

# Technical Language Supervision for Intelligent Fault Diagnosis in Process Industry

Karl Löwenmark<sup>1</sup>, Cees Taal<sup>2</sup>, Stephan Schnabel<sup>3</sup>, Marcus Liwicki<sup>4</sup> and Fredrik Sandin<sup>5</sup>

<sup>1,4,5</sup> *Embedded Intelligent Systems Laboratory (EISLAB), Luleå University of Technology, 971 87 Luleå, Sweden*

*karl.ekstrom@ltu.se*

*marcus.liwicki@ltu.se*

*fredrik.sandin@ltu.se*

<sup>2</sup> *SKF Research & Technology Development, Meidoornkade 14, 3992 AE Houten, P.O. Box 2350, 3430 DT Nieuwegein, The Netherlands*

*cees.taal@skf.com*

<sup>3</sup> *SKF Condition Monitoring Center Luleå AB, 977 75 Luleå, Sweden*

*stephan.schnabel@haw-landshut.de*

## ABSTRACT

In the process industry, condition monitoring systems with automated fault diagnosis methods assist human experts and thereby improve maintenance efficiency, process sustainability, and workplace safety. Improving the automated fault diagnosis methods using data and machine learning-based models is a central aspect of intelligent fault diagnosis (IFD). A major challenge in IFD is to develop realistic datasets with accurate labels needed to train and validate models, and to transfer models trained with labeled lab data to heterogeneous process industry environments. However, fault descriptions and work-orders written by domain experts are increasingly digitised in modern condition monitoring systems, for example in the context of rotating equipment monitoring. Thus, domain-specific knowledge about fault characteristics and severities exists as technical language annotations in industrial datasets. Furthermore, recent advances in natural language processing enable weakly supervised model optimisation using natural language annotations, most notably in the form of *natural language supervision* (NLS). This creates a timely opportunity to develop *technical language supervision* (TLS) solutions for IFD systems grounded in industrial data, for example as a complement to pre-training with lab data to address problems like overfitting and inaccurate out-of-sample generalisation. We surveyed the literature and identify a considerable improvement in the maturity of NLS over the last

two years, facilitating applications beyond natural language; a rapid development of weak supervision methods; and transfer learning as a current trend in IFD which can benefit from these developments. Finally we describe a general framework for TLS and implement a TLS case study based on SentenceBERT and contrastive learning based zero-shot inference on annotated industry data.

## 1. INTRODUCTION

Condition-monitoring (CM) based fault diagnosis of rotating machinery (Carden & Fanning, 2004; A. K. Jardine et al., 2006) is widely used in industry to optimise equipment availability, uniformity of product characteristics and safety in the work environment, and to minimise production losses and material waste. In process industry, this typically requires human expert analysts with years of training and detailed knowledge about the operational states, functional roles and contexts of the machines being monitored. Due to growing demands on production efficiency and the vast amounts of data consequently generated in modern CM systems, automated fault diagnosis systems (Kothamasu et al., 2006) are required to assist human analysis through alarms and policy recommendation. Important tasks for the automated system are fault detection and classification to generate alarms and filter data, and fault severity estimation to predict remaining useful life and recommend policy options. Existing automated systems are mainly based on expert systems (Nan et al., 2008), with a knowledge-base derived from physical properties of analysed components, and a rule-based inference engine with local thresholds set by experts (SKF, 2022). In the

Karl Löwenmark et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<https://doi.org/10.36001/IJPHM.2022.v13i2.3137>

Table 1. Fault Diagnosis Tasks

Task	Question addressed	Output	Added value	Automated?
Detection	Is there a fault present?	Yes/No	Alert human analysis	Yes
Classification	What type of fault?	Class	Guide human analysis	Partially
Severity	How severe is the fault?	Magnitude	Motivate maintenance	No
RUL	Time until maintenance is needed?	Risk vs Time	Maintenance planning	No
Root Cause	What caused the fault?	Description	Preventive policies	No

case of vibration measurements of rotating machinery, signal processing and kinematics based condition indicators are commonly used as knowledge-bases (A. Jardine et al., 2006; Rai & Upadhyay, 2016; Randall & Antoni, 2011). Intelligent fault diagnosis (IFD) (Lei et al., 2020) has been proposed to enhance the automated systems by inferring fault characteristics directly from process or lab data through learning based methods. Improving existing models is vital to meet the increasing demands on CM systems to improve production and equipment life cycle efficiency in process industry (ProcessIT, 2018; Shin & Jun, 2015), and the machine CM market is estimated at \$2.6 billion with a compound annual growth rate estimation of 7.1%<sup>1</sup>. For example, improved IFD algorithms can contribute to: reducing the number of unnecessary interventions; facilitating remanufacturing of components (A. SKF & Kommunikation, 2020); optimising maintenance schedules; and enabling analysts to focus on qualified preventive tasks.

