Research article

Semi-supervised learning for industrial fault detection and diagnosis: A systemic review

José Miguel Ramírez-Sanz, Jose-Alberto Maestro-Prieto, Álvar Arnaiz-González *, Andrés Bustillo

Universidad de Burgos, Avda. Cantabria s/n, Burgos, 09006, Burgos, Spain

ARTICLE INFO

ABSTRACT

The automation of Fault Detection and Diagnosis (FDD) is a central task for many industries today. A myriad of methods are in use, although the most recent leading contenders are data-driven approaches and especially Machine Learning (ML) methods. ML algorithms fall into two main categories: supervised and unsupervised methods, depending on whether or not the instances are labeled with the expected outputs. However, a new approach called Semi-Supervised Learning (SSL) has recently emerged that uses a few labeled instances together with other unlabeled instances for the training process. This new approach can significantly improve the accuracy of conventional ML models for industrial environments where labeled data are scarce. SSL has been tested as a promising solution over the past few years for several FDD problems, although there have been no systemic reviews of this sort of approach up until the present review. In this study, an attempt to organize the existing literature on SSL for FDD using the taxonomy of van Engelen & Hoos is reported. The most and the least frequently used SSL algorithms are identified and considered in terms of different fault detection tasks and their most common dataset structure. Moreover, a set of best practices are proposed in the conclusions of this work for implementation under real industrial conditions, so as to avoid some of the most common faults.

1. Introduction

Nowadays, data is routinely gathered as part of the increasingly digitized industrial processes of Industry 4.0 [1]. Industry can therefore benefit when data are used in several ways to assist with the automation or semi-automation of routine industrial processes and tasks, such as online process monitoring [2], fault detection, and diagnosis [3], and condition monitoring [4], among others.

ML algorithms are usually grouped into: supervised learning—when datasets[1] have feature(s) of interest, commonly referred to as class(es) – and unsupervised learning – when datasets have no class feature(s). Clustering is a popular unsupervised learning task, whereas classification and regression are common supervised learning tasks. On the one hand, the main drawback of supervised learning is that data with the right features are needed: there must be sufficient data that must be properly classified – labeled – in accordance with the model learning process and the purpose of the model, *e.g.*, normal and abnormal (faulty) conditions for industrial-process monitoring, or the different types of faults for Fault Detection and Diagnosis (FDD). On the other hand, unsupervised learning yields less accurate models, due to an absence of labeling. Somewhere in between, there is Semi-Supervised

Learning (SSL), which represents an attempt to use unlabeled information to improve the accuracy of supervised models when insufficient labeled observations are available.

In any industrial process monitored under real industrial conditions, huge volumes of data may be gathered during routine operations; however, only some of these data are properly labeled [5]. It should be noted that the labeling process is usually a time-consuming and costly exercise, which is usually carried out by human experts [6]. In addition, most of the data are examples of non-faulty behavior and there are very few examples of fault conditions. Sometimes data on certain failure modes are not captured and, therefore, there are no data available. Artificial data have been applied in most present-day research projects, to overcome this limitation, because they are generated under laboratory conditions, where faults are intentionally introduced to obtain balanced datasets, in terms of fault and non-fault examples. The industrial requirements of providing data solutions under real industrial conditions commonly implies the use of unlabeled data. Hence the use of SSL that involves unsupervised learning approaches in conjunction with labeled data [7].

---

* Corresponding author.
  E-mail address: alvarag@ubu.es (Á. Arnaiz-González).
[1] A dataset in this context is understood as a tabular matrix composed of rows, called instances/examples/prototypes, and columns, called features/attributes. In industrial processes, experiments are listed in the rows and experimental measurements are listed in the columns.

Within the context of the FDD industrial problem, a set of articles describing methods capable of detecting and diagnosing faults were selected. Among this sample of articles, there are several industrial applications, as will be discussed in Section 6.4. Most of the articles were focused on the utilization of bearing datasets with faults such as ball, inner race, and outer race faults. Additionally, there were articles that addressed faults related to chemicals, surface defects, and wind turbines, among others.

The recent literature on SSL applied to FDD is reviewed in this paper. To do so, an in-depth search of related articles from 2011 to October 2022 was conducted. Scopus and Google Scholar databases were used for finding the articles related with the topic. The Scopus database was queried for articles containing the words 'fault detection' and 'semi-supervised' or 'semisupervised' in the title, abstract, or keywords. Google Scholar was also queried using similar terms that yielded most of the studies on SSL in the context of FDD (in addition, all the cited articles for each paper were checked, so as not to leave any studies overlooked).

More than 300 articles were obtained in an initial search. Their abstracts were read in a preliminary filtering process. Congress papers, articles that were not about industrial FDD, and articles not in English were omitted, leaving 200 articles, all of which were read in full. In that step, a further 50 articles were discarded, either because they were not on the subject of industrial FDD or because SSL methods were not mentioned. Finally, 137 articles were included in the final study, but more than 137 methods were considered, because various methods were presented in some articles.

The main purpose of this article is to address the increasing adoption of SSL methods in several industrial applications, including FDD. Despite their widespread use, the organization and information regarding these SSL methods are often confused. Furthermore, several instances of misconduct in the application of SSL methods, including their application to FDD problems, have been identified. This misuse of SSL methods has sometimes led to misguided outcomes, which emphasizes the need to rectify their use.

Taking into account the aforementioned motivations, a set of objectives have been established. The first one is to clarify the structure of SSL methods applied to FDD, categorizing each method within a well-established taxonomy. Once every method is categorized, a study of the use of every type of SSL method becomes mandatory, in order to assess the relevance of each SSL method type. Furthermore, defining a set of best practices to assist the development of future applications of SSL methods in FDD, as well as other industrial fields, was also a primary objective. Finally, a paramount objective of this article is to identify future trends in the application of SSL methods to FDD, highlighting areas where further research can be conducted to address gaps in the current state-of-the-art.

The remainder of this review is organized as follows: firstly, the main theoretical concepts, namely SSL, FDD, active learning, transfer learning, and safe SSL are described in Section 2; secondly, the reference taxonomy is introduced in Section 3; the articles related to the topic are reviewed, and a proposal for their classification is offered in Sections 4 and 5; the main results and a discussion of the questions that arise are presented in Section 6; and, finally, the conclusions, a set of best practices, and some promising future trends are discussed in Section 7.

2. Background

In this section, the necessary background for following the paper is presented. In particular, fault detection and diagnosis (FDD) and the different learning approaches relevant to the rest of the study are briefly described: *i.e.*, semi-supervised learning (SSL), active learning, transfer learning, and safe SSL.

2.1. Fault detection and diagnosis (FDD)

Nowadays, FDD is one of the cornerstones to ensure the proper operation of complex industrial processes and equipment [8,9]. FDD is one of several processes that can be automated or semi-automated in industrial environments [10]. Among other examples, condition monitoring, predictive maintenance, and process monitoring may all be mentioned. Automated FDD is focused on detecting when an error or bad-function condition occurs within a process or system; once it occurs, the type of fault and even the specific part of the process that is failing may all be identified using FDD [11].

FDD is a difficult task because: (1) fault conditions data are rare compared with normal conditions data; (2) the normal variability of the industrial processes may be responsible for the observed deviations in the features; (3) the complex interrelations between process inputs can make it difficult to identify the source and type of some faults [12]. For all these reasons, the problem has been solved in different ways.

For 20 years, FDD approaches were split into three categories: the model-based approach [13], the knowledge-based approach [14], and the history-based approach [15]. A more recent review of FDD techniques for industrial maintenance [16] keeps this division, while changing the name of the third approach to data-driven [12] and focusing on the industrial requirements of Industry 4.0. Besides, recent reviews have been limited to specific processes where FDD plays a major role: chemical processes [9] and machinery failures [17], among others. In comparison with these previous works, the focus of this systemic review is placed on a promising new set of algorithms: the SSL techniques.

The FDD approaches are classified on the basis of data availability, because data structure, characteristics, and size all play major roles in FDD. The use of data collected for diagnosis from the process or system during its operation means that normal operating conditions may be interrupted and can vary when unexpected events occur, such as failure and even normal wear and tear [10]. A model can therefore be constructed that associates the behavior observed in some features with normal operating conditions and with one or more failure conditions.

As previously noted, ML algorithms have normally been divided into two main approaches, namely supervised and unsupervised learning, depending on whether or not the instances are labeled with information on the expected outcome. However, output availability is often mentioned as a key factor when choosing the most suitable and accurate ML algorithm. For example, approaches using neural networks often produce very accurate models, but they often require a large amount of labeled data with sufficient fault data examples to achieve that accuracy, a rare case in real industrial conditions, as Jiang et al. [18] previously outlined.

In general, it can be said that a key point in the whole ML process is to gather a proper set of representative examples that are labeled with accurate outputs, which can then be used as inputs for training ML algorithms. Today, data from many processes can easily be captured and stored at little or no cost. However, the process of labeling instances with the expected output is usually laborious, time-consuming, and expensive [19]. This industrial limitation is the key issue driving the introduction of SSL for use in FDD tasks.

2.2. Semi-supervised learning (SSL)

Zhou [20] defined supervised learning as a situation where there are sufficient accurately labeled instances available to train a model, while alternative situations are globally classified as weakly supervised learning. The author distinguished three main types of weakly supervised learning: incomplete supervision, inexact supervision, and inaccurate supervision. Incomplete supervision refers to a dataset with an insufficient number of accurately labeled instances, though unlabeled instances are available; in this case, active learning or SSL can be used, depending on whether an accurate label can be requested from a

human operator. Inexact supervision refers to a dataset with accurately labeled instances, though the labels are coarse-grained, and not as precise as needed. Finally, inaccurate supervision refers to a training dataset with some erroneous labels, which must be taken into account during the learning process, usually by updating those erroneous labels.

A possible solution to the problem of not having enough labeled instances for training, an intermediate approach between supervised and unsupervised algorithms, has been proposed: SSL. The goal of SSL is to try to obtain an accurate model using a limited amount of labeled instances together with unlabeled instances, in order to improve the model that only the labeled instances can obtain. SSL is based on four main assumptions [7,21]:

- Smoothness assumption: two instances that are close to each other in the input space have the same label.
- Low-density assumption: the decision boundary should pass through a low-density space.
- Manifold assumption: the high-dimensional input space is composed of multiple lower-density sub-spaces. Instances of the same lower-density sub-space must have the same label.
- Cluster assumption: instances that are located in the same cluster should have the same label.

However, the semi-supervised approach is not often considered in literature reviews on applications of ML-based solutions within industrial environments. For example, Mowbray et al. [22] presented a review of ML approaches in the chemical process industry that included: supervised, unsupervised, and reinforcement learning, although SSL was not considered.

The use of unlabeled data can be done in several ways and at various stages of the process. For example, a common step before the training phase is to perform feature selection or dimensionality reduction of the available data. A common unsupervised method for this purpose is Principal Component Analysis (PCA),[2] which allows direct use of the unlabeled data in that step.

A supervised alternative is Fisher Discriminant Analysis (FDA),[3] which can only use labeled instances.

However, several semi-supervised versions of FDA have been proposed, e.g., Semi-supervised FDA (SFDA) [5], Ensemble Semi-supervised FDA (ESFDA) [23], and others with which both labeled and unlabeled instances can be used at the same time. Similarly, a multitude of algorithms have been developed to include the unlabeled instances in the training step of the model using different approaches.

SSL methods can initially be divided into transductive and inductive methods [7]. On the one hand, transductive algorithms usually develop no model in the training phase and the goal is to label the set of unlabeled instances that are already available. Therefore, the transductive approach is not used with new and unseen instances, for example, in an on-line diagnostic system that performs detection and diagnosis as new data are collected. On the other hand, inductive algorithms usually develop a model during the training phase that can be used later on to label unseen data. The goal in this case is to use the unlabeled instances already available during the training stage to improve the model that could have been obtained, had only labeled instances been used for training. It has been addressed in many ways, as can be seen in [7,24]. A complete explanation of SSL taxonomy will be detailed in Section 3.

---

[2] PCA aims to transform the feature space by creating a new set of orthogonal variables, known as principal components, which capture the maximum variance in the data and improve the identification of data patterns.

[3] FDA aims to find a linear combination of features to maximize the separability between classes in a dataset.

2.3. Active learning

Active learning uses both labeled and unlabeled instances for the training, but instead of using an ML model to pseudo-label the unlabeled instances as SSL does, it selects the most relevant unlabeled items to be queried for labeling, e.g., by a human expert.

A typical criterion for the selection of unlabeled data is the least certain instances to be labeled, or the ones that provide the largest expected error reduction, among others [25,26].

Several active learning applications have been developed for FDD problems, such as in [27], where bearing faults are predicted using a residual network with active learning, enabling similar results to be achieved with 150 instances compared to 1 200.

Active learning is, nonetheless, often seen as a different, perhaps even contrary, approach to semi-supervised learning. Despite these different approaches towards learning, active learning can be applied together with SSL (e.g. [28–30]) in which an expert labels the most uncertain unlabeled instances.

2.4. Transfer learning

Transfer Learning [31], also known as Cross Domain, Domain Generalization, or Cross Machine, is a ML technique that is designed to improve models by adding information from another domain. Two different domains are then considered: the domain of the problem that is to be solved, called the target domain, and the domain of a similar problem where information can be obtained, called the source domain.

All instances of the source domain are usually labeled in transfer learning applied to SSL problems. So, if the source and target domains are similar, significant amounts of information may be obtained.

Several transfer learning applications have been developed for FDD due to the limited availability of labeled data in most datasets. Examples of transfer learning applied to FDD include [32], where different modes from the Tennessee Eastman Process dataset [33] were utilized as distinct views for transfer learning, and [34] where transfer learning was employed to enhance the information in chiller FDD problems.

2.5. Safe semi-supervised learning (SSL)

One of the problems of some SSL techniques is that they do not guarantee that the performance of the model using both labeled and unlabeled data is better, or at least not worse, than the model trained with labeled data alone [35]. This possible collateral effect has meant that SSL is not as popular as other ML techniques in industrial applications.

Bearing the above in mind, the idea behind safe SSL is that it is meant to ensure that a model trained with unlabeled data will not impinge on the classification capabilities of a model trained only with labeled examples. Safe SSL is described in far greater detail in [36].

