Highlights

A Bibliometric Literature Review of Integrated Data and Model Based Diagnosis Approaches for the Industry 4.0

Luis Enciso-Salas, Gustavo Pérez-Zuñiga, Javier Sotomayor-Moriano, Elodie Chanthery, Edgar Hernando Sepúlveda-Oviedo, Audine Subias, Louise Travé-Massuyès, Rodrigo Garcia, Jose Aguilar

- The literature review of hybrid data based diagnosis (DBD) and model based diagnosis (MBD) systems in Industry 4.0 reveals key trends and promising future directions.
- Hybrid diagnostic systems facilitate online monitoring, improve explainability, and enable intelligent feature selection.
- The most widely used hybrid architecture for integrating MBD and DBD techniques is serial integration.
- Serial integration uses MBD for residual generation, and DBDs for fault isolation and detection.
- There is growing interest in integrating hybrid models with explainability and online learning.
- The most common applications are in the automotive and energy sectors, in predictive maintenance and smart monitoring.
- Current hybrid diagnostic systems lack real-time scalability and lack interpretability and reusability.

A Bibliometric Literature Review of Integrated Data and Model Based Diagnosis Approaches for the Industry 4.0

Luis Enciso-Salas[a], Gustavo Pérez-Zuñiga[a], Javier Sotomayor-Moriano[a], Elodie Chanthery[b], Edgar Hernando Sepúlveda-Oviedo[b], Audine Subias[b], Louise Travé-Massuyès[b], Rodrigo Garcia[c,f], Jose Aguilar[c,d,e]

[a]*Engineering Department, Pontificia Universidad Católica del Perú, PUCP, Lima, Perú*
[b]*LAAS-CNRS, Université de Toulouse, CNRS, INSA, Toulouse, France*
[c]*GICOMP, Universidad EAFIT, Medellin, Colombia*
[d]*CEMISID, Universidad de Los Andes, Merida, Venezuela*
[e]*IMDEA Networks Institute, Leganes, Madrid, Spain*
[f]*Engineering Department, Universidad de Córdoba, Mónteria, Colombia*

Abstract

The ubiquity of Cyber-Physical Systems and the Internet of Things has created a data-based environment that has allowed a significant number of applications linked to the "Industry 4.0" paradigm for the manufacturing industry. In this field, Artificial Intelligence is playing an important role for the development of automatic diagnostic tasks. This work carries out a review of the literature on the latest advances in Industry 4.0 around automatic diagnosis based on two specific objectives: (i) Determine the possible combinations between Data Based Diagnosis and Model Based Diagnosis, and (ii) Identify the hybridization opportunities and existing frameworks to integrate machine learning approaches to reuse solutions/algorithms for Industry 4.0 problems. The bibliometric review of the literature was guided by a simplified version of the PRISMA 2020 standard. Some of the main results of the review are that most of the combination schemes use observer-based approaches as the residuals generators. On the other hand, the largest number of machine learning applications for Industry 4.0 have been for smart manufacturing, focusing on failure management. As for the most used techniques, there are those of deep learning and ensemble methods.

Keywords: ~~Integrated model~~ Hybrid model based and data based diagnosis, Data Based Diagnosis, Model Based Diagnosis, Machine

1. Introduction

Based on the significant transformation in the way we produce goods due to the ubiquity of Cyber-Physical Systems (CPS) and Internet of Things (IoT), we can say that the next few years will see a huge increase in manufacturing using technologies of the so-called "Industry 4.0". Today, Industry 4.0 has been established globally as a paradigm for the digital transformation of the manufacturing industry [1], where CPS use computing, communications, and control technologies to achieve new capabilities in areas such as intelligent production, supervision, optimization, diagnosis, monitoring, predictive maintenance, data analysis, real-time applications, and others. On the other hand, with the great progress of IoT, the deployment of sensors has increased in the manufacturing industry to capture the largest amount of data. This new data environment has allowed Artificial Intelligence (AI) to play an important role in the Industry 4.0, where CPS uses AI approaches (including Machine Learning (ML) algorithms) for tasks as automatic diagnosis with the aim of increasing process safety.

In this paper, we present a state-of-the-art study in the field of ML hybrid algorithms for automatic diagnosis and its application in the Industry 4.0, in order to identify current trends and challenges, to extract lessons learned and ideas. Additionally, potential combinations of the Data Based Diagnosis (DBD) and Model Based Diagnosis (MBD) methods are explored to leverage the advantages of each approach while addressing their respective limitations. Note that we restricted MBD to continuous systems, referring to works by the ~~Faults~~ Fault Diagnosis and Isolation (FDI) community, so Discrete Event Systems or Hybrid Systems are not addressed.

1.1. Previous reviews

Silva et al. [2] conducted a systematic literature review in which they propose a conceptual framework that highlights the capabilities and attributes that systems should have to meet common requirements in the Industry 4.0 environment. These include not only self-awareness and self-optimization for continuous improvement, but also the resilience and reliability expected for industrial applications. Such characteristics ensure that the system is capable of handling equipment failures or other unexpected disturbances by quickly

restoring its normal operation state. It is expected that with additional autonomy, the system's intelligent entities should be able to collaboratively resume normal operation by dynamically rescheduling and implementing new plans in a coordinated manner. Tidriri et al. [3] provide a comprehensive overview of model-based and data-based approaches for fault diagnosis, incorporating key concepts proposed by both diagnostic communities. They conduct a thorough review of existing studies that aim to integrate methods from both perspectives, with the ultimate goal of developing more efficient diagnostic systems.

In their study, Wilhelm et al. [4] thoroughly analyze the integration strategies of data-based methods with knowledge-based and physics-based models. They also investigate real-world industry examples to illustrate these strategies. The authors classify these integration approaches based on the different structures of the data flow, as illustrated on Figure 3, and extend their work to analyze the system's behavior. Notably, the authors highlight that several combination strategies for hybrid fault diagnosis approaches remain largely unexplored, particularly in the manufacturing domain, making them highly intriguing for ongoing research endeavors. Tiddi et al. [5] present an overview of the use of knowledge graphs in the context of explainable ML. They define an analytical framework to explore the current landscape of explainable ML. They analyze the integration of the framework with structured knowledge at large scale to analyse a variety of ML domains.

For pure ML, Mohindru et al. [6] review some ML methods used for analyzing data, under popular frameworks. Nguyen et al. [7] present a survey with comparisons and trends in the development and use of AI frameworks, software, libraries. In addition, the survey presents an overview of massive parallelism support. Cioffi et al. [8] present a systematic analysis of the literature relating the application of AI and ML in industry. The biggest number of works were published by USA and China. From the perspective of sustainability, Cioffi et al. [8] found that the new paradigm of smart manufacturing has the potential to improve productivity; and its main applications are the reduction of energy consumption and monitoring of environmental footprint.

A study on recent trends on hybrid modeling for Industry 4.0 [9] shows that both industry and researchers incorporate phenomenological knowledge (e.g., mechanistic modeling approaches) in ML frameworks, leading to more robust and intelligible solutions, capable of assisting stakeholders in decision processes.

Zhong et al. [10] present a similar review of the literature focusing on

intelligent manufacturing, IoT-enabled manufacturing, and cloud manufacturing. They argue that future research should be on a generic framework for intelligent manufacturing and data-driven intelligent manufacturing models.

As we can see, previous literature review works talk about the autonomic capabilities and attributes that software systems should have to meet Industry 4.0 requirements, or about model- and data-driven approaches for fault diagnosis, or about trends in the development and use of frameworks, software and libraries in industry, or about smart factory enabled by IoT and cloud. Thus, no literature review has been done on the advances of automatic FDI in the context of Industry 4.0, emphasizing model, data or hybrid approaches, specifically analyzing the possible combinations between MBD and DBD in the context of Industry 4.0.

1.2. Scope and contributions

The main objective of this work is to conduct a review of the literature on the latest advances in Industry 4.0 around automatic FDI, emphasizing those integrating model and data approaches. To this end, this paper has two specific objectives:

- Determine the possible combinations between MBD and DBD for continuous systems FDI,
- Identify the hybridization opportunities and existing frameworks to integrate ML approaches to reuse solutions/algorithms for Industry 4.0 problems.

The reasons for considering these specific objectives are because MBD is based on representing the behavior of the system using a mathematical model and DBD methods can use classification, clustering or reinforcement algorithms to learn a model of system behavior. In both cases, the model (according to its richness) can be used to predict, detect, identify, and locate failures in the system, so combining both approaches to exploit their advantages, using the large amount of data that the IoT is making available would seem the most natural. For this, it is necessary to be able to study their possible combinations and how ML techniques could be used in this context.

The remainder of this work is organized as follows. The bibliometric literature review methodology used in this work is described in Section 2. The classification and analysis of the literature found are reported in Section

3. ~~And then~~ Then, in Section 4, a visual representation of the interrelations in the literature approaches complements the analysis. Section 5 highlights the possibilities offered by hybridization. Section 6 concludes the document.

2. Bibliometric methodology

This section describes the methodology ~~used~~ adopted for the bibliometric literature review~~analysis. The presented~~ . The systematic review was conducted ~~after a simplification of the standard~~ through a simplified and domain-specific adaptation of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 ~~[11], which has been used in~~ guidelines [11], as applied in previous works such as ~~[12, 13, 14]. This methodology has four steps: (1) identification of the need, (2) definition of a search protocol to constitute the corpus of bibliographic references to be analyzed, (3) review on the corpus, and (4) analysis of the review. Two specific objectiveswere defined: (i)~~ [14, 15, 16]. The aim of this methodology is to ensure transparency, traceability, and reproducibility of the literature selection and analysis process within the scope of engineering research.

The review focused on two specific objectives:

(i) Determine the possible combinations between DBD and MBD, and ~~(ii) Identify the~~

(ii) Identify hybridization opportunities and ~~existent~~ existing frameworks to integrate ML approaches ~~to reuse solutions /algorithms for~~ for reusing solutions or algorithms in Industry 4.0 ~~problems. Based on the two objectives,~~ contexts.

A total of 1327 records were identified through database searches (1316 from Scopus, Web of Science, and Google Scholar, and 11 through backward and forward citation tracking). After duplicate removal and preliminary filtering, 105 studies proceeded to screening. Based on title and abstract evaluation, 3 records were excluded for irrelevance. The full texts of 103 studies were retrieved for further analysis.

Subsequently, 37 full-text articles were excluded for the following reasons: (i) the data-driven component did not involve learning or training (22 studies), (ii) no actual hybridization of physical and data-driven techniques (10 studies), and (iii) the study fell outside Industry 4.0 contexts (5 studies). Finally, 75 ~~articles were obtained, 48 articles about the possible combinations between~~

~~Data Based and Model Based Diagnosis, and 27 articles on the existing frameworks in~~ articles were included in the final synthesis.

~~This section presents the details of the search protocol (step (2))~~

Although inspired by PRISMA, this methodology is structured in three main phases—Identification, Screening, and Inclusion—as detailed in the following subsections. First, the research questions are ~~defined. Next~~ presented. Then, the search strategy is ~~presented. Finally, a summary of~~ explained, and finally, the selection process is ~~provided~~ described in detail.

2.1. ~~Definition of the search questions~~Identification

The ~~general aim is to identify relevant studies according to research questions, which in our case are~~objective of the identification phase was to retrieve relevant literature addressing hybrid diagnostic systems that combine MBD and DBD approaches within Industry 4.0 contexts. To this end, the following research question was formulated:

~~Q1:~~

- **Q1:** Which are the potential combinations between MBD and DBD in the context of Industry 4.0? (datasets/benchmarks and validation strategies)

2.2. ~~Search strategy~~

2.1.1. ~~Search and Query strings~~

~~The search terms were defined using~~ To construct the search strategy, the PICOC method ~~[17] by which Table 1 is obtained~~was employed [17]. Search terms ~~are related to~~ were derived to reflect the diagnostic nature of the ~~diagnosis subject using a combination of model- and data-based approaches in the context of industry 4.0, and later, a search string was proposed for the search question. The question was structured with the logical operators~~ studies, their hybrid structure, and their relevance to Industry 4.0. These were combined using Boolean operators ("~~and~~AND"~~and~~ , "~~or~~OR"~~to define the query stringsthat were used for the search in databases such as Scopus, Web of Science and Google Scholar, in the period starting from 2010 to the present andin English language. Table 2shows the query string, it should be noticed that C1 term was not used for the query string but instead considered in the exclusion criteria~~) to formulate query strings, as shown in Tables 1 and 2.

Table 1: Search terms

ID	Term
	FDI systems
P1	Diagnosis or Detection or Isolation
	Hybrid systems
I1	Hybrid or Combination or Fusion
I2	Enhance or Intelligent or Learn
	Hybridization opportunities
O1	Model or Residual or Parity or Structural or Observer or Parameter
O2	Data or Neural or Artificial Intelligence or Tree or SVM or PCA or ML

Table 2: Query string

Question ID	Question	Query string
1	Combinations between MBD and DBD in the context of Industry 4.0	(P1) and (I1 or I2) and (O1 and O2)

The searches were conducted in three major scientific databases: Scopus, Web of Science, and Google Scholar. Only documents written in English and published from 2010 onward were considered. The review focused exclusively on peer-reviewed literature available in bibliographic databases; no additional registers or gray literature sources were used.

2.2. Screening

After the identification of relevant records, a multi-stage screening process was applied to filter the studies based on their relevance and compliance with predefined criteria.

2.2.1. Screening criteria

~~In the~~ During the initial screening phase, ~~articles meeting one or more of these criteria were excluded~~ studies were excluded if they did not satisfy the following conditions:

- ~~Which area of work is not in~~ The work did not belong to the engineering field.
- ~~Which does not present an FDI~~ The paper did not present a Fault Detection and Isolation (FDI) approach.

2.2.2. Inclusion and exclusion criteria

The inclusion criteria for the ~~selection of publications were~~the ~~following~~full-text review were:

- Articles and book chapters published from 2010 ~~in English language, related to the~~ onward, written in English.
- Publications directly addressing the formulated research question.
- ~~Articles and book chapters are available in electronic form~~Availability of the full text in electronic format.

~~However, the following exclusion criteria were considered~~Exclusion criteria were defined as follows:

- ~~Whose data-based~~ **Reason 1:** Not in engineering or not related to fault diagnosis.
- **Reason 2:** Excluded based on title and abstract: insufficient relevance to hybrid fault diagnosis approaches.
- **Reason 3:** Data-driven component does not ~~employ~~ involve a learning or training ~~phase~~process.
- ~~Which core approaches are not a combination of physical modeling and data-based techniques,~~ **Reason 4:** Does not combine model-based and data-driven approaches.
- ~~Which are not related to the context of~~ **Reason 5:** Not related to Industry 4.0 applications.

