. Denysiuk, D., Geidarova, O., Kapustian, M., Lysenko, S., and Sachenko, A. (2023), "Blockchain-based deep learning algorithm for detecting malware", In: Proceedings of the 4th International Workshop on Intelligent Information Technologies & Systems of Information Security, 22–24 March 2023, Khmelnytskyi, Ukraine, Aachen: CEUR-WS.org. pp. 529–538, available at: https://ceur-ws.org/Vol-3373/paper36.pdf (last accessed 26.02.2026).
11. Morozov, A., Fabarisov, T., Vock, S., Siedel, G., Bolbot, V., and Voß, S. (2026), "Risk and reliability evaluation of future industrial automation systems: a systematic literature review and research agenda", *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, Vol. 12, No. 2, Article 020801. DOI: https://doi.org/10.1115/1.4070764
12. Mochurad, L., and Hladun, Y. (2021), "Modeling of psychomotor reactions of a person based on modification of the tapping test", *International Journal of Computing*, Vol. 20, No. 2, pp. 190–200. DOI: https://doi.org/10.47839/ijc.20.2.2166
13. Pięta, P., and Szmuc, T. (2021), "Applications of rough sets in big data analysis: An overview", *International Journal of Applied Mathematics and Computer Science*, Vol. 31, No. 4, pp. 659–683. DOI: https://doi.org/10.34768/amcs-2021-0046
14. Costa, V. G., and Pedreira, C. E. (2022), "Recent advances in decision trees: An updated survey", *Artificial Intelligence Review*, Vol. 56, No. 5, pp. 4765–4800. DOI: https://doi.org/10.1007/s10462-022-10275-5
15. Javed Mehedi Shamrat, F.M., Ranjan, R., Hasib, K.M., Yadav, A., and Siddique, A.H. (2022), "Performance evaluation among ID3, C4.5, and CART decision tree algorithm", In: *Pervasive Computing and Social Networking*, Singapore: Springer Singapore, Lecture Notes in Networks and Systems, Vol. 317, pp. 127–142. DOI: https://doi.org/10.1007/978-981-16-5640-8_11