However, it is difficult to develop realistic datasets with accurate labels needed to train and validate IFD models, and such data are expected to generalise poorly between process industry plants due to their heterogeneous nature. Recent innovations in natural language processing offer a timely opportunity to address this challenge with methods used in *natural language supervision* (NLS) (F. Chen et al., 2022a) using digitalised technical language fault descriptions and work-orders available in many process industry datasets. Processing technical language poses unique challenges different from natural language, promoting the need for *Technical Language Processing* (TLP) and a technical version of NLS in *Technical Language Supervision* (TLS). Therefore we survey the state of the art in IFD, NLS and TLP, and discuss how TLS can be applied to IFD in a process industry context.

### 1.1. Background

Fault Diagnosis (FD) deals with the mapping of measured signal features to component conditions. The most basic condition is whether a fault is present or not, but more complex estimations such as fault class, fault severity, remaining useful life (RUL) and root cause analysis (RCA) can also be required. Table 1 describes these five major subtasks of FD, ordered in rising complexity based on interviews with condition

monitoring experts from process industry. Fault detection and classification are tasks that are frequently automated in process industry through signal processing (Kothamasu et al., 2006; Nan et al., 2008; SKF, 2022), and for example model-based thresholding. Fault severity estimation, a vital tool in maintenance decisions, is next in line to be automated, but is challenging due to nonlinear relationships between signal features and fault evolution (Cerrada et al., 2018). RUL depends on the evolution of estimated fault severity over time, and predicts the remaining time until a fault is so severe that a component is no longer useful (Lei et al., 2018; D. Wang et al., 2017). RCA is a complex task that may be challenging to automate, but will indirectly be improved if simpler tasks are automated and human experts can invest more time in preventive policies.

The upper part of Figure 1 illustrates an example of a typical FD system (labeled Pipeline 1) implemented in process industry, see for instance (SKF, 2022; PdM, 2021; Cahill, 2021). The system requires no fault history data to learn from, but requires process information and kinematic models for the extraction of condition indicators (Sharma & Parey, 2016). Faults are detected and classified using signal processing (A. K. Jardine et al., 2006), for instance root mean square, peak-to-peak and time synchronous average in the time domain (chung Fu, 2011); spectral density, enveloping and Hilbert transform in the frequency domain; and dictionaries, wavelets and the Wigner-Ville distribution in the time-frequency domain; as well as kinematics based condition indicators, for instance the frequency intensity in the ball pass frequency of the outer race in ballpoint bearings. The decomposed signal is then analysed with typically simple rules based on indicator magnitude defined by experienced analysts. Once a fault is detected by the model, a human analyst is alerted for in-depth diagnosis. The analyst decides whether to further investigate alarms or not, describes eventual faults in the form of natural-language annotations and makes work orders. Thus, the automated FD model acts like a filter between the massive amount of sensor data that is constantly generated, and the accurate but resource-constrained analysis of human experts. Based on cases from two industry collaborations with major process industry actors in Northern Sweden<sup>2</sup>, analysts monitor around 5000 alarms per analyst per year, after filtering, where at most 20% of generated alarms

<sup>1</sup><https://www.marketsandmarkets.com/Market-Reports/machine-health-monitoring-market-29627363.html>

<sup>2</sup>Smurfit Kappa, 700 000 tonnes of Kraftliner per year, and SCA Munksund, 400 000 tonnes of Kraftliner per year