2.3. ~~Selection process~~

~~The search process followed four~~

2.2.1. *Selection process*

The screening was carried out in several stages:

1. ~~Database search and duplicate removal: selection by~~ *Database search and duplicate removal:* Articles were first selected by assessing their title and keywords ~~, evaluating if the title and keywords of the paper are close to the search strings, eliminating the repeated ones.~~ for alignment with the query terms. Duplicate entries were removed at this stage.

2. ~~Screening: based on~~ *Title and abstract screening:* Articles were evaluated based on their abstracts to determine whether they met the screening criteria~~applied to the title and abstract of the papers~~.

3. ~~Retrieval: All the works after the screening phase and related to the question are retrieved for~~ *Full-text retrieval:* Studies deemed relevant were retrieved in full. If the full text ~~review, and those unable to find the full text are marked as~~ could not be accessed, the work was marked as "not retrieved."

4. ~~Full text screening: the eligibility of each work is assessed after reviewing the full text and considering the exclusion criteria~~ *Full-text eligibility assessment:* Each full text was reviewed in detail, and exclusion criteria were applied to determine its eligibility.

5. ~~Identification via other methods: inclusion of additional papersbased on the references to the previously selected documents; exclusion criteriaare also employed.~~

   Reference list check: Additional relevant works were identified by analyzing the references of included papers. These were also screened using the same criteria.

~~The procedure is shown in Figure 1, mentioning at each stage the number of selected papers. At the end of~~

2.3. *Inclusion*

After completing the identification and screening stages, the ~~search process, 48~~ final inclusion process was carried out to define the corpus of references used for the review. This phase involved the integration of all eligible studies that met the criteria after full-text assessment.

A total of 75 articles were selected ~~to be included in the review~~for inclusion in the systematic review. These articles were classified according to the two research objectives established at the beginning of the study:

- 48 articles focus on identifying possible combinations between DBD and MBD approaches in the context of Industry 4.0.
- 27 articles address existing frameworks and hybridization opportunities to integrate ML techniques for the reuse of solutions and algorithms within Industry 4.0 problems.

The selection process is visually summarized in Figure 1, which outlines the number of articles retained at each stage, from initial identification to final inclusion.

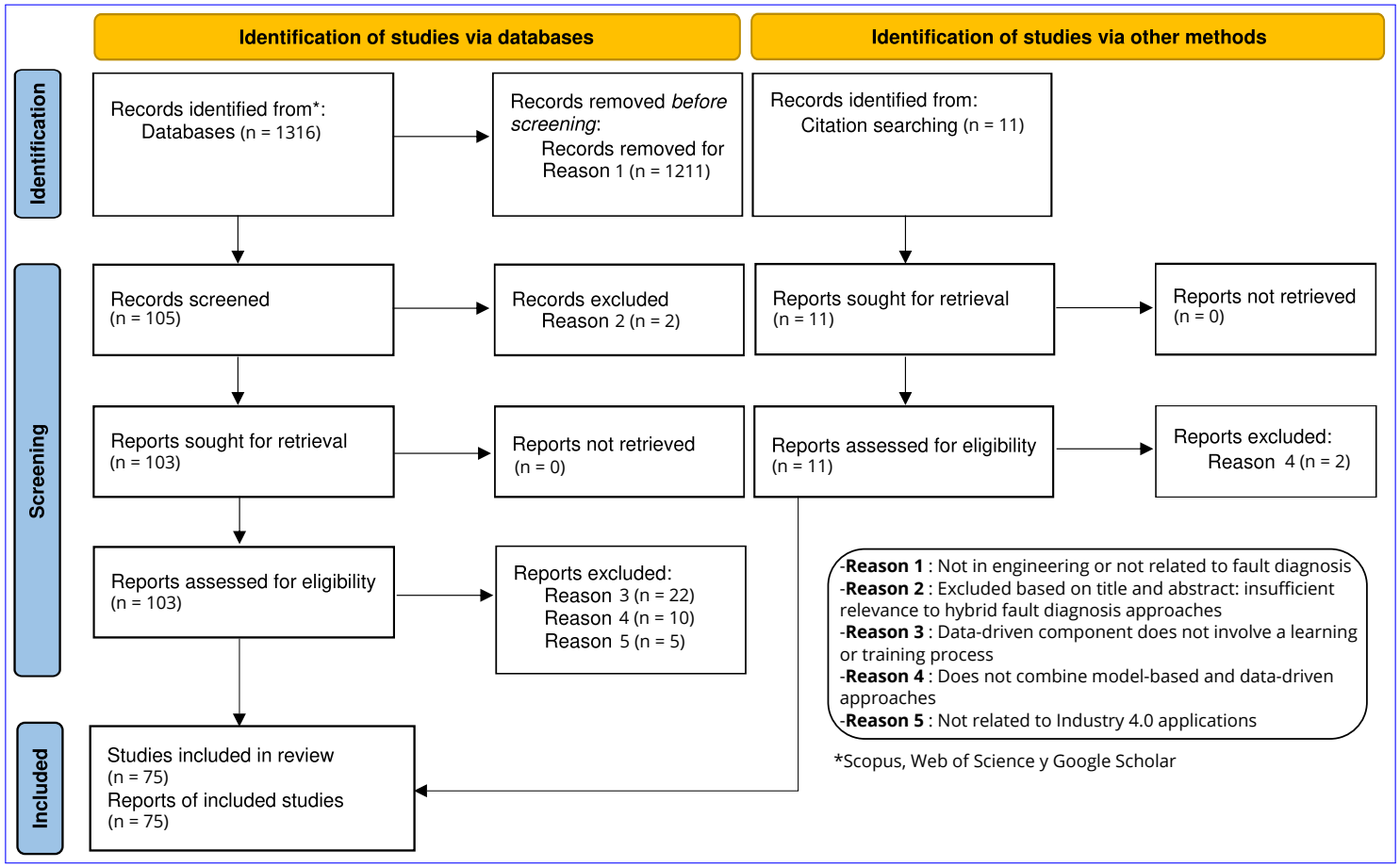


Figure 1: The methodology followed in the bibliography search.

3. Classification and analysis of retrieved papers

~~This section classifies the papers that constitute the corpus of references obtained using the search methodology presented in Section 2~~

This section presents the methodology used to classify the 75 selected studies and provides an analysis of ~~the contents.~~ their contributions. The classification process follows a structured and reproducible approach to group the studies based on the type of hybrid architecture used to integrate MBD and DBD techniques for fault diagnosis in Industry 4.0 systems.

3.1. ~~Classification of the retrieved papers~~

3.1. Conceptual foundations of MDB and DBD approaches

~~In the scope of this review, Industry 4.0 related works are considered as applications with potential interconnectivity in industry-related areas such as manufacturing, energy, electric, chemical and other special activities.~~ MBD and DBD ~~approaches have been historically different paradigms in diagnosis design, albeit with complementary advantages that have ignited combinations in many works in the literature. For~~ represent historically distinct paradigms in diagnostic system design. MBD ~~, the use of~~ approaches use physics-based models ~~enables them to elaborate residualswith sensibility to certain fault conditions without having to see them before-hand, meanwhile,~~ to generate analytical redundancy (residuals), enabling generalization to unseen fault types. DBD ~~can use historic data of past fault conditions in the system to obtain features and train models or classifiers with high accuracy for detecting and identifying failures~~methods, in contrast, leverage historical data and machine learning to build accurate fault classifiers through supervised or unsupervised training.

Figure 2 illustrates a ~~classification of different~~ conceptual taxonomy of these methods and ~~approaches. The middle part of the figure shows the different diagnostic tasks, from "system identification" to "fault isolation (residual interpretation)". The bottom part of the figure focuses on the methods. They can be organized in the following categories: , , , and methods. One or more methods can contribute to several diagnosis tasks, and the same diagnostic task can be solved by different methods. All these methods are part of a more general "systems modeling" method, applied to continuous systems. The figure highlights the fact that methods can also be applied to discrete-event and hybrid systems, which are not considered in this article.~~

~~The upper part of Figure 2 focuses on the methods. Meanwhile, categorization of data-based methods can be a more complicated task due to the large number of methods developed for this end, different classifications of algorithms have been presented, as in Dall et al. [18] or Gao et. al [19] for classification, and Deng et al. [20] for deep learning classification. As for the ones, we propose to classify methods according to different categories : , , , and . Again, one or more methods can contribute to several diagnosis tasks, and the same diagnostic task can be solved by different methods. At a higher level, these methods can participate in two main tasksdefined as "dimension reduction" and "dimension mapping", applied only to time series. The figure also highlights that methods can also be applied on other types of input data (event logs, text or images) that are not taken into account in this article.~~

~~In the lower part of Figure 2, a scheme representing the different and approaches related to the diagnosis steps is presented, considering the main focus of the present article is in applications for time-series continuous systems that use physical modeling. For the model-based methods, they can be organized in the following categories:~~ their roles in various diagnostic tasks. MBD methods are classified into Parameter Estimation (Parameter estimation (PE)), Structural Analysis (Structural Analysis (SA)), Observer-based (Observer-based (OB)), and Parity-Space (Parity-Space (PS)) ~~methods, furthermore a system can be composed by one or many of them. In the diagnosis scheme they are usually utilized in the residual generation, however, other uses such as system modeling, residual selection, fault detection and fault isolation are possible. For latter fault detection and isolation, is usual to employ an additional mechanism to perform such tasks i.e. threshold, fault signature matrix, ; which are included as statistical tools, which do not represent the core of the approaches.~~

~~In the context of diagnosis the~~ approaches. DBD ~~classification is presented in the upper part of Figure 2 furthermore focusing in time-series and its data representations and dimension manipulation subdivisions; thus categories of MDB by applications are~~ methods are grouped into Regression (Regression (REG)), Classification (Classification (CLA)), Clustering (Clustering (CLU)), and Statistical Tools (Statistical tools (STT))~~which can be utilized as well in any of the diagnosis tasks, as is indicated by the connections of the scheme.~~ .

3.2. ~~Analysis~~ Methodology for classification of ~~the retrieved papers classified by architecture~~studies

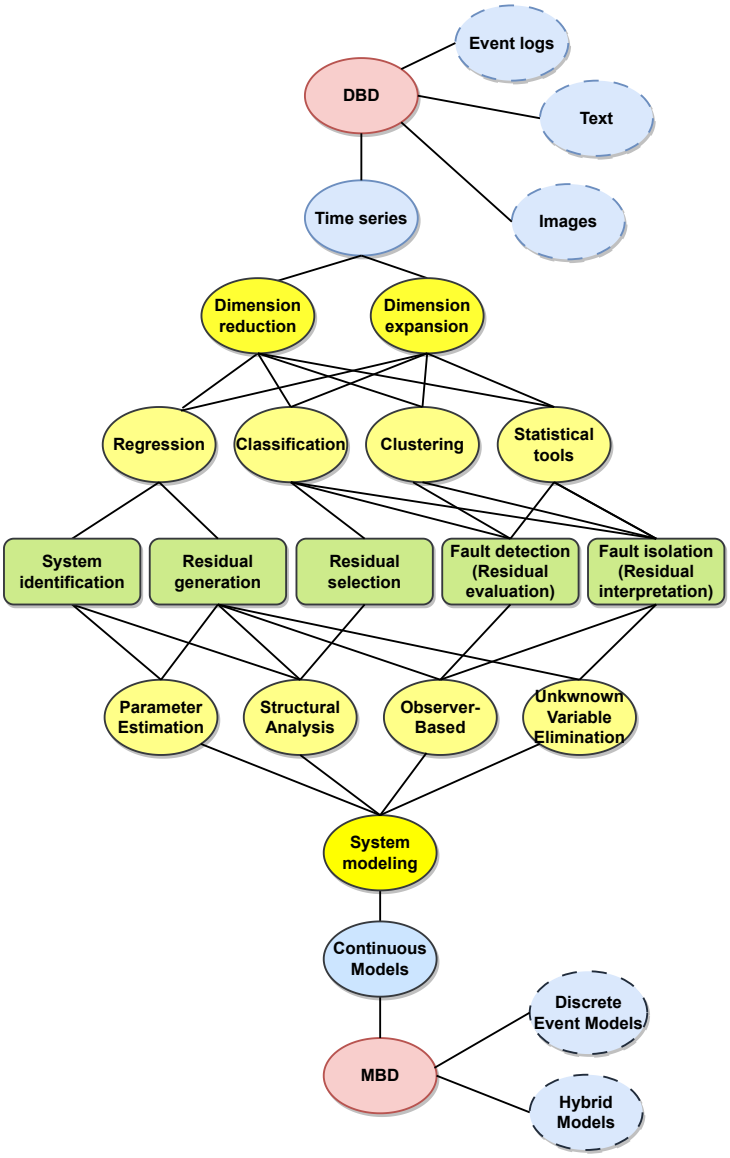


Figure 2: A classification of different MBD and DBD approaches and ~~its possibilities of use~~ their role in ~~the~~ diagnosis tasks.

~~Another way to classify the works is to consider the integration architecture that can be put in place~~The studies were classified according to the integration architecture implemented. Three primary categories were defined:

~~label=()~~

- **Serial** ~~**integration, as presented in Figure 3a, this architecture does the integration by using a combination of methods in a sequential manner. It is, for example, a usual methodology to integrate model-based**~~ **Integration:** MBD and ~~data-based methods to improve the final decision by replacing the traditional decision mechanisms used in~~ DBD methods are combined sequentially, typically using model-based ~~methods with a data-based method approach.~~ residuals as input features for data-driven classifiers.

- ~~**Parallel integration**, as presented in Figure 3b, and consists in running data-based methods in parallel with model-based methods, usually performing each partial diagnosis. Later, all partial results are fused to obtain the improved final decision.~~ **Parallel Integration:** MBD and DBD components operate independently and simultaneously, with results later fused to make a diagnostic decision.

- ~~**Ad-hoc integration**, as presented in Figure 3c, gathers the approaches that do not follow any of the structures given before. Usually, data-based approaches are incorporated into many or more specific tasks in order to improve them. For instance, they can be used to achieve better residual selection, or to carry out certain parts of the system modeling.~~ **Ad-hoc Integration:** Integration occurs in a customized, non-modular way; ML tools are embedded into specific stages such as residual evaluation or model adaptation.

A double-reviewer protocol was followed to ensure objectivity. Reviewers assessed the full text of each paper guided by questions such as:

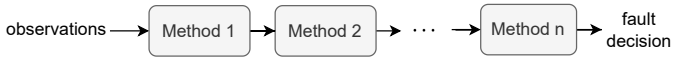
- How are MBD and DBD components combined?
- Is the system architecture modular, sequential, or customized?
- Which diagnostic tasks (e.g., detection, isolation) are addressed?

Disagreements were resolved through discussion. Table 3 presents a summary of the classification results, including the number of papers in each category, the most common techniques, and representative references. . Additionally, Appendix A contains a Table with a more detailed classification by article.

