

Article

A Data-Driven Approach to Extend Failure Analysis: A Framework Development and a Case Study on a Hydroelectric Power Plant

Sara Antomarioni ^{1,*}, Marjorie Maria Bellinello ^{2,3}, Maurizio Bevilacqua ¹, Filippo Emanuele Ciarapica ¹, Renan Favarão da Silva ³ and Gilberto Francisco Martha de Souza ³

¹ Department of Industrial Engineering and Mathematical Science, Università Politecnica delle Marche, Via Brecce Bianche, 12, 60131 Ancona, Italy; m.bevilacqua@staff.univpm.it (M.B.); f.ciarapica@univpm.it (F.E.C.)

² Department of Mechanical and Maintenance Engineering, Federal University of Technology of Paraná, 3165-Rebouças, Curitiba 80230-901, Brazil; belinelli@utfpr.edu.br

³ Department of Mechatronics and Mechanical Systems Engineering, USP—University of São Paulo, Avenida Professor Mello de Moraes 2231, São Paulo 05508-030, Brazil; renanfavarao@usp.br (R.F.d.S.); gfm.souza@usp.br (G.F.M.d.S.)

* Correspondence: s.antomarioni@univpm.it

Received: 18 November 2020; Accepted: 2 December 2020; Published: 3 December 2020



Abstract: Power plants are required to supply the electric demand efficiently, and appropriate failure analysis is necessary for ensuring their reliability. This paper proposes a framework to extend the failure analysis: indeed, the outcomes traditionally carried out through techniques such as the Failure Mode and Effects Analysis (FMEA) are elaborated through data-driven methods. In detail, the Association Rule Mining (ARM) is applied in order to define the relationships among failure modes and related characteristics that are likely to occur concurrently. The Social Network Analysis (SNA) is then used to represent and analyze these relationships. The main novelty of this work is represented by support in the maintenance management process based not only on the traditional failure analysis but also on a data-driven approach. Moreover, the visual representation of the results provides valuable support in terms of comprehension of the context to implement appropriate actions. The proposed approach is applied to the case study of a hydroelectric power plant, using real-life data.

Keywords: maintenance; hydroelectric power plant; reliability; data-driven; association rule; data mining

1. Introduction

Power plants aim to efficiently supply the electric demand, considering the economic, reliability, and environmental aspects [1–3]. The implementation of an accurate maintenance strategy represents a critical issue from many points of view since, for example, power plants are characterized by complex structures [4]. Additionally, an inadequate maintenance strategy may result in energy losses and unpredictable operating conditions [5]. Considering the renewable energy field, hydroelectric sources globally provide the broadest supply [6]; thus, it is fundamental that the maintenance management ensures a smooth operation deployment [7]. This aspect can be critical due to the complex nature of hydroelectric power plants, which requires the analysis of several variables, items, and operating conditions [8] to evaluate how a single failure can trigger a series of cascade effects penalizing the entire production system.

Several techniques have already been applied in the extant body of literature regarding the failure analysis and its extension (e.g., [9,10]). Some of them are applied to improve the automatic identification of the potential failure modes [11], others focus on the improvement of the risk assessment process [12,13], or to integrate the failure analysis and the remaining useful life prediction [14,15]. Some works focus on similar perspectives, such as the study of failure dynamics based on the complex network analysis [16] or the analysis of fault interactions [17]. However, this paper addresses the lack of a unique framework integrating the two aspects and focuses on the lack of research in terms of how the failure analysis results can be explored to derive further information.

To this end, the extension of the failure analysis—typically carried out by the operations and maintenance managers—is proposed by exploring its outcomes through the implementation of data-driven techniques. A framework taking the Failure Mode and Effects Analysis (FMEA) as a starting point is developed by analyzing the results of the FMEA through the implementation of the Association Rule Mining (ARM) and the Social Network Analysis (SNA). The two techniques are devoted to extracting hidden information, patterns, and tendencies in a large amount of data [18], such as the one constituted by the failure analysis of a hydroelectric power plant. Specifically, ARM is used to determine attribute-value relationships frequently verifying contextually [19] during the occurrence (or potential occurrence) of a failure. SNA, instead, aims at providing both a visual representation of the networks constituted by the Association Rules (ARs) and at identifying the most critical actors across the network [20], i.e., the events that must be avoided or limited as much as possible, since their occurrence represents a hazard for the remaining of the plant [21]. Hence, through the interpretation of the results harbored by the proposed techniques, the decision-makers can study the domino effect among the factors influencing the plant's reliability.

The main novelty of this work is represented by support in the maintenance management process based not only on the traditional failure analysis but also on a data-driven approach that allows a complete understanding of the relationships possibly hidden by a large amount of data [22]. Starting from the FMEA harbors different advantages. On the one hand, it allows the company to capitalize on using a traditional approach further and improving its knowledge. On the other hand, the data-driven analysis is carried out basing on the expertise of the multi-functional team that is usually charged with deploying the FMEA—so benefiting from different and comprehensive ranges perspectives. The application to the case study of a hydroelectric power plant supports in understanding the practical introduction in a real-life environment.

The remainder of the paper is organized as follows: the next subsection reports a review of the existing data-driven application to the failure analysis; Section 2 is devoted to describing the material and methods applied in this work, while a case study is presented in Section 3. In Section 4, the theoretical and practical implications derived from implementing the proposed approach are discussed, and the main conclusions are drawn in Section 5.

Data-Driven Failure Analysis Approaches

In the existing literature, several contributions involve the implementation of data-driven approaches to failure analysis. Specifically, the main focus regards the joint implementation of the Failure Mode and Effects Analysis or the Failure Mode Effects and Criticality Analysis and data-driven techniques. For example, the FMEA and fuzzy inference can be applied to perform a thorough criticality evaluation, considering safety issues and production performance [23]. Other authors, instead, propose applications for the automatization of the failure analysis: in [24], for example, the failure modes identification is automatized through knowledge-based fault models, so that the experiences derived from previous projects could be included in new ones, while, in [11] behavior trees are applied for the same objective. Instead, text mining applications can be found to determine all the potential failure modes related to the components [25]. Bayesian networks are often applied to improve the performance of the failure analysis, as testified by [26], who use them to improve the risk and reliability assessment, and [12] that integrates the opinion of the experts in the risk assessment

process. In [14], the remaining useful life of components is predicting through data mining techniques, and the outcomes of the analysis are used to update the risk of failure. In [15], the remaining useful life is predicted in case of multiple failure modes occurrence, comparing a methodology based on logical data analysis and non-parametric cumulative incidence functions with traditional techniques (e.g., neural networks, support vector machines).

Improvements of the traditional failure analysis include prioritizing components based on their risk of failure: multi-criteria decision-making approaches are frequently used to synthesize the experts' judgments and include all the perspectives in the analysis [27,28]. The fuzzy ordered weighted method and DEMATEL are implemented in [29] to calculate each component's risk level and rank them. Risk evaluation and ranking of the risk factors can also be performed by defining the fuzzy digraph and matrix [30] or through the identification of a synthetic failure index that guides the selection of improvement actions to maximize the reliability of the system [31].

