IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 6 October 2022, accepted 22 December 2022, date of publication 29 December 2022, date of current version 5 January 2023.

Digital Object Identifier 10.1109/ACCESS.2022.3233220

SURVEY

Using Data From Similar Systems for Data-Driven Condition Diagnosis and Prognosis of Engineering Systems: A Review and an Outline of Future Research Challenges

MARCEL BRAIG AND PETER ZEILER
Research Group for Reliability Engineering and Prognostics and Health Management, Faculty of Mechanical and Systems Engineering, Esslingen University of Applied Sciences, 73728 Esslingen am Neckar, Germany

Corresponding author: Marcel Braig (marcel.braig@hs-esslingen.de)

This work was supported in part by the Baden-Württemberg Ministry of Science, Research and Culture, Germany; and in part by the Esslingen University of Applied Sciences in the Funding Program Open Access Publishing.

ABSTRACT Prognostics and health management (PHM) is an engineering approach dealing with the diagnosis, prognosis, and management of the health state of engineering systems. Methods that can analyze system behavior, fault conditions, and degradation are crucial for PHM applications, as they create the basis for determining, predicting, and monitoring the health of engineering systems. Data-driven methods have been proven to be suitable for automated diagnosis or prognosis due to their pattern recognition and anomaly detection abilities. Moreover, they do not require knowledge of the underlying degradation process. However, training data-driven methods usually requires a large amount of data, whose collection, cleansing, organization, and preparation are generally very time-consuming and costly. There are usually little or no run-to-failure data available at market launch, especially for new systems such as new machine generations. Nevertheless, related systems, hereinafter referred to as similar systems, often already exist, differing only in some technical characteristics. In this paper, the similar system problem is defined, and explanations of the difficulties that arise when using data from similar systems are presented. Furthermore, it is discussed why the usage of these data offers great potential for condition diagnosis and prognosis of engineering systems. An overview of data-driven methods that can be used to utilize data from similar systems is provided, and the methods that such systems already consider are highlighted. Two related research areas are identified, namely, fleet learning and transfer learning. In the paper, it is shown that similar system approaches will become an important branch of research in PHM. However, some difficulties must be overcome.

INDEX TERMS Condition diagnosis, condition prognosis, data-driven methods, fleet learning, prognostics and health management (PHM), similar system approach, similar system problem, transfer learning.

I. INTRODUCTION

Prognostics and health management (PHM) addresses fault detection, diagnosis of the specific fault, and prognosis of the future degradation of engineering systems. Fig. 1 shows the possible elements of the PHM. Based on current operational data and additional (historical) knowledge about the system, an attempt is made to estimate the system's current state of health. With this estimation, predictions of future degradation are possible. In this way, the time of failure can be predicted, and the remaining useful life (RUL) can be determined. This allows health management processes such as maintenance planning, logistics measures, and performance regulations to be improved. PHM thus forms the basis for advanced maintenance strategies such as condition-based maintenance, predictive maintenance, and prescriptive

The associate editor coordinating the review of this manuscript and approving it for publication was Gongbo Zhou.

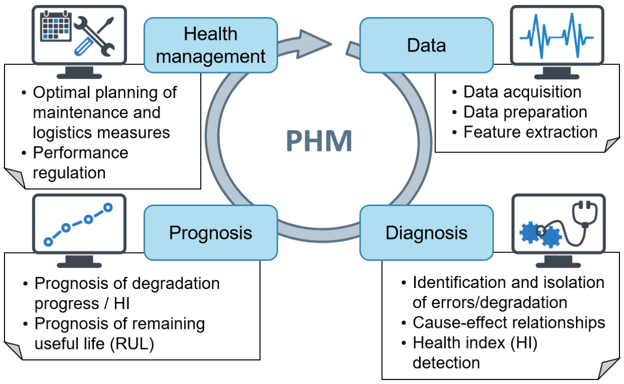

FIGURE 1. Elements of PHM.

maintenance [1], [2], [3]. As seen in Fig. 1, the elements of the PHM are arranged in a circle, where recursive evaluation and optimization can take place.

When using PHM

- downtime and the number of unexpected failures can be reduced,
- uptime between maintenance actions can be maximized without causing serious outages, and
- the life of components can be extended by avoiding unfavorable operating conditions.

This results in reduced maintenance and life cycle costs while increasing the productivity and system reliability [4], [5], [6].

To make statements about the current and future health of an engineering system, both condition diagnosis and condition prognosis methods are necessary. Condition diagnosis methods aim to identify the current health of a system (e.g., the current health index or the presence of a fault), while condition prognosis methods predict the future degradation pattern (e.g., by using the health index) or the RUL. In this paper, a more detailed distinction between fault detection and fault state assignment, e.g., the fault cause or the fault severity [7], is omitted.

In the literature, existing PHM methods for condition diagnosis and prognosis are divided into different categories. The most common distinction is between model-based and data-driven methods. When model-based and data-driven methods are combined, they are referred to as hybrid methods. Literature using this division includes [5], [8], [9], [10], [11], [12], [13], [14]. This categorization is also used in this paper. Depending on the authors, additional divisions of the existing methods are made, e.g., in [15] and [16].

Model-based methods use mathematical models that describe a process, machine, or system in general and are based on physical or chemical principles. In addition to modeling the exact processes in the system, a mathematical description can also be obtained through system identification techniques. For example, mechanically oscillating components behave somewhat like a damped spring-mass oscillator. Once the model is created, condition diagnosis algorithms then monitor, for example, the agreement between the measurements on the real system and the values predicted by the model to identify any deviations that occur and to detect faults that may be present in the system [12], [17], [18], [19]. However, the term "model-based" is not unambiguous. For instance, a neural network could also be seen as a model. To avoid confusion, model-based methods are referred to below as physical methods using physical models. One of the main drawbacks of physical models is the need for a physical understanding to describe the system behavior. In addition, the design of such a model can be very time-consuming and quickly reach its limits for complex systems.

Data-driven methods do not need a physical model. The information about the system is obtained rather exclusively from the data provided [18], [20], [21], whereby machine learning and statistical methods are mostly used. Data-driven approaches are particularly suitable for automated diagnostic systems due to pattern recognition and anomaly detection. In addition, they do not require knowledge of the underlying degradation process [9], [10], [22]. Therefore, they can usually be implemented at low costs. In recent years, data-driven methods for condition diagnosis and prognosis have become increasingly important in PHM. This is partly because today's systems have a large number of sensors, and thus readings of the current system status are generally available [23], [24], [25]. In addition, physical models are becoming increasingly complicated due to the increasing complexity of engineering systems [1], [26].

However, data-driven methods require extensive training data. Data collection and processing are generally very time-consuming and costly. In the case of novel systems, such as new machine generations, the situation is further complicated. Usually, there are little or no runtime and run-to-failure data available at the time of market launch. Due to the typically long life of engineering systems, during which operating conditions change several times due to environmental changes, wear, or the replacement of parts, it is almost impossible to quickly collect a representative set of data [27]. In addition, in many cases, machines are customer designed and only manufactured in small batches, which makes it difficult to create large datasets [28]. All of the above challenges limit the use of data-driven condition diagnosis and prognosis methods for PHM [5], [29], [30].

The aim of this paper is to draw attention to a novel field of PHM research that can contribute to addressing the lack of data in data-driven condition diagnosis and prognosis applications. The basic idea is to use data and knowledge from related systems, herein referred to as similar systems, to increase the amount of available training data. Similar systems may include machines of previous generations or other dimensions. An investigation of how data from similar systems can be used for data-driven diagnosis and prognosis methods of PHM is conducted. The performance of data-driven methods is based directly on available data, suggesting the use of data from similar systems when data are scarce. It should be explicitly noted that the PHM area of health management, i.e., the use of diagnosis and prognosis information such as health state or RUL to manage system health, is not considered here. This is also not currently the focus of the literature dealing with the use of data from similar systems.

Another point worth mentioning is that, depending on the data-driven diagnosis and prognosis method, more or less data are required for model training. This phenomenon is due to the different ways each data-driven method works. For example, deep learning usually requires more data than shallow learning since more model parameters have to be adjusted, and thus, there is a higher risk of overfitting. Additionally, some statistical methods and methods for subspace identification usually manage with relatively little training data. Discussions of the requirements of statistical and machine learning methods in relation to the amount of data can be found, for example, in [31], [32], and [33]. However, the explicit aim of this paper is not the investigation of which data-driven methods require more data and which require less data. Instead, approaches are presented for using similar data from other systems in the case that the data from the system under study are insufficient. This approach can be valuable for all data-driven methods, whether they require more or less data. In addition, there are further possibilities to reduce the amount of data required by the methods, e.g., bootstrapping, cross-validation, or ensemble learning. However, these are also not discussed.

In the further course of this paper, under the key phrase "similar system problems", existing approaches that use data from similar systems are highlighted. First, an overview of the procedure conducted for the literature search is provided in Section II. Subsequently, in Section III, the problem underlying the newly defined field of research in this paper is explained in detail. In addition, the concept of similar system approaches is presented. This is followed by the presentation of related fields of research in Section IV, whose approaches are also suitable for similar system problems, namely transfer learning and fleet learning. In Sections V to IX, for each research field, a review of existing methods for condition diagnosis and prognosis of engineering systems is given. In particular, the existing approaches are clarified, applications are considered, and an explanation of how these approaches currently address similar system problems is presented. Based on this, the applications of the approaches that specifically make use of similar system data are discussed in more detail in Section X. In addition to considering similar systems, there are also approaches that consider similar processes. Such approaches are presented in Section XI. In Section XII, negative transfer is outlined as one of the main problems when using similar system data. Finally, findings are summarized, current challenges are listed, and recommendations for future research are given in Section XIII, and this paper is concluded in Section XIV.

II. LITERATURE SEARCH PROCEDURE

The aim of this paper is to introduce the problem of similar systems and their great potential for condition diagnosis and prognosis, as well as to give an overview of approaches that already consider similar systems or different operating conditions. For the latter, a literature search is conducted covering the period through the end of 2021. The chosen procedure is based on that presented in [34]. Approaches outside the PHM research area, such as in the field of human health, are explicitly excluded.

The keywords from which search strings are formed are listed in Table 1. First, existing relevant research fields in PHM are identified using the keywords in the first row after the header line. This is followed by a targeted search for PHM approaches in the identified research fields of transfer and fleet learning using the keywords also listed in Table 1. Forward and backward snowballing [35] are used to supplement the searches. The searches are performed in Web of Science, Google Scholar, and Science Research. Depending on the search engine, only a subset of the keywords can be used for the search due to character limitations. In these cases, the keywords for the transfer learning and fleet learning searches are split into several subsearches. In the preceding search for related research fields, however, no partial searches are performed, but the search is limited to the central terms in case of character restrictions. The prioritization results from the order in which the terms are mentioned in Table 1. The searches in Web of Science are performed by title, abstract, and keywords. The number of papers to be evaluated is reduced by restricting the search by specifically excluding categories such as medical specialties. The first 400 most relevant results are considered in Google Scholar and Science Research. In Google Scholar, patents and citations are excluded from the search.

The screening of the relevant literature is based on the title and abstract. The decisive factor for selection is that the field of application is in the area of condition diagnosis or prognosis of engineering systems, and a deviation of these systems exists due to either different operating conditions or technical characteristics of the systems themselves.

At this point, references should be made to other related papers that provide overviews of transfer learning approaches in industrial applications. Maschler and Weyrich [36] presented a summary of transfer learning in industrial automation. The approaches were subdivided into anomaly detection, time series prediction, computer vision, fault diagnosis, fault prognosis, and quality management. Some essential approaches were listed for all the areas. Li et al. [8] provided a review of deep transfer learning for machinery fault diagnosis, presenting the most typical deep transfer learning models. Moradi and Groth [37] reviewed central transfer learning approaches specifically in the area of PHM. Yan et al. [38] provided an overview of transfer learning specifically for rotating machinery fault diagnosis. The literature review that is included in this paper gives an updated overview of PHM approaches for condition diagnosis and prognosis in transfer learning, which in principle are also suitable for application areas of similar systems. In contrast to the listed existing reviews, an explicit distinction is made between approaches that "only" consider different operating

TABLE 1. Keywords for literature search (* = right-hand truncation, $ = a wildcard character).

Purpose	Keywords
Identifying related fields of research	Combinations of similar*, alike*, comparab*, akin*, homogen*, heterogen*, relat*, allie*, same-like*, same like*, same-type*, same type*, likeab* system$, machine*, component*, device*, unit*, domain*, fleet*, population*, data, generation*, product*, appliance*, gadget*, apparatus*, equipment*, resource*, instrument* AND prognostics and health management, PHM, condition monitoring, diagnos*, prognos*, predictive maintenance, condition-based maintenance, prescriptive maintenance, degradation, remaining useful life, RUL, health state, health management, performance monitoring
Research transfer learning	transfer learning, knowledge transfer, domain adaptation, domain adaptive transfer AND prognostics and health management, PHM, condition monitoring, diagnos*, prognos*, predictive maintenance, condition-based maintenance, prescriptive maintenance, degradation, remaining useful life, RUL, health state, health management, performance monitoring
Research fleet learning	fleet, older generation, preceding generation, previous generation AND prognostics and health management, PHM, condition monitoring, diagnos*, prognos*, predictive maintenance, condition-based maintenance, prescriptive maintenance, degradation, remaining useful life, RUL, health state, health management, performance monitoring

conditions and those that already consider similar systems. The former means that there are different environmental, operational, or usage conditions.

Contrary to transfer learning, no comprehensive review of fleet learning approaches has been found in the PHM literature. It can therefore be considered that the review given in this paper represents the first review for condition diagnosis and prognosis in fleet learning. One relevant contribution to mention is that by Fink et al. [25], who covered deep learning directions for PHM approaches. In the course of this paper, some transfer and fleet learning PHM approaches were listed.

III. THE SIMILAR SYSTEM APPROACH

Data-driven PHM methods for condition diagnosis and prognosis of engineering systems offer strong potential because detailed system knowledge is not needed. Nevertheless, the lack of sufficient training data is a major drawback to the use of these methods in industrial settings. However, (new) engineering systems are often based on existing systems. For example, new product generations usually have similarities with their predecessors. Even systems with different dimensions are usually closely related. Furthermore, the ongoing standardization of equipment in the course of cost reduction means that machines are becoming increasingly alike through the use of similar components and configurations [39]. Thus, the structure and components as well as the properties, areas of application, and operating conditions (e.g., operating time, mileage, temperature, load, and pressure) of such similar systems are often comparable. As a result, degradation data from similar systems can be extended as training data for data-driven condition diagnosis and prognosis approaches.

Similar systems are defined in this paper as engineering systems that have similar technical characteristics. This includes systems with common components and similar functional principles. Depending on the level of observation, entire machines, devices, or plants can be seen as similar systems but also individual subsystems or components such as installed bearings or gears.

More formally, based on [40], a system can be described as a finite set of technical characteristics. Given two systems with the sets A and B, the level of similarity can be defined by the intersection $U = A \cap B$. The greater the cardinality $|U| = |\{u_1, \ldots, u_n\}| = n$ is, the more similar the systems are. A threshold for the number of elements u_i in U can be set, above which one can speak of similar systems. Possible types of technical characteristics used to describe and compare systems depend on the type of system under consideration as well as its use. For instance, the components or component types that make up a system can be used for comparison. When considering components, all components of the systems to be compared are grouped into the respective sets A and B. Then, it is checked whether elements from the sets are identical, i.e., identical components are installed (e.g., identical gears in two gearboxes). These identical elements then form the set U. Alternatively, the sets can be formed from the installed component types (e.g., bearings and gears). For example, in this case, if bearings of any type are installed in both systems, the component bearing is an element of the set U. In addition to components, the functions or tasks the systems perform can also be considered to determine similarity. For example, both axles and shafts have the common function of carrying rotating components. Therefore, this function forms an element from the set U. However, a function that is not included in U is the torque transmission, which is only a function of the shaft.

When considering component types or function types, it is possible to specify the similarity more precisely by means of a similarity measure. Given A, B, and U defined as above, the similarity can be evaluated as

$$Q = \mathrm{f}(|A|, |B|, |U|, \mathrm{sim}(u_i)), \tag{1}$$

where sim is a similarity measure that evaluates the similarity of the elements in U. For example, bearings as a component type can differ in terms of dimensioning, the number of rolling elements, or the rolling element type. Similarity measures can then be defined to evaluate the similarity between different bearings. These measures must be defined individually depending on the application. A weighting of the individual deviations according to the significance for the deviation of the overall systems can also be included.

As shown in this paper, for the similarity evaluation of systems, instead of comparing technical characteristics, it is also possible to assess the similarity of systems purely on the basis of available measurement data. This possibility offers the potential to automate the similarity assessment without the need for a human expert to evaluate the technical characteristics.

Examples of technical characteristics based on deviations in the structure of a system include the following:

- Dimensioning/shape: Changes in dimensioning and shape have many effects, e.g., other forces, moments, and stresses can occur and dynamic properties and heat distribution can differ. Furthermore, characteristic vibration patterns are affected. In electronic systems, voltages, currents, resistors, capacitances, and inductances can vary.
- Material: The choice of material has a decisive influence on the properties of a system. The material characterizes mechanical properties such as hardness, elasticity, density, strength, and brittleness and other physical properties such as thermal behavior (e.g., thermal conductivity, melting temperature, and heat capacity), electrical conductivity, and optical properties. In addition, chemical properties such as corrosion/oxidation resistance change with the material [41].
- Characteristics depending on manufacturing processes: The technical characteristics of a product can also vary due to different manufacturing processes, e.g., the cohesion, hardness, and surface quality of a material can be affected.
- Characteristics depending on mounting: If several individual parts are assembled, different joining techniques can be used. Therefore, varying characteristics also result. For example, gluing results in a more uniform force transmission than the use of screws.
- Data acquisition: Deviations in data acquisition can also lead to differences. For example, installed sensor types or sensor locations may vary. Since measurements are taken at other locations, differences in data arise. Different sensors may have different sampling rates or different accuracies, for example. Additionally, different quantities can be measured.

Usually, deviations in the functions also lead to deviations in the structure since other requirements are placed on the system. It should be noted that instead of deviations in the technical characteristics of systems, deviations in the operating conditions can also occur. Operating conditions refer to environmental, operational, or usage conditions. In this paper, the sole presence of different operating conditions is not referred to as a similar system but is considered a second case of similar data. This is because different operating conditions can also lead to deviations in the data. For example, the load can have a strong influence on the quantities to be measured [42] and the degradation rate. Different lubricants, contaminants, or environmental conditions, such as ambient temperature or humidity, also lead to different behaviors. Furthermore, it should be noted that structurally identical systems can differ in terms of their system behavior due to manufacturing tolerances. Causes for manufacturing tolerances include dimensional tolerances, manufacturing condition deviations (e.g., temperature and humidity), material tolerances, or assembly tolerances. However, systems that differ by such tolerances are also not considered similar systems in this paper.

Due to the deviations between similar systems, the models and datasets of these systems cannot be readily adopted for the considered system. This problem, which is referred to as the similar system problem in this paper, prevents the promising use of knowledge from similar systems. For example, products of a new generation often have a different design and a wider range of functions [29]. Therefore, quantities that reflect the system health state (e.g., vibrations, sounds, power consumption, and oil condition) depend on the system design and are thus not necessarily comparable. Due to this similar system problem, approaches are needed that enable the utilization of knowledge from similar systems, despite the differences between them. These approaches are summarized in this paper as similar system approaches. The crucial question in the area of similar systems is therefore "When, for what purpose, and how can existing data from similar systems be used?" The scope of this paper is PHM applications for condition diagnosis and prognosis; however, other application areas are also conceivable if the use of data from similar systems could be advantageous.

IV. RELATED RESEARCH FIELDS

Methods for data-driven degradation estimation can be divided into component- and population-based approaches [26]. Component-based methods use a separate prediction model for each system, trained only with the existing data of the system itself. Since each model is trained only with the data of one specific system, the application of these models to other systems is limited. This is because most data-driven methods only yield good results if the test data, i.e., the data during the application of the model, are from the same domain as the training data, meaning that these data have the same characteristics (features, labels, distributions, etc.). If this is not the case, which is usually true for different systems, training typically has to be repeated with new data [43], [44]. This becomes clear, for example, in fault detection based on anomalies. In this type of approach, a model is trained using only operational data recorded in a healthy state. Then, if a deviation from the training examples occurs in an application, it is detected as a potential fault. However, when the trained model is applied to similar systems, the new operating data in a healthy state may differ from that of the original system. As a result, a fault may also be detected [45], [46].

Population-based approaches use the available data of multiple comparable systems that have some similarity to each other. In this way, the model is made universally valid so that it can be applied to any of the systems, whereby some additional adjustments to the unit under consideration may be necessary. The similar system problem studied in this paper falls under these population-based approaches. According to the conducted literature review, two research areas deal with the use of data from related systems for condition diagnosis and prognosis: transfer learning and fleet learning. As the name suggests, fleet learning essentially pursues the idea of collecting as much data as possible from all occurring domains and using it for training. In this way, the risk that the test data originate from a new, unknown domain is reduced. This is achieved by collecting knowledge from a large number of individuals. As a result, the model should be able to represent the current unit under consideration. Whereas fleet learning pursues the idea of creating a model that can be used across the whole fleet, i.e., all domains or at least a significant part of them, transfer learning focuses on a specific domain. Knowledge from other domains is used in such a way that performance in the considered domain is maximized. Accordingly, the knowledge of other individuals is transferred to the domain of the unit under consideration. The aim is to improve the performance of the model specifically on this considered unit and not across all units.

A. TRANSFER LEARNING

Transfer learning provides approaches to transfer previous knowledge to different application fields. Although transfer learning is a relatively new machine learning approach, researchers are currently focusing on this topic [23]. The majority of applications are in the fields of computer vision, natural language processing, and health care, where transfer learning is already being used successfully [47], [48]. In contrast, a relatively small number of publications have focused on PHM, as confirmed in the literature. According to Mao et al. [19], transfer learning is not yet widely used in the PHM of mechanical components. The lack of application in fault diagnosis has been confirmed by Pan et al. [49] and that in fault prognosis has been confirmed by Maschler and Weyrich [36].

However, as in other fields, transfer learning can be used in PHM to transfer between different application fields. In PHM, such a transfer may be necessary due to minor or major deviations caused by different operating conditions or similar systems. Even if the datasets only slightly differ, e.g., when a trained condition diagnosis or prognosis model is applied to an identical machine operating only under slightly different operating conditions, the prediction performance can be significantly increased by transfer learning [25], [30], [50]. Through transfer learning, existing knowledge can be used for new, similar tasks. Therefore, it is not necessary to start from scratch, as is usually the case in practice [16]. Transfer learning can reduce the amount of required new data and the time needed for training data-driven condition diagnosis and prognosis models. Furthermore, it can improve the quality of the models obtained [51], [52]. Thus, transfer learning is very suitable for data-driven condition diagnosis and prognosis PHM applications considering similar systems.

1) FORMAL DEFINITION OF TRANSFER LEARNING

There are several core areas and transfer approaches to transfer learning. To clarify the differences, it is necessary to first formally define transfer learning. Transfer learning aims to make the learning of a prediction function for a target domain more efficient and successful by using knowledge from another (related) source domain. One or more source domains may exist and can be used. However, for the definition below, only one source domain is assumed. The definition is based on [48], [51], and [53].

A domain is formed from the set $D = \{\mathbf{X}, P(X)\}$ with the feature space $\mathbf{X}$ and the marginal distribution $P(X)$. $X = \{\boldsymbol{x_1}, \ldots, \boldsymbol{x_m}\} \in \mathbf{X}$ corresponds to a sample of size m, and $\boldsymbol{x_i}$ is the i-th feature vector. Typically, in a domain, there is a learning task defined as the set $T = \{\mathbf{Y}, P(Y|X)\}$, comprising the label space $\mathbf{Y}$ and the prediction function $P(Y|X)$, which can be seen as a conditional distribution. $Y = \{y_1, \ldots, y_m\} \in \mathbf{Y}$ is the set of labels belonging to the m feature vectors of sample X. $P(Y|X)$ must be learned using the tuples $\{\boldsymbol{x_i}, y_i\}$. Transfer learning is used when either the source domain D_s and target domain D_t or the learning task of the source T_s and target T_t differ. The notation is shown in Fig. 2. $D_s \neq D_t$ means that the domains differ in either the feature space, i.e., $\mathbf{X_s} \neq \mathbf{X_t}$ or the marginal distribution, i.e., $P(X_s) \neq P(X_t)$. $T_s \neq T_t$ occurs when the label spaces in the tasks are different, i.e., $\mathbf{Y_s} \neq \mathbf{Y_t}$ or the prediction functions differ, i.e., $P(Y_s|X_s) \neq P(Y_t|X_t)$. $D_s = D_t$ and $T_s = T_t$ is a problem that can be solved using traditional machine learning methods. However, if at least one of the abovementioned deviations is present, transfer learning should be used to transfer knowledge from the source domain to the target domain and thus make source data usable for the target task.

Returning to the similar system problem, the source represents one or multiple similar systems, and the target represents a new system for which a data-driven model is needed. Based on the above definition, the following differences may occur in the condition diagnosis or prognosis of engineering systems [54]:

- Different types of features, e.g., through different sensors.
- Different marginal distributions, e.g., different probability density functions for one feature. For example, the operating times of systems with different dimensions can significantly vary from each other.
- Different types of labels, e.g., different fault types.

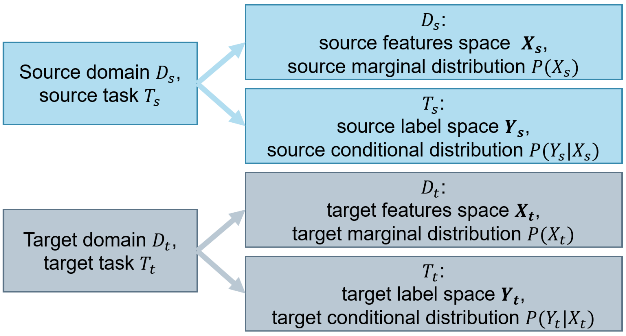

FIGURE 2. Source and target subdivisions in transfer learning.

- Different conditional distributions, e.g., different characteristic frequencies. Systems of different dimensions have different characteristic frequencies. A deflection in the vibration signal at a given frequency that indicates damage in one system may also occur during the healthy operation in the other system.

2) CORE AREAS OF TRANSFER LEARNING

Transfer learning can be divided into three core areas: inductive transfer learning, transductive transfer learning, and unsupervised transfer learning. In this paper, this division is also used for the PHM condition diagnosis and prognosis approaches of transfer learning presented in Sections VI to VIII. The explanations below are based on the definitions in [53], [24], [37], and [55]. One way to divide the areas is by the availability of labeled or unlabeled data in the source and target domains. In addition, each area has its own goal.

a: INDUCTIVE TRANSFER LEARNING

Available Data: Labeled target data & labeled or unlabeled source data.

Goal: Improve the approximation of $P(Y_t|X_t)$ using knowledge in the source domain and knowledge about the source task, while $T_s \neq T_t$. Since the tasks are different, at least some labeled target samples are essential. For the domain difference, both cases $D_s = D_t$ and $D_s \neq D_t$ are conceivable.

Main Transfer Approaches: Instance-based transfer, relation-knowledge-based transfer, parameter-based transfer, and feature-representation-based transfer.

If there are labeled data in the source domain, inductive transfer learning is very similar to multitask learning. However, the latter focuses on learning multiple tasks simultaneously and with balanced quality, whereas inductive transfer learning attempts to maximize the performance of the target task purposefully. If no labeled source data are used, the type of learning is considered to be self-taught learning. Therefore, with a large unlabeled source dataset, a feature representation is learned. This representation is then used to train a supervised learning task with the labeled target data. Multitask learning and inductive transfer learning are not considered in this review.

b: TRANSDUCTIVE TRANSFER LEARNING

Available Data Unlabeled target data & labeled source data.

Goal: Allow the approximation of $P(Y_t|X_t)$ using knowledge in the source domain and knowledge about the source task given $D_s \neq D_t$ and $T_s = T_t$. Accordingly, two types of domain differences could occur: $\mathbf{X_s} \neq \mathbf{X_t}$ or $P(X_s) \neq P(X_t)$. In this paper, the condition $T_s = T_t$ is somewhat softened. The label spaces have to be the same ($\mathbf{Y_s} = \mathbf{Y_t}$), or the source label space has to be at least a subset of the target label space ($\mathbf{Y_s} \subset \mathbf{Y_t}$) and vice versa ($\mathbf{Y_t} \subset \mathbf{Y_s}$). In addition, in condition diagnosis or prognosis applications, the prediction functions will never be identical. Hence, they only have to be similar ($P(Y_s|X_s) \approx P(Y_t|X_t)$) so that the knowledge of the source domain can be applied to the target domain with an acceptable degree of accuracy loss. This means that it is assumed that the tasks are so similar that labeled target data are not necessary.

Main Transfer Approaches: Instance-based transfer, parameter-based transfer, and feature-representation-based transfer.

This transfer learning area also includes domain adaptation. Domain adaptation is often defined as a branch of transfer learning. The source and target have the same task $T_s = T_t$ and the same feature space, i.e., $X_s = X_t$, but different marginal distributions, i.e., $P(X_s) \neq P(X_t)$ [53], [56], [57], [58]. However, this distinction is not consistent in the literature. For example, da Costa et al. [16] and Weiss et al. [51] wrote that domain adaptation is used when domains have different feature spaces. This review also considers domain adaptation approaches, which are an essential part of transductive transfer learning. However, domain adaptation approaches are not explicitly differentiated from other transductive transfer learning approaches.

At this point, it should be highlighted that even with the additional restriction $D_s = D_t$ (so $T_s = T_t$ and $D_s = D_t$), sample selection bias or a covariance shift can still occur. Sample selection bias occurs when the samples of the source and/or target are drawn from the same distribution, albeit not independently. This is typically the case with real datasets [59]. The covariance shift results from the difference between the sample distribution and the distribution of the population due to the sampling scheme [60]. Accordingly, both generally occur in every data-driven PHM application for condition diagnosis or prognosis.

c: UNSUPERVISED TRANSFER LEARNING

Available Data: Unlabeled target data & unlabeled source data.

Goal: Find transferable relationships between $P(X_s)$ and $P(X_t)$. The main difference from the two previous methods is that unsupervised transfer learning deals with unsupervised learning tasks such as dimensionality reduction, clustering, anomaly detection, and pattern recognition. Neither labeled source nor labeled target data are available. Unsupervised transfer learning is also referred to as self-taught learning. One such approach is to cluster a small target dataset using

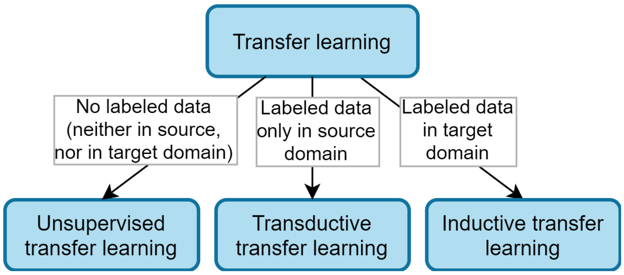

FIGURE 3. **Core areas of transfer learning.**

a large unlabeled source dataset (self-taught clustering). Moreover, a feature representation can be learned using a large unlabeled source dataset. However, the latter alone is usually not sufficient for condition diagnosis or prognosis applications and is therefore often embedded in an inductive or transductive approach.

Main Transfer Approaches: Instance-based transfer and feature-representation-based transfer.

Fig. 3 shows a graphical representation of the core areas of transfer learning, differentiated according to the availability of labeled data in the source and target domains. In addition to the three mentioned classical areas of transfer learning, Niu et al. [61] introduced two additional areas, cross-modality transfer learning and negative transfer learning. To apply classical transfer learning, there must be connections between the source and target feature spaces or the corresponding label spaces. Thus, source and target data must be of the same modality, e.g., text, picture, or audio. Cross-modality transfer learning claims to transfer knowledge between different modalities. For example, the source domain could comprise features of images and the target domain could comprise features of texts. As will be explained in Section XII, there is a risk associated with transfer learning called negative transfer. Briefly, negative transfer occurs when too much uncorrelated information from the source domain is used for the target domain. Therefore, initial negative transfer learning approaches focus on the transfer of knowledge between two widely separated domains and the possibilities to quantify the impact of a negative transfer. However, these areas do not currently play a role in PHM applications for condition diagnosis and prognosis and are therefore not considered further below.

3) TRANSFER APPROACHES IN TRANSFER LEARNING

Transfer learning is also often divided in the literature into different transfer approaches, depending on what is transferred [24], [38], [51], [53].

a: PARAMETER-BASED TRANSFER OR MODEL-BASED TRANSFER–ADOPTION OF THE SOURCE MODEL

In this process, fully trained data-driven models, including model parameters, are transferred from the source to the target. In a weakened form, only the model structure, hyperparameters, or parts of the model are transferred [38], [52], [62]. Depending on the similarity of the data, the transferred model can, for example, be used as initialization for retraining with target data or utilized directly in the target domain.

b: INSTANCE-BASED TRANSFER–ADOPTION OF SOURCE DATA

In these approaches, source data are used directly in the target domain. This means that for the training of data-driven algorithms, data from the source are used in addition to data from the target. By weighting, the impact of the target data can be increased or a part of the source data can be sorted out [53]. Likewise, the source data can be used to train a classifier that assigns pseudo labels to the existing unlabeled target data [63].

c: FEATURE-REPRESENTATION-BASED TRANSFER–ADOPTION OF THE SOURCE FEATURE SPACE

In feature-representation-based transfer, feature spaces are transferred. For example, if a suitable feature space has been found for the source domain through a dimensionality reduction procedure, it may also be possible to use this space in the target domain. The feasibility of this approach depends on the raw data, e.g., from the sensors installed in the system. In addition to the pure transfer of a feature space, a feature space can also be found in which the differences between the source and target data are minimized [38], [51], [53], [64].

d: RELATION-KNOWLEDGE-BASED TRANSFER OR RELEVANCE-BASED TRANSFER–ADOPTION OF RELATIONSHIPS WITHIN THE SOURCE DATA

The idea is that relationships between source data also exist within the target data. If this is the case, these relationships can be transferred to the target [51], [53], [65]. An example from text analysis is sentence structure. Regardless of the informal content of a text, the sentence structure is always very similar. Therefore, if the relationship between words in the source is learned, this knowledge can also be applied to words in the target. In PHM, such relationships could be, e.g., that the RUL essentially always decreases and the degradation usually increases with increasing operating time.

The decision on which transfer learning approach to choose depends on the similarity of the domains and the tasks, the size of the existing datasets, and the specific use case and must be made individually [66]. It is also possible to combine several approaches.