16. Li, H. (2024), "Decision Tree", In: *Machine Learning Methods*, Singapore: Springer, pp. 77–102. DOI: https://doi.org/10.1007/978-981-99-3917-6_5
17. Xu, W., Yan, Y., and Li, X. (2025), "Sequential rough set: a conservative extension of Pawlak's classical rough set", *Artificial Intelligence Review*, Vol. 58, No. 1, Article 9. DOI: https://doi.org/10.1007/s10462-024-10976-z
18. Tetteh, E. T., and Zielosko, B. (2025), "Greedy algorithm for deriving decision rules from decision tree ensembles", *Entropy*, Vol. 27, No. 1, Article 35. DOI: https://doi.org/10.3390/e27010035
19. Słowiński, R., Greco, S., and Matarazzo, B. (2023), "Rough Sets in Decision-Making", In: T.-Y. Lin, C.-J. Liau and J. Kacprzyk, eds., *Granular, Fuzzy, and Soft Computing*, Encyclopedia of Complexity and Systems Science Series, Springer, New York, NY. pp. 419–468. DOI: https://doi.org/10.1007/978-1-0716-2628-3_460
20. Benbrahim, M., Hamaidi, B., Haouassi, H., Mahdaoui, R., and Mouss, L.-H. (2025), "Explainable fault detection and diagnosis based on an IDEOA as applied to an industrial process", *Diagnostyka*, Vol. 26, No. 2, Article 2025203. DOI: https://doi.org/10.29354/diag/203246
21. Badanyuk, I., Nevliudov, I., and Nikitin, D. (2023), "Topological image processing for comprehensive defect and deviation analysis using adaptive binarisation", *Innovative Technologies and Scientific Solutions for Industries*, No. 1(23), pp. 164–173. DOI: https://doi.org/10.30837/itssi.2023.23.164
22. Ma, J., Hu, X., Chu, T., Zhao, H., and Ma, D. (2026), "IPIGN: An interpretable physics-informed graph network multitask cascading failure diagnosis method for distributed energy systems", *IEEE Transactions on Industrial Electronics*, Vol. 73, No. 5, pp. 7960–7971. DOI: https://doi.org/10.1109/tie.2025.3645397
23. Luo, S., Shi, L., Chen, L., and Cao, X. (2025), "Accelerated feature selection via discernibility hashing: A rough set approach", *Entropy*, Vol. 27, No. 12, Article 1222. DOI: https://doi.org/10.3390/e27121222
24. Alves Júnior, J.G.V., Leite, N.F., Guimarães, J.B., Marques, J.A.L., Ribeiro, S.S., and Alexandria, A.R.d. (2025), "Rough Set Theory Applied to Feature Selection", In: Zhang, Q. et al., eds., *Rough Sets (IJCRS 2025)*, Lecture Notes in Computer Science, Vol. 15709, Cham: Springer Nature Switzerland, pp. 91–107. DOI: https://doi.org/10.1007/978-3-031-92744-7_7
25. Błaszczyński, J., Greco, S., Matarazzo, B., and Szeląg, M. (2022), "Dominance-Based Rough Set Approach: Basic Ideas and Main Trends", In: *Intelligent decision support systems*, Cham: Springer International Publishing. pp. 353–382. DOI: DOI: https://doi.org/10.1007/978-3-030-96318-7_18
26. Zhang, P., Li, T., Wang, G., Luo, C., Chen, H., Zhang, J., Wang, D., and Yu, Z. (2021), "Multi-source information fusion based on rough set theory: a review", *Information Fusion*, Vol. 68, pp. 85–117. DOI: https://doi.org/10.1016/j.inffus.2020.11.004
27. Grzymala-Busse, J.W. (2023), "Rule induction", In: *Machine Learning for Data Science Handbook*, Cham: Springer International Publishing, pp. 55–74. DOI: https://doi.org/10.1007/978-3-031-24628-9_4
28. García, S., Luengo, J., Sáez, J. A., López, V., and Herrera, F. (2013), "A survey of discretization techniques: taxonomy and empirical analysis in supervised learning", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 25, No. 4, pp. 734–750. DOI: https://doi.org/10.1109/TKDE.2012.35
29. Toulabinejad, E., Mirsafaei, M., and Basiri, A. (2024), "Supervised discretization of continuous-valued attributes for classification using RACER algorithm", *Expert Systems with Applications*, Vol. 244, Article 121203. DOI: https://doi.org/10.1016/j.eswa.2023.121203
30. Huang, M.-W., Tsai, C.-F., Tsui, S.-C., and Lin, W.-C. (2023), "Combining data discretization and missing value imputation for incomplete medical datasets", *PLOS ONE*, Vol. 18, No. 11, Article e0295032. DOI: https://doi.org/10.1371/journal.pone.0295032
31. Xun, Y., Yin, Q., Zhang, J., Yang, H., and Cui, X. (2021), "A novel discretization algorithm based on multi-scale and information entropy", *Applied Intelligence*, Vol. 51, No. 2, pp. 991–1009. DOI: https://doi.org/10.1007/s10489-020-01850-w
32. Fauzi, R. R., and Maesono, Y. (2023), "Kernel Density Function Estimator", In: *Statistical Inference Based on Kernel Distribution Function Estimators*, Singapore: Springer Nature Singapore, pp. 1–16. DOI: https://doi.org/10.1007/978-981-99-1862-1_1
33. Pelz, M.-T., Schartau, M., Somes, C. J., Lampe, V., and Slawig, T. (2023), "A diffusion-based kernel density estimator (diffKDE, version 1) with optimal bandwidth approximation for the analysis of data in geoscience and ecological research", *Geoscientific Model Development*, Vol. 16, No. 22, pp. 6609–6634. DOI: https://doi.org/10.5194/gmd-16-6609-2023

