Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Systematic exploration of signal-based indicators for failure diagnosis in the context of cyber-physical systems

Santiago RUIZ-ARENAS[‡1,2], Zoltán RUSÁK[1],
Imre HORVÁTH[1], Ricardo MEJÍ-GUTIERREZ[2]

[1]*Faculty of Industrial Design Engineering, Delft University of Technology, Delft 2600AA, the Netherlands*

[2]*Grupo de Investigación en Ingeniería de Diseño, Universidad EAFIT, Medellín 050048, Colombia*

E-mail: s.ruizarenas@tudelft.nl; z.rusak@tudelft.nl; i.horvath@tudelft.nl; rmejiag@eafit.edu.co

Received Apr. 24, 2017; Revision accepted Dec. 16, 2017; Crosschecked Feb. 15, 2019

Abstract: Malfunction or breakdown of certain mission critical systems (MCSs) may cause losses of life, damage the environments, and/or lead to high costs. Therefore, recognition of emerging failures and preventive maintenance are essential for reliable operation of MCSs. There is a practical approach for identifying and forecasting failures based on the indicators obtained from real life processes. We aim to develop means for performing active failure diagnosis and forecasting based on monitoring statistical changes of generic signal features in the specific operation modes of the system. In this paper, we present a new approach for identifying emerging failures based on their manifestations in system signals. Our approach benefits from the dynamic management of the system operation modes and from simultaneous processing and characterization of multiple heterogeneous signal sources. It improves the reliability of failure diagnosis and forecasting by investigating system performance in various operation modes, includes reasoning about failures and forming of failures using a failure indicator matrix which is composed of statistical deviation of signal characteristics between normal and failed operations, and implements a failure indicator concept that can be used as a plug and play failure diagnosis and failure forecasting feature of cyber-physical systems. We demonstrate that our method can automate failure diagnosis in the MCSs and lend the MCSs to the development of decision support systems for preventive maintenance.

Key words: Failure indicators; Failure classification; Failure detection and diagnosis; Complex systems
https://doi.org/10.1631/FITEE.1700277 **CLC number:** TP277

1 Introduction

Certain cyber-physical systems (CPSs) are mission critical systems. Their malfunction or breakdown may cause losses of life, damage the environments, and/or lead to high costs. The risk of having critical malfunction or breakdown of systems exponentially grows in line with the increase of complexity (Somani and Vaidya, 1997). Therefore, preventing failures and faults is crucial for these systems. Preventive maintenance can guarantee that mission critical systems can operate without failure (Ponsard et al., 2005). However, it assumes the recognition of emerging failures, which entails a deep understanding of failure manifestations and development of reliable failure detection, diagnosis, and forecasting techniques.

Failures have been interpreted as permanent interruptions of the operation or as a significant deviation from the expected behavior of a system (Miclea and Sanislav, 2011). They may be preceded by the occurrence of faults that are regarded as deviations from the expected behavior of a sub-system or a component (Witczak, 2014). Depending on the type

‡ Corresponding author
ORCID: Santiago RUIZ-ARENAS, https://orcid.org/0000-0002-4018-7370

of failure, faults may progressively occur before the system breaks down. From the perspective of failure prevention, understanding the process of fault manifestation is essential for developing preventive actions (such as failure forecasting). It requires a proper understanding of relevant features that characterize different faults. These features were considered as failure symptoms and were referred to as fault indicators (Fortuna et al., 2007).

Failure forecasting requires two main steps: (1) recognizing the way in which failure manifests for failure diagnosis; (2) implementing prognosis based on the knowledge acquired about each failure manifestation. We present a failure indicator based approach that aims to tackle the first step, in the context of failure diagnosis of the first- and second-generation CPSs. The first- and second-generation CPSs have a relatively high degree of uncertainty in their operation due to the large number of components, their interaction with each other, and the host environment. CPSs are instrumented with multiple sensors, which provide valuable information about the operation of the system. The acquired and generated information is used as a basis for automated decision-making and as control of the operation of the system.

We define the failure indicators as observable and measurable statistical deviations of system operation from the expected behavior of the system. We take advantage of the high level of instrumentation of a CPS to perform parallel analysis of the multiple system signals representing mass, energy, and information flows. Deviations in the signal characteristics are considered as manifestations of faults, i.e., fault indicators. Recognition of the deviations is facilitated by segmentation of signals based on the operation mode of the system. Our failure analysis method strongly uses the fact that statistical variance of signal characteristics in a specific operation mode is smaller than that in the overall system operation. Our main assumption is that failure symptoms are manifested in system signals differently depending on the system's operation modes, and that both symptom occurrence and the lack of symptoms can be used as indicators to determine the type of failure. This gives a good basis for a failure diagnosis method to distinguish signal deviations with higher accuracy, and to provide the means for conducting failure forecasting in our future research. It is our assumption that fault symptoms and the lack of symptoms form a unique pattern (i.e., failure indicator) for each failure mode, providing a reliable foundation for not only the failure diagnosis but also the understanding of the phenomenon of failure manifestation. Time-dependent analysis of failure indicators provides the means for understanding failure evolution and the basis for implementing failure forecasting methods.

2 Overview of indicators-based failure detection and diagnosis methods

The analysis of failure manifestations entails studying fault occurrence as these precede failures, and constitutes its main symptoms. Multiple fault detection and diagnosis techniques have been developed to keep continuous system operation in a cost-effective way. The first approach of the fault detection and diagnosis techniques directly involves system operators and system experts. They are used to analyze system characteristics, such as component wear, sound, and smell, to determine abnormal events that could be related to failures. In this context, the most elementary approaches are limit checking and redundancy.

Limit checking investigates whether certain parameters are within the pre-defined upper and lower limits (Fujimaki et al., 2005). In this method, the fault indicator is interpreted as a deviation of the observed data from a defined threshold (Johnson, 1996). Due to its simplicity, limit checking has been widely used in rule-based approaches (where experts determine the value of a threshold based on their knowledge and experience), or in distribution-based approaches (He et al., 2016). However, the use of thresholds for analyzing failures in non-linear systems is difficult for two reasons: (1) non-linear systems have an unpredictable behavior, which requires the specification of a wide range of threshold values; (2) a wide range of threshold values makes fault detection unreliable (Lee et al., 2009). To overcome these limitations, the adaptable thresholding methods have been proposed (Qi et al., 2007; Patan, 2008; Cholewa et al., 2010; Rezazadeh et al., 2014). These methods make use of historical data or estimate a residual index (i.e., the difference between the observed behavior and the predicted one) to set the range of acceptable performance of a component in

various application contexts. However, these approaches focus on the isolated problem of component faults but do not consider the dynamic interrelationship between the various signals of the system (Sobhani-Tehrani and Khorasani, 2009a).

Redundancy-based fault detection techniques check the consistency between the known variables, inputs, and measured outputs based on relations derived from a system's model or based on the duplication of hardware components (Tornil-Sin et al., 2014). This consistency is evaluated by estimating a residual index. The first implementation of redundancy implies the duplication of hardware components. It requires using at least two identical sensors or actuators as signal and data sources for characterizing the system performance from the perspective of fault diagnosis (Hwang et al., 2012). Nevertheless, redundant implementation of hardware implies an increase in production and maintenance costs (Shui et al., 2009; Yodo and Wang, 2015). This situation, along with the limitations that limit checking entails, led researchers to develop more sophisticated methods that provide a reliable fault detection and diagnosis in a cost-effective way. The literature distinguishes three main approaches: model-based, data-driven, and signal-based failure diagnosis.

2.1 Model-based fault detection methods

Model-based fault detection simultaneously applies several fault detection and diagnosis techniques, such as state observer-based, parity equations, and parameter estimation. These techniques conduct residual estimation based on analytical models (Isermann, 2006b), and the model's output is compared with the observed system's output (Ding, 2008a).

In general, parity equations are mathematical models of the system which are used to determine the residual between the observed system behavior and the estimated one (Sharifi and Langari, 2013); state observer estimates the residuals based on mathematical models that reconstruct process state variables from available system measurements (Soroush, 1997); parameter estimation uses analytical models to compute the residuals of the system parameters as it presupposes that faults are manifested in the physical parameters of the system.

One of the most common residual-based approaches is the Kalman filter (KF) (Rudin et al., 2014; Zhang et al., 2014; Ghanbari, 2015; Irita and Namerikawa, 2015). KF is a state observer-based fault detection method that considers multiple factors for the estimation of the residual index, such as the effect of noise, uncertainty, and faults (Daum, 2015). Although this method has been widely implemented, the KF does not suffice for dealing with the significant nonlinearities and uncertainties (Qi et al., 2007). Some variations of the KF have been developed to face this problem, such as the extended Kalman filter (EKF) (Gao et al., 2015). However, the stability and accuracy of this method are still questionable (Daum, 2015). In the case of parity equations, analysis of residuals is conducted based on the dynamic input-output models that are used to describe the operation of the system (Gertler and Singer, 1990). These techniques are sensitive to measurement noise and system disturbances caused by the derivatives used to compute the residual index (Sobhani-Tehrani and Khorasani, 2009a).

Realization of the residual index as a fault indicator depends on the reliability of the analytical model of a complex system. Simplifications applied to the analytical models may cause a mismatch between the output of the model and observed system behavior (Puig et al., 2015). This mismatch may be enhanced by the effect of noise and disturbances that cannot be modeled, thus causing the residual to become nonzero during the failure-free operation of the system (Luh et al., 2004). This problem limits the applicability of the model-based failure indicators to well defined operation scenarios that are not subject to uncertainties, unlike non-linear systems (Bocaniala and Palade, 2006; Zweigle et al., 2013). Furthermore, the residual indicator is limited to the specific system functions that are widely known and predictable, hindering its application in highly complex systems where fault propagation across subsystems occurs (Sun et al., 2014). The implementation of the residual index in the context of model-based analysis strongly depends on limit checking. It determines the admissible threshold that triggers a failure alarm (Ding, 2008b) whenever it exceeds the defined threshold (Johnson, 1996). This makes the residual index susceptible to the drawbacks of limit checking. In the case of discrete event-based failure diagnosis approaches, the use of the residual as a failure indicator is complex, because the difference between events is not necessarily well defined (Sobhani-Tehrani and Khorasani, 2009b). Furthermore, the way of

measuring the error between the observed behavior and the expected one cannot be generalized for these cases due to the qualitative nature of the system models.

2.2 Data-driven techniques

Data-driven approaches rely on the historical data for determining the occurrence and type of failures. The main advantage of these approaches is that they do not require specific knowledge about the system and its operation in the way that the model-based approaches do. They operate with a large amount of data (Wang et al., 2013), which makes them suitable for failure detection and diagnosis in large and complex systems (Alzghoul et al., 2014). One subset of the data-driven techniques is the classification techniques.

Classification techniques relate to the reference symptoms of the well known failure mode with the observed ones to conduct failure diagnosis (Ye et al., 1993). Bayes classifiers, support vector machine, artificial neural networks, and k-means use fault indicators, such as probability, residuals, and data distance, for fault detection and diagnosis based on classification approaches. Bayes classifiers have been widely used for fault detection and diagnosis (Hood and Ji, 1997; Krishnamachari and Iyengar, 2004; Mehranbod et al., 2005; Zhou et al., 2011). They estimate the probability distributions of different attributes (among others such as data mean, standard deviation, amplitude, and frequency) that represent symptoms, given the class from a training dataset (Ramoni and Sebastiani, 2001). The main drawback of Bayes classifiers is that they require the data with Gaussian probability density distribution and the definition of class-specific densities (Isermann, 2006a). To determine the probability for classification of failures, polynomial classification is conducted based on the assumption that decision rules can be approximated by polynomials (Leonhardt and Ayoubi, 1997). However, the definition of polynomial order is not straightforward and it constitutes a critical step of this approach.

The residual index is used to conduct classification-based fault diagnosis. Artificial neural network (ANN) can be used for classification of faults based on the evaluation of residual signals (Luh et al., 2004; Patan, 2008; Li et al., 2014; Zarei et al., 2014). Opposite to the approach of the previously discussed analysis of signal residual, ANN does not compare the observed output with that of an analytical model of the system. Instead, it determines a model that explains failures based on the input and output data. This approach provides high flexibility, because no previous knowledge in the relationship between the inputs and outputs is required due to the lack of analytical models describing system behavior (Puig et al., 2015).