B. FLEET LEARNING

Fleet learning attempts to generate knowledge from fleet data that can be used for each individual unit. In PHM, such a fleet may consist of several similar systems or identical systems operating under different operation conditions. In this way, the amount of available data can be significantly increased, which can greatly improve the results of condition diagnosis and prognosis. For example, fleet service life data can be used

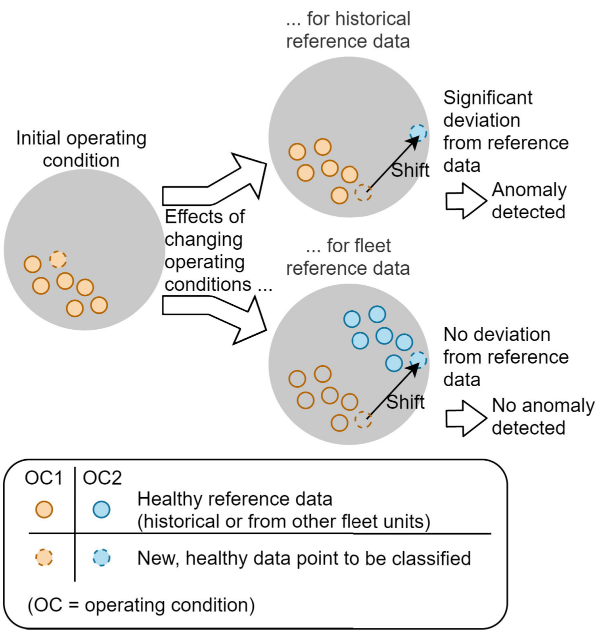

FIGURE 4. Fleet learning under operating condition shift.

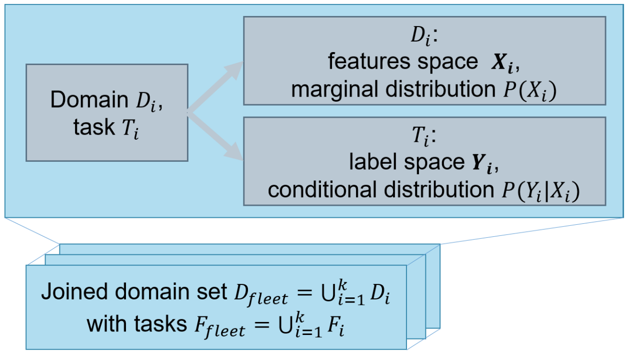

FIGURE 5. Joined domain set in fleet learning.

to make predictions of the degradation profile or RUL of an individual unit [67].

In addition to increasing the amount of data, fleet learning approaches offer a further advantage through the simultaneous consideration of several systems [68]. If there is a fleet whose systems behave similarly in a healthy state, the following approach is possible. It is assumed that most of the systems in the fleet are in a healthy state. If one system deviates significantly from the others, it may be an indication that there is a fault. In contrast to supervised approaches based on historical data (training data), data on all possible faults are not needed. Moreover, unlike unsupervised anomaly detection approaches that rely on historical data from healthy states, operating conditions can change without affecting fault detection functionality if the change occurs for all systems. In contrast, for approaches based on historical healthy state data, a change in operating conditions may cause the new healthy state data to differ from the historical data. Thus, the false detection of a fault is likely. This is illustrated in Fig. 4. In the case of historical data, the change in operating conditions only affects the new data point, resulting in the detection of an anomaly. For fleet data, the shift affects the reference data and the data point to be classified in the same way. Thus, no anomaly is detected, and no false faults are reported.

However, a challenge in fleet learning is that a fleet usually consists of many systems, and even identical systems may behave differently due to different operating conditions [42], [69]. Possible deviations in the technical characteristics further increase the diversity of the fleet.

In the literature, there is no agreement on the name of the fleet learning research field. Different terms, such as "fleet PHM" [70], "fleet-level prognostics" [69], "fleet monitoring" [27], have been used, or the topic is described with expressions such as "fleet-based systems" [71], "sub-fleet knowledge" [72], or "fleet of products" [13]. In the course of this paper, all of these definitions are subsumed under the term "fleet learning".

1) FORMAL DEFINITION OF FLEET LEARNING

Since no formal PHM-related definition of fleet learning can be found in the literature, the term fleet learning will be formally defined here by an analogy with transfer learning to make the differences between them clear. In addition, Fig. 5 illustrates the idea of fleet learning. Compared with transfer learning, fleet learning can be seen as an approach that attempts to create a fleet model that is as general as possible by using the information from the fleet. The model should be able to be used across the whole fleet or at least a major part of it. Thus, the aim is not to create a model that is only tailored and applied to a specific domain (the target domain), as is the case with transfer learning. Therefore, in fleet learning, there are no source and target domains. The domains are seen as one set $D_{fleet} = \bigcup_{i=1}^{k} D_i$. However, in this set, the contained domains D_i and their learning tasks T_i may differ from each other to a greater or lesser extent, as is the case with transfer learning. Therefore, it may be necessary to first identify units from the fleet showing similar behavior, e.g., by clustering, and to train a common model only for this subfleet. This has been confirmed by Medina-Oliva et al. [73]. According to the authors, the formation of subfleets is essential, especially for fleets consisting of significantly different units.

2) SUBDIVISIONS OF FLEET LEARNING

In contrast to transfer learning, fleet learning is not divided into core areas or specific approaches. Therefore, a division, as found in Section IV-A, is not purposeful. Instead, fleet learning approaches can be differentiated according to which units (in PHM, this means systems) make up the fleet under consideration. A distinction is made between fleets considered from the manufacturer's perspective and those considered from the operator's perspective.

a: FLEET FROM THE MANUFACTURER'S PERSPECTIVE

The consideration of a fleet from the manufacturer's perspective is currently the perspective that receives the most attention. Here, a fleet is a set of homogeneous systems with matching characteristics and properties that operate under different but mostly similar conditions. Due to identical systems, it is very likely that the same labels (e.g., faults or RUL values in PHM) occur across all systems $\mathbf{Y_i} = \mathbf{Y_j}$, and the same sensors and features derived from their signals are typically used $\mathbf{X_i} = \mathbf{X_j}$. However, due to variations in the operating conditions, other distributions of features can occur $P(X_i) \neq P(X_j)$. The prediction function may also deviate somewhat, e.g., through production tolerances, although these deviations should be kept within limits $P(Y_i|X_i) \approx P(Y_j|X_j)$.

b: FLEET FROM THE OPERATOR'S PERSPECTIVE

However, a fleet can also be defined from the operator's perspective as a group of systems that perform the same function but do not have to be from the same manufacturer. Thus, their characteristics and properties differ. For example, the same sensors are typically not installed [25], and the general structures of the systems differ significantly from each other. Therefore, in addition, there may also be different features $\mathbf{X_i} \neq \mathbf{X_j}$ due to different sensors, and other labels can occur $\mathbf{Y_i} \neq \mathbf{Y_j}$ due to different degradation behaviors.

In addition to the subdivision based on the manufacturer's and operator's perspectives, there is a further subdivision into identical, homogeneous, and heterogeneous fleets [21], [74].. This is based on the similarity of the systems (i.e., technical characteristics) and the operating conditions (i.e., environmental, operational, and usage conditions) under which they operate. According to this definition, identical fleets consist of systems with identical characteristics operating under the same operating conditions. In contrast, homogeneous and heterogeneous fleets can have different technical characteristics. The difference between homogeneous and heterogeneous fleets lies in their different environmental, operational, and usage conditions. However, for the categorization of fleet learning approaches in this paper, a division into manufacturer's and operator's perspectives is more useful, as it distinguishes between fleets with identical systems and fleets with similar systems.

V. REVIEW OF PHM APPROACHES FOR CONDITION DIAGNOSIS AND PROGNOSIS

In Sections VI to IX, a review of current PHM approaches for condition diagnosis and prognosis in the areas of transfer and fleet learning is provided. The aim is to give a comprehensive overview of approaches that already deal with similar system data and those that could, in principle, be extended to these applications. Accordingly, two cases are considered. The first case is the consideration of identical systems under different operating conditions, i.e., different environmental, operational, or usage conditions. Due to these different conditions, there are differences in the data between the systems, even though they have the same technical characteristics. Approaches that account for differences due to different operating conditions can therefore also be appropriate for, or adapted to, differences arising from similar systems. Deviations in the operating conditions can thus be seen as a preliminary stage to similar systems and are therefore also included in the review. The second case considers approaches that actually deal with similar systems.

The approaches of transfer learning are subdivided according to the three core areas and presented separately in Sections VI to VIII. This subdivision is based on the availability of labeled data in the domains. However, it must be stated that the subdivision assignments are not always unambiguous. In some cases, there may be overlaps between the areas. For example, to further improve performance in the target domain, some approaches that could proceed without target labels additionally perform fine-tuning with a few labeled target data. Therefore, these approaches can no longer be assigned to transductive transfer learning but must be allocated to the inductive transfer learning.

In condition prognosis, for subdivision into inductive and transductive transfer learning, two cases must be distinguished: direct and indirect prognosis. In direct prognosis, the end of life (EOL) or the RUL is directly predicted based on the previously measured values of the system. This corresponds to the application of multivariate pattern mapping. In contrast, indirect prognosis predicts the development of the damage-determining variable over time. This variable is usually called the health index (HI). An example is the decreasing capacity of batteries. The EOL is defined as the point in time in which the HI reaches a certain threshold. The RUL is the time remaining until the EOL [75]. Thus, in direct prognosis, the RUL (or the EOL) can be seen as the label; in indirect prognosis, the HI can be seen as the label. Therefore, direct prognosis is referred to as inductive transfer learning when target RUL labels are available in addition to source RUL labels. If only source data have assigned RUL labels, direct prognosis is referred to as transductive transfer learning. Note that RUL labels can only be assigned if entire degradation runs up to the EOL are available. Accordingly, inductive transfer learning is only possible in direct prognosis if such runs are available in the target domain. For indirect prognosis, the HI values are the labels. Thus, if HI data are available in the target domain, inductive transfer learning is already possible. It is already sufficient if the target HI values are only available for the early degradation state and not until the EOL. This is a clear difference from the direct prognosis.

After the transfer learning sections, the fleet learning approaches are considered in Section IX. These approaches are divided into fleets defined from the manufacturer's perspective and fleets from the operator's perspective. For each of the Sections VI to IX, the existing approaches and applications are presented. In Tables 3 and 5, the approaches are subdivided depending on whether they ''only'' consider different operating conditions or already consider similar systems. Different noise levels and sensor sampling rates are also understood as different operating conditions. In Tables 3 and 5, sources that are marked as * focus especially on different faults, i.e., different fault severities or different fault types. In Section X, the applications of the transfer and fleet learning approaches that specifically consider similar systems are discussed in more detail. It is explained how exactly these systems differ. Tables 3 and 5 also indicate which signal types are considered in the references for condition diagnosis or prognosis and whether real measurement data or simulation data are used. This subdivision takes into account the signals from which the condition diagnosis or prognosis is mainly based. In some references, other auxiliary signals, such as the rotational speed of bearings and gears or the ambient temperature of batteries, are taken into account as additional variables. However, since these signals do not give any direct indication of possible faults and to avoid further expansion of the tables, they are not mentioned in the tables. In addition, it should be noted that in the tables, the term ''vibration'' refers to oscillations of a body, while ''sound'' describes oscillations in gas (air in the applications). As described in [76], simulation data are the numerical outcomes of simulations performed on a computer. The simulation model approximates the behavior of a real-world system or process. In contrast, measurement data originate from real measurements on real-world systems or processes, e.g., on test rigs or in industrial applications.

Figs. 12 and 13 show the main concepts of the existing transfer and fleet learning approaches for condition diagnosis and prognosis, which are discussed in Sections VI to IX. The corresponding references are shown in Tables 4 and 6, which are divided according to condition diagnosis and prognosis. Thus, an overview of the most common concepts is provided. In Sections VI to IX, when approaches are presented, references are made in parentheses to the categories of these two figures and tables. As a supplement to Section X, Table 7 gives a specific overview of all references that actually deal with similar systems. The applications and areas of the approaches are named, and a distinction is made between condition diagnosis and prognosis.

VI. INDUCTIVE TRANSFER LEARNING APPROACHES

In the following, existing inductive transfer learning approaches in the PHM field of condition diagnosis and prognosis are presented. As explained in Section IV, these approaches are characterized by the fact that labeled target data are available. In inductive transfer learning, parameter-based and instance-based transfer are most common; therefore, this subdivision will be used in Sections VI-A and VI-B. In addition, there are other approaches (Section VI-C) as well as combinations of multiple approaches (Section VI-D). The discussion on inductive transfer learning approaches concludes with an interim summary in Section VI-E.

A. PARAMETER TRANSFER APPROACHES

In inductive transfer learning, one of the main approaches is parameter-based transfer from the source to the target model, followed by fine-tuning with labeled target data (A1). Classically, similar to computer vision, a convolutional neural network (CNN) is used for this purpose. For image-based monitoring of systems, this approach can easily be applied, e.g., in the monitoring of crack growth of buildings [77], the observation of visible surface damage in engineering systems such as wind turbine blades [78], or the determination of tool conditions [79]. Pictures of temperature distributions and flow velocity fields can also be used as inputs [80].

However, degradation in machine components such as bearings is not readily visible. Instead of images, for example, structure-borne sound is typically monitored. Therefore, the sensor data of the observed systems first have to be converted into pictures. A wide variety of procedures exist for this purpose. Time-frequency analysis is a common technique that generates two-dimensional data similar to a picture by visualizing the signal over time and frequency. Many procedures exist to carry out time-frequency analysis, e.g., short-time Fourier transform, constant-Q Gabor transform, and Hilbert–Huang transform [81]. The continuous-time wavelet transform is also very popular. In addition to using time and frequency for picture generation, there are approaches that are limited to one of the two, e.g., Gramian angular fields and Markov transition fields [82], [83]. Another solution is, for example, presented in [84], where the values of the signal were directly arranged in a two-dimensional manner, similar to the arrangement of a picture. A color picture with three color channels can also be created in a similar manner [85], [86]. There are other approaches to generate a picture from time series data, e.g., using the plot of the time signals directly [87] or arranging multiple sensor signals into a two-dimensional matrix [88]. This listing is therefore only intended to show a few possibilities and is not considered to be complete.

Thus, essentially, two different source domains can be used to provide the parameters for the transfer, consisting of either natural images, e.g., images of animals and buildings, or artificially generated pictures from the sensor data of engineering systems. In most cases, using natural images as source knowledge means that the source domain has no functional relationship with the target domain. For example, whereas the source domain can contain images of living beings, the target domain can comprise artificially generated pictures from sensor signals of engineering systems. Then, the only commonality is that both domains contain pictures that can be used to train a CNN. Nevertheless, low-level features such as edges, corners, and surfaces exist in both types of pictures. Accordingly, the low-level layers of a CNN trained with the source data can also be utilized in the target domain. This approach is very advantageous since several pretrained CNN architectures already exist in the field of computer vision. Some of the best-known architectures are the VGG architecture introduced by Simonyan

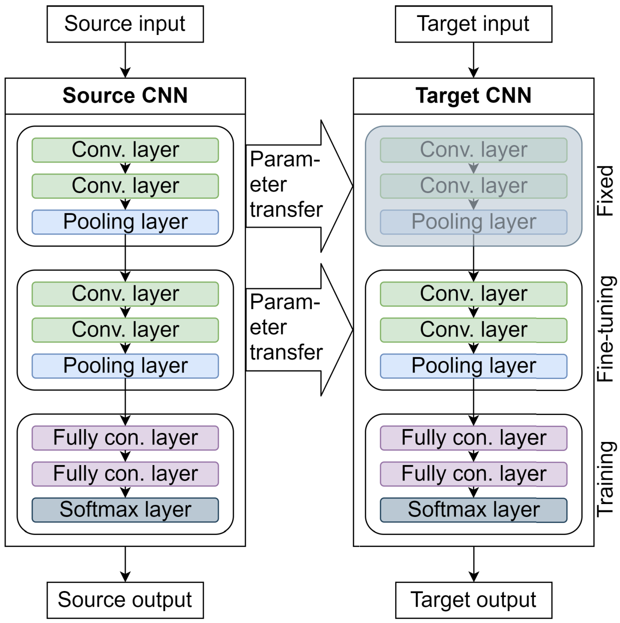

FIGURE 6. Parameter-based transfer approach for CNN.

and Zisserman [89], the ResNet architecture created by He et al. [90], and the Inception architecture developed by Szegedy et al. [91]. Application examples on this type of inductive transfer learning for condition diagnosis include [81] and [92] on bearings, [18] on permanent magnet synchronous motors, [93] on gearboxes, [83] on railway wheels, and [84] and [86] on pumps. Other condition diagnosis approaches were described in [9], [85], [87], [94], [95], and [96] . El-Dalahmeh et al. [97] presented an inductive transfer learning approach for battery capacity prediction and Zhang et al. [98] used an inductive transfer learning approach for bearing and gearbox RUL prognosis.

Fig. 6 shows a principle procedure for parameter transfer for a CNN network. The parameters of the two convolution blocks are transferred to the target domain. The first block is not changed afterward because it mainly extracts low-level features that are very similar in the source and target domains. The second convolution block is fine-tuned with target data to learn problem-specific target features. Unlike the convolution blocks, the fully connected block is not adopted from the source model. This allows its structure (hyperparameters) to be set according to the conditions in the target domain. The fully connected block is then trained from scratch with the target data.

In addition to natural images, artificially generated pictures from the sensor data of similar systems can be used as the source domain. This approach has the advantage that deeper layers that generate deeper features can be adopted in the target domain. The following applications for condition diagnosis can be found in the literature: bearings [66], [99], [100], [101], aircraft engines [102], quadrotor drones [103], batteries [88], [104], [105], gas turbines [106], tanks [107], and gearboxes and rotors [108]. Xu et al. [23] transferred the parameters of shallow CNNs trained with source data to a deeper CNN, which was then fine-tuned with target data. This approach was applied to condition diagnosis of bearings and pumps.

It is also possible to use both natural images and artificially generated pictures for parameter-based transfer. For condition diagnosis of transformer windings, Duan et al. [109] performed several parameter-based transfers, first from a CNN trained with natural images and second with artificial pictures generated from simulation data.

In addition to multidimensional CNNs, there are also CNNs that can process one-dimensional inputs (1D-CNNs). They can therefore process sensor measurement series directly. Kim and Youn [110] used parameter-based transfer on a 1DCNN. The parameters that are fine-tuned after transfer are selected via sensitivity analysis, i.e., based on how much the model output changes when the respective parameter is changed. This approach was applied to bearing condition diagnosis. Other 1D-CNN approaches for condition diagnosis consider transformer rectifier units [111] and gearboxes [112]. Li et al. [24] applied parameter-based transfer on 1D-CNNs and multilayer perceptron (MLP) networks to determine the conditions of bearings and gearboxes. Wang et al. [113] used multiple concatenated 1D-CNNs applied to bearing condition diagnosis.

CNNs are one of the most common types of neural networks used in inductive transfer learning via parameter-based transfer. However, there are other types, such as recurrent neural networks (RNNs). Wang et al. [43] used a parameter-based transfer of a combined network comprising a CNN and a long short-term memory (LSTM) network for bearing condition diagnosis. LSTMs belong to the RNN class and can store and access information over long periods of time [114]. Zhu et al. [115] pretrained an LSTM classification model with source data and fine-tuned the received network with target data. This approach was also applied to condition diagnosis of bearings. Tan and Zhao [116] presented an LSTM approach to forecast the state of health of lithium-ion batteries. Additional approaches for the condition prognosis of batteries using RNNs can be found in [117] and [118]. Although it is not considered a PHM application area in this paper, reference should be made to Liu et al. [119], who presented an approach for state of charge estimation.

A parameter-based transfer with an MLP for the condition diagnosis of gearboxes was presented in [120]. Deng et al. [121] used a stacked autoencoder (SAE) for the condition diagnosis of wind turbine systems and pump trucks. Li et al. [122] performed rolling bearing condition diagnosis with a nonnegativity-constrained sparse autoencoder. Other condition diagnosis approaches using autoencoders for parameterbased transfer learning include the works by He et al. [123] and Chen et al. [124] on gearboxes, Chen et al. [125] on bearings, and Li et al. [126] on wind turbines. Di et al. [127] presented an ensemble-based approach using SAEs. Pan et al. [128] designed an ensemble of random forests to transfer knowledge to the next transistor generation. Gribbestad et al. [129] applied transfer learning to feed forward neural networks, LSTMs, and CNNs to predict the RUL of marine air compressors.

Other parameter-based transfer learning approaches using more machine learning methods include [130], [131], and [132] for condition diagnosis and [133] and [134] for condition prognosis. Guo et al. [135] trained and transferred a data-driven model for parameter identification of a physical model.

B. INSTANCE TRANSFER APPROACHES

For inductive transfer learning, in addition to transferring trained models, there are also approaches that use the source data directly to train the target model, i.e., perform an instance-based transfer (A2). For this purpose, the TrAdaBoost algorithm is popular in PHM applications. TrAdaBoost was introduced in [136]. The core idea is to weigh the source samples during training based on their similarity to the target domain. The better the match to the target data is, the higher the weight and the stronger the influence on training. Source samples that do not match the target domain are weighted very weakly and therefore distort the model only slightly. TrAdaBoost is a boosting algorithm. In each iteration, an iteration model is trained with the weighted data. Based on the errors of this iteration model, the weights are adjusted. Then, the next iteration starts with the adjusted weights. Finally, the iteration models obtained in this way can be weighted and summed, e.g., as shown in [44], to obtain the final model. The iteration model weights are based on the errors of the respective iteration model on the target data. Application examples of TrAdaBoost or similar algorithms include those on bearings [44], [137], [138], high-voltage circuit breakers [49], disks in data centers [139], induction motors [140], gas turbines [141], and self-organizing femtocell networks [142]. All of these approaches considered the condition diagnosis.

In addition to TrAdaBoost, there are additional approaches to transfer instances. Lee et al. [143] also used weighted source samples for the training of a condition diagnosis method for spot-welding machine equipment. Weighting was realized by using statistical similarity measures. Such measures are used in particular for feature matching (see Section VII). Another approach that weights source samples can be found in [144].

C. OTHER APPROACHES

In addition to the inductive transfer learning approaches mentioned thus far, there are others, although they are not as common in PHM applications for condition diagnosis and prognosis. Feature alignment, also known as domain alignment, is one such approach (A3). The maximum mean discrepancy (MMD) is often used to realize feature alignment. The MMD is a distance metric that can be used to measure the distribution discrepancy between two datasets. By minimizing the MMD between the feature values of the source and target datasets during training, the

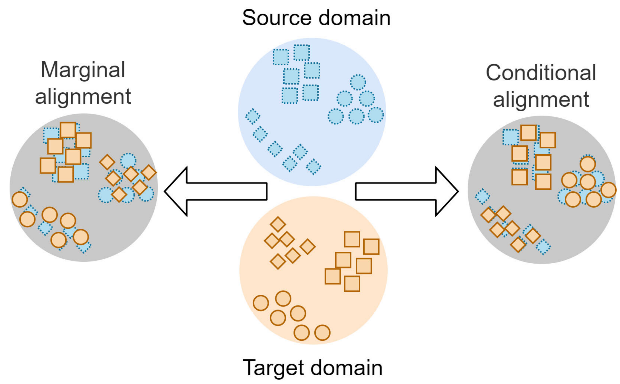

FIGURE 7. Class assignment problem in marginal alignment and the solution through conditional alignment. Adapted from [148, p. 335].

differences between the datasets can be reduced. In principle, no labels are necessary for the alignment of marginal distributions. Therefore, their integration into transductive transfer learning approaches with no labeled target data available is particularly popular. For this reason, MMD will be explained in more detail in Section VII, which focuses on transductive transfer learning. In addition to MMD, there are additional metrics that can be used for feature alignment that will be described in Section VII. However, for conditional distribution alignment, the usage of source and target labels is helpful (inductive transfer learning). If they are available, the source and target distributions can be aligned class by class.

Fig. 7 shows a class assignment problem that could occur when there is only marginal alignment and no conditional alignment. Thus, in marginal alignment, two classes of the target are assigned to the wrong classes of the source. From simply observing the marginal distributions, however, this is not recognizable, as the domains seem to be well aligned to each other. PHM approaches for classwise alignment were presented by Lu et al. [145] and Liu et al. [146] and applied to condition diagnosis of bearings and gearboxes. In addition to feature alignment, labeled target data can be further used for supervised training, as in [147], for the condition diagnosis of bearings.

Moreover, adversarial approaches can be found in inductive transfer learning (A4). Domain-adversarial neural networks (DANNs) are particularly widespread. Although adversarial approaches are popular for transductive transfer learning, they can also benefit from available target labels. As the name suggests, the idea is to train subnetworks using an adversarial approach. In DANNs, a feature extractor and a domain discriminator counteract each other. The discriminator attempts to classify the samples with their domain labels based on the features delivered by the feature extractor. The extractor attempts to deceive the discriminator. A detailed explanation is given in Section VII. In the area of inductive transfer learning, Cheng et al. [149] used a small amount of labeled target data to improve the training results in one of their bearing condition diagnosis approaches. Xu and Li [150] and Han et al. [151] presented a more developed DANN with multiple discriminators, one for each class. These approaches were applied to condition diagnosis of bearings and wind turbines. Due to different label spaces, Li et al. [152] used a separate classifier for each domain (several sources and one target). However, by means of adversarial training, they attempted to find a generalized feature space for the domains that helps the target classifier achieve better condition diagnosis results. Mao et al. [153] and Zhang et al. [154] also presented adversarial approaches for condition diagnosis.

Instead of requiring labeled target data from all or a majority of the existing classes of states of health, there are approaches that only require the availability of labeled target data from one class. These approaches proceed in the direction of semi-supervised learning. In most cases, labeled target data from the healthy state must be available. Since these approaches need the information that the given target data are from the healthy state for supervised training, these approaches are assigned to inductive transfer learning. Li and Zhang [155] used labeled source data and available healthy target data for the initial supervised training. In addition, the distributions of the healthy state source and target data were aligned with MMD. Furthermore, adversarial training was implemented to achieve prediction consistency between multiple classifiers on the target data. Thus, a discriminator attempted to determine from which classifier a prediction originated. In contrast, the classifiers attempted to fool the discriminator. This approach was applied to the condition diagnosis of bearings and a shaft crack test rig. Cody et al. [54] used labeled healthy target data of actuator systems for domain alignment but also for supervised training of a condition diagnosis model. Training was performed with healthy and faulty source data and healthy target data. A more advanced approach was also presented that used only faulty source data and healthy target data.

In additional approaches, attempts have been made to generate artificial faulty target data out of the available healthy data by utilizing the fault knowledge of the source data (A5). For example, no failure data are typically available from new systems. However, data from the healthy state can be collected with relatively little effort. Then, the faulty data of the source can be used to generate artificial faulty target data. For this purpose, data from the undamaged target system must be available. Popular networks to generate artificial data are called generative adversarial networks (GANs). Similar to DANNs, the concept of GANs is based on adversarial training. A data generator generates new artificial samples, and a discriminator attempts to distinguish these artificial data from the real data. The generator adjusts the artificial data to fool the discriminator. In transfer learning, it is also common to minimize the discrepancy between the source and target data, e.g., with the MMD. In the PHM literature, this approach or similar approaches have been used, e.g., for the condition diagnosis of photovoltaic systems [156] and hard disks [157]. Xie and Zhang [158] also presented an approach based on a GAN applied to the condition diagnosis of bearings. The transfer between different health states is the focus. The goal is to transfer knowledge from one health state under several different loads to another health state from which only data under one load are available. Thus, artificial target data under further loads are generated.

In transfer learning, there are other approaches that generate additional artificial data from existing data. For example, Fan et al. [159] used minority oversampling approaches to increase the amount of target data. The field of application was chiller condition diagnosis. In addition to this approach, classical data augmentation procedures can be used to increase the amount of data, as shown in [160]. This approach has been used in both inductive, e.g., [80], [129], and transductive, e.g., [161], [162], transfer learning. However, data augmentation is not a concrete transfer learning approach, which is why it will not be discussed further.

D. COMBINED APPROACHES

It is also possible to combine several transfer learning methods. Typically, certain parts can proceed without labeled target data, while others need them. However, in the approaches described below, there is always at least one part that requires labeled target data, which is why these approaches are listed under inductive transfer learning. Nevertheless, for the purpose of completeness, the following text additionally refers to the respective transductive parts from Fig. 12.

In [163], the MMD was used to align the feature distributions of the source and target domains. In addition, parameter transfer was performed (A1, B2.1). The application was condition diagnosis of bearings. It should be noted that, in addition, a completely transductive approach based on MMD was presented. Zhou et al. [144] presented a combination of parameter-based transfer, instance-based transfer, and feature alignment approaches (A1, A2, B2.1) on gas turbines to improve the accuracy of dynamic simulations of gas turbines. Wu et al. [164] addressed, among other things, the transfer of features using parameters (A1). This approach was used for the condition diagnosis of bearings and gearboxes.

Additional combined approaches for condition diagnosis were presented in [165] for wind turbines (A1, A2), [166] for ball screws (A1, A2), [143] for industrial robots and spot welding (A1, A2), [167] for bearings (A3, A4), and [168] (A1, B2.1) to estimate the health state of cutting tools. A condition prognosis approach was presented in [169]. The authors used a typical DANN structure from transductive transfer learning. In the second step, they fine-tuned the model with labeled target data to extend it to other fault modes (A1, B2.2).

At this point, reference should also be made to an approach for determining the state of charge of batteries. Qin et al. [170] used canonical variate analysis (CVA) for feature extraction in the state of charge estimation of lithium-ion batteries under different ambient temperatures. CVA reduces the dimensionality and maximizes the correlations between the source and target datasets. With the obtained features, an LSTM is trained. During online operation, if the distortion becomes excessive due to temperature fluctuations, a model update is performed using collected data at new temperatures.

There are also other inductive transfer learning approaches without a specific categorization. In addition to feature alignment (A3), Wang et al. [171] implemented a prototype learning approach. The aim was to learn domain invariant prototypes of the individual classes. For the classification of new samples, the class of the nearest neighbor was selected. This approach was applied to the condition diagnosis of bearings. Zhang and Gao [172] presented supervised dictionarybased transfer subspace learning. First, the source and target data were projected into a common subspace, whereby it was possible to represent all data by a shared dictionary matrix. This method was applied to the condition diagnosis of sucker rod pumping systems. Huang et al. [173] also used a diagnosis approach based on transfer dictionary learning applied to wind turbine systems.

Condition prognosis approaches were presented in [174] and [175]. Chehade and Hussein [174] used a collaborative Gaussian process regression to transfer between different battery cells, whereby it was possible to extrapolate the degradation of the capacity of new batteries at their beginning of life. Ma et al. [175] generated an artificial complete degradation trajectory of a target fuel cell until EOL. For this purpose, they used an SAE, which they trained with a complete degradation trajectory of a similar source fuel cell and the current trajectory of the target fuel cell that had not yet reached its EOL.

E. SUMMARY OF THE APPROACHES

As seen from Table 3, bearing and gearbox applications are most commonly used in the literature to evaluate inductive transfer learning-based condition diagnosis and prognosis approaches. Therefore, vibration signals are primarily considered. In addition to these and other mechanical component applications, there are also electrical and electronic component applications, as well as applications for more complex systems. For example, current and voltage signals are often used, and especially in more complex systems, multiple signal types are considered. As further discussed in Section X, there are already approaches that look at similar systems. However, the current focus is on identical systems under different operating conditions.

The main inductive transfer learning concepts used for condition diagnosis and prognosis approaches are listed in Fig. 12, with parameter transfer being the most common. Although many inductive transfer learning approaches, such as parameter-based and instance-based transfer, are in principle suitable for both condition diagnosis and prognosis, it should be pointed out that condition diagnosis applications are the most considered applications, as seen from the listing of the main concepts in Table 4. A major reason for this may be that the classical application fields of transfer learning, such as computer vision or natural language processing, are mostly classification problems. Therefore, it is convenient to apply these approaches to condition diagnosis, which is essentially also concerned with classification, while condition prognosis is mostly a regression approach. This view is also held by Mao et al. [19].

VII. TRANSDUCTIVE TRANSFER LEARNING APPROACHES

According to Table 3 and confirmed by Moradi and Groth [37], transductive transfer learning is the most commonly used PHM transfer learning approach for condition diagnosis and prognosis. The challenge is that there are no labeled target data. This means that although data are available from the system under consideration, they are not labeled with fault classes or RUL values. The supervised learning procedures must therefore proceed with the labels from the source. In transductive transfer learning, the feature-representation-based transfer approach is very common. There are several popular methods, which are described in Sections VII-A to VII-C. In addition, other feature-representation-based and transfer approaches exist (Section VII-D). Combined approaches are presented in Section VII-E. The discussion on transductive transfer learning is concluded with an interim summary in Section VII-F.

A. FEATURE ALIGNMENT BY THE MAXIMUM MEAN DISCREPANCY

One of the main approaches in transductive transfer learning is feature alignment by means of similarity measures (B1, B2.1). As already mentioned in Section VI, the MMD is widely used for this purpose. Therefore, a separate section is provided for approaches using MMD. In principle, however, many of the methods presented here can also be performed with other similarity measures. For orientation, thus, references are made to the structure used in Fig. 12, which is subdivided by the concepts of the methods.

Using the MMD, the distance between two probability distributions of two datasets can be measured. In this context, the MMD can be defined as the difference between two feature centers (means) of two samples [176]. Depending on the class of smooth functions used, different variants of MMD are possible. The unit ball in a reproducing kernel Hilbert space, as explained in [176], [177], and [178], is popular. Due to the strong importance of MMD in the field of transfer learning, it will be formally defined here. This definition is largely adopted from [177] and [178].

Let $\mathcal{F}$ be a class of functions f: $\mathbf{X} \rightarrow \mathbb{R}$ and p and q be Borel probability distributions. Let $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{m_1}\}$ and $\tilde{X} = \{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_{m_2}\}$ be samples composed of independent and identically distributed observations drawn from p and q, respectively. The MMD and its empirical estimation can be defined as

$$\mathrm{MMD}[\mathcal{F}, p, q] := \sup_{\mathrm{f} \in \mathcal{F}} (\mathbf{E}_{\boldsymbol{x}}[\mathrm{f}(\boldsymbol{x})] - \mathbf{E}_{\tilde{\boldsymbol{x}}}[\mathrm{f}(\tilde{\boldsymbol{x}})]) \tag{2}$$

and

$$\mathrm{MMD}[\mathcal{F}, X, \tilde{X}] := \sup_{\mathrm{f}\in\mathcal{F}}\left(\frac{1}{m_1}\sum_{i=1}^{m_1}\mathrm{f}(\boldsymbol{x}_i) - \frac{1}{m_2}\sum_{i=1}^{m_2}\mathrm{f}(\tilde{\boldsymbol{x}}_i)\right). \quad (3)$$

If $\mathcal{F}$ is defined as a unit ball in a universal reproducing kernel Hilbert space $\mathcal{H}$, which is defined on the compact metric space $\mathbf{X}$, an unbiased empirical estimate of the squared MMD can be calculated by the equation

$$\begin{aligned}\mathrm{MMD}_u^2[\mathcal{F}, X, \tilde{X}] = & \frac{1}{m_1(m_1-1)}\sum_{i=1}^{m_1}\sum_{j\neq i}^{m_1}\mathrm{k}(\boldsymbol{x}_i, \boldsymbol{x}_j) \\ & + \frac{1}{m_2(m_2-1)}\sum_{i=1}^{m_2}\sum_{j\neq i}^{m_2}\mathrm{k}(\tilde{\boldsymbol{x}}_i, \tilde{\boldsymbol{x}}_j) \\ & - \frac{2}{m_1 m_2}\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\mathrm{k}(\boldsymbol{x}_i, \tilde{\boldsymbol{x}}_j), \end{aligned} \quad (4)$$

where k is a characteristic kernel (e.g., Gaussian or Laplacian) [179]. The kernel provides a measure of how similar the sample elements are. If the similarity is high, the kernel value is high. For example, if the difference between the sample elements of the samples X and $\tilde{X}$ is large (subtrahend is small) and the difference between the elements within X and $\tilde{X}$ is small (summands are large), the MMD value is large.