The number of studies on Safe SSL for FDD is very scarce, though a detailed explanation on one such study, [37], can be found among this manuscript.

3. Taxonomy of semi-supervised learning

In this section, a taxonomy of SSL is presented for categorizing the literature on FDD, which is applied in the following sections. An up-to-date review of SSL algorithms can be found in [7], in which the taxonomy used in this paper is developed to classify those algorithms.

The taxonomy used as a reference for the classification is depicted in Fig. 1. In addition, three new elements have been added to the taxonomy that are related to SSL in several of the articles that were reviewed: active learning, transfer learning, and safe semi-supervised learning. In the literature, active learning and safe SSL have only been related to inductive learning methods, while transfer learning has been applied to inductive and transductive learning. Nonetheless, active
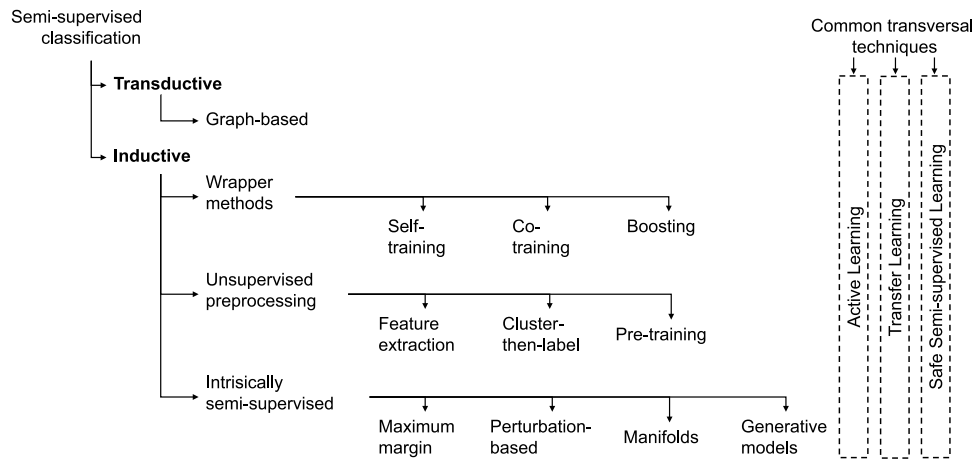
Fig. 1. Taxonomy for semi-supervised classification methods. The original taxonomy proposed in [7], in addition to the three additional techniques found in the review.
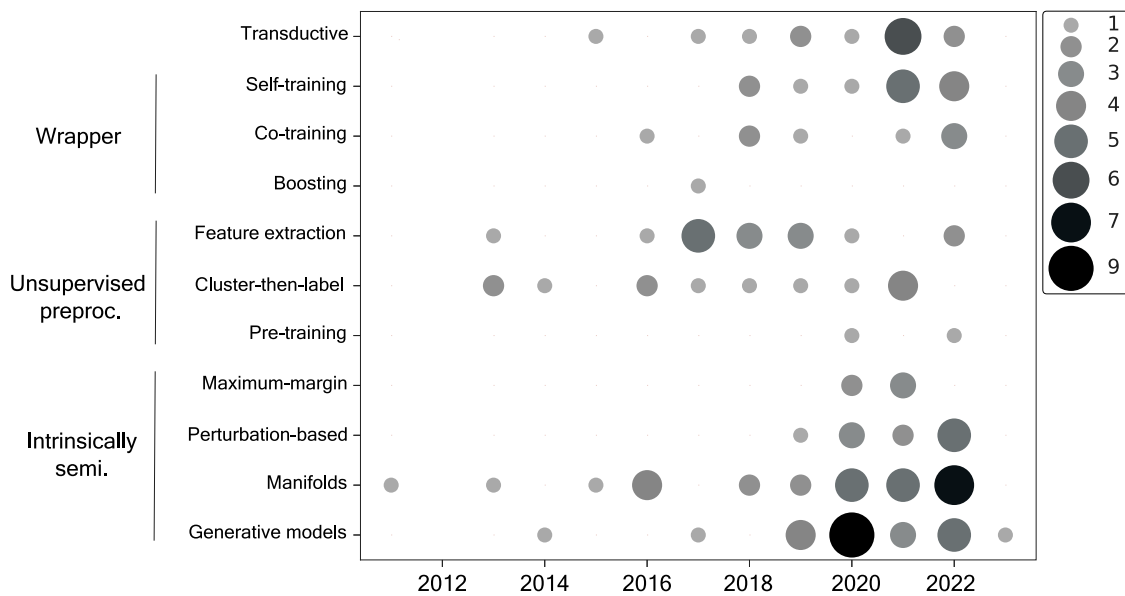

Fig. 2. The different articles under review are sorted by year and classified according to the taxonomy of Engelen & Hoos [7]. The inductive methods are further grouped into the three subcategories: wrapper methods, unsupervised pre-processing, and intrinsically semi-supervised.

learning, safe SSL, and transfer learning are transversal to SSL methods and can be applied to more than one method or type of method.

Although each semi-supervised category can be divided into further subcategories, as was done by Triguero et al. [24], the more general and well-established taxonomy proposed by Engelen & Hoos [7] was used in this review, as the taxonomy proposed in [24] was partial and referred only to wrapper methods. In Fig. 2, the articles are sorted by year and grouped under to their corresponding categories according to the aforementioned taxonomy. The evolution of articles in terms of category distribution will be further explained in subsequent sections.

The following sections compile and arrange the extensive literature on FDD in which SSL is used. Transductive and inductive methods are divided into different sections to enhance comprehension and organization, respectively, Sections 4 and 5.

4. Transductive

The initial group of reviewed SSL methods comprises transductive methods. As previously noted, the only label that transductive algorithms can predict is the label of unlabeled instances already available in the training phase; they are always based on graphs. Methods are typically composed of three phases [7]: graph creation, graph weighting, and inference (label propagation). Among the articles under review, the transductive proposals for FDD are summarized below.

Several authors have researched various ways of building a graph: Zhao et al. [38] used Non-negative Sparse Coding (NSC) to construct their graphs; k-Nearest Neighbors (kNN), which is based on smoothness assumption, was used to build the graph in [39]; while Wang et al. [40] used Non-negative Matrix Factorization (NMF).

Several authors have based their methods on Greedy Gradient Max-Cut (GGMC) for label propagation [41–44]. Zhao et al. [45] proposed Graph Based SSL (GBSSL) applied to photovoltaic faults. Momeni et al. [46] then improved on this method, increasing its accuracy, so as to pinpoint the fault.

Some research was focused on imbalanced datasets, and how to deal with this problem. Qian & Li [47] used the Synthetic Minority Oversampling TEchnique (SMOTE). Likewise, Fang et al. [48] used a Generative Adversarial Network (GAN) in a labeled dataset to correct its proportions. Some other research is also of interest. Chen et al. [49] applied, in addition to the transductive graph, a Random Forest (RF) classifier to verify the graph predictions.

Transfer Learning between datasets was used in [50] to improve graph accuracy. Zhao et al. [51] used a transductive graph method, called Semi-supervised Local Kernel Density Estimation (SLKDE), to

classify historical off-line data and then to create a supervised classifier to detect and to diagnose on-line faults within wireless sensor networks.

As it can be seen, there are scarce few works related to transductive methods applied to FDD. This can be explained because in industrial applications is usually desired the ability of inductive methods to pseudo-label new instances that are not available on the training stage. This makes inductive methods more versatile and useful.

5. Inductive

Subsequently, the focus will shift towards exploring the realm of inductive SSL methods. Most of the articles under review used an inductive approach. Inductive algorithms generate a model, which once it has been trained, can be used to predict the label of new instances. Inductive algorithms can be further divided into three subcategories: wrapper methods (Section 5.1), unsupervised pre-processing (Section 5.2), and intrinsically semi-supervised (Section 5.3).

5.1. Wrapper methods

Wrapper methods are the first inductive method type to be reviewed. These methods first train a classifier (some wrapper methods train more than one classifier) using only the labeled instances for generating the predictions of unlabeled instances. Then, the classifier is (or the classifiers are) retrained using both the original labeled instances and the new labeled (also called pseudo-labeled) instances for improving the model, in a process that can be performed several times. With regard to the wrapper methods, Engelen & Hoos [7] identified three main groups that are listed below.

5.1.1. Self-training

Self-training is the simplest way to obtain a SSL algorithm from a supervised one. Self-training first trains a model with the labeled instances alone, and then the unlabeled instances are pseudo-labeled with the prediction of the trained model. Pseudo-labeled instances, whose prediction confidence is above a certain threshold, are added to the labeled dataset [52]. The training and pseudo-labeling processes are repeated until either there are no unlabeled data or a predefined threshold is reached. The work of Triguero et al. [24] is recommended, for an exhaustive literature review of self-training methods.

Self-training has been applied in various ways, modifying different sections of its implementation. For example, Zheng & Zhao [53] changed the way confidence is calculated for pseudo-labels. The authors used a Temporal-Spatial Confidence Measure (TSCM) to obtain temporal and spatial information from unlabeled and pseudo-labeled data. In [54] , the same authors modified confidence, so that it could be calculated as a distance.

However, the threshold itself was changed in several papers. Zhang et al. [55] determined the threshold using Monte Carlo and Liu et al. [56] used a dual-threshold, in order to avoid overfitting.

Other authors modified the way their pseudo-labels are selected to be added to the labeled dataset. In [2,29] used Active Learning Semi FDA (ALSemiFDA), a special self-training method whose pseudo-labeled instances are not fixed, and with which the pseudo-labeled instances could be corrected when mislabeled. Long et al. [57] presented two selection strategies for self-training method, Gradually Exploiting Mechanism (GEM) and Distance-Based Sampling Criterion (DBSC).

Other authors have focused on datasets of a specific type: imbalanced datasets were investigated in [58], the focus was on noisy datasets in [59], while datasets with new faults that were not available in the training phase were examined in [60].

5.1.2. Co-training

Co-training is the adaptation of self-training from the perspective of an ensemble.[4] Co-training methods are composed of two or more supervised models trained with the same (single view) or different (multiple views) labeled instances [62]. The models then predict the labels of the unlabeled instances that are updated (mainly by voting or by the sum of confidences). As in self-training, the process is repeated until all unlabeled instances have been pseudo-labeled in all views, or a predefined limit has been reached.

An example of single view co-training applied to FDD can be found in [23] where a GAN was used to generate unlabeled data which was then used in the co-training, combining the prediction, by voting, of a conditional Deep Convolutional GAN (cDCGAN)[5] and a Residual Network (ResNet18).[6] Another example of the co-training method with two base estimators, Decision Trees (DT) and kNN, is presented in [63].

In Tri-training [64], as its name suggests, three classifiers are trained. It is a multi-view co-training method that has been extensively used with FDD, as described in [30,65–67].

Other authors have proposed co-training methods with multiple (undefined) base estimators. On the one hand, Liu et al. [68] used Extreme Learning Machines (ELM) as a base estimator to detect and to diagnose milling faults. On the other, Huang et al. [69] proposed a co-training method based on fuzzy rules for demagnetization faults.

5.1.3. Boosting

Boosting [70] is an ensemble method that aims to distribute weights among the model predictions, so that greater weight is assigned to those instances that are mislabeled. When boosting is applied, several weak models are trained, which together generate a strong (good) classifier. The different base classifiers are combined based on weights that depend on their accuracy. Nonetheless, boosting has not been widely used in SSL for FDD.

Razavi-Far et al. [71] compared multiple SSL methods for feature extraction and for classification, noting that the ASSEMBLE [72] classifier achieved the best results. In [73] multiple feature extraction methods and SSL classification methods, among which SemiBoost [74], were also compared.

5.2. Unsupervised pre-processing

Having reviewed wrappers, the focus now shifts to the exploration of unsupervised pre-processing methods. Unsupervised pre-processing methods use unlabeled instances for different purposes, such as to extract features from the unlabeled data, pre-clustering the data, and setting the initial parameters of a supervised learning model in an unsupervised way. It must be noted that this type of methods performs these actions prior to the training of the final model. Unsupervised pre-processing can be further divided into three subcategories: feature extraction, cluster-then-label, and pre-training.

5.2.1. Feature extraction

Feature extraction is an attempt to transform the available data, either to improve the accuracy of the model, or to make its construction more efficient. These reasons are more important for the approaches that do not rely on neural networks to obtain the trained model and, therefore, a common step is to perform a semi-supervised feature

---

[4] Ensembles, also known as multi-classifiers, are methods that use more than one model (classifier or regressor) to improve their prediction capabilities; for further information, the authors would recommend the following work [61].

[5] cDCGAN is a GAN based method with which labeled data can be generated and classified.

[6] Residual Network (ResNet18) is a supervised classification deep neural network with skip connections.

extraction or dimensionality reduction, in an attempt to take advantage of both labeled and unlabeled observations.

Different kinds of Auto-Encoders (AE) have been used for dimensionality reduction and for extracting features from high dimensional datasets as in [18,75–78]. Because EAs can be used for dimensionality reduction and are unsupervised in nature, they provide an easy way to account for unlabeled items.

The different types of FDA variations, as applied in [5,79–84], form another typical approach for semi-supervised feature extraction and dimensionality reduction. Although FDA is a supervised dimensionality reduction algorithm and is therefore based on the output feature(s), there have been several proposals to obtain semi-supervised versions of FDA so that dimensionality reduction can be performed, taking into account both labeled and unlabeled data.

Yet another common approach is to use a manifold-based approach, such as the ones applied in [71,85–89]. Manifold learning methods try to reduce the dimensionality of the data while preserving the nonlinearities present in the data.

Some additional strategies to semi-supervised dimensionality reduction or feature extraction have been tested, such as the Support Vector Machine (SVM), as described in [90,91], and hypergraphs, as described in [92].

5.2.2. Cluster-then-label

A popular practice in SSL is to perform clustering first, either unsupervised or semi-supervised, hoping that the result of the clustering will serve as a guide in the subsequent classification process.

Proposals can be arranged into two main groups that use either any kind of semi-supervised approach or unsupervised clustering.