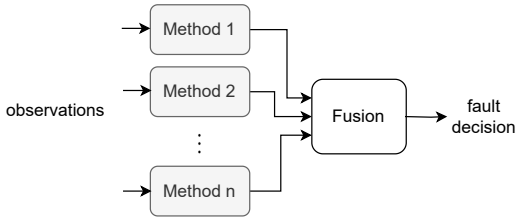
Table 3: Summary of classified studies by integration architecture

Category	# Papers	Common Techniques	Representative Works
Serial Integration	25	SVM, ANN, STT	[21, 22, 23]
Parallel Integration	11	SA, observers, ANN	[24, 25, 26]
Ad-hoc Integration	15	SA, SVM, ANN	[27, 28, 29]

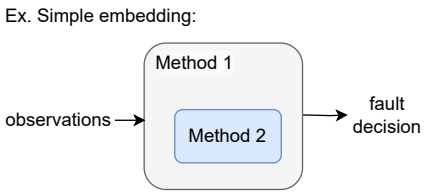
The three integration structures are illustrated in Figure 3.



(a) Serial integration.



(b) Parallel integration.



(c) Ad-hoc integration.

Figure 3: Main types of structures for combinations between MBD and DBD approaches.

3.3. Analysis of retrieved studies by integration type

Building on the methodology described in Subsection 3.2, the 75 studies were grouped into the three integration categories summarized in Table 3. Each category is analyzed in detail in the following subsections, discussing representative works, integration strategies, and diagnostic outcomes.

3.3.1. Serial integration

Kappaganthu and Nataraj [21] propose to use the features generated using both model-based and signal-based methods, and then ranked by using an index called mutual information between features and classes; after that, a selected subset of features is incrementally fed to an Artificial Neural Network (ANN) classifier to find the optimal isolation of the faults.

Benkoudier et al. [22] propose a hybrid approach for fault diagnosis in a chemical reactor. First, an Extended Kalman Filter (EKF) is used to estimate states and also the overall heat transfer coefficient; such coefficients, together with inputs and measurements, are then employed by a probabilistic ANN to elaborate on fault isolation in the system.

Costamagna et al. [23] use a parallel model running in synchronization with the plant conditions. Parity relations are used to obtain residuals that are later evaluated by an Support Vector Machine (SVM) classifier for fault identification. The approach was applied to a solid-oxide fuel cell.

Jung et al. present a series of works. First, in [30], the use of SVM is proposed to improve fault identification. The SVM is used as an algorithm to search for a set of residuals with the best diagnostic performance; the residuals are generated by the SA approach. Jung et al. in [31] also propose a fusion mechanism between residuals and SVM for fault location improvement; for this, a residual generator is proposed based on the process model. A test evaluates which residuals have been activated, and an order of priority among the activated residues is made according to a trained classifier. Meanwhile, Jung and Sundstrom [32] propose a hybrid mechanism that is used to select a set of residual generators through an optimization problem with a relaxed condition, for which L1-logistic regression is used to find the best suitable residuals, and the search by optimization is performed independently for each fault. Finally, in [33], Random Forest (RF) and SVM classifiers are employed and compared for fault isolation tasks for a subset of residual candidates found by the L1-logistic regression optimization.

Sun and Shi [34] employ a Kalman filter to generate estimations that are compared to the measurements to obtain residuals. A set of features is then calculated from the residuals mean. The features are complemented by taking vibration signals and decomposing them into components called intrinsic mode functions, which are generated using the extended empirical mode decomposition method, thus adding signal-based features to the mix. All these features are then used to isolate the faults by means of an ANN.

Khorasgani et al. [35] propose a unifying framework to integrate model-based and data-based approaches. The work describes a methodology that can be applied when combining such approaches in a sequence. It proposes a parallel between residual generation and feature extraction, and furthermore, the use of supervised or unsupervised classifiers for fault isolation. Although the given methodology can be applied to many approaches, the paper demonstrates it for the combination of the data-based Gaussian Kernel Support Vector and the model-based OB method. Both methods provide a list of features that are then used by an SVM to classify the fault.

Pérez-Pérez et al. [36] use a combination of Analytical Redundancy Relations (ARR) and ANN. ARR are employed to generate residuals and then ANN to perform the fault diagnosis. In order to select the most suitable model accuracy metric, a confusion matrix is used to rank the classifiers and choose the best one.

Won et al. [37] first use SA and knowledge information for system modeling, then a gray-model and a black box model are obtained for each output using linear regression and Three based (TB) regression, respectively. These are then used for residual generation. Model-based diagnosis can be obtained using thresholds. To improve accuracy and enable the detection of faults with lower magnitudes, a second diagnostic phase using a classifier is added. The article suggests using an ANN, although other alternatives have also been tried. This approach has been proposed for fault diagnosis in a membrane fuel cell.

Atoui and Cohen [38] work presents a technique that uses a Bayesian classifier with a matrix of fault signatures. In practice, residuals are generated based on structural methods. The system's state is analyzed through a Bayesian classifier, and the methodology is illustrated in a simulated plant.

Cho et al. [39] employ a Kalman filter to obtain estimations. Then residuals are generated by the difference with respect to real measurements. They are evaluated using threshold comparisons to detect the presence of faults. In cases where faults are detected, an ANN classifier is utilized for fault isolation using a window of the signals. The approach is evaluated on the hydraulic pitch system of a wind turbine.

Song et al. [40] utilize a combination of an extended state observer and Deep Forest, which is a TB algorithm, applied to diagnose a Unmanned Aerial Vehicle (UAV) system. First, the observer is employed to obtain residuals for roll and pitch signals by calculating the difference between the estimated and measured accelerations. Subsequently, a fusion of features is performed

by slicing them over time, thereby incorporating location information into the new features. These new features are then supplied to the Deep Forest algorithm that performs the fault isolation.

Van Esch [41] proposes to combine a model-based estimator of faults with a classifier for isolation. An augmented state observer is employed to estimate the fault; such estimation can be impacted under noise conditions and provides only bounded estimation error. Then, such fault estimations are used as fault signatures and utilized in the subsequent detection and isolation stages. This work tests two different algorithms, the SVM and K-nearest neighbors (KNN) in a robotics system.

Xia et. al. [42] propose an approach to utilize a Hyper-graph Convolutional Neural Network (HGCN) for fault diagnosis. At first, an hyper-graph is constructed based on SA, but only using under-determined structures. Thus, hyper-nodes are defined and the links relating the variables. As connection between some hyper-nodes can result ambiguous, then a hyper-matrix is defined to address the connection and direction of the links. From this, a HGCN is defined using the hyper-matrix obtained previously as an adjacent matrix. The HGCN structure is also composed hierarchically by three layers, which are from upper to lower, the product, part and component layers. Thus, lower layers report features to the immediate upper layers. The dataset is obtained from the input signals in the lower layer, which are pre-processed by normalization and signal processing using fast Fourier transform. Finally, the HGCN is trained to provide fault isolation through classification.

Asefi et. al. [43] propose an anomaly detection system in a power grid. For this, two procedures are employed; in first stage, a combination of Weighted Least Squares (WLS) and (EKF) are used to produce an anomaly detection index. This index is based on the difference between the results from the WLS and (EKF), then allowing to distinguish between normal operation an existence of anomaly. In the second stage, a ML method is used to classify two different sources of anomaly with better accuracy. Features given as inputs to the classifier include the measurements and the estimations. The tested ML methods were RF, Logistic Regression (LR) and KNN.

Jung et. al. [44] elaborate a methodology for using data-driven generated residuals in the fault diagnosis. For this, test validity region is defined as the region in which the training set allows distinguishing between nominal and fault operation. Although no definitive methodology is proposed, some insights into the design of data-driven residuals using structural information are given such as the consideration of integral or derivative causality, and its

impact in the test validity region.

Jiang et. al. [45] propose a fault tolerant control system based on a model predictive controller (MPC), and also a hybrid Fault Detection and Diagnosis (FDD) system to detect the occurrence of a fault. For this, a state error observer generated from the MPC controller is first employed to obtain residuals, and later, a Principal Component Analysis (PCA) algorithm is used to reduce the dimensionality of the dataset. Furthermore, a SVM is trained and later employed to classify the fault source.

El Mawas et. al. [46] propose the combination of several models, an odometric model to predict, a generalized information residual to produce residuals based on a Jensen Shannon divergence metric. Fault detection is performed using the Decision Tree (DT) algorithm, followed by RF to perform the fault diagnosis.

3.3.2. Parallel integration

Some of the works developed by using this architecture are dedicated to define a generic framework that can accommodate any model-based or data-based technique suitable for such a framework and the scope of functionality defined within it. Such articles are presented in [47, 48, 49]. In these cases, specific combinations are presented only for validating the approach or as exemplary cases within it.

Luo et al. [24] present an FDD algorithm for an anti-break system using parity relations and a non-linear observer to obtain residuals. Simultaneously, a statistical method is applied to isolate the most sensitive faults. In parallel, an SVM is used to isolate faults for two less sensitive faults. Additionally, a parameter estimation method is employed, using the output error, to assess the fault severity.

Narasinham et al. [50] employ an observer and analytical redundancy relations to perform diagnosis by examining fault signatures. Furthermore, they identify ambiguity groups (groups of faults with the same fault signatures). To address this, they propose a feature-based classifier, with the paper suggesting that such a classifier can be implemented using an ANN. However, in their work, a tree-based method is employed having as input a feature table for the ambiguity groups.

Ghimire et al. [51] uses an approach that combines the parity space method for generating residuals using signals from the system. Such a combination is then passed to a data reduction phase through techniques such as PCA. Fault detection is accomplished using the Cumulative Sum method

(CUSUM). Additionally, Fault isolation is achieved using data-driven techniques, which include SVM, KNN or ANN.

Lu et al. [52] propose a hierarchical fusion of model-based and data-based generated evidence using a weighted Dempster-Shaffer algorithm. In this approach, a model-based particle filter generates one piece of evidence and its corresponding fault diagnosis through the use of fuzzy logic. In parallel, a kernel Extreme Learning Machine (ELM) method generates the second piece of evidence as a vector of probabilities, which is considered as a fault isolation decision. Finally, the weighted Dempster-Shaffer algorithm combines both results to make the final diagnosis decision. This approach is developed for an application in the in-flight gas path of an airplane.

Slimani et al. [47] present a method that combines diagnosis based on both data and models. To achieve this, they propose a generic representation and framework to express the different diagnosis results obtained from model-based and data-based algorithms in a probabilistic form. Subsequently, they carry out a fusion process using a Bayesian strategy to derive a posterior probability, which then gives the final diagnostic decision.

Yang et al. [25] present a fault diagnosis method that comprises two parts. In the first part, an ARR method is used for residual generation, and a fault signature matrix is used for identification. In parallel, signals are reconstructed using Auto Associative Kernel Regression (AAKR); for this, an index called prediction interval is also monitored for robustness checking. Estimations generated with the AAKR are then compared with measurements and used for online fault detection, which is then combined with the ARR-based results.

Yang et al. [53] conduct diagnosis through two distinct methods. In the first method, they use structural analysis to design state observers and generate residuals, making decisions based on predefined thresholds. For the second method, features are generated using a wavelet transformation, and then a back-propagation ANN is used for diagnosis. This dual approach is applied to an electro-pneumatic brake system.

Imbassahy et al. [48] use a framework, known as the Generic Anomaly Detection Hybridization Algorithm framework, for combining various ML algorithms with model-based generated residuals. The fusion of results is done by an ensemble algorithm based on triads of ML methods and a voting system for the decision. Additionally, the residuals, obtained by comparing estimated and actual states, are included in the ensemble. The final decision can be performed using either a linear combination or an Artificial Neural

Fuzzy Inference System (ANFIS) classifier. The method is demonstrated on an airplane landing gear.

Huang et al. [49] propose a framework for diagnosis that uses in parallel a state observer and Kalman observers for residual generation. ELM is used in two distinct ways: one to improve the residual generation based on the state-observer, and the other to obtain a parallel residual that is fused with the model-based result. This solution is tested for the diagnosis of an aero-engine.

Yao et al. [26] perform fault detection for distributed energy resource inverters using a dual approach. For MBD, they use weighted least squares to elaborate initial residuals, which are subsequently post-processed using the largest normalized residual. An identification threshold is used to detect faults. In parallel, Autoencoder (AE)s are deployed in a decentralized setup to identify anomalous behavior in each inverter. This is achieved by reconstructing the residuals, with the hypothesis in the paper being that abnormal data will have larger reconstruction errors. The reconstructed residuals are also evaluated using threshold comparison for fault detection, and a centralized mechanism is used for the AE update. Finally, the results from both approaches are fused for more robustness and accuracy, with the final decision elaborated by an OR mechanism.

Liu et. al. [54] propose a fault diagnosis system for the braking system of a vehicle, which combines a model-based method using an observer in parallel with a data-based method using Gated Recurrent Network (GRU). The observer is developed for the hydraulic control unit whose behavior can be described by hydraulic equations, meanwhile, the GRU is used to estimate a data-based model for the wheel cylinder pressure which has strong non-linearity and coupling with the other units. The data-based model is composed by an encoder, followed by two GRU modules and a decoder, and receives pre-processed signals from diverse sensors, and a scheme for data augmentation is also proposed. Both observer and the data-based model generate estimations for the wheel cylinder pressure. These estimations and the measurements are used to generate residuals, and the detection is performed by the CUSUM algorithm, meanwhile isolation is performed by using a fault matrix.

3.3.3. Ad-hoc integration

Svard et al. [55] propose a learning approach for residual evaluation, while residual generation is performed using ARR approach. In this work,

residual evaluation involves the off-line training of a probabilistic distribution for no-fault scenarios. Learning these no-fault distributions is based on the application of KNN. Furthermore, test statistic hypotheses are designed using generalized likelihood ratio, which serve to generate the residual evaluators. Svard et al. [27] expand the residual evaluation methodology and propose a new algorithm to learn automatic distributions from no-fault data, employing Kullback-Liebler divergence optimization. One notable advantage of Svard's proposed solution is its ability to detect smaller faults compared with other classical evaluators.

Sheibat-Ohman et al. [56] use a combination of SVM and an observer. First, an SVM algorithm is used for fault detection. When no fault is detected, the model is corrected at each step. However, if a fault is detected, the correction process is stopped and a set of residuals is generated by the observer. The isolation is then performed using an SVM classifier.

Frisk and Krysander [57] propose a system to select the best residuals in a full set of candidates without drop of performance if possible. To perform this selection, a RF classifier is used to order the residuals by predictive power. The diagnosis is then carried out using lighter algorithms, such as consistency-based isolation, using a subsetof the ranked residuals.

Yang et al. [25] perform fault diagnosis by integrating a fault isolation mechanism using ARR with fault matrix for the decision. Additionally, they employ a method for online monitoring that addresses robust fault detection. For the latter, a form of kernel regression (auto-associative) is used to estimate the outputs based on historical data and measurements. Then, prediction intervals are calculated to get a measure of the system's uncertainty.