A simple way to use MMD is to evaluate the similarity of degradation datasets. Then, for example, the most similar datasets can be selected as training data. Such approaches are presented in [180], [181], and [182]. However, by actively minimizing the MMD, source and target distributions can also be aligned. Depending on the specific use of the MMD to align the distributions, a distinction is made between different procedures.

1) TRANSFER COMPONENT ANALYSIS AND JOINT DISTRIBUTION ADAPTATION

One popular MMD-based approach is called transfer component analysis (TCA), which was introduced by Pan et al. [183]. It attempts to minimize the distance between the source and target marginal probability distributions, which is measured by the MMD. This means that a transformation must be determined for which $\mathrm{f}(P(X_s)) \approx \mathrm{f}(P(X_t))$. By using the kernel trick, the MMD is expressed by a kernel and coefficient matrix. On this basis, a transformation matrix can be found that maps the source and target data into a feature space in which the distance between the distributions is minimized. In addition, a second optimization constraint aims to preserve or maximize the variance of the source and target data in this new feature space [184]. Fig. 8 illustrates TCA visually. The TCA approaches used in PHM include those by Chen et al. [185] and Xu et al. [147] on condition diagnosis of bearings, Xie et al. [186] on gearboxes, and Xiao et al. [187] on bearings and induction motors. Mao et al. [19] predicted the RUL of bearings.

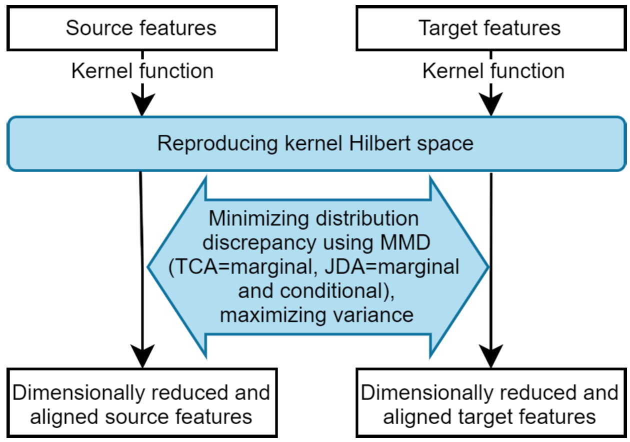

FIGURE 8. Visualization of TCA and JDA. Adapted from [44, p. 2].

Joint distribution adaptation (JDA) is another MMD-based feature extraction approach [188] and is also shown in Fig. 8. When using JDA, the marginal and conditional distribution differences of the source and target are minimized by mapping the data in a latent space, i.e., $\mathrm{f}(P(X_s)) \approx \mathrm{f}(P(X_t))$ and $\mathrm{f}(P(Y_s|X_s)) \approx \mathrm{f}(P(Y_t|X_t))$. Unlike the approximation of the conditional target distribution with available target labels in inductive transfer learning, no target labels are necessary in JDA. Instead, a classifier trained on the source data is used to assign pseudo labels to the target data. Applications of JDA, similar methods, or further developed approaches, e.g., balanced distribution adaptation for condition diagnosis, can be found in [189], [190], [191], and [192] on bearings, [193] on bearings and gearboxes, [194] additionally on wind turbines, and [195] on structures. Approaches for condition prognosis are presented in [196] on bearings and [197] on gearboxes of wind turbines.

As shown in [188], TCA is in some ways a special case of JDA that considers only the marginal distributions and not labels, i.e., conditional distributions. In addition to JDA, there are extensions of TCA that also use pseudo labels to approximate the conditional distributions. Ma et al. [198] used weighted TCA (WTCA) to align the conditional distributions of the source and target. Thus, the MMD was minimized class by class. For this purpose, pseudo labels were used for the unlabeled target data. These pseudo labels were assigned by a classifier trained on the transformed labeled source data. The alignment was performed iteratively. During each iteration, the transformation matrix and the classifier were adapted. This approach was validated on bearing condition diagnosis. van de Sand et al. [199] realized conditional alignment by adapting the decision boundaries of the classifier to the target domain. First, a classifier was trained by the TCA-transformed labeled source data; then, some of the labeled source data were replaced by pseudo labeled target data, and the training of the classifier was continued. This approach was used for the condition diagnosis of chillers.

2) DEEP ADAPTATION NETWORKS

TCA and JDA are feature-representation-based approaches that attempt to align the domains by identifying a common feature space in which the domain discrepancy is minimized. In this common space, a classification or regression model trained with the source data can be applied to the target data [185], [193]. In both approaches, domain alignment during feature extraction and the training of the classification or regression model are completed in a sequential manner; the generation of pseudo labels (if any) occurs recursively. In contrast, in a deep adaptation network (DAN), domain alignment and the training of the classification or regression model are performed simultaneously [200]. For this purpose, these networks integrate layerwise MMD terms into the loss function

$$\mathcal{L} = \mathcal{L}_{CR} + \lambda \mathcal{L}_{MMD}. \tag{5}$$

$\mathcal{L}_{CR}$ is the loss function of the classification or regression accuracy, $\mathcal{L}_{MMD}$ is the MMD term, and λ is a weighting parameter.

Thus, in addition to the classification or regression task, the network learns to find transferable features between the source and target domains simultaneously during training by minimizing the MMD. Typically, a CNN with attached fully connected layers is used as the basic network type. In contrast to parameter-based transfer in inductive transfer learning followed by fine-tuning with labeled target data, the DAN approach does not require labeled target data. The convolution layers of the DAN are simply taken over during transfer. The convolution layers, which are located deeper in the network, extract high-level features that may already be domain dependent. Therefore, if available, fine-tuning with possibly existing labeled target data can be beneficial, but this is omitted in the case of transductive approaches. The fully connected layers at the back end of the DAN are highly adapted to the specific domain and therefore not easily transferable from the source to the target. The approach taken during training of the DAN is therefore to ensure that the domain discrepancy in these layers is minimized. The upper branch in Fig. 9 is adapted with the help of the labeled source data. If labeled data from the target domain are available, they can be used to train the lower branch. In addition, the MMD between the distributions of the layers in the upper and lower branches is minimized. Accordingly, the layers of the two domains can be aligned. Thus, the source data indirectly help to train the target layers [200].

Networks of this functional principle are also used in PHM, even if the designation is not uniform in the literature. Examples include Zhu et al. [201], who used condition diagnosis on bearings or Lin et al. [202], who detected structural damage. Wu et al. [203] and Yu et al. [204] presented a slightly altered DAN architecture combining a CNN with LSTM layers and applied it to RUL prognosis of aircraft engines.

3) OTHER MMD APPROACHES

Other approaches that include MMD terms in the training or fine-tuning loss function can be found in [50], [163], [205], [206], [207], [208], [209], and [210] on condition diagnosis of bearings and in [211] and [212] on gearbox condition diagnosis. Other condition diagnosis approaches are presented in [213] and [214] on bearings and gearboxes, [215] on bearings and a crack rotating machinery dataset, and [216] on induction motors. Yang et al. [179], Yang et al. [217], and Guo et al. [218] added an MMD term and a pseudo label term to the loss function. They applied their approaches to the condition diagnosis of bearings or reciprocating compressor valves. Tang et al. [219] presented a similar approach for the condition diagnosis of bearings and gearboxes. An approach for RUL prognosis of bearings was shown in [220].

B. FEATURE ALIGNMENT BY OTHER SIMILARITY MEASURES

In addition to the MMD, there are many more similarity measures in the literature that can be used for feature alignment (B1, B2.1). They are discussed in this section. However, MMD is by far the most frequently used measure in PHM approaches for condition diagnosis and prognosis. Wang and Jin [57] utilized the Wasserstein distance to minimize the domain distribution difference. This approach was applied to the condition diagnosis of a feedwater heater system of a coal-fired power generation unit. Liu et al. [221] also used the Wasserstein distance to calculate the distribution discrepancy of bearing condition diagnosis datasets. Zhao et al. [222] used both the MMD and Wasserstein distance for the condition diagnosis of bearings. Another similarity measure utilized in transfer learning is Kullback-Leibler divergence. Qian et al. [223] aligned the source and target distributions via high-order Kullback-Leibler divergence. This approach was utilized for the condition diagnosis of bearings and gearboxes. Qian et al. [224] utilized correlation alignment (CORAL) to find domain invariant features, whereby the difference between the covariance matrices of the source and target domains was minimized. This approach was applied to condition diagnosis of gearboxes. An et al. [161] also used a CORAL loss term for the condition diagnosis of bearings. Central moment discrepancy (CMD) can also be applied to measure and minimize the domain discrepancy. Li et al. [225] used CMD for the condition diagnosis of bearings and Xiong et al. [226] used it for the condition diagnosis of gearboxes. Wang et al. [227] measured the similarity with Pearson correlation coefficients to adapt the length of the source and target time windows. This approach is necessary to classify faults in an electric power plant due to the different sampling rates of the source and target domains. Another metric for similarity measurement is the maximum variance discrepancy (MVD). It is based on the same principle as MMD but focuses on second-order statistics instead of first-order statistics. PHM approaches were presented by Zhang et al. [228] and Zhang et al. [229] for the condition

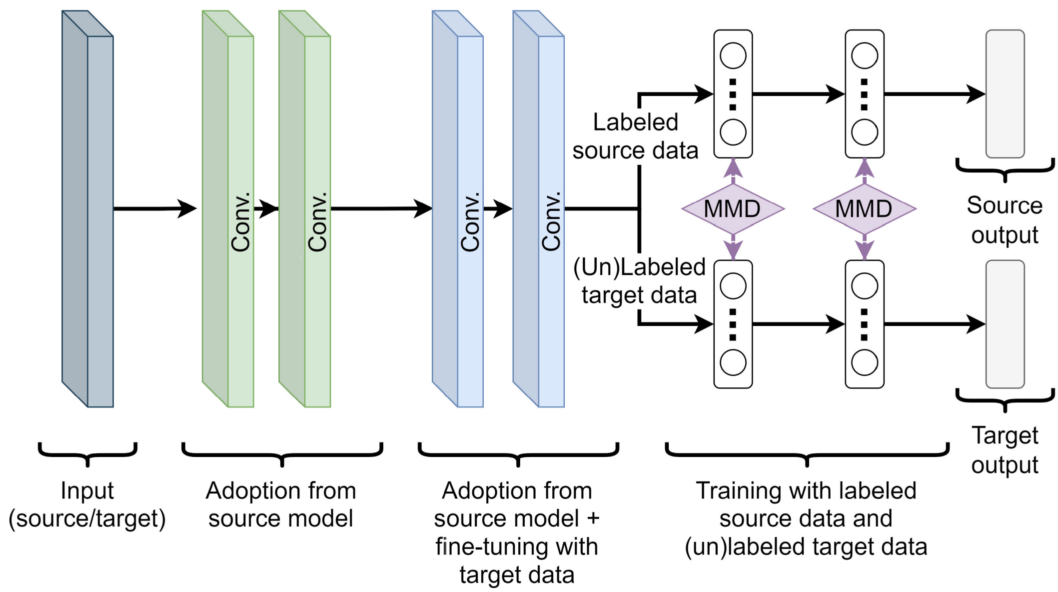

FIGURE 9. Structure of a DAN. Adapted from [200, p. 3].

diagnosis of bearings and gearboxes. Other measures used in PHM to compare the similarity between datasets are the cosine distance [137], [144], the log-Euclidean metric of second-order statistics [230], and the a-distance [192].

In some approaches, multiple similarity measures were used simultaneously or in combination. As already mentioned, Zhao et al. [222] applied the MMD and Wasserstein distance for condition diagnosis. In addition, Cao et al. [231] used these two measures for RUL prognosis of bearings. Ma et al. [232] determined the most similar degradation trajectory in historical test-to-failure data (source domain). This approach considers battery degradation and predicts its RUL. The similarity between historical samples and the actual degrading battery data from which the RUL has to be determined is measured by several measures that compare the capacity degradation trajectories. The idea is that further degradation of the actual battery will mainly follow the trend of the most similar batteries. The use of knowledge from the most similar degrading batteries is undertaken by parameter- and instancebased transfer. Liu et al. [233] used the Pearson correlation coefficient for similarity measures to select the appropriate source samples. Subsequently, a procedure similar to JDA was used to align the domains. The process was tested for functionality by means of condition diagnosis of wind turbine bearings. In addition to MMD for marginal domain alignment, [234] used pseudo labels to minimize the difference between the source and target conditional distributions. Based on an entropy penalty, the classification boundaries were shifted to low-density regions, and thus, the performance of the classifier in the target domain can be improved in an approach called classifier adaptation. Wang et al. [235] presented a similar concept. Both approaches were applied to condition diagnosis of bearings. Other condition diagnosis approaches using multiple similarity measures were presented in [54], [236], and [237]. Siahpour et al. [236] aligned the marginal distributions by MMD and the conditional distributions by the Manhattan distance between paired source and target data. Pandhare et al. [237] applied the Euclidean distance on paired source and target data for conditional alignment. Cody et al. [54] used the Euclidean distance and the maximum absolute difference. The first distance was calculated between the medians of the marginal samples, and the second was calculated between their empirical cumulative distribution functions.

Dong et al. [238], Dong et al. [239], and Dong et al. [240] presented joint geometrical and statistical alignment (JGSA). In this approach, two transformation matrices that project the source and the target data into a common feature space are learned. The source matrix is optimized to maximize the variance between the classes of the source and minimize the variance within each class. The target matrix is optimized to minimize the variance in the unlabeled target data. Furthermore, MMD is used to adapt both matrices to align the marginal and conditional distributions. For the latter, pseudo labels are used in the target domain. Additionally, the difference between the two matrices is penalized. Yu et al. [241] combined JGSA and sparse coding with an integrated MMD penalty term. All approaches were applied to condition diagnosis of bearings.

Other condition diagnosis approaches include those proposed by Pang et al. [242] and Gardner et al. [243]. Pang et al. [242] used MMD and manifold regularization to align marginal and conditional domain distributions. In addition to TCA and JDA, [243] applied adaptation regularization-based transfer learning (ARTL) [244]. ARTL combines MMD-based marginal distribution adaptation, conditional distribution adaptation by pseudo labels, and manifold regularization. Zhang et al. [245] presented a multiple alignment of source and target data for RUL prognosis of bearings and aircraft engines. First, the difference between the healthy state data of the source and target is reduced by minimizing the variance. In the next step, the degradation data are projected on an identical straight line in a high-level subspace with several alignment measures, including MMD minimization. If this succeeds, the degradation information of the source can also be used in the target.

In addition to the similarity measures listed, there are others that are used in transfer learning but for which no PHM applications were found. Table 2 gives a brief overview of some popular measures.

C. ADVERSARIAL APPROACHES

In addition to the similarity-measure-based methods presented above, adversarial approaches are alternative approaches for minimizing the deviation between the source and target distributions (B2.2). Accordingly, DANNs are one of the most popular PHM approaches for condition diagnosis and prognosis. DANNs, which follow the same idea as DANs, simultaneously determine a common feature space and train the machine learning model. Both networks combine these two steps through deep learning, which provides the possibility to choose the feature space in such a way that it minimizes the domain discrepancy and is also appropriate for the actual machine learning task. This may prevent the actual machine learning task from becoming more difficult or impossible in the new common feature space [246], [247]. The underlying operating principle of DANNs originates from Ajakan et al. [248] and Ganin et al. [249]. They showed that the DANN approach is applicable to any feedforward architecture that can be trained with backpropagation.

Fig. 10 illustrates the structure of a DANN using the example of a CNN architecture. Similar to classical feedforward deep networks, DANNs comprise a feature extractor G_f and a label predictor G_y. Both are adjusted such that G_y can fulfill the prediction task on the source data as accurately as possible. DANNs were initially developed as classifiers, which explains why they are currently mainly used for classification tasks in the field of PHM. In addition to G_f and G_y, a domain discriminator G_d is added as a third component. The features are chosen by G_f in such a way that G_d cannot distinguish between domains based on the feature values, whereas G_y can still separate the classes of the source well. This is achieved by the adversarial training of G_f and G_d as well as the classic feedforward training of G_f and G_y. In the former, G_d is adapted to distinguish the domains and G_f to find a feature space in which G_d cannot distinguish them. Ideally, multiple iterations result in a G_f under whose features the domains cannot be separated, even by a very good G_d. The idea behind this is equivalent to the similarity measurement approaches; if the source and target domains do not differ in the chosen feature space, the model trained with the source data can probably also be applied to the target data. In summary, the loss of a DANN can be described as [250]

$$\mathcal{L}_{DANN} = \mathcal{L}_y(G_y(G_f(X_s)), Y_s) - \lambda \mathcal{L}_d(G_d(G_f(X_{s+t})), D_{s+t}), \tag{6}$$

where $\mathcal{L}_y$ quantifies the label prediction loss over all source data X_s, and $\mathcal{L}_d$ quantifies the domain classification loss over all source and target data X_{s+t}. Y_s are the true source labels of X_s, and D_{s+t} are the true domains of X_{s+t}. The training process can be formulated as

$$(\hat{\theta}_f, \hat{\theta}_y) = \arg\min_{\theta_f, \theta_y} \mathcal{L}_{DANN}(\theta_f, \theta_y, \hat{\theta}_d) \tag{7}$$

$$\hat{\theta}_d = \arg\max_{\theta_d} \mathcal{L}_{DANN}(\hat{\theta}_f, \hat{\theta}_y, \theta_d), \tag{8}$$

where θ_f, θ_y and θ_d are the parameters of the feature extractor, the label predictor, and the domain discriminator, respectively.

Although the DANN approach is still relatively new, there are already several applications in the PHM literature; however, some of them are referred to by different names or use slightly different network architectures. Liu et al. [250] built a DANN with an SAE as a feature extractor applied to the condition diagnosis of bearings. Wang et al. [50] used a DANN with a CNN feature extractor for the condition diagnosis of bearings. Other CNN-based DANNs were used to diagnose the health of rolling bearings, as presented in [149] and [251]. In addition to bearings, Wang et al. [252] applied a DANN based on a CNN architecture for the condition diagnosis of hard disks. Zhu et al. [253] used a DANN built on an MLP. This approach was applied to the condition diagnosis of building chillers. For RUL prognosis, da Costa et al. [16] presented a DANN that uses an LSTM as a feature extractor. With this LSTM-DANN, information from time series data in a source domain with observed RUL values was used to estimate the RUL values for the unlabeled target data. This network was applied to engine data. Liu and Gryllias [254] used a similar approach for the RUL prognosis of bearings. Another approach for RUL prognosis can be found in [255]. A DANN was applied with convolution layers for the RUL prognosis of bearings.

DANNs with different architectures were described in [12], [256], and [257]. These DANNs were used for the condition diagnosis of bearings and wind turbine gearboxes and the RUL prognosis of aircraft engines. There are also other adjusted DANNs. Among other optimization approaches, Yu et al. [258] integrated label predictions into the input of the domain discriminator. With this adaptation of the DANN structure, the conditional distribution discrepancy of the domains can also be minimized. Yu et al. [259] separated the training process into two stages. In the first stage, a classifier and a source feature extractor are trained with source data. In the next step, adversarial training is performed with two feature extractors, one for the source domain and one for the target domain. Both approaches are applied to condition diagnosis. Deng et al. [260] considered a condition diagnosis

TABLE 2. Similarity measurement methods. Adapted from [232, p. 4].

Category	Method	Measurement object
Distance	Mahalanobis distance, city block distance, Minkowski distance, Euclidean distance	Distance between vectors
Similarity	Cosine similarity Pearson correlation coefficient Jaccard similarity	Angle between vectors Correlation between random variables Correlation between sets
Information entropy	Mutual information, Kullback–Leibler divergence, Jensen–Shannon divergence	Distance between probability distributions
Distance in mapping space	Maximum mean discrepancy (MMD), principal angle, a-distance Hilbert–Schmidt independence criterion	Distance between probability distributions Independence of two sets of data
Transmission distance	Wasserstein distance	Distance between probability distributions

problem with different source and target label spaces, which is referred to as partial transfer learning. In this case, only a few of the source classes exist in the target label space. Without any adaptation, any transfers would fail because some source classes have no counterpart in the target label space. Therefore, several subdomain discriminators are used. This means that, in contrast to the classic DANN, a separate subdomain discriminator is used for each source class. Each discriminator attempts to keep the source and target data of a class apart from each other. Another difficulty is that, as is generally the case in transductive transfer learning, the target data are unlabeled. An additional approach for partial adversarial transfer learning was presented in [261] and applied to the condition diagnosis of bearings. To make the task more difficult, gearbox data were included.

Li et al. [148] used a DANN with multiple classifiers as the label predictor. In addition to the adversarial training of the feature extractor and the label predictor, a second adversarial training process is thus possible (B2.2). Thus, the discrepancy between the classifiers is maximized by adjusting their parameters. However, at the same time, the feature extractor is adjusted to minimize this discrepancy. Through this second adversarial training, an additional conditional domain alignment is achieved. The application area is the condition diagnosis of bearings and of a crack test rig. Jiao et al. [262] presented a similar approach using two classifiers to further refine a common feature space found with a DANN. Different outputs of the classifiers on a target sample were treated as an indication that the sample had not yet been adapted correctly and, for example, was between two source classes. The approach was applied to condition diagnosis of bearings and gearboxes.

Other condition diagnosis approaches that integrate multiple classifiers into the DANN structure but do not perform adversarial training between them and the feature extractor were proposed in [263] and [264] (B2.3). Liu et al. [263] integrated two classifiers in a DANN structure. The classifiers were trained with different features to obtain two classification boundaries. In this way, the samples in the target domain for which the classifiers disagree could be identified. Zhang et al. [264] used a DANN architecture with multiple classifiers, one for each of the multiple source domains. This approach was used to adjust the feature extractor and the classifiers to minimize the classification difference of the classifiers.

It is also possible to adversarially train multiple classifiers with a feature extractor without using a domain discriminator, as in a DANN (B2.2). Zhao et al. [265] used two classifiers trained in three training steps, which are iteratively repeated. First, the classifiers are trained to minimize the classification error on the source data. In the next step, a discrepancy term is added to the loss function, which promotes differences between the classifiers on the target data. Finally, the feature extractor is adapted such that the classification difference between the two classifiers on the target data is minimized. Through the adversarial training processes in steps two and three, features are extracted for which the outputs of the two classifiers on the target data are as identical as possible despite a maximum difference of the classifiers on these data. This approach was used to determine the condition of bearings and gearboxes. Yu et al. [266] introduced a similar adversarial approach based on two classifiers for the condition diagnosis of bearings. First, the feature extractor and both classifiers are trained with source data. Subsequently, the training is continued in an adversarial manner. The classifiers are trained such that their disagreement on the target data increases, and the feature extractor is adapted to minimize this disagreement.

D. OTHER APPROACHES

As with inductive transfer learning, there are some parameter-based transfer approaches (B3) to transductive transfer learning, but in much smaller numbers. One such approach is adaptive batch normalization (AdaBN) [301]. The benefit of this approach is that a trained source network can be adapted to a target domain by simply changing the AdaBN parameters. No retraining is necessary. AdaBN is based on the same idea as batch normalization (BN), although it is applied to multiple datasets of different domains. BN normalizes the output of the activation functions

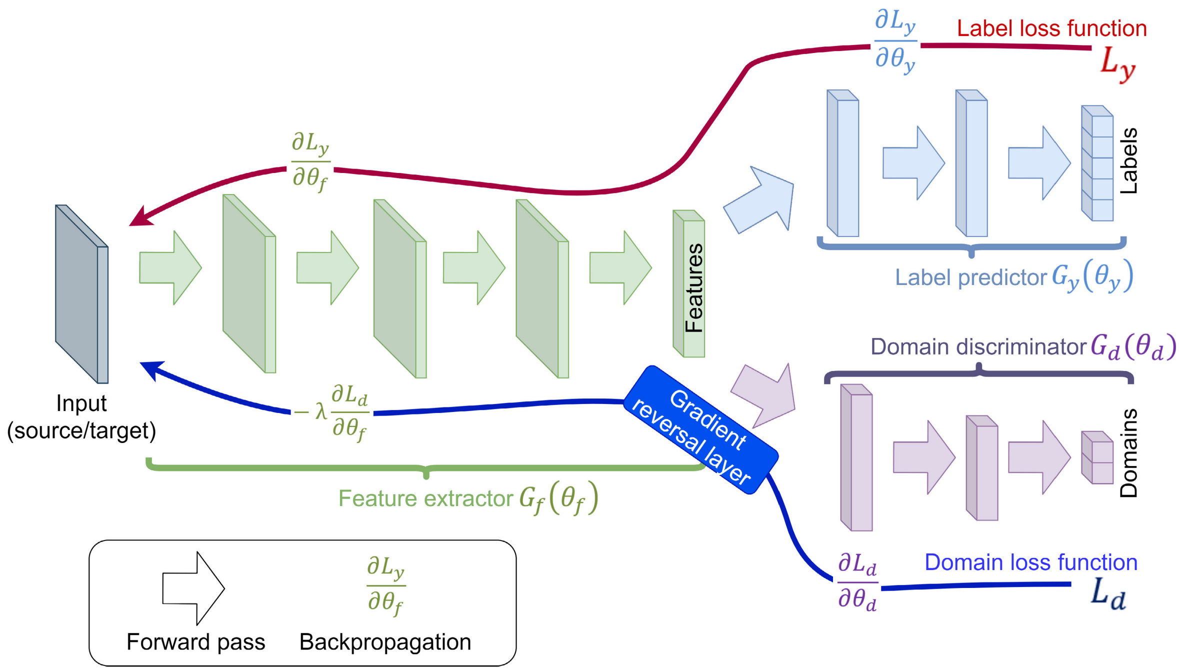

FIGURE 10. Structure of a DANN. Adapted from [249, p. 1182].

of neurons in each layer using BN statistics (mean and variance). Thus, the input of each subsequent layer is normally distributed. This normalization is also useful for domain alignment. Typically, there is a distribution discrepancy between the source and target data on the intermediate layers. This discrepancy causes a network trained on the source data to perform worse on the target data. By normalizing with AdaBN, each layer receives data from a similar distribution, regardless of whether it comes from the source or target domain. This minimizes the distribution discrepancy and improves the performance on the target data. Specifically, a network is first trained using the labeled source data and then transferred to the target domain by adjusting the BN statistics. There are applications of AdaBN in PHM, e.g., on condition diagnosis of bearings and gearboxes [50], [268] and on RUL prognosis of aircraft engines [295].

Other parameter-based transfer approaches exist in transductive transfer learning. Shen et al. [292] presented a penalty domain selection machine (PDSM) for gearbox condition diagnosis. This approach is based on a domain selection machine [62]. The basic idea is to select the most relevant source domains from multiple source domains for the target domain. Each classifier trained with the data from a selected source domain is then combined, and an adaptation loss function is added to obtain the target classifier. PDSM extends the approach by domain penalty and signal penalty factors. This gives stronger importance to the new samples and sensors close to the fault location. There are other parameter-based transfer approaches for the condition diagnosis of bearings [269], [270].

In addition to parameter-based transfer approaches, there are also instance-based transfer approaches (B4). In the field of transductive transfer learning, however, these are usually combined with other approaches, so references to them are presented in Section VII-E, i.e., the combined approaches.

Originally presented in [302] and [303], consensus self-organized modeling (COSMO) can also be used for transductive transfer learning to find transferable features (B5). The underlying idea is to compare single individuals with the set of all units. It is assumed that most units are healthy. As a result, as in the original approaches and in [304] and [305], unsupervised anomaly detection for fault detection can be implemented. However, there is another possible application of COSMO for feature selection in the area of transfer learning, as explained by Fan et al. [296]. First, two reference groups are created, one for the source domain and one for the target domain. In each case, only the healthy system data of the domains form the reference group. Instead of using only source data for the source reference group and only target data for the target reference group, it is also possible to mix samples from the source and target domains. After the reference groups are formed, the aim is to generate new features that represent the difference between the systems and their corresponding reference group. With these new features, a machine learning model is trained using labeled source data. Then, the model can be transferred to the target data by parameter transfer. This approach was applied to the RUL prediction of aircraft engines.

There are also approaches using kernels for domain alignment based on geodesic flow kernels [306], [307] (B6). In these approaches, the datasets of the source and target

TABLE 3. Application examples of transfer learning for condition diagnosis and prognosis in PHM. (doc = different operating conditions, sis = similar systems, * = key focus on faults, — = no literature found).

Application on		Transfer learning					
		Inductive		Transductive		Unsupervised	
Type	Signals	Simulated	Measured	Simulated	Measured	Simulated	Measured
Bearings							
- doc:	vibration	—	[23], [24], [43], [44], [66], [101], [113], [115], [123], [130], [131], [137], [145]–[147], [150], [153]–[155], [171], [110]*, [125]*, [138]*, [158]*, [164]*, [100]*	—	[50], [58], [147]–[149], [161]–[163], [185], [189]–[193], [196], [198], [201], [205]–[210], [213]–[215], [219]–[223], [225], [228]–[231], [234], [235], [238]–[242], [245], [250]–[252], [254]–[256], [258]–[260], [262]–[283], [194]*, [284]*	—	[45], [46], [285]
- sis:	vibration	[167]	[24], [66], [99], [112], [113], [122], [149], [151], [152], [163], [167], [171]	—	[30], [58], [149], [179], [198], [217], [260], [261], [278], [283], [284], [286]–[289]	—	—
Gears/gearboxes							
- doc:	vibration	—	[24], [112], [120], [123], [124], [127], [145], [146], [154], [164]	—	[186], [193], [211]–[214], [219], [223], [224], [226], [229], [242], [262], [264], [265], [267], [268], [272], [280], [289]–[291], [194]*	—	—
	sound	—	—	—	[187]	—	—
- sis:	vibration	—	[24], [108]	—	[261], [292]	—	—
Ball screws							
- doc:	vibration	—	[132], [166]	—	—	—	—
- sis:	vibration	—	—	—	[237]	—	—
Tools for machines							
- doc:	images	—	[79]	—	—	—	—
	multiple	—	[169]*	—	[169]	—	—
Structures							
- sis:	vibration	—	—	[195], [202], [243]	[195], [202], [243]	—	—
Transistors							
- sis:	multiple	[128]	—	—	—	—	—
Transformers/rectifiers							
- sis:	temperature, oil flow	[80]	[80]	—	—	—	—
	voltage	[109], [111]	[109], [111]	—	—	—	—
Circuit breakers							
- sis:	current	[49]	[49]	—	—	—	—
Electric motors							
- doc:	vibration	—	[140]	—	[216]	—	—
	sound	—	—	—	[187]	—	—
Other actuators							
- doc:	multiple	—	[54]	—	—	—	—
- sis:	vibration	—	—	—	[236]	—	—
Other rotating components							
- doc:	vibration	—	[155]	—	[148], [215], [251], [286]	—	—
- sis:	vibration	—	[108], [152]	—	—	—	—
Tanks							
- doc:	vibration	—	[107]	—	—	—	—
Boilers							
- sis:	multiple	—	—	—	[293]	—	—
Batteries (V=voltage, I=current, C=capacity, Q=charge)							
- doc:	V	—	[116], [117]	—	—	—	—
	C	—	[174]	—	—	—	—
	V,I	—	[135]	—	—	—	—
	V,I,Q	—	[88], [105]	—	—	—	—
- sis:	C	—	[118]	—	[232]	—	—
	V,I,Q	—	[104]	—	—	—	—

TABLE 3. ***(Continued.)*** **Application examples of transfer learning for condition diagnosis and prognosis in PHM. (doc = different operating conditions, sis = similar systems, * = key focus on faults, — = no literature found).**

Application on		Transfer learning					
		Inductive		Transductive		Unsupervised	
Type	Signals	Simulated	Measured	Simulated	Measured	Simulated	Measured
Fuel cells							
- doc:	voltage	—	[175]	—	—	—	—
Photovoltaic systems							
- sis:	current	—	[156]	—	—	—	—
Components of power plants							
- doc:	multiple	[106]	—	—	[57]	—	[294]
- sis:	multiple	[106], [144]	[144]	—	[227]	—	—
Wind turbines							
- doc:	vibration	—	[151]	—	[194], [233], [257]	—	—
	multiple	—	[121], [126], [165], [173]	—	—	—	—
- sis:	vibration	—	—	—	[197]	—	—
Compressors							
- sis:	temperature, pressure	—	—	—	[218]	—	—
	multiple	—	[129]	—	—	—	—
Chillers							
- sis:	multiple	—	[159]	—	[199], [253]	—	—
Pumps							
- doc:	vibration	—	[23]	—	—	—	—
- sis:	force	—	[172]	—	—	—	—
Pump trucks							
- doc:	multiple	—	[121]	—	—	—	—
Aircraft engines							
- doc:	multiple	[134], [141]	[102]	[12], [16], [203], [204], [245], [295]–[297]	—	[45], [46]	—
- sis:	multiple	[129], [133]	[133]	—	—	—	—
Quadrotor drones							
- sis:	sound	—	[103]	—	—	—	—
Industrial robots							
- doc:	vibration	—	[143]	—	—	—	—
	torque	—	—	—	—	—	[298]
Locomotives / turnouts							
- doc:	vibration	—	—	—	[258]	—	—
	current	—	—	—	—	—	[299]
Cellular networks							
- sis:	radio signal	[142]	—	—	[300]	—	[300]
Disk systems							
- doc:	multiple	—	—	—	[252]	—	—
- sis:	multiple	—	[139], [157]	—	—	—	—

domains are embedded into a Grassmann manifold. By constructing a geodesic flow between the resulting points, infinite intermediate subspaces, gradually changing from the source to the target, are generated. By projecting the source and target features into each of these subspaces, feature vectors with infinite dimensions can be obtained. With a kernel defined by the inner product of these vectors, low-dimensional domain invariant representations can be learned. In this feature space, a model with labeled source data is trained and then applied to the target data. A PHM application was presented in [267] for the condition diagnosis of bearings and gearboxes.