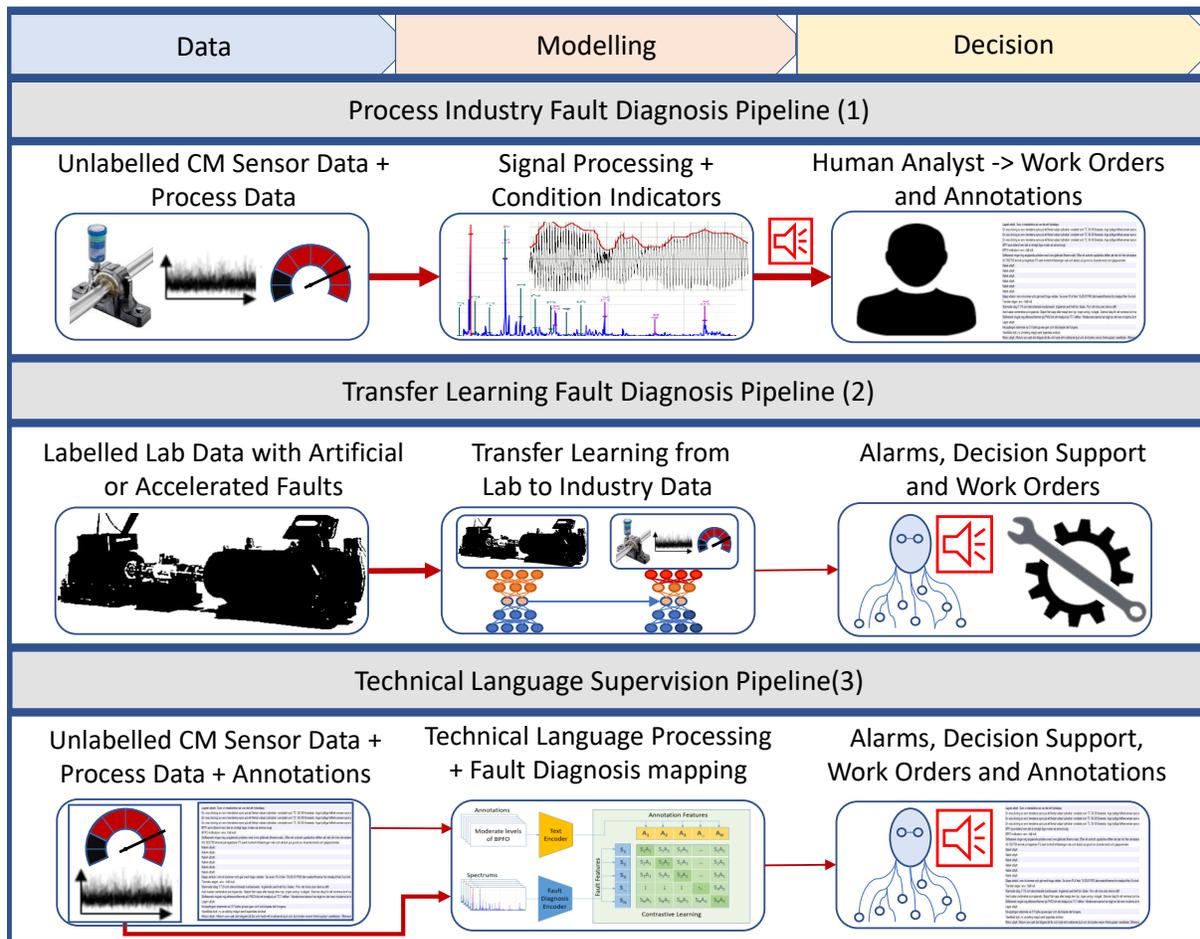


Figure 1. An overview of a typical process industry fault diagnosis pipeline (1), possible transfer learning IFD pipeline additions (2), and our suggested natural language supervision pipeline (3). Both (2) and (3) can provide considerable contributions to (1), with the strongest contributions coming from both pipelines implemented in symbiosis.

point to component faults and the rest are due to temporary or constant signal malfunctions.

With improved automated FD, analysts could focus on more advanced fault diagnosis tasks beyond the current capabilities of IFD. Considerable research has been invested in automated FD, and many learning-based methods have shown promising results on test datasets (R. Liu et al., 2018; Stetco et al., 2019; Hoang & Kang, 2019). However, the accurate deep learning models used in many IFD publications require vast amounts of training data in the form of labelled datasets, sets that typically do not exist in process industry cases (Khan & Yairi, 2018). Instead, training and test datasets are created in lab environments with artificial or accelerated fault development, such as the Case Western Reserve University bearing dataset (*Case Western Reserve University Bearing Data Center Website*, n.d.), the Intelligent Maintenance System (IMS) by the University of Cincinnati dataset (*NASA prognostic data repository*, n.d.), and the Machinery Failure Prevention Technology (MFPT) dataset (*Condition Based Maintenance Fault Database*

*for Testing of Diagnostic and Prognostics Algorithms*, n.d.), but typically generalise poorly to heterogeneous environments (Smith & Randall, 2015; S. Zhang et al., 2019) such as process industries. Thus, despite the maturity of IFD methods in terms of literature, supervised IFD lacks wide-spread implementation in industry.

Industry datasets suitable for IFD can in some cases potentially be created, but it is difficult and costly to define high-quality labels that are accurately connected to relevant data. Therefore, transfer learning (Schwendemann et al., 2021), illustrated in Pipeline 2 in Figure 1, has become an increasingly popular approach to develop IFD methods without requiring a large labelled dataset in the target domain (Lei et al., 2020). Ideally, a model could be developed/trained with data from a lab environment, then transferred to similar components in an industrial environment. However, this remains a challenging goal due to differences between developing faults, heterogeneous environments, varying sensor and signal-to-noise conditions, and complex coupling of signal components. Thus,

a method for the extraction of labels for industry data would be valuable and can facilitate implementations of current IFD models, as well as transfer learning by providing access to labels in the target domain.