ANN is typically implemented as a black box approach, in which the relationships between the symptoms and failure modes remain hidden. As a result, ANN cannot provide insight into failure manifestations and failure forming processes. Moreover, the application of ANN has an important drawback since it is easily trapped into a local minimum (Wu et al., 2016), thus preventing the residual reaching its real minimum. Several approaches have been developed to solve this problem. These approaches include the modification of the network parameters (such as the weight and size of the network), introduction or subtraction of neurons, and implementation of statistical parameters (such as mean and variance) for evaluating the weight, error, and output (Bellido and Fernández, 1991). The introduction of a momentum parameter is made to avoid the algorithm converging to a local minimum by modifying the step change of the weight (Shukla et al., 2010).

Support vector machine (SVM) uses a classification distance as a criterion to determine the occurrence of a failure mode (Liu et al., 2010; Muralidharan et al., 2014; Yin et al., 2014; Hang et al., 2016; Swetapadma and Yadav, 2016). In its classical version, this method determines a hyper-plane that optimally divides data corresponding to two different classes through training datasets (Kishore et al., 2016). Although this method has a high predictive accuracy, it is very sensitive to parameter selection (Stockman et al., 2012). It is also sensitive to overlapping groups, as observations near one another are treated alike (Berk, 2008). Although the classical implementation is limited to a two-class classification problem, multiclass classification can be mostly used by decomposing the problem into several binary analyses that involve the multiple implementation of binary SVM (Mathur and Foody, 2008). One-vs-all SVM, pairwise SVM, error-correcting output code (ECOC), and all-at-once SVM are the most common methods for conducting multiclass

prediction (Abe, 2010). Nevertheless, these are computationally more expensive. Although the nearest neighbor-based fault detection is sensitive to overlapping groups and uneven data density, cluster density information has been proved useful in improving the accuracy of the classification, overcoming the uneven data density problem (Kang et al., 2016).

Data distance-based fault indicators can be implemented without previous specification of the class during the learning process, as it is the case of k-means and ANN based on the radial basis functions. These types of methods are known as the "unsupervised failure classification," as they can recognize hidden relationships in unlabeled data (Abdolsamadi et al., 2015). They enable the classification of not only the known failures but also the emerging ones (Zhang, 2016). They determine the distance of any new observation with respect to the center of each of the available classified failure clusters. However, these methods are not as powerful or reliable as the supervised ones, because the distance index used as the failure indicator presents problems for managing overlapping classes (Isermann, 2006b). Additionally, supervised classification methods raise the stability-plasticity dilemma. This refers to the behavior that causes the system to forget the already learned categories when it learns new clusters and patterns of relationships (Fernando and Surgenor, 2017).

Data-driven techniques are the very flexible methods, as they can discover variable relationships based on historical data. The learning capabilities they provide make them very suitable for failure analysis, particularly for facing the emergence of the unknown failure modes. Nevertheless, these methods do not provide information that allows understanding of failure manifestations. Classification methods have proved very useful for pattern recognition and for discriminating data coming from different classes. However, the relationship between predictors is difficult to understand, hindering failure analysis.

2.3 Signal-based analysis techniques

Signal-based analysis is underpinned by the development of signal models for determining the deviations from a reference signal. These deviations are manifested on signal attributes, which can be extracted in time, frequency, or time-frequency domains. Signal analysis in the time domain studies the measured variables as a function of time by focusing on the geometrical factors, such as amplitudes, peaks, and statistical parameters. Cross-correlation analysis and statistical features-based analysis are usually conducted in the time domain (Lei, 2017). The cross-correlation analysis can be used to evaluate the similarity between two signals, e.g., when comparing a failure-free reference signal with an observed one with regard to detecting failures. The statistical features-based analysis makes use of signal statistical properties for failure detection. However, some sorts of distortions and small disturbances of signals are difficult to detect in the time domain. Because of these limitations, frequency-domain analysis (FDA) is presented as another option for feature extraction. FDA transforms the captured signal from the time domain to the frequency domain. The most popular technique is Fourier analysis, which represents an arbitrary function in a finite interval as a sum of sinusoids (Alencar and da Rocha, 2005). While the representation of signals in the time domain indicates the evolution of the signal amplitude over time, representation in the frequency domain shows how quickly such changes take place (Montaño et al., 2007). This approach has, however, an important limitation. FDA does not allow determining the time in which failure symptoms occur, thus preventing the interpretation of the failure manifestations and the analysis of their root cause.

The need for time-frequency domain analysis (TFDA) typically arises because of the lack of time information in frequency-domain analysis (Ramos et al., 2009). Information about time may be required in some cases, as the actual signal frequency composition may change with time. For this reason, TFDA analysis captures both frequency and time information about the processed signal. Among others, the widely known methods for performing TFDA analysis are short time Fourier (STF) analysis, Wigner-Ville distribution (WVD) analysis, and wavelets analysis. The wavelets analysis method is the most popular one. As with other signal-based analyses, it represents signals as a composition of a set of basis functions. Nevertheless, the selection of such functions implies minimizing the complexity of its representation (Albertos and Mareels, 2010).

Signal-based prediction is a very useful technique for evaluating system performance. The reason is that signals typically have features that may be indicators of failure onset and their manifestation.

However, these methods rely on prior knowledge of the way in which signal characteristics are influenced by the occurring failures. This makes it difficult to use these methods to evaluate the emerging failures and to automate the failure diagnosis process. Control actions can have a negative effect on signal-based failure analysis. They modify system settings, thus altering signal properties. This situation can lead to false failure alarms, as variations in such properties can be misinterpreted as failure effects.

In conclusion, all the aforementioned fault diagnosis methods are typically embedded into complex algorithms that aim at detecting failures. Their implementation is limited to the analysis of specific parameters or component functions. Therefore, they do not facilitate reasoning on the effect of faults at the system level. They still face four important challenges: (1) scalability of the fault indicator to deal with emerging functions and their possible faults without losing autonomy in failure diagnosis; (2) automated failure diagnosis based on reasoning with multiple signals; (3) understanding of failure manifestations and failure evolution; (4) scalability of the method for forecasting purposes. To address these challenges, we introduce a novel concept for exploring failure indicators based on heterogeneous signal processing. This concept is a data-driven approach to failure diagnosis and forecasting. It allows itself to learn the relationships between symptoms and failure modes; therefore, unknown failures can be learned. However, unlike the aforementioned methods that conduct failure classification based on data density, the proposed approach performs classification based on the pattern formed by the combination of symptoms and the lack of symptoms that each failure mode produces.

3 Elements of the theory underpinning the proposed failure indicator

3.1 Theoretical considerations

The proposed method intends to obtain failure indicators (**FI**). It relies on the segmentation of signals which belong to a set of sensor signals (S_{S}) obtained from sensors installed in the system:

$$S_{\mathrm{S}} = \{S_{\mathrm{S}_j} | j = 1, 2, \ldots, p\}. \tag{1}$$

In addition to these measures, the actuator information is collected through a set of actuator signals (S_{A}) that are used as the trigger for the segmentation process. Actuator signals are symbolically represented as

$$S_{\mathrm{A}} = \{S_{\mathrm{A}_j} | j = 1, 2, \ldots, n\}. \tag{2}$$

Hence, S_{S} and S_{A} together form the set of system signals (S), $S = S_{\mathrm{A}} \cup S_{\mathrm{S}}$.

The natural variations in the surrounding environment and the frequent changes of the use conditions demand that the system should present multiple operative behaviors. Every operative behavior is enabled by a particular combination of system settings. A system operation mode (SOM) describes system behavior at time t based on the actual system settings. Our formal definition of an SOM is a singular combination of all the component operation modes (COM) of the system at a particular time t. COMs can be regarded as the component state at time t. From now on, the set of all potential COMs (ζ) of $\mathrm{S}_{\mathrm{A}j}$ can be expressed as

$$E_{S_{\mathrm{A}_j}} = \{\zeta_1, \zeta_2, \ldots, \zeta_u\}. \tag{3}$$

All the above-presented definitions are necessary for the technical introduction of the concept of "system level failure indicator." Denote ζ_d for any particular SOM at time t, where $\zeta(t) = \{\zeta_d(t) | d = 1, 2, \ldots, l\}$. Considering that SOM denotes a single combination of COM at time t, ζ_d can be expressed as

$$\zeta_d = \{\zeta_{S_{\mathrm{A}_1}}(t), \zeta_{S_{\mathrm{A}_2}}(t), \ldots, \zeta_{S_{\mathrm{A}_k}}(t)\}. \tag{4}$$

For instance, let us consider the example of a greenhouse where the water tank is irrigating and heating up the water in the reservoir at the same time t. Considering the aforementioned notation, the system operation mode ζ_d at time t can be represented as $\zeta_d = \{\zeta_{S_{\mathrm{A}_1}}(t), \zeta_{S_{\mathrm{A}_2}}(t), \zeta_{S_{\mathrm{A}_3}}(t), \zeta_{S_{\mathrm{A}_4}}(t)\}$, where, for example, it may represent a situation such as $\zeta_{S_{\mathrm{A}_1}}(t) = \mathrm{ValveOpen}$, $\zeta_{S_{\mathrm{A}_2}}(t) = \mathrm{ValveClose}$, $\zeta_{S_{\mathrm{A}_3}}(t) = \mathrm{ValveClose}$, and $\zeta_{S_{\mathrm{A}_4}}(t) = \mathrm{HeaterOn}$.

Some other important concepts should be introduced for defining a system-level failure indicator. These are "failure mode" (F_{r}) and "signal segment" (S_{g}). Failure mode is a particular type of failure that can occur in a system. Every failure mode presents certain characteristic symptoms (ϕ) that can be used in failure diagnosis. We consider a symptom any deviation of a signal from its expected behavior, which can be explained by the effect of a failure.

A "signal segment" (S_{g}) is a part of a signal that is measured during a particular time interval. It includes all measured/sampled values of a signal (S_{S_j}) between a start point in time t_{s} and an end point in time t_{e}. The start and end points are determined by any variations on any COM. Therefore, S_{g} represents a system variable during a specific SOM (ζ_d). The same SOM (ζ_d) can occur several times during the whole signal measurement period, and may be detected several times in the whole signal during the system operation. Therefore, each individual segment, representing a SOM, is expressed as

$$S_{\mathrm{g_h}} = [S_{\mathrm{S}}(t_{\mathrm{s}}), \ldots, S_{\mathrm{S}}(t_{\mathrm{e}})], \tag{5}$$

where $S_{\mathrm{S}}(t)$ is the measured value of a particular sensor signal (S_{S}) in a specific time t.

3.2 Description of the proposed failure indicator (FI)

We define the "failure indicator" (**FI**) as an $m \times n$ matrix whose rows are system sensor signals (S_{S}) and columns are system actuator signals (S_{A}). In the **FI** matrix, each item $\mathrm{FI}_{i,j}$ is composed by the presence or absence of a symptom ϕ in a segment $S_{\mathrm{g_h}}$ from the signal S_{S_j} corresponding to a particular SOM (ζ_d), expressed as

$$\mathbf{FI} = \begin{bmatrix} \phi_{1,1} & \cdots & \phi_{1,n} \\ \vdots & & \vdots \\ \phi_{m,1} & \cdots & \phi_{m,n} \end{bmatrix}, \tag{6}$$

where $\phi(S_{\mathrm{S}_j}, \zeta_d)$ equals 1 if there is a symptom, or 0 if there is a lack of symptom.

To build a failure indicator, it can be inferred that failures are derived because of the deviation between the signal features obtained from the regular operation (O) (i.e., without failure) and irregular operation of the system (i.e., under fault/failure conditions). The reference behavior (∂) is introduced to describe the characteristic behavior of either a particular "failure mode" (F_{r}) or a failure-free system. In every SOM (ζ_d), data corresponding to the observed signal (S_{S_j}) are compared with data of the system's reference behavior ∂, corresponding to a given signal (S_{S_j}) during ζ_d. A detailed description of the process to be followed to build an **FI** is given in the following sections.