Saez et al. [58] establish a framework to model Cyber-Physical Manufacturing Systems using models based on both physical equations and data. The IoT was used to obtain a computer numerical control tool with hybrid data (discrete and continuous). For discrete treatment, the framework uses global operational states. It includes a context-sensitive modeling strategy to establish thresholds for fault detection from residuals. Furthermore, it incorporates diagnosis using two potential methods, SVM and a rule-based method derived from defined context-sensitive diagnosis rules based on magnitudes and patterns from key variables of the plant.

Jung [59] proposes the use of a structural approach combined with residual generation from Recurrent Neural Network (RNN). First, the structural approach is used to incorporate physical insights into the residuals design, using Minimal Structurally Overdetermined subsystem (MSO) sets. The ex-

pressions of the residuals are then found using RNN, trained using nominal data. The RNN-generated residuals can be later used for fault diagnosis. Additionally, in [60], Jung proposes a slight modification, this time using greybox RNN whose models are used for residual generation. The structural information is interpreted in bipartite graphs, which are then used for the RNN networks. Since training in the grey RNN is performed with nominal data, large values in residuals can be attributed to the presence of faults or to operations not covered in the training conditions. A related work by Mohammadi et al. [61] analyzes the impact of limited data and suggests a modification for improving robustness by including decoupled fault data in the training.

Wang et al. [62] propose an offline data-based Local Linear Model Trees (LOLIMOT) model that is used in parallel to the model-based parity-space approach. The former produces features, while the latter produces residuals. Both features and residuals are transformed into symptoms, which are then evaluated by a fault signature matrix to achieve fault diagnosis.

Chen et al. [28] propose to use Graph Convolutional Neural network (GCN) for diagnosis purposes. In the approach, SA is utilized first to obtain prediagnosis results, which are used to form an adjacency matrix (A) that together with the nodes (N), composed by measurements, form an association graph for the GCN. The GCN is then augmented with an ANN classifier, and trained to make the final diagnosis decision. Furthermore, Chen et al. [63] propose an improvement for the GCN-SA method to better handle the limited availability of labeled data. To achieve this, a parameter θ is incorporated to weight the structure of the association matrix A. It is optimized using a particle swarm algorithm.

Gienger et al. [29] employ a set of parity relations for generating residuals. As these residuals can be coupled with noise, a PCA is proposed for decoupling components and reducing the residuals dimension. Additionally, the Kullback-Leibler (KL) algorithm is used as a measure of the isolation performance, considering a specific number of principal components. Detection is then carried out using a cumulative sum of uncorrelated signals derived from PCA, and isolation is performed by using a Bayes classifier (a class of Bayessian Classifier (BN)) on the estimated probabilities.

Galve et al. [64] select a set of MSO sets using SA followed by a heuristic minimization. Then, ML models are trained to learn the relationship between a measured variable and those linked to it by the feasible MSO sets. The diagnosis is performed by first using a kernel density estimator for probability

distribution estimation and confidence level of deviations. Later, a Bayesian method is employed to isolate the fault. The paper shows the application of this approach in a water cooling system.

Enciso et al. [65] present an approach that uses a pre-trained model implemented by a Neural ODE in order to generate residuals. In this case, the model for the measurements is incorporated in the training, and an additional ANN is used for fault classification.

Oromi et al. [66] combine SA with ANFIS and a Bayesian classifier for fault diagnosis. The SA is used to obtain the structure of the residuals without considering a mathematical model. Then, the expressions of the residuals are obtained from historical data. These expressions are refined using an ANN. For fault isolation, a Bayesian classifier, equipped with a fault signature matrix obtained from the structural analysis, is employed. The approach is demonstrated on a brushless DC motor.

Pérez-Pérez et al. [67] use SA to establish several structural relations (SR) static SR, meaning not recursive, are estimated using equations, while dynamic SR are estimated using zonotopic Takagi-Sugeno system (TS) observers. To achieve this, ANFIS networks are employed to learn the TS relations, and robust LMI optimization is used to design the zonotopic observers. Finally, fault detection is carried out by analyzing the intersection between the measured and observer sets.

Sacchi et al. [68] propose a sliding mode observer for implementing the FDD system. The sliding mode observer is employed as an unknown input estimator for the fault, and detection is performed by comparing the estimation to a predefined threshold. In cases where only partial dynamics are known, the sliding mode observer is adapted by introducing an ANN to estimate the unknown drift term.

4. Visualization of the results

A preliminary look to the reviewed studies indicates that the selection of a specific hybrid approach has been primarily influenced by the availability of resources. The serial integration approach, where MBD is followed by DBD, was typically employed when a prior model was available, enabling comparison with data-driven classifiers. In contrast, parallel integration approaches could be adopted to validate diagnostic outcomes obtained from data, although also for complementing the initial diagnosis. Lastly, the ad-hoc integration approach was utilized in scenarios where it was necessary

to enhance or complete a model using DBD techniques, which was subsequently applied following the principles of the MBD approach.

Furthermore, to enhance the understanding of how different diagnostic approaches are related and combined within the reviewed studies, two visual representations were developed. These tools aim to systematically illustrate the integration patterns and highlight the methodological relationships among the employed techniques. A detailed explanation of each visual representation is provided in the following.

Figure 4 illustrates a Sankey diagram depicting the relationships between MBD and DBD approaches. ~~It complements~~ This figure complements the conceptual overview provided in Figure 2. ~~In~~ At the center of the ~~figure~~diagram, the different diagnostic tasks corresponding to the green rectangles in Figure 2, are presented. ~~On the leftside,~~ To the left, the diagram shows the different categories of model-based methods ~~included~~ identified in the articles ~~studied in the bibliometric literaturereview are depicted. On the rightside, various categories of data-based methodsare shown.~~ included in the reviewed literature. To the right, it displays the corresponding categories of DBD methods. The diagram thus provides a visual synthesis of how different methodological categories contribute to each diagnostic task.

The width of the arrows in the diagram is proportional to the flow quantity they represent: this means that the wider the line in the figure, the more articles there are on this topic in the studied articles. The figure reveals that the majority of model-based approaches pertain to observers (OB). The primary diagnostic task performed by model-based approaches is the residual generation. There is a scarcity of articles in which model-based approaches engage in fault identification or fault isolation. In this study, no model-based approach is involved in fault detection. Conversely, data-based approaches appear to focus on fault isolations. Some are also employed for residual generation and faut detection. Regarding the approaches, the primary data-based approach is classification (CLA), for fault isolation or (to a lesser extent) for fault detection. Regression (REG) on the other hand, is mainly used for residual generation.

Of all the approaches employed in the proposed hybrid approaches, the following diagnostic tasks were identified in the review:

a) Residual generation (mainly for MBD part)

b) Fault isolation (mainly for the DBD part)

c) Residual selection (DBD approaches)

d) Fault detection (DBD approaches)

e) System identification (both by MBD and DBD approaches)

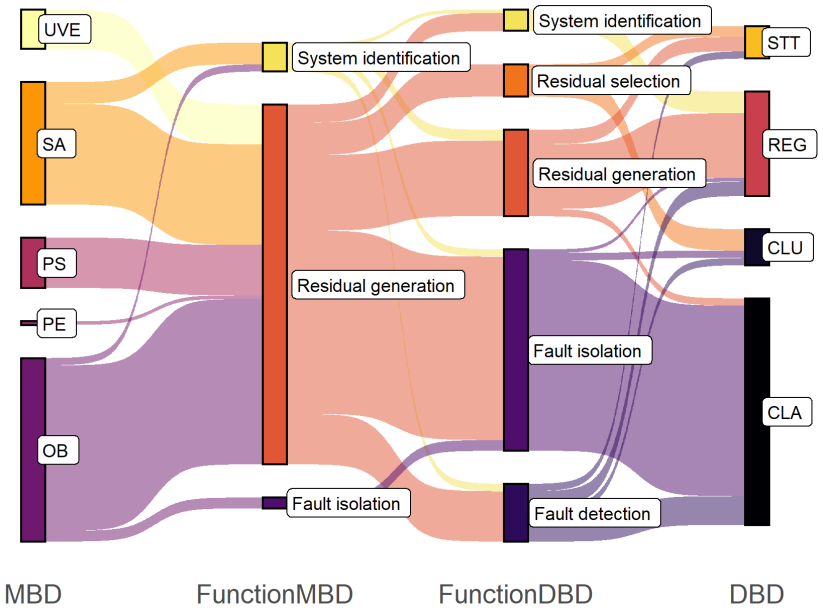


Figure 4: Roles of MBD and its relation with DBD methods. The connecting ribbons and their thicknesses show how often specific MBD and DBD approaches are used for a specific diagnostic task (specified in figure 2) and how are they linked or combined with its counterpart.

Figure 5 is a chord diagram. It represents the relationships between a group of entities, here, the pair of diagnostic approaches. Entities are typically arranged around the circumference of a circle, and chords connect pairs of entities to represent the relationships between them. The thickness of the chords is proportional to the number of articles where the two nodes are paired in a hybrid diagnostic approach. This form of visualization provides an intuitive means of identifying which combinations of diagnostic approaches are most commonly explored and which remain relatively rare, offering insights into prevailing trends and potential gaps in the research landscape.

Figure 5 illustrates that the most significant connections exist between observer-based approaches (OB) and classification (CLA), ~~as well as between~~ this strong linkage suggests a notable trend in the literature toward combining model-based residual generation, based on observers, with data-driven classification techniques for fault interpretation and decision-making. Similarly, Structural Analysis (SA) ~~and~~ is frequently integrated with both classification (CLA) ~~or~~ and regression (REG) methods; these combinations highlight a growing interest in leveraging the analytical rigor of structural modeling alongside the flexibility and learning capabilities of machine learning algorithms to enhance diagnostic performance. It is observed that clustering (CLU) is combined with various techniques ~~but without a distinct preference or a large number of articles.~~ yet without a strong or consistent pattern of co-occurrence. The corresponding chords in the diagram are relatively thin and dispersed, indicating that while clustering is present in the hybrid diagnostic literature, it is not a dominant component in any specific pairing. This may reflect the exploratory nature of clustering in fault diagnosis, where it is often used for unsupervised fault grouping or anomaly detection, rather than being a core component of well-established hybrid frameworks.

The hybrid diagnostic approaches that combine MBD and DBD techniques have started to be implemented in real industrial environments, particularly in sectors where sensor availability, connectivity, and computational power enable the deployment of advanced cyberphysical systems. For example, in smart manufacturing plants, MBD-DBD approaches have been applied to monitor the condition of rotating machinery by integrating state observers with classification models to detect and isolate faults such as imbalance, misalignment, or component wear. In energy systems, such as smart grids, physical models of load behavior have been combined with clustering and classification algorithms to detect anomalous patterns indicative of technical or non-technical losses. Similarly, in the aerospace industry, where systems are highly complex and safety-critical, hybrid schemes have been used for early fault detection in hydraulic or propulsion systems, leveraging residual generation from dynamic models and deep learning techniques to interpret these residuals and predict potential failures.

5. Opportunities for hybridization

The potential for hybridization hinges on two key aspects. Firstly, it depends on the presence of frameworks capable of streamlining development

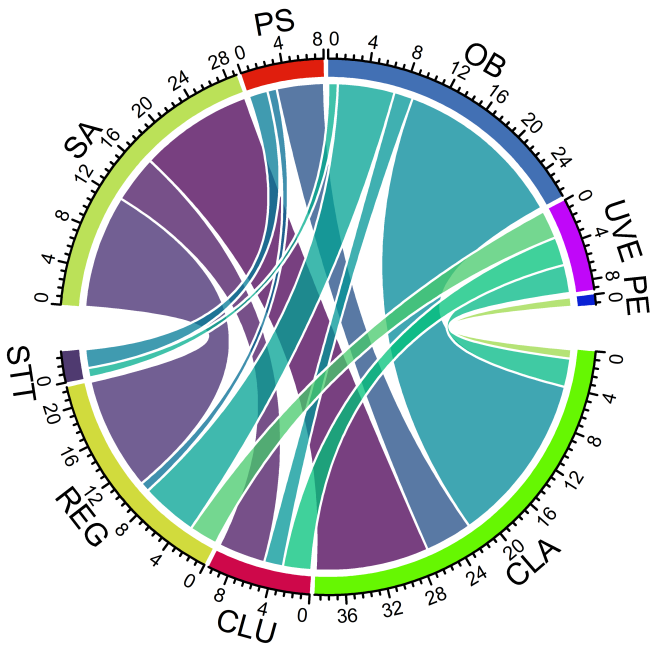


Figure 5: Related methods of the hybrid approaches. The connecting ribbons and their thicknesses show how often specific MBD approaches are linked or combined with particular DBD techniques, helping to highlight common pairings and methodological trends.

processes, improving code maintainability, and offering a cohesive structure for developers engaged in a project. Secondly, it is influenced by the conceptual flexibility of methods, which may either permit or restrict the integration of modules inherited from both sides of the hybrid approach. This section is organized in three subsections that present:

- Existing frameworks for model-based approaches and data-based approaches.
- Example applications in the context of industry 4.0. In particular, applications targeting manufacturing production, Prognostics and Health Management, and Industrial-IoT because they involve the integration of smart devices, sensors, software, and other technologies to collect, exchange, and analyze data in industrial environments, all of which can improve efficiency, productivity, and safety.
- Perspectives for integration based on our in-house tools Learning Algo-

rithm for Multivariable Data Analysis (LAMDA) and Dynamic Clustering algorithm for tracking Evolving Environments (DYCLEE)[1].

5.1. Frameworks

In this paper, a "framework" refers to a pre-built structure or set of tools, libraries, and conventions designed to facilitate the development of software applications. It provides a foundation that developers can build upon, offering a reusable and customizable structure for creating various types of software.In summary, frameworks provide a foundation of reusable components, pre-built modules, and established patterns that accelerate the development process and enable rapid prototyping.

5.1.1. Frameworks for model-based approaches

There are several popular frameworks and libraries used for model-based fault diagnosis. These tools are designed to develop, implement, and analyze fault diagnosis models. Classifying them by the programming language employed, we can mention ~~:~~

~~*Matlab:*~~

- ~~Symbolic COntrol of Timed Systems (SCOTS) [69]: is a toolbox for the analysis and control synthesis of discrete-time and continuous-time systems. It can be applied to model-based fault diagnosis problems.~~
- ~~Diagnosis and Monitoring in Systems and Processes (DMSP), is a toolbox that focuses on diagnosis and monitoring in systems and processes. It includes tools for building diagnostic models and performing fault diagnosis.~~
- ~~Fault Diagnosis Toolbox [70], is a toolbox for design of fault diagnosis systems for dynamic systems.~~
- ~~FDITOOLS [71], is a toolbox for the synthesis of model-based filters employing null-space methods. It can perform also tasks such as evaluating fault detection specifications, performing feasibility analysis, evaluating models via pairwise distances, evaluating fault detection sensitivity conditions, and other methods to evaluate the performance of the filters.~~

[1] https://gitlab.laas.fr/project-dyclee/dyclee

~~*Python:*~~

- ~~DIAGNOSYS [72], is a library for fault diagnosis and prognosis. It provides tools for building diagnostic models, analyzing data, and making predictions about the health of a system.~~

- ~~Scikit-FME [73], is an open-source Python library for fault modeling and diagnosis. It provides tools for constructing fault models and performing diagnosis based on these models.~~

- ~~PHAVer [74]: is a Python library for the analysis and verification of hybrid systems. It can be used for fault diagnosis in systems with continuous and discrete dynamics.~~

~~*Java:* FAME (Fault Adaptive Modeling Environment), is~~ a ~~Java-based framework for fault modeling and diagnosis. It provides a platform for developing and testing fault diagnosis algorithms.~~ variety of toolboxes and libraries, each tailored to specific modeling environments such as Matlab, Python, Java, Modelica, and Julia. A summary of these frameworks, along with their purposes and programming languages, is provided in Table 4.