Semi-supervised clustering is a kind of clustering that at some point uses information (from labeled instances or constraints) that improves its performance. Semi-supervised clustering was used in the following proposals [4,45,93–97]. On the other hand, there are several proposals using an unsupervised (*i.e.*, k-Means, density peak clustering, among others) clustering algorithm [98–102]. In this case, the unsupervised clustering process is usually a prior step that clusters the data before any semi-supervised labeling process is performed, or the process is used as a pseudo-labeling step performed after a semi-supervised process such as feature extraction or a dimensionality reduction step.

It is worth mentioning that new types of faults, which were not present in the training dataset, may be detected with some of these proposed methods, [95,96,100].

5.2.3. Pre-training

Pre-training uses unlabeled data primarily for approximating weights in deep learning networks, such as Deep Belief Networks (DBN) and stacked autoencoders, before applying supervised learning for fine tuning those weights.

Ding et al. [103] used a modified ResNet for incipient fault detection using vibration signals. The neural network was firstly pre-trained and then a semi-supervised fine tuning was performed. Finally, it was able to classify under either normal or (incipient) fault conditions.

Liao et al. [104] proposed the use of an optimized DBN through the particle swarm approach for avoiding the optimization problems associated with DBN. The DBN must be pretrained in an unsupervised way before fine tuning using labeled data.

5.3. Intrinsically semi-supervised

Continuing with the discussion, the focus will now shift to the examination of the last category of inductive methods, known as intrinsically semi-supervised. These methods incorporate unlabeled instances into the objective or optimization function of the learning method. Intrinsically semi-supervised methods can be further divided into: maximum-margin, perturbation-based, manifolds, and generative models.

5.3.1. Maximum-margin

Maximum-margin methods represent an attempt to maximize the distance between the data points and the decision boundary, based on the low-density assumption. As in a supervised learning approach to margin maximization, SVM adaptations are often used. Several adaptations of SVM to SSL have been proposed throughout the past few decades [105], in which semi-supervised SVMs leverage information from unlabeled data to achieve better class separation.

Wang et al. [106] used a tree structure created with the Semi-Supervised Gaussian Mixture Model (SGMM) in combination with a Semi-Supervised SVM (S3VM). Jia et al. [37] proposed a Dynamic Active Safe Semi-Supervised SVM (DAS4VM) with active learning and safe SSL to detect faults using PCA as pre-processing. Mao et al. [107] also applied a S4VM where an online perspective was used to distinguish between faulty and normal states, among others. Other authors applied the low-density assumption to their neural networks, in order to separate data of different classes [108,109].

5.3.2. Perturbation-based

Based on the smoothness assumption, perturbation-based methods try to add noise to the data or apply noise directly to the model, since if the noisy data and the real data are similar they should have the same label.

Neural networks are one of the most commonly used techniques in perturbation-based methods, for example in [110] a Semi-Supervised Deep Ladder Network (SSDLN) with information fusion on a gear failure dataset was tested. Chen et al. [111] also used the Ladder Network,[7] this time applied to photovoltaic systems.

Another type of neural network that has been modified to be an SSL perturbation-based method is Long Short Term Memory (LSTM).[8] Zhang & Qiu [112], for example, applied LSTM to a chemical dataset; while Tang et al. [113] used LSTM with transfer learning to address bearing faults.

Some authors used Data Augmentation/Generative methods to improve dataset information and then used the improved dataset to train a perturbation-based neural network [114–116].

Other types of modifications have been proposed in various studies, in order to apply perturbation-based methods to the FDD problem. Hu et al. [117] used two networks, one for inter-instance information and another for intra-temporal information, which were then mixed in the final loss function to classify bearing faults. Bearing faults were also taken into account in [118] where a Deep AutoEncoder (DAE) was used to preprocess bearing data, and in [119] where an attention method called squeeze-and-excitation was used. Shim et al. [120] proposed a three-phase framework for wafer semiconductor manufacturing. The framework was special, in so far as different types of ML (supervised, unsupervised, and semi-supervised) methods were applied in each phase, depending on the amount of labeled data available.

5.3.3. Manifolds

Centering on the manifold assumption, manifold methods usually modify the input feature space to calculate distances between data points, which can also be achieved with graphs. Manifolds methods have been widely applied to different FDD problems.

Most of the research on manifold SSL applied to FDD is graph-based. It is common to generate a graph and then calculate classes from a distance type [121–125]. Multiple graphs are used in some manifolds, in order to represent the information [126–128]. While others used different techniques to add more information to the graph [129].

---

[7] Semi-Supervised Ladder Neural Networks are an accurate representation of perturbation-based methods, as this type of network modifies its cost function to incorporate both labeled and unlabeled data, all with backpropagation.

[8] LSTM is a type of recurrent neural network architecture designed to effectively capture and model long-term dependencies in sequential data.

Some research can be grouped by the way the graph that represents both labeled and unlabeled data is generated. The most common graph creation method, kNN, is used in several papers, such as [130–132].

Other methods are characterized in different ways. Kernel functions were used to modify input data, so that the data were distributed over the different manifolds, as described in [133] for the production process of fused magnesia and in [3] for the Hot Galvanizing Pickling Waste Liquor Treatment Process (HGPWLTP). Fan & Zhang [134] also approached the HGPWLTP problem, though they used Laplacian Regularization.

In some other studies, the techniques were applied to improve the original dataset, such as active learning, as used in [135]. In yet others, transfer learning was used to add information from the source domain to improve the classification of the target domain [136–139].

Different methods to perform interesting types of pre-processing were used in some other investigations, which consisted of transforming the vibration data into images [19,128,140].

Liu et al. [141] focused on the Riemannian space instead of the Euclidean space that is typically used. Gao et al. [142] created a manifold method combining a Convolutional Neural Network (CNN) and AE to create Pseudo-Label CNN (PLCNN) for FDD. Razavi-Far et al. [73] proposed Semi-Supervised Smooth Alpha Layering (S3AL), performing a comparative study between different SSL algorithms.

5.3.4. Generative models

Generative models are designed to create new instances from labeled data. A discriminative function of the generative model classifies most of the new unlabeled data. Several generative methods were proposed in [7] to deal with FDD SSL, all of which rely on the three subsets: mixture models, generative adversarial networks, and variational AEs.

Mixture models are very useful when the distribution of the data is known and a mixture model is built based on several distributions. Several pieces of research have applied mixture models to FDD [143–146].

The most commonly used model is nevertheless the semi-supervised GAN. Typical unsupervised GANs are mainly composed of a generator model, which creates new data from training data, and a discriminator that serves to detect which data are real (an original labeled data) and which were created by the generator. In semi-supervised versions of GAN, the discriminator is modified, so that it can predict the data labels [147–161].

A Variational AE (VAE) [162] is a latent variable model in which data are treated as they are generated from a vector of latent variables [163–166].

As has been seen in other SSL method types, several works transform their data, mainly vibration-based, into images, in order to be able to use CNN in the generation process, some examples of which can be found in [148,151,154,155,157,160].

It is also remarkable that some of the methods proposed in these articles are able to detect new types of faults, not seen in the training phase [143,146,152].

6. Results and discussion

Following the thorough review of papers on SSL for FDD, it was found that the results could not be directly compared, due to differences between the experiments and the implementations. Instead, a journal analysis was performed, to show the top journals in which articles on SSL for FDD have been published (Section 6.1). In addition, the evolution of publications on SSL for FDD has been studied, to observe the patterns over time and the possible future of the research field (Section 6.2). Finally, a popularity comparison was performed, to find out which methods appear to be the most widely used (Section 6.3).

6.1. Journals

The papers under review were published in different journals, some of which were more generic while others were more specific. The journals were divided into three different groups: industry, artificial intelligence, and mixed.

As can be seen in Fig. 3, industry theme journals are the most popular (59.85%), which may show how industry issues are more relevant for selecting a journal in which to publish research than the ML methods themselves. But this approach also explains some limitations in the use of SSL in industrial environments; a field where most of the researchers involved in most of the works are closer to the industry than to computer science.

IEEE Transactions on Instrumentation and Measurement is the industry journal with the most publications (6.57%). At the other extreme, artificial intelligence journals hardly pay enough attention to real solutions for industry, restricting themselves to publications on the development of *ad-hoc* SSL solutions for FDD. Therefore, a major role is expected from hybrid journals, that can build a bridge between these two disciplines in the near future. It is worth mentioning that *IEEE Transactions on Industrial Informatics and Chemometrics* and *IEEE Access* are the journals with the most FDD SSL publications (8.76% and 8.03% respectively).

6.2. Publications throughout time

At the beginning, from 2011 to 2015, several SSL method types were used for FDD, as it can be seen in Fig. 2. Something that stands out from these years is the absence of wrapper methods, despite the fact that those methods are the easiest transition from supervised learning to SSL. This gap may be due to the late application of SSL to FDD, after the development of other types of methods [21].

From 2015 to 2018, feature extraction methods were the most popular. However, intrinsically semi-supervised methods have been the most widely used since 2018, and mainly methods using the manifold approach and generative type methods. The latter because of their ability to generate new data and their good performance and popularity in other fields [167]. Furthermore, transductive and self-training methods have also gained popularity quite recently.

It is worth noting that there are some types of methods, such as manifolds and cluster-then-label, that have been widely used since first introduced.

Overall, focusing on Fig. 4, it is clear that the application of SSL methods to FDD problems has been an emerging topic, since its first application up until today. It is thought that this evolution will be maintained over coming years.

6.3. Popularity

As previously noted, inductive methods are by far the most frequently used compared with the transductive ones. Over 90% of the articles reviewed for this study are related with inductive methods. The possible main reason why inductive methods are more popular than transductive ones is because of their capability to predict new data that were not been used when training the model. For this reason, there is no need for retraining with inductive methods whenever new unlabeled data are processed.

As can be seen in Fig. 5, wrappers were the least used (18.18% of inductive methods) inductive methods, closely followed by unsupervised pre-processing (25.62%). Finally, intrinsically semi-supervised methods were the most popular among the inductive methods (56.20%). The reason why intrinsically semi-supervised models are so popular is because they are the most complex and can be adapted to different problems.

Regarding the wrapper methods, self-training and co-training were the most used (10.74% and 6.61% respectively). It was a bit surprising
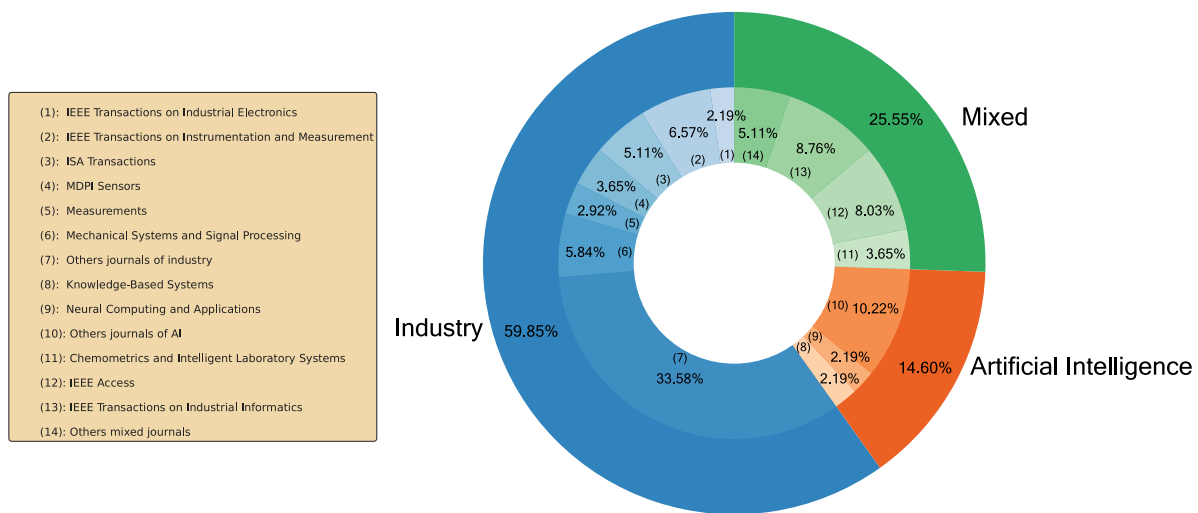
Fig. 3. Distribution of papers, according to their aim and scope by journal of publication: industry, artificial intelligence, and a mixture of both.
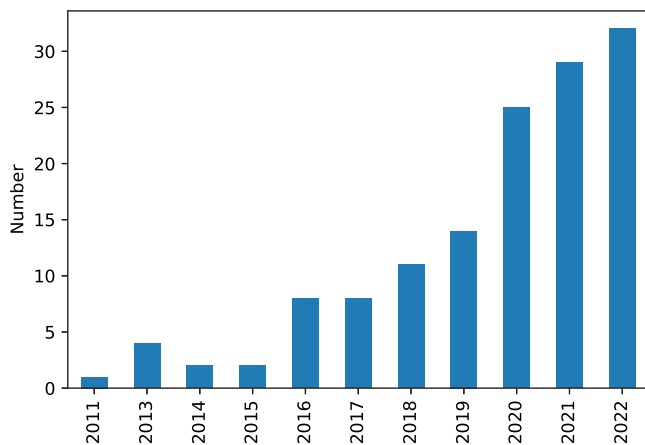

Fig. 4. Evolution of the number of publications of SSL applied to FDD over the years.

that self-training was more popular than co-training, because the use of ensembles should bring better results [168]. Nonetheless, boosting methods could be applied more often to FDD problems, to obtain better results and to add to the popularity of wrapper methods, as was shown in [71].

Not all unsupervised pre-processing methods have been equally applied to FDD, as few journal publications that used pre-training methods to approach this problem were found (1.65%). The scarce use of pre-training methods is probably due to the low impact that can be achieved compared to training a whole model based on SSL. The other unsupervised pre-processing methods are widely used: cluster-then-label methods were mentioned in 10.74% of the articles, meanwhile feature extraction methods, the third most used subgroup, were slightly more popular at 13.22%. The main reason for the popularity of feature extraction methods is their ease of application, in preparation for the use of SSL.

Intrinsically semi-supervised methods are the most widely used, but not all of their different types were applied with the same frequency. On the one hand, the maximum-margin and the perturbation-based methods are the least used (4.13% and 9.09% respectively). On the other hand, generative models and manifolds are widely used. The capability to generate new unlabeled data is the main reason why generative models are so popular (19.83%). Manifold methods that are based on the manifold assumption are the most popular subgroup (23.14%), due to their ability to modify the data space and to classify.