While labels are lacking in realistic CM datasets, technical language fault descriptions are written by analysts when documenting and monitoring the development of for example bearing faults over long periods of time (several months). Thus, the text-annotations produced as outputs of Pipeline 1 in Figure 1 contain valuable albeit noisy information about fault development characteristics and severities. This motivates the question, can such domain-specific annotations and related knowledge be used for training and fine-tuning of IFD methods as a substitute for regular labels?

Language has been used to train machine learning models for image recognition and object detection through recent breakthroughs in natural language supervision (Radford et al., 2021a; Ramesh et al., 2021). Can a similar approach be used to train IFD models on industry data using annotations and work orders as zero-shot labels?

## 1.2. Contribution

We propose the usage of TLS on technical language fault descriptions to overcome the lack of labels in industry CM datasets. TLS is grounded in three fields, IFD, TLP and NLS, and we briefly survey all three to motivate the purpose and benefits of TLS. Potential TLS contributions supervision are improved support for human analysts and automation of simpler tasks by augmenting the label domain for transfer learning or zero-shot learning.

Pipeline 3 in Figure 1 illustrates the concept of a technical language supervision framework for process industry data. Unlabelled CM sensor data and process data are used to extract features through methods already used in IFD models, and the features are mapped to annotation embeddings. In the implementation stage, an unannotated signal is thereby mapped to the closest language fault queries in the joint embedding space, and with a sufficiently good model and well chosen queries, the fault class and severity can be estimated and described. Besides alarms and work orders, a language based model could also retrieve spectra from queries and generate new annotations and descriptions of detected faults. We implement a TLS model based on process industry signals and annotations, and show an example of spectrum retrieval from free form queries, as well as zero-shot fault classification of spectra.

## 1.3. Research Trends

We also surveyed the fault diagnosis literature and recent publications on language-based learning in the context of natural language supervision and image captioning to identify the

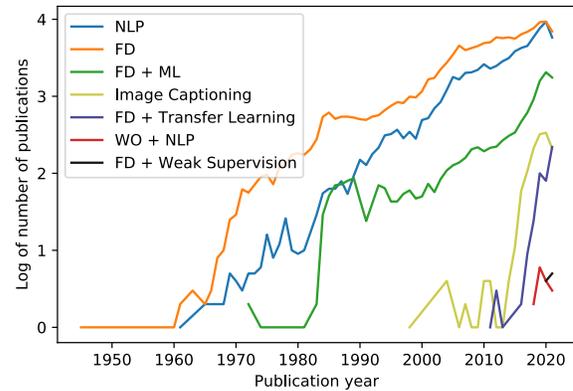


Figure 2. Trends of publications between 1967 and 2020, obtained through Scopus queries looking for publications with the targeted keywords in the article title, the abstract or the keywords. For instance, a query for fault diagnosis related keywords and transfer learning is designed as follows: ("condition monitoring" OR "fault diagnosis" OR "fault classification" OR "fault detection") AND "transfer learning". The annual number of articles about the application of machine learning (ML) to condition monitoring (CM) and fault diagnosis (FD) increases exponentially. That is also the case for the annual number of natural language processing (NLP) articles, which now equates the total annual number of FD-related articles. A total of 15 articles that use NLP on work orders (WO) were found, but no implementations of natural language supervision on fault diagnosis problems were identified. Weak supervision, or weakly supervised learning, is also not yet commonly used, with 4 articles in 2020 and 5 articles so far in 2021.

trends of publications that combine these concepts. Figure 2 shows the number of published articles per year according to Scopus for search queries containing keywords related to fault diagnosis and machine learning. We present the publication trends of natural language processing (NLP), fault diagnosis (FD), fault diagnosis with machine learning (FD + ML), fault diagnosis with transfer learning (FD + Transfer Learning), image captioning, work orders with natural language processing (WO + NLP), and finally fault diagnosis with weak supervision. For fault diagnosis, a query including "condition monitoring" OR "fault diagnosis" OR "fault detection" OR "fault classification" was used. For machine learning, "machine learning" OR "data driven" OR "deep learning" OR "artificial intelligence" were used. The queries "transfer learning", "work order", "natural language processing" and "image captioning" were used explicitly as is. Weak supervision was queried as "weak supervision" OR "weakly supervised".