~~*Modelica:*~~

- ~~OpenModelica [75]: is an open-source modeling and simulation environment. It can be used for modeling physical systems and performing fault diagnosis by simulating system behavior under different conditions.~~

Table 4: Model-based fault diagnosis frameworks by programming language

Language	Framework	Description
Matlab	SCOTS [69]	Toolbox for the analysis and control synthesis of discrete-time and continuous-time systems. It can be applied to model-based fault diagnosis problems.
	DMSP	Toolbox that focuses on diagnosis and monitoring in systems and processes. It includes tools for building diagnostic models and performing fault diagnosis.
	Fault Diagnosis Toolbox [70]	Toolbox for design of fault diagnosis systems for dynamic systems.
	FDITOOLS [71]	Toolbox for the synthesis of model-based FDI filters employing null-space methods. It can perform also tasks such as evaluating fault detection specifications, performing feasibility analysis, evaluating models via pairwise distances, evaluating fault detection sensitivity conditions, and other methods to evaluate the performance of the FDI filters.
Python	DIAGNOSYS [72]	Library for fault diagnosis and prognosis. It provides tools for building diagnostic models, analyzing data, and making predictions about the health of a system.
	Scikit-FME [73]	Open-source Python library for fault modeling and diagnosis. It provides tools for constructing fault models and performing diagnosis based on these models.
	PHAVer [74]	Python library for the analysis and verification of hybrid systems. It can be used for fault diagnosis in systems with continuous and discrete dynamics.
Java	FAME	Java-based framework for fault modeling and diagnosis. It provides a platform for developing and testing fault diagnosis algorithms.
Modelica	OpenModelica [76]	Open-source modeling and simulation environment. It can be used for modeling physical systems and performing fault diagnosis by simulating system behavior under different conditions.
Julia	FaultDetectionTools.jl [71]	Modern Julia version of FDITOOLS, that can be used for the synthesis of model-based FDI filters employing null-space methods.

~~*Julia:*~~

- ~~FaultDetectionTools.jl [71]: is a modern Julia version of FDITOOLS, that can be used for the synthesis of model-based filters employing null-space methods.~~

5.1.2. Frameworks for data-based approaches

Frameworks and libraries for data-based fault diagnosis are numerous, with an emphasis on the development of applications based on ML. Here, data ~~is~~ are typically preprocessed, a suitable ML algorithm is chosen, the model is trained on historical data, and then ~~,~~ the trained model is used for fault diagnosis on new data. ~~Classifying them by the programming languageemployed, we can mention:~~ These frameworks, organized by programming language, are listed in Table 5.

~~*Python* [77]: Scikit-learn, Tensorflow, PyTorch, Caffe, XGBoost, LightGBM, H2O. ai, Manifold [78], Hadoop [79], to mention a few.~~

Table 5: Data-based fault diagnosis frameworks by programming language

Language	Frameworks	Description
Python	Scikit-learn, Tensorflow, PyTorch, Caffe, XGBoost, LightGBM, H2O.ai, Manifold [78], Hadoop [79]	Frameworks and libraries for data-based fault diagnosis are numerous, with emphasis on the development of applications based on ML. Here, data is typically preprocessed, a suitable ML algorithm is chosen, the model is trained on historical data, and then, the trained model is used for fault diagnosis on new data.
Java	WEKA [80], Pache OpenNLP	Same process as above, using Java-based tools for data-based fault diagnosis.
Matlab	Various toolboxes [81]	Includes multiple built-in toolboxes that support ML-based fault detection workflows.
Julia	Flux, Knet, GLM, MultivariateStats, Clustering, LIBSVM, DecisionTree [19]	Julia's ML stack includes libraries for deep learning, statistical learning, clustering, and SVM-based fault classification.
R	Caret, e1071, Mlr, Nnet, KernLab, XGBoost [82]	R provides extensive support for machine learning, model tuning, and fault classification.
C++	Tensorflow, mlpack, Shogun, OpenNN, SHARK, FANN, DLib [83]	High-performance libraries implemented in C++ for scalable ML-based fault diagnosis.
Multiplatform	Tensorflow (Python, Julia, R, C++), Torch (Python, Julia, R, C++), Caffe (Python, C++), ScikitLearn (Python, Julia), XGBoost (Python, R, Java, C++, Julia)	Several frameworks are designed to be used across multiple programming languages, facilitating integration in diverse environments.

~~*Java:* WEKA [80], Pache OpenNLP.~~

~~*MatLab* [81]: Various toolboxes.~~

~~*Julia* [19]: Flux, Knet, GLM, MultivariateStats, Clustering, LIBSVM, and DecisionTree, to mention a few.~~

~~*R* [82]: caret, e1071, Mlr, Nnet, KernLab, and XGBoost, to mention a few.~~

~~*C++* [83]: Tensorflow, mlpack, Shogun, OpenNN, SHARK, FANN, and DLib, to mention a few.~~

~~Multiplatform: Tensorflow (Python, Julia, R, C++), Torch (Python, Julia, R, C++), Caffe (Python, C++), ScikitLearn (Python, Julia), XGBoost (Python, R, Java, C++, Julia).~~

5.2. Example applications

The various application domains of hybrid methods studied in the literature are displayed in Figure 6.

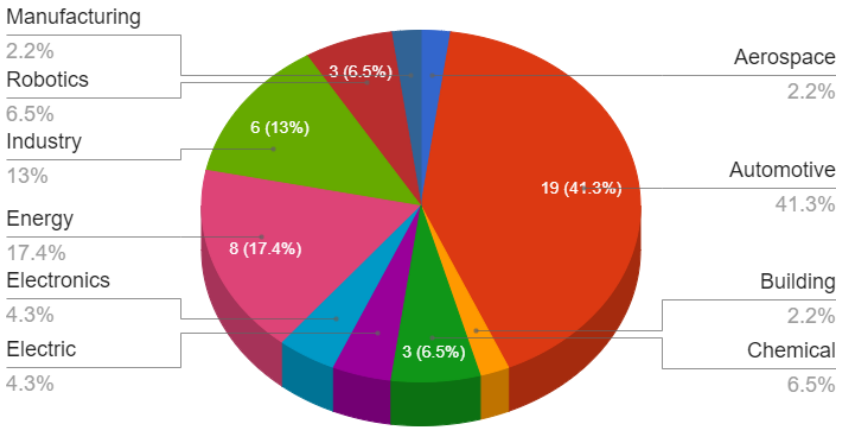


Figure 6: Application areas of hybrid methods.

Figure 6 shows that the application domains where hybrid methods have been applied the most are Automotive (41.3%) and Energy (17.4%) but there are many. The majority of applications in the Automotive, Aerospace, Robotics, Industrial, Chemical and Electrical fields target IoT and/or Prognostics objectives. Hence, in the following, we chose to deepen the applications in Manufacturing production, Prognostics and Health Management, IoT, and Energy Consumption and air quality, grouping together several items of Figure 6.

5.2.1. Manufacturing Production applications

Lin et al. [84] propose a framework for Manufacturing 4.0 of an Integrated Cyber Physical Environment. This framework uses concepts such as ML, distributed autonomous control, distributed artificial intelligence Distributed Artificial Intelligence (DAI) and digital twin, to define applications of systems modelling and simulation for manufacturing Industry 4.0 systems. It has a scenario of cyber space simulation to test discrete controllers for manufacturing 4.0.

Zheng et al. [85] also present a study that examines smart manufacturing systems for Industry 4.0. They propose a conceptual framework of smart manufacturing systems for Industry 4.0. In addition, demonstrative scenarios related to smart design, smart machining, smart control, smart monitoring, and smart scheduling, are presented.

Jiang et al. [86] propose a ML integrated design for an additive manufacturing framework, to learn the relationships between the design and performance spaces. ML techniques are used to model input–output relationships in both directions (property–structure relationships using structure–property input–output data). A case study was carried out to design a customized ankle brace that has a tunable mechanical performance.

Qureshi et al. [87] define a theoretical framework about ML methods for detection in the context of the Industry 4.0. The designed theoretical framework emphasizes the relevance of the monitoring operations based on ML approaches, and highlights the challenges to overcome, such as interoperability issues between technologies, financial impacts of technological adoption, etc.

Paraschos et al. [88] describe an integrated design and operation management based framework to obtain joint policies for the authorization of production, maintenance, remanufacturing, and recycling activities in circular manufacturing systems (which are manufacturing/remanufacturing systems). This framework can be used to add smart manufacturing devices in the manufacturing industry. Particularly, the framework utilizes a reinforcement learning technique and ad-hoc production control policies to improve the flexibility and the resilience of the examined systems to the ever-changing customer demand.

5.2.2. Prognostics and Health Management applications

Salehi et al. [89] present a data interpretation framework for structural health monitoring with incomplete (i.e., not continuous in time) and noisy

signals, using a sensing technology based on AI. They integrate AI methods, such as pattern recognition and ML methods, within the data interpretation framework developed to be used in a practical engineering problem: data-driven structural health monitoring of plate-like structures. They carry out finite element simulations on an aircraft stabilizer wing to validate the framework. The ML algorithms are integrated within the framework for performance assessment of the monitored structures for different levels of harvested energy.

Shcherbakov et al. [90] present a hybrid deep learning multitask framework integrating the Convolutional Neural Network (CNN) and Long Short-Term Memory model (LSTM) neural network to reflect the relatedness of remaining useful life prediction with health status detection process for complex multi-object systems in proliferation of CPS environments. The experimental results suggest that the proposed hybrid CNN-LSTM model is superior to existing methods, including traditional ML and Deep Learning (DL)-based methods. Their framework can provide strong support for the health management and maintenance strategy development of complex multi-object systems.

5.2.3. IoT applications

Boyes et al. [91] present an analysis framework for Industrial-IoT that can be used to enumerate and characterise Industrial-IoT devices when studying system architectures and analysing security threats and vulnerabilities.

Latif et al. [92] conducted a study that evaluates the potential of DL in Industrial-IoT applications, and presents a brief Industrial-IoT architecture with key enabling technologies. In addition, they discuss various DL algorithms (Deep Feedforward Neural Networks, Restricted Boltzmann Machines, Deep Belief Networks, AE, CNN, RNN, Generative Adversarial Networks), along with their theoretical foundations, and several software for their developments (TensorFlow, Microsoft CNTK, Keras, Caffe, Caffe2, MXNet, Torch, PyTorch, Theano, Chainer) and hardware (Raspberry Pi 4, NVIDIA Jetson Xavier, NVIDIA Jetson Nano, NVIDIA Jetson AGX Xavier, Google Coral, Google Coral Dev Board Mini, Rock Pi N10, HiKey).

Lee et al. [93] propose an integrated sensor/ML system for IoT. They develop an efficient real-time realization of a Gaussian mixture model for execution on the NXP FRDM-K64F embedded sensor board. The embedded algorithm automatically determines an appropriate set of GMM parameters using the embedded processor/sensor board in real time and can continuously

provide classification results. The integrated sensor/ML system provided real-time analytics with on-demand training. The ability of the system to classify signals and provide analytics directly from the board enables lower sensor data traffic.

Carvajal-Soto et al. [94] present an IoT-ML orchestration framework, applied to detect failures of surface mount devices during production. This framework is evaluated by using a discrete-event simulation of a production line for electronic devices, and it is based on the open-source tool LinkSmart, which makes its reutilization possible and extendable.

5.2.4. Energy Consumption and Air Quality applications

Jogunola et al. [95] design a hybrid DL framework to predict the energy consumption of different building types (e.g., residential and commercial) for different countries. The framework uses a CNN, an AE and a LSTM. The CNN layer extracts the features from the dataset, and AE and LSTM are used for prediction.

Ajagekar and You, in 2021 [96], propose a cutting-edge Quantum Computing (QC) and deep learning framework, specifically tailored for fault diagnosis in electrical power systems. This approach fuses the feature extraction strengths of conditional restricted Boltzmann machines with the classification capabilities of deep networks. A pivotal aspect of their work is addressing the computational challenges inherent in such complex deep learning models. This is achieved through QC-based training methodologies that effectively combine the benefits of quantum-assisted learning with classical training techniques.

Mohamed et al. [97] present an approach for optimizing the topology of ANN in the context of photovoltaic energy systems. Their method involves the use of a Genetic Algorithm (GA) for both diagnosis and online repair of these systems. This hybrid approach marks a significant advancement in fault detection capabilities. The authors report an impressive accuracy range of 97.6% to 98.9%, which surpasses the results achieved with conventional methods for identifying various types of faults, such as open circuits, degradation, short circuits, and shading faults.

Khan et al. [98] discuss the significant impact of DL and AI in energy. In their article, they introduce a hybrid energy forecasting model that leverages ML techniques. This model uniquely integrates three ML algorithms:

extreme gradient boosting, categorical boosting, and the RF method. Unlike typical ML approaches that primarily focus on hyperparameter tuning, this model emphasizes preprocessing through feature engineering to enhance forecasting accuracy. The efficacy of the proposed model is validated using a regression score, showcasing an R-squared value of 0.9212, indicating superior accuracy compared to existing models.

In the domain of hybrid models, Aggarwal et al. [99] use Particle Swarm Optimization to solve the issue of optimal hyperparameter calibration of LSTM networks for air quality forecasting. Additional works have explored hybrid and ensemble approaches for photovoltaic diagnosis. Sepúlveda-Oviedo et al. [100] propose a two-step method that uses Partial Least Squares (PLS) to extract latent components, which are then used to train a Linear Discriminant Analysis (LDA) model—significantly improving diagnostic performance (F1-score: 0.875). This work highlights how combining dimensionality reduction with classification boosts fault diagnosis. Further improvement is achieved through ensemble learning strategies, as shown in a complementary study [101], which integrates SVM, KNN, and Decision Trees using time–frequency features. This method remains robust across seasonal variations, with F1-scores ranging from 0.7 to 0.9, demonstrating the scalability and adaptability of ensemble approaches under real-world conditions.