Data-driven failure analysis applications are frequently implemented on power plants, as testified by several works on the topic. Early works mainly involve the implementation of hybrid approaches, namely the integration of model-driven and data-driven techniques, for the prognostic health management of power systems ([32,33]). Some papers, instead, focus on the fault diagnosis aspects: in [34,35], principal component analysis and independent component analysis are respectively applied to diagnose the failures in thermal power plants components, while in [36] a dynamic, uncertain causality graph is optimized through a genetic algorithm to diagnose the failures in a nuclear power plant. Artificial neural networks can also be applied to detect failures in thermal plants (e.g., [37,38]) or in wind turbines [39]. Nonetheless, the remaining useful life estimation is widely addressed by combining different algorithms to improve overall performance [40].

Noteworthy, the existing contributions mainly deal with the implementation of techniques and approaches for diagnosing or detecting failures. Two interesting contributions can be compared to the one proposed in this work: specifically, in [16], early fault detection on a photovoltaic power plant is performed by analyzing complex networks of sensors to discover hidden dynamics non-observable from the observation of a single sensor. Additionally, in [17], the implementation of different data-driven techniques is proposed to detect faults and cluster them to analyze the interactions among faults. In the proposed approach, the interactions among failures, items, and failure modes are investigated through the Association Rule Mining and then further explored by analyzing a network. Moreover, the network is built based on the data of the failure analysis and does not reflect the plant's physical structure.

2. Materials and Methods

In this section, the proposed approach is detailed. Specifically, this work aims to provide a framework for improving the potential failures' analysis to extract previously unknown information and adapt the maintenance strategies accordingly. From this perspective, three main steps can be highlighted:

1. Data collection and understanding
2. Determination of the relevant associations
3. Social network analysis and insights definition.

2.1. Data Collection and Understanding

The procedure relies on the results of the analysis of the past or potential failures; thus, the first step is collecting all the possible data on this matter. Failure Modes and Effects Analysis [41] represents a valid starting point for applying the present framework. The system under investigation is broken down to identify its elementary items (sub-systems or parts) separately analyzed. The objective of the decomposition is to anticipate all the potential failure modes and effects. The failure analysis is, in this way, carried out collaboratively, involving interdisciplinary groups (e.g., Operations and

Maintenance engineers, managers, technicians, on-field personnel) in the discussion of the main features of the system.

The main advantage of taking the FMEA as a starting point is that several perspectives are questioned so that a complete understanding of the potential failures and effects is achieved. Additionally, due to the multi-disciplinary team's contributions in the FMEA, it is possible to limit the subjective bias related to each role and avoid the related uncertainty. Finally, in carrying out the FMEA, a dataset containing the system's equipment under investigation, the potential failure modes, and the associated effects are created and can be analyzed through the association rule mining. Additional information can be added, such as the mean time to repair (MTTR) or the failure modes' criticalities. Starting from the FMEA has different advantages: on the one hand, it allows the company to improve the plant's knowledge further. On the other hand, the data-driven analysis is carried out basing on the expertise of the multi-functional team that is usually charged with deploying the FMEA—so benefiting from different and wide-ranges perspectives.

2.2. Determination of the Relevant Associations

The second step of the procedure requires defining the relevant associations among the events extracted from the FMEA dataset. Specifically, relevant information may regard the failure modes frequently occurring on different items or the same effects deriving from different failure modes. This exploratory analysis aims to extend the existing knowledge of the analyzed system. The larger the dataset, the more complex the data analysis is: in this sense, data-driven techniques overcome the traditional statistical ones, which are no longer able to provide useful insights alone, without the need for formulating hypotheses. Hence, the ARM selection represents a valid alternative [42] since it allows both the simultaneous analysis of a large amount of data and an intuitive results interpretation [43] due to the structure of the outcomes. In this sense, it is also easier to involve the non-expert of the data analytics field to understand and implement the insights obtained in the data-driven analysis.

The applications of the ARM are widespread and can be found in different fields, such as the operations and production-related ones; however, the first one regards the extraction of hidden patterns from large datasets for marketing scopes [44]. In the following, a formal definition of ARs and ARM is provided.

Let $K = \{k_1, k_2, \dots, k_n\}$ be a set of data, called items, and $T = \{t_1, t_2, \dots, t_m\}$ the set of transactions; each transaction is composed of a set of items, namely an itemset, taken from K . An Association Rule (AR) is an implication $I \rightarrow J$ such that I and J are item sets ($I, J \subseteq K$). They have not items in common ($I \cap J = \emptyset$). The itemset I is called the body or left-hand side of the rule, while J is the head or right-hand side. The two principal metrics for evaluating the quality of a rule are support (1) and confidence (2) [26].

$$\text{Support } \{I \rightarrow J\} = \frac{\#\{I, J\}}{\#\{T\}} \quad (1)$$

$$\text{Confidence } \{I \rightarrow J\} = \frac{\text{Support } \{I \rightarrow J\}}{\text{Support } \{I \rightarrow \text{true}\}} \quad (2)$$

Since the function $\#\{*\}$ represents the cardinality of the itemset, the support can be defined as the number of itemsets taken from T in which there are both I and J , i.e., the probability of finding both I and J in the transaction set. Instead, the confidence determines the number of itemsets containing both I and J among the ones containing I . Hence, it is a measure of the conditional probability of finding J , given the fact that a transaction contains I .

The ARM can be performed through several algorithms: in this application, the Frequent Pattern-growth (FP-growth) [45] is selected due to its better efficiency [46]. This algorithm requires the scan of the transaction set $T = \{t_1, t_2, \dots, t_m\}$ to identify the items appearing more frequently than a user-defined threshold, i.e., the minimum support (min_sup). Those that do not meet the min_sup requirement are excluded from the scan: in this way, itemsets composed of several items are considered only if each of the single items has the support higher than min_sup. Starting from

the selected itemsets, the rules meeting a minimum confidence threshold (`min_conf`) are generated according to the following procedure:

1. Define `min_sup`: the minimum support threshold required to consider a rule;
2. Define `min_conf`: the minimum confidence threshold required to consider a rule;
3. Use the FP-growth algorithm [27] to determine the frequent itemsets;
4. Combine pairs of frequent itemsets to create the association rules; delete rules having confidence lower than `min_conf`.

The association rules mined are used as input for the third step of the research approach.

2.3. Social Network Analysis and Insights Definition

The SNA is usually applied to investigate social structures relying on the network and graph theory [47]. A network is defined by an ordered pair of nodes (N) connected by edges (E), $G = (N, E)$. The traditional application of SNA is the study of the interactions among a set of actors, respectively represented by edges and nodes. In the current work, the frequent itemsets identified through the FP-growth algorithm are the social network actors, while the ARs describe their interactions. Indeed, in this framework, the aim is to deploy an SNA to display failure modes, effects, and criticalities frequently occurring concurrently in order to clarify the interpretation of the association rules extracted. In this way, an overview of the patterns to take into account is provided, and proper insights can be defined based on the nature of network structure.

For example, if the rule $a \rightarrow b$ is extracted, a and b will be nodes of the network, and they will be connected. The confidence of the rule $a \rightarrow b$ is the weight of the edge connecting them. If the rule $b \rightarrow a$ is defined too, then the connection between the two nodes will be double arrowed; however, the weight of the edge between b and a (i.e., the confidence of the rule $b \rightarrow a$) can be different than the other one. For each node, the Out-Degree (OD) [48] is determined as follows:

$$OD_j = \sum_{i=1}^n w_i \quad (3)$$

Specifically, the OD of a node j is the weighted sum of the n edges outgoing from j : a high OD indicates a strong influence of a node on its successors, highlighting the need to control that node. Basing on the OD metric and having a complete visualization of the interrelations among the items identified during the failure analysis, it is possible to define useful insights to extend the plant's knowledge.