The results on the popularity of different SSL types are not random; instead, they are aligned with the advantages and disadvantages associated with each type.

A significant distinction arises between inductive and transductive methods. As mentioned earlier, over 90% of the articles in this review are based on inductive methods. The primary reason for such a high imbalance proportion is the key advantage of inductive methods over transductive methods – the ability to predict new data that was not seen during the training phase –.

Focusing more deeply on the popularity of the different types of inductive methods, a more comprehensive analysis can be obtained. To do so, a summary of the advantages and disadvantages of these types is presented in Table 1. As mentioned earlier, wrapper methods are the least used inductive methods, primarily due to the fact that most of them are prone to add noise, to misclassify pseudo-labels and to add them to labeled dataset and are unable to correct it. On the other hand, the use of unsupervised pre-processing methods is astonishing due to their lower potential impact when utilizing unlabeled data. Finally, the popularity of intrinsically semi-supervised methods is normal, considering their ease of development from supervised methods and the ability of unlabeled data to directly influence the objective function or the optimization process.

Taking into account the aforementioned popularity and considering the possible reasons (including the different pros and cons of the SSL methods), certain gaps may be identified in the application of SSL to FDD. One potential gap lies in the gain that can be obtained by applying boosting methods on wrappers: which can improve their performance through the use of ensembles and the weight distribution of these models. Another potential gap is the limited application of maximum margin methods within intrinsically semi-supervised methods, particularly the underuse of semi-supervised SVM. A broader use of semi-supervised SVM could be beneficial, given its good performance in applications across various fields [169].

Unsupervised preprocessing methods are not widely used, primarily due to their limited potential impact when utilizing unlabeled data. However, one approach to applying deep learning methods from a SSL perspective is to employ a pre-training method that enables the unlabeled data to influence the pre-configuration of the deep network.

One additional gap, irrespective of the method type, pertains to the application of methods capable of predicting new fault types that have not been observed in labeled data.

6.4. Industrial issues

The industrial applications of FDD collected in this review are varied, but most of them are focused on two main topics: chemical
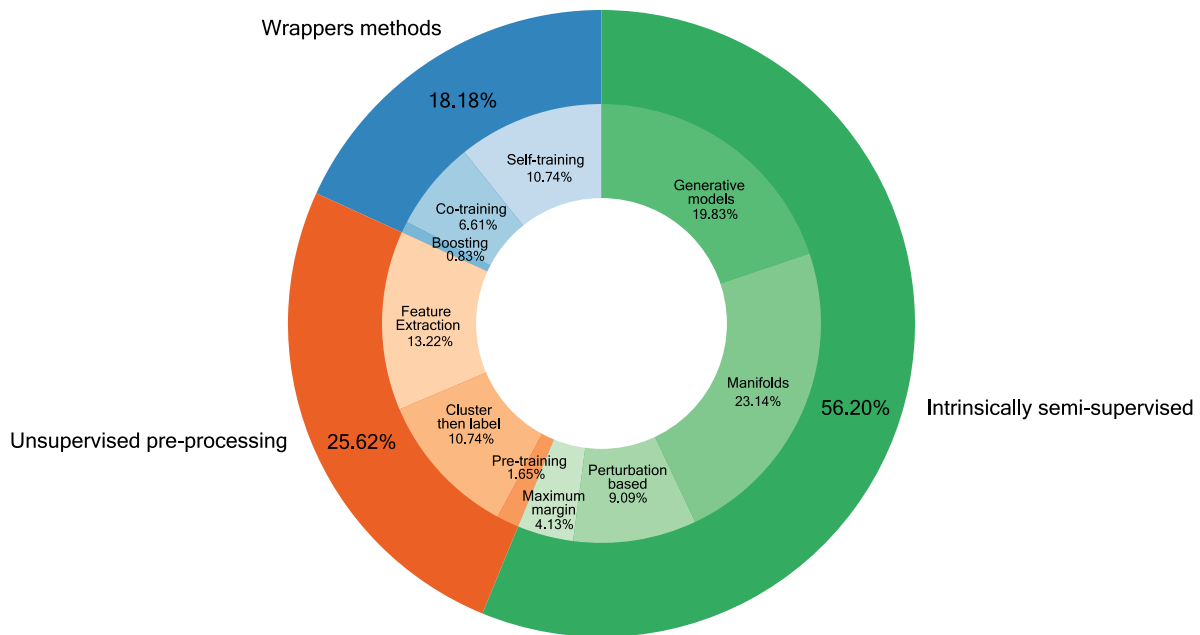
Fig. 5. Distribution of inductive SSL FDD publications found in the review, grouped by type of method as *per* the taxonomy in [7].

Table 1
Pros and cons of inductive methods.

Method	Pros	Cons
Wrapper methods	1. Easy to implement. 2. Configurable, easy to change base estimator/s. 3. Can be used with almost any supervised method.	1. Prone to add noise. 2. Dependent on supervised methods
Uns. preprocessing	1. Can be used with almost any supervised method.	2. Less impact of unlabeled data.
Intrinsically semi.	1. Unlabeled data is used on the lower level (objective function or optimization procedure). 2. Usually easy to develop from its supervised version.	1. More complex models Harder to train. 2. Most of them require large amounts of data.

and mechanical processes. The mechanical processes are more common and form part of different fields, from energy generation to machining. They all merge in so far as they serve: mechanical chain-based problems that are conventionally evaluated in terms of vibration analysis (*e.g.*, bearing and gearbox wear and tear). More than 80% of the research has been directed at vibration-based problems , due to the high capability of vibrations to collect useful information on machine performance levels. In these cases, after an ordinary vibration analysis, it is very common to apply Fourier transform or discrete wavelet transform techniques to analyze vibration information. However, some new approaches, passing vibration behaviors from numbers to images use a different solution, such as Continuous Wavelet Transform. As both approaches are based on the extraction of information from vibrations, the richest source of information on mechanical chain dysfunctions, both should be considered optimal.

As previously noted in Section 2.1, FDD data are difficult to obtain. Most industrial FDD datasets are imbalanced, as data on normal/healthy state are easier to obtain than data on failures. There are even datasets that contain no data on all possible failures, but only on some of the most common ones. Other datasets only contain data on the early degradation states associated with catastrophic failures (*e.g.*, data on high levels of imbalance in a wind-turbine gearbox are extremely rare). As the industry is reluctant to produce datasets with significant numbers of failures, researchers are obliged to generate those data in laboratories. Considering this industrial fact, the datasets tested in each piece of research can be classified into three types: open access benchmark, laboratory, and industrial datasets.

The first type, open access benchmark datasets, are publicly available datasets that refer to different industrial processes and include

different types of failures. They are not only useful for the validation of new SSL techniques, but also to compare them with the accuracy of state-of-the-art SSL methods. Some of the most common benchmark datasets are:

- **CWRU**[9] (Case Western Reserve University): a bearing defect dataset whose features are vibrations in different (motor loads and different fault levels) study cases. There are 4 classes in total with balanced numbers of instances, 3 faulty (ball, inner race, and outer race), and the normal/healthy state [170].
- **IMS**[10] (Intelligent Maintenance Systems): a bearing database available at the University of Cincinnati, consisting of 3 different datasets. As in CWRU, these datasets focus on vibration data and have 4 classes with the same 3 fault states as CWRU, but in this case classes are more unbalanced than CWRU dataset [171].
- **Paderborn University Dataset**[11]: another bearing dataset with vibrations features. There are 32 different classes, of which 6 are healthy states, and they are roughly balanced. The damaged bearing classes can be joined into inner faults and outer faults.
- **TEP**[12] (Tennessee Eastman Process): a chemical industry dataset on a process that involves 5 main elements, namely a reactor, a

---

vapor–liquid separator, a recycle compressor, a product stripper, and a product condenser. The dataset is composed of 52 features, and there are 21 different fault types, that are unbalanced represented on the dataset, 5 of which are unknown [33,172].

- **NEU**[13] (Northeastern University): a dataset of steel surface defect images. There are 1 800 images with 6 different fault types (crazing, patches, rolled-in scale, inclusion, pitted surface, and scratches). Despite the presence of multiple fault types, it is remarkable that the classes are balanced.

It should be noted that the first three datasets refer to the degradation of bearings in mechanical chain analyzed on the basis of vibration data. This kind of industrial problem is usually disengaged (the degradation of one bearing will not generate degradation in other bearings of the mechanical chain) and each bearing will have a natural frequency that is easily isolated within the FFT spectrum. More complex problems in mechanical chains like axis misalignment are not considered in these datasets. Only the TEP dataset includes failure in continuous no-rotatory process, such as chemical reactions, with an interconnectivity between failures, revealing a more complex relation between failures, although they only take into account a limited set of possible failures. Finally, NEU dataset, based in images, is a very different approach to FDD and cannot be directly compared with the other benchmark datasets. These types of datasets are mainly close to balance between different classes, as was noted in their presentation.

The second type, laboratory datasets are designed to obtain data, mainly from a testbed, including as many machine states as possible. This dataset type is really useful in FDD problems where some kinds of fault are very rare, because under laboratory conditions, all kinds of faults can be caused and different levels of severity can be characterized and tested (*e.g.*, imbalance in a rotor can be tested at different well-characterized levels). These datasets have three main drawbacks that limit the direct use of the results obtained under real industrial conditions. First, they are extracted from very well and over-sensorized testbeds, far away from industrial conditions where the sensorization is very much more limited (*e.g.*, an accelerometer placed close to the cutting tool was used to evaluate cutting-tool wear in a machining process, and a Kistler table was included in the experiment to measure cutting forces; under industrial machining conditions, coolant fluids and chip production make it necessary to distance the accelerometer from the tool tip, while a Kistler table is usually never used, due to its cost). Second, the testbeds are a scale reproduction of real machines (*e.g.*, a lab testbed for a windmill mechanical chain is usually around 1/10 of an industrial one) and their performance might differ from an industrial one. Finally, the cost of setting up these industrial experiments complicates the repeatability of the research. Besides, it should be outlined that these types of datasets are mainly balanced, because under laboratory conditions, testing failure conditions could be more easily performed than under real industrial conditions. This fact creates an extra-difference between industrial datasets and these datasets.

The third type, real industrial datasets, are the least common, because they are the most difficult to obtain and industries are very reluctant to let them be used for research. In addition, this type of dataset will not often have all possible failures represented or will have a very limited number of failures levels, which complicates the ability to create a model that is capable of predicting states across the entire data space (*e.g.*, mechanical models usually use experimental data from no failure and light failure levels to fix thresholds, but ML models also require data from more severe failure states to achieve highly accurate models). Besides, this kind of datasets are usually strongly unbalanced, because under industrial conditions, test of failure conditions is difficult and very expensive.
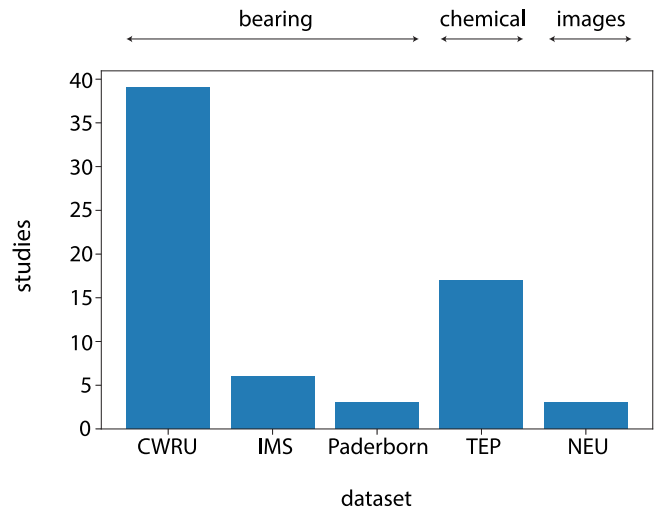

Fig. 6. Frequency of use of the different datasets in the articles under analysis. The diagram depicts the most frequently used datasets, in addition to the other databases referenced in 113 articles considering the nature of the industrial problem.

In most FDD research, at least one benchmark dataset and one laboratory or real industrial dataset in a second stage are usual, to overcome the limitations of the different type of datasets. If no extensive public industrial dataset is available for research within a specific field, this solution might be considered best practice.

When considering the level of labeled instances, usually the first two types have all their instances labeled, while the industrial ones are the only ones that have many unlabeled instances. Most researchers therefore only apply some of the labels when using open access benchmark or laboratory datasets under SSL conditions. These datasets are not actually semi-supervised, because all the instances are labeled (labels are removed from some usually randomly chosen instances). A rare counter example is found in [78], which used a completely unlabeled set of data collected over several years relating to the operation of an electrical network.

Fig. 6 shows the papers under review that are classified by the dataset used for either training or testing purposes. As can be seen, the most commonly used dataset for testing is CWRU. However, articles using other datasets are the most numerous with 113 appearances. Among these articles, some refer to publicly available datasets, while others use their own dataset for training and testing. Table A.2 in Appendix shows the references classified by the dataset in use.

When analyzing the use of SSL methods in terms of the industrial problem to be solved, we can see that there is no clear relationship between any particular method and a specific problem; a fact that might be expected, due to the novelty of SSL techniques and their application to FDD problem. The popularity of the methods was almost identical for each industrial issue, which shows that the decision to apply SSL methods to FDD problems is based on the most commonly used FDD methods, and not on specific methods for each industrial problem. Moreover, no baseline method has been found with which they may all be compared. These two conclusions demonstrate that the use of SSL techniques for FDD problems is still far from being a standard solution and will require extensive research in coming years to establish reliable reference solutions for each industrial process against which to test the new proposals. However, at the same time, that fact outlines the interest and possibilities of future research into the application of SSL to FDD.