5.3. Opportunities for hybridization of in-house project tools

The motivation for hybridization is to make the best use of data and knowledge to produce a hybrid AI-based system with high-level efficiency and explainability. In our case, we are interested in producing hybrid AI-based monitoring and on-line diagnosis systems. This section overviews the opportunities for hybridization of our in-house AI-ML systems, namely LAMDA and DYCLEE.

5.3.1. DYCLEE

DYCLEE implements a distance and density based algorithm that features several properties like handling non-convex, multi-density clustering with outlier rejection, and it achieves full dynamicity. All these properties are not generally found together in the same algorithm, and DYCLEE hence pushes forward the state of the art in this respect. The first step of the algorithm, qualified as distance-based step, operates at the rate of the data stream generated by the sensors and creates micro-clusters putting together data samples that are close in the sense of the L1-norm. Micro-clusters are

stored in the form of summarized representations, including statistical and temporal information. The second step, qualified as a density-based step, uses a density-based approach applied to the micro-clusters to create the final clusters by putting together dense micro-clusters: thanks to a forgetting process, clusters may emerge, drift, merge, split or disappear, hence following the evolution of the environment. Interestingly, DYCLEE has been recently extended to be able to deal with both numeric and categorical data.

While DYCLEE has previously been used for automatic generation of discrete event models [102, 103, 104], it has not resulted in hybridization with knowledge models. This represents a real opportunity for development. Two main lines of work can be mentioned. The first one is introducing constraints so that some samples fall into the same cluster may be an interesting idea in some contexts where the user knows that some targetted clusters must include dissimilar samples in terms of their features. The second is introducing constraints that represent meta-knowledge about diagnosis reasoning as is done in [105] with diagnostic trees.

5.3.2. LAMDA

LAMDA is a fuzzy algorithm that uses the maximum value of a membership degrees matrix, called Global Adequacy Degrees (GAD), to determine the membership of an individual X to a class k [106]. GADs in turn depend on the Marginal Adequacy Degrees matrix (MAD), computed using probability density functions (e.g., binomial function, Gaussian function, etc.) to measure the similarity of a descriptor of X with the same descriptor in each class k. One of the main advantages of LAMDA over other ML algorithms is that it can be used in classification/clustering tasks and can create new classes even after its training stage. LAMDA has been used in engineering processes. For instance, it has been used as a FDI approach to detect operating states, to identify normal and abnormal states in real time, or to determine the fault location ~~[107, 108, 109, 110]~~[107, 108, 109, 111, 110].

On the other hand, a significant number of extensions have been made to the LAMDA algorithm to improve its performance in the context of classification and clustering tasks, and also to perform control tasks. Some of those recent extensions for classification tasks are LAMDA-FAR [110] and LAMDA-HAD [112], and for clustering tasks are LAMDA-RD [113] and LAMDA-TIGAD [114]. Finally, different dynamic controllers have been designed based on this fuzzy algorithm, such as LAMDA-PID, LAMDA-Sliding-Mode Control (LSMC) and adaptive LAMDA controllers [115].

As it is possible to see, LAMDA has been used in different ML tasks (classification, clustering), being able to develop DBD models with performance qualities very similar to the best ML algorithms. Now, three interesting possible aspects to study with LAMDA are the following. First of all, develop semi-supervised diagnostic algorithms (with known and unknown states), recent results motivate this possible application [116]. On the other hand, the development of explainability methods based on LAMDA family algorithms, to make the diagnostic models more expressive, is another challenge. In this case, a specific reasoning method about the diagnosis made could be developed to generate meta-knowledge. Finally, using LAMDA in the feature engineering phases, to analyze the behavior of each variable/descriptor, is another possible application. These last two applications (methods of explainability and feature engineering based on LAMDA) could be defined at the level of each class, something innovative in those ML fields.

6. Conclusions

This work has carried out a systematic review of the literature related to the development of automatic diagnostic tasks in the context of "Industry 4.0 " for the manufacturing industry. The ~~literature review sought to determine the works that have combined and~~ objective was to identify the current state of hybrid diagnostic approaches that combine Data-Based Diagnosis (DBD) and Model-Based Diagnosis (MBD), and to ~~identify existing frameworks for~~ examine the existing machine learning (ML) frameworks relevant to Industry 4.0 ~~problems~~contexts. The reason for considering these aspects was, on the one hand, to exploit the advantages of both approaches, and on the other hand, to exploit the large amount of data that the IoT is making available.

~~Among the main results is that most of the combination schemes use~~ The bibliometric methodology applied (see Section 2) enabled the structured selection and classification of 75 relevant studies. From the analysis of retrieved papers (Section 3), we found that the most frequently explored architecture is serial integration, particularly combining observer-based ~~approaches or that there are not many frameworks for the development of applicationsbased on .~~ model components with classification-based ML algorithms. This reflects a tendency to leverage the strengths of each approach: residual generation through physical models and classification through data-driven techniques.

Visualization of the methods and roles (Section 3) confirmed that model-based techniques are primarily responsible for residual generation, while data-based

methods are more involved in fault isolation and detection. Furthermore, the bibliometric analysis demonstrated a growing interest in integrating hybrid models with explainability and online capabilities, which remains an open challenge for real-time industrial applications.

The review of frameworks and application areas (Section 5) highlighted that while many ML frameworks exist, there are relatively few mature, integrated environments tailored to hybrid diagnostic systems. One of the reasons is ~~because~~ that there is currently a high development of new ML techniques, and that the adaptation of each one to the operational context is an almost artisanal job to get the most out of them. Recent approaches, such as meta-learning, transfer learning, among others, will help to reuse many of the existing solutions based on ML, under middleware that can autonomously handle the processes of self-adaptation to these new contexts.

~~Finally, we have been able to see that our~~ Finally, applications are most prevalent in the automotive and energy sectors, often targeting predictive maintenance and smart monitoring.

Based on these findings, the following contributions and insights are emphasized:

- This is the first bibliometric review, to our knowledge, specifically analyzing hybrid DBD–MBD diagnostic systems within Industry 4.0 (see motivation and scope in Section 3).
- We proposed a new classification of integration strategies (serial, parallel, ad-hoc) and mapped them to diagnostic roles, facilitating future analysis and system design (see Section 3.2 and Figure 3).
- The review identified important open challenges, such as the lack of real-time scalability, insufficient support for interpretability, and the absence of reusable frameworks (see frameworks analysis in Section 5).

Our own machine learning algorithms, LAMDA and DYCLEE, ~~can be used in the hybridization opportunities that currently exist, in order to develop hybrid online monitoring and diagnosis systems based on . These systems could achieve high levels of efficiency,~~ also offer opportunities to address these challenges, as discussed in Section 5. Their potential integration in hybrid diagnostic pipelines could support online monitoring, improve explainability, and ~~could add functionalities such as feature engineering,~~ enable

intelligent feature selection through meta-learning ~~, transfer learning , among others, in their architecture~~and transfer learning mechanisms. The preliminary results obtained in the context of DBD with both algorithms motivate the performance of these future works.

In summary, this review offers a structured roadmap for advancing hybrid diagnostic systems in Industry 4.0 by clarifying integration strategies, identifying methodological trends, and highlighting key challenges already emerging in current literature.

6.1. Limitations

As with any bibliometric study, this review is subject to methodological limitations that must be acknowledged to ensure transparency and academic rigor. However, each limitation also opens up promising directions for future work.

First, the analysis was restricted to publications indexed in major bibliographic databases such as Scopus, Web of Science, and Google Scholar. While these databases offer extensive coverage, they are not exhaustive. For example, Jacobsen (2011) emphasizes that relying solely on indexed databases can lead to the omission of studies published in journals with limited visibility or from regions with less developed research infrastructure [117]. Similarly, Zhao (2014) highlights that the use of a single or limited set of databases increases the risk of selection bias in systematic reviews [118]. To address this, future studies could expand their scope by incorporating emerging repositories, preprint servers, and specialized engineering indexes to achieve greater coverage and inclusivity.

Second, the review was limited to articles published in English. This language restriction may have led to the exclusion of relevant studies published in other languages, particularly those addressing region-specific problems or developed by non-English-speaking research communities. As noted by Song et al. (2010), excluding non-English literature introduces a language bias that can distort the overall understanding of a research field by underrepresenting culturally specific contributions [119]. Future work could consider multilingual search strategies, supported by automatic translation tools, to mitigate this bias.

Third, grey literature such as theses, technical reports, and conference proceedings was not systematically included in this review. These sources often contain early-stage or practical research results not yet published in journals. Affengruber et al. (2021) highlight that many grey literature

documents are excluded due to limited metadata or non-standard indexing, which can result in skewed bibliometric analyses [120]. Including selected grey literature using clear quality and relevance criteria would enhance the comprehensiveness of future reviews.

Additionally, while the classification of studies followed a double-reviewer protocol with predefined criteria, the process remains manual and may not scale efficiently to larger corpora. To improve reproducibility and scalability, future efforts could explore the integration of computational techniques such as topic modeling and automatic classification to support reviewer decisions. For instance, the use of bibliometric mapping and co-occurrence analysis—as demonstrated in [14]—has shown strong potential for identifying clusters of research and improving the structure of literature reviews. Likewise, topic modeling approaches such as Latent Dirichlet Allocation (LDA) and their recent applications in hybrid diagnostic contexts have proven effective in uncovering hidden themes and research gaps across large corpora [121].

Finally, advances in generative AI offer promising prospects for enhancing the reach and synthesis capabilities of future reviews. Large language models (e.g., ChatPDF) allow for questions to be asked about the content, specific aspects to be summarized and explained, key findings and methodologies to be identified, or even taxonomies and visualization schemes to be proposed, thus improving the depth of the analysis.

Acknowledgments

This project has been carried out as part of the SticAmSud project HAMADI 4.0 "Hybrid Algorithms Based on Models and Data in Industry 4.0 ", n°22-STIC-06. It has also been supported by ANITI within the French "Investing for the Future – PIA3" program under the Grant agreement n°ANR-19-PI3A-0004. It is also one of the contributions of the IFAC Working Group "AI enhanced Safeprocess" of TC 6.4 SAFEPROCESS "Fault Detection, Supervision & Safety of Technical Processes". ~~Also, the authors want to~~ The authors would also like to acknowledge Minciencias of Colombia, Grant No. 144-2023~~of Minciencias of Colombia, for the support of~~ , for supporting this work.

Data Availability Statement

The data will be made available upon request, under reasonable request.

Disclosure Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Detailed table of reviewed articles

Ref	MBM	Function MBM	Function DBM	DBM	Area
[21]	OB	Residual generation	Residual generation, Fault detection, Fault isolation	REG, CLA, CLA	Industry
[22]	OB	Residual generation	Fault isolation	CLA	Chemical
[23]	PS	Residual generation	Fault isolation	CLA	Energy
[30]	SA	Residual generation	Fault isolation	CLA	Automotive
[31]	SA	Residual generation	Fault isolation	CLA	Automotive
[32]	SA, UVE	Residual generation, Residual generation	Residual selection	CLU	Automotive
[33]	SA, UVE	Residual generation, Residual generation	Residual selection, Fault detection, Fault isolation	CLU, CLA, CLA	Automotive
[34]	OB	Residual generation	Fault detection, Fault isolation	CLA, CLA	Electronics
[35]	OB	Residual generation	Residual generation, Fault isolation	CLA, CLA	Energy
[36]	SA	Residual generation	Fault isolation	CLA	Energy

[37]	SA	System identification	Residual generation, Residual generation, Fault isolation	REG, REG, CLA	Chemical
[38]	OB	Residual generation	Fault isolation	CLA	Industry
[39]	OB	Residual generation	Fault isolation	CLA	Energy
[40]	OB	Residual generation	Fault detection, Fault isolation	CLA, CLA	Robotics
[41]	OB, OB	Residual generation, Fault isolation	Fault isolation, Fault isolation	CLA, CLA	Robotics
[42]	SA	Residual generation	Fault isolation	CLA	Energy
[43]	OB, OB	Residual generation, Residual generation	Fault isolation, Fault isolation, Fault isolation	CLA, CLA, CLA	Energy
[44]	SA	System identification	Fault detection	REG	General
[45]	OB	Residual generation	Fault isolation, Fault isolation	CLU, CLA	Robotics
[46]	OB	Residual generation	Fault detection, Fault isolation	CLA, CLA	Robotics
[47]	OB, PS	Residual generation, Residual generation	Fault isolation, Fault isolation	CLA, CLA	Automotive
[48]	OB	Residual generation	Residual generation, Fault isolation	CLA, CLU	Aerospace
[49]	OB, OB	System identification, Residual generation	System identification, Residual generation	REG, REG	Robotics

[24]	PS, OB	Residual generation, Residual generation	Fault isolation	CLA	Automotive
[50]	OB	Residual generation	Fault isolation, Fault isolation	CLA, CLA	Electronics
[51]	PS, OB	Residual generation, Residual generation	Residual generation, Residual generation, Fault isolation, Fault isolation, Fault isolation, Fault detection, Residual selection	STT, STT, CLA, CLA, CLA, STT, STT	Automotive
[52]	OB	Fault isolation	Fault isolation	CLA	Energy
[25]	SA, UVE	Residual generation, Residual generation	System identification	REG	Automotive
[53]	SA, UVE	Residual generation, Residual generation	Fault detection, Fault isolation	CLA, CLA	Automotive
[26]	OB	Residual generation	Fault detection	REG	Energy
[54]	OB	Residual generation	System identification, Residual generation	REG, REG	Automotive
[55]	SA	Residual generation	Residual generation, Fault detection	REG, CLU	Automotive
[27]	SA	Residual generation	Residual generation, Fault detection	REG, CLU	Automotive

[56]	OB	Residual generation	Fault detection, Fault isolation	CLA, CLA	Chemical
[57]	SA, UVE	Residual generation, Residual generation	Residual selection	CLU	Automotive
[58]	PE	Residual generation	Fault isolation	CLA	Manufacturing
[59]	SA, UVE	Residual generation, Residual generation	Residual generation	REG	Industry
[60]	SA	Residual generation	Residual generation	REG	Automotive
[61]	SA	Residual generation	Residual generation, Fault detection, Fault isolation	REG, REG, REG	Industry
[62]	PS	Residual generation	Residual generation	REG	Automotive
[28]	SA, UVE	Residual generation, Residual generation	Residual generation, Fault isolation	REG, CLA	Electric
[122]	SA	Residual generation	Residual generation, Fault isolation	REG, CLA	Electric
[29]	PS	Residual generation	Residual selection, Fault isolation	STT, CLA	Building
[64]	SA	Residual generation	Fault isolation	CLA	Industrial
[65]	OB	Residual generation	System identification, Fault isolation	REG, CLA	Industry

[66]	SA	Residual generation	Residual generation, Fault isolation	REG, CLA	Automotive
[67]	SA	Residual generation	Fault detection, Fault isolation	REG, CLA	Automotive
[68]	OB	Residual generation	Residual generation	REG	Automotive