34. Francisco-Fernández, M., Barreiro-Ures, D., and Cao, R. (2026), "Subagging for bandwidth selection: A computationally efficient approach to kernel density estimation", *Computational Statistics*, Vol. 41, No. 1, Article 29. DOI: https://doi.org/10.1007/s00180-025-01712-4
35. Lysyi, A., Sachenko, A., Radiuk, P., Lysyi, M., Melnychenko, O., and Zahorodnia, D. (2025), "Method of UAV-based inspection of photovoltaic modules using thermal and RGB data fusion", *Radioelectronic and Computer Systems*, Vol. 2025, No. 4, pp. 186–205. DOI: https://doi.org/10.32620/reks.2025.4.13
36. Shannon, C. E. (1948), "A mathematical theory of communication", *The Bell System Technical Journal*, Vol. 27, No. 3, pp. 379–423. DOI: https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
37. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B. et al. (2011), "Scikit-learn: Machine Learning in Python", *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830. https://jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf (last accessed 26.02.2026)
38. Fisher, R. A. (1936), "The use of multiple measurements in taxonomic problems", *Annals of Eugenics*, Vol. 7, No. 2, pp. 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x
39. Aeberhard, S., Coomans, D., and de Vel, O. (1994), "Comparative analysis of statistical pattern recognition methods in high dimensional settings", *Pattern Recognition*, Vol. 27, No. 8, pp. 1065–1077. DOI: https://doi.org/10.1016/0031-3203(94)90145-7
40. Kohavi, R. (1995), "A study of cross-validation and bootstrap for accuracy estimation and model selection", *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Vol. 2, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. pp. 1137–1143. DOI: https://doi.org/10.5555/1643031.1643047

Received (Надійшла) 26.02.2026
Accepted for publication (Прийнята до друку) 21.05.2026
Publication date (Дата публікації) 27.06.2026

Відомості про авторів / About the Authors

Радюк Павло Михайлович – доктор філософії, Хмельницький національний університет, доцент кафедри комп'ютерних наук, Хмельницький, Україна;
Pavlo Radiuk – PhD, Khmelnytskyi National University, Associate Professor of the Computer Science Department, Khmelnytskyi, Ukraine;
e-mail: radiukp@khmnu.edu.ua
ORCID ID: https://orcid.org/0000-0003-3609-112X
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57216894492

Саченко Анатолій Олексійович – доктор технічних наук, професор, Західноукраїнський національний університет, директор Науково-дослідного інституту інтелектуальних комп'ютерних систем, Тернопіль, Україна; Радомський університет імені Казімежа Пуласького, Радом, Польща;
Anatoliy Sachenko – Doctor of Technical Sciences, Professor, West Ukrainian National University, Director of the Research Institute for Intelligent Computer Systems, Ternopil, Ukraine; Kazimierz Pulaski University of Radom, Radom, Poland;
e-mail: as@wunu.edu.ua
ORCID ID: https://orcid.org/0000-0002-0907-3682
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=35518445600

Мельниченко Олександр Вікторович – доктор філософії, Хмельницький національний університет, старший викладач кафедри комп'ютерної інженерії та інформаційних систем, Хмельницький, Україна;
Oleksandr Melnychenko – PhD, Khmelnytskyi National University, Senior Lecturer of the Computer Engineering and Information Systems Department, Khmelnytskyi, Ukraine;
e-mail: melnychenko@khmnu.edu.ua
ORCID ID: https://orcid.org/0000-0001-8565-7092
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=45961382900

ISSN 2522-9818 (print) *Сучасний стан наукових досліджень та технологій в промисловості. 2026. № 2 (36)*
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2026. No. 2 (36)*

Бруханський Руслан Феоктистович – доктор економічних наук, професор, Західноукраїнський національний університет, завідувач кафедри енергетичних систем та бізнес-аналітики, Тернопіль, Україна;
Ruslan Brukhanskyi – Doctor of Economic Sciences, Professor, West Ukrainian National University, Head of the Energy Systems and Business Analytics Department, Ternopil, Ukraine;
e-mail: r.brukhanskyi@wunu.edu.ua
ORCID ID: https://orcid.org/0000-0002-9360-1109
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57203579103

