A Novel Approach for Evaluating Datasets Similarities Based on Analytical Hierarchy Process in the Industrial PHM Context

Mohamed Aziz Zaghdoudi[1], Christophe Varnier[2], Sonia Hajri-Gabouj[3], and Noureddine Zerhouni[4]

[1,2,4] *SUPMICROTECH, CNRS, institut FEMTO-ST, F-25000 Besançon, France*
mohamed.zaghdoudi@femto-st.fr
christophe.varnier@ens2m.fr
noureddine.zerhouni@ens2m.fr

[1,3] *LISI, Institut National des Sciences Appliquées et de Technologie,*
Université de Carthage, Centre Urbain Nord, BP 676, 1080, Tunis, Tunisia
sonia.hajri@insat.ucar.tn

Abstract

In prognostics and health management (PHM), data-driven approaches are crucial for performing prognostics based on historical data, relying on the analysis of extensive datasets to identify patterns and relationships that contribute to predicting or optimizing variables. However, their efficiency is contingent upon the availability of large, high-quality datasets tailored to the specific task at hand.
Yet, real-world applications frequently face challenges as data may not always be readily available due to limitations in data acquisition systems or confidentiality concerns. Paradoxically, the contemporary era witnesses an unprecedented surge in the availability of online databases across various fields. These databases offer a plethora of data that can be harnessed to develop, prototype, and test PHM solutions.
This study endeavors to introduce an innovative approach for assessing the similarity between datasets, specifically tailored for prognostic and health management applications. The objective is to empower the development of PHM solutions for predefined systems without relying on data generated from the system itself, but rather by leveraging analogous datasets. To quantify the similarity between different datasets, we propose a set of criteria and sub-criteria based on the characteristics of datasets. Subsequently, the analytic hierarchy process (AHP), a well-established multi-criteria decision-making approach, is employed to systematically compare the importance of criteria and sub-criteria for each elementary process within the PHM cycle. This dynamic process considers the varying importance of criteria across different phases, acknowledging that a criterion may not be uniformly significant for all elementary processes. The evaluation of dataset similarity incorporates the proposed criteria and sub-criteria, utilizing a fundamental scale of importance intensity and weights assigned through AHP. This holistic approach yields a comprehensive similarity score, enabling a nuanced understanding of dataset compatibility.
To exemplify the efficiency of our proposed approach, we applied it to a practical case study. The study involves assessing the similarity between a run-to-stop database of mechanical bearings and a set of online databases dedicated to the same application. Our solution facilitated the identification of criteria pertinent to the case study, the determination of criterion weights, and ultimately, the calculation of a similarity score for each database. This process proved instrumental in selecting the most similar database, showcasing the practical utility of our proposed approach in real-world PHM scenarios.

1. Introduction

Prognostics and Health Management (PHM) is an engineering and research field that aims to study fielded systems conditions, predict their possible failures, and take appropriate actions to mitigate those malfunctions effects (Bougacha, Varnier, & Zerhouni, 2022). In this context, data-driven approaches are being increasingly used to convert historical data into models that accurately represent the physical systems' degradation behavior (Tobon-Mejia, Medjaher, Zerhouni, & Tripot, 2012). To perform efficiently, those approaches require the presence of extensive datasets, adhering to established data quality standards, and accurately reflecting the characteristics of the system under study. However, for real systems, data collection is a complicated process that requires setting up sometimes costly acquisition devices, overcoming confidentiality issues, and selecting the characteristics of the data to be collected (data format, relevant variables, data quality requirements...). This has led to a problem of insufficient amount of data for some PHM applications and uncertainty regarding the characteristics of the data to be collected.
Conversely, the current era is experiencing a proliferation in

both the quantity and diversity of online databases, with approximately 31 million databases accessible on the Internet as of August 2020 (Benjelloun, Chen, & Noy, 2020). These publicly accessible datasets span a broad spectrum of domains, encompassing around 4600 domains in August 2020 (Benjelloun et al., 2020), and are amenable to adaptation for analogous problem-solving scenarios.

This theme has motivated this research work. We are interested in finding an approach for datasets similarity evaluation that makes it possible to find, among freely accessible datasets, the most similar dataset to a sample of data from a system studied in order to overcome the problem of lack of data for PHM applications.

In pursuit of this goal, we have introduced a set of criteria grounded in data characteristics to assess the similarity between datasets. Subsequently, we presented a methodology employing the Analytical Hierarchy Process (AHP), a widely recognized multi-criteria decision-making technique. This methodology serves to determine criteria weights and evaluate datasets similarity on the base of those criteria.

The remainder of this paper is organized into four sections. Section 2 summarizes previous works related to data insufficiency, data characterization, and the AHP technique. Section 3 describes the proposed methodology. Section 4 presents an illustrative case study evaluating the similarity between different bearing datasets. In section 5, a reliability evaluation approach is proposed to assess the consistency of the results. Finally, section 6 summarizes the main findings and outlines future directions for research.