3.3 Methods for deriving failure indicators

Signals collected through data from the system is collected in a vector $\boldsymbol{S}$, where signals from actuators S_{A} and sensors S_{S} are arranged as

$$\boldsymbol{S} = [S_{\mathrm{A}_1}, \ldots, S_{\mathrm{A}_n}, S_{\mathrm{S}_1}, \ldots, S_{\mathrm{S}_p}], \tag{7}$$

where n denotes the number of actuators, and p is the number of sensors. All signals have the same time t; therefore, they are stored as follows:

$$\boldsymbol{S}(t) = [S_{\mathrm{A}_1}(t), \ldots, S_{\mathrm{A}_n}(t), S_{\mathrm{S}_1}(t), \ldots, S_{\mathrm{S}_p}(t)]. \tag{8}$$

The entire system database $\boldsymbol{D}$ consists of the storage of $\boldsymbol{S}$ in each sample time t:

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{S}(t=1) \\ \boldsymbol{S}(t=2) \\ \vdots \\ \boldsymbol{S}(t=k) \end{bmatrix}. \tag{9}$$

3.3.1 Identification and representation of the operation modes of a system

Each S_{A} from the vector $\boldsymbol{S}(t)$ corresponds to a COM. The SOMs are the results of a unique combination of the COMs (S_{A_j}). The progression in time t enables the SOM identification by analyzing if $S_{\mathrm{A}_j}(t) \neq S_{\mathrm{A}_j}(t+1)$, which implies a change in the SOM. For this reason, a collection of COM is gathered in an operation mode matrix $\mathbf{OM}_{l \times n}$ with l rows (representing the SOM in the sampling time t) and n columns (corresponding to the number of actuator's signals S_{A}), and it can be denoted as

$$\mathbf{OM}_{l \times n} = \begin{bmatrix} S_{\mathrm{A}_1}^{\zeta_1} & \cdots & S_{\mathrm{A}_n}^{\zeta_1} \\ \vdots & & \vdots \\ S_{\mathrm{A}_1}^{\zeta_l} & \cdots & S_{\mathrm{A}_n}^{\zeta_l} \end{bmatrix}. \tag{10}$$

The time instant t in $S_{\mathrm{A}j}(t) \neq S_{\mathrm{A}_j}(t+1)$ is extracted to determine the moment in which the SOM changes. These times are collected in a $\boldsymbol{T}_{2 \times l}$ matrix, where the starting time of a ζ_d is denoted as t_{s}, and its end time SOM is denoted as t_{e}. Therefore, t_{s_1} denotes the time instant when ζ_1 starts, and t_{e_1} denotes the time instant when ζ_1 switches to the next SOM.

$$\boldsymbol{T} = \begin{bmatrix} t_{\mathrm{s}_1} & t_{\mathrm{e}_1} \\ \vdots & \vdots \\ t_{\mathrm{s}_l} & t_{\mathrm{e}_l} \end{bmatrix}. \tag{11}$$

Considering that each ζ_d may occur several times during the system's operation, **OM** contains rows that are repeated. A matrix extracts all possible ζ's by filtering **OM**, avoiding repetition created. This matrix is $\mathbf{SOM}_{v\times n}$, where v is the total number of ζ's that may occur in the system, and n is the number of actuator's signals S_{A}.

3.3.2 Segmentation of signals

Each row of $\boldsymbol{T}$ is used as a reference for signal segmentation. Data represented in $\boldsymbol{D}$ are extracted and stored in vectors denoted as $\boldsymbol{S}_{\mathrm{g}_h}$, where $h=1,2,\ldots,w$ (w is the total number of signal segments). Time instants t_{s} and t_{e} stored in row h of matrix $\boldsymbol{T}$ are used as references to extract all data corresponding to S_{S} from $\boldsymbol{D}$, in the time interval between t_{s_h} and t_{e_h}. We thus have

$$S_{\mathrm{g}_h}=\{S_{\mathrm{S}}(t_{\mathrm{s}}),\ldots,S_{\mathrm{S}}(t_{\mathrm{e}})\}. \tag{12}$$

Therefore, there is a dataset DS that contains all signal segments:

$$\mathrm{DS}=\{S_{\mathrm{g}_1},S_{\mathrm{g}_2},\ldots,S_{\mathrm{g}_w}\}. \tag{13}$$

The segmented data are stored in an array called the "segment" that organizes each S_{g_h} depending on the SOM occurring in the period of time when signal segment S_{g_h} is sensed, so that

$$\mathrm{Segment.OM_g}\leftarrow S_{\mathrm{g}_h}. \tag{14}$$

For example, if segments S_{g_7}, $S_{\mathrm{g}_{15}}$, and $S_{\mathrm{g}_{30}}$ are sensed in the periods of time when the system is in the operation mode OM_3, and the segments S_{g_5}, $S_{\mathrm{g}_{40}}$, $S_{\mathrm{g}_{58}}$, and $S_{\mathrm{g}_{72}}$ are sensed in periods when the system is in the operation mode OM_1, then $\mathrm{Segment.OM}_3=\{S_{\mathrm{g}_7},S_{\mathrm{g}_{15}},S_{\mathrm{g}_{30}}\}$ and $\mathrm{Segment.OM}_1=\{S_{\mathrm{g}_5},S_{\mathrm{g}_{40}},S_{\mathrm{g}_{58}},S_{\mathrm{g}_{72}}\}$.

3.3.3 Characterization of signal segments

A signal feature, $a(\mathrm{Segment.OM_g})$, provides a simple representation that can capture the trends and characteristic features of signal segments. Depending on the type of signal, the feature may be determined by the parameters relevant to time- or frequency-domain signal analysis. Given $a(\mathrm{Segment.OM_g})$, there is a vector $\boldsymbol{A}$ that groups the features of all segments S_{g_h} stored in the Segment.OM, and we have

$$\boldsymbol{A}_{\zeta_d}=\begin{bmatrix} a(\mathrm{Segment.OM}_1)\\ \vdots\\ a(\mathrm{Segment.OM}_z)\end{bmatrix}, \tag{15}$$

where z is the total number of segments corresponding to ζ_d. Among others, features, such as the derivative of the whole segment $\left(\frac{\mathrm{d}S}{\mathrm{d}t}\right)$, area of the segment $\left(\int_{t_{\mathrm{e}}}^{t_{\mathrm{s}}}S_{\mathrm{S}}\right)$, change of the signal's slope $\left(\frac{\mathrm{d}^2S}{\mathrm{d}t^2}\right)$, mean of the slope, number of zero crossings, and mean of the derivative (point by point), can be used to characterize signal segments. The selection of α depends on the particular characteristics of each signal.

3.3.4 Evaluation of the deviation of signals

The evaluation of the deviation of signals aims to identify significant anomalies, which can be caused by failures, in the operation of the system. This process is based on the application of a statistical test (ST), in which a sample coming from the system's reference behavior $A^{\partial}_{\zeta_d}$ during the system's operation mode ζ_d, is compared with the system's observed behavior $A^{O}_{\zeta_d}$ during ζ_d. The observed behavior may be or may not be under the effect of a particular failure. The aim of ST is to determine whether $A^{O}_{\zeta_d}$ presents any deviation from the reference behavior ∂ due to the effect of a failure. Considering $\mathrm{ST}(A^{\partial}_{\zeta_d},A^{O}_{\zeta_d})$, our null hypothesis H_0 states that samples $A^{\partial}_{\zeta_d}$ and $A^{O}_{\zeta_d}$ have the same distribution Θ, expressed as

$$H_0:\mu_{\partial}\in\Theta\wedge\mu_0\in\Theta, \tag{16}$$

where μ_{∂} is the mean of sample $A^{\partial}_{\zeta_d}$ and μ_0 is the mean of the observed data $A^{O}_{\zeta_d}$. Conversely, our alternative hypothesis H_1 states that samples $A^{\partial}_{\zeta_d}$ and $A^{O}_{\zeta_d}$ belong to different distributions:

$$H_1:\mu_{\partial}\in\Theta\wedge\mu_0\notin\Theta. \tag{17}$$

Therefore,

$$\Phi(A^{\partial}_{\zeta_d},A^{O}_{\zeta_d})=\begin{cases}1, & \text{if } \mathrm{ST}(A^{\partial}_{\zeta_d},A^{O}_{\zeta_d})\in\Omega,\\ 0, & \text{if } \mathrm{ST}(A^{\partial}_{\zeta_d},A^{O}_{\zeta_d})\notin\Omega,\end{cases} \tag{18}$$

where Ω is the rejection region. When $\Phi(A^{\partial}_{\zeta_d},A^{O}_{\zeta_d})=1$, it means that H_0 is rejected, and therefore $\phi(S_{\mathrm{S}_j},\zeta_D)=1$. When $\Phi(A^{\partial}_{\zeta_d},A^{O}_{\zeta_d})=0$, it means that H_0 is accepted, and thus $\phi(S_{\mathrm{S}_j},\zeta_D)=0$. To determine the result of Φ and $\phi(S_{\mathrm{S}_j},\zeta_D)$, the concept of p-value is used. The p-value determines the

probability that $\mathrm{ST}(A^{\partial}_{\zeta_d}, A^{O}_{\zeta_d}) \in \Omega$ due to the effect of failure. In our case, we consider a p-value of 0.05 as the threshold that determines the significance of the results:

$$\begin{cases} \mathrm{ST}(A^{\partial}_{\zeta_d}, A^{O}_{\zeta_d}) \in \Omega, & \text{if } p < 0.05, \\ \mathrm{ST}(A^{\partial}_{\zeta_d}, A^{O}_{\zeta_d}) \notin \Omega, & \text{if } p > 0.05. \end{cases} \tag{19}$$

Considering that this process is conducted for all signal segments, there is a p-value $p_{i,j}$ for each element of matrix **FI**, so $\phi(S_{\mathrm{S}_j}, \zeta_D)$ can be determined. The result of this process is matrix **FI**, where $\mathrm{FI}_{i,j} = 1$ if $p_{i,j} < 0.05$, and $\mathrm{FI}_{i,j} = 0$ if $p_{i,j} < 0.05$. Therefore, **FI** is a matrix composed by zeros and ones.

Different statistical tests can be implemented depending on the type of data distribution. For example, the t-test can be used for normal distribution, while the Kruskal-Wallis test can be used whenever the analyzed data do not present a parametric distribution. In addition, the decision algorithms, such as Monte Carlo, can be used to determine the significance of the difference between $A^{\partial}_{\zeta_d}$ and $A^{O}_{\zeta_d}$.

3.3.5 Deriving operation mode-dependent failure indicators

System-level failure indicators are generated based on the frequency of symptom occurrence in the multiple operating scenarios (S_{c}) subject to the same failure mode r. Each scenario is an operation context determined by the combination of different system initial conditions or system settings caused by the surrounding environment and the user's manipulation. S_{c} are considered as they influence system signals and SOMs. A resultant **FI** matrix $\mathbf{FI}^r$ is formed as

$$\begin{cases} \text{if } \sum_{S_{\mathrm{c}1}}^{S_{\mathrm{c}}\psi} \mathrm{FI}_{i,j} < 0.95\psi, & \text{then } \mathrm{FI}^r_{i,j} = 0, \\ \text{if } \sum_{S_{\mathrm{c}1}}^{S_{\mathrm{c}}\psi} \mathrm{FI}_{i,j} > 0.95\psi, & \text{then } \mathrm{FI}^r_{i,j} = 1, \end{cases} \tag{20}$$

where r denotes the reference behaviors that **FI** represents ($r = 0, 1, \ldots, u$), u is the total number of reference failure modes that are available in the system (for clarity, we denote $\mathbf{FI}^r$ as the reference indicators and **FI** as the observed indicator corresponding to the current failure performance), S_{c} is each of the datasets corresponding to the scenarios analyzed, and ψ is the total number of the cases analyzed. If 95% of the $\mathrm{FI}_{i,j}$ considered for deriving $\mathrm{FI}^r_{i,j}$ satisfy $\phi(S_{\mathrm{S}_j}, \zeta_D) = 1$, then $\mathrm{FI}^r_{i,j}$ is red. Conversely, if 95% of the $\mathrm{FI}_{i,j}$ considered for deriving $\mathrm{FI}^r_{i,j}$ satisfy $\phi(S_{\mathrm{S}_j}, \zeta_D) = 0$, then it is green. Yellow cells are those that do not reach a 95% either in $\phi(S_{\mathrm{S}_j}, \zeta_D) = 1$ or in $\phi(S_{\mathrm{S}_j}, \zeta_D) = 0$ (Fig. 1).

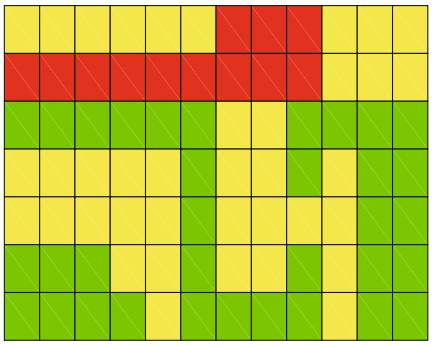

Fig. 1 Example of a colored R matrix

References to color refer to the online version of this figure

3.3.6 Matching equation

The matching of matrix **FI** of the actual system operation to identify failure indicators stored in a library is achieved based on the comparison of the similarity with a failure indicator reference $\mathbf{FI}^r$. It is a pairwise comparison process during which their degree of similarity, MD, is measured, so that

$$\mathrm{MD} = 100 - \frac{C}{(C + \mathrm{NC}) - W} \times 100, \tag{21}$$

where C is the number of the symptoms and the lack of symptoms in common. The failure indicator reference $\mathbf{FI}^r$ and the observed failure indicator correspond to the actual system behavior **FI**. NC is the number of symptoms that differ from those in C, and W is the number of operation modes with no data.