Каштальян Антоніна Сергіївна – доктор технічних наук, доцент, Хмельницький національний університет, професор кафедри комп'ютерної інженерії та інформаційних систем, Хмельницький, Україна;
Antonina Kashtalian – Doctor of Technical Sciences, Associate Professor, Khmelnytskyi National University, Professor of the Computer Engineering and Information Systems Department, Khmelnytskyi, Ukraine;
e-mail: yantonina@ukr.net
ORCID ID: https://orcid.org/0000-0002-4925-9713
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57218242499

WEDD-RST: ІНТЕРПРЕТОВНИЙ ПІДХІД ДО ВИЯВЛЕННЯ ДЕФЕКТІВ ФОТОЕЛЕКТРИЧНИХ ПАНЕЛЕЙ У КІБЕРФІЗИЧНИХ СИСТЕМАХ

Предметом дослідження є інтерпретованість механізмів виявлення дефектів фотоелектричних панелей у кіберфізичних системах. Поширення інфраструктури відновлюваної енергетики генерує значні обсяги безперервних сенсорних даних, що вимагає розширеного моніторингу для зниження ризиків безпеки, як-от пожежі на фотоелектричних станціях. Однак поширені підходи глибокого навчання працюють як непрозорі "чорні ящики", яким бракує прозорості логічних висновків для ухвалення критично важливих рішень. **Метою цієї роботи** є покращення інтерпретованості систем виявлення дефектів у кіберфізичних середовищах способом побудови виробничої бази знань на основі правил безпосередньо за результатами безперервних вимірювань датчиків. **Завдання дослідження**: розроблення методу адаптивної дискретизації, ідентифікація мінімальних підмножин ознак та вирішення суперечливих шаблонів за допомогою інтегральної оцінки підтримки класу. **Методи** передбачають інтеграцію теорії наближених множин (RST) з новим алгоритмом зваженої ентропійно-щільнісної дискретизації (WEDD). Метод оптимізує пороги на основі подвійного критерію інформаційної ентропії та локальної щільності ймовірності, використовуючи оцінку ядра щільності для розміщення точок зрізу в природних западинах даних. Детерміновані правила вилучаються з нижнього наближення, а ймовірнісні правила синтезуються з граничної ділянки. **Результати** обчислювальних експериментів демонструють високу ефективність запропонованого методу. Перевірена на змодельованому наборі даних SCADA для виявлення пожежної небезпеки, система досягає загальної точності 96,2% і макро-F1-оцінки 0,960. Зокрема, вона дає 100% точність для детермінованих правил і забезпечує повноту 98,0% для критичного класу пожежної небезпеки, сприяючи високій надійності в аварійних сценаріях. Отримана база знань є самодостатнім і легким артефактом у форматі JSON, придатним для розгортання на периферійних пристроях з обмеженими ресурсами. **Висновки:** дослідження утверджує підхід WEDD-RST як строгу структуру для перетворення необроблених даних датчиків на повністю перевірювані правила IF-THEN із явними оцінками впевненості, пропонуючи високонадійне та інтерпретоване рішення для автоматизованого моніторингу безпеки в кіберфізичних середовищах.

Ключові слова: виробнича база знань; теорія наближених множин; дискретизація; грануляція інформації; редукт; класифікація; кіберфізичні системи; фотоелектричний моніторинг; SCADA.

Бібліографічні описи / Bibliographic descriptions

Радюк П. М., Саченко А. О., Мельниченко О. В., Бруханський Р. Ф., Каштальян А. С. WEDD-RST: інтерпретовний підхід до виявлення дефектів фотоелектричних панелей у кіберфізичних системах. *Сучасний стан наукових досліджень та технологій в промисловості*. 2026. № 2 (36). С. 153–172. DOI: https://doi.org/10.30837/2522-9818.2026.2.153

Radiuk, P., Sachenko, A., Melnychenko, O., Brukhanskyi, R., Kashtalian, A. (2026), "WEDD-RST: An interpretive approach to detecting defects in photovoltaic panels in cyber-physical systems", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (36), P. 153–172. DOI: https://doi.org/10.30837/2522-9818.2026.2.153