Transductive transfer learning based on feature alignment is also possible with transfer factor analysis (TFA) (B7). TFA combines two FA models, one for the source domain and the other for the target domain. The factor loading matrix is shared, which enables the transfer between the source and target. In addition, each domain has its own noise covariance matrix that can cover the differences. With TFA, features can be found that reduce the difference between the source and target domains. Wang et al. [290] and Wang et al. [291] applied TFA for the condition diagnosis of gearboxes.

As previously mentioned with adversarial training using DANN, there are approaches that use the outputs of multiple classifiers for transductive transfer learning (B2.3). Another approach to this, but without adversarial training, was presented by Wen et al. [271]. They utilized multiple source domains for the condition diagnosis of bearings. The network comprises a common feature extractor for all source domains and the target domain. The extractor is based on a CNN pretrained with natural images. A domain-specific feature extractor and classifier are attached to this common extractor for each source domain. Each of these specific feature extractors attempts to find a feature space to align the respective source domain with the target domain. For this purpose, the MMD is used. It is expected that, especially for the target data near the classification boundaries, the source classifiers will predict different labels. Accordingly, the discrepancy between these classifiers on the target data is minimized. The final prediction on the target data is the averaged output of all classifiers.

E. COMBINED APPROACHES

To increase transfer performance, there are several approaches that combine multiple transductive transfer learning methods. The most common is the combination of adversarial DANN approaches (B2.2) and similarity-measure-based methods (B1, B2.1). PHM application examples include the condition diagnosis of bearings [30], [272], [273], [276], [286], gearboxes [224], [272], and fog radio access networks [300]. Mao et al. [275] and Mao et al. [274] presented other approaches for condition diagnosis and health index construction for bearings. Jia et al. [277] combined the adversarial training of two classifiers with a feature extractor using the minimization of a similarity measure to determine the condition of bearings.

Transfer joint matching (TJM) [308] combines feature alignment by MMD as in TCA (B1.1) and instance reweighting (B4). In the latter, the source samples are weighted by their relevance for the target domain. Zhang et al. [228] enhanced TJM by the additional use of MVD and applied the approach to bearing condition diagnosis. Zhang et al. [229] also assigned weights to the source samples to reduce the influence of irrelevant samples and used MVD as a similarity measure. In addition, a manifold regularization term was added. This method was applied to condition diagnosis of bearings and gearboxes.

Jin et al. [162] combined AdaBN with an MMD approach (B2.1, B3), and Jin et al. [58] added adversarial learning (B2.2, B3). Both considered the condition diagnosis of bearings. Combinations including a model transfer of networks pretrained with natural images (B3), which is well known in the field of inductive transfer learning, were presented in [278], [279], and [284]. The former two combined it with a similarity measure (B2.1, B3) and the latter additionally with adversarial training (B2.2, B3). The field of application is bearing condition diagnosis. A model transfer from source to target combined with MMD (B2.1, B3) was presented in [210] for condition diagnosis and with Kullback-Leibler divergence in [309] for RUL prognosis.

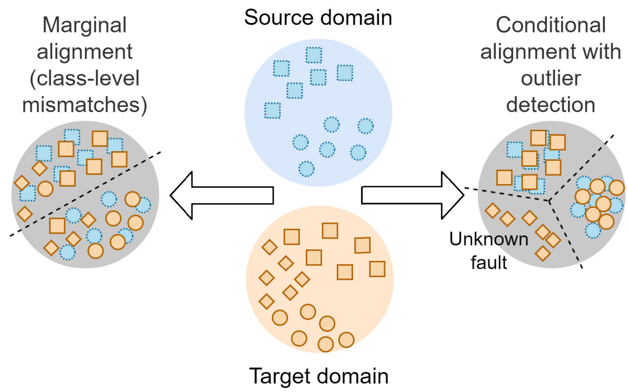

FIGURE 11. Partial transfer learning problem and solution through conditional alignment with outlier detection.

Other combined approaches that considered adversarial training (B2.2) include those in [283], [287], [288], [293], and [297].

Combined approaches for transductive transfer learning also play a major role in partial transfer learning. As previously explained, partial transfer learning is used if there are different label spaces. Specifically, the source label space can be a subset of the target label space and vice versa. Since at least some parts of the label spaces match, transductive transfer learning is applicable. Fig. 11 illustrates the problem. The goal is to detect the samples whose classes do not occur in both domains. Then, for example, domain alignment can be performed with only samples from the common classes. In PHM, this problem plays a major role, especially for condition diagnosis. For example, different types of faults can occur in engineering systems due to different operating conditions or deviating technical characteristics.

All approaches considered below deal with condition diagnosis. Combined approaches for partial transfer learning with a target label space that is a subset of the source label space are presented in [280] and [281]. In the PHM, this case occurs, for example, when data of the target system cannot be generated for all fault types due to limited capabilities. The applications are on condition diagnoses of bearings and gearboxes. Jiao et al. [280] trained two classifiers with source data. The target data are then classified with these classifiers. Source classes, for which much of the target data are assigned, probably also occur in the target label space. However, source classes for which no or very little target data are assigned probably do not occur in the target label space (outlier source data). Therefore, only the source samples whose labels probably also occur in the target data are mainly considered for the next transfer step. This is realized by weighting the source samples. It follows a domain alignment based on the inconsistency of the two classifiers and further training of the classifiers with the weighted source data (B2.3, B4). Zhang et al. [281] also assigned weights to the source samples, and the network structure resembled a DANN. This approach neglects the outlier classes of the source that do not exist in the target data during domain alignment.

For this purpose, each source sample is assigned a weight that indicates how similar the sample is to the target data. To determine the size of the weight, the output of the domain discriminator is considered. Because there will quite likely be a discrepancy between the outlier samples of the source and the target samples after alignment, the discriminator can distinguish the outlier source data well (B2.2, B4).

The other case of partial transfer learning is if the source label space is a subset of the target label space. This could occur if additional fault types occur at the target system that do not exist at the source system. Yang et al. [289] and Zhang et al. [282] considered this case and attempted to address it using combined transfer learning approaches. To achieve this, Yang et al. [289] added an additional output neuron that indicates the probability that a target sample does not belong to any source class. The alignment of the source and target domains takes place via the domain discriminator and MMD (B2.1, B2.2, B4). In addition, multisource fusion, which handles multiple source domains, is presented. Similar to Li et al. [281], Zhang et al. [282] assigned weights based on the output of a domain discriminator (B2.2, B4). The difference is that the weights are not assigned to the source data but are assigned to the target data. In addition, an outlier classifier is added that is trained to identify the target samples that have classes that are not known in the source.

F. SUMMARY OF THE APPROACHES

Table 3 shows that in condition diagnosis and prognosis, bearings and gearboxes are the main applications in transductive transfer learning, similar to inductive transfer learning. Here, too, mainly vibration signals are used. In addition, there are again further applications to mechanical, electrical, and electronic components as well as more complex systems. These include approaches that consider similar systems; however, the current research is focused on identical systems under different operating conditions (see Section X).

The main transductive transfer learning concepts are listed in Fig. 12. Feature alignment is the most popular. Although there are some approaches to condition prognosis in the field of transductive transfer learning, it must be emphasized that, similar to inductive transfer learning, the focus is on approaches to condition diagnosis, as seen from Table 4. Again, however, most approaches, such as feature alignment or adversarial training, can also be applied to condition prognosis.

A. Inductive transfer learning

- A1 Parameter transfer from the source domain and fine-tuning based on target data
- A2 Instance transfer with sample weighting using target labels (e.g., TrAdaBoost)
- A3 Feature alignment by similarity measures with respect to conditional distributions using source and target labels
- A4 Adversarial approaches (e.g., DANN, promote differences between classifiers)
- A5 Generation of artificial target data (e.g., GAN)

B. Transductive transfer learning

- B1 Preceding feature alignment by similarity measures
  - B1.1 Marginal distribution alignment (e.g., TCA, TJM)
  - B1.2 Marginal and conditional distribution alignment based on pseudo labels in the target domain (e.g., JDA, JGSA)
- B2 Feature alignment during the training process via the adapted loss function
  - B2.1 Measure similarity (e.g., DAN, ARTL)
  - B2.2 Adding adversarial components to the model structure (e.g., DANN, promote differences between classifiers)
  - B2.3 Minimize differences between classifiers
- B3 Parameter transfer from the source domain without fine-tuning with target data (e.g., AdaBN, DSM)
- B4 Instance transfer with sample weighting without target labels (e.g., TJM)
- B5 Forming reference groups and introducing relative features (e.g., COSMO)
- B6 Incremental learning of intermediate domains (e.g., geodesic flow kernels)
- B7 Separation of commonalities and differences among the domains (e.g., TFA)

C. Unsupervised transfer learning

Marginal feature alignment with ensuing anomaly detection or clustering

FIGURE 12. Main transfer learning concepts for condition diagnosis and prognosis in PHM.

VIII. UNSUPERVISED TRANSFER LEARNING APPROACHES

As noted by Moradi and Groth [37], in unsupervised transfer learning, there are only a few PHM approaches. Not much has changed since this publication, as seen from Table 3. Unsupervised learning tasks include anomaly detection, density estimation, clustering, and appropriate feature space generation. In this respect, domain alignment procedures for finding a common feature space that do not use label values to approximate the conditional distributions could be described as unsupervised. However, finding a common feature space alone is not a complete transfer learning approach that can be used to diagnose faults or predict the RUL. A diagnosis or prognosis procedure must also be added. This usually requires at least labeled source data (transductive transfer learning) and optionally additional labeled target data (inductive transfer learning). Nevertheless, there are few approaches that have proceeded without labels. Unsupervised anomaly-based approaches have the advantage that faults that did not occur during training can also be detected. In industrial systems, many potential causes for failure can occur, and in most cases, data are not available for every cause of failure (even when using data from similar systems). Therefore, supervised models can rarely cover all possible fault causes. Anomaly detectors have the advantage that they only need to be trained with healthy operational data. Then, if a case occurs in the application that deviates from the healthy training data, there is an alarm. However, there is the risk that data from similar systems or other operating conditions will also be falsely recognized as an anomaly and thus as a fault. Therefore, transfer learning approaches are indispensable in these cases.

Michau and Fink [294] was concerned with detecting anomalies in the early life stage of gas turbines. Before anomaly detection, the domains are aligned, whereby three possibilities are presented: autoencoding, homothety loss, and adversarial domain discrimination (C). For fault detection, a one-class classifier is trained with the healthy state data of the source and target domains. Mao et al. [285] aligned the domains with an autoencoder by including an MMD penalty term in the loss function. Anomaly detection is then carried out (C). This approach was applied to the fault detection of bearings. Guo et al. [299] used an unsupervised method for detecting faults in railway turnouts. The first step is to cluster the data to determine the fault-free modes. In the next step, both a global fault detection deep autoencoder model with data from all modes and local models for each mode are trained. In addition, in parameter-based transfer, knowledge is transferred from modes with much data to those with little data. Michau and Fink [45] and [46] presented anomaly detection approaches to detect faults of bearings and aircraft engines. They used adversarial training for domain alignment (C). Another unsupervised transfer learning approach was presented by Mahyari et al. [298], who utilized manifold alignment to align the domains (C). This approach was applied to industrial robots. Even though the main focus is on transductive transfer learning, Wu et al. [300] also presented a density-based spatial clustering approach for fault detection in fog radio access networks. This enables fault detection to be performed because healthy state data usually occur much more frequently than faulty data. Therefore, samples located in areas of the feature space with high density are very likely to be from healthy conditions. However, samples located in sparse areas are likely to be fault data.

As seen from the references listed in Table 4, all the approaches found in unsupervised transfer learning deal with condition diagnosis. Specifically, anomaly detection and clustering are used, usually combined with feature alignment (see Fig. 12). According to Table 3, applications include bearings, components of power plants, industrial robots, turnouts, and cellular networks. With one exception, all approaches consider identical systems, i.e., no similar systems.

In Sections VI to VIII, transfer learning approaches were considered. An overall summary of the findings is provided in Section XIII. In Section IX, the condition diagnosis and prognosis approaches of fleet learning are presented as other types of approaches that are also promising for similar system problems.

IX. FLEET LEARNING APPROACHES

Existing fleet learning approaches in the PHM field of condition diagnosis and prognosis are collated in this section.

TABLE 4. References to the main transfer learning concepts in Fig. 12 (Diagnosis = condition diagnosis, Prognosis = condition prognosis, — = no literature found).

A1	Diagnosis: [9], [18], [23], [24], [43], [66], [77]–[81], [83]–[88], [92]–[96], [99]–[113], [115], [120]–[128], [130]–[132], [135], [143], [163]–[166], [168] Prognosis: [97], [98], [116]–[118], [129], [133], [134], [144], [169]
A2	Diagnosis: [44], [49], [137]–[143], [165], [166] Prognosis: [144]
A3	Diagnosis: [145], [146], [155], [167], [171] Prognosis: —
A4	Diagnosis: [149]–[155], [167] Prognosis: —
A5	Diagnosis: [156]–[158] Prognosis: —
B1.1	Diagnosis: [147], [185]–[187], [228], [229] Prognosis: [19]
B1.2	Diagnosis: [189]–[195], [198], [233], [238]–[241] Prognosis: [196], [197]
B2.1	Diagnosis: [30], [50], [57], [161]–[163], [179], [201], [202], [205]–[219], [221]–[226], [230], [234]–[237], [243], [272]–[279], [284], [286], [289], [300] Prognosis: [203], [204], [220], [309]
B2.2	Diagnosis: [30], [50], [58], [148], [149], [224], [250]–[253], [256]–[266], [272]–[278], [281]–[283], [286]–[289], [293], [300] Prognosis: [12], [16], [254], [255], [297]
B2.3	Diagnosis: [263], [264], [271], [280] Prognosis: —
B3	Diagnosis: [50], [58], [162], [210], [268]–[270], [278], [279], [284], [292] Prognosis: [295], [297], [309]
B4	Diagnosis: [228], [229], [280]–[282], [289] Prognosis: —
B5	Diagnosis: — Prognosis: [296]
B6	Diagnosis: [267] Prognosis: —
B7	Diagnosis: [290], [291] Prognosis: —
C	Diagnosis: [45], [46], [285], [294], [298] Prognosis: —

A division is made between approaches from the manufacturer's and operator's perspectives, as previously introduced in Section IV-B. The latter can be seen as approaches that already consider similar systems. The discussion on fleet learning approaches concludes with an interim summary in Section IX-C. Similar to Fig. 12 and Table 4 for transfer learning, Fig. 13 and Table 6 list the main concepts of the fleet learning approaches to provide an overview.

A. APPROACHES FROM THE MANUFACTURER'S PERSPECTIVE

A common approach in fleet learning is to combine the fleet data into one dataset and then use that dataset for training without changing the methods themselves (D1). This merged dataset is intended to ensure that the data-driven method has already seen data from as many domains as possible during training. Thus, it should be unlikely in the application that data from a new, unknown domain are encountered. In the case of fleets defined from the manufacturer's point of view, domains refer to operating conditions since the systems are identical. Basora et al. [310] identified faults of cooling units installed in several aircrafts by anomaly detection with different variants of autoencoders. Bull et al. [311] detected faults of wind turbines in a wind farm with Gaussian process regression. The predicted mean can be seen as the mean representation of the wind turbine, and the variance tracks the variations in the population. Simone and Subanatarajan [67] used a Monte Carlo algorithm to perform a statistical analysis of fleet degradation data and predict the RUL of individual voltage breakers and switch gears. In [312], a Monte Carlo simulation or a neural network was used. Bouzidi et al. [313] applied different data-driven learning algorithms to a fleet of aircraft engines. The aim was to predict the RUL.

A model trained with all fleet data might tend to provide only a rough estimate of degradation without detailing the component-specific variations. Therefore, another fleet learning concept is to use a general model that basically emulates the behavior of the entire fleet but then adapts it to the specific system under consideration (D2). In [314], a framework for the condition diagnosis of a fleet of micro gas turbines with production variances was presented. The idea is to expand a model that depicts the behavior of an averaged turbine by tuning parameters that can adapt to each individual turbine. A prognosis approach to specify the downtime predictions of a model trained with all fleet data was presented in [315]. A correction model was trained with the deviations of the true degradation of each unit from the general model. This approach was applied to identical pneumatic valves used under different operating conditions.

There are also fleet learning approaches that combine multiple models into an ensemble. Models of the same type (E1) as well as models of different types (E2) can be combined. Regardless of the types of models combined, another approach is to train the models with different signals (E3). Rigamonti et al. [26] created a diagnosis model for bipolar transistors that can be used for different operating conditions. For this purpose, a separate model for each operating condition is trained, and then all the models are combined into an ensemble (E1). In [316], an approach for the condition diagnosis of aircraft engines was presented, whereby a data-driven and physical model are combined (E2).

Another fleet learning concept is to deal with subfleets that contain the units of the entire fleet that are most similar to each other (F1). Using a hierarchical extreme learning machine, Michau et al. [70] measured similarities between datasets of gas turbines of the same type installed in several power plants and operated under different conditions. If similar sets were found, a neural network was trained further with the associated data. Ultimately, the network can then be used to determine the health state of individual units. In another similar approach applied to the same case study, feature alignment for multiple units was performed using an unsupervised feature alignment network. This allows the features of the units to be combined and a neural network to be trained [27]. Jin et al. [69] used a clustering algorithm to monitor the temporal evolution of control valve damage in the oil and gas industry under different operating conditions. The idea is to cluster valves based on the feature vectors. For each cluster, a predictive model can be used to estimate the health state. If units in a cluster differ too much during the life cycle, all units are reclustered. Leone et al. [72] and Leone et al. [13] presented algorithms to estimate the RUL of industrial circuit breakers. The basic idea is to select from a fleet exactly those circuit breakers that show the highest similarity to the degradation pattern of the unit under consideration. With this subfleet, the RUL of the circuit breaker under consideration can be estimated. A number of approaches are used to select the subfleet, such as the root mean squared distance between the degradation trajectories or the two-sample Kolmogorov- Smirnov test. A Monte Carlo simulation serves as a datadriven prognosis method. Liu [71] attempted, to predict the RUL of aircraft engines under different failure modes and engine operation settings. For this purpose, a similarity comparison of the present trajectory with similar historical trajectories was also performed (F1).

The last common fleet learning concept addresses the advantage of fleet learning arising from the simultaneous consideration of multiple systems. As described in Section IV-B, when considering several units in a fleet, historical data can be omitted, and instead, a comparison of data between fleet units can be performed (F2). An unsupervised anomaly detection approach was presented by Hendrickx et al. [317]. During operation, real-time data from several electric power trains are compared with each other. Cluster analysis is used to detect anomalies and thus potential damage. This method does not require historical data. Only deviating machine behavior is used as a degradation indicator. Even if different engines are used for the load simulation, the approach is assigned to the fleet type defined from the manufacturer's perspective, as the same drive engine type is always used. Only the load differs due to different load motors. Hendrickx et al. [68] and Hendrickx et al. [318] presented a similar approach for the same application. Among other approaches, Liu [71] performed time series cluster analysis with operational data from an onshore wind turbine farm comprising 24 turbines to establish an optimal maintenance schedule (F2).

There are also approaches that combine several of the concepts in Fig. 13. An ensemble-based anomaly detection approach was presented in [319] (E1, E3, F1). Wind turbines that are close to each other were compared. If the sensor signals of one turbine deviate from the others, this indicates a fault. To increase the robustness of the condition diagnosis, multiple models trained with different sensor signals are combined into an ensemble and then a vote is taken on the presence of an anomaly.

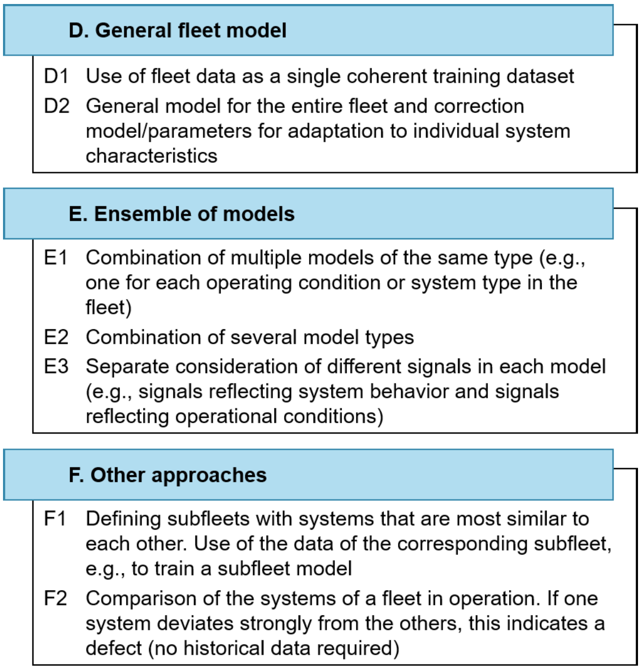

FIGURE 13. Main fleet learning concepts for condition diagnosis and prognosis in PHM.

Al-Dahidi et al. [21] utilized fleet information for the progress of degradation and the RUL prognosis. Therefore, a homogeneous discrete-time finite-state semi-Markov model (HDTFSSMM) is used. An unsupervised ensemble clustering algorithm determines the number of necessary states of the HDTFSSMM. For this purpose, several clusterings are performed, each based on different signals (E1, E3). For example, one clustering can use signals that characterize the behavior of the system, and another clustering can use signals that are related to the operating condition. In this way, the data belonging to the same degradation state (considering different operating conditions) can be grouped. The approach is applied to aluminum electrolytic capacitor data under different temperature profiles as well as aircraft engines. Al-Dahidi et al. [320] and Al-Dahidi et al. [321] took a similar approach as in Al-Dahidi et al. [21]. In addition, they used an ensemble-based approach comprising an HDTFSSMM and a fuzzy similarity-based model for the RUL prognosis of aluminum electrolytic capacitors under different operating conditions. Each model is assigned a weight and bias depending on its local performance. Furthermore, the similarity of the actual degradation trajectory is compared with historical trajectories to select similar trajectories (E1, E2, E3, F1).

As seen from the references listed, there are already several condition diagnosis and prognosis approaches that address fleets from a manufacturer's perspective. However, these approaches only consider a preliminary stage of similar systems, since the systems under consideration are regarded as identical. The differences arise primarily from different operating conditions. However, it may well be that some approaches are applicable to similar systems.

B. APPROACHES FROM THE OPERATOR'S PERSPECTIVE

Fleets defined from the operator's perspective correspond to fleets of similar systems. Although the individual units are similar, they differ in technical characteristics. Examples include similar electric motors with different powers, similar manufacturing machines with slightly different structures, or similar batteries with different capacities or even different chemical compositions. Although this case has enormous importance for the industry, according to Fink et al. [25], this type of fleet has not yet been investigated in PHM. Jia et al. [323] also viewed fleet-based forecasting using data from similar machines as an open but very important question. Nevertheless, there are initial condition diagnosis and prognosis approaches that proceed in this direction. However, in most cases, the similarities between the systems under consideration are still very distinctive, and often only components and not complex engineering systems are considered.

An anomaly based fault detection approach for generators of wind turbines was presented in [322]. The turbines were from four different manufacturers and were located in several places in Europe. For anomaly detection, an autoencoder was used (D1). Electrical circuits usually comprise standardized components, which are only arranged differently. Samie et al. [15] therefore pursued the idea of setting up an algorithm to predict the RUL of DC/DC converters with the same components but different topologies (D1).

In addition to these references, there are approaches that consider multiple concepts of Fig. 13. Al-Dahidi et al. [74] clustered data from similar turbines of different nuclear power plants. This approach was applied to two different turbines installed in two different nuclear power plants. However, the power plants were highly standardized, so the technical differences are rather small. Similar to Al-Dahidi et al. [21], [320], [321], who considered fleets from the manufacturer's perspective, several clusterings were performed based on different signals to characterize the behavior and operating condition of the systems (E1, E3). In addition to the ensemble approach already discussed in Section IX-A (E1, E3, F1), Helsen et al. [319] also used artificially generated data to improve condition diagnosis. These simulated data can be seen as similar system data.

Although they did not present a procedure for condition diagnosis or prognosis, the approach of Medina-Oliva et al. [73] should be mentioned. They used an ontology-based approach to divide a fleet of distinctly different diesel engines into subfleets in which the systems are most similar to each other.

C. SUMMARY OF THE APPROACHES

According to Table 5, common condition diagnosis and prognosis applications for fleet learning are circuit breakers or switchgears, electric motors, wind turbines, and aircraft engines. Unlike transfer learning, the focus is on electrical or electronic applications as well as more complex systems

TABLE 5. Application examples of fleet learning for condition diagnosis and prognosis. (— = no literature found).

Application on, signals	Fleet learning			
	Different operating conditions (manufacturer's perspective)		Similar systems (operator's perspective)	
	Simulated	Measured	Simulated	Measured
Transistors				
- current, voltage, temperature	—	[26]	—	—
Electrolytic capacitors				
- resistance	[21], [320], [321]	—	—	—
DC/DC converters				
- voltage, current	—	—	[15]	—
Circuit breakers / switchgears				
- unclear	—	[13], [67], [72], [312]	—	—
Valves				
- crack depth	[315]	—	—	—
- multiple	—	[69] (scheduled)	—	—
Electric motors				
- current	—	[68], [317], [318]	—	—
- vibration	—	[317]	—	—
Micro gas turbines				
- unclear	—	[314] (scheduled)	—	—
Power plant turbines				
- vibration	—	—	—	[74]
- multiple	—	[27], [70]	—	—
Wind turbines				
- power	—	[311]	—	—
- vibration	—	—	—	[319]
- power, temperature	—	—	—	[322]
- multiple	—	[71], [319]	—	—
Cooling units				
- unclear	—	[310]	—	—
Aircraft engines				
- multiple	[21], [71], [313], [316]	—	—	—

TABLE 6. References to the main fleet learning concepts in Fig. 13 (Diagnosis = condition diagnosis, Prognosis = condition prognosis, — = no literature found).

D1	Diagnosis: [310], [311], [322] Prognosis: [15], [67], [312], [313]
D2	Diagnosis: [314] Prognosis: [315]
E1	Diagnosis: [26], [319] Prognosis: [21], [74], [320], [321]
E2	Diagnosis: [316] Prognosis: [320], [321]
E3	Diagnosis: [319] Prognosis: [21], [74], [320], [321]
F1	Diagnosis: [27], [69], [70], [319] Prognosis: [13], [71], [72], [320], [321]
F2	Diagnosis: [68], [71], [317], [318] Prognosis: —

such as aircraft engines rather than individual mechanical components such as bearings. The most frequently used signals are current and vibration. However, there are also some approaches that use several signals simultaneously. Overall, fleet learning represents a research field whose approaches can be used to exploit data from similar systems. In particular, fleets from the operator's perspective can be considered as a collection of similar systems. Nevertheless, there are currently very few approaches dealing with this type of fleet. These applications are discussed further in Section X.

Fig. 13 lists the main fleet learning concepts used for condition diagnosis and prognosis. According to Table 6, the three most popular concepts are to use the fleet data as one set for training without making specific adjustments to the data-driven methods, to combine knowledge from several operating conditions, and to define subfleets to train several models. In contrast to transfer learning, where the focus is mainly on condition diagnosis, fleet learning has approximately the same number of approaches to condition diagnosis and prognosis, as seen in Table 6).

X. PHM APPLICATIONS FOR CONDITION DIAGNOSIS AND PROGNOSIS CONSIDERING SIMILAR SYSTEMS

In Sections VI to IX, currently used PHM approaches for condition diagnosis and prognosis related to a similar system problem in the field of transfer and fleet learning were presented and explained. These include approaches that consider different operating conditions as well as those that deal specifically with similar systems. Tables 3 and 5 divide the existing approaches accordingly. In the following, the approaches applied to similar systems will be considered again in more detail. Whereas the previous sections considered the functionalities and the applications of the approaches, the focus is now placed on how similar systems differ in concrete terms. This should illustrate the current state of research in the field of similar systems. As Tables 3 and 5 show, approaches that consider similar systems are currently in the minority. The focus of both transfer learning and fleet learning is on identical systems under different operating conditions. However, there are already similar system applications.

It must be mentioned at the outset that the parameter-based approaches that use natural images as a source domain are not listed in Tables 3 and 5. Although natural images form a very different domain, they are usually not similar engineering systems. Furthermore, these approaches cannot be assigned to a transfer between different operating conditions. Approaches to which this applies are those in [9], [81], [84], [85], [86], [92], [93], [94], [95], [96], and [98] on bearings, [87], [93], [98] on gearboxes, [168] on tools for machinery, [83] on railway wheels, [18], [93] on electric motors, [97] on batteries, [84], [86], [95] on pumps, and [78] on wind turbines.

Furthermore, it should be mentioned that there are often different operating conditions between similar systems. However, in the course of the assignment, deviating operating conditions are seen as a consequence of the deviations in the systems. Such approaches are therefore only assigned in Tables 3 and 5 to the category of similar systems. An exception are approaches that explicitly examine different operating conditions and similar systems separately. These are listed in both categories.

In the following Sections X-A to X-E the similar system approaches of transfer and fleet learning, subdivided by the applications, are discussed in more detail. In conclusion, the key findings are summarized in Section X-F. Additionally, Table 7 lists the condition diagnosis and prognosis applications in Sections VI to IX, which consider similar systems. In contrast to Tables 3 and 5, a distinction is made between whether condition diagnosis or prognosis is considered. As previously noted in Sections VI to IX, condition diagnosis approaches currently predominate, even and especially when focusing on similar system approaches.

A. SIMILAR BEARINGS

The most common application for similar system approaches are bearings. These represent the main use case of inductive and transductive transfer learning approaches on similar system data. Inductive and transductive approaches exist in roughly equal numbers. First, the inductive approaches are listed. Yang et al. [66] used bearing data from Case Western Reserve University (CWRU). The two domains are formed by different types of bearings at different positions in the test rig. The first type is a deep groove ball bearing double-sided sealed (6205-2RS JEM) with nine rolling elements installed on the drive end. The second bearing type is also a double-sided sealed deep groove ball bearing (6203-2RS JEM), albeit with eight rolling elements and a different size, and is installed at the fan end. Both bearings are manufactured by SKF. The operating conditions in the form of the load are kept constant. Cheng et al. [149] also utilized the CWRU dataset and transferred knowledge between two sensor locations, at the drive end and at the fan end. Zheng et al. [99] used bearing data from two different test rigs. The first rig is the MFS-MG rig in their laboratory, and the second dataset is a bearing dataset contributed by the Center for Intelligent Maintenance Systems (IMS) of the University of Cincinnati. In addition to the different test rigs, the signals were recorded under different sampling frequencies and rotation speeds. In [112], data from two different bearing test rigs were also used. The first dataset is from the CWRU 6205-2RS JEM bearing test rig, and the second is from another test rig. Between these datasets, in addition to the test rigs, the operation conditions also differed (e.g., different speeds, fault types, fault diameters). Another approach where knowledge was transferred between two bearing test rigs was presented in [113]. Thus, the first domain is formed again by the CWRU 6205-2RS JEM bearing data. The second domain includes data from an internal bearing test rig with deep groove ball bearings of type CBS6209. Again, the operation conditions varied. Han et al. [151] also used the CWRU dataset and a bearing dataset from Paderborn University; thus, two different test rigs with different bearing types were considered for the transfer. The data originated from different operating conditions. Two additional approaches in which data were transferred between different bearing test rigs include those presented by Zhou et al. [163], who transferred knowledge between the CWRU dataset, the IMS dataset, and data from their own test rig, and Wang et al. [171], who transferred knowledge between the CWRU and the IMS bearing datasets. Li et al. [122] transferred the knowledge learned from the CWRU dataset to a real-world railway locomotive bearing dataset recorded on another test rig. In addition to the test rigs, the bearings and the operation conditions differed.

There are also approaches that transfer knowledge between bearings and distinctly different similar systems. Li et al. [152] presented an approach using four different datasets. Three of these datasets were bearing datasets, specifically the CWRU bearing dataset, the IMS bearing dataset, and data from a train bogie test rig. In addition, another dataset from a significantly different area of application was used, which comprises data from a shaft crack test rig. Transfers were carried out between these four datasets. In [24], an approach that uses completely different components was presented. Knowledge was transferred from bearings to gearboxes and vice versa. The CWRU dataset and a gearbox dataset of the 2009 PHM data challenge were used.

TABLE 7. Application examples for condition diagnosis and prognosis considering similar systems. (— = no literature found).

Application on	Approach	Condition diagnosis	Condition prognosis
Bearings	Inductive transfer learning	[24], [66], [99], [112], [113], [122], [149], [151], [152], [163], [167], [171]	—
	Transductive transfer learning	[30], [58], [149], [179], [198], [217], [260], [261], [278], [283], [284], [286]–[289]	
Gears/gearboxes	Inductive transfer learning	[24], [108]	—
	Transductive transfer learning	[261], [292]	—
Ball screws	Transductive transfer learning	[237]	—
Structures	Transductive transfer learning	[195], [202], [243]	—
Transistors	Inductive transfer learning	[128]	—
Electric power converters	Inductive transfer learning	[80], [109], [111]	—
	Fleet learning	—	[15]
Circuit breakers	Inductive transfer learning	[49]	—
Other actuators	Transductive transfer learning	[236]	—
Other rotating components	Inductive transfer learning	[108], [152]	—
Boilers	Transductive transfer learning	[293]	—
Batteries	Inductive transfer learning	[104]	[118]
	Transductive transfer learning	—	[232]
Photovoltaic systems	Inductive transfer learning	[156]	—
Components of power plants	Inductive transfer learning	[106]	[144]
	Transductive transfer learning	[227]	—
	Fleet learning	[74]	—
Wind turbines	Transductive transfer learning	—	[197]
	Fleet learning	[319], [322]	—
Compressors	Inductive transfer learning	—	[129]
	Transductive transfer learning	[218]	—
Chillers	Inductive transfer learning	[159]	—
	Transductive transfer learning	[199], [253]	—
Pumps	Inductive transfer learning	[172]	—
Aircraft engines	Inductive transfer learning	—	[129], [133]
Quadrotor drones	Inductive transfer learning	[103]	—
Cellular networks	Inductive transfer learning	[142]	—
	Transductive transfer learning	[300]	—
	Unsupervised transfer learning	[300]	—
Disk systems	Inductive transfer learning	[139], [157]	—

In addition to using data from real applications, simulation data can also be used. These usually do not exactly match the real-world system, so simulations also represent a similar system problem. Yu et al. [167] used a simulation model of a rotor-bearing system as the source domain. The knowledge was transferred to the data of two different bearing test rigs. One of the datasets was provided by Xi'an Jiaotong University (XJTU).