During the analysis of the SNs, it is noteworthy to consider an additional metric, the Betweenness Centrality (BC) [49]: the shortest weighted paths among all the couples of nodes are determined, and the BC_j equals the sum of the shortest weighted paths on which node j appears. This metric measures the influence of node j across the network [50] since a node having a high BC value can be considered as a bridge among separate portions of the network.

3. Application to a Case Study

3.1. Case Study

The proposed approach is applied to a Brazilian hydroelectric power plant (HPP). It is equipped with three hydro generators type Kaplan units, which operate at 166.25 MW. Kaplan hydro generators units can work where a small head of water is involved; the turbines are applied in sites having a head range of 2–40 m. Since the angles of their blades can be modified to adapt to the water flow, Kaplan turbines can also work efficiently at a wider range of water head, allowing for variations in the dam's water level. Three principal systems compose the hydro-generator Kaplan unit: speed governor, turbine system, and axis. In all, 152 components have been identified during the FMEA analysis of the HPP; thus, they are treated in the failure analysis.

3.2. Data-Driven Framework Application

The hydroelectric industry requires a high level of availability and reliability. The FMEA is regularly carried out on the system to identify components' criticality and prioritize their maintenance. In this way, the risk involved in the production process is monitored; however, further knowledge of the HPP can be extracted through the implementation of the proposed approach. The FMEA is performed following the US Military Standard's recommendation, adopting a bottom-up approach: the system under investigation is broken down to analyze its elementary components separately. Through the breaking-down, the objective is to provide an accurate description of the failure modes, effects, and impact on safety, environment, and assets. A collaborative approach is adopted to deploy the FMEA so that the HPP's main features are discussed by interdisciplinary groups of people involved in the system's operations at different levels (e.g., maintenance engineers, managers, on-field technical personnel).

The dataset structure used as a starting point for the data-driven analysis is reported in Table 1. Specifically, data refer to the FMEA traditionally carried out by the company and regard:

1. System: one of the three main systems composing the HPP;
2. Name: one of the 152 components relevant for the study;
3. PFM: potential failure mode occurring on the component;
4. Main functions: effect of the PFM on the main functionality of the component;
5. FR: the failure rate of the component (it can be actual if the FM has already occurred or theoretical if the FM is potential);
6. MTTR: the mean time to repair, expressed in hours;
7. SAI: the impact of the FM occurrence on the availability of the system;
8. IOP: the impact of the FM occurrence on people;
9. EI: the impact of the FM occurrence on the environment.

Table 1. Structure of the Failure Mode and Effects Analysis (FMEA) dataset.

System	Name	PFM	Main Functions	FR	MTTR	SAI	IOP	EI
AXIS	Generator Shaft	Break	Provide rotation for electricity generation	0.000001	168	9	7	1

Attributes 7–9 are evaluated by the multi-disciplinary team members responsible for performing the FMEA on a 1:9 scale.

The second step of the approach regards the determination of the relevant associations through the ARM. The dataset, whose structure is presented in Table 1, comprises 432 transactions (rows of the dataset). The components analyzed are 152, while the distinct PFM is 113: this means that the same failure mode can affect different components. The software selected for this case study is RapidMiner studio: its main strength is the graphical interface that does not require any programming language knowledge, making it easier to implement in an industrial context.

First, to identify the association rules worthy of investigation and not limit their extraction, null support, and confidence threshold are set ($\text{min_sup} = 0$; $\text{min_conf} = 0$). The ARs among all the nine attributes explained in Table 1 are mined. Indeed, min_sup and min_conf thresholds have to be set based on the specific case study, considering the dimensions of the dataset and relying on the decision-maker's expertise, since there is no absolute value suitable for all the cases [51].

In all, 4147 associations among 362 itemsets are extracted and are represented using the open-source software Gephi. To limit the study to the relevant associations and to be able to analyze them properly, the following procedure is applied:

1. Create the SN using all the ARs;
2. Determine the most interesting node based on the *OD*;
3. Filter the ARs and create more specific SNs, limiting the analysis to the nodes considered more relevant;
4. Formalize the information extracted.

The turbine node has the highest *OD* (4.645) if compared to the axis (4.301) and the speed governor (4.419); hence, the ARs referring to this portion of the HPP is extracted. Therefore, the ARs referring to the turbine are extracted to focus the analysis on this branch of the system primarily. This filter leads to the mining of 1248 ARs (127 itemsets). To focus on the most relevant portions of the network, the attributes Item, PFM, and Functions are taken into account, creating an SN composed of 102 nodes and 308 arcs.

Interestingly, as reported in Figure 1, 13 communities of nodes originated, considering these ARs. This structure indicates that not all the nodes are connected among the others, thus limiting the potentiality of spreading their occurrence across the network. Indeed, if the nodes are not connected among them, there is no relation among the events represented by such nodes. This aspect limits the attention that the maintenance managers have to pay to the so-called domino effect. In particular, 8 networks simply represent the connection among the item, the related function, and failure modes: this information is not new since it can be derived from the FMEA with no reason for extending the analysis through the data-driven framework. Indeed, the proposed approach aims to extend the current body of knowledge on the existing plant by extracting previously unknown relationships. On the contrary, there are 3 networks (Figure 1d,e,i) in which relevant and previously unknown relationships are displayed. Indeed. These relationships involve more than one item and several PFM, supporting the maintenance managers in identifying potential combined inspections and actions to anticipate the potential failures across the plant.

For example, in Figure 2a—which deploys Figure 1i in detail, it can be noticed that the node PFM = External leak acts as a bridge among the two portions of the network: indeed, its BC is the highest in the SN (74.67). In this sense, the occurrence of an external leak may have an impact on control valves, the oil pump, and the pump drainage system, as evidenced in Table 2. The confidence associated with the three rules (PFM = external leak → Item = pump drainage system; PFM = external leak → Item = oil pump; PFM = external leak → Item = control valves) is 0.333 since it is equiprobable that, when an external leak occurs, the item is one of those listed.

These connections highlight the need for establishing a protocol for the inspection of the item when an external leak occurs. Specifically, such protocol should require the verification of the normal functioning of the items, e.g., the flow of the fluid at the desired pressure (Function = promote the flow of fluid at the desired pressure → Item = Oil pump), the drainage of the water (Function = Drain the water that eventually passes through the inner cover seal → Item = Pump drainage system) and the control of the oil flow (Function = Check the oil flow for actuating the gate → Item = Control valves). The confidence is 100% for the three cases since each function is associated with a single item.

Similarly, in Figure 2b, the communities of nodes reported in Figure 1e are reported. The considerations drawn for Figure 2a can be extended to this community too. Indeed, the two items noted in this network (i.e., gate and adduction grid) share a common potential failure mode (PFM = deterioration of concrete) that acts as a bridge for the two portions of the network. When this failure mode occurs is then essential to check whether both the items are normally functioning or if an intervention is needed. As noticeable from Table 3, when the potential failure mode “deterioration of concrete” occurs, the confidence of 50% indicates that it regards either the gate or the adduction grid (see the first two rules reported in Table 3).