List of Terms and Abbreviations

AAKR Auto Associative Kernel Regression

AE Autoencoder

AI Artificial Intelligence

ANFIS Artificial Neural Fuzzy Inference System

ANN Artificial Neural Network

ARR Analytical Redundancy Relations

ARX Autoregressive with Exogenous Input

BC Bayessian Classifier

BN Bayessian Classifier

CNN Convolutional Neural Network

CLA Classification

CLU Clustering

CPS Cyber-Physical Systems

CUSUM Cumulative Sum method

DAI Distributed Artificial Intelligence

DBD Data Based Diagnosis

DL Deep Learning

DT Decision Tree

EKF Extended Kalman Filter

ELM Extreme Learning Machine

FFT Fast Fourier Transform

FDD Fault Detection and Diagnosis

FDI Fault Detection and Isolation

GCN Graph Convolutional Neural network

GRU Gated Recurrent Network

HGCN Hyper-graph Convolutional Neural Network

IoT Internet of Things

KF Kalman Filter

KNN K-nearest neighbors

KL Kullback-Leibler

KR Kernel Regression

LOLIMOT Local Linear Model Trees

LR Logistic Regression

LS Least Squares

LSTM Long Short-Term Memory model

MBD Model Based Diagnosis

ML Machine Learning

MSO Minimal Structurally Overdetermined subsystem

PCA Principal Component Analysis

PB Parameter-based

PE Parameter estimation

PS Parity-Space

PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses

OB Observer-based

REG Regression

RF Random Forest

RT Random Tree

RNN Recurrent Neural Network

STT Statistical tool

SVM Support Vector Machine

SA Structural Analysis

TB Three based

TS Takagi-Sugeno system

UAV Unmanned Aerial Vehicle

WLS Weighted Least Squares

QC Quantum Computing

GA Genetic Algorithm

DYCLEE Dynamic Clustering algorithm for tracking Evolving Environments

LAMDA Learning Algorithm for Multivariable Data Analysis

ODE An equation involving a function and its derivatives with respect to a single independent variable

SR Logical or mathematical relations between components of a system used to model its structure and behavior

LMI A convex optimization technique where constraints are expressed as linear matrix inequalities, commonly used in control theory and robust optimization

GMM A probabilistic model that assumes all data points are generated from a mixture of several Gaussian distributions with unknown parameters

PLS A statistical method that projects predictors and responses to a new space to extract latent variables that maximize the covariance between them, often used for dimensionality reduction and regression

LDA A supervised classification method that projects data onto a lower-dimensional space while maximizing class separability

GAD A scalar measure used to quantify the overall consistency between observed data and reference patterns in pattern recognition or classification systems

MAD A matrix that represents the degree of adequacy of each data point with respect to each class prototype or pattern, often used in classification algorithms such as LAMDA

LDA A generative probabilistic model used to discover topics from large collections of text or data, where each document is represented as a mixture of topics and each topic as a mixture of words

LSMC A robust nonlinear control technique that forces system trajectories to reach and stay on a predefined sliding surface, ensuring stability and performance under model uncertainties or disturbances

PID A widely used feedback control mechanism that combines proportional, integral, and derivative actions to maintain the desired output of a system

FAR A post-recognition classification step used to evaluate the functional condition or degradation level of a system

HAD A measure used to determine the most representative or dominant class/component based on adequacy degrees

RD A statistical distance metric designed to be less sensitive to outliers, often used for anomaly detection or classification

TIGAD A combined measure integrating typicality and intuitionistic fuzzy logic to assess global adequacy in classification or decision-making systems

References

[1] Kagermann H, Wahlster W. Industrie 4.0: Mit dem Internet der Dinge auf dem Weg zur 4. Industriellen Revolution. Sci. 2022;4(3):26.

[2] Peres RS, Jia X, Lee J, Sun K, Colombo AW, Barata J. Industrial Artificial Intelligence in Industry 4.0 - Systematic Review, Challenges and Outlook. IEEE Access. 2020;8:220121-39.

[3] Tidriri K, Chatti N, Verron S, Tiplica T. Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges. Annual Reviews in Control. 2016;42:63-81.

[4] Wilhelm Y, Reimann P, Gauchel W, Mitschang B. Overview on hybrid approaches to fault detection and diagnosis: Combining data-driven, physics-based and knowledge-based models. In: Procedia CIRP. vol. 99. Elsevier B.V.; 2021. p. 278-83.

[5] Tiddi I, Schlobach S. Knowledge graphs as tools for explainable machine learning: A survey. Artif Intell. 2022;302:103627.

[6] Mohindru G, Mondal K, Dutta P, Banka H. Chapter 12 - Mining challenges in large-scale IoT data framework – a machine learning perspective. In: De S, Dey S, Bhattacharyya S, Bhatia S, editors. Advanced Data Mining Tools and Methods for Social Computing. Hybrid Computational Intelligence for Pattern Analysis. Academic Press; 2022. p. 239-59.

[7] Nguyen G, Dlugolinsky S, Bobák M, Tran V, Álvaro López García, Heredia I, et al. Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey. Artif Intell Rev. 2019;52:77–124.

[8] Cioffi R, Travaglioni M, Piscitelli G, Petrillo A, De Felice F. Artificial Intelligence and Machine Learning Applications in Smart Production: Progress, Trends, and Directions. Sustainability. 2020;12(2).

[9] Aggarwal A, Toshniwal D. A hybrid deep learning framework for urban air quality forecasting. J Clean Prod. 2021;329:129660.

[10] Zhong RY, Xu X, Klotz E, Newman ST. Intelligent Manufacturing in the Context of Industry 4.0: A Review. Engineering. 2017;3(5):616-30.

[11] Page MJ, McKenzie JE, Bossuyt P, Grimshaw JM, Hróbjartsson A, Lalu MM, et al. The prisma 2020 statement: An updated guideline for reporting systematic reviews. Syst Rev. 2021;10(89).

[12] Madhavan M, Wangtueai S, Sharafuddin MA, Chaichana T. The Precipitative Effects of Pandemic on Open Innovation of SMEs: A Scientometrics and Systematic Review of Industry 4.0 and Industry 5.0. JOItmC. 2022;8(3):152. Available from: https://doi.org/10.3390/joitmc8030152.

[13] Hoyos W, Aguilar J, Toro M. Dengue models based on machine learning techniques: A systematic literature review. Artificial Intelligence in Medicine. 2021;119:102157. Available from: https://www.sciencedirect.com/science/article/pii/S0933365721001500.

[14] Sepúlveda-Oviedo EH, Travé-Massuyès L, Subias A, Pavlov M, Alonso C. Fault diagnosis of photovoltaic systems using artificial intelligence: A bibliometric approach. Heliyon. 2023;9(11):e21491.

[15] Toscano-Miranda R, Toro M, Aguilar J, Caro M, Marulanda A, Trebilcok A. Artificial-intelligence and sensing techniques for the management of insect pests and diseases in cotton: a systematic literature review. The Journal of Agricultural Science. 2022;160(1–2):16–31.

[16] Salazar C, Aguilar J, Monsalve-Pulido J, Montoya E. Affective recommender systems in the educational field. A systematic literature review. Computer Science Review. 2021;40:100377.

[17] Kitchenham B, Charters S. Guidelines for performing systematic literature reviews in software engineering; 2007. January 2007.

[18] Dhall D, Kaur R, Juneja M. Machine learning: A review of the algorithms and its applications. In: Lecture Notes in Electrical Engineering. vol. 597. Springer; 2020. p. 47-63. Available from: https://link.springer.com/chapter/10.1007/978-3-030-29407-6{_}5.

[19] Gao K, Mei G, Piccialli F, Cuomo S, Tu J, Huo Z. Julia language in machine learning: Algorithms, applications, and open issues. Comput Sci Rev. 2020;37:100254.

[20] Deng L. A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA Trans Signal Inf Process. 2014;3.

[21] Kappaganthu K, Nataraj C. Mutual information based feature selection from data driven and model based techniques for fault detection in rolling element bearings. Proceedings of the ASME 2011 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference. 2011.

[22] Benkouider AM, Kessas R, Yahiaoui A, Buvat JC, Guella S. A hybrid approach to faults detection and diagnosis in batch and semi-batch reactors by using EKF and neural network classifier. JLPPI. 2012;25(4):694-702.

[23] Costamagna P, De Giorgi A, Magistri L, Moser G, Pellaco L, Trucco A. A Classification Approach for Model-Based Fault Diagnosis in Power Generation Systems Based on Solid Oxide Fuel Cells. IEEE Trans Energy Convers. 2016;31(2):676-87.

[24] Luo J, Namburu M, Pattipati KR, Qiao L, Chigusa S. Integrated model-based and data-driven diagnosis of automotive antilock braking systems. IEEE Trans Syst Man Cybern Syst. 2010;40(2):321-36.

[25] Yang Y, Zhu C, Li S, Zhang S, Xiao P, Huang Z, et al. Robust Fault Diagnosis for DK-2 Brake Based on Both Data-driven and Model. Chinese Control Conference, CCC. 2018;2018-July:6055-60.

[26] Yao Y, Ding F, Liu W. A hybrid data-driven and model-based anomaly detection scheme for DER operation. 2022 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT). 2022:1-5.

[27] Svärd C, Nyberg M, Frisk E, Krysander M. Data-driven and adaptive statistical residual evaluation for fault detection with an automotive application. Mech Syst Signal Process. 2014;45(1):170-92.

[28] Chen Z, Xu J, Peng T, Yang C. Graph Convolutional Network-Based Method for Fault Diagnosis Using a Hybrid of Measurement and Prior Knowledge. IEEE Trans Cybern. 2022;52(9):9157-69.

[29] Gienger A, Wagner J, Böhm M, Sawodny O, Member S, Tarín C. Robust Fault Diagnosis for Adaptive Structures With Unknown Stochastic Disturbances. IEEE Trans Control Syst Technol. 2021;29(3):1131-46.

[30] Jung D, Ng KY, Frisk E, Krysander M. A combined diagnosis system design using model-based and data-driven methods. Conference on Control and Fault-Tolerant Systems, SysTol. 2016;2016-Novem:177-82.

[31] Jung D, Ng KY, Frisk E, Krysander M. Combining model-based diagnosis and data-driven anomaly classifiers for fault isolation. Control Eng Pract. 2018;80:146-56.

[32] Jung D, Sundstrom C. A Combined Data-Driven and Model-Based Residual Selection Algorithm for Fault Detection and Isolation. IEEE Trans Control Syst Technol. 2019;27(2):616-30.

[33] Jung D. Engine Fault Diagnosis Combining Model-based Residuals and Data-Driven Classifiers. IFAC-PapersOnLine. 2019;52(5):285-90.

[34] Sun L, Shi J. A hybrid fault diagnosis method of electromechanical actuator combining model-based and data-driven. In: IET Conference Publications. vol. 2018; 2018. p. 1528-32.

[35] Khorasgani H, Farahat A, Ristovski K, Gupta C, Biswas G. A framework for unifying model-based and data-driven fault diagnosis. Proceedings of the Annual Conference of the Prognostics and Health Management Society, PHM. 2018:1-10.

[36] Pérez-Pérez E, López-Estrada FR, Puig V. Diagnosis of faults in a wind turbine using Analytical Redundancy Relations and an Artificial Neural Network. In: Congreso Nacional de Control Automático CNCA 2020. 2013; 2020. p. 1-6.

[37] Won J, Oh H, Hong J, Kim M, Lee WY, Choi YY, et al. Hybrid diagnosis method for initial faults of air supply systems in proton exchange membrane fuel cells. Renew Energ. 2021;180:343-52.

[38] Atoui MA, Cohen A. Coupling data-driven and model-based methods to improve fault diagnosis. Comput Ind. 2021;128:103401.

[39] Cho S, Choi M, Gao Z, Moan T. Fault detection and diagnosis of a blade pitch system in a floating wind turbine based on Kalman filters and artificial neural networks. Renew Energ. 2021;169:1-13.

[40] Song J, Shang W, Ai S, Zhao K. Model and Data-Driven Combination: A Fault Diagnosis and Localization Method for Unknown Fault Size of Quadrotor UAV Actuator Based on Extended State Observer and Deep Forest. Sensors. 2022;22(19):7355.

[41] Esch V, Tim AJ. Hybrid Model-based and Data-based Fault Detection , Isolation , and Estimation for Wafer Handler Robots [Ph.D. thesis]. Eindhoven University of Technology; 2022.

[42] Xia L, Liang Y, Zheng P, Huang X. Residual-Hypergraph Convolution Network: A Model-Based and Data-Driven Integrated Approach for Fault Diagnosis in Complex Equipment. IEEE Trans Instrum Meas. 2023;72:1-11.

[43] Asefi S, Mitrovic M, Ćetenović D, Levi V, Gryazina E, Terzija V. Anomaly detection and classification in power system state estimation: Combining model-based and data-driven methods. SEGAN. 2023;35:101116.

[44] Jung D, Krysander M, Mohammadi A. Fault diagnosis using data-driven residuals for anomaly classification with incomplete training data. In: IFAC World Congress. Yokohama; 2023. p. 3234-9.

[45] Jiang H, Xu F, Wang X, Wang S. Active Fault-Tolerant Control Based on MPC and Reinforcement Learning for Quadcopter with Actuator Faults. In: IFAC World Congress. Yokohama; 2023. p. 1-8.

[46] El M, El B. Decision Tree based diagnosis for hybrid model-based / data-driven fault detection and exclusion of a decentralized multi-vehicle cooperative localization system. In: IFAC World Congress. 1994. Yokohama; 2023. p. 1-6.

[47] Slimani A, Ribot P, Chanthery E, Rachedi N. Fusion of Model-based and Data-based Fault Diagnosis Approaches. IFAC-PapersOnLine. 2018;51(24):1205-11.

[48] Imbassahy DWD, Marques HC, Rocha GC, Martinetti A. Empowering predictive maintenance: a hybrid method to diagnose abnormal situations. Appl Sci. 2020;10(19):1-27.

[49] Huang Y, Sun G, Tao J, Hu Y, Yuan L. A modified fusion model-based/data-driven model for sensor fault diagnosis and performance degradation estimation of aero-engine. Meas Sci Technol. 2022;33(8).

[50] Narasimhan S, Roychoudhury I, Balaban E, Saxena A. Combining Model-Based and Feature-Driven Diagnosis Approaches - A Case Study on Electromechanical Actuators. International Workshop on the Principles of Diagnosis. 2010:1-8.

[51] Ghimire R, Sankavaram C, Ghahari A, Pattipati K, Ghoneim Y, Howell M, et al. Integrated model-based and data-driven fault detection and diagnosis approach for an automotive electric power steering system. AUTOTESTCON (Proceedings). 2011;1001445:70-7.

[52] Lu F, Jiang C, Huang J, Wang Y, You C. A novel data hierarchical fusion method for gas turbine engine performance fault diagnosis. Energies. 2016;9(10).