The transductive approaches applied to bearings deal with mostly identical similar system datasets as those in inductive approaches. A transfer between the degradation data of bearings from different test rigs has often been considered. Thus, the test rig architectures, the bearings, the sensors, the sensor positions, and the fault modes, often differ from one another. Deng et al. [260] utilized the CWRU dataset, the Paderborn University dataset, and the XJTU dataset. Wang et al. [287] conducted examinations with the CWRU and IMS data and a bearing test rig dataset provided by Jiangnan University. Xiang et al. [288] used the CWRU, XJTU, and Paderborn University bearing datasets, and Jin et al. [58] used the CWRU data and data from an internal test rig. Feng et al. [283] also transferred knowledge between datasets from three bearing test rigs. Furthermore, the CWRU dataset was used to transfer data between different sensor positions in [149], [198], [278], and [286] and between the drive end and fan end bearings in [284]. The knowledge gained from test rigs has also been used for bearings in real operation, especially locomotive bearings. Yang et al. [179] utilized the knowledge of laboratory bearing data (CWRU) to determine the health state of locomotive bearings operated in the real world. Yang et al. [217] transferred knowledge from two laboratory bearing datasets, one of which was the

CWRU dataset, to locomotive bearings. In [30], the CWRU, IMS, and a railway locomotive bearing dataset were used as the domains. Yang et al. [289] transferred knowledge from the XJTU dataset and a dataset provided by the Society for Machinery Failure Prevention Technology to data from a test rig for locomotive bearings. Liu et al. [261] utilized CWRU data and the 2009 PHM gearbox dataset. The gearbox dataset served as a strongly deviating domain to achieve negative transfer (see Section XII). The algorithm should detect this ill-fitting domain and prevent its negative effect.

B. SIMILAR GEARBOXES AND OTHER ROTATING COMPONENTS

Although bearings are currently the most commonly used similar system applications, other similar systems are also being considered. Gearboxes are one such example. Kumar et al. [108] utilized a bevel gearbox dataset as similar system data for spur gearboxes from the IEEE PHM Challenge Competition 2009. The transfer between bearings and gearboxes already listed in the course of the bearing applications should also be mentioned [24]. Both of the previously referenced papers considered inductive transfer learning. A transductive approach to bearings and gearboxes, which is also referred to in bearing applications, was presented in [261]. Shen et al. [292] also addressed transductive transfer learning and the transfer between different sensor positions in a test rig simulating a drivetrain with gear faults.

Approaches using other rotating component systems also exist. As already mentioned, Li et al. [152] utilized a shaft crack dataset in addition to the bearing datasets. Kumar et al. [108] also applied their approach to rotor defect test rig data. The faults in both setups were misalignment, unbalance, and rotor rub. Both approaches can be assigned to inductive transfer learning. Pandhare et al. [237] transferred knowledge between different sensor positions on ball screws.

C. SIMILAR ELECTRICAL AND ELECTRONIC COMPONENTS

Batteries are also a similar system application field that has already been explored. Li et al. [104] transferred knowledge between the same type of batteries ($LiFePO_4$), albeit with different specifications, specifically different nominal capacities and nominal voltages. Kim et al. [118] considered different types of batteries with different capacities, packaging styles, chemistry, and numbers of cells. In addition to these approaches using inductive transfer, transductive approaches have also been used. In [232], batteries with the same cathode and separator material, albeit different anode materials, electrolyte solutions, and a varying number of contained cells, were considered. Furthermore, it is worth mentioning that Liu et al. [119] transferred knowledge between different battery types, namely, nickel cobalt manganese and lithium cobalt oxide batteries. However, they performed state of charge estimation, which is not seen as a PHM application area in the course of this paper. It should be noted that in applications, the operating conditions of batteries usually differ greatly, which can lead to strong deviations in the degradation behavior. Therefore, transfer learning approaches can already be significantly beneficial in these cases.

There are inductive transfer learning approaches on similar transformers. Duan et al. [80] generated artificial fault patterns (temperature and velocity field) through simulation and used them as the source domain. Duan et al. [109] also used a simulation dataset. Chen et al. [111] considered different similar transformer rectifier unit structures. Samie et al. [15] presented a fleet learning approach considering DC-to- DC converters. Converters are considered to have the same components but different topologies.

In [49], knowledge was transferred from circuit breaker simulations to a real-world experimental domain. Pan et al. [128] transferred knowledge between different transistor generations. In detail, the source domain comprises data from fin field-effect transistors and the target domain from gate-allaround field-effect transistors.

D. SIMILAR COMPONENTS FOR POWER GENERATION

Similar system applications can also be found on the components of power plants. Two inductive transfer learning approaches have been found. Yang et al. [106] transferred knowledge between simulation data of GE9FA and Siemens V64.3 gas turbines. Both are single-shaft turbines comprising a compressor, a combustion chamber, and a turbine. Zhou et al. [144] used gas turbine data based on a physical model as the source domain to transfer the knowledge to a real-world application. Al-Dahidi et al. [74] utilized fleet learning on two different turbines installed in two different power plants. A transductive transfer learning approach was presented in [227]. Knowledge was transferred between two different components of a power plant, a boiler dataset, and an electricity generator dataset.

Lu et al. [156] considered a photovoltaic emulator and a photovoltaic system on a rooftop comprising four JINKO JKM350M-72 monocrystalline panels. A fleet learning approach was presented in [322] on wind turbine generators. The wind turbines were from different manufacturers and were located at different sites. Overall, they considered four turbine brands from seven wind farms scattered across Europe. Pan et al. [197] transferred knowledge between different sensor positions of a wind turbine gearbox, and Helsen et al. [319] used not only measured data but also additionally generated simulation data.

E. OTHER SIMILAR SYSTEMS

There are also other applications considering similar systems. Similar structures with geometric and material differences were considered in [195] and [243]. In addition, physical models were used as source domains. Lin et al. [202] also transferred knowledge from simulations to real-world structures. All of these approaches are based on transductive transfer learning. Similar aircraft engines are another field of application. Liu et al. [133] used turbofan engine data from the 2008 PHM conference competition as the source domain. These data were based on a simulator called CMAPSS and are part of the C-MAPSS dataset. The target domain comprises real measurement data of aircraft gas turbine engines. Gribbestad et al. [129] also utilized the 2008 PHM dataset as the source domain. However, the target domain comprises marine air compressors. Both approaches are part of inductive transfer learning.

Transfer learning approaches on chillers also exist. In [253], building chillers with different cooling capacities, powers, and structures were considered. van de Sand et al. [199], Fan et al. [159], and Li et al. [293] also transferred knowledge between two different chiller types. However, the latter focuses on energy optimization rather than condition diagnosis or prognosis. Further possible differences between chillers are different refrigerants, compressors, or heat exchangers. In addition to the energy optimization of chillers, in [293], the transfer between different boilers for condition diagnosis was considered. Wu et al. [300] presented an unsupervised approach to fault detection and a transductive approach to condition diagnosis in fog radio access networks between similar nodes. Wang et al. [142] also addressed radio networks, specifically with configurations of femtocells. However, neither approach clarified how similar the nodes were.

Zhang et al. [139] used inductive transfer learning to utilize knowledge from similar disk systems in data centers. The types studied are HDDs, SATA SSDs, and NVMe SSDs from different manufacturers. Another approach was presented in [157], in which knowledge was transferred between hard disks of different manufacturers. Reference [236] transferred knowledge between different sensor positions on electromechanical actuators. As previously mentioned, Gribbestad et al. [129] used knowledge from aircraft engines on marine air compressors, and Guo et al. [218] transferred knowledge between two valves installed in two different reciprocating compressor models (IODM-115-5-3-16 and IODM 70-5-4R). Propeller damage of similar quadrotor drones that differ, e.g., in the drone architecture, propeller size, or propeller rotation speed, was considered in [103]. Data from similar sucker rod pumping systems were utilized in [172].

F. SUMMARY OF THE APPLICATIONS

As seen from the applications listed in Table 7 and explained in this section, there are already some similar system approaches for condition diagnosis and prognosis in PHM, with the diagnosis being considered significantly more often. Existing applications range from mechanical components such as bearings and gearboxes to electromechanical components and electrochemical components such as batteries or photovoltaic systems. There is also a wide range in the complexity of the systems considered, ranging from individual components such as bearings to more complex systems such as aircraft engines or wind turbines. The already existing and potentially possible scope of applications is therefore very large. This reinforces the assumption that transfer and fleet learning can play a significant role in the use of data from similar systems in the future. Table 7 clearly shows that most of the existing similar system approaches belong to inductive and transductive transfer learning. There are few approaches to unsupervised transfer learning and fleet learning in the field of similar systems. This can be explained in part by the fact that unsupervised transfer learning and fleet learning are generally not as widely used in PHM for condition diagnosis and prognosis as inductive and transductive transfer learning. Therefore, the current state of the literature suggests that transfer learning approaches are likely to be more relevant than fleet learning for similar system problems in the future.

XI. SIMILAR PROCESSES

Although this literature review focuses on engineering systems, the presented approaches can generally be applied whenever similar datasets are available. In the industrial environment, the application of these approaches to similar processes is therefore also very promising. The focus here is not primarily on the degradation of the systems that are underlying the process. Instead, the central challenges are the diagnosis of process errors that occur, the determination of production quality, or the prediction of production progress. Here, a distinction can also be made between different levels of similarity. For example, different process parameters can be set, different products can be produced, or processes with different substeps can be considered. There are already a few transfer learning approaches dealing with such challenges, which are briefly listed below.

Huang et al. [173] used transfer dictionary learning for process monitoring of a stirred tank heater under different process parameters. Xu et al. [324] presented an MMDbased approach for condition diagnosis in a car body-side production line. Simulation data were used as source data. A process condition diagnosis method based on MMD was presented in [210]. Wang et al. [325] applied TCA to minimize the distribution discrepancy between domains. This method was applied to condition diagnosis of a simulation of the Tennessee-Eastman process and an ore-grinding process.

A transfer learning approach for production progress prediction can be found in [326]. The aim was to predict the degree of fulfillment of a production plan in the future. Accordingly, the models were pretrained with historical data and subsequently fine-tuned. A similar approach was used in [327] for modeling lithography simulation processes with different configurations. Lin et al. [328] presented a further development of this approach. Tercan et al. [52] predicted the production quality of injection molding based on machine settings, whereby pretraining with production data from previous parts was also used. A feature extraction network pretrained with natural images was used in [329] and finetuned to diagnose the condition of a cylindrical metal shell. Yang et al. [330] and Kim et al. [331] also used such pretrained networks to detect the water-binder ratio of concrete and to monitor the quality in laser-assisted micromilling of glass, respectively. Similar approaches were also presented in [332] for the quality assessment of flat glass using images of the glass produced and in [333], [334], and [335] for detecting defects in welded joints.

Gong et al. [336] combined a DANN and a similaritymeasure method based on the Euclidean distance. The aim was to improve the defect detection of composite materials in X-ray images by using X-ray images of welding defects as the source domain. Li et al. [337] combined a DANN and the MMD as a similarity measure for the condition diagnosis of a stirred tank reactor and a pulp mill plant. There are also initial approaches using transfer learning in semiconductor manufacturing. For example, Kang [338] and Azamfar et al. [339] aimed to use machine state data to infer possible defect types. Accordingly, Kang [338] implemented a parameterbased transfer, and Azamfar et al. [339] added an MMD term to the loss function. Other transfer learning approaches were presented in [340] to create a regression model for time series from industrial manufacturing, [341] for process monitoring of an ore grinding and grading process, [342] to monitor the Tennessee-Eastman process, and [343] on a manufacturing process in which several similar products are produced.

In summary, the following application scenarios of similar system and process approaches are conceivable:

- Classical PHM, i.e., condition diagnosis and prognosis of engineering systems by using sensor and control data
- Determination of production quality using measurement data from an end-of-line test
- Determination of production quality using sensor and control data from production
- Condition diagnosis and prognosis of manufacturing machines using sensor and control data from production
- Condition diagnosis and prognosis of manufacturing machines using measurement data from an end-of-line test
- Optimization of process parameters using measurement data from an end-of-line test

XII. AVOIDING NEGATIVE TRANSFER

As the literature listed in the previous sections shows, using data from similar systems can add significant value to data-driven learning methods. In particular, if only a small amount of data from the system under consideration are available, the performance of the algorithms can be strongly increased. Moreover, if data from similar systems already exist, they are usually available without much further effort. However, the similarity of the datasets has a decisive influence on the added value that arises from their usage. The greater the deviations in the domains and tasks, the less promising the use. If the differences are excessive, there can even be a deterioration in performance. In the case of transfer learning, this is referred to as the negative transfer. If a negative transfer occurs, the transferred information from

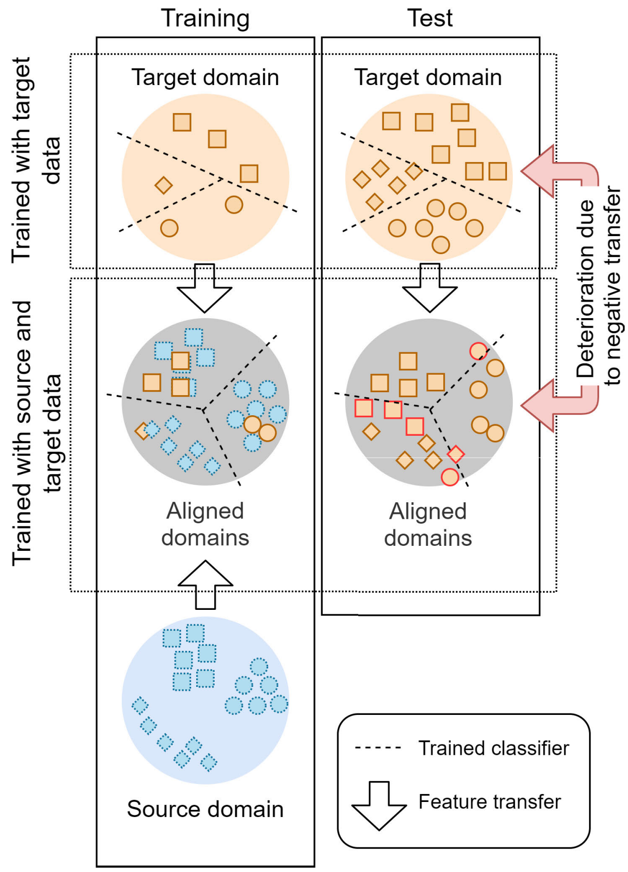

FIGURE 14. Example of a negative transfer situation.

the source domain(s) negatively influences the learning of the target prediction function, meaning that the training without the usage of the source data would be more successful [51], [54], [55], [344]. An example of a negative transfer situation is shown in Fig. 14. When using additional samples from the source domain for training, misclassifications occur on the target test data that do not occur when using only the target data. That is, the performance on the target test data deteriorates when the samples from the source domain are used for training. A negative transfer can also occur in fleet learning if the system under consideration strongly deviates from the other fleet units. Then, the usage of the knowledge about the fleet (other domains) deteriorates the prediction function of the considered system.

The detection and prediction of negative transfers have since been and currently remain an open problem in all application areas of transfer learning [53], [55], [62], [345], [346], [347]. According to Wang et al. [348], open questions include the following. How can negative transfer be measured? What should be used as a baseline for assessing improvement or deterioration? Moreover, it is not clear what exactly causes a negative transfer. According to them, the discrepancy between the domains is probably a strong factor. However, it is not known which discrepancies cause negative transfers. Additionally, there could be other influencing factors that favor negative transfers. The authors saw the main problem in how to detect a negative transfer by utilizing unlabeled target data.

The negative transfer problem is fundamental since it can occur whenever data from other domains are used that do not exactly match the domain under consideration. Thus, it is also crucial for the industrial usage of similar system data in PHM. It lies beyond the scope of this review to give an overview of all existing approaches for detecting and preventing negative transfers. However, due to the strong significance, some basic considerations will be made in the following regarding similar system data. Transfer learning approaches in PHM that draw attention to negative transfer include, e.g., [195], [260], [261], [269], [270], [288]. Furthermore, in principle, all approaches using system data from different operating conditions or similar systems (transfer and fleet learning) aim to avoid negative transfers as best as possible. As examples, domain alignment or the weighting of the source samples can be mentioned. Nevertheless, in the current PHM literature using transfer learning and fleet learning for condition diagnosis and prognosis, the fundamental suitability of the data from different domains is usually simply assumed. For example, domain alignment methods are applied in the hope that the domains are sufficiently similar to be approximated at all. However, as in other fields of transfer learning, this is too optimistic [346]. For reliable statements in the field of PHM, the basic suitability of the available data of similar systems should first be verified. This is particularly necessary because not only different operating conditions but also different systems are considered.

One straightforward method for verification could be to see how the use of similar system data affects the training result by means of extensive test runs. However, in the industrial environment, there is often insufficient time or it is too costly to carry out extensive training runs to compare the performance, with and without the use of similar system data. Another problem with evaluating the suitability of similar systems based on classification performance is that there is usually an imbalance between classes in the condition diagnosis. This is because faults occur much less frequently, and therefore, fault data are underrepresented. Therefore, the performance evaluation is mainly determined by the healthy state data, as they form the largest part of the datasets [346]. However, in the application, it is crucial to classify the faults correctly. In addition, in practice, there is often not only one potential similar system but several. Then, another question arises. Which of these systems should be used? This would require evaluating the performance of each possible training data combination. For these reasons, it is necessary to find a way to identify the most similar systems without having to train data-driven methods.

Approaches that are based on the structure of systems seem to be promising for similar system approaches. For example, from several types of ball bearings, those with the same number of rolling elements and similar dimensions could be selected. In the case of batteries, for example, it would make sense to use source batteries made of the same (chemical) materials as the target battery. However, the evaluation is then very subjective. For example, is the dimensioning or the number of rolling elements more important for bearings? What influence does the type of rolling element have on the similarity? In addition, it is necessary to define individual evaluation criteria for each system type. The more complex the system type, the more complex the definition. Above a certain complexity, such as that of production machines, this approach quickly reaches its limits.

Therefore, data-driven methods that decide how similar systems are purely on the basis of the measurement data are particularly suitable. For example, the similarity measures presented in Table 2 can be used for the initial similarity measurement of similar systems. In addition to comparing the marginal distributions, the conditional distributions should also be compared. In this way, the most similar systems considered from the data can be identified. This approach would be consistent with that of Wang et al. [348], who observed the core of the negative transfer in the distribution shift. However, there are challenges with this approach. To make a statement about the usefulness of similar system data, a kind of threshold value is necessary that indicates when the data are sufficiently similar for similar system approaches. There are already approaches in statistics that introduce such a threshold, e.g., [177], [349]. This threshold is used to check whether two samples are drawn from the same population. However, in the field of similar systems, this is not yet an indication that the systems are too different for similar system approaches. Therefore, approaches must be found to relax this usually too strict threshold. In addition, there are usually few degradation data available from the system under consideration that could be used for similarity analysis. Therefore, it will be crucial for the industrial utilization of similar system data to identify suitable similarity metrics that can deal with these difficulties and define appropriate thresholds. At this point, negative transfer learning should be mentioned once again as another area in the transfer learning literature [61]. However, due to the lack of approaches in the field of PHM, it will not be discussed further.

XIII. SUMMARY AND OUTLINE OF FURTHER RESEARCH CHALLENGES

Data-driven PHM methods for condition diagnosis and prognosis offer strong potential for industrial use. However, sufficient training data are required. Especially in industrial applications, data from individual systems are usually not available in the desired quantity because data generation is usually associated with large time and monetary expenditures. In addition, especially at the time of market launch, there are usually only small amounts of runtime and run-to-failure data available, as the system has not yet been used in industrial applications. These problems are exacerbated by the increasing importance of a variant-rich product portfolio. Therefore, there is no doubt that in PHM, the use of data from similar systems can add great value to the data-driven condition diagnosis and prognosis of engineering systems.

A. PRINCIPLE PROBLEM IN USING SIMILAR SYSTEM DATA

Nevertheless, data from similar systems cannot simply be used as is. Similar systems do not behave exactly identically due to variations in their technical characteristics and show different degradation behaviors. Additionally, the installed sensor types or the sensor locations may vary. Different operating conditions also lead to further differences in the data. All of this can result in:

- Measurement and feature differences
  - Measurement quantities
  - Signal properties (sampling rate, accuracy, etc.)
  - Interferences (bias and noise)
  - Marginal distributions
  - Conditional distributions
- Label differences
  - Fault types
  - Fault severities
  - Fault locations
  - Health index values (e.g., RUL)

Therefore, the data from similar systems must be treated ''differently'' than the data from the system under consideration. For this, similar system approaches, with the help of which knowledge about similar systems can be used advantageously, are imperative.

B. SUMMARY OF CURRENT APPROACHES

The use of data from systems under different operating conditions or similar systems is very promising for data-driven condition diagnosis and prognosis approaches in PHM. The literature listed in Tables 3 and 5 shows that transfer learning and fleet learning can harness these data.

Although PHM is currently an underrepresented application area of transfer learning, there are already approaches that address knowledge transfer for condition diagnosis and prognosis between systems under different operating conditions or similar systems. The majority can be categorized under the core areas of inductive and transductive transfer learning. In unsupervised transfer learning, very few approaches exist. Furthermore, the current focus of transfer learning approaches in PHM is on condition diagnosis rather than prognosis. A main reason for this is that the literature on transfer learning in general deals mainly with classification tasks and little with regression tasks. However, condition prognosis, such as RUL prediction, is generally a regression problem. In addition, most publicly available PHM datasets suitable for transfer learning applications consider condition diagnosis problems (primarily fault classification). As seen from Fig. 12 and Table 4, the most common approach in inductive transfer learning addresses parameter transfer. In transductive transfer learning, feature alignment is most popular. Nevertheless, there are several more approaches in transfer learning.

Fleet learning is another approach in addition to transfer learning that can be applied to similar system problems. According to Table 6, unlike transfer learning, in fleet learning, the number of condition diagnosis and prognosis approaches is roughly the same proportion. As seen from Fig. 13 and Table 6, the three most common fleet learning concepts are to use the fleet data as one set for training, to combine several models, and to define subfleets. With the fleet defined from the operator's perspective, a concrete branch of research already exists that addresses similar systems. However, most approaches consider fleets from the manufacturer's perspective, and unlike transfer learning, fleet learning currently receives less attention in PHM research. The majority of transfer learning approaches in PHM have been published in recent years, especially as of 2019. The speed of development is mainly because there are already many approaches in other application areas that can be adapted. Unlike transfer learning, fleet learning cannot be based on the variety of findings from other fields of application.

Looking at the applications in existing approaches, mechanical components such as bearings and gearboxes are mainly considered in transfer learning. In most cases, the vibration signal is used to obtain the degradation information. In addition to mechanical components, transfer learning is already applied to electrical and electronic components such as circuit breakers, transformers, and batteries. Applications involving systems with multiple components, such as wind turbines, aircraft engines, or disk systems, are also considered. Fleet learning focuses on electrical or electronic applications such as circuit breakers or capacitors, as well as more complex systems such as aircraft engines or wind turbines. However, taking all transfer and fleet learning approaches into account, in most cases, data are generated under laboratory conditions on test rigs or by means of simulations and do not originate from industrial use.

In both transfer and fleet learning, the focus is currently on the condition diagnosis and prognosis of identical systems operating under different operating conditions. This can be seen as a precursor to the consideration of similar systems. However, that the transfer and fleet learning approaches are also suitable for similar systems is confirmed by existing approaches that consider similar system problems (see Table 7). Fleet learning already has a branch that deals with data from similar systems, but due to the dynamic in transfer learning, it is expected that in the next few years, transfer learning in particular will deal with PHM applications that use data from similar systems.

C. CHALLENGES IN USING SIMILAR SYSTEM DATA

From the authors' perspective, the similar systems approach will be a decisive future research branch within PHM. Based on the review conducted in this paper, some potential challenges that arise when data from similar systems are to be used can be identified. These are described in more detail below, along with what further research is needed to use data from similar systems. The order in which the points are mentioned represents a sequence of steps, at the end of which is the successful use of similar system approaches in industrial applications. Finally, Fig. 15 shows a summary of the key research questions.

1) AVAILABLE DEGRADATION DATASETS FOR RESEARCH

For the development of new similar system approaches and the adaptation of existing approaches, datasets are needed that represent similar system problems. However, there are currently very few such condition diagnosis and prognosis datasets publicly available. Even if the datasets that only consider identical systems with different operating conditions are also included, there exist significantly fewer datasets in the field of PHM than in the field of image processing, for example [350]. Maschler and Weyrich [36] also noted this as a key obstacle to the use of transfer learning for condition diagnosis and prognosis. In addition, the majority of these PHM datasets only contain data from laboratory setups or simulations and not from real industrial applications. The first challenge, therefore, is the lack of suitable, publicly available degradation datasets considering similar system problems. It is necessary that, in the future, more such datasets be generated and made publicly available. These datasets are essential to enable and promote research on similar system methods.

For frequent application areas such as bearings or gearboxes, it may be useful to use several datasets in the meantime and to transfer knowledge between these datasets. In doing so, each dataset does not have to contain data from similar systems. It is sufficient if the systems differ between the datasets. An overview of publicly available datasets on PHM applications can be found in [351].

2) UTILIZATION OF SIMILAR SYSTEM DATA

The second challenge will be the beneficial utilization of similar system data for condition diagnosis and prognosis. Approaches from the areas of transfer and fleet learning are suitable for this purpose. However, since the differences between similar systems are typically larger than those between identical systems, which have mostly been considered thus far, an adaptation of the existing approaches may be necessary. Moreover, due to the expected major differences between similar systems, it cannot be assumed that similar system data can always be used advantageously. In summary, the challenges of using similar system data can be described by answering the questions of when, for what purpose, and how existing data from similar systems can be used. Moradi and Groth [37] realized that these points are essential for the use of transfer learning in PHM, but they are also crucial for using similar system data in PHM in general, whether through transfer or fleet learning.

To clarify when knowledge of similar systems should be used, the following questions need to be addressed:

- Are data from similar systems needed in the current application?
- Are the existing data or trained models of similar systems suitable for usage?
- If there are several similar systems, which are most similar and most promising to consider?

These considerations are important to efficiently use similar system data and avoid negative transfer. However, most of the current literature on transfer and fleet learning does not clarify them but simply assumes the necessity and suitability of available data or models. If identical systems are considered under different operating conditions, this trial and error approach may work, but in the area of similar systems, this is usually not the case. Therefore, it is necessary to analyze whether the data-driven method needs more data from similar systems to provide acceptable results for the current application, and the similarity of existing similar systems must be evaluated in advance.

To determine whether consideration of similar systems is even necessary, it is essential to evaluate the quality of data-driven methods that they can achieve with the currently available dataset of the system under consideration. One approach to this would be to examine the prediction uncertainty, e.g., by analyzing the confidence and prediction intervals of the trained models. However, to save computational cost, an evaluation without having to train the models would be more practical. For example, the coverage of the feature space could be investigated. Another option could be to use approaches that address the explainability or interpretability of data-driven methods. This research area is concerned with understanding the decisions of methods and thus being able to explain, for example, why a model makes incorrect predictions for certain inputs. Based on these findings, similar system data can be targeted to improve the performance of the model. A survey on explainable or interpretable machine learning is presented in [352] and [353].

Assessing the similarity of similar systems is essential to make conclusions about the suitability of similar system data and to prevent negative transfers. One approach to such similarity assessment is statistical similarity measures such as those listed in Table 2. These measures can be used to evaluate the similarity based on the data of the systems. There are statistical tests in the literature that use such similarity measures. These tests originate from statistics and check whether two samples were drawn from the same population [177], [349]. However, the thresholds used for this purpose are typically very strict and may not be suitable for similar system problems. This is because methods from the area of transfer and fleet learning are developed to deal with datasets that were not drawn from the same population. Research regarding suitable threshold values is therefore essential for a data-driven assessment of the similarity of similar systems and to infer similar system data usability from similarity. One drawback of data-driven similarity assessments is that sufficient data must

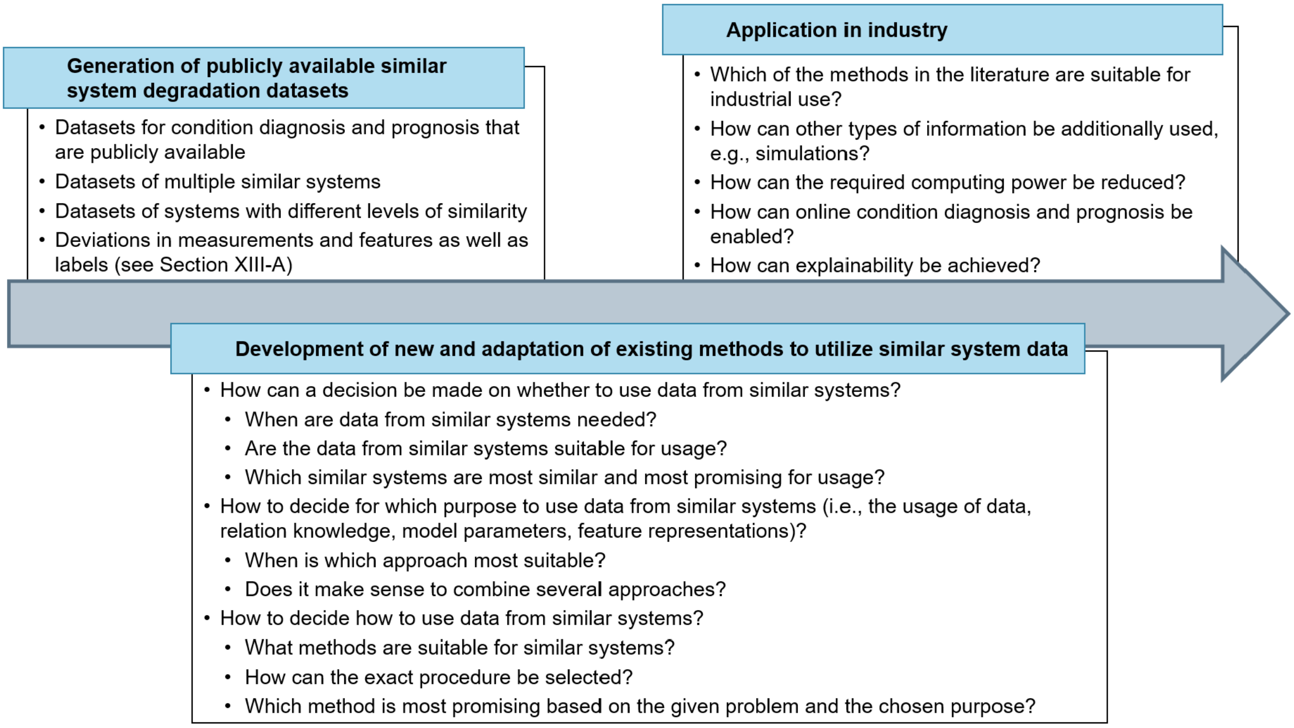

FIGURE 15. Timeline of key tasks and research questions.

be available from all the systems to be compared, i.e., also from the system under consideration itself. If this is not the case, the methods do not provide reliable results. Therefore, in addition to data-driven similarity evaluations, another approach is to evaluate the similarity of similar systems based on their technical characteristics. This has the advantage that no operating data of the systems are necessary. For example, the number of cylinders in a combustion engine is an essential characteristic that determines similarity. For structuring, concepts of ontology can be used, as shown, for example, in [73]. However, this type of similarity assessment requires expert knowledge to identify the key characteristics that determine similarity. Moreover, this approach quickly reaches its limits, especially in the case of complex technical systems. In addition to similarity checking as a means of preventing negative transfer, the research field of negative transfer learning should also be highlighted to monitor and ensure that data from similar systems do not adversely affect the training results [61]. Since these approaches are not yet widely used in the field of PHM, they have not been discussed in detail in this review.

Once it has been determined that there is promise in using data from similar systems, consideration should be given to the purpose of using the data from similar systems. The transfer approaches in transfer learning summarize the possibilities well. Pure data can be used as another data source, or knowledge already gained from the processed data can be used, i.e., relation knowledge, trained models, or feature representations. If knowledge already gained from the data is used, the training effort can also be reduced. However, there are currently no rules to decide what is most suitable under which constraints. One point to consider is the type of data available. If, for example, only unlabeled similar system data are available, it makes sense to use these data to find a suitable feature space. In addition, it can also be advantageous to combine several possibilities. There is already a trend in this direction in the literature. Nevertheless, existing approaches are also based on trial and error, without investigating why a combination of methods is promising for the problem under consideration. However, this question needs to be investigated further. Only then can an informed decision be made as to whether it makes sense to combine several possibilities.

In addition to deciding for what purpose the data from similar systems should be used, the exact procedure for doing so must also be selected. In this paper, it has been shown that approaches from transfer and fleet learning are suitable for similar system problems. However, the fact that the question of how to use knowledge from similar systems is not trivial to answer has already been shown by the large number of existing approaches to transfer and fleet learning. One example is deciding whether to perform a domain alignment using similarity measures or adversarial training, e.g., by a DANN. Similar to classical data-driven methods, there will likely be no universal answer. Nevertheless, whether a statement can be made about which procedures are most promising for certain boundary conditions should be investigated.

3) APPLICATION IN INDUSTRY

The successful application of similar system approaches in industry is another challenge. Current transfer and fleet learning methods for condition diagnosis and prognosis are usually evaluated using artificially generated data from test rigs or simulations. Thus, the similarity of the operating conditions is specifically controlled, or the deviations of the similar systems are limited to a few characteristics. However, conditions in industry usually differ significantly from these laboratory conditions. It is to be expected that the deviations in industrial applications will be significantly larger, and usually several deviations will occur simultaneously (see Section XIII-A). For instance, different sensors are typically installed in machines of different generations. Additionally, other physical quantities may even be measured, e.g., vibration data could be measured on one system and acoustic data on another. In addition, different fault classes or RUL values of different magnitudes may occur due to design differences. Although there are already initial approaches that address such problems in a mitigated form, research in this direction is still at an early stage. This has also been confirmed by Li et al. [8] and Fink et al. [25]. Future research directions include cross-modality transfer learning and partial transfer learning [61], [280], [281]. However, such approaches are currently not or are only sporadically used in PHM.

For industrial applications, there may also be few data available from similar systems. Therefore, it will be necessary to use data from several similar system types. It must be clarified how these different domains are to be addressed. For example, similarity measures could be used to control the influence of the different types on the training of a data-driven model. In the literature, approaches that utilize several source domains are referred to as multisource [56], [340], [354] and multidomain [355], [356] transfer learning. In industrial practice, artificially generated data derived from simulations or other physical models can also be a valuable source of data. In general, physical models cannot exactly reproduce the system behavior or the degradation processes of the real-world system. Therefore, similar to data from similar systems, they do not exactly fit the system under consideration. Therefore, similar system approaches are also promising for the use of simulation data. In this way, data-driven methods can benefit from existing physical models that only partially describe the system under consideration but enable the generation of much source data. Other authors also see major promise in using simulation data as another domain. For example, Fink et al. [25] noted that transfer learning is important to make better use of the data gained from simulations. In addition to using physical models to generate artificial data, there are other possibilities for integrating the knowledge contained in these models into data-driven methods. In the literature, the combination of data-driven and physical models is referred to as the hybrid approach. An overview is provided in [357]. As part of future research, it is reasonable to combine similar system and hybrid approaches.