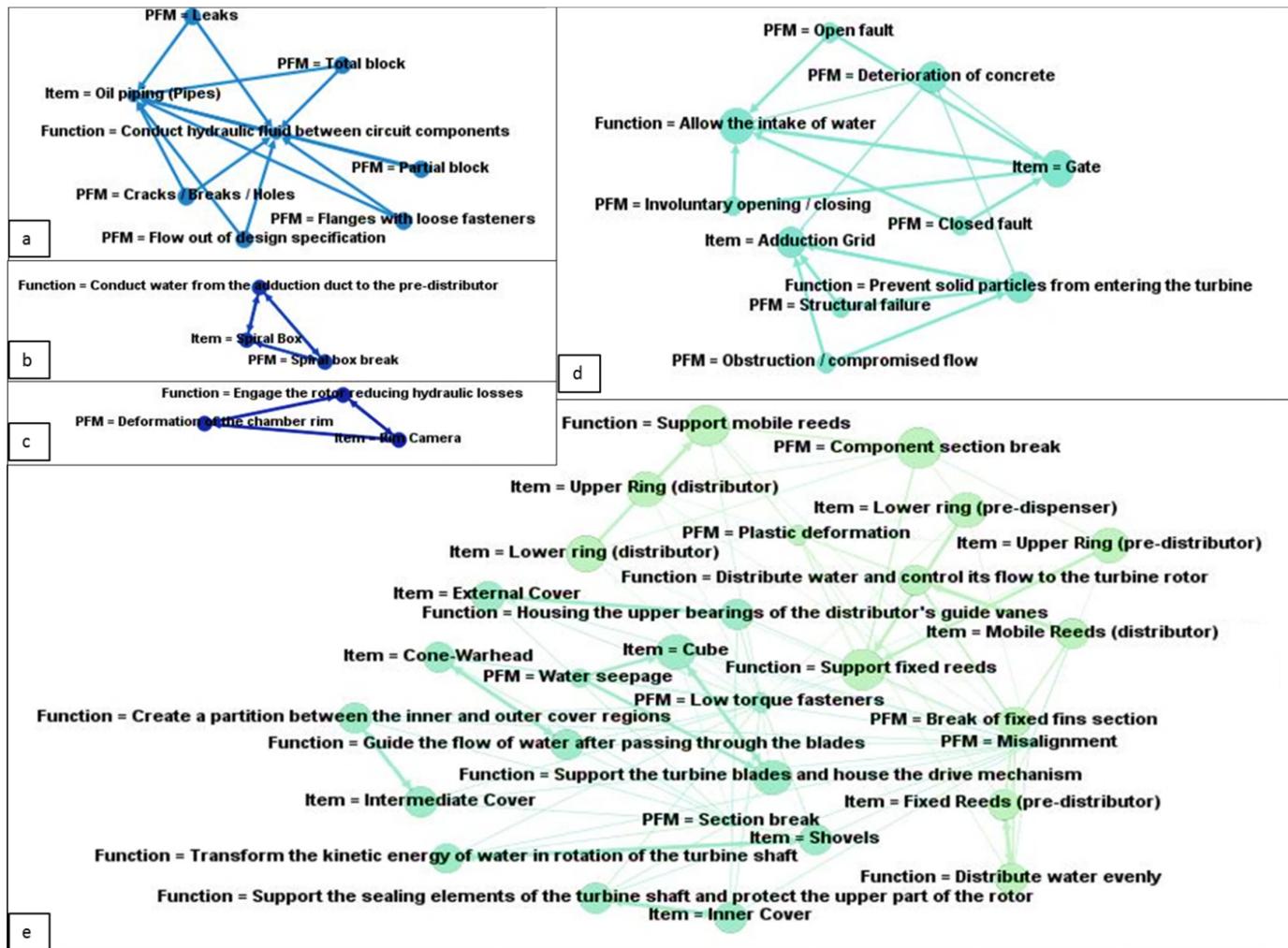


Figure 1. Cont.

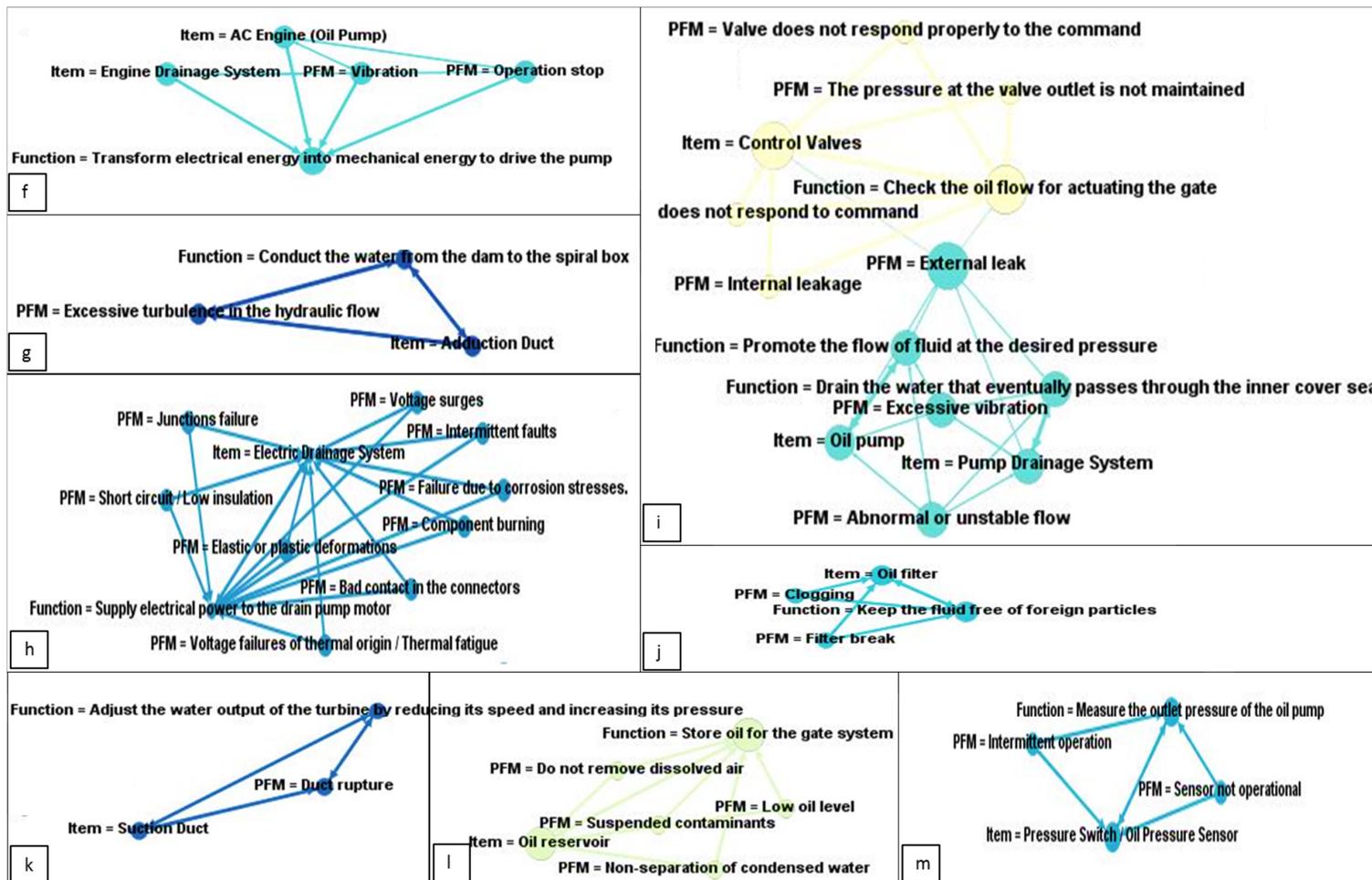


Figure 1. Social Network representing the relationships between Item, Potential Failure Mode (PFM) and Functions of the turbine system: (a–m); represent the thirteen communities of nodes originated from the analysis.