2. RELATED WORK

2.1. Solving the data insufficiency problem

The data insufficiency problem was the subject of several research works. Indeed, (Guo, Lei, Xing, Yan, & Li, 2018) require the existence of two conditions for the success of machine diagnosis data-driven intelligent approaches : Labeled data containing fault information is available and training and test data are drawn from the same probability distribution. However, for some systems, it is difficult to obtain massive labeled data (Guo et al., 2018).

One of the solutions proposed in the literature is Transfer Learning. It is defined as follows: Given a source domain $\mathrm{D_S}$ with a corresponding source task $\mathrm{T_S}$ and a target domain $\mathrm{D_T}$ with a corresponding task $\mathrm{T_T}$, transfer learning is the process of improving the target predictive function $\mathrm{f_T}(\cdot)$ using related information from $\mathrm{D_S}$ and $\mathrm{T_S}$, where $\mathrm{D_S} \neq \mathrm{D_T}$ or $\mathrm{T_S} \neq \mathrm{T_T}$ (Weiss, Khoshgoftaar, & Wang, 2016).

The transfer learning approach has been applied to several industrial systems. (Wen, Gao, & Li, 2017) applied deep transfer learning method for fault diagnosis in a big data environment. Their approach was tested on a Case Western Reserve University bearing dataset (Smith & Randall, 2015). (Shao, McAleer, Yan, & Baldi, 2018) developed a deep transfer learning framework for mechanical fault diagnosis and classification, and created a repository of several reference datasets.

Despite its ability to solve the data gap problem, the transfer learning technique requires that the source and target data are similar and of the same distribution.

Another widely used approach is data augmentation. This technique consists in increasing the amount of training data by using the information contained within it (Perez & Wang, 2017).

Various data augmentation techniques have been applied to specific problems. The main techniques fall under the category of data warping, which is an approach to directly augment the input data to the model in the data space. This technique has been applied for several industrial applications and on various types of data. (Li, Zhang, Ding, & Sun, 2020) employed it for fault diagnosis of rotating machines. They applied 5 techniques for data augmentation in the form of digital signals, namely, Gaussian noise, masking noise, signal translation, amplitude shift, and time stretching.

Moreover, this technique is widely used with image data. As an example, we cite the work of (Wang, Yang, Jiang, & Fan, 2020) on image augmentation for crack detection using 9 different techniques.

Certainly, the data augmentation technique is useful to overcome the problem of lack of data for different applications and data types. However, this approach requires the existence of a minimal amount of data to be augmented.

On the other hand, other alternatives are used by researchers and industrialists to generate artificial data, such as physical model-based simulation (Saxena, Goebel, Simon, & Eklund, 2008) or test bench fabrication (Nectoux et al., 2012).

2.2. Analytical Hierarchy Process

The Analytical Hierarchy Process (AHP) was developed by Saaty in the 1970s (Saaty, 1980). This method, used in many fields related to multiple criteria decision-making (MCDM) is considered one of the most useful decision-making techniques (Ahmadi, Arasteh Khouy, Kumar, & Schunnesson, 2009). It's a methodology for relative measurement (Brunelli, 2014) where the focus is on proportions between some quantities rather than their exact measurement.

In AHP, The problem is divided into a hierarchy of qualitative and quantitative criteria, and then, using experience, the degree of relative importance is deducted. According to (Nydick & Hill, 1992), the AHP method is based on 4 steps :

1. Problem structuring
2. Data collection and measurement
3. Normalized weights determination
4. Application and problem-solution-finding

The Analytical Hierarchy Process has been used in several industrial applications to make decisions in different areas.

(Cabrita & Frade, 2016) proposed an AHP-based solution to the supplier selection problem using fourteen different criteria. (Ren & Lützen, 2015) used AHP for fuel evaluation and selection under nine criteria for emission reduction from shipping. (Kilic, Zaim, & Delen, 2014) evaluated and selected the best ERP system using an AHP-based solution to solve this MCDM problem.
Hence, the analytical hierarchy process can be considered as a strong decision-making tool that can be used to evaluate and select the best action/alternative in multiple criteria decision-making problems.

2.3. Data Characterization

Databases similarity assessment first requires the establishment of data characterization criteria. Several previous works have addressed the issue of database characterization. However, the definitions and criteria proposed differ from one work to another, and the research has not resulted in unified criteria.
In this context, (Alelyani, Liu, & Wang, 2011) proposed 4 characteristics and studied their effects on feature selection stability. The proposed characteristics are the number of samples, features and classes, and the data distribution. (Bhatt, Thakkar, & Ganatra, 2012) divided thirteen characterization criteria into 2 different groups: phenotype characteristics dealing with entropy and the noise-signal ratio, and characteristics concerning the genotype of a dataset, divided into 2 categories:

- Simple Characteristics concern the attributes and instances' numbers
- Statistical Characteristics that deal with the statistical aspect of data.

(Oreski, Oreski, & Klicek, 2017) characterized data by 11 characteristics in 5 different groups, consisting mainly of standard, data sparsity, statistical, information-theoretic, and noise measures.
On the other hand, data quality has emerged as a fundamental notion for characterizing data. (Strong, Lee, & Wang, 1997) have defined high-quality data as data that is suitable for data consumers. Thus, we can conclude that data with different degrees of quality will lead to different results. (Redman, 1997) proposed four data quality characteristics most studied in the literature: accuracy, consistency, completeness, and timeliness. (Omri, Al Masry, Mairot, Giampiccolo, & Zerhouni, 2021) suggest that for PHM applications, data quality is characterized by volume, accuracy and completeness.