4 Operationalization of the proposed failure indicator concept

The proposed concept of deriving failure indicators is implemented as a smart (software-based) augmentation module of a cyber-physical system and operationalized as a transversal process that is concurrently executed with the regular system operation. While the traditional complex system transforms the flows of material, energy, and information to provide its regular functions and services, the proposed indicator-based failure detection mechanism takes the system signals as inputs, along with

the report of the component operation modes corresponding to each S_{A_j} to conduct failure diagnosis and forecasting.

After a complicated system has been tuned to proper operation, our failure indicator based diagnosis and forecasting method can be applied. Fig. 2 shows the operationalization of the proposed method. The first step is to determine the failure indicator matrices $\mathbf{FI}^r$, in which the system signals are recorded and segmented based on the system's operation modes. This process is initially conducted when the system is operating under regular conditions, i.e., without the presence of faults or failures. The result of this step is the system's reference behavior ∂, which is stored in the failure indicator library to be used as a basis for further failure diagnosis processes. This process is not static, as the reference behavior can be updated whenever there is a system upgrade or the system is reconfigured.

4.1 Failure diagnosis

The reference behavior ∂ is used as a basis for a continuous failure detection and diagnosis process. System signals are recorded during a specified time window and used as inputs to determine the $\mathbf{FI}^r$ matrix of the reference behavior. The time window should be defined for each individual system based on its dynamic behavior and typical use cases in which the system operates. **FI** is derived through the comparison of the observed dataset A^{∂} with the reference dataset A^{∂} to determine the possible symptoms of faults (Fig. 3). If no symptoms are identified (i.e., the statistical deviation of the signal features is not significant), the system will continue to operate. The occurrence of the symptoms (significant deviations) stops the loop and starts the execution of the failure indicator matching process to identify the failure modes. Fig. 4 shows the pseudo algorithm matching the failure indicator matrices to the failure indicator library.

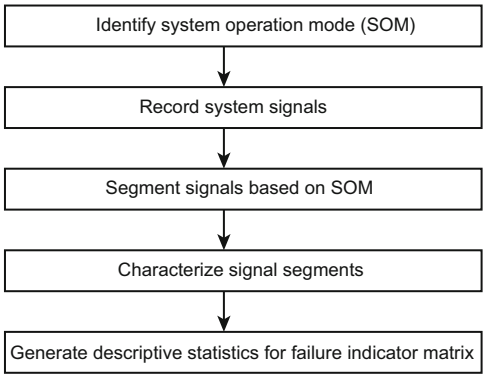

Fig. 3 Operationalization of the identification of FI

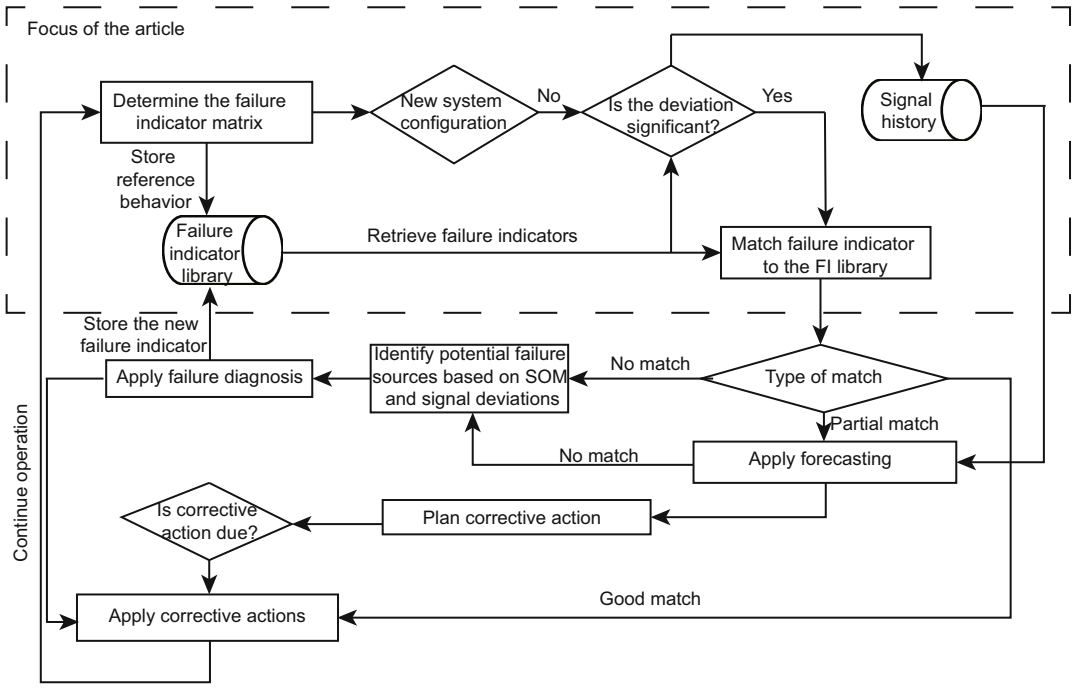

Fig. 2 Operationalization of the proposed method

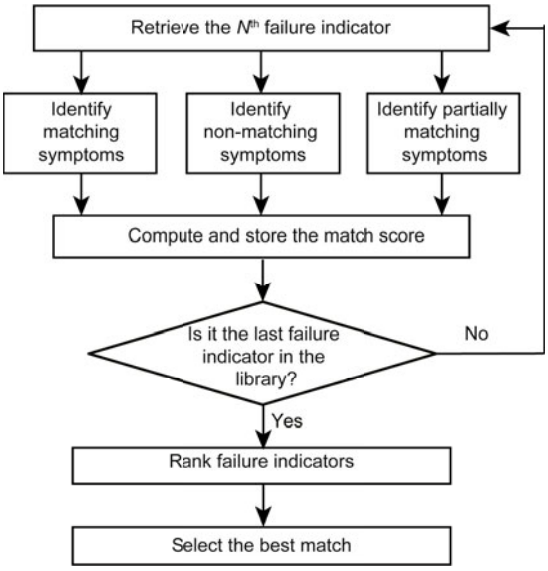

Fig. 4 Operationalization of the matching process

The failure indicator matching algorithm retrieves and evaluates each failure indicator $\mathbf{FI}^r$ stored in the library and computes its similarity score based on the equation presented in Section 3.3.6. The degree of the matching MD helps rank possible failure modes and select the best match for the actual failure indicator **FI**. If the MD is higher than a preset threshold (such as 70%), then the match is considered to be a "good match." If it is lower than 30%, the match is considered to be "no match," and anything between these threshold values is classified as a "partial match." In the case of a "good match," the observed behavior is associated to a particular failure mode, and a relevant maintenance action is scheduled. In the case of a "partial match," the failure indicator shows either a forming failure (already known) or an unknown failure, which triggers the execution of the forecasting algorithm or the process of identification of a new failure.

4.2 Learning process of new failure indicators

Whenever there is no match and the forecasted matrix does not correspond to any of the failure modes available in the system's library, potential failure sources should be identified. This situation occurs when the actual failure mode is unknown and its corresponding failure indicator is not stored in the system's library. The observed symptoms in the failure indicator matrix, however, can provide relevant information about the potential source of errors as they are associated with the system components. The types of sensors and their positions in the system architecture contain valuable information about the origin of the failure. The probability of the failure for each identified component is analyzed and ranked. This process should be conducted by a maintenance expert. Once it is possible to provide a reasonable explanation to the pattern of symptoms, the observed failure indicator is stored in the failure indicator library (Fig. 5). The extension of the failure indicator library enables the determination of the corrective action during a new occurrence of such a failure mode and preventive actions, when the failure forming process has started.

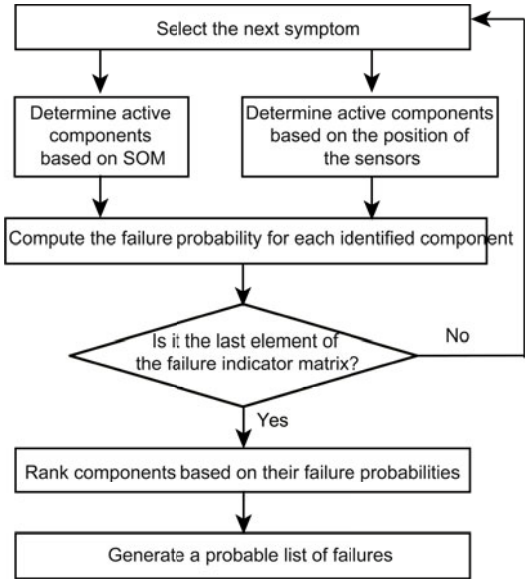

Fig. 5 Description of the learning process

5 Experimental validation of signal-based indicator exploration in failure diagnosis

5.1 Description of the experiment

To validate the proposed failure diagnosis method in a realistic environment, we have built a testbed of a cyber-physical greenhouse (Fig. 6). This system consists of two plant bed units and a central unit. Each of the units is equipped with an independent control, an XBee-based transmission unit, and a real-time clock (RTC), besides its corresponding sensors and actuators. A detailed description of the system components and their locations is presented

in Fig. 7. All data sensed are sent to a coordinator node, which redirects it to a faster processing unit where the signal segmentation-based failure diagnosis algorithm is executed. This processing unit gathers all system data, synchronizes them, facilitates their visualization, and executes the failure diagnosis algorithm implemented in Matlab. The collected data are arranged in a matrix $\boldsymbol{S}$, where actuator signals S_{A} and sensor signals S_{S} are recorded. A detailed description of the signals considered for failure diagnosis is shown in Tables 1 and 2.

Fig. 6 Setup of the experiment

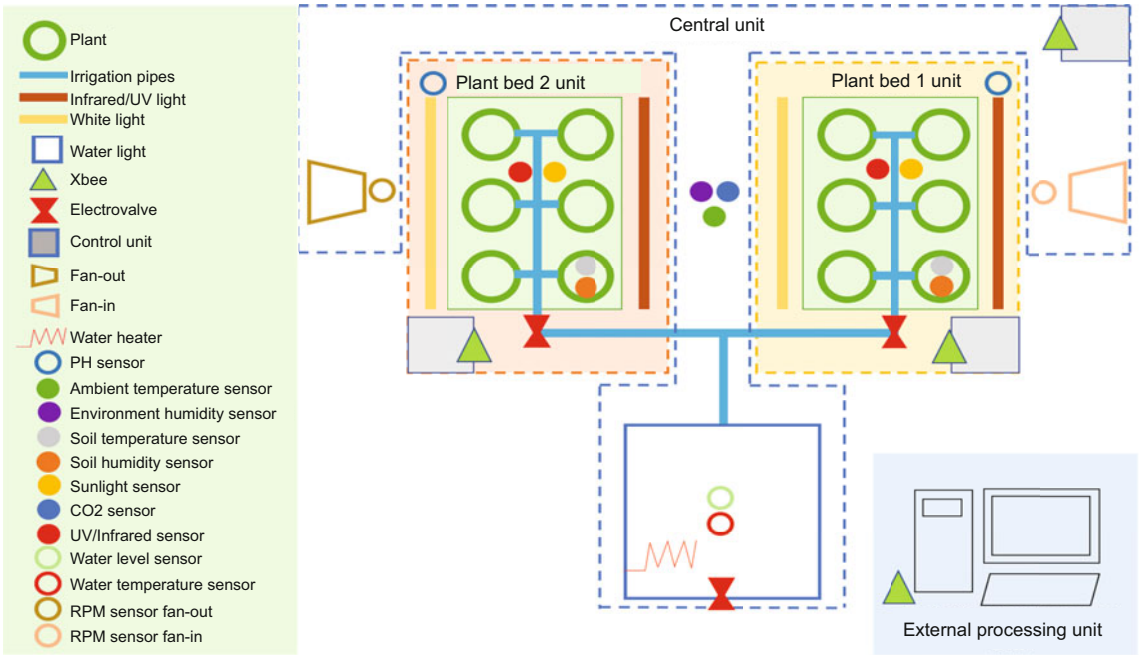

Fig. 7 Configuration of the testbed

Table 1 Description of the system variables

System component	Variable	Description
LDR sensor	S_{S_1}	White light level on plant bed 1
Sensor	S_{S_2}	Lighting power consumption on plant bed 1
Soil moisture sensor	S_{S_3}	Soil humidity on plant bed 1
Soil temperature sensor	S_{S_4}	Soil temperature on plant bed 1
PAR light sensor	S_{S_5}	PAR light on plant bed 1
Sunlight sensor	S_{S_6}	UV light level on plant bed 11
PH sensor	S_{S_7}	PH level on plant bed 1
Water level sensor	S_{S_8}	Water level in tank
Water temperature sensor	S_{S_9}	Water temperature in tank
Thermohygrometer	$S_{S_{10}}$	Greenhouse temperature
Thermohygrometer	$S_{S_{11}}$	Relative humidity into the greenhouse
CO_2 sensor	$S_{S_{12}}$	CO_2 level in the greenhouse
RPM sensor	$S_{S_{13}}$	RPM fan-in
RPM sensor	$S_{S_{14}}$	RPM fan-out
LDR sensor	$S_{S_{15}}$	White light level on plant bed 2
Sensor	$S_{S_{16}}$	Lighting power consumption on plant bed 2
Soil moisture sensor	$S_{S_{17}}$	Soil humidity on plant bed 2
Soil temperature sensor	$S_{S_{18}}$	Soil temperature on plant bed 2
PAR light sensor	$S_{S_{19}}$	PAR light on plant bed 2
Sunlight sensor	$S_{S_{20}}$	UV light level on plant bed 2
PH sensor	$S_{S_{21}}$	PH level on plant bed 2