Besides the challenges already mentioned, the typically limited computing power in industrial applications and the suitability of the approaches for online condition diagnosis or prognosis pose further challenges. Therefore, in addition to developing and training a data-driven model, other issues include integrating it into the online diagnosis or prognosis system and designing an embedded system [358]. Another point that was previously mentioned, which is also essential for industrial use, is the explainability of data-driven methods. As with all data-driven approaches, one drawback with current similar system approaches is the lack of traceability of decisions. However, to ensure that no wrong decisions are made in industrial applications, such traceability is essential. This problem is well known from classical data-driven condition diagnosis and prognosis approaches and is therefore not discussed in detail here.

Despite the challenges mentioned in this section, two promising approaches to use data from similar systems for condition diagnosis and prognosis already exist: transfer learning and fleet learning. Although the current focus is on different operating conditions and variations are relatively small when considering similar systems, they have already proven their potential. Therefore, it seems very promising to stick to these approaches and develop them further.

XIV. CONCLUSION

The use of data from similar systems offers the potential for data-driven condition diagnosis and prognosis in applications for which there are actually insufficient data from the system under consideration. However, due to the differences between similar systems, the data cannot simply be utilized. With transfer and fleet learning, research fields that are suitable for harnessing similar system data have been identified. Some of these approaches have already been applied to similar system problems. Nevertheless, most approaches currently consider identical systems under different operating conditions.

Especially in the industrial use of data-driven methods for condition diagnosis and prognosis, a sufficient database is often lacking. Similar system approaches can solve this problem by utilizing existing data from other systems. With this approach, the widespread industrial usage of data-driven methods can be made possible.

However, there remain some challenges in utilizing similar system data. This includes the lack of publicly available degradation datasets containing similar system data. Such datasets are necessary to develop and evaluate similar system approaches. In addition, solutions must be found to clarify when, for what purpose, and how existing data from similar systems can be used. At present, most approaches that can harness similar system knowledge are developed and evaluated using data recorded under laboratory conditions on test rigs or generated by simulations. Another challenge will be to transfer approaches that have proven themselves under these conditions to industrial application scenarios. It will be important to address the expected significant variations between some similar systems.

While research in the field of similar system approaches is just beginning, due to its large potential, the use of data from similar systems is seen as an important future research area in PHM.

REFERENCES

[1] Z.-Q. Wang, C.-H. Hu, and H.-D. Fan, "Real-time remaining useful life prediction for a nonlinear degrading system in service: Application to bearing data," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 211–222, Feb. 2018, doi: 10.1109/TMECH.2017.2666199.

[2] J. Chang and F. Wei, "Prognostics and health management of mechanical systems," *Adv. Mater. Res.*, vols. 694–697, pp. 872–875, May 2013, doi: 10.4028/www.scientific.net/AMR.694-697.872.

[3] D. Wang, K.-L. Tsui, and Q. Miao, "Prognostics and health management: A review of vibration based bearing and gear health indicators," *IEEE Access*, vol. 6, pp. 665–676, 2018, doi: 10.1109/ACCESS.2017.2774261.

[4] J. B. Coble, "Merging data sources to predict remaining useful life: An automated method to identify prognostic parameters," Ph.D. dissertation, Dept. Nucl. Eng., Univ. Tennessee, Knoxville, TN, USA, Jan. 2010. [Online]. Available: https://trace.tennessee.edu/utkgraddiss/683

[5] G. Niu, "Overview of data-driven PHM," in *Data-Driven Technology for Engineering Systems Health Management*, G. Niu, Ed. Singapore: Springer, 2017, pp. 35–48, doi: 10.1007/978-981-10-2032-2_3.

[6] H.-B. Jun, "An introduction to PHM and prognostics case studies," in *Advances in Production Management Systems Smart Manufacturing for Industry 4.0* (IFIP Advances in Information and Communication Technology), vol. 536, I. Moon, G. M. Lee, J. Park, D. Kiritsis, and G. Von Cieminski, Eds. Cham, Switzerland: Springer, 2018, pp. 304–310, doi: 10.1007/978-3-319-99707-0_38.

[7] K. L. Tsui, N. Chen, Q. Zhou, Y. Hai, and W. Wang, "Prognostics and health management: A review on data driven approaches," *Math. Probl. Eng.*, vol. 2015, no. 6, pp. 1–17, 2015, doi: 10.1155/2015/793161.

[8] C. Li, S. Zhang, Y. Qin, and E. Estupinan, "A systematic review of deep transfer learning for machinery fault diagnosis," *Neurocomputing*, vol. 407, pp. 121–135, Sep. 2020, doi: 10.1016/j.neucom.2020.04.045.

[9] M. Hemmer, H. Van Khang, K. Robbersmyr, T. Waag, and T. Meyer, "Fault classification of axial and radial roller bearings using transfer learning through a pretrained convolutional neural network," *Designs*, vol. 2, no. 4, p. 56, Dec. 2018, doi: 10.3390/DESIGNS2040056.

[10] T. Tinga and R. Loendersloot, "Physical model-based prognostics and health monitoring to enable predictive maintenance," in *Predictive Maintenance in Dynamic Systems*, E. Lughofer and M. Sayed-Mouchaweh, Eds. Cham, Switzerland: Springer, 2019, pp. 313–353, doi: 10.1007/978-3-030-05645-2_11.

[11] A. Mosallam, K. Medjaher, and N. Zerhouni, "Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction," *J. Intell. Manuf.*, vol. 27, no. 5, pp. 1037–1048, Oct. 2016, doi: 10.1007/s10845-014-0933-4.

[12] M. Ragab, Z. Chen, M. Wu, C. K. Kwoh, and X. Li, "Adversarial transfer learning for machine remaining useful life prediction," in *Proc. IEEE Int. Conf. Prognostics Health Manage. (ICPHM)*, Q. Miao, Ed. Jun. 2020, pp. 1–7, doi: 10.1109/ICPHM49022.2020.9187053.

[13] G. Leone, L. Cristaldi, and S. Turrin, "A data-driven prognostic approach based on statistical similarity: An application to industrial circuit breakers," *Measurement*, vol. 108, no. 1, pp. 163–170, Oct. 2017, doi: 10.1016/j.measurement.2017.02.017.

[14] L. Liao, W. Jin, and R. Pavel, "Enhanced restricted Boltzmann machine with prognosability regularization for prognostics and health assessment," *IEEE Trans. Ind. Electron.*, vol. 63, no. 11, pp. 7076–7083, Nov. 2016, doi: 10.1109/TIE.2016.2586442.

[15] M. Samie, S. Perinpanayagam, A. Alghassi, A. M. S. Motlagh, and E. Kapetanios, "Developing prognostic models using duality principles for Dc-to-Dc converters," *IEEE Trans. Power Electron.*, vol. 30, no. 5, pp. 2872–2884, May 2015, doi: 10.1109/TPEL.2014.2376413.

[16] P. R. D. O. Da Costa, A. Akçay, Y. Zhang, and U. Kaymak, "Remaining useful lifetime prediction via deep domain adaptation," *Rel. Eng. Syst. Saf.*, vol. 195, Mar. 2020, Art. no. 106682, doi: 10.1016/j.ress.2019.106682.

[17] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3757–3767, Jun. 2015, doi: 10.1109/TIE.2015.2417501.

[18] Z. Ullah, B. A. Lodhi, and J. Hur, "Detection and identification of demagnetization and bearing faults in PMSM using transfer learning-based VGG," *Energies*, vol. 13, no. 15, p. 3834, Jul. 2020, doi: 10.3390/EN13153834.

[19] W. Mao, J. He, and M. Zuo, "Predicting remaining useful life of rolling bearings based on deep feature representation and transfer learning," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1594–1608, Apr. 2020, doi: 10.1109/TIM.2019.2917735.

[20] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—Part II: Fault diagnosis with knowledge-based and hybrid/active approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3768–3774, Jun. 2015, doi: 10.1109/TIE.2015.2419013.

[21] S. Al-Dahidi, F. Di Maio, P. Baraldi, and E. Zio, "Remaining useful life estimation in heterogeneous fleets working under variable operating conditions," *Rel. Eng. Syst. Saf.*, vol. 156, pp. 109–124, Dec. 2016, doi: 10.1016/J.RESS.2016.07.019.

[22] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mech. Syst. Signal Process.*, vol. 104, pp. 799–834, May 2018, doi: 10.1016/j.ymssp.2017.11.016.

[23] G. Xu, M. Liu, Z. Jiang, W. Shen, and C. Huang, "Online fault diagnosis method based on transfer convolutional neural networks," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 2, pp. 509–520, Feb. 2020, doi: 10.1109/TIM.2019.2902003.

[24] X. Li, Y. Hu, M. Li, and J. Zheng, "Fault diagnostics between different type of components: A transfer learning approach," *Appl. Soft Comput.*, vol. 86, Jan. 2020, Art. no. 105950, doi: 10.1016/J.ASOC.2019.105950.

[25] O. Fink, Q. Wang, M. Svensén, P. Dersin, W.-J. Lee, and M. Ducoffe, "Potential, challenges and future directions for deep learning in prognostics and health management applications," *Eng. Appl. Artif. Intell.*, vol. 92, Jun. 2020, Art. no. 103678, doi: 10.1016/J.ENGAPPAI.2020.103678.

[26] M. Rigamonti, P. Baraldi, A. Alessi, E. Zio, D. Astigarraga, and A. Galarza, "An ensemble of component-based and population-based self-organizing maps for the identification of the degradation state of insulated-gate bipolar transistors," *IEEE Trans. Rel.*, vol. 67, no. 3, pp. 1304–1313, Sep. 2018, doi: 10.1109/TR.2018.2834828.

[27] G. Michau and O. Fink, "Unsupervised fault detection in varying operating conditions," in *Proc. IEEE Int. Conf. Prognostics Health Manage. (ICPHM)*, Jun. 2019, pp. 1–10, doi: 10.1109/ICPHM.2019.8819383.

[28] M. Baur, P. Albertelli, and M. Monno, "A review of prognostics and health management of machine tools," *Int. J. Adv. Manuf. Technol.*, vol. 107, nos. 5–6, pp. 2843–2863, Mar. 2020, doi: 10.1007/s00170-020-05202-3.

[29] S. Cho, G. May, and D. Kiritsis, "A predictive maintenance approach toward industry 4.0 machines," in *Engineering Assets and Public Infrastructures in the Age of Digitalization* (Lecture Notes in Mechanical Engineering), J. P. Liyanage, J. Amadi-Echendu, and J. Mathew, Eds. Cham, Switzerland: Springer, 2020, pp. 646–652, doi: 10.1007/978-3-030-48021-9_72.

[30] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, "Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data," *IEEE Trans. Ind. Electron.*, vol. 66, no. 9, pp. 7316–7325, Sep. 2018, doi: 10.1109/TIE.2018.2877090.

[31] V. Cerqueira, L. Torgo, and C. Soares, "Machine learning vs statistical methods for time series forecasting: Size matters," 2019, *arXiv:1909.13316*.

[32] T. Van Der Ploeg, P. C. Austin, and E. W. Steyerberg, "Modern modelling techniques are data hungry: A simulation study for predicting dichotomous endpoints," *BMC Med. Res. Methodol.*, vol. 14, no. 1, p. 137, Dec. 2014, doi: 10.1186/1471-2288-14-137.

[33] H. S. R. Rajula, G. Verlato, M. Manchia, N. Antonucci, and V. Fanos, "Comparison of conventional statistical methods with machine learning in medicine: Diagnosis, drug development, and treatment," *Medicina*, vol. 56, no. 9, p. 455, Sep. 2020, doi: 10.3390/medicina56090455.

[34] B. Kitchenham. *Procedures for Performing Systematic Reviews*. Keele, U.K. [Online]. Available: https://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf

[35] C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in *Proc. 18th Int. Conf. Eval. Assessment Softw. Eng. (EASE)*, 2014, pp. 1–10, doi: 10.1145/2601248.2601268.

[36] B. Maschler and M. Weyrich, "Deep transfer learning for industrial automation: A review and discussion of new techniques for data-driven machine learning," *IEEE Ind. Electron. Mag.*, vol. 15, no. 2, pp. 65–75, Jun. 2021, doi: 10.1109/MIE.2020.3034884.

[37] R. Moradi and K. M. Groth, "On the application of transfer learning in prognostics and health management," in *Proc. Annu. Conf. PHM Soc. (PHM)*, A. Saxena, Ed., 2020, pp. 1–8.

[38] R. Yan, F. Shen, C. Sun, and X. Chen, "Knowledge transfer for rotary machine fault diagnosis," *IEEE Sensors J.*, vol. 20, no. 15, pp. 8374–8393, Aug. 2020, doi: 10.1109/JSEN.2019.2949057.

[39] A. Yunusa-Kaltungo, J. K. Sinha, and A. D. Nembhard, "A novel fault diagnosis technique for enhancing maintenance and reliability of rotating machines," *Structural Health Monitor., An Int. J.*, vol. 14, no. 6, pp. 604–621, Nov. 2015, doi: 10.1177/1475921715604388.

[40] Z. Meili, "Principles and practice of similarity system theory," *Int. J. Gen. Syst.*, vol. 23, no. 1, pp. 39–48, Oct. 1994, doi: 10.1080/03081079408908028.

[41] E. Hornbogen, G. Eggeler, and E. Werner, *Werkstoffe: Aufbau und Eigenschaften Von Keramik-, Metall-, Polymer- und Verbundwerkstoffen* (Springer-Lehrbuch), 10th ed. Berlin, Germany: Springer-Verlag, 2012, doi: 10.1007/978-3-642-22561-1.

[42] R. Patrick, M. J. Smith, C. S. Byington, G. J. Vachtsevanos, K. Tom, and C. Ly, "Integrated software platform for fleet data analysis, enhanced diagnostics, and safe transition to prognostics for helicopter component CBM," in *Proc. Annu. Conf. Prognostics Health Manage. Soc. (PHM)*, 2010, pp. 1–15.

[43] Z. Wang, Q. Liu, H. Chen, and X. Chu, "A deformable CNN-DLSTM based transfer learning method for fault diagnosis of rolling bearing under multiple working conditions," *Int. J. Prod. Res.*, vol. 59, no. 16, pp. 1–15, 2020, doi: 10.1080/00207543.2020.1808261.

[44] Z. Du, B. Yang, Y. Lei, X. Li, and N. Li, "A hybrid transfer learning method for fault diagnosis of machinery under variable operating conditions," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Qingdao)*, Oct. 2019, pp. 1–5, doi: 10.1109/PHM-Qingdao46334.2019.8942974.

[45] G. Michau and O. Fink, "Transferring complementary operating conditions for anomaly detection," ETH Zürich, Zürich, Switzerland, Tech. Rep., 2020. [Online]. Available: https://arxiv.org/abs/2008.07815v1

[46] G. Michau and O. Fink, "Unsupervised transfer learning for anomaly detection: Application to complementary operating condition transfer," *Knowl.-Based Syst.*, vol. 216, Mar. 2021, Art. no. 106816, doi: 10.1016/j.knosys.2021.106816.

[47] R. Gupta, K. K. Bhardwaj, and D. K. Sharma, "Transfer learning," in *Machine Learning and Big Data*, U. N. Dulhare, K. Ahmad, and K. A. B. Ahmad, Eds. Hoboken, NJ, USA: Wiley-Scrivener, 2020, pp. 337–360, doi: 10.1002/9781119654834.ch13.

[48] A. Farahani, B. Pourshojae, K. Rasheed, and H. R. Arabnia, "A concise review of transfer learning," in *Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI)*, Dec. 2020, pp. 344–351.

[49] Y. Pan, F. Mei, H. Miao, J. Zheng, K. Zhu, and H. Sha, "An approach for HVCB mechanical fault diagnosis based on a deep belief network and a transfer learning strategy," *J. Elect. Eng. Technol.*, vol. 14, no. 1, pp. 407–419, Jan. 2019, doi: 10.1007/S42835-018-00048-Y.

[50] Q. Wang, G. Michau, and O. Fink, "Domain adaptive transfer learning for fault diagnosis," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Paris)*, May 2019, pp. 279–285, doi: 10.1109/PHM-Paris.2019.00054.

[51] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, p. 9, 2016, doi: 10.1186/S40537-016-0043-6.

[52] H. Tercan, A. Guajardo, and T. Meisen, "Industrial transfer learning: Boosting machine learning in production," in *Proc. IEEE 17th Int. Conf. Ind. Informat. (INDIN)*, Jul. 2019, pp. 274–279, doi: 10.1109/INDIN41052.2019.8972099.

[53] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Dec. 2009, doi: 10.1109/TKDE.2009.191.

[54] T. Cody, S. Adams, P. A. Beling, S. Polter, K. Farinholt, N. Hipwell, A. Chaudhry, K. Castillo, and R. Meekins, "Transferring random samples in actuator systems for binary damage detection," in *Proc. IEEE Int. Conf. Prognostics Health Manage. (ICPHM)*, Jun. 2019, pp. 1–7, doi: 10.1109/ICPHM.2019.8819393.

[55] N. Agarwal, A. Sondhi, K. Chopra, and G. Singh, "Transfer learning: Survey and classification," in *Proc. Smart Innov. Commun. Comput. Sci., Proc. Int. Conf. Smart Innov. Commun. Comput. Sci. (ICSICCS)*, S. Tiwari, M. C. Trivedi, K. K. Mishra, A. K. Misra, K. K. Kumar, and E. Suryani, Eds. Singapore: Springer, 2020, pp. 145–155, doi: 10.1007/978-981-15-5345-5_13.

[56] S. Sun, H. Shi, and Y. Wu, "A survey of multi-source domain adaptation," *Inf. Fusion*, vol. 24, pp. 84–92, Jul. 2015, doi: 10.1016/j.inffus.2014.12.003.

[57] X. Wang and M. Jin, "Convolutional domain adaptation network for fault diagnosis of thermal system under different loading conditions," in *Proc. 39th Chin. Control Conf. (CCC)*, Jul. 2020, pp. 4193–4197, doi: 10.23919/CCC50068.2020.9189493.

[58] G. Jin, K. Xu, H. Chen, Y. Jin, and C. Zhu, "A novel multi-adversarial cross-domain neural network for bearing fault diagnosis," *Meas. Sci. Technol.*, vol. 32, no. 5, May 2021, Art. no. 055102, doi: 10.1088/1361-6501/abd900.

[59] B. Zadrozny, "Learning and evaluating classifiers under sample selection bias," in *Proc. 41st Annu. Design Autom. Conf. (DAC)*, C. Brodley, Ed. 2004, p. 114, doi: 10.1145/1015330.1015425.

[60] H. Shimodaira, "Improving predictive inference under covariate shift by weighting the log-likelihood function," *J. Statist. Planning Inference*, vol. 90, no. 2, pp. 227–244, 2000, doi: 10.1016/S0378-3758(00)00115-4.

[61] S. Niu, Y. Liu, J. Wang, and H. Song, "A decade survey of transfer learning (2010–2020)," *IEEE Trans. Artif. Intell.*, vol. 1, no. 2, pp. 151–166, Oct. 2020, doi: 10.1109/TAI.2021.3054609.

[62] L. Duan, D. Xu, and S.-F. Chang, "Exploiting web images for event recognition in consumer videos: A multiple source domain adaptation approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1338–1345, doi: 10.1109/CVPR.2012.6247819.

[63] R. Chattopadhyay, Q. Sun, W. Fan, I. Davidson, S. Panchanathan, and J. Ye, "Multisource domain adaptation and its application to early detection of fatigue," *ACM Trans. Knowl. Discovery Data*, vol. 6, no. 4, pp. 1–26, 2012, doi: 10.1145/2382577.2382582.

[64] Z. Wang, Y. Song, and C. Zhang, "Transferred dimensionality reduction," in *Proc. Mach. Learn. Knowl. Discovery Databases, Eur. Conf. Mach. Learn. Princ. Pract. Knowl. Discovery Databases (ECML PKDD)*, in Lecture notes in computer science Lecture notes in artificial intelligence, vol. 5212, W. Daelemans, B. Goethals, and K. Morik, Eds. Cham, Switzerland: Springer, 2008, pp. 550–565, doi: 10.1007/978-3-540-87481-2_36.

[65] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jul. 2020, doi: 10.1109/JPROC.2020.3004555.

[66] F. Yang, W. Zhang, L. Tao, and J. Ma, "Transfer learning strategies for deep learning-based PHM algorithms," *Appl. Sci.*, vol. 10, no. 7, p. 2361, Mar. 2020, doi: 10.3390/APP10072361.

[67] S. Subbiah and S. Turrin, "Extraction and exploitation of R&M knowledge from a fleet perspective," in *Proc. Annu. Rel. Maintainability Symp. (RAMS)*, Jan. 2015, pp. 1–6.

[68] K. Hendrickx, W. Meert, B. Cornelis, K. Janssens, K. Gryllias, and J. Davis, "A fleet-wide approach for condition monitoring of similar machines using time-series clustering," in *Proc. 6th Int. Conf. Condition Monitor. Machinery Non-Stationary Operations (CMMNO)*, in Applied Condition Monitoring, vol. 15, A. F. Del Rincon, F. V. Rueda, F. Chaari, R. Zimroz, and M. Haddar, Eds. Cham, Switzerland: Springer, 2019, pp. 101–110, doi: 10.1007/978-3-030-11220-2_11.

[69] C. Jin, D. Djurdjanovic, H. D. Ardakani, K. Wang, M. Buzza, B. Begheri, P. Brown, and J. Lee, "A comprehensive framework of factory-to-factory dynamic fleet-level prognostics and operation management for geographically distributed assets," in *Proc. IEEE Int. Conf. Autom. Sci. Eng. (CASE)*, Aug. 2015, pp. 225–230, doi: 10.1109/CoASE.2015.7294066.

[70] G. Michau, T. Palmé, and O. Fink, "Fleet PHM for critical systems: Bi-level deep learning approach for fault detection," in *Proc. 4th Eur. Conf. PHM Soc. (PHME)*, vol. 4, C. S. Kulkarni and T. Tinga, Eds., 2018, p. 403.

[71] Z. Liu, "Cyber-physical system augmented prognostics and health management for fleet-based systems," Ph.D. dissertation, Dept. Mech. Mater. Eng., Univ. Cincinnati, Cincinnatim, OH, USA, 2018. [Online]. Available: https://etd.ohiolink.edu/apexprod/rws_olink/r/1501/10?clear=10&p10_accession_num=ucin1522321192371536

[72] G. Leone, L. Cristaldi, and S. Turrin, "A data-driven prognostic approach based on sub-fleet knowledge extraction," in *Proc. 14th IMEKO TC10 Workshop Tech. Diag.*, 2016, pp. 417–422.

[73] G. Medina-Oliva, A. Voisin, M. Monnin, and J.-B. Leger, "Predictive diagnosis based on a fleet-wide ontology approach," *Knowl.-Based Syst.*, vol. 68, pp. 40–57, Sep. 2014, doi: 10.1016/j.knosys.2013.12.020.

[74] S. Al-Dahidi, F. Di Maio, P. Baraldi, E. Zio, and R. Seraoui, "A framework for reconciliating data clusters from a fleet of nuclear power plants turbines for fault diagnosis," *Appl. Soft Comput.*, vol. 69, pp. 213–231, Aug. 2018, doi: 10.1016/j.asoc.2018.04.044.

[75] K. Goebel, B. Saha, and A. Saxena, "A comparison of three data-driven techniques for prognostics," in *Proc. 62nd Meeting Soc. Machinery Failure Prevention Technol. (MFPT)*. Dresher, PA, USA: Society for Machinery Failure Prevention Technology, 2008, pp. 119–131.

[76] L. Von Rueden, S. Mayer, K. Beckh, B. Georgiev, S. Giesselbach, R. Heese, B. Kirsch, M. Walczak, J. Pfrommer, A. Pick, R. Ramamurthy, J. Garcke, C. Bauckhage, and J. Schuecker, "Informed machine learning—A taxonomy and survey of integrating prior knowledge into learning systems," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 1, pp. 614–633, Jan. 2021, doi: 10.1109/TKDE.2021.3079836.

[77] Y. Gao and K. M. Mosalam, "Deep transfer learning for image-based structural damage recognition," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 9, pp. 748–768, 2018, doi: 10.1111/mice.12363.

[78] Y. Yu, H. Cao, X. Yan, T. Wang, and S. S. Ge, "Defect identification of wind turbine blades based on defect semantic features with transfer feature extractor," *Neurocomputing*, vol. 376, pp. 1–9, Feb. 2020, doi: 10.1016/j.neucom.2019.09.071.

[79] H. Mamledesai, M. A. Soriano, and R. Ahmad, "A qualitative tool condition monitoring framework using convolution neural network and transfer learning," *Appl. Sci.*, vol. 10, no. 20, p. 7298, Oct. 2020, doi: 10.3390/app10207298.

[80] J. Duan, Y. He, B. Du, R. M. R. Ghandour, W. Wu, and H. Zhang, "Intelligent localization of transformer internal degradations combining deep convolutional neural networks and image segmentation," *IEEE Access*, vol. 7, pp. 62705–62720, 2019, doi: 10.1109/ACCESS.2019.2916461.

[81] J. Wang, Z. Mo, H. Zhang, and Q. Miao, "A deep learning method for bearing fault diagnosis based on time-frequency image," *IEEE Access*, vol. 7, pp. 42373–42383, 2019, doi: 10.1109/ACCESS.2019.2907131.

[82] Z. Wang and T. Oates, "Imaging time-series to improve classification and imputation," in *Proc. 24th Int. Joint Conf. Artif. Intell. (IJCAI)*, Q. Yang and M. Wooldridge, Eds., 2015, p. 3939.

[83] Y. Bai, J. Yang, J. Wang, and Q. Li, "Intelligent diagnosis for railway wheel flat using frequency-domain Gramian angular field and transfer learning network," *IEEE Access*, vol. 8, pp. 105118–105126, 2020, doi: 10.1109/ACCESS.2020.3000068.

[84] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, Jul. 2018, doi: 10.1109/TIE.2017.2774777.

[85] L. Wen, X. Li, X. Li, and L. Gao, "A new transfer learning based on VGG-19 network for fault diagnosis," in *Proc. IEEE 23rd Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD)*, W. Shen, Ed., May 2019, pp. 205–209, doi: 10.1109/CSCWD.2019.8791884.

[86] L. Wen, X. Li, and L. Gao, "A transfer convolutional neural network for fault diagnosis based on ResNet-50," *Neural Comput. Appl.*, vol. 32, no. 10, pp. 6111–6124, May 2020, doi: 10.1007/s00521-019-04097-w.

[87] P. Cao, S. Zhang, and J. Tang, "Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning," *IEEE Access*, vol. 6, pp. 26241–26253, 2018, doi: 10.1109/ACCESS.2018.2837621.

[88] Y. Li and J. Tao, "CNN and transfer learning based online SOH estimation for lithium-ion battery," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Aug. 2020, pp. 5489–5494, doi: 10.1109/CCDC49329.2020.9164208.

[89] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, Y. Bengio and Y. LeCun, Eds., 2015, pp. 1–14.

[90] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[91] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[92] L. Wen, L. Gao, Y. Dong, and Z. Zhu, "A negative correlation ensemble transfer learning method for fault diagnosis based on convolutional neural network," *Math. Biosci. Eng.*, vol. 16, no. 5, pp. 3311–3330, 2019, doi: 10.3934/mbe.2019165.

[93] S. Shao, S. McAleer, R. Yan, and P. Baldi, "Highly accurate machine fault diagnosis using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2446–2455, Apr. 2019, doi: 10.1109/TII.2018.2864759.

[94] P. Ma, H. Zhang, W. Fan, C. Wang, G. Wen, and X. Zhang, "A novel bearing fault diagnosis method based on 2D image representation and transfer learning-convolutional neural network," *Meas. Sci. Technol.*, vol. 30, no. 5, May 2019, Art. no. 055402, doi: 10.1088/1361-6501/ab0793.

[95] D. Zhang and T. Zhou, "Deep convolutional neural network using transfer learning for fault diagnosis," *IEEE Access*, vol. 9, pp. 43889–43897, 2021, doi: 10.1109/ACCESS.2021.3061530.

[96] W. Mao, L. Ding, S. Tian, and X. Liang, "Online detection for bearing incipient fault based on deep transfer learning," *Measurement*, vol. 152, Feb. 2020, Art. no. 107278, doi: 10.1016/j.measurement.2019.107278.

[97] M. El-Dalahmeh, M. Al-Greer, M. El-Dalahmeh, and M. Short, "Time-frequency image analysis and transfer learning for capacity prediction of lithium-ion batteries," *Energies*, vol. 13, no. 20, p. 5447, Oct. 2020, doi: 10.3390/en13205447.

[98] H. Zhang, Q. Zhang, S. Shao, T. Niu, X. Yang, and H. Ding, "Sequential network with residual neural network for rotatory machine remaining useful life prediction using deep transfer learning," *Shock Vibrat.*, vol. 2020, pp. 1–16, Sep. 2020, doi: 10.1155/2020/8888627.

[99] Z. Zheng, J. Fu, C. Lu, and Y. Zhu, "Research on rolling bearing fault diagnosis of small dataset based on a new optimal transfer learning network," *Measurement*, vol. 177, Jun. 2021, Art. no. 109285, doi: 10.1016/j.measurement.2021.109285.

[100] M. J. Hasan, M. Sohaib, and J.-M. Kim, "A multitask-aided transfer learning-based diagnostic framework for bearings under inconsistent working conditions," *Sensors*, vol. 20, no. 24, p. 7205, Dec. 2020, doi: 10.3390/s20247205.

[101] M. J. Hasan and J.-M. Kim, "Bearing fault diagnosis under variable rotational speeds using Stockwell transform-based vibration imaging and transfer learning," *Appl. Sci.*, vol. 8, no. 12, p. 2357, Nov. 2018, doi: 10.3390/APP8122357.

[102] Z. Shi-Sheng, F. Song, and L. Lin, "A novel gas turbine fault diagnosis method based on transfer learning with CNN," *Measurement*, vol. 137, pp. 435–453, Apr. 2019, doi: 10.1016/j.measurement.2019.01.022.

[103] W. Liu, Z. Chen, and M. Zheng, "An audio-based fault diagnosis method for quadrotors using convolutional neural network and transfer learning," in *Proc. Amer. Control Conf. (ACC)*, M. Oishi, Ed., Jul. 2020, pp. 1367–1372, doi: 10.23919/ACC45564.2020.9148044.

[104] Y. Li, K. Li, X. Liu, Y. Wang, and L. Zhang, "Lithium-ion battery capacity estimation—A pruned convolutional neural network approach assisted with transfer learning," *Appl. Energy*, vol. 285, Mar. 2021, Art. no. 116410, doi: 10.1016/j.apenergy.2020.116410.

[105] S. Shen, M. Sadoughi, M. Li, Z. Wang, and C. Hu, "Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries," *Appl. Energy*, vol. 260, Feb. 2020, Art. no. 114296, doi: 10.1016/j.apenergy.2019.114296.

[106] X. Yang, M. Bai, J. Liu, J. Liu, and D. Yu, "Gas path fault diagnosis for gas turbine group based on deep transfer learning," *Measurement*, vol. 181, Aug. 2021, Art. no. 109631, doi: 10.1016/j.measurement.2021.109631.

[107] M. J. Hasan, M. M. M. Islam, and J.-M. Kim, "Multi-sensor fusion-based time-frequency imaging and transfer learning for spherical tank crack diagnosis under variable pressure conditions," *Measurement*, vol. 168, Jan. 2021, Art. no. 108478, doi: 10.1016/j.measurement.2020.108478.

[108] A. Kumar, G. Vashishtha, C. P. Gandhi, H. Tang, and J. Xiang, "Sparse transfer learning for identifying rotor and gear defects in the mechanical machinery," *Measurement*, vol. 179, Jul. 2021, Art. no. 109494, doi: 10.1016/j.measurement.2021.109494.

[109] J. Duan, Y. He, and X. Wu, "Serial transfer learning (STL) theory for processing data insufficiency: Fault diagnosis of transformer windings," *Int. J. Electr. Power Energy Syst.*, vol. 130, Sep. 2021, Art. no. 106965, doi: 10.1016/j.ijepes.2021.106965.

[110] H. Kim and B. D. Youn, "A new parameter repurposing method for parameter transfer with small dataset and its application in fault diagnosis of rolling element bearings," *IEEE Access*, vol. 7, pp. 46917–46930, 2019, doi: 10.1109/ACCESS.2019.2906273.

[111] S. Chen, H. Ge, H. Li, Y. Sun, and X. Qian, "Hierarchical deep convolution neural networks based on transfer learning for transformer rectifier unit fault diagnosis," *Measurement*, vol. 167, Jan. 2021, Art. no. 108257, doi: 10.1016/j.measurement.2020.108257.

[112] Z. Chen, K. Gryllias, and W. Li, "Intelligent fault diagnosis for rotary machinery using transferable convolutional neural network," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 339–349, Jan. 2020, doi: 10.1109/TII.2019.2917233.

[113] G. Wang, M. Zhang, L. Xiang, Z. Hu, W. Li, and J. Cao, "A multi-branch convolutional transfer learning diagnostic method for bearings under diverse working conditions and devices," *Measurement*, vol. 182, Sep. 2021, Art. no. 109627, doi: 10.1016/j.measurement.2021.109627.

[114] A. Graves, "Long short-term memory," in *Supervised Sequence Labelling With Recurrent Neural Networks* (Studies in Computational Intelligence), vol. 385. Berlin, Germany: Springer, 2012, pp. 37–45, doi: 10.1007/978-3-642-24797-2_4.

[115] D. Zhu, X. Song, J. Yang, Y. Cong, and L. Wang, "A bearing fault diagnosis method based on L1 regularization transfer learning and LSTM deep learning," in *Proc. IEEE Int. Conf. Inf. Commun. Softw. Eng. (ICICSE)*, Mar. 2021, pp. 308–312, doi: 10.1109/ICICSE52190.2021.9404081.

[116] Y. Tan and G. Zhao, "Transfer learning with long short-term memory network for state-of-health prediction of lithium-ion batteries," *IEEE Trans. Ind. Electron.*, vol. 67, no. 10, pp. 8723–8731, Oct. 2020, doi: 10.1109/TIE.2019.2946551.

[117] Y. Che, Z. Deng, X. Lin, L. Hu, and X. Hu, "Predictive battery health management with transfer learning and online model correction," *IEEE Trans. Transport. Electrific.*, vol. 70, pp. 1269–1277, 2021, doi: 10.1109/TVT.2021.3055811.

[118] S. Kim, Y. Y. Choi, K. J. Kim, and J.-I. Choi, "Forecasting state-of-health of lithium-ion batteries using variational long short-term memory with transfer learning," *J. Energy Storage*, vol. 41, Sep. 2021, Art. no. 102893, doi: 10.1016/j.est.2021.102893.