3. PROPOSED APPROACH

The proposed methodology (Fig. 1) is composed of four different phases. The first phase includes the proposal of similarity criteria and sub-criteria. The second phase is linked to the PHM cycle and the processes that make it up. The third phase details the criteria and sub-criteria weights calculation using AHP technique. The final phase is dedicated to decision-making using the established methodology.

3.1. Problem modeling / Criteria setting

The first step consists of proposing similarity criteria according to which the similarity will be evaluated. This step is also called 'Problem modeling' for AHP applications (Ishizaka & Labib, 2011). In fact, it is recommended to structure the criteria in a hierarchical structure to be able to focus on their importance when assigning their weights (Ishizaka & Labib, 2011). A structure of sub-criteria assembled in clusters (criteria) helps describe the problem more conveniently and reduces bias (Ishizaka, 2004).
To define criteria that are in line with this problem, we mainly rely on the data characterization criteria proposed in the literature. In (Table 1), a non-inclusive list of 17 sub-criteria divided into four criteria is proposed to evaluate the similarity between databases. These criteria can be used fully or partially, depending on the application or case study under consideration.
In addition to the attributes outlined in existing literature, we have introduced two supplementary sub-criteria, namely 'Data extension' and 'Data format.' Specifically, within the context of a given system and application, data representing the system state may manifest in various types and formats, such as images, signals, or tabular data. Disparities in data format and extension necessitate distinct characterizations and treatments.
Furthermore, our research proposes a novel set of application-related criteria, consisting of two sub-criteria. These criteria aim to evaluate the domain (e.g., manufacturing, medical, transportation) of the system depicted in the dataset, along with discerning the data source—whether it originates from a real-world application, a simulation, or a test bench.

3.2. PHM cycle modeling

In order to assign weights to each similarity criterion, we propose to, firstly, divide the studied PHM cycle into elementary processes. In fact, the PHM cycle is composed of seven elementary processes according to (Omri, Al Masry, Mairot, Giampiccolo, & Zerhouni, 2020), namely data acquisition, data processing, data assessment, diagnostic, prognostics, decision support, and HMI. From data acquisition to decision support and HMI, the importance of each of the established criteria depends on the process.
For example, the data distribution a negligible impact on the data acquisition process. However, this characteristic is very important in the data processing and exploitation processes (diagnostic and prognostic). Thus, the importance of each of the criteria will be judged with respect to every PHM process separately.