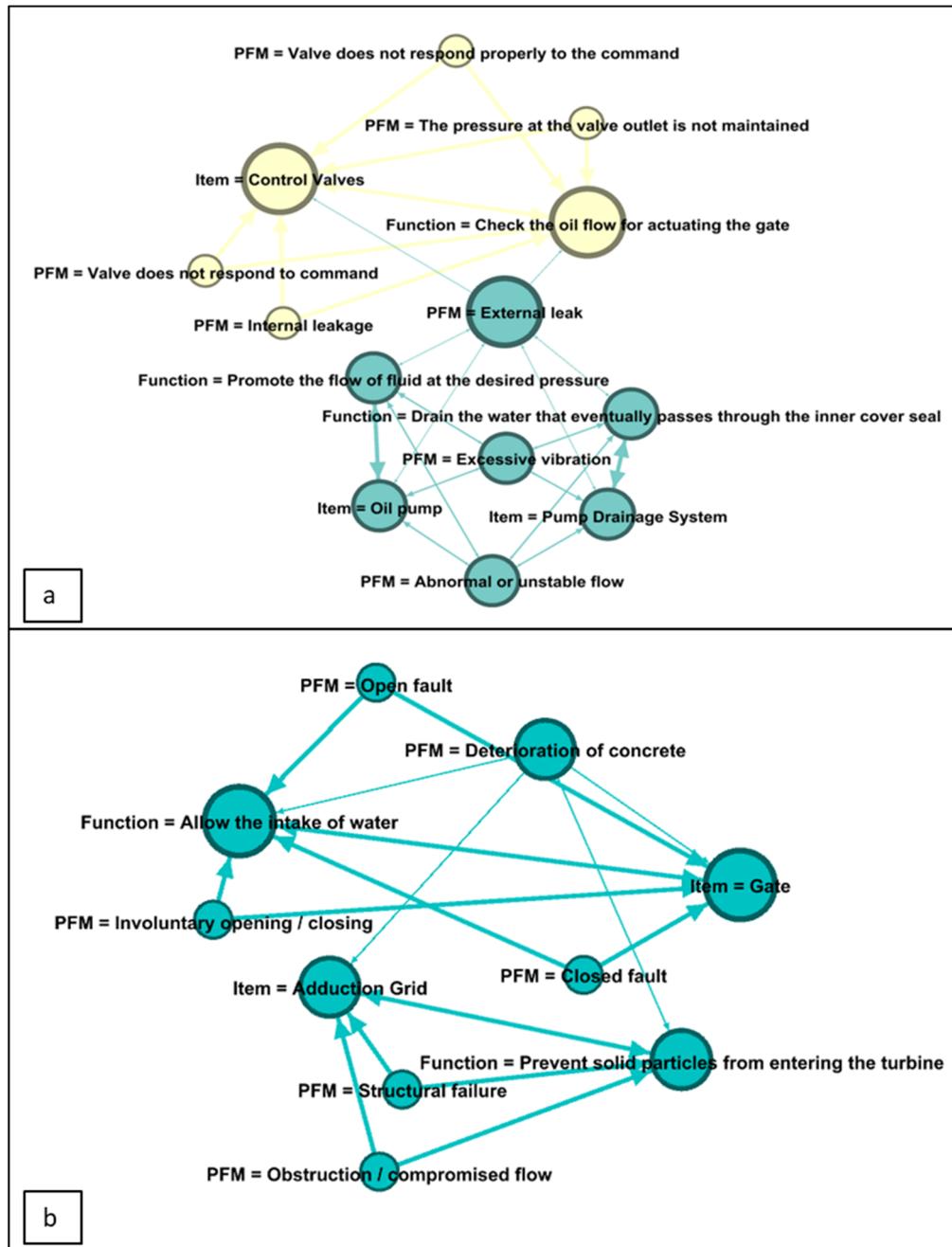


Figure 2. Relationships among potential failure modes, items, and functions. (a) refers to Figure 1i; (b) refers to Figure 1d.

On the contrary, when a malfunctioning on the gate occurs, only in 25% of cases, the failure mode is the deterioration of concrete. Indeed, other PFMs are related to this item, as reported in Figure 2b. The same consideration can be assumed for the adduction grid, but the rule's confidence is 33.3%. At the same time, when the gate experiences a malfunctioning, the compromised function is indeed “Allow intake of water”, as testified by a confidence value 1 of the rule Item = Gate → Function = Allow the intake of water. Accordingly, when the maintenance department members notice a lack in this function, they should immediately check the gate since it is surely damaged.

Table 2. Association Rules (ARs) among the PFM, item, and function of the network's portion reported in Figure 2a.

Left-Hand Side	Right-Hand Side	Support	Confidence
PFM = External leak	Item = Pump drainage system	0.011	0.333
PFM = External leak	Item = Oil pump	0.011	0.333
PFM = External leak	Item = Control valves	0.011	0.333
PFM = External leak	Function = Promote the flow of fluid at the desired pressure	0.011	0.333
PFM = External leak	Function = Drain the water that eventually passes through the inner cover seal	0.011	0.333
PFM = External leak	Function = Check the oil flow for actuating the gate	0.011	0.333
Function = Check the oil flow for actuating the gate	Item = Control valves	0.056	1
Function = Drain the water that eventually passes through the inner cover seal	Item = Pump drainage system	0.033	1
Function = Promote the flow of fluid at the desired pressure	Item = Oil pump	0.033	1

Table 3. Excerpt of the ARs among the PFM, item of the portion of the network reported in Figure 2b.

Left-Hand Side	Right-Hand Side	Support	Confidence
PFM = Deterioration of concrete	Item = Gate	0.011	0.500
PFM = Deterioration of concrete	Item = Adduction grid	0.011	0.500
PFM = Deterioration of concrete	Function = Allow the intake of water	0.011	0.500
PFM = Deterioration of concrete	Function = Prevent solid particles from entering the turbine	0.011	0.500
Item = Gate	PFM = Deterioration of concrete	0.011	0.250
Item = Addduction grid	PFM = Deterioration of concrete	0.011	0.333
Function = Prevent solid particles from entering the turbine	PFM = Deterioration of concrete	0.011	0.333
Function = Allow the intake of water	PFM = Deterioration of concrete	0.011	0.250
Item = Gate	Function = Allow the intake of water	0.044	1.000
Item = Addduction grid	Function = Prevent solid particles from entering the turbine	0.033	1.000
Function = Allow the intake of water	Item = Gate	0.044	1.000
Function = Prevent solid particles from entering the turbine	Item = Addduction grid	0.033	1.000

It is noteworthy to evaluate the impact of a failure on the related items, taking Figure 2a as a reference: the ARs involving the item, the measures of the impacts on people, system availability, and environment are taken into consideration to create the SN reported in Figure 3. According to the experts' opinion, failures on the three items cause low impact at a system availability (Item = Control Valves → System_Availability_Impact = 1; Item = Oil pump → System_Availability_Impact = 1; Item = Pump Drainage System → System_Availability_Impact = 1) in all cases, since the confidence associated with these rules is 100%. At an environmental level, instead, the pump drainage system and the oil pump are associated with a value of 3 on the 1:9 scale, while control valves are less critical (1 out of 9). A score of 3 is assigned to the pump drainage system and the control valves, while the oil pump is less critical. These evaluations support the decision-makers in defining which areas should be monitored first after the occurrence of a malfunctioning, prioritizing the interventions in the area where the impact is higher: referring to Figure 3, for example, people safety is the primary concern (hence the first aspect to be investigated) in case of a failure on control valves, while both people and environment have the priority over the impact on system availability in case of a failure of the pump drainage system. In this way, the areas characterized by a higher risk are controlled and repaired firstly.