Finally, two main industrial impediments have been found in this research. Firstly, neither the SSL methods nor the use of unlabeled data are clearly explained in several research papers, even though such an explanation might be essential for an understanding of SSL methods.
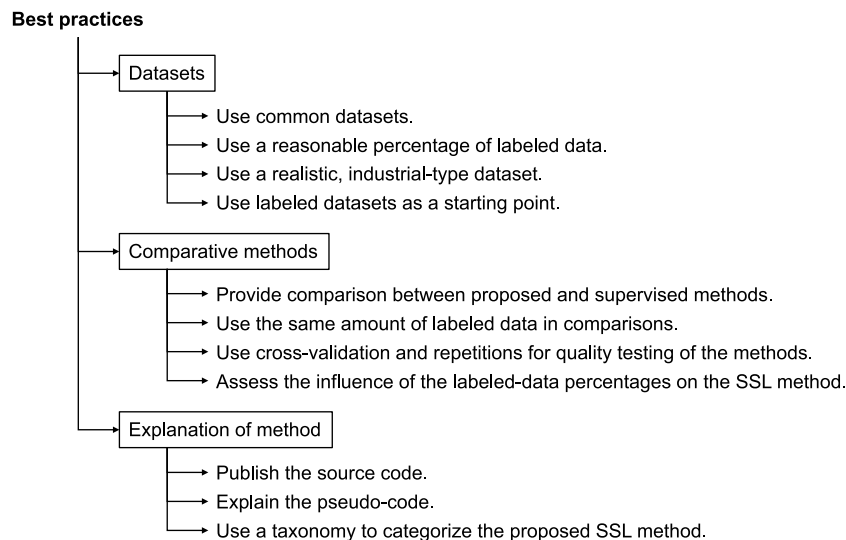
Best practices

Fig. 7. Proposed set of best practices for Semi-Supervised Learning (SSL) in industrial scenarios.

Secondly, several pieces of research use high labeled percentages, which hardly correspond with the real state of an industrial problem. SSL is meant to be applied in situations where labels are scarce, and FDD industrial problems are one of these situations, so testing high labeled percentage makes no sense. For these reasons, neither strategy should be applied in any future research.

7. Conclusions, best practices, and future trends

In this section, the analysis of SSL for FDD is concluded and a set of best practices are proposed, which have been established as successful solutions within this field over the last decade. The future within this field is promising and some research trends that are likely to gain attention over coming years are also described.

7.1. Best practices

The first attempt to apply a new Machine Learning technique is often far removed from the most optimal application method. Therefore, during this first decade of using SSL for FDD, some of the proposed strategies have shown some inconsistencies. The most common ones are: (1) considering ML techniques and strategies that cannot be considered as SSL (*e.g.*, mixing supervised and unsupervised methods and referring to them as semi-supervised), (2) the use of too-high a percentage of labeled instances makes it difficult to justify the use of any SSL method, for instance, in [88] 25% of the observations were unlabeled, and in [86] only 10% of the observations were unlabeled, whereas in other proposals fewer than 50% of observations were unlabeled, and (3) the presentation of new though somewhat vague SSL methods, reduced the chance of experimental repeatability.

The following set of good practices (see Fig. 7) are proposed, to overcome these limitations in future works:

- Dataset related:

    Use of common datasets. The use of a publicly available standard or reference datasets, such as those discussed in Section 2.1 (CWRU, TEP, NEU, Paderborn…), makes it possible to compare the results obtained with different models and to replicate the experiments.

    As different approaches and methods are usually proposed, and as there is no one algorithm better than all the others, the only way to compare in which specific cases or type of problem one proposal may be better or worse than

another is to have a reference for their comparison. A natural benchmark would be the use of common datasets (problems). It can be especially important in industrial environments, where the presence of noise in the data, insufficient variables available in the observations, and the need to perform certain transformations on the data can ruin the learning capabilities of some algorithms.

Use of a realistic, industrial-type dataset. Datasets from real industrial environments often have the characteristics of being imbalanced, typically containing fewer observations relating to error conditions than to normal operating conditions, and perhaps no observations of some interesting types of failures at all. The use of datasets with other characteristics might yield unexpected results or only an upper, perhaps unrealistic results boundary, which can be obtained when used in a real-world situation.

Combine open access benchmark datasets with industrial/laboratory datasets to validate new SSL techniques. In the first place, the use of an open access benchmark dataset to assure improvements under general FDD conditions of any new SSL technique is advisable. Then, any new research can test the performance of the new SSL technique in either a laboratory or an industrial dataset, to show the unique advantages of this technique for a well-defined FDD industrial task.

Use of a reasonable percentage of labeled data. Ideally, using a small number of labeled instances along with unlabeled instances (presumably easier to obtain and less expensive) is one way the problems associated with labeling the instances needed to train the models can to some extent be alleviated in SSL. However, some of the proposals under review use high percentages of labeled instances for training, *e.g.*, in [81], the authors used 50% labeled instances for training, making it difficult to justify the use of SSL.

Although the size of the datasets for training can greatly affect the accuracy of the model obtained, it should be considered whether the use of a SSL algorithm makes sense or not. The use of small, but almost completely labeled datasets or the use of a large enough number of labeled instances to train a supervised learning algorithm – even if it represents a small percentage of a large dataset –, may be the main reason for using the SSL approach: increased accuracy while reducing the number of labeled instances is unfeasible.

Use of labeled datasets as a starting point. It is interesting that a fully labeled dataset could be used as a starting point. Although the process of selecting observations for label removal may introduce random behavior or be biased, it will allow the results predicted by the model to be checked against known results. An aspect that is especially important in industrial processes, where similar systems may behave very differently and identifying the presence of a fault, the specific way in which a system may fail, or the degree of failure might even be impossible for people with no industrial background.

• Comparative methods:

Provide comparisons between the proposed method and supervised methods. For reference purposes, the results should be obtained from a supervised version of the proposed method or from one or more typical supervised methods trained with only the corresponding percentages of labeled data. In this way, the improvement of incorporating unlabeled data over the use of labeled data alone can be tested.

Using the same amount of labeled data to compare. Using the fully labeled dataset to train a supervised model and comparing its results with a semi-supervised model trained with a low percentage of labeled data is not a fair comparison from which valid conclusions may not easily be reached.

Quality testing of the methods by using cross-validation and repetitions. The use of k-fold cross-validation and repeat experiments are common ML practices that yield better estimations of the results than the model can achieve in a real environment. Practices that, unfortunately, are not usually found in industry publications.

Assess the influence of the percentage of labeled data on the SSL method. The number of labeled items can greatly affect the performance of learning methods, which often happens with industrial process datasets. It is interesting to vary the number/percentage of labeled data, in order to evaluate how the learning method behaves.

• Explanation of the method:

Publish the source code. The availability of the source code is paramount in any method that is either proposed or compared. It facilitates any use of the methods and comparisons between proposals. Publishing the source code is a common practice in other research fields [173].

Explain the pseudo-code. Pseudo-code is a good tool for explaining new methods, but they must be accompanied by a proper explanation.

Use a taxonomy to categorize the proposed SSL method. Classifying the proposed methods according to a taxonomy facilitates the search for specific methods, helps readers to categorize the proposals, and facilitates the classification of the methods once reviewed. Some proposed methods may be difficult to classify and help from the authors might be appreciated. The use of a taxonomy (*e.g.*, such as the one proposed in [7]) may assist authors with the selection of references and researchers with the selection of relevant articles.

7.2. Future trends

On the one hand, as has been previously discussed in Section 6, the most popular type of methods are generative, manifolds, and feature extraction. Methods of this type will continue to be widely used because of their proven effectiveness in the field. Nonetheless, other methods that are rarely used nowadays, might become more popular in the future. As with supervised learning [174], SSL can take advantage of the improvement offered by boosting to enhance the performance of co-training methods. An interesting boosting technique that could be applied to SSL is statistical boosting [175] where boosting is no longer a black-box method.

A potential future trend could be the extensive use of semi-supervised SVM, which falls under the category of maximum margin methods. This would be a plausible development considering the relatively low popularity of maximum margin methods compared to other intrinsically semi-supervised methods, alongside the growing popularity of SVM.

Moreover, some other SSL methods that have not been applied to the field, such as pre-training, represent a a research pathway that, coincident with the growing popularity of deep learning methods, could yield promising results. A recent and up-to-date review of Deep Semi-Supervised Learning can be found in [176].

Recently, the use of meta-learning in SSL FDD has been presented in some works [119,137]. Nevertheless, its use has not been properly explored and FDD can benefit from meta-learning via deep learning [177] among other approaches.

Besides, most of the reviewed articles are concerned with FDD in bearings and gearboxes. The conventional proposal for such problems is vibration measurement, frequency transformation, and statistical analysis, although complex failures and industrial conditions with unlabeled data are a natural border to this solution that new SSL-based approaches can overcome. As previously outlined, bearing analysis using frequency spectrum might be the easiest FDD task in mechanical chains, where coupled failures (e.g., axis misalignment) might lead to failure states of more complex diagnosis. However, SSL might be a proper solution, due to the lack of labeled instances in those complex failures. Besides, the extension of the process information to other sources of signals (e.g., electrical or power signals) of easy-to-measure under industrial conditions, rather than vibration data, which are usually more complex to be measure in industry, could open a more generalized use of machine learning techniques and SSL in Industry 4.0. As rotary-based processes seem to be the most interesting and complex in industry, solutions to rotary-based problems where failures are more difficult to detect also have a promising future (e.g., surface defect detection or breakage of machining cutters). Any industrial process working under conditions that are far from stable (e.g., windmill electrical generation under real-variable wind conditions and machining processes such as milling where the cutting paths cannot assure a stable material removal rate) are promising fields of application for SSL techniques. Those characteristics converge in other industrial tasks where SSL has recently been tested for the first time, such as failures in air separation units [28], automotive assembly [80], stirred tank heater processes [81], solar photovoltaic arrays [45], laser powder-bed fusion [145], chemical batch processes [143], and a wireless sensor network [51]. It therefore shows that other industrial processes with highly variable working conditions, complex behaviors and industrial limitation on the labeling process, can benefit from the use of SSL for FDD. Besides, the consideration of data imbalance combined with unlabeled data will be a promising research line for FDD, due to the high industrial demand for solutions under these conditions, specially if costly and sensitive sensors, such accelerometers, can be avoided and more reliable and easy-to-measure signals can be used in their place.

Finally, a problem that has frequently arisen in this research is the difficulty of establishing a category for some methods. It shows that a further revision of the Engelen & Hoos [7] taxonomy may be necessary in this field for further development of the topic.

Table A.2
References grouped by the datasets used. It must be noted that a reference can be in more than one row if it uses more than one dataset.

Dataset	References using the dataset
CWRU	[19,30,38,39,41,50,57,66,71,73,76,90,93–95,98,104,109,113,114,119,122,128,129,131,132,135,136,138,140,141,151,154,157–160,166,178]
IMS	[30,66,107,116,158,166]
Paderborn	[76,118,140]
TEP	[2,5,18,28,29,37,53,54,77,80,81,83,86,112,121,125,148]
NEU	[23,126,142]
Other	[3,4,19,28,29,39–51,55–60,63,65–69,75–78,80,81,84,85,87–92,94,96–104,104,106–117,119–125,127–131,133–137,139–147,149–153,155,156,158–161,163–165,178–182]

Declaration of competing interest

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

Acknowledgments

Appendix. References using each dataset

See Table A.2.

References

[1] Zhang W, Wang Z, Li X. Blockchain-based decentralized federated transfer learning methodology for collaborative machinery fault diagnosis. Reliab Eng Syst Saf 2023;229:108885.

[2] Yin L, Wang H, Fan W, Zhou J, Yu G. Combining active learning and Fisher discriminant analysis for the semi-supervised process monitoring. IFAC-PapersOnLine 2018;51(21):147–51. http://dx.doi.org/10.1016/j.ifacol.2018.09.407, 5th IFAC Workshop on Mining, Mineral and Metal Processing MMM 2018.

[3] Jia Q, Du W, Zhang Y. Semi-supervised kernel partial least squares fault detection and identification approach with application to HGPWLTP. J Chemometrics 2016;30(7):377–85. http://dx.doi.org/10.1002/cem.2803.

[4] Leturiondo U, Salgado O, Ciani L, Galar D, Catelani M. Architecture for hybrid modelling and its application to diagnosis and prognosis with missing data. Measurement 2017;108:152–62. http://dx.doi.org/10.1016/j.measurement.2017.02.003.

[5] Zhong S, Wen Q, Ge Z. Semi-supervised Fisher discriminant analysis model for fault classification in industrial processes. Chemometr Intell Lab Syst 2014;138:203–11. http://dx.doi.org/10.1016/j.chemolab.2014.08.008.

[6] Li X, Yu S, Lei Y, Li N, Yang B. Intelligent machinery fault diagnosis with event-based camera. IEEE Trans Ind Inf 2023;1–10. http://dx.doi.org/10.1109/TII.2023.3262854.

[7] van Engelen JE, Hoos HH. A survey on semi-supervised learning. Mach Learn 2020;109(2):373–440. http://dx.doi.org/10.1007/s10994-019-05855-6.

[8] Chen KY, Lim CP, Lai WK. Application of a neural fuzzy system with rule extraction to fault detection and diagnosis. J Intell Manuf 2005;16(6):679–91. http://dx.doi.org/10.1007/s10845-005-4371-1.

[9] Abid A, Khan MT, Iqbal J. A review on fault detection and diagnosis techniques: basics and beyond. Artif Intell Rev 2021;54(5):3639–64. http://dx.doi.org/10.1007/s10462-020-09934-2.

[10] Qin SJ, Chiang LH. Advances and opportunities in machine learning for process data analytics. Comput Chem Eng 2019;126:465–73. http://dx.doi.org/10.1016/j.compchemeng.2019.04.003.

[11] Juez-Gil M, Saucedo-Dorantes JJ, Arnaiz-González Á, López-Nozal C, García-Osorio C, Lowe D. Early and extremely early multi-label fault diagnosis in induction motors. ISA Trans 2020;106:367–81. http://dx.doi.org/10.1016/j.isatra.2020.07.002.

[12] Ji C, Sun W. A review on data-driven process monitoring methods: Characterization and mining of industrial data. Processes 2022;10(2):335. http://dx.doi.org/10.3390/pr10020335.

[13] Venkatasubramanian V, Rengaswamy R, Yin K, Kavuri SN. A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. Comput Chem Eng 2003;27(3):293–311. http://dx.doi.org/10.1016/S0098-1354(02)00160-6.

[14] Venkatasubramanian V, Rengaswamy R, Kavuri SN. A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies. Comput Chem Eng 2003;27(3):313–26. http://dx.doi.org/10.1016/S0098-1354(02)00161-8.