The trends show that machine learning is increasingly applied in the FD literature, and that transfer learning has become increasingly popular, going from 2 publications in 2016 to 178 publications in 2020. NLP is a rapidly evolving field of research, with significant practical advancements in the last

Table 2. Fault Diagnosis Model Frameworks

Framework	Data requirements	Challenges
Unsupervised Learning	CM data	Applications beyond fault detection
Supervised Learning	Labelled CM dataset	Lack of labelled industry data
Transfer Learning	Labelled lab dataset, CM data	Lab features different from industry features
Weak Supervision	Weakly Labelled CM dataset	Still requires labels
Language Based	CM dataset, CM annotations	Not yet applied in IFD

decade. This is also reflected in the swift growth of image captioning publications starting in 2015, increasing from 11 to 322 publications in four years. 15 publications that use natural language processing with work orders were found, but NLP was employed for information retrieval, and no publications combining natural language supervision with IFD were found. Weak supervision only appeared nine times (with one valid article scheduled for 2022 not counted) in our queries, but notably three articles were cited more than ten times; X. Li, Zhang, et al. (2020) with 64 and 30 (X. Li, Li, & Ma, 2020) citations, and Yu, Fu, et al. (2021) with 12, showing that the interest far outweighs the current publication number. Articles citing weak supervision articles were mainly focused on transfer learning, but we predict an increase in direct mentions of weak supervision methods.

#### 1.4. Outline of article

In Section II, we describe the application of FD in process industry, which is subject to constraints related to the high cost of unplanned stops that can affect the whole production process. Five principal FD tasks are described, and the related methods and algorithms used for automated FD are also presented. In section III we briefly review natural language supervision and related fields such as image captioning, and discuss how natural language can be integrated in an IFD framework. Section IV describes a case study implementation of a TLS solution for IFD based on theories from section III, using process industry data for training and illustrations of model performance. We focus on rotating machinery in process industry, but in principle the framework of technical language supervision is expected to generalise to other fault diagnosis applications where fault descriptions are also present.

## 2. DEEP LEARNING IN INTELLIGENT FAULT DIAGNOSIS

Table 2 summarises different data-driven methods used for IFD, besides the kinematic rule-based method already discussed in the background. The methods are ordered roughly by maturity and data requirements. Unsupervised learning applies directly to unlabelled CM data, and it is partially implemented in process industries (SKF, n.d.; Monitron, n.d.; Simon, n.d.; Emerson, 2021). Supervised learning requires a labelled dataset in the application environment, and is widely investigated in the literature (Yin et al., 2014; Khan & Yairi,

2018; R. Liu et al., 2018; Helbing & Ritter, 2018; Stetco et al., 2019; Zhang et al., 2020), but not in process industry. Transfer learning requires a labelled dataset for pre-training, and data from the application environment, ideally labelled, for fine-tuning. The number of articles on transfer learning has increased rapidly in the last decade, but although transfer between lab environments show great results, we find no articles that apply transfer learning methods directly on process industry data. Finally, natural language supervision based learning only requires unlabelled CM data with associated annotations, but this method remains to be adapted and investigated for fault diagnosis tasks. The first mentions of natural language processing for in an IFD context was 2020 in the name of "technical language processing", though natural language supervision is yet to be introduced to IFD.

#### 2.1. Unsupervised Learning

Unsupervised learning, i.e learning patterns without labels, is connected to the modelling module of Figure 1 and is primarily used for *clustering, encoding, feature extraction* and *anomaly detection fault detection* (Lei et al., 2016). Models commonly used for clustering are k-means, Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE). Auto-Encoders and variational auto-encoders (Jiang et al., 2018; Haidong et al., 2018) and Dictionary Learning (Papayan et al., 2018; H. Liu et al., 2011) are commonly used for Encodings and Anomaly Detection. Virtually all models can be used to reduce dimensionality and extract features depending on the data, with PCA and t-SNE being more direct dimensionality reductions and auto-encoders serving as a more complex reconstruction model, often with encoders/decoders based on convolutions, recurrence or transformers.

Clustering, encodings and feature extraction can be valuable ways of understanding, simplifying or visualising data. A CM dataset with healthy and unhealthy data can with the right methods and data be divisible in to two clusters, which can then be manually labelled healthy and unhealthy, thus detecting faults (Yiakopoulos et al., 2011). Likewise, encodings or extracted features can serve as values in a simple rule-based system for fault detection or classification, and extracted features in particular can give valuable insight in feature importance. Regardless, the lack of a supervision signal necessitates a human in the last step to validate or assign meaning

to clusters, encodings or features, before the model is ready to automatically detect faults. Anomaly detection can work more autonomously by learning the healthy state of a signal, then classifying deviations from this state as detected faults (del Campo & Sandin, 2017) or into fault classes (C. Lu et al., 2017).