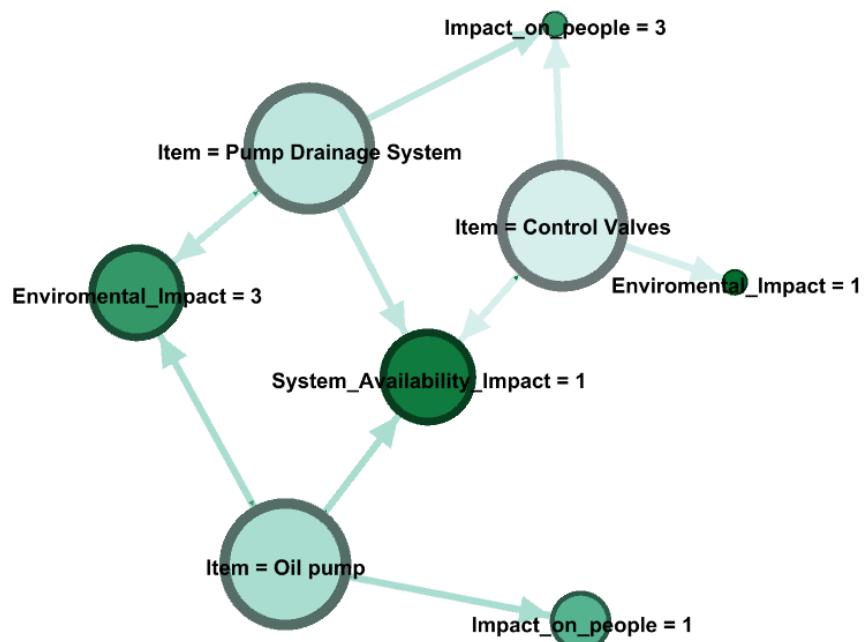


Figure 3. Relationships among items, potential failure modes, impact on safety, environment and system availability.

4. Discussion

The approach proposed in this work aims to extend the failure analysis usually carried out through the FMEA by introducing data-driven techniques. Some theoretical and practical contributions can be extracted from the implementation of the proposed data-driven framework.

4.1. Theoretical Implications

From a theoretical point of view, a comprehensive analysis of large systems' failures can be critical since traditional statistical techniques are not suitable to deal with a large amount of data effectively. Indeed, the ARM implementation allows the definition of the relationships among data, highlighting both the pattern was already known and unknown ones, i.e., those objects of the study. An important feature characterizing the ARM is that there is no need for hypothesis formulation since the whole dataset is explored and the possible connections among items are made [44]. The definition of all the possible itemsets requires the combination of $2k - 1$ items (k being the total number of

items), making the dimension of the dataset a possible critical issue. However, selecting the FP-growth algorithm supports the approach in this sense since the scanning of the dataset is only necessary twice during the whole procedure [45].

The creation of the networks through the SNA, instead, helps in the visualization of all the connection extracted through the ARM and allows the identification of the communities of nodes, facilitating the understanding of the interactions by making them more intuitive. Due to the latter characteristic, besides, it is easier for the analysts to define whether there are missing connections related to the first phases of the failure analysis (e.g., during the deployment of the FMEA). This step is also strategic from a managerial point of view. It allows the definition of whether the analysis of possible failures is performed accurately or if amendments in the process are necessary.

4.2. Practical Implications

Practical implications, instead, regard the possibility of having a closer control of maintenance management. First of all, this approach extends the system's knowledge by using as a starting point the failure analysis usually carried out by the company. In this way, the resources employed are further capitalized without requiring additional investments (i.e., open-source tools are widely available for developing the proposed approach), with a positive impact from an economic point of view. Additionally, the plant on which the analysis is carried out benefits from a complete knowledge of the potential failure modes and a better response to failure occurrence. It is possible to prioritize the interventions based on the impact of the failure itself. Moreover, making the maintenance processes more controllable and predictable harbors benefits from a resource allocation perspective.

From an engineering point of view, the visualization obtained with the SNA is useful to identify the critical chain possibly triggered by the occurrence of a failure mode. In this way, it is easier to understand which items have to be monitored and the areas more hit by the failure. It is also easier to define the items that should be inspected when a failure mode occurs or when malfunctioning is noticed since the network structure presented by the SNA is clear and comprehensive of the whole plant. Specific resources can be destined to maintain and control the critical areas or define interventions for their structural change. In addition, this framework supports the definition of items criticality from a classical risk assessment perspective and considers the unexpected relationships extracted by the ARM and visualized through the SNA.

A further aspect that should be pointed out regards the acceptance of the proposed approach by the plant personnel: indeed, when introducing new methods, there is the risk of resistance to change that can compromise their development. Starting from the FMEA, as presented in the case study, introduces only a partial change in the maintenance management procedures, facilitating the acceptance of the new technology introduction.

5. Conclusions

In this work, a framework for extending the failure analysis through data-driven techniques is proposed. Specifically, the approach offers to consider the failure analysis carried out by the companies (e.g., the Failure Modes Effects Analysis) as a starting point for the application of the Association Rule Mining and Social Network Analysis. The former one aims to identify the co-occurrence of events, such as potential failure modes on specific items, functions compromised, or impacts on the process. Instead, the latter is used for representing these co-occurrences to make them more understandable and intuitive, using a network structure. The failure modes, items and all the attributes analyzed through the Association Rule Mining are the nodes of the network, while the Association Rules are the edges. Using these techniques jointly, it is easier to extend the plant's knowledge and capitalize on the failure analysis's information more extensively. Moreover, the two data-driven techniques enable the exploration of a large amount of data without the need for formulating a research hypothesis.

The proposed approach is applied to the case study of a hydroelectric power plant, using the real failure modes effects analysis as a starting point. The implementation of the policy helps understand

the process, identifying critical nodes of the network that are worthy of monitoring in case of failure occurrence and communities of nodes interacting among them. Indeed, starting from the communities of nodes allows the maintenance managers to define which nodes interact among them and have to be monitored jointly. Thus, support to manage the maintenance process is provided by implementing the data-driven failure analysis, giving the maintenance managers both the chance of enhancing the knowledge of the system and capitalizing on traditional analysis usually carried out.

Future research directions may involve developing further case studies so that the results of different applications in the same research area can be compared and useful insights can be extended to the hydroelectric power plant industry.