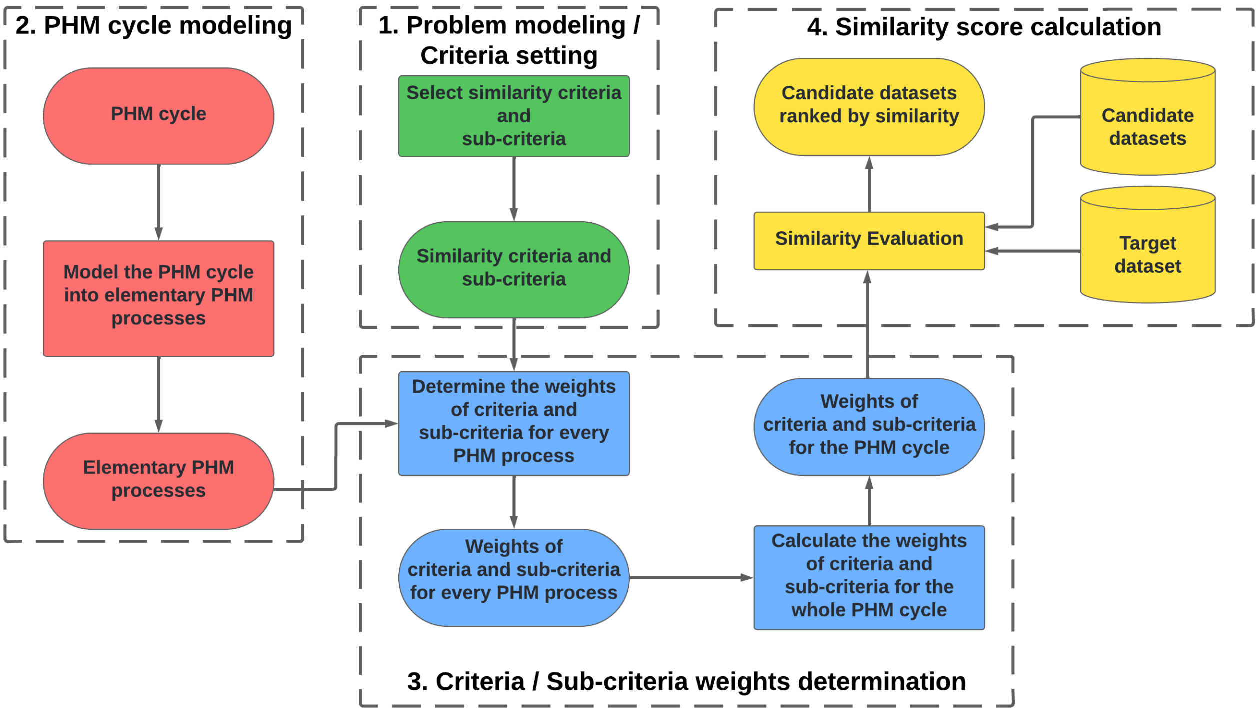

Figure 1. AHP based approach for datasets similarity evaluation

Table 1. Data similarity criteria

Criteria	Sub-criteria
Standard criteria	1. Number of attributes (Alelyani et al., 2011; Bhatt et al., 2012; Oreski et al., 2017) 2. Number of instances (Alelyani et al., 2011; Bhatt et al., 2012; Oreski et al., 2017) 3. Number of classes (Alelyani et al., 2011; Bhatt et al., 2012; Oreski et al., 2017) 4. Number of binary features (Bhatt et al., 2012) 5. Data format 6. Data extension
Statistical criteria	7. Data distribution (Alelyani et al., 2011) 8. Features correlation (Bhatt et al., 2012; Oreski et al., 2017) 9. Multivariate normality (Oreski et al., 2017) 10. Mean Kurtosis of attributes (Bhatt et al., 2012) 11. Mean skewness of attributes (Bhatt et al., 2012)
Data quality criteria	12. Accuracy (Omri et al., 2021; Redman, 1997) 13. Completeness (Omri et al., 2021; Oreski et al., 2017) 14. Consistency (Redman, 1997) 15. Timeliness (Redman, 1997)
Application related criteria	16. Field of application 17. Data source

3.3. Criteria / Sub-criteria weights determination

Notation:

- P_i : Elementary process i (i=1,..., L)
- D_h : Similar dataset h (h=1,..., Q)
- Cj : Criterion j (j=1,..., N)
- $X_{j,i}$: Weight of criterion j for process i
- SCk : Sub-criterion k (k=1,..., M)
- $Y_{k,i}$: Weight of sub-criterion k for process i
- W_k : Weight of sub-criterion k
- M_j : Number of sub-criterion related to the criterion j
- $Z_{h,k}$: Similarity score of the candidate dataset h with the target dataset with respect to the sub-criterion k
- R_h : Similarity score of the candidate dataset h with the target dataset.

In the AHP technique, a ratio scale is used to derive, two by two, the criteria's and sub-criteria's importance. This comparison, unlike techniques that use interval scales, requires no units (Ishizaka & Labib, 2011) and assures a more accurate decision than comparing all the criteria at once.
The pairwise comparison of criteria, and every group of sub-criteria, is realized using Saaty's 1-9 scale for pairwise comparison (Saaty, 2005) described in Table 2.

Table 2. Saaty's 1-9 scale for pairwise comparison

Intensity of importance	Definition
1	Equal importance
3	Moderate importance
5	Strong importance
7	Very strong importance
9	Extreme importance

To determine the weights of N criteria for the elemental process Pi, An NxN matrix is created, where $a_{j1,j2}$ describes the importance of criterion $Cj1$ over criterion $Cj2$. Therefore, for all j1 and j2, $a_{j1,j2}$ is the inverse of $a_{j2,j1}$ and $a_{j1,j1} = 1$.

$$\begin{bmatrix} 1 & a_{1,2} & .. & a_{1,N} \\ a_{2,1} & 1 & .. & a_{2,N} \\ .. & .. & .. & .. \\ a_{N,1} & .. & a_{N,N-1} & 1 \end{bmatrix} \tag{1}$$

This procedure is carried out to deduce the relative importance of the criteria by comparing them two by two using the fundamental scale of importance intensity. The weight $X_{j,i}$ of criterion Cj in relation to the process Pi is calculated using equation 2.

$$X_{j,i} = \frac{\sum_{j2=1}^{N} \left(\frac{a_{j1,j2}}{\sum_{j1=1}^{N} a_{j1,j2}} \right)}{N} \tag{2}$$

Similarly, the sub-criteria relating to each criterion are compared two by two, and the weight of each sub-criterion in relation to the Pi process is calculated using equation 4

$$\begin{bmatrix} 1 & b_{1,2} & .. & b_{1,Mj} \\ b_{2,1} & 1 & .. & b_{2,Mj} \\ .. & .. & .. & .. \\ b_{Mj,1} & .. & b_{Mj,Mj-1} & 1 \end{bmatrix} \tag{3}$$

$$Y_{k,i} = \frac{\sum_{k2=1}^{Mj} \left(\frac{b_{k1,k2}}{\sum_{k1=1}^{Mj} b_{k1,k2}} \right)}{Mj} \times X_{j,i} \tag{4}$$

At the end of this procedure, the weight of each criterion/sub-criterion is given, showing their importance for each elementary process of the PHM cycle.
In order to deduce the weight of a sub-criterion for the whole cycle, an average of these weights is calculated (equation 5).