[53] Yang Y, Zhu C, Xiao P, Zheng C, Huang Z. Fault Diagnosis of Direct Electro-pneumatic Brake Based on Model and Data-driven. Proceedings of the World Congress on Intelligent Control and Automation (WCICA). 2019;2018-July:1584-9.

[54] Liu N, Hu M, Wang J, Ren Y, Tian W. Fault detection and diagnosis using Bayesian network model combining mechanism correlation analysis and process data: Application to unmonitored root cause variables type faults. Process Saf Environ Prot. 2022;164(June):15-29.

[55] Svärd C, Nyberg M, Frisk E, Krysander M. Automotive engine FDI by application of an automated model-based and data-driven design methodology. Control Eng Pract. 2013;21(4):455-72.

[56] Sheibat-Othman N, Laouti N, Valour JP, Othman S. Support vector machines combined to observers for fault diagnosis in chemical reactors. Can J Chem Eng. 2014;92(4):685-95.

[57] Frisk E, Krysander M. Residual Selection for Consistency Based Diagnosis Using Machine Learning Models. IFAC-PapersOnLine. 2018;51(24):139-46.

[58] Saez MA, Maturana FP, Barton K, Tilbury DM. Context-Sensitive Modeling and Analysis of Cyber-Physical Manufacturing Systems for Anomaly Detection and Diagnosis. IEEE Trans Autom Sci Eng. 2020;17(1):29-40.

[59] Jung D. Isolation and localization of unknown faults using neural network-based residuals. In: Proceedings of the Annual Conference of the Prognostics and Health Management Society, PHM. vol. 11; 2019. p. 1-8.

[60] Jung D. Residual Generation Using Physically-Based Grey-Box Recurrent Neural Networks For Engine Fault Diagnosis. Cornell University. 2020:1-12.

[61] Mohammadi A, Krysander M, Jung D. Analysis of grey-box neural network-based residuals for consistency-based fault diagnosis. IFAC-PapersOnLine. 2022;55(6):1-6.

[62] Wang Y, Cui D, Guo F, Wang R, Cheng Z, Zhang W, et al. A Detection and Isolation of Faults Technique in Automotive Engines Using a Data-Driven and Model-Based Approach. In: Proceedings of the 11th International Conference on Modelling, Identification and Control (ICMIC2019); 2020. p. 503-20.

[63] Chen Z, Xu J, Ke H, Fan X, Peng T. Graph convolution network-based fault diagnosis method for the rectifier of the high-speed train. In: Proceedings - 2021 4th IEEE International Conference on Industrial Cyber-Physical Systems, ICPS 2021. IEEE; 2021. p. 491-7.

[64] Galve S, Vilajosana X, Puig V. Fault Diagnosis using a Combined Model and Data Based Approach: Application to a Water Cooling Machine. Conference on Control and Fault-Tolerant Systems, SysTol. 2021;2021-Septe:151-6.

[65] Enciso-Salas L, Pérez-Zuñiga G, Sotomayor-Moriano J. Fault Detection and Isolation for UAVs using Neural Ordinary Differential Equations. IFAC-PapersOnLine. 2022;55(6):643-8.

[66] Oromi A, Puig V, Galve S, Trapiello C. Robust Fault Diagnosis using a Data-based Approach and Structural Analysis. IFAC-PapersOnLine. 2022;55(6):211-6.

[67] Pérez-Pérez EJ, Puig V, López-Estrada FR, Valencia-Palomo G, Santos-Ruiz I. Neuro-fuzzy Takagi Sugeno observer for fault diagnosis in wind turbines. In: IFAC world Congress. Yokohama; 2023. p. 1-10.

[68] Sacchi N, Paolo G, Antonella I, Industriale I, Milano P. Actuator Fault Diagnosis With Neural Network-Integral Sliding Mode Based Unknown Input Observers. In: IFAC World Congress. Yokohama; 2023. p. 1-6.

[69] Rungger M, Zamani M. SCOTS: A Tool for the Synthesis of Symbolic Controllers. In: Proceedings of the 19th International Conference on Hybrid Systems: Computation and Control; 2016. p. 1-6.

[70] Frisk E, Krysander M, Jung D. A toolbox for analysis and design of model based diagnosis systems for large scale models. In: IFAC-PapersOnLine; 2017. p. 1-7.

[71] Varga A. Varga A, editor. Fault Detection and Isolation Tools (FDITOOLS) User's Guide. Andreas Varga; 2018.

[72] JXu Z, Jiang X. DIAGNOSYS: A Diagnostic Approach to Model-Based Diagnosis. In: Proceedings of the International Workshop on Systems Diagnosis (IWSD); 2017. .

[73] Nagaraja K, Bianchini R, Martin R, Nguyen TD. Using fault model enforcement to improve availability. In: Proceedings of the Second Workshop on Evaluating and Architecting System dependabilitY (EASY); 2002. p. 1-10.

[74] Frehse G. PHAVer: Algorithmic verification of hybrid systems past HyTech. In: International workshop on hybrid systems: computation and control; 2005. p. 1-17.

[75] Tinnerholm J, Pop A, Sjölund M. A modular, extensible, and modelica-standard-compliant openmodelica compiler framework in julia supporting structural variability. Electronics. 2022;11(11).

[76] Sepúlveda-Oviedo EH, Bermeo Clavijo LE, Méndez Córdoba LC. Effect of timing of umbilical cord clamping and birth on fetal to neonatal transition: OpenModelica-based virtual simulator-based approach. Biocybern Biomed Eng. 2024;44(3):716-30.

[77] Stančin I, Jović A. An overview and comparison of free Python libraries for data mining and big data analysis. In: 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO); 2019. p. 977-82.

[78] Zhang J, Wang Y, Molino P, Li L, Ebert DS. Manifold: A Model-Agnostic Framework for Interpretation and Diagnosis of Machine Learning Models. IEEE Trans Vis Comput Graph. 2019;25(1):364-73.

[79] Zhou H, Lascaud B. An Integrated Machine Learning Framework for Optimizing Unconventional Resources Development. In: SPE/AAPG/SEG Unconventional Resources Technology Conference; 2019. p. D023S041R003.

[80] van Rijn JN, Holmes G, Pfahringer B, Vanschoren J. The online performance estimation framework: heterogeneous ensemble learning for data streams. In: Machine Learning. vol. 107(1); 2018. p. 149-76.

[81] Paluszek M, Thomas S. 1. In: MATLAB machine learning recipes: a problem-solution approach. Apress; 2019. p. 325-9.

[82] Boehmke B, Greenwell BM. Hands-on machine learning with R. CRC press; 2019.

[83] Kolodiazhnyi K. Hands-On Machine Learning with C++: Build, train, and deploy end-to-end machine learning and deep learning pipelines. Packt Publishing Ltd; 2020.

[84] Lin WD, Low YH, Chong YT, Teo CL. Integrated Cyber Physical Simulation Modelling Environment for Manufacturing 4.0. In: 2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM); 2018. p. 1861-5.

[85] Zheng P, Sang Z, Zhong RY, Liu Y, Liu C, Mubarok K, et al. Smart manufacturing systems for Industry 4.0: Conceptual framework, scenarios, and future perspectives. Frontiers of Mechanical Engineering. 2018;13(2):137-50.

[86] Jiang J, Xiong Y, Zhang Z, Rosen DW. Machine learning integrated design for additive manufacturing. J Intell Manuf. 2022;33:1073—1086.

[87] Qureshi AH, Alaloul WS, Manzoor B, Musarat MA, Saad S, Ammad S. Implications of Machine Learning Integrated Technologies for Construction Progress Detection Under Industry 4.0 (IR 4.0). In: 2020 Second International Sustainability and Resilience Conference: Technology and Innovation in Building Designs(51154); 2020. p. 1-6.

[88] Paraschos PD, Xanthopoulos AS, Koulinas GK, Koulouriotis DE. Machine learning integrated design and operation management for resilient circular manufacturing systems. Comput Ind Eng. 2022;167:107971.

[89] Salehi H, Biswas S, Burgueño R. Data interpretation framework integrating machine learning and pattern recognition for self-powered data-driven damage identification with harvested energy variations. Eng Appl Artif Intell. 2019;86:136-53.

[90] Shcherbakov M, Sai C. A Hybrid Deep Learning Framework for Intelligent Predictive Maintenance of Cyber-Physical Systems. ACM Trans Cyber-Phys Syst. 2021.

[91] Boyes H, Hallaq B, Cunningham J, Watson T. The industrial internet of things (IIoT): An analysis framework. Comput Ind. 2018;101:1-12.

[92] Latif S, Driss M, Boulila W, Huma Ze, Jamal SS, Idrees Z, et al. Deep learning for the industrial internet of things (iiot): A comprehensive survey of techniques, implementation frameworks, potential applications, and future directions. Sensors. 2021;21(22):7518.

[93] Lee J, Stanley M, Spanias A, Tepedelenlioglu C. Integrating machine learning in embedded sensor systems for Internet-of-Things applications. In: 2016 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT); 2016. p. 290-4.

[94] Soto JAC, Tavakolizadeh F, Gyulai D. An online machine learning framework for early detection of product failures in an Industry 4.0 context. Int J Comput Integr Manuf. 2019;32(4-5):452-65.

[95] Jogunola O, Adebisi B, Hoang KV, Tsado Y, Popoola SI, Hammoudeh M, et al. CBLSTM-AE: A Hybrid Deep Learning Framework for Predicting Energy Consumption. Energies. 2022;15(3).

[96] Ajagekar A, You F. Quantum computing based hybrid deep learning for fault diagnosis in electrical power systems. Appl Energy. 2021;303:117628.

[97] mohamed Ah, nassar Am. New Algorithm for Fault Diagnosis of Photovoltaic Energy Systems. Int j comput. 2015;114(9):26-31.

[98] Waqas Khan P, Byun YC, Lee SJ, Park N. Machine Learning Based Hybrid System for Imputation and Efficient Energy Demand Forecasting. Energies. 2020;13(11).

[99] Liu A, Lu S. Functional design framework for innovative design thinking in product development. CIRP J Manuf Sci Technol. 2020;30:105–117.

[100] Sepúlveda-Oviedo EH, Travé-Massuyès L, Subias A, Alonso C, Pavlov M. Feature extraction and health status prediction in PV systems. Adv Eng Inform. 2022;53:101696.

[101] Sepúlveda-Oviedo EH, Travé-Massuyès L, Subias A, Pavlov M, Alonso C. An ensemble learning framework for snail trail fault detection and diagnosis in photovoltaic modules. Eng Appl Artif Intell. 2024;137:109068.

[102] Vignolles A, Ribot P, Chanthery E. Modeling complex systems with Heterogeneous Petri Nets (HtPN). In: 33rd International Workshop on Principle of Diagnosis–DX 2022; 2022. p. 1-10.

[103] Vignolles A, Chanthery E, Ribot P. Hybrid Model Learning for System Health Monitoring. IFAC-PapersOnLine. 2022;55(6):7-14.

[104] Barbosa NA, Travé-Massuyès L, Grisales VH. Diagnosability improvement of dynamic clustering through automatic learning of discrete event models. IFAC-PapersOnLine. 2017;50(1):1037-42.

[105] Goupil L, Chanthery E, Travé-Massuyès L, Delautier S. Tree based diagnosis enhanced with meta knowledge. In: 34th International Workshop on Principles of Diagnosis (DX'23); 2023. p. 1-10.

[106] Aguilar-Martín J, López R. The process of classification and learning the meaning of linguistic descriptors of concept. Approximate reasoning in decision analysi. 1982:165—175.

[107] Hernandez H, Camas J, Medina A, Perez M, Le Lann V. Fault Diagnosis by LAMDA methodology Applied to Drinking Water Plant. IEEE Latin America Transactions. 2014;12(6):985-90.

[108] Waissman J, Sarrate R, Escobet T, Aguilar J, Dahhou B. Wastewater treatment process supervision by means of a fuzzy automaton model. In: Proceedings of the 2000 IEEE International Symposium on Intelligent Control. Held jointly with the 8th IEEE Mediterranean Conference on Control and Automation; 2000. p. 163-8.

[109] Mora-Florez J, Barrera-Nunez V, Carrillo-Caicedo G. Fault Location in Power Distribution Systems Using a Learning Algorithm for

Multivariable Data Analysis. IEEE Transactions on Power Delivery. 2007;22(3):1715-21.

[110] Ruiz FA, Isaza CV, Agudelo AF, Agudelo JR. A new criterion to validate and improve the classification process of LAMDA algorithm applied to diesel engines. Engineering Applications of Artificial Intelligence. 2017;60:117-27.

[111] Araujo M, Aguilar J, Aponte H. Fault detection system in gas lift well based on artificial immune system. In: Proceedings of the International Joint Conference on Neural Networks, 2003.. vol. 3; 2003. p. 1673-7 vol.3.

[112] Morales L, Aguilar J, Chavez Garcia G, Isaza C. LAMDA-HAD, An Extension to the LAMDA Classifier in the Context of Supervised Learning. International Journal of Information Technology & Decision Making. 2019 12;19.

[113] Morales L, Aguilar J. An Automatic Merge Technique to Improve the Clustering Quality Performed by LAMDA. IEEE Access. 2020;8:162917-44.

[114] Botía Valderrama JF, Botía Valderrama DJL. On LAMDA clustering method based on typicality degree and intuitionistic fuzzy sets. Expert Systems with Applications. 2018;107:196-221.

[115] Morales L, Aguilar J, Camacho O, Rosales A. An intelligent sliding mode controller based on LAMDA for a class of SISO uncertain systems. Information Sciences. 2021;567:75-99.

[116] Quintero-Gull C, Aguilar J. LAMDA-HSCC: A semi-supervised learning algorithm based on the multivariate data analysis. Expert Systems with Applications. 2022;202:117479.

[117] Jacobsen K. P1-182 A taxonomy of bias in systematic reviews. Journal of Epidemiology & Community Health. 2011;65(Suppl 1):A116-7.

[118] Zhao JG. Combination of multiple databases is necessary for a valid systematic review. International orthopaedics. 2014 10;38:1.

[119] Song F, Parekh S, Hooper L, Loke Y, JJ R, Sutton A, et al. Dissemination and publication of research findings: An updated review of related biases. Health technology assessment (Winchester, England). 2010 02;14:iii, ix-xi, 1.

[120] Affengruber L, Dobrescu A, Persad E, Klerings I, Wagner G, Sommer I, et al. Characteristics and recovery methods of studies falsely excluded during literature screening—a systematic review. Systematic Reviews. 2022 11;11.

[121] Mustak M, Salminen J, Plé L, Wirtz J. Artificial intelligence in marketing: Topic modeling, scientometric analysis, and research agenda. J Bus Res. 2021;124:389-404.

[122] Chen C, Seo H, Zhao Y. A novel pavement transverse cracks detection model using WT-CNN and STFT-CNN for smartphone data analysis. International Journal of Pavement Engineering. 2021;0(0):1-13.