Author Contributions: Conceptualization, S.A., M.M.B., F.E.C. and R.F.d.S.; data curation, S.A., M.M.B. and R.F.d.S.; formal analysis, S.A., M.M.B., M.B., F.E.C. and R.F.d.S.; funding acquisition, S.A., M.M.B., F.E.C., R.F.d.S. and G.F.M.d.S.; investigation, S.A., M.M.B., M.B., F.E.C. and R.F.d.S.; methodology, S.A., M.M.B., M.B., F.E.C. and R.F.d.S.; project administration, M.M.B., F.E.C. and G.F.M.d.S.; resources, S.A., M.B., F.E.C., R.F.d.S. and G.F.M.d.S.; software, S.A. and F.E.C.; supervision, M.B., F.E.C. and G.F.M.d.S.; validation, S.A.; writing—original draft, S.A.; writing—review and editing, S.A., M.M.B., M.B., F.E.C., R.F.d.S. and G.F.M.d.S. All authors have read and agreed to the published version of the manuscript.

Funding: The authors thank the financial support of FDTE (Fundação para o Desenvolvimento Tecnológico da Engenharia) and MEC (Brazil's Ministry of Education) for the funding for the development of the present research.

Acknowledgments: G.F.M.d.S., R.F.d.S. and M.M.B. thank EDP Brasil for partially supporting this research through an ANEEL R&D project. These authors thank the financial support of FDTE (Fundação para o Desenvolvimento Tecnológico da Engenharia) and MEC (Brazil's Ministry of Education) for the development of the present research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Perez-Canto, S.; Rubio-Romero, J.C. A model for the preventive maintenance scheduling of power plants including wind farms. *Reliab. Eng. Syst. Saf.* **2013**, *119*, 67–75. [[CrossRef](#)]
2. Allan, G.; Eromenko, I.; Gilmartin, M.; Kockar, I.; McGregor, P. The economics of distributed energy generation: A literature review. *Renew. Sustain. Energy Rev.* **2015**, *42*, 543–556. [[CrossRef](#)]
3. Murad, C.A.; Melani, A.H.D.A.; Michalski, M.A.D.C.; Netto, A.C.; De Souza, G.F.M.; Nabeta, S.I. Fuzzy-FMSA: Evaluating Fault Monitoring and Detection Strategies Based on Failure Mode and Symptom Analysis and Fuzzy Logic. *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part B Mech. Eng.* **2020**, *6*. [[CrossRef](#)]
4. Özcan, E.C.; Ünlüsoy, S.; Eren, T. A combined goal programming–AHP approach supported with TOPSIS for maintenance strategy selection in hydroelectric power plants. *Renew. Sustain. Energy Rev.* **2017**, *78*, 1410–1423. [[CrossRef](#)]
5. Darabnia, B.; Demichela, M. Maintenance an Opportunity for Energy Saving. *Chem. Eng. Trans.* **2013**, *32*, 259–264.
6. Li, R.; Arzaghi, E.; Abbassi, R.; Chen, D.Y.; Li, C.; Li, H.; Xu, B. Dynamic maintenance planning of a hydro-turbine in operational life cycle. *Reliab. Eng. Syst. Saf.* **2020**, *204*, 107129. [[CrossRef](#)]
7. Pisacane, O.; Potena, D.; Antomarioni, S.; Bevilacqua, M.; Ciarapica, F.E.; Diamantini, C. Data-driven predictive maintenance policy based on multi-objective optimization approaches for the component repairing problem. *Eng. Optim.* **2020**, *1*–20. [[CrossRef](#)]
8. Kouigas, I.; Aggidis, G.; Avellan, F.; Deniz, S.; Lundin, U.; Moro, A.; Muntean, S.; Novara, D.; Pérez-Díaz, J.I.; Quaranta, E.; et al. Analysis of emerging technologies in the hydropower sector. *Renew. Sustain. Energy Rev.* **2019**, *113*, 109257. [[CrossRef](#)]
9. Bevilacqua, M.; Braglia, M.; Gabbielli, R. Monte Carlo simulation approach for a modified FMECA in a power plant. *Qual. Reliab. Eng. Int.* **2000**, *16*, 313–324. [[CrossRef](#)]
10. Carpitella, S.; Certa, A.; Izquierdo, J.; La Fata, C.M. A combined multi-criteria approach to support FMECA analyses: A real-world case. *Reliab. Eng. Syst. Saf.* **2018**, *169*, 394–402. [[CrossRef](#)]
11. Grunske, L.; Lindsay, P.; Yatapanage, N.; Winter, K. An Automated Failure Mode and Effect Analysis Based on High-Level Design Specification with Behavior Trees. *Comput. Vis.* **2005**, *3771 LNCS*, 129–149.