[119] Y. Liu, X. Shu, H. Yu, J. Shen, Y. Zhang, Y. Liu, and Z. Chen, "State of charge prediction framework for lithium-ion batteries incorporating long short-term memory network and transfer learning," *J. Energy Storage*, vol. 37, May 2021, Art. no. 102494, doi: 10.1016/j.est.2021.102494.

[120] J. Li, X. Li, D. He, and Y. Qu, "A domain adaptation model for early gear pitting fault diagnosis based on deep transfer learning network," *Proc. Inst. Mech. Engineers, O, J. Risk Rel.*, vol. 234, no. 1, pp. 168–182, Feb. 2020, doi: 10.1177/1748006X19867776.

[121] Z. Deng, Z. Wang, Z. Tang, K. Huang, and H. Zhu, "A deep transfer learning method based on stacked autoencoder for cross-domain fault diagnosis," *Appl. Math. Comput.*, vol. 408, Nov. 2021, Art. no. 126318, doi: 10.1016/j.amc.2021.126318.

[122] X. Li, H. Jiang, K. Zhao, and R. Wang, "A deep transfer nonnegativity-constraint sparse autoencoder for rolling bearing fault diagnosis with few labeled data," *IEEE Access*, vol. 7, pp. 91216–91224, 2019, doi: 10.1109/ACCESS.2019.2926234.

[123] Z. He, H. Shao, X. Zhang, J. Cheng, and Y. Yang, "Improved deep transfer auto-encoder for fault diagnosis of gearbox under variable working conditions with small training samples," *IEEE Access*, vol. 7, pp. 115368–115377, 2019, doi: 10.1109/ACCESS.2019.2936243.

[124] D. Chen, S. Yang, and F. Zhou, "Transfer learning based fault diagnosis with missing data due to multi-rate sampling," *Sensors*, vol. 19, no. 8, p. 1826, Apr. 2019, doi: 10.3390/s19081826.

[125] D. Chen, S. Yang, and F. Zhou, "Incipient fault diagnosis based on DNN with transfer learning," in *Proc. Int. Conf. Control, Autom. Inf. Sci. (ICCAIS)*, C. Wen, Ed., Oct. 2018, pp. 303–308, doi: 10.1109/ICCAIS.2018.8570702.

[126] Y. Li, W. Jiang, G. Zhang, and L. Shu, "Wind turbine fault diagnosis based on transfer learning and convolutional autoencoder with small-scale data," *Renew. Energy*, vol. 171, pp. 103–115, Jun. 2021, doi: 10.1016/j.renene.2021.01.143.

[127] Z. Di, H. Shao, and J. Xiang, "Ensemble deep transfer learning driven by multisensor signals for the fault diagnosis of bevel-gear cross-operation conditions," *Sci. China Technol. Sci.*, vol. 64, no. 3, pp. 481–492, Mar. 2021, doi: 10.1007/s11431-020-1679-x.

[128] J. Pan, K. L. Low, J. Ghosh, S. Jayavelu, M. M. Ferdaus, S. Y. Lim, E. Zamburg, Y. Li, B. Tang, X. Wang, J. F. Leong, S. Ramasamy, T. Buonassisi, C.-K. Tham, and A. Thean, "Transfer learning-based artificial intelligence-integrated physical modeling to enable failure analysis for 3 nanometer and smaller silicon-based CMOS transistors," *ACS Appl. Nano Mater.*, vol. 4, no. 7, pp. 6903–6915, 2021, doi: 10.1021/acsanm.1c00960.

[129] M. Gribbestad, M. U. Hassan, and I. A. Hameed, "Transfer learning for prognostics and health management (PHM) of marine air compressors," *J. Mar. Sci. Eng.*, vol. 9, no. 1, p. 47, Jan. 2021, doi: 10.3390/jmse9010047.

[130] C. Zhang, L. Xu, X. Li, and H. Wang, "A method of fault diagnosis for rotary equipment based on deep learning," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Chongqing)*, Oct. 2018, pp. 958–962, doi: 10.1109/PHM-Chongqing.2018.00171.

[131] R. Zhang, H. Tao, L. Wu, and Y. Guan, "Transfer learning with neural networks for bearing fault diagnosis in changing working conditions," *IEEE Access*, vol. 5, pp. 14347–14357, 2017, doi: 10.1109/ACCESS.2017.2720965.

[132] S. H. Cho, S. Kim, and J.-H. Choi, "Transfer learning-based fault diagnosis under data deficiency," *Appl. Sci.*, vol. 10, no. 21, p. 7768, Nov. 2020, doi: 10.3390/app10217768.

[133] X. Liu, L. Liu, L. Wang, and X. Peng, "Improved condition monitoring of aircraft auxiliary power unit based on transfer learning," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Qingdao)*, Oct. 2019, pp. 1–6, doi: 10.1109/PHM-Qingdao46334.2019.8942809.

[134] A. Zhang, H. Wang, S. Li, Y. Cui, Z. Liu, G. Yang, and J. Hu, "Transfer learning with deep recurrent neural networks for remaining useful life estimation," *Appl. Sci.*, vol. 8, no. 12, p. 2416, Nov. 2018, doi: 10.3390/app8122416.

[135] D. Guo, G. Yang, X. Han, X. Feng, L. Lu, and M. Ouyang, "Parameter identification of fractional-order model with transfer learning for aging lithium-ion batteries," *Int. J. Energy Res.*, vol. 45, no. 9, pp. 12825–12837, Jul. 2021, doi: 10.1002/er.6614.

[136] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, "Boosting for transfer learning," in *Proc. 24th Int. Conf. Mach. Learn. (ICML)*, Z. Ghahramani, Ed. 2007, pp. 193–200, doi: 10.1145/1273496.1273521.

[137] F. Shen, C. Chen, R. Yan, and R. X. Gao, "Bearing fault diagnosis based on SVD feature extraction and transfer learning classification," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM)*, Oct. 2015, pp. 1–6, doi: 10.1109/PHM.2015.7380088.

[138] S. Dong, J. Wang, B. Tang, X. Xu, and J. Liu, "Bearing faults recognition based on the local mode decomposition morphological spectrum and optimized support vector machine," in *Proc. 13th Int. Conf. Ubiquitous Robots Ambient Intell. (URAI)*, Aug. 2016, pp. 1012–1016, doi: 10.1109/URAI.2016.7734128.

[139] J. Zhan, K. Zhou, Y. H. Wang, and Y. H. Wang, "Minority disk failure prediction based on transfer learning in large data centers of heterogeneous disk systems," *IEEE Trans. Parallel Distrib. Syst.*, vol. 31, no. 9, pp. 2155–2169, Apr. 2020, doi: 10.1109/TPDS.2020.2985346.

[140] D. Xiao, Y. Huang, C. Qin, Z. Liu, Y. Li, and C. Liu, "Transfer learning with convolutional neural networks for small sample size problem in machinery fault diagnosis," *Proc. Inst. Mech. Eng. C, J. Mech. Eng. Sci.*, vol. 233, no. 14, pp. 5131–5143, Jul. 2019, doi: 10.1177/0954406219840381.

[141] S. Tang, H. Tang, and M. Chen, "Transfer-learning based gas path analysis method for gas turbines," *Appl. Thermal Eng.*, vol. 155, pp. 1–13, Jun. 2019, doi: 10.1016/j.applthermaleng.2019.03.156.

[142] W. Wang, J. Zhang, and Q. Zhang, "Transfer learning based diagnosis for configuration troubleshooting in self-organizing femtocell networks," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2011, pp. 1–5, doi: 10.1109/GLOCOM.2011.6133802.

[143] K. Lee, S. Han, V. H. Pham, S. Cho, H.-J. Choi, J. Lee, I. Noh, and S. W. Lee, "Multi-objective instance weighting-based deep transfer learning network for intelligent fault diagnosis," *Appl. Sci.*, vol. 11, no. 5, p. 2370, Mar. 2021, doi: 10.3390/app11052370.

[144] D. Zhou, J. Hao, D. Huang, X. Jia, and H. Zhang, "Dynamic simulation of gas turbines via feature similarity-based transfer learning," *Frontiers Energy*, vol. 14, no. 4, pp. 817–835, Dec. 2020, doi: 10.1007/s11708-020-0709-9.

[145] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang, "Deep model based domain adaptation for fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 64, no. 3, pp. 2296–2305, Mar. 2017, doi: 10.1109/TIE.2016.2627020.

[146] Z.-H. Liu, B.-L. Lu, H.-L. Wei, X.-H. Li, and L. Chen, "Fault diagnosis for electromechanical drivetrains using a joint distribution optimal deep domain adaptation approach," *IEEE Sensors J.*, vol. 19, no. 24, pp. 12261–12270, Sep. 2019, doi: 10.1109/JSEN.2019.2939360.

[147] W. Xu, Y. Wan, T.-Y. Zuo, and X.-M. Sha, "Transfer learning based data feature transfer for fault diagnosis," *IEEE Access*, vol. 8, pp. 76120–76129, 2020, doi: 10.1109/ACCESS.2020.2989510.

[148] X. Li, W. Zhang, H. Ma, Z. Luo, and X. Li, "Deep learning-based adversarial multi-classifier optimization for cross-domain machinery fault diagnostics," *J. Manuf. Syst.*, vol. 55, pp. 334–347, Apr. 2020, doi: 10.1016/j.jmsy.2020.04.017.

[149] C. Cheng, B. Zhou, G. Ma, D. Wu, and Y. Yuan, "Wasserstein distance based deep adversarial transfer learning for intelligent fault diagnosis with unlabeled or insufficient labeled data," *Neurocomputing*, vol. 409, pp. 35–45, Oct. 2020, doi: 10.1016/j.neucom.2020.05.040.

[150] N.-X. Xu and X. Li, "Cross-domain machinery fault diagnosis using adversarial network with conditional alignments," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Qingdao)*, Oct. 2019, pp. 1–5, doi: 10.1109/PHM-Qingdao46334.2019.8943041.

[151] T. Han, C. Liu, R. Wu, and D. Jiang, "Deep transfer learning with limited data for machinery fault diagnosis," *Appl. Soft Comput.*, vol. 103, May 2021, Art. no. 107150, doi: 10.1016/j.asoc.2021.107150.

[152] X. Li, W. Zhang, Q. Ding, and X. Li, "Diagnosing rotating machines with weakly supervised data using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 1688–1697, Jul. 2020, doi: 10.1109/TII.2019.2927590.

[153] W. Mao, L. Ding, Y. Liu, S. S. Afshari, and X. Liang, "A new deep domain adaptation method with joint adversarial training for online detection of bearing early fault," *ISA Trans.*, vol. 122, pp. 444–458, Mar. 2022, doi: 10.1016/j.isatra.2021.04.026.

[154] X. Zhang, B. Han, J. Wang, Z. Zhang, and Z. Yan, "A novel transfer-learning method based on selective normalization for fault diagnosis with limited labeled data," *Meas. Sci. Technol.*, vol. 32, no. 10, Oct. 2021, Art. no. 105116, doi: 10.1088/1361-6501/ac03e5.

[155] X. Li and W. Zhang, "Deep learning-based partial domain adaptation method on intelligent machinery fault diagnostics," *IEEE Trans. Ind. Electron.*, vol. 68, no. 5, pp. 4351–4361, May 2021, doi: 10.1109/TIE.2020.2984968.

[156] S. Lu, T. Sirojan, B. T. Phung, D. Zhang, and E. Ambikairajah, "DA-DCGAN: An effective methodology for DC series arc fault diagnosis in photovoltaic systems," *IEEE Access*, vol. 7, pp. 45831–45840, 2019, doi: 10.1109/ACCESS.2019.2909267.

[157] C. Shi, Z. Wu, X. Lv, and Y. Ji, "DGTL-Net: A deep generative transfer learning network for fault diagnostics on new hard disks," *Exp. Syst. Appl.*, vol. 169, May 2021, Art. no. 114379, doi: 10.1016/j.eswa.2020.114379.

[158] Y. Xie and T. Zhang, "A transfer learning strategy for rotation machinery fault diagnosis based on cycle-consistent generative adversarial networks," in *Proc. Chin. Autom. Congr. (CAC)*, Nov. 2018, pp. 1309–1313, doi: 10.1109/CAC.2018.8623346.

[159] Y. Fan, X. Cui, H. Han, and H. Lu, "Chiller fault detection and diagnosis by knowledge transfer based on adaptive imbalanced processing," *Sci. Technol. Built Environ.*, vol. 26, no. 8, pp. 1082–1099, Sep. 2020, doi: 10.1080/23744731.2020.1757327.

[160] X. Wang, Z. Chu, B. Han, J. Wang, G. Zhang, and X. Jiang, "A novel data augmentation method for intelligent fault diagnosis under speed fluctuation condition," *IEEE Access*, vol. 8, pp. 143383–143396, 2020, doi: 10.1109/ACCESS.2020.3014340.

[161] J. An, P. Ai, and D. Liu, "Deep domain adaptation model for bearing fault diagnosis with domain alignment and discriminative feature learning," *Shock Vibrat.*, vol. 2020, pp. 1–14, Mar. 2020, doi: 10.1155/2020/4676701.

[162] T. Jin, C. Yan, C. Chen, Z. Yang, H. Tian, and J. Guo, "New domain adaptation method in shallow and deep layers of the CNN for bearing fault diagnosis under different working conditions," *Int. J. Adv. Manuf. Technol.*, pp. 1–12, Jun. 2021, doi: 10.1007/s00170-021-07385-9.

[163] J. Zhou, L.-Y. Zheng, Y. Wang, and C. Gogu, "A multistage deep transfer learning method for machinery fault diagnostics across diverse working conditions and devices," *IEEE Access*, vol. 8, pp. 80879–80898, 2020, doi: 10.1109/ACCESS.2020.2990739.

[164] J. Wu, Z. Zhao, C. Sun, R. Yan, and X. Chen, "Few-shot transfer learning for intelligent fault diagnosis of machine," *Measurement*, vol. 166, Dec. 2020, Art. no. 108202, doi: 10.1016/j.measurement.2020.108202.

[165] W. Chen, Y. Qiu, Y. Feng, Y. Li, and A. Kusiak, "Diagnosis of wind turbine faults with transfer learning algorithms," *Renew. Energy*, vol. 163, pp. 2053–2067, Jan. 2021, doi: 10.1016/j.renene.2020.10.121.

[166] L. Zhang, L. Guo, H. Gao, D. Dong, G. Fu, and X. Hong, "Instance-based ensemble deep transfer learning network: A new intelligent degradation recognition method and its application on ball screw," *Mech. Syst. Signal Process.*, vol. 140, Jun. 2020, Art. no. 106681, doi: 10.1016/j.ymssp.2020.106681.

[167] K. Yu, Q. Fu, H. Ma, T. R. Lin, and X. Li, "Simulation data driven weakly supervised adversarial domain adaptation approach for intelligent cross-machine fault diagnosis," *Struct. Health Monitor.*, vol. 20, no. 4, pp. 2182–2198, Dec. 2020, doi: 10.1177/1475921720980718.

[168] M. Marei, S. E. Zaatari, and W. Li, "Transfer learning enabled convolutional neural networks for estimating health state of cutting tools," *Robot. Comput.-Integr. Manuf.*, vol. 71, Oct. 2021, Art. no. 102145, doi: 10.1016/j.rcim.2021.102145.

[169] C. Liu, L. Zhang, J. Li, J. Zheng, and C. Wu, "Two-stage transfer learning for fault prognosis of ion mill etching process," *IEEE Trans. Semicond. Manuf.*, vol. 34, no. 2, pp. 185–193, May 2021, doi: 10.1109/TSM.2021.3059025.

[170] Y. Qin, S. Adams, and C. Yuen, "Transfer learning-based state of charge estimation for lithium-ion battery at varying ambient temperatures," *IEEE Trans. Ind. Informat.*, vol. 17, no. 11, pp. 7304–7315, Nov. 2021, doi: 10.1109/TII.2021.3051048.

[171] H. Wang, X. Bai, J. Tan, and J. Yang, "Deep prototypical networks based domain adaptation for fault diagnosis," *J. Intell. Manuf.*, vol. 33, pp. 1–11, Nov. 2020, doi: 10.1007/s10845-020-01709-4.

[172] A. Zhang and X. Gao, "Supervised dictionary-based transfer subspace learning and applications for fault diagnosis of sucker rod pumping systems," *Neurocomputing*, vol. 338, pp. 293–306, Apr. 2019, doi: 10.1016/j.neucom.2019.02.013.

[173] K. Huang, H. Wen, C. Zhou, C. Yang, and W. Gui, "Transfer dictionary learning method for cross-domain multimode process monitoring and fault isolation," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 11, pp. 8713–8724, Nov. 2020, doi: 10.1109/TIM.2020.2998875.

[174] A. A. Chehade and A. A. Hussein, "A collaborative Gaussian process regression model for transfer learning of capacity trends between li-ion battery cells," *IEEE Trans. Veh. Technol.*, vol. 69, no. 9, pp. 9542–9552, Sep. 2020, doi: 10.1109/TVT.2020.3000970.

[175] J. Ma, X. Liu, X. Zou, M. Yue, P. Shang, L. Kang, S. Jemei, C. Lu, Y. Ding, N. Zerhouni, and Y. Cheng, "Degradation prognosis for proton exchange membrane fuel cell based on hybrid transfer learning and intercell differences," *ISA Trans.*, vol. 113, pp. 149–165, Jul. 2021, doi: 10.1016/j.isatra.2020.06.005.

[176] A. Gretton, K. Borgwardt, M. J. Rasch, B. Scholkopf, and A. J. Smola, "A kernel method for the two-sample problem," May 2008, *arXiv:0805.2368v1*.

[177] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 723–773, 2012.

[178] K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, B. Schölkopf, and A. J. Smola, "Integrating structured biological data by kernel maximum mean discrepancy," *Bioinformatics*, vol. 22, no. 14, pp. e49–e57, Jul. 2006, doi: 10.1093/bioinformatics/btl242.

[179] B. Yang, Y. Lei, F. Jia, and S. Xing, "A transfer learning method for intelligent fault diagnosis from laboratory machines to real-case machines," in *Proc. Int. Conf. Sens., Diagnostics, Prognostics, Control (SDPC)*, Aug. 2018, pp. 35–40, doi: 10.1109/SDPC.2018.8664814.

[180] X. Jia, M. Zhao, Y. Di, Q. Yang, and J. Lee, "Assessment of data suitability for machine prognosis using maximum mean discrepancy," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5872–5881, Jul. 2018, doi: 10.1109/TIE.2017.2777383.

[181] H. Cai, X. Jia, J. Feng, W. Li, L. Pahren, and J. Lee, "A similarity based methodology for machine prognostics by using kernel two sample test," *ISA Trans.*, vol. 103, pp. 112–121, Aug. 2020, doi: 10.1016/j.isatra.2020.03.007.

[182] H. Cai, J. Feng, W. Li, Y.-M. Hsu, and J. Lee, "Similarity-based particle filter for remaining useful life prediction with enhanced performance," *Appl. Soft Comput.*, vol. 94, Sep. 2020, Art. no. 106474, doi: 10.1016/j.asoc.2020.106474.

[183] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2010, doi: 10.1109/TNN.2010.2091281.

[184] G. Matasci, M. Volpi, D. Tuia, and M. Kanevski, "Transfer component analysis for domain adaptation in image classification," in *Proc. SPIE*, L. Bruzzone, Ed., 2011, Art. no. 81800F, doi: 10.1117/12.898229.

[185] C. Chen, Z. Li, J. Yang, and B. Liang, "A cross domain feature extraction method based on transfer component analysis for rolling bearing fault diagnosis," in *Proc. 29th Chin. Control Decis. Conf. (CCDC)*, May 2017, pp. 5622–5626, doi: 10.1109/CCDC.2017.7978168.

[186] J. Xie, L. Zhang, L. Duan, and J. Wang, "On cross-domain feature fusion in gearbox fault diagnosis under various operating conditions based on transfer component analysis," in *Proc. IEEE Int. Conf. Prognostics Health Manage. (ICPHM)*, Jun. 2016, pp. 1–6, doi: 10.1109/ICPHM.2016.7542845.

[187] D. Xiao, C. Qin, H. Yu, Y. Huang, C. Liu, and J. Zhang, "Unsupervised machine fault diagnosis for noisy domain adaptation using marginal denoising autoencoder based on acoustic signals," *Measurement*, vol. 176, May 2021, Art. no. 109186, doi: 10.1016/j.measurement.2021.109186.

[188] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2200–2207, doi: 10.1109/ICCV.2013.274.

[189] Z. Tong, W. Li, B. Zhang, and M. Zhang, "Bearing fault diagnosis based on domain adaptation using transferable features under different working conditions," *Shock Vibrat.*, vol. 2018, pp. 1–12, Jun. 2018, doi: 10.1155/2018/6714520.

[190] Z. Tong, W. Li, B. Zhang, F. Jiang, and G. B. Zhou, "Bearing fault diagnosis under variable working conditions based on domain adaptation using feature transfer learning," *IEEE Access*, vol. 6, pp. 76187–76197, 2018, doi: 10.1109/ACCESS.2018.2883078.

[191] J. Gu and Y. Wang, "A cross domain feature extraction method for bearing fault diagnosis based on balanced distribution adaptation," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Qingdao)*, Oct. 2019, pp. 1–5, doi: 10.1109/PHM-Qingdao46334.2019.8942996.

[192] N. Cao, Z. Jiang, J. Gao, and B. Cui, "Bearing state recognition method based on transfer learning under different working conditions," *Sensors*, vol. 20, no. 1, p. 234, Dec. 2019, doi: 10.3390/s20010234.

[193] W. Qian, S. Li, P. Yi, and K. Zhang, "A novel transfer learning method for robust fault diagnosis of rotating machines under variable working conditions," *Measurement*, vol. 138, pp. 514–525, May 2019, doi: 10.1016/j.measurement.2019.02.073.

[194] T. Han, C. Liu, W. Yang, and D. Jiang, "Deep transfer network with joint distribution adaptation: A new intelligent fault diagnosis framework for industry application," *ISA Trans.*, vol. 97, pp. 269–281, Feb. 2020, doi: 10.1016/j.isatra.2019.08.012.

[195] P. Gardner, L. A. Bull, J. Gosliga, N. Dervilis, and K. Worden, "Foundations of population-based SHM, part III: Heterogeneous populations—Mapping and transfer," *Mech. Syst. Signal Process.*, vol. 149, Feb. 2021, Art. no. 107142, doi: 10.1016/j.ymssp.2020.107142.

[196] H. Shi, Y. Shang, X. Zhang, and Y. Tang, "Research on the initial fault prediction method of rolling bearings based on DCAE-TCN transfer learning," *Shock Vibrat.*, vol. 2021, pp. 1–15, Mar. 2021, doi: 10.1155/2021/5587756.

[197] Y. Pan, R. Hong, J. Chen, J. Feng, and W. Wu, "Performance degradation assessment of wind turbine gearbox based on maximum mean discrepancy and multi-sensor transfer learning," *Structural Health Monitor.*, vol. 20, no. 1, pp. 118–138, Jan. 2021, doi: 10.1177/1475921720919073.

[198] P. Ma, H. Zhang, W. Fan, and C. Wang, "A diagnosis framework based on domain adaptation for bearing fault diagnosis across diverse domains," *ISA Trans.*, vol. 99, pp. 465–478, Apr. 2020, doi: 10.1016/j.isatra.2019.08.040.

[199] R. Van De Sand, S. Corasaniti, and J. Reiff-Stephan, "Data-driven fault diagnosis for heterogeneous chillers using domain adaptation techniques," *Control Eng. Pract.*, vol. 112, Jul. 2021, Art. no. 104815, doi: 10.1016/j.conengprac.2021.104815.

[200] M. Long, Y. Cao, J. Wang, and M. I. Jordan, "Learning transferable features with deep adaptation networks," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, F. Bach and D. Blei, Eds. 2015, pp. 97–105.

[201] J. Zhu, N. Chen, and C. Shen, "A new deep transfer learning method for bearing fault diagnosis under different working conditions," *IEEE Sensors J.*, vol. 20, no. 15, pp. 8394–8402, Aug. 2019, doi: 10.1109/JSEN.2019.2936932.

[202] Y. Lin, Z. Nie, and H. Ma, "Dynamics-based cross-domain structural damage detection through deep transfer learning," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 37, no. 1, pp. 24–54, Jan. 2022, doi: 10.1111/mice.12692.

[203] Z. Wu, S. Yu, X. Zhu, Y. Ji, and M. Pecht, "A weighted deep domain adaptation method for industrial fault prognostics according to prior distribution of complex working conditions," *IEEE Access*, vol. 7, pp. 139802–139814, 2019, doi: 10.1109/ACCESS.2019.2943076.

[204] S. Yu, Z. Wu, X. Zhu, and M. Pecht, "A domain adaptive convolutional LSTM model for prognostic remaining useful life estimation under variant conditions," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Paris)*, May 2019, pp. 130–137, doi: 10.1109/PHM-Paris.2019.00030.

[205] M. Sun, H. Wang, P. Liu, S. Huang, and P. Fan, "A sparse stacked denoising autoencoder with optimized transfer learning applied to the fault diagnosis of rolling bearings," *Measurement*, vol. 146, pp. 305–314, Nov. 2019, doi: 10.1016/j.measurement.2019.06.029.

[206] X. Li, W. Zhang, Q. Ding, and J.-Q. Sun, "Multi-layer domain adaptation method for rolling bearing fault diagnosis," *Signal Process.*, vol. 157, pp. 180–197, Apr. 2019, doi: 10.1016/j.sigpro.2018.12.005.

[207] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 1, pp. 136–144, Oct. 2017, doi: 10.1109/TSMC.2017.2754287.

[208] Z. An, S. Li, J. Wang, Y. Xin, and K. Xu, "Generalization of deep neural network for bearing fault diagnosis under different working conditions using multiple kernel method," *Neurocomputing*, vol. 352, pp. 42–53, Aug. 2019, doi: 10.1016/j.neucom.2019.04.010.

[209] C. Che, H. Wang, Q. Fu, and X. Ni, "Deep transfer learning for rolling bearing fault diagnosis under variable operating conditions," *Adv. Mech. Eng.*, vol. 11, no. 12, Dec. 2019, Art. no. 168781401989721, doi: 10.1177/1687814019897212.

[210] Y. Tian, Y. Tang, and X. Peng, "Cross-task fault diagnosis based on deep domain adaptation with local feature learning," *IEEE Access*, vol. 8, pp. 127546–127559, 2020, doi: 10.1109/ACCESS.2020.3006250.

[211] M. Azamfar, J. Singh, X. Li, and J. Lee, "Cross-domain gearbox diagnostics under variable working conditions with deep convolutional transfer learning," *J. Vibrat. Control*, vol. 27, nos. 7–8, pp. 854–864, Apr. 2021, doi: 10.1177/1077546320933793.

[212] C. Chen, F. Shen, J. Xu, and R. Yan, "Domain adaptation-based transfer learning for gear fault diagnosis under varying working conditions," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021, doi: 10.1109/TIM.2020.3011584.

[213] Z. Chen, R. Huang, Y. Liao, J. Li, G. Jin, and W. Li, "Simultaneous fault type and severity identification using a two-branch domain adaptation network," *Meas. Sci. Technol.*, vol. 32, no. 9, Sep. 2021, Art. no. 094014, doi: 10.1088/1361-6501/abead1.

[214] Z. Zhang, H. Chen, S. Li, and Z. An, "Sparse filtering based domain adaptation for mechanical fault diagnosis," *Neurocomputing*, vol. 393, pp. 101–111, Jun. 2020, doi: 10.1016/j.neucom.2020.02.049.

[215] W. Zhang and X. Li, "Data privacy preserving federated transfer learning in machinery fault diagnostics using prior distributions," *Structural Health Monitor.*, vol. 21, no. 4, Jul. 2022, Art. no. 147592172110292, doi: 10.1177/14759217211029201.

[216] D. Xiao, Y. Huang, L. Zhao, C. Qin, H. Shi, and C. Liu, "Domain adaptive motor fault diagnosis using deep transfer learning," *IEEE Access*, vol. 7, pp. 80937–80949, 2019, doi: 10.1109/ACCESS.2019.2921480.

[217] B. Yang, Y. Lei, F. Jia, and S. Xing, "An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings," *Mech. Syst. Signal Process.*, vol. 122, pp. 692–706, May 2019, doi: 10.1016/j.ymssp.2018.12.051.

[218] F.-Y. Guo, Y.-C. Zhang, Y. Wang, P.-J. Ren, and P. Wang, "Fault diagnosis of reciprocating compressor valve based on transfer learning convolutional neural network," *Math. Problems Eng.*, vol. 2021, pp. 1–13, Jan. 2021, doi: 10.1155/2021/8891424.

[219] Z. Tang, L. Bo, X. Liu, and D. Wei, "An autoencoder with adaptive transfer learning for intelligent fault diagnosis of rotating machinery," *Meas. Sci. Technol.*, vol. 32, no. 5, May 2021, Art. no. 055110, doi: 10.1088/1361-6501/abd650.

[220] Y. Ding, M. Jia, Q. Miao, and P. Huang, "Remaining useful life estimation using deep metric transfer learning for kernel regression," *Rel. Eng. Syst. Saf.*, vol. 212, Aug. 2021, Art. no. 107583, doi: 10.1016/j.ress.2021.107583.

[221] Z.-H. Liu, L.-B. Jiang, H.-L. Wei, L. Chen, and X.-H. Li, "Optimal transport-based deep domain adaptation approach for fault diagnosis of rotating machine," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021, doi: 10.1109/TIM.2021.3050173.

[222] C. Zhao, G. Liu, and W. Shen, "A dual-view alignment-based domain adaptation network for fault diagnosis," *Meas. Sci. Technol.*, vol. 32, no. 11, Nov. 2021, Art. no. 115102, doi: 10.1088/1361-6501/ac100e.

[223] W. Qian, S. Li, and J. Wang, "A new transfer learning method and its application on rotating machine fault diagnosis under variant working conditions," *IEEE Access*, vol. 6, pp. 69907–69917, 2018, doi: 10.1109/ACCESS.2018.2880770.

[224] Q. Qian, Y. Qin, Y. Wang, and F. Liu, "A new deep transfer learning network based on convolutional auto-encoder for mechanical fault diagnosis," *Measurement*, vol. 178, Jun. 2021, Art. no. 109352, doi: 10.1016/j.measurement.2021.109352.

[225] X. Li, Y. Hu, J. Zheng, M. Li, and W. Ma, "Central moment discrepancy based domain adaptation for intelligent bearing fault diagnosis," *Neurocomputing*, vol. 429, pp. 12–24, Mar. 2021, doi: 10.1016/j.neucom.2020.11.063.

[226] P. Xiong, B. Tang, L. Deng, M. Zhao, and X. Yu, "Multi-block domain adaptation with central moment discrepancy for fault diagnosis," *Measurement*, vol. 169, Feb. 2021, Art. no. 108516, doi: 10.1016/j.measurement.2020.108516.

[227] H. Wang, W. Lu, S. Tang, and Y. Song, "Predict industrial equipment failure with time Windows and transfer learning," *Int. J. Speech Technol.*, vol. 52, no. 3, pp. 2346–2358, Feb. 2022, doi: 10.1007/s10489-021-02441-z.

[228] Z. Zhang, H. Chen, S. Li, and Z. An, "Unsupervised domain adaptation via enhanced transfer joint matching for bearing fault diagnosis," *Measurement*, vol. 165, Dec. 2020, Art. no. 108071, doi: 10.1016/j.measurement.2020.108071.

[229] Z. Zhang, M. Shao, L. Wang, S. Shao, and C. Ma, "A novel domain adaptation-based intelligent fault diagnosis model to handle sample class imbalanced problem," *Sensors*, vol. 21, no. 10, p. 3382, May 2021, doi: 10.3390/s21103382.

[230] J. An and P. Ai, "Deep domain adaptation model for bearing fault diagnosis with Riemann metric correlation alignment," *Math. Problems Eng.*, vol. 2020, pp. 1–12, Mar. 2020, doi: 10.1155/2020/4302184.

[231] Y. Cao, M. Jia, P. Ding, and Y. Ding, "Transfer learning for remaining useful life prediction of multi-conditions bearings based on bidirectional-GRU network," *Measurement*, vol. 178, Jun. 2021, Art. no. 109287, doi: 10.1016/j.measurement.2021.109287.

[232] J. Ma, P. Shang, X. Zou, N. Ma, Y. Ding, J. Sun, Y. Cheng, L. Tao, C. Lu, Y. Su, J. Chong, H. Jin, and Y. Lin, "A hybrid transfer learning scheme for remaining useful life prediction and cycle life test optimization of different formulation Li-ion power batteries," *Appl. Energy*, vol. 282, Jan. 2021, Art. no. 116167, doi: 10.1016/j.apenergy.2020.116167.

[233] W. Liu, H. Ren, M. A. Shaheer, and J. A. Awan, "A novel wind turbine health condition monitoring method based on correlative features domain adaptation," *Int. J. Precis. Eng. Manuf.-Green Technol.*, vol. 9, no. 1, pp. 191–200, Jan. 2022, doi: 10.1007/s40684-020-00293-5.

[234] H. Wang and L. Wang, "An improved transfer learning method for bearing diagnosis under variable working conditions based on dilated convolution," in *Proc. IEEE 16th Int. Conf. Control Autom. (ICCA)*, Oct. 2020, pp. 1612–1617, doi: 10.1109/ICCA51439.2020.9264479.

[235] K. Wang, W. Zhao, A. Xu, P. Zeng, and S. Yang, "One-dimensional multi-scale domain adaptive network for bearing-fault diagnosis under varying working conditions," *Sensors*, vol. 20, no. 21, p. 6039, Oct. 2020, doi: 10.3390/s20216039.

[236] S. Siahpour, X. Li, and J. Lee, "Deep learning-based cross-sensor domain adaptation for fault diagnosis of electro-mechanical actuators," *Int. J. Dyn. Control*, vol. 8, no. 4, pp. 1054–1062, Dec. 2020, doi: 10.1007/s40435-020-00669-0.

[237] V. Pandhare, X. Li, M. Miller, X. Jia, and J. Lee, "Intelligent diagnostics for ball screw fault through indirect sensing using deep domain adaptation," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021, doi: 10.1109/TIM.2020.3043512.

[238] S. Dong, G. Wen, and Z. Zhang, "Bearing fault diagnosis under different operating conditions based on cross domain feature projection and domain adaptation," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2019, pp. 1–6, doi: 10.1109/I2MTC.2019.8826993.

[239] S. Dong, K. He, and B. Tang, "The fault diagnosis method of rolling bearing under variable working conditions based on deep transfer learning," *J. Brazilian Soc. Mech. Sci. Eng.*, vol. 42, no. 11, Nov. 2020, doi: 10.1007/s40430-020-02661-3.