However, healthy states that lack sufficient presence in the training set has a risk of being classified as unhealthy at deployment, and unhealthy states that are present during training might be considered healthy, which is difficult to detect due to the lack of labels in the dataset. Furthermore, deviations might occur due to healthy states, or in directions relatively orthogonal to previous deviations. Such issues fall within the scope of *zero-shot learning*, wherein a model is required to observe and predict samples from a previously unseen class or distribution. (T. Zhang et al., 2021). For zero-shot learning to work, there has to be a distinct characteristic of faults and healthy states that is true for previously unseen faults or healthy states, which can be leveraged to assign these distributions to the correct class. NLS is sometimes discussed in the scope of zero-shot learning, and zero-shot learning techniques are often used in NLS. Likewise, zero-shot learning can be used to augment supervised learning methods beyond classes present in the supervision signal, but it is best described under the umbrella term of unsupervised learning or through the lens of weak supervision, as discussed in Section 2.4.

## 2.2. Supervised Learning

Supervised Learning can be employed for any FD task, as long as sufficient data and good labels are present. Transfer Learning, Weak Supervision and Language Supervision are all arguably subgroups of supervised learning explicitly designed to circumvent the limitation of requiring good labels. Architectures used in supervised learning are thus also employed in its derivatives, though with different learning procedures, just as how for instance auto-encoders from unsupervised learning can be used together with an output layer in a supervised paradigm.

Supervised learning architectures used in IFD range from shallow models such as tree-based models, e.g. random forest (D. Zhang et al., 2018); support vector machines (Yin et al., 2014; Qin, 2012); probabilistic models such as Bayesian statistics (Stief et al., 2019; H. Zhang et al., 2018); and deep architectures such as fully connected feed forward deep neural networks (F. Jia et al., 2016); (variational) auto-encoders with classification layers (Yan et al., 2021; Haidong et al., 2018); convolutional neural networks (F. Jia et al., 2018; Pan et al., 2018), commonly used in image analysis; recurrent neural networks (H. Liu et al., 2018; Qiao et al., 2020; X. Chen et al., 2021), commonly used in language analysis but applicable on sequential data in general. Importantly, supervised learning

has been employed for fault severity estimation (Cerrada et al., 2018) and RUL prediction (Babu et al., 2016; X. Li et al., 2018; Ben Ali et al., 2015; Guo et al., 2017; Lei et al., 2018; D. Wang et al., 2017).

Labelling industry datasets for supervised learning can facilitate implementations in that industry environment, but the labelling process is costly, and requires analyst efforts. Furthermore, some faults have stochastic features, for example due to the varying nature of the source geometry or signal transfer function, and are thus difficult to generalise with supervised classifiers. In general, faults are undesirable and therefore relatively scarce in industrial datasets, but are required in training datasets for supervised learning. Consequently, producing a labelled industry dataset for supervised learning would require considerable resources and potentially occupy analyst time necessary for condition monitoring. Therefore, fault classification models described in the literature are typically trained on labelled data from lab environments, where faults are generally either artificially induced or provoked through intense loads, as it might take several years until faults develop naturally. The development of the fault is then accelerated by e.g high loads or starved lubrication, which increase fault development per revolution, and high speeds to increase revolutions per minute (RPM). High RPM also produce higher signal-to-noise ratios as some noise is stationary and fault features increase more in magnitude than noise features.

Ideally, a model supervised on a component in a lab environment would then be deployable in an industry environment, but there are two issues that makes this difficult. Firstly, artificial or accelerated fault developments result in fault characteristics that are different compared to faults in industry environments. Therefore, the decision boundaries do not necessarily generalise well from lab to industry environments, and the feature space can differ due to different fault development processes. Secondly, signals generated in a lab setting differ greatly from signals in an industry environment where a component is connected to several other components in a larger system, and signal components are combined and masked by noise. The signal to noise-ratio will consequently be lower in the industry environment, and the coupling with surrounding components can shift the true feature space as well. Thus, direct supervised learning works best in the environment where it has been trained, and generalisation can be difficult unless labels are preserved in the target space.

## 2.3. Transfer Learning

Recently, the research focus in IFD has shifted to include methods to overcome the lack of labels in industry datasets such as transfer learning and weak supervision.

Transfer learning seeks to develop methods for training of a model in one environment, then fine-tuning the feature space

and decision boundaries to suit implementation in another environment (C. Li et al., 2020). In situations with sparse data optimisation limits, transfer learning can use domains with rich data, such as lab datasets, to infer necessary knowledge (Q. Zhang et al., 2021). The research on transfer learning in fault diagnosis applications has increased rapidly over the last few years, with many successful transfers between different lab datasets (Lei et al., 2020). As models improve, transfer learning can enable broader implementation of these models in process industry with a lower demand for labeled instances compared to supervised learning (Cao et al., 2018a), while solving the same tasks.