Table 2 Description of the actuator variables of the greenhouse

System component	Variable	Description	Domain/Set-point
Electro valve Plant bed 1	S_{A_1}	Irrigation valve of plant bed 1	$E_{S_{A_1}}$ = {ValveClose, ValveOpen}
Electro valve Water reservoir	S_{A_2}	Inlet tank valve	$E_{S_{A_2}}$ = {ValveClose, ValveOpen}
Heater	S_{A_3}	Water resistance for the heater	$E_{S_{A_3}}$ = {ResistanceOff, ResistanceOn}
Fan-in	S_{A_4}	Fan-in of the central unit	$E_{S_{A_4}}$ = {Fan-inOff, Fan-inOn}
Fan-out	S_{A_5}	Fan-out of the central unit	$E_{S_{A_5}}$ = {Fan-inOff, Fan-inOn}
Electro valve Plant bed 2	S_{A_6}	Irrigation valve of plant bed 2	$E_{S_{A_6}}$ = {ValveClose, ValveOpen}

5.2 Signal pre-processing for failure diagnosis

Actuator signals S_A are binary signals with possible values of 0 or 1, where 0 represents the inactive state and 1 presents the active state. For example, the <ResistanceOff> component operation mode corresponding to S_{A_3} is reported as 0, while the <ResistanceOn> component operation mode corresponding to the same component is reported as 1. Sensor signals S_S require sophisticated signal processing techniques. A digital filter (moving average) is implemented to clean the noise and disturbances presented during the signal's sensing of S_S. The filter averages the last 40 data measurements before delivering them to the external processing unit. The data are cleaned from the noise caused by the communication problems between the coordinator node and the processing unit. This filtering process is conducted by analyzing the signal derivatives, identifying the measurements $S_{S_j}(t)$ in which $S_{S_j}(t+1) - S_{S_j}(t) < \mathrm{SD}_{S_{S_j}}$ (where $\mathrm{SD}_{S_{S_j}}$ denotes the standard deviation of S_{S_j}), and replacing them with the previous value of $S_{S_j}(t-1)$. All system signals have a sampling time of 1 s.

5.3 Validation process

The goals of this experiment are: (1) analyzing the distinctive power of failure indicators, i.e., how the reliable failure indicators can distinguish different types of failures; (2) evaluating to what extent

the derived failure indicators are sensitive to the selected signal features; (3) exploring the effectiveness of the matching algorithm for operationalizing the failure diagnosis; (4) demonstrating the benefits of the signal segmentation based on operation modes in the failure diagnosis process. Three different failure modes were injected to the system: (1) a leak on the water reservoir, (2) an obstruction in the irrigation pipes, and (3) an irregular fan operation. A set of data obtained during the regular operation of the system in different periods of the day were collected and stored as reference A^{∂} in the different failure injection experiments. All experiments were conducted with the goal to identify what is the corresponding failure indicator of each failure mode and to evaluate to what extent it can be diagnosed through the proposed method. The coherence of the obtained indicators was studied by analyzing 9400 independent data samples measured during five different injection processes for each failure mode over five days. The procedure of failure injection is explained in the following:

1. Failure 1 (F_1): tank leak. To inject the failure of a leak in a controlled manner, a drain valve was installed on one of the walls of the tank, close to its bottom and below the inlet and outlet valves. This installation enabled the manipulation regulation of the outflow of leaking water that leaves the tank through this drain value. For this purpose, four different opening levels of the valve were installed to produce a failure with different intensities. Each opening level constitutes a different experiment.

2. Failure 2 (F_2): blockage in the irrigation pipes. The irrigation holes corresponding to plant bed 2 were partially obstructed with teflon tape. This failure forming process was conducted by covering different numbers of irrigation holes; therefore, the failure could be repeated with different but realistic scenarios. The manipulated variable was the flow rate of irrigation.

3. Failure 3 (F_3): irregular fan operation. A resistance reducing the electrical current that feeds the fan was installed to alter the regular speed of rotation of the inlet air fan. The manipulated variable was the number of revolutions per minute (RPM) of the fan.

External factors, such as sunlight, ambient temperature, and inflowing water temperature, can have a strong influence on the system operation and affect the results. Therefore, in our testbed setup, we monitored these external factors and included them in our experiments. All experiments corresponding to the same failure mode were performed under the following conditions: (1) sunlight 3000–4500 lx; (2) ambient temperature 27–30 °C; (3) water temperature 29–32 °C. Failures were compared with reference data A^{∂} obtained in days whose sunlight level, ambient temperature, and water temperature fell into the above-mentioned ranges.

5.3.1 Data sorting

The Kruskal-Wallis test was used to evaluate the deviation of signals to derive the failure indicator. The different datasets corresponding to the reference data were merged into a single dataset A^{∂}. In this manner, a dataset per failure mode was generated. $A^{O}_{F_1}$ is the dataset corresponding to the tank leak, $A^{O}_{F_2}$ is the dataset corresponding to the obstruction in the irrigation pipes, and $A^{O}_{F_3}$ is the dataset corresponding to the irregular fan operation. Each dataset $A^{O}_{F_r}$ was randomly divided into a training set Γ_{F_r} and a test set T_{F_r}.

5.3.2 Deriving reference failure indicator matrix

A total of 500 **FI** matrices were generated per F_r to derive its corresponding $\mathbf{FI}^r$ in the process described in Section 3.3.5. Each **FI** matrix was derived from matching one randomly selected subsample $\Gamma_{F_{ij}}$ and one randomly selected subsample A^{∂}_{k}. This process was conducted for each signal feature considering the derivative, mean, standard deviation, RMS, and area. The obtained matrices $\mathbf{FI}^r$ ($r = 1, 2, 3$), $\mathbf{FI}^1$ for tank leak, $\mathbf{FI}^2$ for obstruction in the irrigation pipes, and $\mathbf{FI}^3$ for irregular fan operation, are shown in Fig. 8.

Theoretically, our testbed can have 64 system operation modes (SOMs). However, only 10 were actually activated during our experiment. The activated SOMs are presented in Table 3. The SOMs that did not occur are represented through white cells in the resultant matrices $\mathbf{FI}^r$. In our experiment, we found that some operation modes were triggered by the emerging failures, pushing the system into an "abnormal" operation mode (i.e., a combination of component operation modes not typical under regular circumstances). These will be called the "failure induced operation mode" (FIOM).

Table 3 Occurring system's operation modes

SOM	S_{A_1}	S_{A_2}	S_{A_3}	S_{A_4}	S_{A_5}	S_{A_6}
9	Off	Off	Off	On	Off	Off
11	Off	On	Off	On	Off	Off
12	Off	On	On	On	Off	Off
13	Off	Off	On	On	Off	Off
15	Off	On	On	On	Off	Off
33	Off	Off	Off	Off	Off	On
41	Off	Off	Off	On	Off	On
43	Off	On	Off	On	Off	On
45	Off	Off	On	On	Off	On
47	Off	On	On	On	Off	On

5.3.3 Exploring the distinctive power of failure indicators

We made a pairwise comparison of the $\mathbf{FI}^r$ matrices for each feature to determine their dissimilarity. This comparison allows us to evaluate if the $\mathbf{FI}^r$ corresponding to a failure mode is a subset of another failure indicator or could be a cause of misclassification. Therefore, we can obtain

$$D = \frac{C}{(C + \mathrm{NC}) - W} \times 100. \tag{22}$$

The elements of Eq. (22) correspond to the same elements of MD. These were presented in Section 3.3.6. W cells were not considered in the estimation of the level of dissimilarity, as the corresponding system operation modes did not occur during our experiment. Table 4 shows the dissimilarity analysis of failure indicators for our testbed. Failure indicators had a dissimilarity in the range of 27%–44.76%, where a maximum of 44.76% was found in the case of F_2 vs. F_1 with signal feature mean. Area and slope are signal features that present the lowest levels of dissimilarity, and therefore it is expected that these signal features are more likely to present misclassifications during the failure diagnosis process.

It can be concluded that the slope and area features have weaker distinctive power than the rest of the signal features analyzed. In contrast, RMS and mean are the features that have better distinctive power. Dissimilarities between F_1 and F_3 in area and F_2 and F_3 in RMS are graphically presented in Fig. 9. F_1 and F_3 selected in area present the lowest dissimilarity level, and F_2 and F_3 in RMS present a high level of dissimilarity. Black cells represent the signal cells that do not coincide between the pairwise-based analysis of failure indicators.

From this study, it can be inferred that a wide range of signal features should be used to define failure indicators to guarantee reliable failure diagnosis. Using dissimilarity analysis of failure indicators, signal attributes with high distinctive power can be identified and used in failure classification. Such a selective approach can further improve the reliability of failure diagnosis.

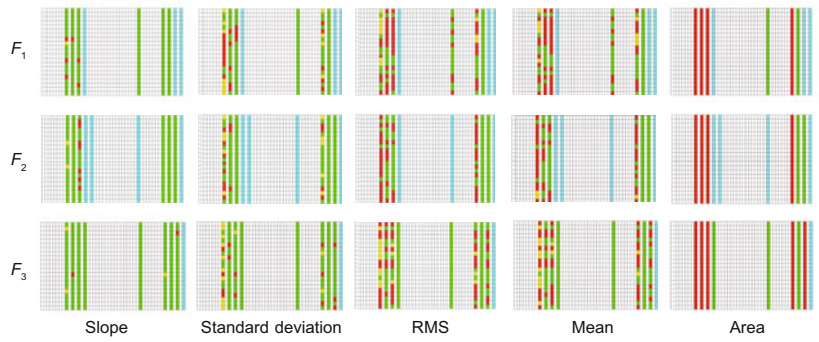

Fig. 8 Obtained failure indicators

Cells corresponding to the FIOM are highlighted in light blue. References to color refer to the online version of this figure

Table 4 Dissimilarity level between failure indicators

Failure mode	Slope (%)			Standard deviation (%)			RMS (%)			Mean (%)			Area (%)		
	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
F_1	0	38	27.51	0	41	35.4	0	39.68	42.32	0	44.76	43.38	0	30	22.22
F_2	38	0	36.5	41	0	42.3	39.68	0	41.26	44.76	0	44.76	30	0	40
F_3	27.51	36.5	0	35.4	42.3	0	42.32	41.26	0	43.38	44.76	0	22.22	40	0

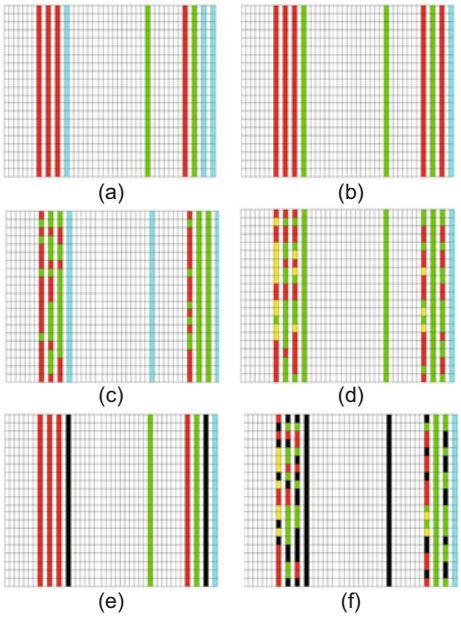

Fig. 9 Failure indicators overlapped for comparison: (a) area feature for F_1; (b) area feature for F_3; (c) RMS feature for F_2; (d) RMS feature for F_3; (e) area feature for F_1 and F_3; (f) RMS feature for F_2 and F_3

5.3.4 Sensitivity to signal features

Another aspect of failure differentiation is the dissimilarity of $\mathbf{FI}^r$ between signal features. It is desirable to have high dissimilarity between the $\mathbf{FI}^r$ of signal features to reduce the redundancy of failure indicators. Although analyzing several signal features improves the reliability of failure diagnosis, it comes at the cost of computing. Therefore, it is important to evaluate to what extent each signal feature conveys similar and dissimilar fault symptoms of the analyzed failure modes. The results are presented in Table 5. Each table cell presents the dissimilarity level corresponding to two different features. They are grouped by failure modes as each type of failure is represented by a failure indicator matrix $\mathbf{FI}^r$ per signal feature.