[15] Venkatasubramanian V, Rengaswamy R, Kavuri SN, Yin K. A review of process fault detection and diagnosis: Part III: Process history based methods. Comput Chem Eng 2003;27(3):327–46. http://dx.doi.org/10.1016/S0098-1354(02)00162-X.

[16] Park Y-J, Fan S-KS, Hsu C-Y. A review on fault detection and process diagnostics in industrial processes. Processes 2020;8(9). http://dx.doi.org/10.3390/pr8091123.

[17] Tang M, Zhao Q, Wu H, Wang Z, Meng C, Wang Y. Review and perspectives of machine learning methods for wind turbine fault diagnosis. Front Energy Res 2021;9. http://dx.doi.org/10.3389/fenrg.2021.751066.

[18] Jiang L, Ge Z, Song Z. Semi-supervised fault classification based on dynamic Sparse Stacked auto-encoders model. Chemometr Intell Lab Syst 2017;168:72–83. http://dx.doi.org/10.1016/j.chemolab.2017.06.010.

[19] Wu X, Zhang Y, Cheng C, Peng Z. A hybrid classification autoencoder for semi-supervised fault diagnosis in rotating machinery. Mech Syst Signal Process 2021;149:107327. http://dx.doi.org/10.1016/j.ymssp.2020.107327.

[20] Zhou Z-H. A brief introduction to weakly supervised learning. Natl Sci Rev 2018;5(1):44–53. http://dx.doi.org/10.1093/nsr/nwx106.

[21] Chapelle O, Schölkopf B, Zien A. Semi-Supervised Learning. Introduction to Semi-Supervised Learning, Cambridge: MIT Press; 2006, p. 1–12. http://dx.doi.org/10.7551/mitpress/9780262033589.001.0001.

[22] Mowbray M, Vallerio M, Perez-Galvan C, Zhang D, Del Rio Chanona A, Navarro-Brull FJ. Industrial data science – a review of machine learning applications for chemical and process industries. React Chem Eng 2022;7:1471–509. http://dx.doi.org/10.1039/D1RE00541C.

[23] He Y, Song K, Dong H, Yan Y. Semi-supervised defect classification of steel surface based on multi-training and generative adversarial network. Opt Lasers Eng 2019;122:294–302. http://dx.doi.org/10.1016/j.optlaseng.2019.06.020.

[24] Triguero I, García S, Herrera F. Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study. Knowl Inf Syst 2015;42(2):245–84. http://dx.doi.org/10.1007/s10115-013-0706-y.

[25] Settles B. Active learning. Synthesis lectures on artificial intelligence and machine learning, Morgan & Claypool Publishers; 2012, http://dx.doi.org/10.2200/S00429ED1V01Y201207AIM018.

[26] Aggarwal CC, Kong X, Gu Q, Han J, Yu PS. Data classification. CRC Press; 2014, p. 571–605. http://dx.doi.org/10.1201/b17320.

[27] Jin Y, Qin C, Huang Y, Liu C. Actual bearing compound fault diagnosis based on active learning and decoupling attentional residual network. Measurement 2021;173:108500. http://dx.doi.org/10.1016/j.measurement.2020.108500.

[28] Liu J, Song C, Zhao J. Active learning based semi-supervised exponential discriminant analysis and its application for fault classification in industrial processes. Chemometr Intell Lab Syst 2018;180:42–53. http://dx.doi.org/10.1016/j.chemolab.2018.07.003.

[29] Yin L, Wang H, Fan W, Kou L, Lin T, Xiao Y. Incorporate active learning to semi-supervised industrial fault classification. J Process Control 2019;78:88–97. http://dx.doi.org/10.1016/j.jprocont.2019.04.008.

[30] Jian C, Yang K, Ao Y. Industrial fault diagnosis based on active learning and semi-supervised learning using small training set. Eng Appl Artif Intell 2021;104:104365. http://dx.doi.org/10.1016/j.engappai.2021.104365.

[31] Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. J Big Data 2016;3(1):9. http://dx.doi.org/10.1186/s40537-016-0043-6.

[32] Wu H, Zhao J. Fault detection and diagnosis based on transfer learning for multimode chemical processes. Comput Chem Eng 2020;135:106731. http://dx.doi.org/10.1016/j.compchemeng.2020.106731.

[33] Downs J, Vogel E. A plant-wide industrial process control problem. Comput Chem Eng 1993;17(3):245–55. http://dx.doi.org/10.1016/0098-1354(93)80018-I.

[34] Zhu X, Chen K, Anduv B, Jin X, Du Z. Transfer learning based methodology for migration and application of fault detection and diagnosis between building chillers for improving energy efficiency. Build Environ 2021;200:107957. http://dx.doi.org/10.1016/j.buildenv.2021.107957.

[35] Singh A, Nowak RD, Zhu X. Unlabeled data: Now it helps, now it doesn't. In: Koller D, Schuurmans D, Bengio Y, Bottou L, editors. Proceedings of the 21st international conference on neural information processing systems. NIPS '08, Red Hook, NY, USA: Curran Associates, Inc.; 2008, p. 1513–20. http://dx.doi.org/10.5555/2981780.2981969.

[36] Li Y-F, Liang D-M. Safe semi-supervised learning: a brief introduction. Front Comput Sci 2019;13(4):669–76. http://dx.doi.org/10.1007/s11704-019-8452-2.

[37] Jia X, Tian W, Li C, Yang X, Luo Z, Wang H. A dynamic active safe semi-supervised learning framework for fault identification in labeled expensive chemical processes. Processes 2020;8(1). http://dx.doi.org/10.3390/pr8010105.

[38] Zhao M, Li B, Qi J, Ding Y. Semi-supervised classification for rolling fault diagnosis via robust sparse and low-rank model. In: 2017 IEEE 15th international conference on industrial informatics (INDIN). 2017, p. 1062–7. http://dx.doi.org/10.1109/INDIN.2017.8104920.

[39] Gao Y, Yu D. Intelligent fault diagnosis for rolling bearings based on graph shift regularization with directed graphs. Adv Eng Inform 2021;47:101253. http://dx.doi.org/10.1016/j.aei.2021.101253.

[40] Wang X, Wang T, Ming A, Zhang W, Li A, Chu F. Semi-supervised hierarchical attribute representation learning via multi-layer matrix factorization for machinery fault diagnosis. Mech Mach Theory 2022;167:104445. http://dx.doi.org/10.1016/j.mechmachtheory.2021.104445.

[41] Chen X, Wang Z, Zhang Z, Jia L, Qin Y. A semi-supervised approach to bearing fault diagnosis under variable conditions towards imbalanced unlabeled data. Sensors 2018;18(7):2097. http://dx.doi.org/10.3390/s18072097.

[42] Zaman SMK, Liang X, Zhang L. Greedy-gradient max cut-based fault diagnosis for direct online induction motors. IEEE Access 2020;8:177851–62. http://dx.doi.org/10.1109/ACCESS.2020.3027322.

[43] Zaman SMK, Liang X, Li W. Fault diagnosis for variable frequency drive-fed induction motors using wavelet packet decomposition and greedy-gradient max-cut learning. IEEE Access 2021;9:65490–502. http://dx.doi.org/10.1109/ACCESS.2021.3076149.

[44] Zaman SMK, Liang X. An effective induction motor fault diagnosis approach using graph-based semi-supervised learning. IEEE Access 2021;9:7471–82. http://dx.doi.org/10.1109/ACCESS.2021.3049193.

[45] Zhao Y, Ball R, Mosesian J, de Palma J-F, Lehman B. Graph-based semi-supervised learning for fault detection and classification in solar photovoltaic arrays. IEEE Trans Power Electron 2015;30(5):2848–58. http://dx.doi.org/10.1109/TPEL.2014.2364203.

[46] Momeni H, Sadoogi N, Farrokhifar M, Gharibeh HF. Fault diagnosis in photovoltaic arrays using GBSSL method and proposing a fault correction system. IEEE Trans Ind Inf 2019;16(8):5300–8. http://dx.doi.org/10.1109/TII.2019.2908992.

[47] Qian M, Li Y-F. A weakly supervised learning-based oversampling framework for class-imbalanced fault diagnosis. IEEE Trans Reliab 2022;71(1):429–42. http://dx.doi.org/10.1109/TR.2021.3138448.

[48] Fang J, Yang F, Tong R, Yu Q, Dai X. Fault diagnosis of electric transformers based on infrared image processing and semi-supervised learning. Glob Energy Interconnect 2021;4(6):596–607. http://dx.doi.org/10.1016/j.gloei.2022.01.008.

[49] Chen S, Yang R, Zhong M. Graph-based semi-supervised random forest for rotating machinery gearbox fault diagnosis. Control Eng Pract 2021;117:104952. http://dx.doi.org/10.1016/j.conengprac.2021.104952.

[50] Hu C, Zhou Z, Wang B, Zheng W, He S. Tensor transfer learning for intelligence fault diagnosis of bearing with semisupervised partial label learning. J Sensors 2021;2021:6205890. http://dx.doi.org/10.1155/2021/6205890.

[51] Zhao M, Tian Z, Chow TWS. Fault diagnosis on wireless sensor network using the neighborhood kernel density estimation. Neural Comput Appl 2019;31(8):4019–30. http://dx.doi.org/10.1007/s00521-018-3342-3.

[52] Yarowsky D. Unsupervised word sense disambiguation rivaling supervised methods. In: Proceedings of the 33rd annual meeting on association for computational linguistics. ACL '95, USA: Association for Computational Linguistics; 1995, p. 189–96. http://dx.doi.org/10.3115/981658.981684.

[53] Zheng S, Zhao J. A self-adaptive temporal-spatial self-training algorithm for semisupervised fault diagnosis of industrial processes. IEEE Trans Ind Inf 2022;18(10):6700–11. http://dx.doi.org/10.1109/TII.2021.3120686.

[54] Zheng S, Zhao J. High-fidelity positive-unlabeled deep learning for semi-supervised fault detection of chemical processes. Process Saf Environ Prot 2022;165:191–204. http://dx.doi.org/10.1016/j.psep.2022.06.058.

[55] Zhang X, Su Z, Hu X, Han Y, Wang S. Semisupervised momentum prototype network for gearbox fault diagnosis under limited labeled samples. IEEE Trans Ind Inf 2022;18(9):6203–13. http://dx.doi.org/10.1109/TII.2022.3154486.

[56] Liu Q, Wang X, Yang B, Wang Z, Liu Y, Guan X. FSR-SSL: A fault sample rebalancing framework based on semi-supervised learning for PV fault diagnosis. IET Renew Power Gener 2022;16(12):2667–81. http://dx.doi.org/10.1049/rpg2.12458.

[57] Long J, Chen Y, Yang Z, Huang Y, Li C. A novel self-training semi-supervised deep learning approach for machinery fault diagnosis. Int J Prod Res 2022;1–14. http://dx.doi.org/10.1080/00207543.2022.2032860.

[58] Yan K, Zhong C, Ji Z, Huang J. Semi-supervised learning for early detection and diagnosis of various air handling unit faults. Energy Build 2018;181:75–83. http://dx.doi.org/10.1016/j.enbuild.2018.10.016.

[59] Nie X, Xie G. A two-stage semi-supervised learning framework for fault diagnosis of rotating machinery. IEEE Trans Instrum Meas 2021;70:1–12. http://dx.doi.org/10.1109/TIM.2021.3091489.

[60] Fan C, Liu Y, Liu X, Sun Y, Wang J. A study on semi-supervised learning in enhancing performance of AHU unseen fault detection with limited labeled data. Sustainable Cities Soc 2021;70:102874. http://dx.doi.org/10.1016/j.scs.2021.102874.

[61] Kuncheva LI. Combining pattern classifiers: methods and algorithms. John Wiley & Sons; 2004, http://dx.doi.org/10.1002/0471660264.

[62] Blum A, Mitchell T. Combining labeled and unlabeled data with co-training. In: Proceedings of the eleventh annual conference on computational learning theory. COLT' 98, New York, NY, USA: Association for Computing Machinery; 1998, p. 92–100. http://dx.doi.org/10.1145/279943.279962.

[63] Abdelgayed TS, Morsi WG, Sidhu TS. Fault detection and classification based on co-training of semisupervised machine learning. IEEE Trans Ind Electron 2018;65(2):1595–605. http://dx.doi.org/10.1109/TIE.2017.2726961.

[64] Zhou Z-H, Li M. Tri-training: exploiting unlabeled data using three classifiers. IEEE Trans Knowl Data Eng 2005;17(11):1529–41. http://dx.doi.org/10.1109/TKDE.2005.186.

[65] Jing D, Zhenya W. Incremental fault diagnosis: Exploiting unlabelled data with semi-supervised ensemble learning. In: 2017 13th IEEE international conference on electronic measurement instruments (ICEMI). 2017, p. 7–14. http://dx.doi.org/10.1109/ICEMI.2017.8265696.

[66] Jian C, Ao Y. Imbalanced fault diagnosis based on semi-supervised ensemble learning. J Intell Manuf 2022;1–16. http://dx.doi.org/10.1007/s10845-022-01985-2.

[67] Qin W-L, Zhang W-J, Wang Z-Y. Improvement of roller bearing diagnosis with unlabeled data using cut edge weight confidence based tritraining. Shock Vib 2016;2016:1646898. http://dx.doi.org/10.1155/2016/1646898.

[68] Liu W, Wang P, You Y. Ensemble-based semi-supervised learning for milling chatter detection. Machines 2022;10(11):1013. http://dx.doi.org/10.3390/machines10111013.

[69] Huang F, Zhang X, Qin G, Xie J, Peng J, Huang S, Long Z, Tang Y. Demagnetization fault diagnosis of permanent magnet synchronous motors using magnetic leakage signals. IEEE Trans Ind Inf 2023;19(4):6105–16. http://dx.doi.org/10.1109/TII.2022.3165283.

[70] Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. J Comput System Sci 1997;55(1):119–39. http://dx.doi.org/10.1006/jcss.1997.1504.

[71] Razavi-Far R, Hallaji E, Saif M, Rueda L. A hybrid scheme for fault diagnosis with partially labeled sets of observations. In: 2017 16th IEEE international conference on machine learning and applications (ICMLA). 2017, p. 61–7. http://dx.doi.org/10.1109/ICMLA.2017.0-177.