$$W_k = \frac{\sum_{i=1}^{L} Y_{k,i}}{L} \tag{5}$$

3.4. Similarity score calculation

In this final step, the similarity $Z_{h,k}$ of every candidate dataset D_h with the studied dataset regarding each sub-criterion k is evaluated. The assessment is done using the fundamental scale of importance intensity (Table 2).
For quantitative criteria, an odd number between 1 and 9 is assigned, depending on the decision-maker's expertise. On the other hand, for qualitative criteria, only two possible values can be given, 9 for two data sets with similar attributes and 1 otherwise.
Finally, a normalized similarity score of each candidate dataset R_h is calculated using equation 6. The higher the similarity score, the more the concerned dataset is similar to the target dataset. A similarity score of 1 means that the two compared datasets have identical characteristics.

$$R_h = \frac{\sum_{k=1}^{M} Z_{h,k} \times W_k}{9} \tag{6}$$

4. ILLUSTRATIVE CASE STUDY

The proposed database similarity assessment methodology will be applied to a case study of bearing failure databases available online.
A bearing is a machine component that lessens friction between moving elements in mechanical engineering. It is frequently used in wheels or axles to support and guide a rotating or oscillating shaft. Bearings can be subject to various failures, manifested by cracks, wear marks, chips, and abnormal noises. These failures can significantly affect the mechanical and energy sectors' capacity to operate, level of safety, and financial aspect (Nectoux et al., 2012).
In the context of PHM applications for bearing condition prognosis, (Nectoux et al., 2012) provided a database for the IEEE PHM 2012 Prognostic Challenge. The experiments were carried out on the PRONOSTIA platform at the Femto-ST Institute, and the results present 9 features relating to run-to-failure tests of 17 bearings.

4.1. Proposed criteria and PHM cycle modeling

For the application under consideration, based on the criteria summary table (Table 1), eleven sub-criteria split among three criteria were proposed. The sub-criteria relating to the standard criteria were retained, except for the number of classes. This selection is justified by the studied databases, which were not originally designed for classification purposes and lack class labels. In addition, the application-related criteria were also retained with the proposal of two additional criteria specific to this application, namely the number of operating conditions applied and the number of tested bearings. Moreover, two of the data quality sub-criteria were used in this case study. The completeness was evaluated as the ratio of non-empty cells over all available cells, and the accuracy was assessed as the presence or absence of noise.

For this application, the PHM cycle was simplified to 3 elementary processes, namely the data acquisition, the data preprocessing, and the prognostics processes.

4.2. Similar databases collection

A collection of four databases, available online, for the same applications, was carried out.

The first dataset (Kaggle, 2023) is provided by Quantum company in collaboration with Kharkiv Polytechnic Institute. It consists of 3-axis vibration measurements of 112 rotating bearings.

The second dataset (Qiu, Lee, Lin, & Yu, 2006) is a run-to-failure dataset of four bearings under one operating condition, provided by Qiu et al. Eight features related to the vibration and the temperature of the bearings were collected to study their health state.

The third data set (CWRU, .) is provided by Case Western Reserve University and presents ten statistical features related to measurements of 21 bearings under fixed operating conditions that manifested ten possible types of faults.

Finally, the fourth database presents recordings of the acceleration of a high-speed bearing used for wind turbines over 30 days (6 seconds daily). These recordings were made under two operating conditions.

Table 3 details the selected dataset characteristics in relation to the criteria and sub-criteria chosen for the study.

4.3. Criteria and sub-criteria weights calculation

As mentioned in section 3, and in order to determine the sub-criteria weights, a pairwise comparison of the importance of the criteria for every elementary PHM process was performed.

Table 4 details the process of comparing importance and calculating criteria weights for the data acquisition process.

The criteria weights were calculated using equation 2, after constructing the comparison matrix. The application criterion contributes the most to selecting a similar dataset for the data acquisition process. In addition, a similarity in the application criterion is strongly preferred to the quality criterion. In fact, a different application may require another data acquisition system. Moreover, as seen in Table 4, no two criteria are of equal importance for the acquisition process.

Table 5 compares the criteria importance and weights in the data preprocessing process. In contrast to the data acquisition process, the application criterion has an insignificant weight compared to standard and quality criteria, indicating a lower priority in this context. Conversely, the quality criterion holds the highest significance in selecting an appropriate dataset in the preprocessing process, holding nine times more importance than the application criterion and three times more significance than the standard criterion. These findings align with expectations, as the preprocessing process rarely depends on applications and focuses mainly on data quality and standard characteristics.

The weights of each family of sub-criteria were then determined for each elementary process of the PHM cycle. This is done by comparing them two by two using the 1-9 scale for pairwise comparison and then, by applying equation 4 to incorporate the weights of the associated criteria.

Table 6 shows the weights of the standard sub-criteria for the data acquisition process. The number of features is found to be the most important sub-criterion to assess the similarity between two datasets concerning the data acquisition process. In fact, features (variables) are collected using acquisition devices like sensors. These devices are costly and require studies to set them up and to ensure data acquisition. This sub-criterion is therefore the most important for this PHM process. The number of features sub-criterion is considered to be very strongly important than the number of instances, extremely important than the data extension sub-criterion, and moderately important than the data format sub-criterion.