The RMS and mean present the lowest levels of dissimilarity, as low as 1.06% in the case of F_1 and a maximum of 10.95% for F_2. This implies that the inclusion of both failure indicators does not provide any added value to the failure diagnosis process, as they are rather redundant. On the other hand, area presents the highest level of dissimilarity to the other signal features. However, this feature, along with slope, presents the lowest levels of distinctive power between different failure modes.

Table 5 Dissimilarity level (%) between failure indicators using signal feature as the evaluation criterion

Failure mode 1

Feature	Slope	SD	RMS	Mean	Area
Slope	0.00	18.52	28.57	29.63	40.21
SD	18.52	0.00	20.63	20.63	24.87
RMS	28.57	20.63	0.00	1.06	19.05
Mean	29.63	20.63	1.06	0.00	17.99
Area	40.21	24.87	19.05	17.99	0.00

Failure mode 2

Feature	Slope	SD	RMS	Mean	Area
Slope	0.00	15.24	31.90	22.86	35.71
SD	15.24	0.00	22.38	11.43	28.10
RMS	31.90	22.38	0.00	10.95	28.57
Mean	22.86	11.43	10.95	0.00	17.62
Area	35.71	28.10	28.57	17.62	0.00

Failure mode 3

Feature	Slope	SD	RMS	Mean	Area
Slope	0.00	13.76	32.80	31.22	52.91
SD	13.76	0.00	26.46	24.87	41.27
RMS	32.80	26.46	0.00	4.76	21.16
Mean	31.22	24.87	4.76	0.00	22.75
Area	52.91	41.27	21.16	22.75	0.00

5.3.5 Effectiveness of the matching equation

Twenty **FI** matrices were generated for each signal feature for each failure mode. Data for these matrices were randomly taken from the test set T. The obtained matrices **FI** were matched with $\mathbf{FI}^1$, $\mathbf{FI}^2$, and $\mathbf{FI}^3$. We averaged the results concerning the similarity level obtained for the 20 considered cases per failure mode. To evaluate the effect of FIOM (blue cells), we performed the matching with and without the inclusion of failure-induced operation modes. In Fig. 10, red bars represent the averaged results of the analysis of the similarity level without FIOM, while green bars represent the averaged results with FIOM. They were placed together to ease visualization of the difference between them. Each of the plots was divided into three main groups organized as follows: the first three bars (of the same color) corresponded to the match of the test cases in which F_1 was injected; the second three bars corresponded to the comparison of the test cases in which F_2 was injected; the last three bars corresponded to the match of the cases in which F_3 was injected. In each of these groups, the first bar of each color was the average similarity of the analyzed test cases with $\mathbf{FI}^1$, the second bar was the average similarity of the test cases corresponding to a group with $\mathbf{FI}^2$, and

the third bar was the average similarity of the test cases belonging to this group with $\mathbf{FI}^3$.

We analyzed the standard deviation of the results of the matching degree for both cases, with and without FIOM (Fig. 11). Fig. 11 shows that without FIOM, results for F_1 and F_2 are satisfactory with correct matches, higher than 72%. However, it is not the case for F_3, as it presented a higher match with $\mathbf{FI}^2$ than with $\mathbf{FI}^3$ for the slope and SD features. Since the average dissimilarity was 33% between F_2 and F_3, there is a chance of misclassification as the rest of the fault symptoms were the same. Then, we can infer that the effect of F_3 on the slope and SD features is not very consistent; thus, symptoms change when the failure occurs. Nevertheless, it is not the case for RMS, in which the mean and area features presented a high degree of matching with the failure indicator matrix $\mathbf{FI}^3$.

In the case for failure diagnosis with FIOM, significant improvements were found as shown by the MD boxplots. This improvement was clearly observable even for F_3, when analyzed through SD which was initially misclassified by failure diagnosis without FIOM. Although cases injected with F_3 were still misclassified by the slope feature, the inclusion of FIOM reduced the difference between the match conducted with $\mathbf{FI}^3$ and the one corresponding to $\mathbf{FI}^2$.

The frequency analysis of correct matches (i.e., the percentage of data samples that were properly

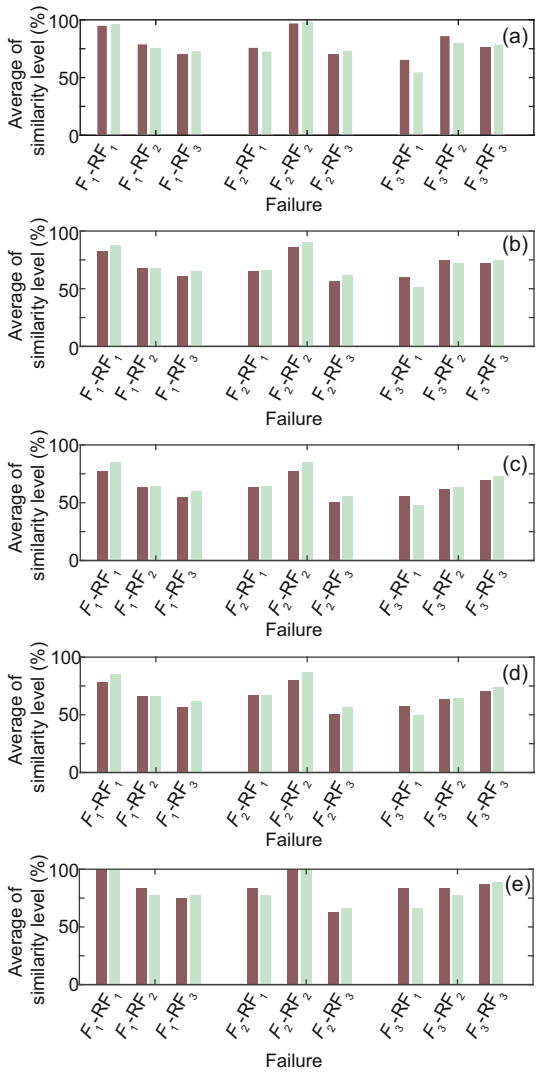

Fig. 10 Averaged similarity level for slope (a), standard deviation (b), RMS (c), mean (d), and area (e)

References to color refer to the online version of this figure

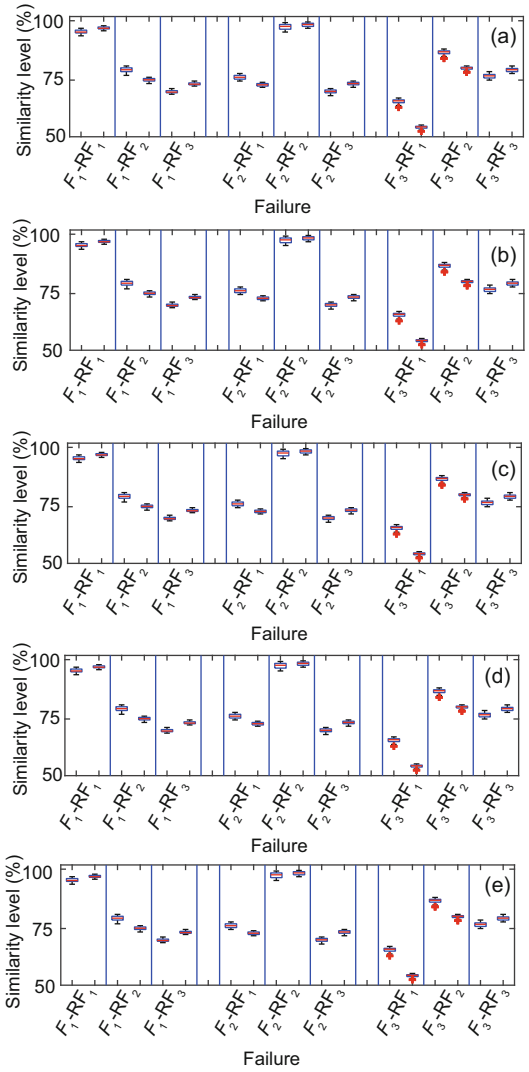

Fig. 11 Boxplots of the similarity levels for slope (a), standard deviation (b), RMS (c), mean (d), and area (e)

matched) corresponding to the analyses with FIOM is presented in Table 6. It can be seen that 100% of the analyzed samples were successfully classified except for F_3, where only 5% of the analyzed samples were properly matched when analyzed with the slope feature.

Table 6 Frequency of correct matches with FIOM

Failure mode	Frequency of correct matches (%)				
	Slope	SD	RMS	Mean	Area
F_1	100	100	100	100	100
F_2	100	100	100	100	100
F_3	5	100	100	100	100

From these results, it can be concluded that FIOM provides valuable information about each of the analyzed failure modes; thus, it should be considered during the diagnosis process. It has a positive effect on the matching process by significantly reducing the number of cases of misclassification (Fig. 10). The analysis of the standard deviation in Fig. 11 shows that it is possible to validate the matching process as its standard deviation is low (the highest standard deviation obtained is 0.96).

6 Contrasting the proposed method with data-driven failure analysis methods

Using the same datasets as used in the previous experiment, we investigated the accuracy of failure classification of some existing failure diagnosis methods. In our analysis, we aimed to compare our method with the methods reported in the literature. Considering that this research addressed failure diagnosis in the first- and second-generation CPSs, we discarded the model-based techniques and focused on data-driven techniques.

To derive classification models, we generated a predictor vector $\boldsymbol{P}$,

$$\boldsymbol{P}=\left[a_1^{S_{S_8}}, a_2^{S_{S_8}}, \ldots, a_5^{S_{S_8}}, \ldots, a_1^{S_{S_{18}}}, a_2^{S_{S_8}}, \ldots, a_5^{S_{S_{18}}}\right], \tag{23}$$

where a_1 is the derivative feature, a_2 standard deviation, a_3 RMS, a_4 mean, and a_5 area, as the failure indicator feature derived from sensor signals without segmentation. In this analysis, we evaluated only the following signals: water level in tank (S_{S_8}), water temperature (S_{S_9}), RPM fan-in ($S_{S_{13}}$), RPM fan-out ($S_{S_{14}}$), soil humidity of plat bed 2 ($S_{S_{17}}$), and soil temperature in plant bed 2 ($S_{S_{18}}$).

These signals were considered most likely to present failure symptoms due to their strong relation to the induced failures. A class vector $\boldsymbol{C}$ composed of the injected failure modes and the failure-free system state was generated, and $\boldsymbol{C}=[F_1, F_2, F_3]$.

Failure diagnosis was conducted by classification methods of the decision tree, support vector machines, k-nearest neighbors, boosted trees, and discriminant analysis implemented in the classification learner of Matlab 2017b. A summary of the implemented techniques is presented in Table 7. We used k-fold cross-validation (5-folds) to avoid overfitting and enable generalization.

The results presented a high level of accuracy for most of the analyzed methods. The accuracy of the classification of different methods ranged from 34.5% to 98.10% for true positives. The highest level of accuracy was obtained by the cubic SVM classifier, where 99% samples corresponding to the failure-free behavior were correctly classified, 97% data samples from F_1 were successfully diagnosed, 98% data samples from F_2 presented satisfactory results, and all the data samples corresponding to F_3 were correctly classified. The confusion matrix corresponding to cubic SVM is presented in Fig. 12.