[72] Bennett KP, Demiriz A, Maclin R. Exploiting unlabeled data in ensemble methods. In: Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining. KDD '02, New York, NY, USA: Association for Computing Machinery; 2002, p. 289–96. http://dx.doi.org/10.1145/775047.775090.

[73] Razavi-Far R, Hallaji E, Farajzadeh-Zanjani M, Saif M. A semi-supervised diagnostic framework based on the surface estimation of faulty distributions. IEEE Trans Ind Inf 2019;15(3):1277–86. http://dx.doi.org/10.1109/TII.2018.2851961.

[74] Mallapragada PK, Jin R, Jain AK, Liu Y. SemiBoost: Boosting for semi-supervised learning. IEEE Trans Pattern Anal Mach Intell 2009;31(11):2000–14. http://dx.doi.org/10.1109/TPAMI.2008.235.

[75] Zhao X, Jia M, Liu Z. Semisupervised deep sparse auto-encoder with local and nonlocal information for intelligent fault diagnosis of rotating machinery. IEEE Trans Instrum Meas 2021;70:1–13. http://dx.doi.org/10.1109/TIM.2020.3016045.

[76] Luo S, Huang X, Wang Y, Luo R, Zhou Q. Transfer learning based on improved stacked autoencoder for bearing fault diagnosis. Knowl-Based Syst 2022;256:109846. http://dx.doi.org/10.1016/j.knosys.2022.109846.

[77] Ma Y, Shi H, Tan S, Tao Y, Song B. Consistency regularization auto-encoder network for semi-supervised process fault diagnosis. IEEE Trans Instrum Meas 2022;71:1–15. http://dx.doi.org/10.1109/TIM.2022.3184346.

[78] Kim S, Jo S-H, Kim W, Park J, Jeong J, Han Y, et al. A semi-supervised autoencoder with an auxiliary task (SAAT) for power transformer fault diagnosis using dissolved gas analysis. IEEE Access 2020;8:178295–310. http://dx.doi.org/10.1109/ACCESS.2020.3027830.

[79] Jiang L, Xuan J, Shi T. Feature extraction based on semi-supervised kernel Marginal Fisher analysis and its application in bearing fault diagnosis. Mech Syst Signal Process 2013;41(1):113–26. http://dx.doi.org/10.1016/j.ymssp.2013.05.017.

[80] Zeng X, Yin S-B, Guo Y, Lin J-R, Zhu J-G. A novel semi-supervised feature extraction method and its application in automotive assembly fault diagnosis based on vision sensor data. Sensors 2018;18(8):2545. http://dx.doi.org/10.3390/s18082545.

[81] Ge Z, Zhong S, Zhang Y. Semisupervised kernel learning for FDA model and its application for fault classification in industrial processes. IEEE Trans Ind Inf 2016;12(4):1403–11. http://dx.doi.org/10.1109/TII.2016.2571680.

[82] Liu J, Song C, Zhao J. Active learning based semi-supervised exponential discriminant analysis and its application for fault classification in industrial processes. Chemometr Intell Lab Syst 2018;180:42–53. http://dx.doi.org/10.1016/j.chemolab.2018.07.003.

[83] Zheng J, Wang H, Song Z, Ge Z. Ensemble semi-supervised Fisher discriminant analysis model for fault classification in industrial processes. ISA Trans 2019;92:109–17. http://dx.doi.org/10.1016/j.isatra.2019.02.021.

[84] Bi F, Li X, Lin J, Bi X, Ma T, Yang X, et al. Knock detection based on recursive variational mode decomposition and multilevel semi-supervised local Fisher discriminant analysis. IEEE Access 2019;7:122028–40. http://dx.doi.org/10.1109/ACCESS.2019.2937571.

[85] Jiang Q, Zhu Q, Wang B, Guo L. Nonlinear machine fault detection by semi-supervised Laplacian Eigenmaps. J Mech Sci Technol 2017;31:3697–703. http://dx.doi.org/10.1007/s12206-017-0712-1.

[86] Park D, Na J, Lee JM. Clustered manifold approximation and projection for semisupervised fault diagnosis and process monitoring. Ind Eng Chem Res 2021;60(26):9521–31.

[87] Luo J, Xu H, Su Z, Xiao H, Zheng K, Zhang Y. Fault diagnosis based on orthogonal semi-supervised LLTSA for feature extraction and Transductive SVM for fault identification. J Intell Fuzzy Systems 2018;34(6):3499–511. http://dx.doi.org/10.3233/JIFS-169529.

[88] Wang Z, Yao L, Cai Y, Zhang J. Mahalanobis semi-supervised mapping and beetle antennae search based support vector machine for wind turbine rolling bearings fault diagnosis. Renew Energy 2020;155:1312–27. http://dx.doi.org/10.1016/j.renene.2020.04.041.

[89] Xiao Y, Wang Q, Wang S, Chen W. Research on a multisource domain improved fault diagnosis method of the rotor system. IEEE Access 2022;10:85399–415. http://dx.doi.org/10.1109/ACCESS.2022.3197898.

[90] Hu Y, Baraldi P, Di Maio F, Zio E. A systematic semi-supervised self-adaptable fault diagnostics approach in an evolving environment. Mech Syst Signal Process 2017;88:413–27. http://dx.doi.org/10.1016/j.ymssp.2016.11.004.

[91] Tan X, Guo C, Wang K, Wan F. A novel two-stage Dissolved Gas Analysis fault diagnosis system based semi-supervised learning. High Voltage 2022;7(4):676–91. http://dx.doi.org/10.1049/hve2.12195.

[92] Yuan J, Zhao R, He T, Chen P, Wei K, Xing Z. Fault diagnosis of rotor based on Semi-supervised Multi-Graph Joint Embedding. ISA transactions 2022;131:516–32. http://dx.doi.org/10.1016/j.isatra.2022.05.006.

[93] Sun T, Yu G. A multi-stage semi-supervised improved deep embedded clustering (MS-SSIDEC) method for bearing fault diagnosis under the situation of insufficient labeled samples. 2021, CoRR abs/2109.13521.

[94] Li W, Zhang S, He G. Semisupervised distance-preserving self-organizing map for machine-defect detection and classification. IEEE Trans Instrum Meas 2013;62(5):869–79. http://dx.doi.org/10.1109/TIM.2013.2245180.

[95] Li C, Zhou J. Semi-supervised weighted kernel clustering based on gravitational search for fault diagnosis. ISA Trans 2014;53(5):1534–43. http://dx.doi.org/10.1016/j.isatra.2014.05.019, ICCA 2013.

[96] Lemos A, Caminhas W, Gomide F. Adaptive fault detection and diagnosis using an evolving fuzzy classifier. Inform Sci 2013;220:64–85. http://dx.doi.org/10.1016/j.ins.2011.08.030.

[97] Zhai L, Jia Q. Simultaneous fault detection and isolation using semi-supervised kernel nonnegative matrix factorization. Can J Chem Eng 2019;97(12):3025–34. http://dx.doi.org/10.1002/cjce.23580.

[98] Li X, Li X, Ma H. Deep representation clustering-based fault diagnosis method with unsupervised data applied to rotating machinery. Mech Syst Signal Process 2020;143:106825. http://dx.doi.org/10.1016/j.ymssp.2020.106825.

[99] Yu K, Lin TR, Ma H, Li X, Li X. A multi-stage semi-supervised learning approach for intelligent fault diagnosis of rolling bearing using data augmentation and metric learning. Mech Syst Signal Process 2021;146:107043. http://dx.doi.org/10.1016/j.ymssp.2020.107043.

[100] Cerrada M, Sánchez R-V, Cabrera D. A semi-supervised approach based on evolving clusters for discovering unknown abnormal condition patterns in gearboxes. J Intell Fuzzy Systems 2018;34(6):3581–93. http://dx.doi.org/10.3233/JIFS-169535.

[101] Tao X, Ren C, Li Q, Guo W, Liu R, He Q, et al. Bearing defect diagnosis based on semi-supervised kernel Local Fisher Discriminant Analysis using pseudo labels. ISA Trans 2021;110:394–412. http://dx.doi.org/10.1016/j.isatra.2020.10.033.

[102] Xiao W, Yang Y, Wang H, Li T, Xing H. Semi-supervised hierarchical clustering ensemble and its application. Neurocomputing 2016;173:1362–76. http://dx.doi.org/10.1016/j.neucom.2015.09.009.

[103] Ding Y, Zhuang J, Ding P, Jia M. Self-supervised pretraining via contrast learning for intelligent incipient fault detection of bearings. Reliab Eng Syst Saf 2022;218:108126. http://dx.doi.org/10.1016/j.ress.2021.108126.

[104] Liao Y, Huang R, Li J, Chen Z, Li W. Deep semisupervised domain generalization network for rotary machinery fault diagnosis under variable speed. IEEE Trans Instrum Meas 2020;69(10):8064–75. http://dx.doi.org/10.1109/TIM.2020.2992829.

[105] Bennett K, Demiriz A. Semi-supervised support vector machines. In: Kearns M, Solla S, Cohn D, editors. Advances in neural information processing systems. vol. 11, MIT Press; 1998, p. 368–74.

[106] Wang L, Tian H, Zhang H. Soft fault diagnosis of analog circuits based on semi-supervised support vector machine. Analog Integr Circuits Signal Process 2021;108(2):305–15. http://dx.doi.org/10.1007/s10470-021-01851-w.

[107] Mao W, Tian S, Fan J, Liang X, Safian A. Online detection of bearing incipient fault with semi-supervised architecture and deep feature representation. J Manuf Syst 2020;55:179–98. http://dx.doi.org/10.1016/j.jmsy.2020.03.005.

[108] Tang M, Hu J, Wu H, Wang Z. Wind turbine pitch system fault detection using ssODM-DSTA. Front Energy Res 2021;9:750983. http://dx.doi.org/10.3389/fenrg.2021.750983.

[109] Wu Y, Zhao R, Jin W, He T, Ma S, Shi M. Intelligent fault diagnosis of rolling bearings using a semi-supervised convolutional neural network. Appl Intell 2021;51(4):2144–60. http://dx.doi.org/10.1007/s10489-020-02006-6.

[110] Razavi-Far R, Hallaji E, Farajzadeh-Zanjani M, Saif M, Kia SH, Henao H, et al. Information fusion and semi-supervised deep learning scheme for diagnosing gear faults in induction machine systems. IEEE Trans Ind Electron 2019;66(8):6331–42. http://dx.doi.org/10.1109/TIE.2018.2873546.

[111] Chen S-Q, Yang G-J, Gao W, Guo M-F. Photovoltaic fault diagnosis via semisupervised ladder network with string voltage and current measures. IEEE J Photovolt 2021;11(1):219–31. http://dx.doi.org/10.1109/JPHOTOV.2020.3038335.

[112] Zhang S, Qiu T. Semi-supervised LSTM ladder autoencoder for chemical process fault diagnosis and localization. Chem Eng Sci 2022;251:117467. http://dx.doi.org/10.1016/j.ces.2022.117467.

[113] Tang Z, Bo L, Liu X, Wei D. A semi-supervised transferable LSTM with feature evaluation for fault diagnosis of rotating machinery. Appl Intell 2022;52(2):1703–17. http://dx.doi.org/10.1007/s10489-021-02504-1.

[114] Fu W, Jiang X, Tan C, Li B, Chen B. Rolling bearing fault diagnosis in limited data scenarios using feature enhanced generative adversarial networks. IEEE Sens J 2022;22(9):8749–59. http://dx.doi.org/10.1109/JSEN.2022.3160762.

[115] Yu K, Ma H, Lin TR, Li X. A consistency regularization based semi-supervised learning approach for intelligent fault diagnosis of rolling bearing. Measurement 2020;165:107987. http://dx.doi.org/10.1016/j.measurement.2020.107987.

[116] Zhang Y, Ren Z, Zhou S. An intelligent fault diagnosis for rolling bearing based on adversarial semi-supervised method. IEEE Access 2020;8:149868–77. http://dx.doi.org/10.1109/ACCESS.2020.3016314.

[117] Hu C, Wu J, Sun C, Yan R, Chen X. Interinstance and intratemporal self-supervised learning with few labeled data for fault diagnosis. IEEE Trans Ind Inf 2023;19(5):6502–12. http://dx.doi.org/10.1109/TII.2022.3183601.

[118] Toma RN, Piltan F, Kim J-M. A deep autoencoder-based convolution neural network framework for bearing fault classification in induction motors. Sensors 2021;21(24). http://dx.doi.org/10.3390/s21248453.

[119] Feng Y, Chen J, Zhang T, He S, Xu E, Zhou Z. Semi-supervised meta-learning networks with squeeze-and-excitation attention for few-shot fault diagnosis. ISA Trans 2022;120:383–401. http://dx.doi.org/10.1016/j.isatra.2021.03.013.

[120] Shim J, Cho S, Kum E, Jeong S. Adaptive fault detection framework for recipe transition in semiconductor manufacturing. Comput Ind Eng 2021;161:107632. http://dx.doi.org/10.1016/j.cie.2021.107632.

[121] Wang X, Feng H, Fan Y. Fault detection and classification for complex processes using semi-supervised learning algorithm. Chemometr Intell Lab Syst 2015;149:24–32. http://dx.doi.org/10.1016/j.chemolab.2015.10.019.

[122] Yuan J, Liu X. Semi-supervised learning and condition fusion for fault diagnosis. Mech Syst Signal Process 2013;38:615–27. http://dx.doi.org/10.1016/j.ymssp.2013.03.008.

[123] Zhao X, Jia M, Liu Z. Semisupervised graph convolution deep belief network for fault diagnosis of electormechanical system with limited labeled data. IEEE Trans Ind Inf 2021;17(8):5450–60. http://dx.doi.org/10.1109/TII.2020.3034189.

[124] Zhao X, Jia M, Ding P, Yang C, She D, Liu Z. Intelligent fault diagnosis of multichannel motor–rotor system based on multimanifold deep extreme learning machine. IEEE/ASME Trans Mechatronics 2020;25(5):2177–87. http://dx.doi.org/10.1109/TMECH.2020.3004589.

[125] Pu X, Li C. Online semisupervised broad learning system for industrial fault diagnosis. IEEE Trans Ind Inf 2021;17(10):6644–54. http://dx.doi.org/10.1109/TII.2020.3048990.