The data format is the second most important standard sub-criterion to assess datasets similarity in relation to the data acquisition process. It is strongly more important than the number of instances and the number of binary features, and moderately more important than the data extension sub-criterion.

The Standard sub-criteria importance and weights for the prognostic process are described in Table 7. Similarly to the acquisition process, the number of features criterion is the most important factor in determining databases' similarity in relation to the prognostic process. Additionally, the data extension is the least impacting factor in both processes. The second most important standard sub-criterion is the number of binary features in this context. It is moderately more important than the number of instances and data format criterion and highly more important than the data extension criterion.

The final weights of all the considered sub-criteria for the whole PHM cycle are detailed in Table 8.

The accuracy sub-criterion is found to be the most important. Since the scores are normalized, then a weight of 0,3603 means that this sub-criterion contributes by 36,03% to the final decision about datasets similarity. The following sub-criteria are the number of tested bearings, the number of operating conditions, and the number of collected features. These four sub-criteria contribute by more than 75% to the final decision.

Table 3. Collected datasets characteristics

	Target dataset	Candidate Dataset 1	Candidate Dataset 2	Candidate Dataset 3	Candidate Dataset 4
Number of attributes	7	13	8	10	2
Number of instances	18196480	10265700	4415488	2048	29,286,800
Number of binary features	0	0	0	0	0
Data format	tabular	tabular	text	tabular	binary data container
Data extension	.csv	.csv	text	.csv	.mat
Field of application	Academic	Industrial	Academic	Industrial	energy industry
Data source	test bench	test bench	test bench	test bench	real life
Number of operating conditions	3	3	1	1	2
Number of bearings tested	17	112	4	21	1
Completeness	100,00 %	100,00 %	100,00 %	100,00 %	100,00 %
Accuracy	noised	X	noised	X	noised

Table 4. Criteria matrix and weights for the data acquisition process

	Standard	Application	Quality	**Criteria weights**
Standard	1	1/3	3	**0,2605**
Application	3	1	5	**0,6333**
Quality	1/3	1/5	1	**0,1062**

Table 5. Criteria matrix and weights for the preprocessing process

	Standard	Application	Quality	**Criteria weights**
Standard	1	7	1/3	**0,2946**
Application	1/7	1	1/9	**0,0567**
Quality	3	9	1	**0,6486**

4.4. Similarity score calculation and decision-making

In this final step, the similarity of every candidate dataset with the target dataset is evaluated with respect to every sub-criterion using the fundamental scale of importance intensity (Table 2).

Similarity based on qualitative criteria is assessed using the Saaty scale. In other words, if both datasets have the same attribute, a rating of 9 is assigned; otherwise, a rating of 1 is assigned.

For example, for the data format sub-criterion, two candidate datasets are of the same format as the target dataset, so they got a similarity score of 9. The other two datasets are of different formats (text and binary data container), leading to a weak similarity score of 1.

A score of similarity, according to every sub-criterion, between each candidate dataset and the target dataset is given. Afterward, the weights deducted in the previous step are used to get a similarity score for every candidate dataset 8.

The second dataset (Qiu et al., 2006) is found to be the most similar dataset to the target dataset (Nectoux et al., 2012) with a similarity score of 0,7143. However, the first candidate dataset (Kaggle, 2023) is the least similar dataset to the target dataset. This is mainly caused by the difference in the accuracy sub-criterion, the number of tested bearings and the number of features. These sub-criterion were found to be three of the four most important comparison sub-criteria. Therefore, a low score in these attributes leads to a weak overall similarity score.

If a simple normalized mean of the similarity scores is calculated, the first candidate dataset will obtain a higher score of 0,7172, meaning that it is the most similar dataset. This shows the importance of assigning weights to the comparison sub-criteria.

It is important to note that, although the fourth dataset is the only one originating from the real world, it was not selected. This decision stems from the fact that the 'data source' criterion is just one of several simulation criteria used in the selection process. Moreover, the target dataset itself is derived from a simulation, rendering dataset number 4 dissimilar in terms of data source. Our aim is to select the dataset that most closely resembles the target dataset, rather than simply identifying the best dataset.

5. DECISION RELIABILITY

The methodology outlined in this study hinges upon conducting pairwise comparisons of both criteria and sub-criteria to ascertain their respective weights. These comparisons are based on subjective judgments provided by the decision-maker. Consequently, it becomes important to assess the consistency of these judgments. Consistency, within the context of the Analytic Hierarchy Process (AHP), denotes the extent to which the pairwise comparisons rendered by decision-makers exhibit logical coherence and absence of contradictions. Inconsistencies in judgments bear the risk of yielding unreliable weight assignments, thereby potentially skewing the subsequent similarity evaluations.