Methods used in transfer learning vary; many publications use transferrable convolutional neural networks (Cao et al., 2018b; Shao et al., 2019; B. Yang et al., 2019; Guo et al., 2019; Zhong et al., 2019; Xu et al., 2020; T. Han et al., 2020; Z. He et al., 2020; Wen et al., 2020; Z. Chen et al., 2020; Shao et al., 2021), occasionally employed with adversarial networks (Q. Wang et al., 2019; T. Han et al., 2019; X. Li et al., 2020); some use recurrent neural networks (A. Zhang et al., 2018; An et al., 2019; Zhao et al., 2020); auto-encoders are also used (Wen et al., 2019), and recently weak supervision (Li et al., 2020) and digital twin-based transfer learning (Xu et al., 2019) have been successfully implemented.

Transferring knowledge from one environment to another adds an additional benefit to symbol-feature relation graphs besides illustrating the process of the reasoning module. Humans learn concepts in a highly transferable manner, and it is for instance highly feasible that an experienced analyst could diagnose faults in a previously unseen environment with good accuracy, while a learning based model would certainly fail at adapting unless optimised through transfer learning. The underlying concepts of fault developments are likely the same in both environments, which is what humans use to generalise knowledge. Optimizing not only direct mappings, but symbol-feature relation graphs as well, can thus create models with stronger generalisability by mimicking human knowledge (Y. Li et al., 2020).

## 2.4. Weak Supervision

Weak supervision is an umbrella term for a set of methods developed to perform supervised tasks on data where labels are insufficient for regular supervised learning (Z.-H. Zhou, 2017). It can work in conjunction with transfer learning to enhance fine-tuning on the target dataset, or stand-alone to facilitate direct optimisation in the target environment. Table 3 illustrates three major ways in which labels can be insufficient, the cause, and proposed methods to amend the issue.

### 2.4.1. Incomplete supervision

*Incomplete labels* are characterised by a dataset where most data points are unlabelled. In a CM dataset, faults that have

not been discovered yet are a cause for incompleteness, as this prevents the assumption that all unlabelled data is healthy data.

The main strategy for dealing with incomplete datasets is called *semi-supervised learning* (van Engelen & Hoos, 2020; Zhai et al., 2019; Jian et al., 2021), which aims to create clusters of features that correspond to the available labels, and to estimate the probability that an unseen feature belongs to one of the identified clusters. Semi-supervised learning has been employed in IFD settings on lab datasets with partial (Razavi-Far et al., 2019) or limited labels (Yu, Lin, et al., 2021). By implementing semi-supervised learning on a CM dataset with natural language supervision, it is possible to include all time series data for a prediction, where particularly noisy samples would be less likely to affect the model optimisation process, as they are likely distributed far away from the cluster centres. The diagnosis of faults in unlabelled samples also belong to the domain of semi-supervised learning, albeit with the additional challenge associated with many unique components and features. This challenge can necessitate *active learning* (Aghdam et al., 2019; Jian et al., 2021), in which a model identifies selected unlabelled datapoints and alerts a human expert to label them. Active learning requires human intervention, but aims to make use of human efforts as efficiently as possible to improve the model accuracy (Q. Zhang et al., 2021). Another scheme used to overcome incomplete labels is *few-shot learning* (Y. Wang et al., 2020), where a model is optimised to perform supervision tasks with insufficient data for normal supervision training (D. Zhou et al., 2018; A. Zhang et al., 2019; Ren et al., 2020). Few-shot learning provides an interesting opportunity to learn fault features with only a few instances in a training datasets, as can be the case for many rare faults or components. In the case where no labels exist, supervision algorithms might still be applicable through *zero-shot learning* (T. Zhang et al., 2021). In zero-shot learning, the model seeks to generalise knowledge from seen classes to unseen classes with similar behaviour, much like how humans can see images of house-cats and dogs and then correctly categorise lions to felines and wolves to canines (Gao et al., 2020; Feng & Zhao, 2021).

### 2.4.2. Inexact supervision

*Inexact labels* coarsely describe some aspects of the ground truth for a set of features, but do not accurately define it. In general, symbols like labels can not fully represent physical processes of unknown dimensions. Instead, labels define semantics at a certain level of approximation and scale. Thus, labels of physical processes are by nature incomplete semantical descriptions of reality. CM annotations do not describe the properties of each recording in a faulty component, only that from a large bag of recording a fault has been diagnosed. The fault features from each recording were likely not equally important for the diagnosis however, and learning which fea-

Table 3. Different weak supervision challenges, causes and solutions

Weak supervision group	Cause	Solution
Incomplete labels	Missing labels from datapoints	Active learning Semi-supervised learning Few-shot learning Zero-shot learning
Inexact labels	Multiple datapoints per label	Multi-instance learning Contrastive learning
Inaccurate labels	Label is wrong	Regularisation Re-labelling

tures imply which diagnoses would be easier if each recording had its own label.