[126] Wang Y, Gao L, Gao Y, Li X. A new graph-based semi-supervised method for surface defect classification. Robot Comput-Integr Manuf 2021;68:102083. http://dx.doi.org/10.1016/j.rcim.2020.102083.

[127] Feng X, Jiao Y, Lv C, Zhou D. Label consistent semi-supervised non-negative matrix factorization for maintenance activities identification. Eng Appl Artif Intell 2016;52:161–7. http://dx.doi.org/10.1016/j.engappai.2016.02.016.

[128] Wang H, Wang J, Zhao Y, Liu Q, Liu M, Shen W. Few-shot learning for fault diagnosis with a dual graph neural network. IEEE Trans Ind Inf 2023;19(2):1559–68. http://dx.doi.org/10.1109/TII.2022.3205373.

[129] Xie Z, Chen J, Feng Y, He S. Semi-supervised multi-scale attention-aware graph convolution network for intelligent fault diagnosis of machine under extremely-limited labeled samples. J Manuf Syst 2022;64:561–77. http://dx.doi.org/10.1016/j.jmsy.2022.08.007.

[130] Yan S, Shao H, Xiao Y, Zhou J, Xu Y, Wan J. Semi-supervised fault diagnosis of machinery using LPS-DGAT under speed fluctuation and extremely low labeled rates. Adv Eng Inform 2022;53:101648. http://dx.doi.org/10.1016/j.aei.2022.101648.

[131] Gao Y, Chen M, Yu D. Semi-supervised graph convolutional network and its application in intelligent fault diagnosis of rotating machinery. Measurement 2021;186:110084. http://dx.doi.org/10.1016/j.measurement.2021.110084.

[132] Gao Y, Yu D. Fault diagnosis of rolling bearing based on Laplacian regularization. Appl Soft Comput 2021;111:107651.

[133] Zhang Y, Fu Y, Wang Z, Feng L. Fault detection based on modified kernel semi-supervised locally linear embedding. IEEE Access 2018;6:479–87. http://dx.doi.org/10.1109/ACCESS.2017.2767698.

[134] Fan S, Zhang Y. Application of independent component analysis with semi-supervised Laplacian regularization kernel density estimation. Can J Chem Eng 2018;96(6):1327–36. http://dx.doi.org/10.1002/cjce.23067.

[135] Zhao X, Li M, Xu J, Song G. An effective procedure exploiting unlabeled data to build monitoring system. Expert Syst Appl 2011;38(8):10199–204. http://dx.doi.org/10.1016/j.eswa.2011.02.078.

[136] Zhao K, Jiang H, Wu Z, Lu T. A novel transfer learning fault diagnosis method based on Manifold Embedded Distribution Alignment with a little labeled data. J Intell Manuf 2020;33:151–65. http://dx.doi.org/10.1007/s10845-020-01657-z.

[137] Lin J, Shao H, Min Z, Luo J, Xiao Y, Yan S, et al. Cross-domain fault diagnosis of bearing using improved semi-supervised meta-learning towards interference of out-of-distribution samples. Knowl-Based Syst 2022;252:109493. http://dx.doi.org/10.1016/j.knosys.2022.109493.

[138] Kavianpour M, Ramezani A, Beheshti MT. A class alignment method based on graph convolution neural network for bearing fault diagnosis in presence of missing data and changing working conditions. Measurement 2022;199:111536. http://dx.doi.org/10.1016/j.measurement.2022.111536.

[139] Mao W, Wei B, Xu X, Chen L, Wu T, Peng Z, et al. Fault diagnosis for power transformers through semi-supervised transfer learning. Sensors 2022;22(12). http://dx.doi.org/10.3390/s22124470.

[140] Li X, Zhang Z, Gao L, Wen L. A new semi-supervised fault diagnosis method via deep CORAL and transfer component analysis. IEEE Trans Emerg Top Comput Intell 2022;6(3):690–9. http://dx.doi.org/10.1109/TETCI.2021.3115666.

[141] Liu Y, Hu Z, Zhang Y. Symmetric positive definite manifold learning and its application in fault diagnosis. Neural Netw 2022;147:163–74. http://dx.doi.org/10.1016/j.neunet.2021.12.013.

[142] Gao Y, Gao L, Li X, Yan X. A semi-supervised convolutional neural network-based method for steel surface defect recognition. Robot Comput-Integr Manuf 2020;61:101825. http://dx.doi.org/10.1016/j.rcim.2019.101825.

[143] Yan Z, Huang C-C, Yao Y. Semi-supervised mixture discriminant monitoring for chemical batch processes. Chemometr Intell Lab Syst 2014;134:10–22. http://dx.doi.org/10.1016/j.chemolab.2014.03.002.

[144] Liang P, Deng C, Wu J, Yang Z-X, Zhu J, Zhang Z. Single and simultaneous fault diagnosis of gearbox via a semi-supervised and high-accuracy adversarial learning framework. Knowl-Based Syst 2020;198:105895. http://dx.doi.org/10.1016/j.knosys.2020.105895.

[145] Okaro IA, Jayasinghe S, Sutcliffe C, Black K, Paoletti P, Green PL. Automatic fault detection for laser powder-bed fusion using semi-supervised machine learning. Addit Manuf 2019;27:42–53. http://dx.doi.org/10.1016/j.addma.2019.01.006.

[146] Yan H-C, Zhou J-H, Pang CK. Gaussian mixture model using semisupervised learning for probabilistic fault diagnosis under new data categories. IEEE Trans Instrum Meas 2017;66(4):723–33. http://dx.doi.org/10.1109/TIM.2017.2654552.

[147] He R, Li X, Chen G, Chen G, Liu Y. Generative adversarial network-based semi-supervised learning for real-time risk warning of process industries. Expert Syst Appl 2020;150:113244. http://dx.doi.org/10.1016/j.eswa.2020.113244.

[148] Ko T, Kim H. Fault classification in high-dimensional complex processes using semi-supervised deep convolutional generative models. IEEE Trans Ind Inf 2020;16(4):2868–77. http://dx.doi.org/10.1109/TII.2019.2941486.

[149] Di H, Ke X, Peng Z, Zhou D. Surface defect classification of steels with a new semi-supervised learning method. Opt Lasers Eng 2019;117:40–8. http://dx.doi.org/10.1016/j.optlaseng.2019.01.011.

[150] Li B, Cheng F, Zhang X, Cui C, Cai W. A novel semi-supervised data-driven method for chiller fault diagnosis with unlabeled data. Appl Energy 2021;285:116459. http://dx.doi.org/10.1016/j.apenergy.2021.116459.

[151] Liang P, Deng C, Wu J, Li G, Yang Z, Wang Y. Intelligent fault diagnosis via semisupervised generative adversarial nets and wavelet transform. IEEE Trans Instrum Meas 2020;69(7):4659–71. http://dx.doi.org/10.1109/TIM.2019.2956613.

[152] Zhou K, Diehl E, Tang J. Deep convolutional generative adversarial network with semi-supervised learning enabled physics elucidation for extended gear fault diagnosis under data limitations. Mech Syst Signal Process 2023;185:109772. http://dx.doi.org/10.1016/j.ymssp.2022.109772.

[153] Li B, Cheng F, Cai H, Zhang X, Cai W. A semi-supervised approach to fault detection and diagnosis for building HVAC systems based on the modified generative adversarial network. Energy Build 2021;246:111044. http://dx.doi.org/10.1016/j.enbuild.2021.111044.

[154] Che C, Wang H, Lin R, Ni X. Semi-supervised multitask deep convolutional generative adversarial network for unbalanced fault diagnosis of rolling bearing. J Braz Soc Mech Sci Eng 2022;44(7):1–10. http://dx.doi.org/10.1007/s40430-022-03576-x.

[155] Wang Z, Xuan J, Shi T. A novel semi-supervised generative adversarial network based on the actor-critic algorithm for compound fault recognition. Neural Comput Appl 2022;1–19. http://dx.doi.org/10.1007/s00521-022-07011-z.

[156] Zhu X, Ye Y. Optimized fault diagnosis algorithm under GAN and CNN hybrid model. Wirel Commun Mob Comput 2022;2022. http://dx.doi.org/10.1155/2022/2433673.

[157] Xu M, Wang Y. An imbalanced fault diagnosis method for rolling bearing based on semi-supervised conditional generative adversarial network with spectral normalization. IEEE Access 2021;9:27736–47. http://dx.doi.org/10.1109/ACCESS.2021.3058334.

[158] Wang X, Liu F, Zhao D. Cross-machine fault diagnosis with semi-supervised discriminative adversarial domain adaptation. Sensors 2020;20(13). http://dx.doi.org/10.3390/s20133753.

[159] Pan T, Chen J, Xie J, Chang Y, Zhou Z. Intelligent fault identification for industrial automation system via multi-scale convolutional generative adversarial network with partially labeled samples. ISA Trans 2020;101:379–89. http://dx.doi.org/10.1016/j.isatra.2020.01.014.

[160] Xu J, Shi Y, Shi L, Ren Z, Lu Y. Intelligent deep adversarial network fault diagnosis method using semisupervised learning. Math Probl Eng 2020;2020:1–13. http://dx.doi.org/10.1155/2020/8503247.

[161] Zhong C, Yan K, Dai Y, Jin N, Lou B. Energy efficiency solutions for buildings: Automated fault diagnosis of air handling units using generative adversarial networks. Energies 2019;12(3). http://dx.doi.org/10.3390/en12030527.

[162] Kingma DP, Welling M. Auto-encoding variational Bayes. 2022, arXiv:1312.6114.

[163] Zhang K, Tang B, Qin Y, Deng L. Fault diagnosis of planetary gearbox using a novel semi-supervised method of multiple association layers networks. Mech Syst Signal Process 2019;131:243–60. http://dx.doi.org/10.1016/j.ymssp.2019.05.049.

[164] Wang C, Tao L, Ding Y, Lu C, Ma J. An adversarial model for electromechanical actuator fault diagnosis under nonideal data conditions. Neural Comput Appl 2022;34(8):5883–904. http://dx.doi.org/10.1007/s00521-021-06732-x.

[165] San Martin G, Droguett EL. Temporal variational auto-encoders for semi-supervised remaining useful life and fault diagnosis. IEEE Access 2022;10:55112–25. http://dx.doi.org/10.1109/ACCESS.2022.3174860.

[166] Zhang S, Ye F, Wang B, Habetler TG. Semi-supervised bearing fault diagnosis and classification using variational autoencoder-based deep generative models. IEEE Sens J 2021;21(5):6476–86. http://dx.doi.org/10.1109/JSEN.2020.3040696.

[167] Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative adversarial networks: An overview. IEEE Signal Process Mag 2018;35(1):53–65. http://dx.doi.org/10.1109/MSP.2017.2765202.

[168] Polikar R. Ensemble based systems in decision making. IEEE Circuits Syst Mag 2006;6(3):21–45. http://dx.doi.org/10.1109/MCAS.2006.1688199.

[169] Wan Z, Dong Y, Yu Z, Lv H, Lv Z. Semi-supervised support vector machine for digital twins based brain image fusion. Front Neurosci 2021;15. http://dx.doi.org/10.3389/fnins.2021.705523.

[170] Neupane D, Seok J. Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review. IEEE Access 2020;8:93155–78. http://dx.doi.org/10.1109/ACCESS.2020.2990528.

[171] Gousseau W, Antoni J, Girardin F, Griffaton J. Analysis of the rolling element bearing data set of the center for intelligent maintenance systems of the university of cincinnati. In: CM2016. Charenton, France; 2016, p. 1–13, URL https://hal.archives-ouvertes.fr/hal-01715193.

[172] Russell EL, Chiang LH, Braatz RD. Data-driven methods for fault detection and diagnosis in chemical processes. London: Springer London; 2000, p. 99–108. http://dx.doi.org/10.1007/978-1-4471-0409-4_8.

[173] Barnes N. Publish your computer code: it is good enough. Nature 2010;467(7317):753. http://dx.doi.org/10.1038/467753a.

[174] Schapire RE. The boosting approach to machine learning: An overview. Nonlinear Estim Classif 2003;149–71.

[175] Mayr A, Binder H, Gefeller O, Schmid M. The evolution of boosting algorithms. Methods Inform Med 2014;53(06):419–27. http://dx.doi.org/10.3414/ME13-01-0122.

[176] Yang X, Song Z, King I, Xu Z. A survey on deep semi-supervised learning. IEEE Trans Knowl Data Eng 2022;1–20. http://dx.doi.org/10.1109/TKDE.2022.3220219.

[177] Feng Y, Chen J, Xie J, Zhang T, Lv H, Pan T. Meta-learning as a promising approach for few-shot cross-domain fault diagnosis: Algorithms, applications, and prospects. Knowl-Based Syst 2022;235:107646. http://dx.doi.org/10.1016/j.knosys.2021.107646.

[178] Jiang W, Zhang Z, Li F, Zhang L, Zhao M, Jin X. Joint label consistent dictionary learning and adaptive label prediction for semisupervised machine fault classification. IEEE Trans Ind Inf 2016;12(1):248–56. http://dx.doi.org/10.1109/TII.2015.2496272.

[179] Tang Y, Zhang X, Zhai Y, Qin G, Song D, Huang S, et al. Rotating machine systems fault diagnosis using semisupervised conditional random field-based graph attention network. IEEE Trans Instrum Meas 2021;70:1–10. http://dx.doi.org/10.1109/TIM.2021.3091212.

[180] Ruan H, Wang Y, Li X, Qin Y, Tang B, Wang P. A relation-based semisupervised method for gearbox fault diagnosis with limited labeled samples. IEEE Trans Instrum Meas 2021;70:1–13. http://dx.doi.org/10.1109/TIM.2021.3052010.

[181] Pang S, Yang X. A cross-domain stacked denoising autoencoders for rotating machinery fault diagnosis under different working conditions. IEEE Access 2019;7:77277–92. http://dx.doi.org/10.1109/ACCESS.2019.2919535.

[182] Lv Z, Tang B, Zhou Y, Zhou C. A novel method for mechanical fault diagnosis based on variational mode decomposition and multikernel support vector machine. Shock Vib 2016;1–11. http://dx.doi.org/10.1155/2016/3196465.