Several works have addressed the problem of consistency of AHP matrices. One way to deal with this is by determining the Consistency Ratio (CR) (Pant, Kumar, Ram, Klochkov, & Sharma, 2022; Franek & Kresta, 2014). First, the Consis-

Table 6. Standard sub-criteria matrix and weights for the data acquisition process

	Number of instances	Number of features	Number of binary features	Data format	Data extension	**Criteria weights**
Number of instances	1	1/7	1/3	1/5	3	**0,0740**
Number of features	7	1	5	3	9	**0,5048**
Number of binary features	3	1/5	1	1/5	3	**0,1163**
Data format	5	1/3	5	1	3	**0,2581**
Data extension	1/3	1/9	1/3	1/3	1	**0,0468**

Table 7. Standard sub-criteria matrix and weights for the prognostic process

	Number of instances	Number of features	Number of binary features	Data format	Data extension	**Criteria weights**
Number of instances	1	1/5	1/3	3	7	**0,1449**
Number of features	5	1	3	7	9	**0,4992**
Number of binary features	3	1/3	1	3	7	**0,2298**
Data format	1/3	1/7	1/3	1	7	**0,0962**
Data extension	1/7	1/9	1/7	1/7	1	**0,0299**

Table 8. Final weights and similarly scores of the candidate datasets

Sub-criteria	Sub-criteria weights	**Candidate dataset 1**	**Candidate dataset 2**	**Candidate dataset 3**	**Candidate dataset 4**
Number of instances	**0,0156**	5	3	1	5
Number of features	**0,1008**	3	7	5	1
Number of binary features	**0,0224**	9	9	9	9
Data format	**0,0584**	9	1	9	1
Data extension	**0,0177**	9	5	9	1
Accuracy	**0,3603**	1	9	1	9
Completeness	**0,0721**	9	9	9	9
Field of application	**0,0354**	5	9	5	3
Data source	**0,0276**	9	9	9	5
Number of operating conditions	**0,1229**	9	3	3	7
Number of bearings	**0,1668**	3	3	7	1
Similarity score		**0,4787**	**0,7143**	**0,4863**	**0,6244**

tency Index (CI) is computed according to equation 7:

$$CI = \frac{\lambda_{max} - N}{N - 1} \tag{7}$$

with λ_{max} representing the largest eigenvalue of the pairwise comparison matrix and N indicating the matrix size (number of criteria or sub-criteria).

Using pre-defined tables (Table 9), the Random Index (RI) corresponding to the matrix size is determined. The Consistency Ratio (CR) is then calculated by dividing CI by RI. A CR value below 0.1 signifies acceptable consistency in judgments, while values exceeding 0,1 may indicate potential inconsistencies requiring further scrutiny or adjustment.

As an example, the consistency ratio of the criteria pairwise comparison matrix is 0,03 for the data acquisition process and 0,07 for the preprocessing process. These results demonstrate that the weights of the resulting criteria are consistent and can be used to reliably determine criteria weights.

On the other hand, the consistency ratio of the standard sub-criteria matrix for the prognostic process is 0,11 meaning that the comparison need to be adjusted in order to get a consistent judgment of the sub-criteria weights.

6. CONCLUSION

In this work, a database comparison approach was proposed to find a solution to the problem of lack of data for PHM applications. Indeed, for this field of study, and in order to develop a data-driven PHM solution, datasets need to be available, containing all the variables describing the system under study and complying with quality standards. In reality, this is not always the case.

Therefore, we have proposed an approach for assessing the similarity between a target dataset and a set of datasets available online. A set of criteria has been proposed, based on data characteristics. As the criteria are not equally important for judging similarity, a weight for each criterion is determined using the analytical hierarchy process. The similarity of the datasets is then scored against each criterion, and a normalized score is calculated for each dataset.

The proposed approach has been applied to an illustrative

Table 9. Random index for the AHP consistency ratio (Saaty, 1980)

Number of rows	1	2	3	4	5	6	7	8	9
RI	0	0	0,58	0,90	1,12	1,24	1,32	1,41	1,45

case, where the similarity of four datasets with a bearing operating database has been evaluated. The application leads to calculating similarity scores for each dataset and selecting the most similar one.

This work presents a first step towards solving the problem of lack of data for PHM applications. It makes it possible to design a PHM solution for a given system without the need to use data directly from that system.

On the other hand, this proposal is limited by the subjectivity of the decision-maker. The latter, responsible for rating similarity and judging the importance of criteria, may be biased and lead to subjective decisions. We therefore recommend that weights and scores be allocated by several experts at the same time, in order to limit the subjectivity of decision-makers.

In addition, considering the limitations of our current methodology, future studies may employ fuzzy techniques to reduce the uncertainty of the decision. Furthermore, in this work, we proposed a non-exhaustive list of criteria, other criteria can also be used, namely the maturity of the data for example, which leads to the generalization of the approach to various fields and applications.

REFERENCES

Ahmadi, A., Arasteh Khouy, I., Kumar, U., & Schunnesson, H. (2009). Selection of maintenance strategy, using analytical hierarchy process. *Communications in Dependability and Quality Management*, *12*(1), 121–132.

Alelyani, S., Liu, H., & Wang, L. (2011). The effect of the characteristics of the dataset on the selection stability. In *2011 ieee 23rd international conference on tools with artificial intelligence* (pp. 970–977).

Benjelloun, O., Chen, S., & Noy, N. (2020). Google dataset search by the numbers. In *International semantic web conference* (pp. 667–682).

Bhatt, N., Thakkar, A., & Ganatra, A. (2012). A survey & current research challenges in meta learning approaches based on dataset characteristics. *International Journal of soft computing and Engineering*, *2*(10), 234–247.