[240] S. Dong, G. Wen, Z. Lei, and Z. Zhang, "Transfer learning for bearing performance degradation assessment based on deep hierarchical features," *ISA Trans.*, vol. 108, pp. 343–355, Feb. 2021, doi: 10.1016/j.isatra.2020.09.004.

[241] Y. Yu, C. Zhang, Y. Li, and Y. Li, "A new transfer learning fault diagnosis method using TSC and JGSA under variable condition," *IEEE Access*, vol. 8, pp. 177287–177295, 2020, doi: 10.1109/ACCESS.2020.3025956.

[242] S. Pang and X. Yang, "A cross-domain stacked denoising autoencoders for rotating machinery fault diagnosis under different working conditions," *IEEE Access*, vol. 7, pp. 77277–77292, 2019, doi: 10.1109/ACCESS.2019.2919535.

[243] P. Gardner, X. Liu, and K. Worden, "On the application of domain adaptation in structural health monitoring," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106550, doi: 10.1016/J.YMSSP.2019.106550.

[244] M. Long, J. Wang, G. Ding, S. J. Pan, and P. S. Yu, "Adaptation regularization: A general framework for transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 5, pp. 1076–1089, May 2014, doi: 10.1109/TKDE.2013.111.

[245] W. Zhang, X. Li, H. Ma, Z. Luo, and X. Li, "Transfer learning using deep representation regularization in remaining useful life prediction across operating conditions," *Rel. Eng. Syst. Saf.*, vol. 211, Jul. 2021, Art. no. 107556, doi: 10.1016/j.ress.2021.107556.

[246] M. B. Bejiga, F. Melgani, and P. Beraldini, "Domain adversarial neural networks for large-scale land cover classification," *Remote Sens.*, vol. 11, no. 10, p. 1153, May 2019, doi: 10.3390/rs11101153.

[247] A. Elshamli, G. W. Taylor, A. Berg, and S. Areibi, "Domain adaptation using representation learning for the classification of remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 9, pp. 4198–4209, Sep. 2017, doi: 10.1109/JSTARS.2017.2711360.

[248] H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, and M. Marchand, "Domain-adversarial neural networks," 2015, *arXiv:1412.4446*.

[249] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, F. Bach and D. Blei, Eds., 2015, pp. 1180–1189.

[250] Z.-H. Liu, B.-L. Lu, H.-L. Wei, L. Chen, X.-H. Li, and M. Ratsch, "Deep adversarial domain adaptation model for bearing fault diagnosis," *IEEE Trans. Syst., Man, Cybern. Syst.*, pp. 1–10, 2020, doi: 10.1109/TSMC.2019.2932000.

[251] X. Li, W. Zhang, H. Ma, Z. Luo, and X. Li, "Domain generalization in rotating machinery fault diagnostics using deep neural networks," *Neurocomputing*, vol. 403, pp. 409–420, Aug. 2020, doi: 10.1016/j.neucom.2020.05.014.

[252] Y. Wang, X. Sun, J. Li, and Y. Yang, "Intelligent fault diagnosis with deep adversarial domain adaptation," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021, doi: 10.1109/TIM.2020.3035385.

[253] X. Zhu, K. Chen, B. Anduv, X. Jin, and Z. Du, "Transfer learning based methodology for migration and application of fault detection and diagnosis between building chillers for improving energy efficiency," *Building Environ.*, vol. 200, Aug. 2021, Art. no. 107957, doi: 10.1016/j.buildenv.2021.107957.

[254] C. Liu and K. Gryllias, "Unsupervised domain adaptation based remaining useful life prediction of rolling element bearings," in *Proc. 5th Eur. Conf. Prognostics Health Manage. Soc. (PHME)*, vol. 5, A. Bregon and K. Medjaher, Eds., 2020, pp. 1–10, doi: 10.36001/phme.2020.v5i1.1208.

[255] F. Zeng, Y. Li, Y. Jiang, and G. Song, "An online transfer learning-based remaining useful life prediction method of ball bearings," *Measurement*, vol. 176, May 2021, Art. no. 109201, doi: 10.1016/j.measurement.2021.109201.

[256] M. Zhang, D. Wang, W. Lu, J. Yang, Z. Li, and B. Liang, "A deep transfer model with Wasserstein distance guided multi-adversarial networks for bearing fault diagnosis under different working conditions," *IEEE Access*, vol. 7, pp. 65303–65318, 2019, doi: 10.1109/ACCESS.2019.2916935.

[257] J. Guo, J. Wu, S. Zhang, J. Long, W. Chen, D. Cabrera, and C. Li, "Generative transfer learning for intelligent fault diagnosis of the wind turbine gearbox," *Sensors*, vol. 20, no. 5, p. 1361, 2020, doi: 10.3390/s20051361.

[258] X. Yu, Z. Zhao, X. Zhang, C. Sun, B. Gong, R. Yan, and X. Chen, "Conditional adversarial domain adaptation with discrimination embedding for locomotive fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2020, doi: 10.1109/TIM.2020.3031198.

[259] Y. Ying, Z. Jun, T. Tang, W. Jingwei, C. Ming, W. Jie, and W. Liang, "Wasserstein distance based asymmetric adversarial domain adaptation in intelligent bearing fault diagnosis," *Meas. Sci. Technol.*, vol. 32, no. 11, Jun. 2021, Art. no. 115019, doi: 10.1088/1361-6501/ac0a0c.

[260] Y. Deng, D. Huang, S. Du, G. Li, C. Zhao, and J. Lv, "A double-layer attention based adversarial network for partial transfer learning in machinery fault diagnosis," *Comput. Ind.*, vol. 127, May 2021, Art. no. 103399, doi: 10.1016/j.compind.2021.103399.
[261] Z.-H. Liu, B.-L. Lu, H.-L. Wei, L. Chen, X.-H. Li, and C.-T. Wang, "A stacked auto-encoder based partial adversarial domain adaptation model for intelligent fault diagnosis of rotating machines," *IEEE Trans. Ind. Informat.*, vol. 17, no. 10, pp. 6798–6809, Oct. 2021, doi: 10.1109/TII.2020.3045002.
[262] J. Jiao, J. Lin, M. Zhao, and K. Liang, "Double-level adversarial domain adaptation network for intelligent fault diagnosis," *Knowl.-Based Syst.*, vol. 205, Oct. 2020, Art. no. 106236, doi: 10.1016/j.knosys.2020.106236.
[263] Y. Z. Liu, K. M. Shi, Z. X. Li, G. F. Ding, and Y. S. Zou, "Transfer learning method for bearing fault diagnosis based on fully convolutional conditional Wasserstein adversarial networks," *Measurement*, vol. 180, Aug. 2021, Art. no. 109553, doi: 10.1016/j.measurement.2021.109553.
[264] Y. Zhang, Z. Ren, S. Zhou, and T. Yu, "Adversarial domain adaptation with classifier alignment for cross-domain intelligent fault diagnosis of multiple source domains," *Meas. Sci. Technol.*, vol. 32, no. 3, Mar. 2021, Art. no. 035102, doi: 10.1088/1361-6501/abcad4.
[265] B. Zhao, X. Zhang, Z. Zhan, and Q. Wu, "Deep multi-scale adversarial network with attention: A novel domain adaptation method for intelligent fault diagnosis," *J. Manuf. Syst.*, vol. 59, pp. 565–576, Apr. 2021, doi: 10.1016/j.jmsy.2021.03.024.
[266] K. Yu, H. Han, Q. Fu, H. Ma, and J. Zeng, "Symmetric co-training based unsupervised domain adaptation approach for intelligent fault diagnosis of rolling bearing," *Meas. Sci. Technol.*, vol. 31, no. 11, Nov. 2020, Art. no. 115008, doi: 10.1088/1361-6501/ab9841.
[267] Z. Zhang, H. Chen, S. Li, Z. An, and J. Wang, "A novel geodesic flow kernel based domain adaptation approach for intelligent fault diagnosis under varying working condition," *Neurocomputing*, vol. 376, pp. 54–64, Feb. 2020, doi: 10.1016/j.neucom.2019.09.081.
[268] W. Qian, S. Li, J. Wang, Y. Xin, and H. Ma, "A new deep transfer learning network for fault diagnosis of rotating machine under variable working conditions," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Chongqing)*, Oct. 2018, pp. 1010–1016, doi: 10.1109/PHM-Chongqing.2018.00180.
[269] P. Chen, Y. Li, K. Wang, and M. J. Zuo, "A novel knowledge transfer network with fluctuating operational condition adaptation for bearing fault pattern recognition," *Measurement*, vol. 158, Jul. 2020, Art. no. 107739, doi: 10.1016/j.measurement.2020.107739.
[270] S. S. Udmale, S. K. Singh, R. Singh, and A. K. Sangaiah, "Multi-fault bearing classification using sensors and ConvNet-based transfer learning approach," *IEEE Sensors J.*, vol. 20, no. 3, pp. 1433–1444, Feb. 2020, doi: 10.1109/JSEN.2019.2947026.
[271] T. Wen, R. Chen, and L. Tang, "Fault diagnosis of rolling bearings of different working conditions based on multi-feature spatial domain adaptation," *IEEE Access*, vol. 9, pp. 52404–52413, 2021, doi: 10.1109/ACCESS.2021.3069884.
[272] J. Jiao, M. Zhao, J. Lin, and K. Liang, "Residual joint adaptation adversarial network for intelligent transfer fault diagnosis," *Mech. Syst. Signal Process.*, vol. 145, Nov. 2020, Art. no. 106962, doi: 10.1016/j.ymssp.2020.106962.
[273] F. Li, T. Tang, B. Tang, and Q. He, "Deep convolution domain-adversarial transfer learning for fault diagnosis of rolling bearings," *Measurement*, vol. 169, Feb. 2021, Art. no. 108339, doi: 10.1016/j.measurement.2020.108339.
[274] W. Mao, J. Chen, Y. Chen, S. S. Afshari, and X. Liang, "Construction of health indicators for rotating machinery using deep transfer learning with multiscale feature representation," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021, doi: 10.1109/TIM.2021.3057498.
[275] W. Mao, B. Sun, and L. Wang, "A new deep dual temporal domain adaptation method for online detection of bearings early fault," *Entropy*, vol. 23, no. 2, p. 162, Jan. 2021, doi: 10.3390/e23020162.
[276] Y. Zhang, Z. Ren, and S. Zhou, "A new deep convolutional domain adaptation network for bearing fault diagnosis under different working conditions," *Shock Vibrat.*, vol. 2020, pp. 1–14, Jul. 2020, doi: 10.1155/2020/8850976.
[277] S. Jia, J. Wang, B. Han, G. Zhang, X. Wang, and J. He, "A novel transfer learning method for fault diagnosis using maximum classifier discrepancy with marginal probability distribution adaptation," *IEEE Access*, vol. 8, pp. 71475–71485, 2020, doi: 10.1109/ACCESS.2020.2987933.
[278] J. Shao, Z. Huang, and J. Zhu, "Transfer learning method based on adversarial domain adaption for bearing fault diagnosis," *IEEE Access*, vol. 8, pp. 119421–119430, 2020, doi: 10.1109/ACCESS.2020.3005243.
[279] X. Wang, C. Shen, M. Xia, D. Wang, J. Zhu, and Z. Zhu, "Multi-scale deep intra-class transfer learning for bearing fault diagnosis," *Rel. Eng. Syst. Saf.*, vol. 202, Oct. 2020, Art. no. 107050, doi: 10.1016/j.ress.2020.107050.
[280] J. Jiao, M. Zhao, J. Lin, and C. Ding, "Classifier inconsistency-based domain adaptation network for partial transfer intelligent diagnosis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 5965–5974, Sep. 2020, doi: 10.1109/TII.2019.2956294.
[281] X. Li, W. Zhang, H. Ma, Z. Luo, and X. Li, "Partial transfer learning in machinery cross-domain fault diagnostics using class-weighted adversarial networks," *Neural Netw.*, vol. 129, pp. 313–322, Sep. 2020, doi: 10.1016/j.neunet.2020.06.014.
[282] W. Zhang, X. Li, H. Ma, Z. Luo, and X. Li, "Open-set domain adaptation in machinery fault diagnostics using instance-level weighted adversarial learning," *IEEE Trans. Ind. Informat.*, vol. 17, no. 1, pp. 7445–7455, Nov. 2021, doi: 10.1109/TII.2021.3054651.
[283] Y. Feng, J. Chen, Z. Yang, X. Song, Y. Chang, S. He, E. Xu, and Z. Zhou, "Similarity-based meta-learning network with adversarial domain adaptation for cross-domain fault identification," *Knowl.-Based Syst.*, vol. 217, Apr. 2021, Art. no. 106829, doi: 10.1016/j.knosys.2021.106829.
[284] P. Wang and R. X. Gao, "Transfer learning for enhanced machine fault diagnosis in manufacturing," *CIRP Ann.*, vol. 69, no. 1, pp. 413–416, 2020, doi: 10.1016/j.cirp.2020.04.074.
[285] W. Mao, D. Zhang, S. Tian, and J. Tang, "Robust detection of bearing early fault based on deep transfer learning," *Electronics*, vol. 9, no. 2, p. 323, Feb. 2020, doi: 10.3390/electronics9020323.
[286] X. Li, W. Zhang, N.-X. Xu, and Q. Ding, "Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places," *IEEE Trans. Ind. Electron.*, vol. 67, no. 8, pp. 6785–6794, Aug. 2020, doi: 10.1109/TIE.2019.2935987.
[287] X. Wang, F. Liu, and D. Zhao, "Cross-machine fault diagnosis with semi-supervised discriminative adversarial domain adaptation," *Sensors*, vol. 20, no. 13, p. 3753, Jul. 2020, doi: 10.3390/s20133753.
[288] S. Xiang, J. Zhang, H. Gao, D. Shi, and L. Chen, "A deep transfer learning method for bearing fault diagnosis based on domain separation and adversarial learning," *Shock Vibrat.*, vol. 2021, pp. 1–9, Jun. 2021, doi: 10.1155/2021/5540084.
[289] B. Yang, S. Xu, Y. Lei, C.-G. Lee, E. Stewart, and C. Roberts, "Multi-source transfer learning network to complement knowledge for intelligent diagnosis of machines with unseen faults," *Mech. Syst. Signal Process.*, vol. 162, Jan. 2022, Art. no. 108095, doi: 10.1016/j.ymssp.2021.108095.
[290] J. Wang, J. Xie, L. Zhang, and L. Duan, "A factor analysis based transfer learning method for gearbox diagnosis under various operating conditions," in *Proc. Int. Symp. Flexible Autom. (ISFA)*, Aug. 2016, pp. 81–86, doi: 10.1109/ISFA.2016.7790140.
[291] J. Wang, R. Zhao, and R. X. Gao, "Probabilistic transfer factor analysis for machinery autonomous diagnosis cross various operating conditions," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 8, pp. 5335–5344, Aug. 2020, doi: 10.1109/TIM.2019.2963731.
[292] F. Shen, Y. Hui, R. Yan, C. Sun, and J. Xu, "A new penalty domain selection machine enabled transfer learning for gearbox fault recognition," *IEEE Trans. Ind. Electron.*, vol. 67, no. 10, pp. 8743–8754, Oct. 2020, doi: 10.1109/TIE.2020.2988229.
[293] Z. Li, R. Cai, H. W. Ng, M. Winslett, T. Z. J. Fu, B. Xu, X. Yang, and Z. Zhang, "Causal mechanism transfer network for time series domain adaptation in mechanical systems," *ACM Trans. Intell. Syst. Technol.*, vol. 12, no. 2, pp. 1–21, Apr. 2021, doi: 10.1145/3445033.
[294] G. Michau and O. Fink, "Domain adaptation for one-class classification: Monitoring the health of critical systems under limited information," *Int. J. Prognostics Health Manage.*, vol. 10, no. 11, pp. 1–13, 2019, doi: 10.3929/ETHZ-B-000390680.
[295] J. Li and D. He, "A Bayesian optimization AdaBN-DCNN method with self-optimized structure and hyperparameters for domain adaptation remaining useful life prediction," *IEEE Access*, vol. 8, pp. 41482–41501, 2020, doi: 10.1109/ACCESS.2020.2976595.
[296] Y. Fan, S. Nowaczyk, and T. Rögnvaldsson, "Transfer learning for remaining useful life prediction based on consensus self-organizing models," *Rel. Eng. Syst. Saf.*, vol. 203, Nov. 2020, Art. no. 107098, doi: 10.1016/j.ress.2020.107098.

[297] M. Ragab, Z. Chen, M. Wu, C. S. Foo, C. K. Kwoh, R. Yan, and X. Li, "Contrastive adversarial domain adaptation for machine remaining useful life prediction," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5239–5249, Aug. 2021, doi: 10.1109/TII.2020.3032690.
[298] A. G. Mahyari and T. Locker, "Domain adaptation for robot predictive maintenance systems," 2018, *arXiv:1809.08626*.
[299] Z. Guo, Y. Wan, and H. Ye, "An unsupervised fault-detection method for railway turnouts," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 11, pp. 8881–8901, Nov. 2020, doi: 10.1109/TIM.2020.2998863.
[300] W. Wu, M. Peng, W. Chen, and S. Yan, "Unsupervised deep transfer learning for fault diagnosis in fog radio access networks," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8956–8966, Sep. 2020, doi: 10.1109/JIOT.2020.2997187.
[301] Y. Li, N. Wang, J. Shi, J. Liu, and X. Hou, "Revisiting batch normalization for practical domain adaptation," in *Proc. 5th Int. Conf. Learn. Represent. (ICLR)*, 2017, pp. 1–12.
[302] S. Byttner, T. Rognvaldsson, M. Svensson, G. Bitar, and W. Chominsky, "Networked vehicles for automated fault detection," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2009, pp. 1213–1216, doi: 10.1109/ISCAS.2009.5117980.
[303] S. Byttner, T. Rögnvaldsson, and M. Svensson, "Consensus self-organized models for fault detection (COSMO)," *Eng. Appl. Artif. Intell.*, vol. 24, no. 5, pp. 833–839, Aug. 2011, doi: 10.1016/j.engappai.2011.03.002.
[304] Y. Fan, S. Nowaczyk, and T. Rögnvaldsson, "Evaluation of self-organized approach for predicting compressor faults in a city bus fleet," *Proc. Comput. Sci.*, vol. 53, pp. 447–456, 2015, doi: 10.1016/j.procs.2015.07.322.
[305] Y. Fan, S. Nowaczyk, T. Rögnvaldsson, and E. A. Antonelo, "Predicting air compressor failures with echo state networks," in *Proc. 13rd Eur. Conf. Prognostics Health Manage. Soc. (PHME)*, I. Eballard and A. Bregon, Eds., 2016, pp. 568–578.
[306] R. Gopalan, R. Li, and R. Chellappa, "Domain adaptation for object recognition: An unsupervised approach," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 999–1006, doi: 10.1109/ICCV.2011.6126344.
[307] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2066–2073, doi: 10.1109/CVPR.2012.6247911.
[308] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer joint matching for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1410–1417, doi: 10.1109/CVPR.2014.183.
[309] C. Sun, M. Ma, Z. Zhao, S. Tian, R. Yan, and X. Chen, "Deep transfer learning based on sparse autoencoder for remaining useful life prediction of tool in manufacturing," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2416–2425, Apr. 2019, doi: 10.1109/TII.2018.2881543.
[310] L. Basora, P. Bry, X. Olive, and F. Freeman, "Aircraft fleet health monitoring with anomaly detection techniques," *Aerospace*, vol. 8, no. 4, p. 103, Apr. 2021, doi: 10.3390/aerospace8040103.
[311] L. A. Bull, P. A. Gardner, J. Gosliga, T. J. Rogers, N. Dervilis, E. J. Cross, E. Papatheou, A. E. Maguire, C. Campos, and K. Worden, "Foundations of population-based SHM, part I: Homogeneous populations and forms," *Mech. Syst. Signal Process.*, vol. 148, Feb. 2021, Art. no. 107141, doi: 10.1016/j.ymssp.2020.107141.
[312] L. Cristaldi, G. Leone, R. Ottoboni, S. Subbiah, and S. Turrin, "A comparative study on data-driven prognostic approaches using fleet knowledge," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. Proc.*, May 2016, pp. 1–6, doi: 10.1109/I2MTC.2016.7520371.
[313] Z. Bouzidi, L. S. Terrissa, N. Zerhouni, and S. Ayad, "QoS of cloud prognostic system: Application to aircraft engines fleet," *Eur. J. Ind. Eng.*, vol. 14, no. 1, p. 34, 2020, doi: 10.1504/EJIE.2020.105080.
[314] I. Aslanidou, M. Rahman, V. Zaccaria, and K. G. Kyprianidis, "Micro gas turbines in the future smart energy system: Fleet monitoring, diagnostics, and system level requirements," *Frontiers Mech. Eng.*, vol. 7, Jun. 2021, doi: 10.3389/fmech.2021.676853.
[315] J. Liu and E. Zio, "A framework for asset prognostics from fleet data," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Chengdu)*, Oct. 2016, pp. 1–5, doi: 10.1109/PHM.2016.7819824.
[316] V. Zaccaria, A. D. Fentaye, M. Stenfelt, and K. G. Kyprianidis, "Probabilistic model for aero-engines fleet condition monitoring," *Aerospace*, vol. 7, no. 6, p. 66, May 2020, doi: 10.3390/aerospace7060066.
[317] K. Hendrickx, W. Meert, Y. Mollet, J. Gyselinck, B. Cornelis, K. Gryllias, and J. Davis, "A general anomaly detection framework for fleet-based condition monitoring of machines," *Mech. Syst. Signal Process.*, vol. 139, May 2020, Art. no. 106585, doi: 10.1016/J.YMSSP.2019.106585.
[318] K. Hendrickx, W. Meert, B. Cornelis, K. Gryllias, and J. Davis, "Similarity-based anomaly score for fleet-based condition monitoring," in *Proc. Annu. Conf. PHM Soc. (PHM)*, vol. 12, A. Saxena, Ed., 2020, p. 9, doi: 10.36001/phmconf.2020.v12i1.1178.
[319] J. Helsen, C. Peeters, T. Verstraeten, J. Verbeke, and A. Nowé, "Fleet-wide condition monitoring combining vibration signal processing and machine learning rolled out in a cloud-computing environment," in *Proc. Int. Conf. Noise Vibrat. Eng. (ISMA)*, 2018, pp. 17–19.
[320] S. Al-Dahidi, F. Di Maio, P. Baraldi, and E. Zio, "A locally adaptive ensemble approach for data-driven prognostics of heterogeneous fleets," *Proc. Inst. Mech. Eng. O, J. Risk Reliab.*, vol. 231, no. 4, pp. 350–363, 2017, doi: 10.1177/1748006X17693519.
[321] S. M. A. Al-Dahidi, F. Di Maio, P. Baraldi, and E. Zio, "A switching ensemble approach for remaining useful life estimation of electrolytic capacitors," in *Proc. 26th Eur. Saf. Rel. Conf. (ESREL)*, in A Balkema book, L. Walls, M. Revie, and T. Bedford, Eds. Boca Raton, FL, USA: CRC Press, 2017, pp. 325–332.
[322] M. Beretta, J. J. Cárdenas, C. Koch, and J. Cusidó, "Wind fleet generator fault detection via SCADA alarms and autoencoders," *Appl. Sci.*, vol. 10, no. 23, p. 8649, Dec. 2020, doi: 10.3390/app10238649.
[323] X. Jia, B. Huang, J. Feng, H. Cai, and J. Lee, "Review of PHM data competitions from 2008 to 2017: Methodologies and analytics," in *Proc. Annu. Conf. PHM Soc. (PHM)*, vol. 10, A. Bregon and M. Orchard, Eds., 2018, pp. 1–10, doi: 10.36001/phmconf.2018.v10i1.462.
[324] Y. Xu, Y. Sun, X. Liu, and Y. Zheng, "A digital-twin-assisted fault diagnosis using deep transfer learning," *IEEE Access*, vol. 7, pp. 19990–19999, 2019, doi: 10.1109/ACCESS.2018.2890566.
[325] X. Wang, X. Liu, and Y. Li, "An incremental model transfer method for complex process fault diagnosis," *IEEE/CAA J. Autom. Sinica*, vol. 6, no. 5, pp. 1268–1280, Sep. 2019, doi: 10.1109/JAS.2019.1911618.
[326] S. Huang, Y. Guo, D. Liu, S. Zha, and W. Fang, "A two-stage transfer learning-based deep learning approach for production progress prediction in IoT-enabled manufacturing," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10627–10638, 2019, doi: 10.1109/JIOT.2019.2940131.
[327] Y. Lin, Y. Watanabe, T. Kimura, T. Matsunawa, S. Nojima, M. Li, and D. Z. Pan, "Data efficient lithography modeling with residual neural networks and transfer learning," in *Proc. Int. Symp. Phys. Design*, C. Chu and I. Bustany, Eds., Mar. 2018, pp. 82–89, doi: 10.1145/3177540.3178242.
[328] Y. Lin, M. Li, Y. Watanabe, T. Kimura, T. Matsunawa, S. Nojima, and D. Z. Pan, "Data efficient lithography modeling with transfer learning and active data selection," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 38, no. 10, pp. 1900–1913, Oct. 2019, doi: 10.1109/TCAD.2018.2864251.
[329] Y. Gong, J. Luo, H. Shao, K. He, and W. Zeng, "Automatic defect detection for small metal cylindrical shell using transfer learning and logistic regression," *J. Nondestruct. Eval.*, vol. 39, no. 1, Mar. 2020, doi: 10.1007/s10921-020-0668-4.
[330] H. Yang, S. Jiao, and P. Sun, "Bayesian-convolutional neural network model transfer learning for image detection of concrete water-binder ratio," *IEEE Access*, vol. 8, pp. 35350–35367, 2020, doi: 10.1109/ACCESS.2020.2975350.
[331] Y. Kim, T. Kim, B. D. Youn, and S.-H. Ahn, "Machining quality monitoring (MQM) in laser-assisted micro-milling of glass using cutting force signals: An image-based deep transfer learning," *J. Intell. Manuf.*, vol. 33, no. 6, pp. 1813–1828, Aug. 2022, doi: 10.1007/s10845-021-01764-5.
[332] D. Jin, S. Xu, L. Tong, L. Wu, and S. Liu, "End image defect detection of float glass based on faster region-based convolutional neural network," *Traitement Signal*, vol. 37, no. 5, pp. 807–813, Nov. 2020, doi: 10.18280/ts.370513.
[333] P. Sassi, P. Tripicchio, and C. A. Avizzano, "A smart monitoring system for automatic welding defect detection," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9641–9650, Dec. 2019, doi: 10.1109/TIE.2019.2896165.
[334] X. Le, J. Mei, H. Zhang, B. Zhou, and J. Xi, "A learning-based approach for surface defect detection using small image datasets," *Neurocomputing*, vol. 408, pp. 112–120, Sep. 2020, doi: 10.1016/j.neucom.2019.09.107.
[335] Y. Yang, L. Pan, J. Ma, R. Yang, Y. Zhu, Y. Yang, and L. Zhang, "A high-performance deep learning algorithm for the automated optical inspection of laser welding," *Appl. Sci.*, vol. 10, no. 3, p. 933, Jan. 2020, doi: 10.3390/app10030933.

[336] Y. Gong, H. Shao, J. Luo, and Z. Li, "A deep transfer learning model for inclusion defect detection of aeronautics composite materials," *Composite Struct.*, vol. 252, Nov. 2020, Art. no. 112681, doi: 10.1016/j.compstruct.2020.112681.

[337] W. Li, S. Gu, X. Zhang, and T. Chen, "Transfer learning for process fault diagnosis: Knowledge transfer from simulation to physical processes," *Comput. Chem. Eng.*, vol. 139, Aug. 2020, Art. no. 106904, doi: 10.1016/j.compchemeng.2020.106904.

[338] S. Kang, "On effectiveness of transfer learning approach for neural network-based virtual metrology modeling," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 1, pp. 149–155, Feb. 2018, doi: 10.1109/TSM.2017.2787550.

[339] M. Azamfar, X. Li, and J. Lee, "Deep learning-based domain adaptation method for fault diagnosis in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 3, pp. 445–453, Aug. 2020, doi: 10.1109/TSM.2020.2995548.

[340] W. Zellinger, T. Grubinger, M. Zwick, E. Lughofer, H. Schöner, T. Natschläger, and S. Saminger-Platz, "Multi-source transfer learning of time series in cyclical manufacturing," *J. Intell. Manuf.*, vol. 31, no. 3, pp. 777–787, Mar. 2020, doi: 10.1007/s10845-019-01499-4.

[341] X. Wang, J. Zhao, and X. Guo, "A process monitoring method based on robust collective non-negative matrix factorization," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Jun. 2019, pp. 2650–2655, doi: 10.1109/CCDC.2019.8832577.

[342] X. Wang, J. Zhao, and D. Li, "A fault monitoring method based on transfer subspace learning," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Jun. 2019, pp. 2656–2661, doi: 10.1109/CCDC.2019.8832926.

[343] R. Luis, L. E. Sucar, and E. F. Morales, "Inductive transfer for learning Bayesian networks," *Mach. Learn.*, vol. 79, pp. 227–255, May 2009, doi: 10.1007/s10994-009-5160-4.

[344] C.-W. Seah, Y.-S. Ong, and I. W. Tsang, "Combating negative transfer from predictive distribution differences," *IEEE Trans. Cybern.*, vol. 43, no. 4, pp. 1153–1165, Aug. 2013, doi: 10.1109/TSMCB.2012.2225102.

[345] M. T. Rosenstein, Z. Marx, L. P. Kaelbling, and T. G. Dietterich, "To transfer or not to transfer," Dept. Comput. Sci. Artif. Intell. Lab., Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep., 2005.

[346] L. Ge, J. Gao, H. Ngo, K. Li, and A. Zhang, "On handling negative transfer and imbalanced distributions in multiple source transfer learning," in *Proc. SIAM Int. Conf. Data Mining (SDM)*, J. Gosh, Z. Obradovic, J. Dy, Z.-H. Zhou, C. Kamath, and S. Parthasarathy, Eds. 2013, pp. 261–269, doi: 10.1137/1.9781611972832.29.

[347] L. Gui, R. Xu, Q. Lu, J. Du, and Y. Zhou, "Negative transfer detection in transductive transfer learning," *Int. J. Mach. Learn. Cybern.*, vol. 9, no. 2, pp. 185–197, Feb. 2018, doi: 10.1007/s13042-016-0634-8.

[348] Z. Wang, Z. Dai, B. Poczos, and J. Carbonell, "Characterizing and avoiding negative transfer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, L. Davis, Ed. Jun. 2019, pp. 11285–11294.

[349] A. Gretton, K. Fukumizu, Z. Harchaoui, and B. K. Sriperumbudur, "A fast, consistent kernel two-sample test," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 673–681.

[350] J. Zhang, W. Li, P. Ogunbona, and D. Xu, "Recent advances in transfer learning for cross-dataset visual recognition: A problem-oriented perspective," *ACM Comput. Surveys*, vol. 52, no. 1, pp. 1–38, Feb. 2019, doi: 10.1145/3291124.

[351] F. Mauthe, S. Hagmeyer, and P. Zeiler, "Creation of publicly available data sets for prognostics and diagnostics addressing data scenarios relevant to industrial applications," *Int. J. Prognostics Health Manage.*, vol. 12, no. 2, Nov. 2021, doi: 10.36001/ijphm.2021.v12i2.3087.

[352] N. Burkart and M. F. Huber, "A survey on the explainability of supervised machine learning," *J. Artif. Intell. Res.*, vol. 70, pp. 245–317, Jan. 2021, doi: 10.1613/jair.1.12228.

[353] S. Vollert, M. Atzmueller, and A. Theissler, "Interpretable machine learning: A brief survey from the predictive maintenance perspective," in *Proc. 26th IEEE Int. Conf. Emerg. Technol. Factory Autom. (ETFA)*, Sep. 2021, pp. 1–8, doi: 10.1109/ETFA45728.2021.9613467.

[354] Q. Gu and Q. Dai, "A novel active multi-source transfer learning algorithm for time series forecasting," *Int. J. Speech Technol.*, vol. 51, no. 3, pp. 1326–1350, Mar. 2021, doi: 10.1007/s10489-020-01871-5.

[355] X. Ding, Q. Shi, B. Cai, T. Liu, Y. Zhao, and Q. Ye, "Learning multi-domain adversarial neural networks for text classification," *IEEE Access*, vol. 7, pp. 40323–40332, 2019, doi: 10.1109/ACCESS.2019.2904858.

[356] H. Nam and B. Han, "Learning multi-domain convolutional neural networks for visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4293–4302.

[357] K. Javed, R. Gouriveau, and N. Zerhouni, "State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels," *Mech. Syst. Signal Process.*, vol. 94, pp. 214–236, Sep. 2017, doi: 10.1016/j.ymssp.2017.01.050.

[358] J. S. Lal Senanayaka, H. Van Khang, and K. G. Robbersmyr, "Online fault diagnosis system for electric powertrains using advanced signal processing and machine learning," in *Proc. 13th Int. Conf. Electr. Mach. (ICEM)*, Sep. 2018, pp. 1932–1938, doi: 10.1109/ICELMACH.2018.8507171.

MARCEL BRAIG received the B.Sc. degree in technology development from the University of Applied Sciences Ravensburg–Weingarten (RWU), Germany, and the joint M.Eng. degree in mechatronics/systems engineering from the Aalen University of Applied Sciences and the Esslingen University of Applied Sciences, Germany. He is currently pursuing the Ph.D. degree in cooperation with the University of Stuttgart. He is also working as a Research Assistant with the Reliability Engineering and Prognostics and Health Management Research Group, Esslingen University of Applied Sciences. His research interests include diagnostics and prognostics, condition monitoring, data-driven methods, and machine learning with a focus on application to similar systems.

PETER ZEILER received the Dipl.-Ing and Dr.-Ing. degrees in mechanical engineering from the University of Stuttgart. He worked as the Team Leader of Research and Development at Robert Bosch GmbH. Thereafter, he became the Head of the Department of Reliability Engineering and the Department of Drive Technology, Institute of Machine Components, University of Stuttgart. In 2017, he joined the Esslingen University of Applied Sciences, Germany. Since 2021, he has been associated with the Faculty of Engineering Design, Production Engineering and Automotive Engineering, University of Stuttgart. He is currently a Professor with the Faculty of Mechanical and Systems Engineering, Esslingen University of Applied Sciences, where he is the Head of the Reliability Engineering and Prognostics and Health Management Research Group. His research interests include prognostics and health management, methods for calculating reliability under operating conditions, modeling and simulation of reliability, and availability for complex systems. He is committed to Verein Deutscher Ingenieure (VDI) (English: Association of German Engineers). Then, he is the Chairperson of the Safety and Reliability Advisory Board, and the Technical Committees for Prognostics and Health Management and Monte Carlo Simulation. Furthermore, he is the Conference Chair of the German Conference Fachtagung Technische Zuverlässigkeit (TTZ) (English: Technical Reliability Conference).