12. Ben Said, A.; Shahzad, M.K.; Zamai, E.; Hubac, S.; Tollenaere, M. Experts' knowledge renewal and maintenance actions effectiveness in high-mix low-volume industries, using Bayesian approach. *Cogn. Technol. Work.* **2016**, *18*, 193–213. [[CrossRef](#)]
13. Belinelli, M.; Federal University of Technology; Zattar, I.C.; Da Silva, M.M.; Seleme, R.; De Souza, G.F.; Rodrigues, M.; De Oliveira, C.C.; Savoldi, A. Prioritization of the Industrial Maintenance Activities According its Ergonomics Risks using Multi-criteria Analysis AHP. *Int. J. Eng. Trends Technol.* **2018**, *59*, 214–222. [[CrossRef](#)]
14. Aggogeri, F.; Adamini, R.; Aivaliotis, P.; Borboni, A.; Eytan, A.; Merlo, A.; Németh, I.; Taesi, C.; Pellegrini, N. Robotic System Reliability Analysis and RUL Estimation Using an Iterative Approach. In *Advances in Intelligent Systems and Computing*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2020; Volume 980, pp. 134–143.
15. Ragab, A.; Yacout, S.; Ouali, M.S.; Osman, H. Prognostics of multiple failure modes in rotating machinery using a pattern-based classifier and cumulative incidence functions. *J. Intell. Manuf.* **2016**, *30*, 255–274. [[CrossRef](#)]
16. Bonacina, F.; Corsini, A.; Cardillo, L.; Lucchetta, F. Complex Network Analysis of Photovoltaic Plant Operations and Failure Modes. *Energies* **2019**, *12*, 1995. [[CrossRef](#)]
17. Han, R.; Zhou, Q. Data-driven Solutions for Power System Fault Analysis and Novelty Detection. In Proceedings of the 2016 11th International Conference on Computer Science & Education (ICCSE), Nagoya, Japan, 23–25 August 2016; IEEE: Nagoya, Japan, 2016; pp. 86–91.
18. Shahbaz, M.; Srinivas, M.; Harding, J.A.; Turner, M. Product Design and Manufacturing Process Improvement Using Association Rules. *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.* **2006**, *220*, 243–254. [[CrossRef](#)]
19. Buddhakulsomsiri, J.; Siradeghyan, Y.; Zakarian, A.; Li, X. Association rule-generation algorithm for mining automotive warranty data. *Int. J. Prod. Res.* **2006**, *44*, 2749–2770. [[CrossRef](#)]
20. Ciarapica, F.E.; Bevilacqua, M.; Antomarioni, S. An approach based on association rules and social network analysis for managing environmental risk: A case study from a process industry. *Process. Saf. Environ. Prot.* **2019**, *128*, 50–64. [[CrossRef](#)]
21. Antomarioni, S.; Pisacane, O.; Potena, D.; Bevilacqua, M.; Ciarapica, F.E.; Diamantini, C. A predictive association rule-based maintenance policy to minimize the probability of breakages: Application to an oil refinery. *Int. J. Adv. Manuf. Technol.* **2019**, *105*, 3661–3675. [[CrossRef](#)]
22. Crespo, A.; Carmona, A.D.L.F.; Antomarioni, S. A Process to Implement an Artificial Neural Network and Association Rules Techniques to Improve Asset Performance and Energy Efficiency. *Energies* **2019**, *12*, 3454. [[CrossRef](#)]
23. Savino, M.M.; Brun, A.; Riccio, C. Integrated system for maintenance and safety management through FMECA principles and fuzzy inference engine. *Eur. J. Ind. Eng.* **2011**, *5*, 132. [[CrossRef](#)]
24. Tso, K.S.; Tai, A.T.; Chau, S.N.; Alkalai, L. On Automating Failure Mode Analysis and Enhancing its Integrity. In Proceedings of the 11th Pacific Rim International Symposium on Dependable Computing (PRDC'05), Hunan, China, 12–14 December 2005; IEEE: Hunan, China, 2006; Volume 2005, pp. 287–292.
25. Xu, Z.; Dang, Y.; Munro, P.; Wang, Y. A data-driven approach for constructing the component-failure mode matrix for FMEA. *J. Intell. Manuf.* **2019**, *31*, 249–265. [[CrossRef](#)]
26. Liu, Y.; Hu, X.; Zhang, W. Remaining useful life prediction based on health index similarity. *Reliab. Eng. Syst. Saf.* **2019**, *185*, 502–510. [[CrossRef](#)]
27. Li, X.; Ran, Y.; Zhang, G.; He, Y. A failure mode and risk assessment method based on cloud model. *J. Intell. Manuf.* **2020**, *31*, 1339–1352. [[CrossRef](#)]
28. Lv, J.; Xu, S.; Zhang, R.; Xiao, H.; Chen, Z. Safety Analysis of Metro Turnouts Based on Fuzzy FMECA. In Proceedings of the 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Halifax, NS, Canada, 30 July–3 August 2018; IEEE: Halifax, NS, Canada, 2018; pp. 599–606.
29. Chang, K.H.; Cheng, C.-H. Evaluating the risk of failure using the fuzzy OWA and DEMATEL method. *J. Intell. Manuf.* **2009**, *22*, 113–129. [[CrossRef](#)]
30. Liu, H.C.; Chen, Y.Z.; You, J.; Li, H. Risk evaluation in failure mode and effects analysis using fuzzy digraph and matrix approach. *J. Intell. Manuf.* **2014**, *27*, 805–816. [[CrossRef](#)]
31. Khorshidi, H.A.; Gunawan, I.; Ibrahim, M.Y. Data-Driven System Reliability and Failure Behavior Modeling Using FMECA. *IEEE Trans. Ind. Inform.* **2015**, *12*, 1253–1260. [[CrossRef](#)]

32. Li, F.; Upadhyaya, B.R.; Coffey, L.A. Model-based monitoring and fault diagnosis of fossil power plant process units using Group Method of Data Handling. *ISA Trans.* **2009**, *48*, 213–219. [CrossRef]
33. Byington, C.S.; Matt, W.; Bharadwaj, S.P. Automated Health Management for Gas Turbine Engine Accessory System Components. In Proceedings of the 2008 IEEE Aerospace Conference, Big Sky, MT, USA, 1–8 March 2008; IEEE: Big Sky, MT, USA, 2008; pp. 1–12.
34. Ying-Min, W.; Feng-Bin, Y. Fault Diagnosis of Electric Actuator in the Thermal Power Plant Based on Data-Driven. In Proceedings of the 2009 International Conference on Energy and Environment Technology, Guilin, China, 16–18 October 2009; IEEE: Guilin, China, 2009; pp. 667–670.
35. Ajami, A.; Daneshvar, M. Data driven approach for fault detection and diagnosis of turbine in thermal power plant using Independent Component Analysis (ICA). *Int. J. Electr. Power Energy Syst.* **2012**, *43*, 728–735. [CrossRef]
36. Zhao, Y.; Di Maio, F.; Zio, E.; Dong, C.; Zhang, Q. Optimization of a Dynamic Uncertain Causality Graph (DUCG) for Fault Diagnostics in Nuclear Power Plants by Genetic Algorithm. In Proceedings of the 24th ICONE Conference—International conference on Nuclear Engineering, Houston, TX, USA, 26–30 June 2016; ASME International: Houston, TX, USA, 2016; pp. 1–8.
37. Banjanovic-Mehmedovic, L.; Hajdarevic, A.; Kantardzic, M.; Mehmedovic, F.; Dzananovic, I. Neural network-based data-driven modelling of anomaly detection in thermal power plant. *Autom.* **2017**, *58*, 69–79. [CrossRef]
38. Dhini, A.; Kusumoputro, B.; Surjandari, I. Neural Network Based System for Detecting and Diagnosing Faults in Steam Turbine of Thermal Power Plant. In Proceedings of the 2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST), Taichung, Taiwan, 8–10 November 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 149–154.
39. Jiménez, A.A.; Gómez Muñoz, C.Q.; García Márquez, F.P. Machine learning for wind turbine blades maintenance management. *Energies* **2018**, *11*, 1–16.
40. Hu, C.; Youn, B.D.; Wang, P.; Taek Yoon, J. Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life. *Reliab. Eng. Syst. Saf.* **2012**, *103*, 120–135. [CrossRef]
41. US Military Standard. *MIL-STD-1629a. Procedures for Performing a Failure Mode, Effect and Criticality Analysis*; Department of Defense: Washington, DC, USA, 1980.
42. Hand, D.J. Data Mining: Statistics and More? *Am. Stat.* **1998**, *52*, 112. [CrossRef]
43. Chen, W.C.; Tseng, S.S.; Wang, C.Y. A novel manufacturing defect detection method using association rule mining techniques. *Expert Syst. Appl.* **2005**, *29*, 807–815. [CrossRef]
44. Agrawal, R.; Imieliński, T.; Swami, A. Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, DC, USA, 26–28 May 1993.
45. Han, J.; Pei, J.; Yin, Y. Mining frequent patterns without candidate generation. *SIGMOD* **2000**, *29*, 1–12. [CrossRef]
46. Xiao, S.; Hu, Y.; Han, J.; Zhou, R.; Wen, J. Bayesian Networks-based Association Rules and Knowledge Reuse in Maintenance Decision-Making of Industrial Product-Service Systems. *Procedia CIRP* **2016**, *47*, 198–203. [CrossRef]
47. Otte, E.; Rousseau, R. Social network analysis: A powerful strategy, also for the information sciences. *J. Inf. Sci.* **2002**, *28*, 441–453. [CrossRef]
48. Knoke, D.; Yang, S. *Social Network Analysis*; SAGE Publications: Newbury Park, CA, USA, 2008.
49. Brandes, U. A faster algorithm for betweenness centrality. *J. Math. Sociol.* **2001**, *25*, 163–177. [CrossRef]
50. Scott, J. Social Network Analysis. *Sociology* **1998**, *22*, 109–127. [CrossRef]
51. Tan, P.N.; Kumar, V.; Srivastava, J. Selecting the right objective measure for association analysis. *Inf. Syst.* **2004**, *29*, 293–313. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).