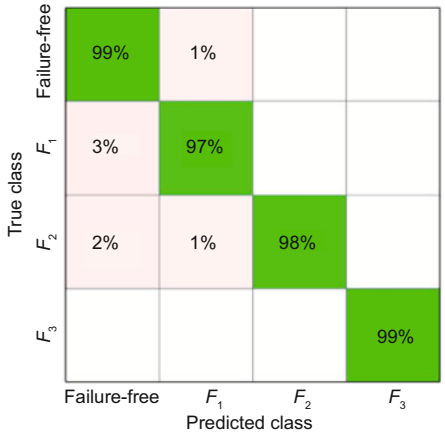

Fig. 12 Confusion matrix corresponding to cubic SVM

The results demonstrated that the current standard techniques can properly classify the studied failure modes. However, the interpretability of the obtained results is still an issue. These methods deliver the predicted class and the classification model. Nevertheless, these techniques do not provide explanatory information about failure manifestations (such as the magnitude of the symptoms, the signals

Table 7 Data-driven methods implemented

Method	Variation	Specification	Overall accuracy (%)
Decision tree	Complex tree	Number of splits: 100 Split criterion: Gini's diversity index	94.50
	Medium tree	Number of splits: 20 Split criterion: Gini's diversity index	94.50
	Simple tree	Number of splits: 4 Split criterion: Gini's diversity index	90.70
SVM	Linear	Multi-class method: one vs. one	97.40
	Quadratic		**98.10**
	Cubic		98.30
	Fine Gaussian		57.40
	Medium Gaussian		97.40
	Coarse Gaussian		95.90
KNN	Fine KNN	Number of neighbors: 1 Distance metric: Euclidean	94.30
	Medium KNN	Number of neighbors: 10 Distance metric: Euclidean	94.30
	Coarse KNN	Number of neighbors: 100 Distance metric: Euclidean	77
	Cosine KNN	Number of neighbors: 10 Distance metric: cosine	94.30
	Cubic KNN	Number of neighbors: 100 Distance metric: Minkowski	93.30
Ensemble	Boosted tree	Ensemble method: AdaBoost Number of splits: 20	34.40
	Bagged tree	Ensemble method: bag	97.80
	Subspace discriminant	Ensemble method: subspace Subspace dimension: 15	94.30
Discriminant analysis	Linear	Covariance structure: diagonal	87.30
	Quadratic		34.50

indicating that component failures are manifested, and the approximate time of occurrence). The confusion matrix that represents the results of classification does not provide sufficient information to determine the root cause of the failures and the progress of system degradation. This is a crucial limitation of data-driven failure classification techniques that can affect a further implementation of failure forecasting.

In contrast to the above-mentioned results, our failure diagnosis concept provides information about the possible root cause of failures. A combination of the signal sources and system operation modes offers a ground for reasoning about the potential source of failures at the component level.

7 Discussions

The proposed signal segmentation based failure diagnosis presented satisfactory results during the execution of the case study. Our experiment showed that the proposed concept of failure indicator is useful for not only diagnosing failure modes but also understanding failure manifestation. We found that there is a unique pattern for each failure mode that is reliably reproducible each time for this type of failure occurring in the system. This concept meets two challenges: automated failure diagnosis based on the reasoning of multiple signals and understanding failure manifestations.

The results showed that in all evaluated cases, the proposed approach can diagnose the injected failures. It presented accuracy results equivalent to those presented by the tested set of data-driven techniques. Nevertheless, the combination of all system signals, along with the system operation modes, does not just provide the means for conducting diagnosis, but helps understand which particular combination of signals and SOM failures are manifested. From this perspective, the proposed method is superior to the existing methods, since they lack support for

interpretability of root cause, thus preventing the analysis of factors, such as location, magnitude, and transient/steady manifestation of symptoms. This additional information about failure manifestation can lay the ground for the development of failure forecasting methods for cyber-physical systems.

Although the scalability of the proposed method for dealing with emerging functions and unknown failures was not tested in this study, our definition of system operation modes as a combination of component operation modes enables the implementation of an extendible failure indicator matrix, which can be used to represent and handle emerging system failures. Failure indicator matrices are extendible by adding/removing signals or SOMs as new rows or columns. The failure indicator needs to be updated every time the system configuration changes. This is another aspect of difference from the tested data-driven techniques, as the incorporation of new system functions or the emergence of new failure modes in traditional methods would require re-training of classifiers, thus raising the stability-plasticity dilemma. One of the major findings of our research is that the failure-induced system operation modes FIOM significantly improve the distinctive power of failure indicators. These SOMs occur only as a consequence of a particular failure mode of the system. They were initially not only considered in our failure diagnosis approach, but they significantly improved the results of the failure classification process in our experiments. It was found that the frequency of the occurrence of certain SOM is altered because of the failure mode, and thus this fault symptom should be considered as a meaningful information source for failure diagnosis and forecasting. Nevertheless, the proposed failure indicator concept is sensitive to the definition of the matching equation. Any variation on the equation considerably influences the results. Although the last issue affected failure diagnosis, it did not disqualify the concept of our failure indicator. There was a consistent pattern of symptoms and a lack of symptoms for each of the analyzed failure modes.

We evaluated how the consistency of failure indicators is influenced by the alterations in failure manifestation. Tank leakage and obstruction in the irrigation pipes failure modes were found to be more consistent than failure of the fan in terms of their fault symptoms. We found that the failure indicators generated by the signal features of slope and standard deviation present less consistent pattern of symptoms. This can be explained by the unreliable signal of the RPM sensor of the fan, causing disparity in the obtained data. Nevertheless, we found that the failure indicators of the other signal features of F_3 are consistent and they compensate for the inconsistency of the slope and SD features. Multiple signal features not only allow the overcoming of the inconsistency of failure indicators, but also automate the diagnosis and learning process, because not all failure modes manifest in all signal features. Traditional classification methods used in failure diagnosis require an expert determine which signal feature(s) to consider for identifying a specific failure. Our approach offers a generic solution which does not require a specialist specify a particular signal feature. It can affect the processing time of the method, as the consideration of the multiple signal features involves running the diagnosis process several times, one per signal feature.

The proposed method was successfully applied to failure diagnosis in an experimental setup. Our experiment demonstrated how the failure indicators can help understand the failure manifestations and how they can provide the means to automate failure diagnosis. They can be potentially used to identify emergent system functions and learn failure indicators corresponding to unknown failures. As the proposed method is a data-driven approach with the potential capability to learn new failure modes, it can cope with emergent system behaviors and the occurrence of unknown failure modes. Although the last capability was not evaluated in this study, Section 3 presented a concept of this learning approach. If the fault symptoms represented by the red cells and the failure-induced operation modes gradually occur as the failure manifestation progresses, the analysis of historical data of the failure indicators can provide a sufficient basis for the forecasting methods. We expect that the matching equation applied to failure diagnosis can be used in failure forecasting to predict the type of forming failure. This claim will be evaluated in our future research.

8 Conclusions and future work

In this paper, we have proposed a new failure diagnosis method based on the concept of failure

indicators. The presented method concurrently analyzes multiple signals to identify statistical deviations of signal features that are characterizing changes of the system behavior caused by the occurrence of failures. These deviations have been defined as fault symptoms. A novel element of this method is the signal segmentation based on the system operation modes. This concept has been introduced to improve the consistency of generic signal features and the reliability of failure diagnosis. The failure indicator is a matrix whose rows are the signal features and columns are the system operation modes. Failure diagnosis has been operationalized by identifying unique patterns of failure indicators that are a combination of symptoms and the lack of symptoms. We have aimed to evaluate the diagnostic capability of our failure indicator concept and explore its potential application in failure forecasting for our future research.

A case study has been used to evaluate the benefits and limitations of the proposed methods for failure diagnosis. We have built a testbed of a cyber-physical greenhouse to evaluate failure diagnostic capabilities of the proposed method in a realistic environment. The results showed that the proposed failure diagnosis method is effective for diagnosing all the induced failures. It demonstrated similar accuracy and prediction to the existing classification methods. In terms of interpretability of failure manifestations, it has been found to be superior. The results were satisfactory and the application of failure indicators can be extended to failure forecasting.

We have found that the introduction of system operation modes for signal segmentation and signal feature characterization has a positive effect on the outcome of failure diagnosis. Our study showed that the parallel use of multiple signal features for characterizing the system operations provides reliable failure diagnosis. At the same time, we have found that the failure indicators of some signal features have a strong similarity, thereby representing redundant information on the failure manifestation process. This redundancy can either be interpreted as an unnecessary computational load for the failure diagnosis process, or be used to increase the confidence and reliability of failure diagnosis. Our experiment has explored whether failure-induced system operation modes improve the outcome of failure diagnosis as they represent additional fault symptoms and improve the distinctive power of failure indicators by up to 15% in the best case.

Considering that the proposed concept is a data-driven method, it can be used to derive the failure indicators corresponding to the unknown and unpredictable failure modes. It not only can be trained for learning new failure modes, but also enables failure diagnosis in systems with self-tuning capabilities by handling and interpreting the entrance and exit of new signals and operation modes. However, strong similarities of the failure indicators may be a source of misclassification. This can be overcome by a larger number of signal sources and operation modes for deriving the failure indicator matrix. It is operationalized through the inclusion of new rows and columns in the failure indicator matrix as the new signal sources and operation modes enter and update the failure indicator.

References

Abdolsamadi A, Wang PF, Tamilselvan P, 2015. A generic fusion platform of failure diagnostics for resilient engineering system design. Proc ASME Int Design Engineering Technical Conf and Computers and Information in Engineering Conf, p.1-10. https://doi.org/10.1115/DETC2015-47009

Abe S, 2010. Multiclass support vector machines. In: Abe S (Ed.), Support Vector Machines for Pattern Classification. Springer, London. https://doi.org/10.1007/978-1-84996-098-4_3

Albertos P, Mareels I, 2010. Signal analysis. In: Albertos P, Mareels I (Eds.), Feedback and Control for Everyone. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03446-6_4

Alencar MS, da Rocha VCJr, 2005. Signal analysis. In: Alencar MS, da Rocha VCJr (Eds.), Communication Systems. Springer, Boston. https://doi.org/10.1007/0-387-27097-3_1

Alzghoul A, Backe B, Löfstrand M, et al., 2014. Comparing a knowledge-based and a data-driven method in querying data streams for system fault detection: a hydraulic drive system application. *Comput Ind*, 65(8):1126-1135. https://doi.org/10.1016/j.compind.2014.06.003

Bellido I, Fernández G, 1991. Backpropagation growing networks: towards local minima elimination. Int Workshop on Artificial Neural Networks, p.130-135. https://doi.org/10.1007/BFb0035887

Berk RA, 2008. Support vector machines. In: Berk RA (Ed.), Statistical Learning from a Regression Perspective. Springer, New York. https://doi.org/10.1007/978-0-387-77501-2_7

Bocaniala CD, Palade V, 2006. Computational intelligence methodologies in fault diagnosis: review and state of the art. In: Palade V, Jain L, Bocaniala CD (Eds.), Computational Intelligence in Fault Diagnosis. Springer, London. https://doi.org/10.1007/978-1-84628-631-5_1

Cholewa W, Korbicz J, Kościelny JM, et al., 2010. Diagnostic methods. In: Korbicz J, Kościelny JM (Eds.), Modeling, Diagnostics and Process Control: Implementation in the DiaSter System. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-16653-2_5

Daum FE, 2015. Kalman filters. In: Baillieul J, Samad T (Eds.), Encyclopedia of Systems and Control. Springer, London. https://doi.org/10.1007/978-1-4471-5058-9_61

Ding SX, 2008a. Introduction. In: Ding SX (Ed.), Model-Based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-76304-8_1

Ding SX, 2008b. Norm based residual evaluation and threshold computation. In: Ding SX (Ed.), Model-Based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-76304-8_9

Fernando H, Surgenor B, 2017. An unsupervised artificial neural network versus a rule-based approach for fault detection and identification in an automated assembly machine. *Rob Comput Integr Manuf*, 43:79-88. https://doi.org/10.1016/j.rcim.2015.11.006

Fortuna L, Graziani S, Rizzo A, et al., 2007. Fault detection, sensor validation and diagnosis. In: Fortuna L, Graziani S, Rizzo A, et al. (Eds.), Soft Sensors for Monitoring and Control of Industrial Processes. Springer, London. https://doi.org/10.1007/978-1-84628-480-9_9

Fujimaki R, Yairi T, Machida K, 2005. Adaptive limit-checking for spacecraft using relevance vector autoregressive model. Proc 8th Int Symp on Artificial Intelligence, Robotics and Automation in Space, p.1-7.