Bougacha, O., Varnier, C., & Zerhouni, N. (2022). Impact of decision horizon on post-prognostics maintenance and missions scheduling: a railways case study. *International Journal of Rail Transportation*, *10*(4), 516–546.

Brunelli, M. (2014). *Introduction to the analytic hierarchy process*. Springer.

Cabrita, M. D. R., & Frade, R. (2016). Supplier selection approach: integrating analytic hierarchy process and supplier risk analysis. *International Journal of Business and Systems Research*, *10*(2-4), 238–261.

CWRU. (.). *Case western reserve university bearing data center dataset.* `https://engineering.case.edu/bearingdatacenter`. (Accessed on March 12, 2024)

Franek, J., & Kresta, A. (2014). Judgment scales and consistency measure in ahp. *Procedia economics and finance*, *12*, 164–173.

Guo, L., Lei, Y., Xing, S., Yan, T., & Li, N. (2018). Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data. *IEEE Transactions on Industrial Electronics*, *66*(9), 7316–7325.

Ishizaka, A. (2004). The advantages of clusters in ahp. In *The 15th mini-euro conference, mudsm.*

Ishizaka, A., & Labib, A. (2011). Review of the main developments in the analytic hierarchy process. *Expert systems with applications*, *38*(11), 14336–14345.

Kaggle. (2023). *Bearing classification dataset.* `www.kaggle.com/datasets/isaienkov/bearing-classification`. (Accessed on March 12, 2024)

Kilic, H. S., Zaim, S., & Delen, D. (2014). Development of a hybrid methodology for erp system selection: The case of turkish airlines. *Decision Support Systems*, *66*, 82–92.

Li, X., Zhang, W., Ding, Q., & Sun, J.-Q. (2020). Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation. *Journal of Intelligent Manufacturing*, *31*(2), 433–452.

Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-Morello, B., Zerhouni, N., & Varnier, C. (2012). Pronostia: An experimental platform for bearings accelerated degradation tests. In *Ieee international conference on prognostics and health management, phm'12.* (pp. 1–8).

Nydick, R. L., & Hill, R. P. (1992). Using the analytic hierarchy process to structure the supplier selection procedure. *International Journal of purchasing and materials management*, *28*(2), 31–36.

Omri, N., Al Masry, Z., Mairot, N., Giampiccolo, S., & Zerhouni, N. (2020). Industrial data management strategy towards an sme-oriented phm. *Journal of Manufacturing Systems*, *56*, 23–36.

Omri, N., Al Masry, Z., Mairot, N., Giampiccolo, S., & Zerhouni, N. (2021). Towards an adapted phm approach:

Data quality requirements methodology for fault detection applications. *Computers in industry*, *127*, 103414.

Oreski, D., Oreski, S., & Klicek, B. (2017). Effects of dataset characteristics on the performance of feature selection techniques. *Applied Soft Computing*, *52*, 109–119.

Pant, S., Kumar, A., Ram, M., Klochkov, Y., & Sharma, H. K. (2022). Consistency indices in analytic hierarchy process: a review. *Mathematics*, *10*(8), 1206.

Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*.

Qiu, H., Lee, J., Lin, J., & Yu, G. (2006). Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *Journal of sound and vibration*, *289*(4-5), 1066–1090.

Redman, T. C. (1997). *Data quality for the information age*. Artech House, Inc.

Ren, J., & Lützen, M. (2015). Fuzzy multi-criteria decision-making method for technology selection for emissions reduction from shipping under uncertainties. *Transportation Research Part D: Transport and Environment*, *40*, 43–60.

Saaty, T. L. (1980). The analytic hierarchy process (ahp). *The Journal of the Operational Research Society*, *41*(11), 1073–1076.

Saaty, T. L. (2005). *Theory and applications of the analytic network process: decision making with benefits, opportunities, costs, and risks*. RWS publications.

Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. In *2008 international conference on prognostics and health management* (pp. 1–9).

Shao, S., McAleer, S., Yan, R., & Baldi, P. (2018). Highly accurate machine fault diagnosis using deep transfer learning. *IEEE Transactions on Industrial Informatics*, *15*(4), 2446–2455.

Smith, W. A., & Randall, R. B. (2015). Rolling element bearing diagnostics using the case western reserve university data: A benchmark study. *Mechanical systems and signal processing*, *64*, 100–131.

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM*, *40*(5), 103–110.

Tobon-Mejia, D. A., Medjaher, K., Zerhouni, N., & Tripot, G. (2012). A data-driven failure prognostics method based on mixture of gaussians hidden markov models. *IEEE Transactions on reliability*, *61*(2), 491–503.

Wang, Z., Yang, J., Jiang, H., & Fan, X. (2020). Cnn training with twenty samples for crack detection via data augmentation. *Sensors*, *20*(17), 4849.

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big data*, *3*, 1–40.

Wen, L., Gao, L., & Li, X. (2017). A new deep transfer learning based on sparse auto-encoder for fault diagnosis. *IEEE Transactions on systems, man, and cybernetics: systems*, *49*(1), 136–144.