Gao ZW, Cecati C, Ding SX, 2015. A survey of fault diagnosis and fault-tolerant techniques—part I: fault diagnosis with model-based and signal-based approaches. *IEEE Trans Ind Electron*, 62(6):3757-3767. https://doi.org/10.1109/TIE.2015.2417501

Gertler J, Singer D, 1990. A new structural framework for parity equation-based failure detection and isolation. *Automatica*, 26(2):381-388. https://doi.org/10.1016/0005-1098(90)90133-3

Ghanbari T, 2015. Kalman filter based incipient fault detection method for underground cables. *IET Gener Transm Distrib*, 9(14):1988-1997. https://doi.org/10.1049/iet-gtd.2015.0040

Hang J, Zhang JZ, Cheng M, 2016. Application of multi-class fuzzy support vector machine classifier for fault diagnosis of wind turbine. *Fuzzy Sets Syst*, 297:128-140. https://doi.org/10.1016/j.fss.2015.07.005

He P, Liu G, Tan C, et al., 2016. Nonlinear fault detection threshold optimization method for RAIM algorithm using a heuristic approach. *GPS Sol*, 20(4):863-875. https://doi.org/10.1007/s10291-015-0494-9

Hood CS, Ji C, 1997. Proactive network-fault detection [telecommunications]. *IEEE Trans Reliab*, 46(3):333-341. https://doi.org/10.1109/24.664004

Hwang W, Han K, Huh K, 2012. Fault detection and diagnosis of the electromechanical brake based on observer and parity space. *Int J Autom Technol*, 13(5):845-851. https://doi.org/10.1007/s12239-012-0085-5

Irita T, Namerikawa T, 2015. Decentralized fault detection of multiple cyber attacks in power network via Kalman filter. Proc European Control Conf, p.3180-3185. https://doi.org/10.1109/ECC.2015.7331023

Isermann R, 2006a. Fault diagnosis with classification methods. In: Isermann R (Ed.), Fault-Diagnosis Systems: an Introduction from Fault Detection to Fault Tolerance. Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-30368-5_16

Isermann R, 2006b. Supervision and fault management of processes—tasks and terminology. In: Isermann R (Ed.), Fault-Diagnosis Systems: an Introduction from Fault Detection to Fault Tolerance. Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-30368-5_2

Johnson DM, 1996. A review of fault management techniques used in safety-critical avionic systems. *Prog Aerosp Sci*, 32(5):415-431. https://doi.org/10.1016/0376-0421(96)82785-0

Kang M, Ramaswami GK, Hodkiewicz M, et al., 2016. A sequential k-nearest neighbor classification approach for data-driven fault diagnosis using distance- and density-based affinity measures. Proc 1st Int Conf on Data Mining and Big Data, p.253-261. https://doi.org/10.1007/978-3-319-40973-3_25

Kishore B, Satyanarayana MRS, Sujatha K, 2016. Efficient fault detection using support vector machine based hybrid expert system. *Int J Syst Assur Eng Manag*, 7(S1):34-40. https://doi.org/10.1007/s13198-014-0281-y

Krishnamachari B, Iyengar S, 2004. Distributed Bayesian algorithms for fault-tolerant event region detection in wireless sensor networks. *IEEE Trans Comput*, 53(3):241-250. https://doi.org/10.1109/TC.2004.1261832

Lee C, Lee D, Koo J, et al., 2009. Proactive fault detection schema for enterprise information system using statistical process control. *LNCS*, 5617:113-122. https://doi.org/10.1007/978-3-642-02556-3_13

Lei Y, 2017. Signal processing and feature extraction. In: Lei Y (Ed.), Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery. Elsevier, Amsterdam.

Leonhardt S, Ayoubi M, 1997. Methods of fault diagnosis. *Contr Eng Pract*, 5(5):683-692. https://doi.org/10.1016/S0967-0661(97)00050-6

Li WL, Monti A, Ponci F, 2014. Fault detection and classification in medium voltage DC shipboard power systems with wavelets and artificial neural networks. *IEEE Trans Instrum Meas*, 63(11):2651-2665. https://doi.org/10.1109/TIM.2014.2313035

Liu YM, Ye LB, Zheng PY, et al., 2010. Multiscale classification and its application to process monitoring. *J Zhejiang Univ-Sci C (Comput & Electron)*, 11(6):425-434. https://doi.org/10.1631/jzus.C0910430

Luh GC, Wu CY, Cheng WC, 2004. Artificial immune regulation (AIR) for model-based fault diagnosis. Proc 3rd Int Conf on Artificial Immune Systems, p.28-41. https://doi.org/10.1007/978-3-540-30220-9_3

Mathur A, Foody GM, 2008. Multiclass and binary SVM classification: implications for training and classification users. *IEEE Geosci Remot Sens Lett*, 5(2):241-245. https://doi.org/10.1109/LGRS.2008.915597

Mehranbod N, Soroush M, Panjapornpon C, 2005. A method of sensor fault detection and identification. *J Process Contr*, 15(3):321-339. https://doi.org/10.1016/j.jprocont.2004.06.009

Miclea L, Sanislav T, 2011. About dependability in cyber-physical systems. Proc 9th East-West Design & Test Symp, p.17-21. https://doi.org/10.1109/EWDTS.2011.6116428

Montaño JC, Bravo JC, Borrás MD, 2007. Joint time-frequency analysis of the electrical signal. In: Moreno-Muñoz A (Ed.), Power Quality: Mitigation Technologies in a Distributed Environment. Springer, London. https://doi.org/10.1007/978-1-84628-772-5_3

Muralidharan V, Sugumaran V, Indira V, 2014. Fault diagnosis of monoblock centrifugal pump using SVM. *Eng Sci Technol Int J*, 17(3):152-157. https://doi.org/10.1016/j.jestch.2014.04.005

Patan K, 2008. Decision making in fault detection. In: Patan K (Ed.), Artificial Neural Networks for the Modelling and Fault Diagnosis of Technical Processes. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-79872-9_7

Ponsard C, Massonet P, Rifaut A, et al., 2005. Early verification and validation of mission critical systems. *Electron Note Theor Comput Sci*, 133:237-254. https://doi.org/10.1016/j.entcs.2004.08.067

Puig V, Escobet T, Sarrate R, et al., 2015. Fault diagnosis and fault tolerant control in critical infrastructure systems. In: Kyriakides E, Polycarpou M (Eds.), Intelligent Monitoring, Control, and Security of Critical Infrastructure Systems. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-44160-2_10

Qi JT, Zhao XG, Jiang Z, et al., 2007. An adaptive threshold neural-network scheme for rotorcraft UAV sensor failure diagnosis. Proc 4th Int Symp on Advances in Neural Networks, p.589-596. https://doi.org/10.1007/978-3-540-72395-0_73

Ramoni M, Sebastiani P, 2001. Robust Bayes classifiers. *Artif Intell*, 125(1-2):209-226. https://doi.org/10.1016/S0004-3702(00)00085-0

Ramos PM, Martins RC, Rapuano S, et al., 2009. Frequency and time-frequency domain analysis tools in measurement. In: Pavese F, Forbes AB (Eds.), Data Modeling for Metrology and Testing in Measurement Science. Birkhäuser, Boston. https://doi.org/10.1007/978-0-8176-4804-6_6

Rezazadeh AS, Koofigar HR, Hosseinnia S, 2014. Robust leakage detection for electro hydraulic actuators using an adaptive nonlinear observer. *Int J Prec Eng Manuf*, 15(3):391-397. https://doi.org/10.1007/s12541-014-0349-2

Rudin K, Ducard GJJ, Siegwart RY, 2014. A sensor fault detection for aircraft using a single Kalman filter and hidden Markov models. Proc IEEE Conf on Control Applications, p.991-996. https://doi.org/10.1109/CCA.2014.6981464

Sharifi R, Langari R, 2013. Sensor fault diagnosis with a probabilistic decision process. *Mech Syst Signal Process*, 34(1-2):146-155. https://doi.org/10.1016/j.ymssp.2012.07.014

Shui AS, Chen WM, Zhang P, et al., 2009. Review of fault diagnosis in control systems. Proc Chinese Control and Decision Conf, p.5324-5329. https://doi.org/10.1109/CCDC.2009.5195065

Shukla A, Tiwari R, Kala R, 2010. Artificial neural networks. In: Shukla A, Tiwari R, Kala R (Eds.), Towards Hybrid and Adaptive Computing: a Perspective. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-14344-1_2

Sobhani-Tehrani E, Khorasani K, 2009a. Fault detection and diagnosis. In: Sobhani-Tehrani E, Khorasani K (Eds.), Fault Diagnosis of Nonlinear Systems Using a Hybrid Approach. Springer, Boston. https://doi.org/10.1007/978-0-387-92907-1_2

Sobhani-Tehrani E, Khorasani K, 2009b. Introduction. In: Sobhani-Tehrani E, Khorasani K (Eds.), Fault Diagnosis of Nonlinear Systems Using a Hybrid Approach. Springer, Boston. https://doi.org/10.1007/978-0-387-92907-1_1

Somani AK, Vaidya NH, 1997. Understanding fault tolerance and reliability. *Computer*, 30(4):45-50. https://doi.org/10.1109/MC.1997.585153

Soroush M, 1997. Nonlinear state-observer design with application to reactors. *Chem Eng Sci*, 52(3):387-404. https://doi.org/10.1016/S0009-2509(96)00391-0

Stockman M, El Ramli RS, Awad M, et al., 2012. An asymmetrical and quadratic support vector regression loss function for Beirut short term load forecast. Proc IEEE Int Conf on Systems, Man, and Cybernetics, p.651-656. https://doi.org/10.1109/ICSMC.2012.6377800

Sun B, Luh PB, Jia QS, et al., 2014. Building energy doctors: an SPC and Kalman filter-based method for system-level fault detection in HVAC systems. *IEEE Trans Autom Sci Eng*, 11(1):215-229. https://doi.org/10.1109/TASE.2012.2226155

Swetapadma A, Yadav A, 2016. Directional relaying using support vector machine for double circuit transmission lines including cross-country and inter-circuit faults. *Int J Electr Power Energy Syst*, 81:254-264. https://doi.org/10.1016/j.ijepes.2016.02.034

Tornil-Sin S, Ocampo-Martinez C, Puig V, et al., 2014. Robust fault diagnosis of nonlinear systems using interval constraint satisfaction and analytical redundancy relations. *IEEE Trans Syst Man Cybern Syst*, 44(1):18-29. https://doi.org/10.1109/TSMC.2013.2238924

Wang DW, Yu M, Low CB, et al., 2013. Health monitoring of engineering systems. In: Wang DW, Yu M, Low CB, et al. (Eds.), Model-Based Health Monitoring of Hybrid Systems. Springer, New York. https://doi.org/10.1007/978-1-4614-7369-5_1

Witczak M, 2014. Introduction. In: Witczak M (Ed.), Fault Diagnosis and Fault-Tolerant Control Strategies for Non-linear Systems: Analytical and Soft Computing Approaches. Springer, Cham. https://doi.org/10.1007/978-3-319-03014-2_1

Wu W, Liu M, Liu Q, et al., 2016. A quantum multi-agent based neural network model for failure prediction. *J Syst Sci Syst Eng*, 25(2):210-228. https://doi.org/10.1007/s11518-016-5308-2

Ye N, Zhao BJ, Salvendy G, 1993. Neural-networks-aided fault diagnosis in supervisory control of advanced manufacturing systems. *Int J Adv Manuf Technol*, 8(4):200-209. https://doi.org/10.1007/BF01748629

Yin G, Zhang YT, Li ZN, et al., 2014. Online fault diagnosis method based on incremental support vector data description and extreme learning machine with incremental output structure. *Neurocomputing*, 128:224-231. https://doi.org/10.1016/j.neucom.2013.01.061

Yodo N, Wang PF, 2015. Resilience analysis and allocation for complex systems using Bayesian network. Proc ASME Int Design Engineering Technical Conf and Computers and Information in Engineering Conf, p.1-10. https://doi.org/10.1115/DETC2015-46999

Zarei J, Tajeddini MA, Karimi HR, 2014. Vibration analysis for bearing fault detection and classification using an intelligent filter. *Mechatronics*, 24(2):151-157. https://doi.org/10.1016/j.mechatronics.2014.01.003

Zhang JH, Chen YN, Xiong J, et al., 2014. Sensor fault detection and estimation for heat exchanger using unscented Kalman filter. Proc 9th IEEE Conf on Industrial Electronics and Applications, p.540-545. https://doi.org/10.1109/ICIEA.2014.6931223

Zhang W, 2016. Fault diagnosis method based on artificial immune system. In: Zhang W (Ed.), Failure Characteristics Analysis and Fault Diagnosis for Liquid Rocket Engines. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-49254-3_7

Zhou ZJ, Hu CH, Xu DL, et al., 2011. Bayesian reasoning approach based recursive algorithm for online updating belief rule based expert system of pipeline leak detection. *Expert Syst Appl*, 38(4):3937-3943. https://doi.org/10.1016/j.eswa.2010.09.055

Zweigle O, Keil B, Wittlinger M, et al., 2013. Recognizing hardware faults on mobile robots using situation analysis techniques. In: Lee S, Cho H, Yoon KJ, et al. (Eds.), Intelligent Autonomous Systems 12. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-33926-4_37