The challenges of inexact labels have been proposed to be overcome through multi-instance learning (Dietterich et al., 1997; Hoffmann et al., 2011; Zeng et al., 2015) and contrastive learning. In multi-instance learning, the optimisation algorithms seeks to find the common denominators in the label "bags" that are present for learning. By learning from which components were replaced and which were not, correlations in underlying features such as fault severity or deterioration speed can be associated as parts of the bag and used for predictions.

### 2.4.3. Inaccurate supervision

*Inaccurate labels* occur when analysts make fault diagnosis mistakes. This is unlikely to occur with fault classification, but possible with fault severity due to the higher complexity of that task. An analyst can for example assume that a fault may be severe and order a replacement of the component to avoid failure, while the fault actually is minor.

Inaccurate labels are characterised by not conforming to the ground truth, in other words being wrong. To learn with noisy or inaccurate labels, a model seeks to identify and potentially correct incorrect labels (Tanaka et al., 2018). Thus, the model maintains some trust in its predictions, capable of deeming the label inaccurate when confidence in prediction is high and label features deviate from similar labels (J. Li et al., 2019). This trust can be reinforced with physics induced machine learning to maintain a baseline estimate of how labels and signals should correlate, based on physical knowledge of the problem.

## 3. TECHNICAL LANGUAGE SUPERVISION

The direction of research in IFD points towards finding ways to transfer the success on lab datasets to successful applications on industry datasets (Lei et al., 2020; Fink et al., 2020). Both transfer learning and weak supervision can create the opportunity to implement successful algorithms on new datasets without requiring an expensive labelling process. Inspired by recent innovations in TLP and NLS, TLS present a third, yet unused direction to integrate the annotations present

in CM datasets as labels, learning directly from technical language.

The potential effects of TLS can be summarised as

- Opportunities
  - Facilitates direct optimisation on heterogeneous industry data
  - Methods are available and developed in other research areas
  - Language data is commonly associated with condition monitoring data-bases
- Challenges
  - Language annotations are uncertain, and require technical language processing and weak supervision techniques to use
  - Processing of technical language jointly with industry signals is a novel area of research yet to be developed
  - Rapid progress requires open industry datasets containing potentially sensitive information

In this section, we briefly describe the state of TLP and NLS, then combine these into an outline of how TLP can be implemented.

### 3.1. Natural Language Supervision

Natural Language Supervision (F. Chen et al., 2022b) is a recent term introduced to describe machine learning optimisation based on free-form text descriptions rather than predefined labels, though language has been used in a similar fashion to labels before. Labutov et al. (2019) trained semantic parsers that interpret questions and feedback from user natural language responses. Hancock et al. (2018), used natural language explanations of human labelling decision to create BabbleLable, which converts explanations to noisy labels through a semantic parser. Murty et al. (2020) introduced ExpBERT, which is a BERT variation that forms representations using BERT with natural language explanations of the inputs.

Text-encoding is a crucial part of NLP and has seen rapid development recent years. Language models (Peters et al.,

2018) based on the transformer have increased the representational powers of text encoders drastically (Radford, 2018a; Devlin et al., 2019; Radford et al., 2019a; Z. Yang et al., 2020; Microsoft, 2020; Brown et al., 2020a). Early examples of text-image pairings used simpler encoding methods, such as Bag of Words (BoW) and TF-IDF, or recursive encodings derived from the word2vec model (Mikolov et al., 2013), the predecessor of current transformer-based language models. The choice of text-encoder depends on data size and computational power; a larger model can produce better representations, but requires more data and computational power to train. Pre-trained language models with general natural language representational capacity, such as BERT (Devlin et al., 2019), have successfully been fine-tuned on specific tasks with significantly smaller datasets, based on the assumption that the target language and source language has similar underlying distributions.

Optimizing mappings between natural language and images has been done before natural language supervision was introduced; for example, image captioning (Zakir Hossain et al., 2019; X. Lu et al., 2018; S. He et al., 2020) and visual question answering (Antol et al., 2015) have both trained mappings between images and text through top-down or bottom-up mappings (Anderson et al., 2018) and semantic attention (Zhang et al., 2019; Ding et al., 2020). Knowledge and concepts can also be integrated using language as a supervision tool through neuro-symbolic concept learning (Mao et al., 2019), where visual concepts, word representations, and semantic parsing of sentences are jointly learned.

Image recognition generally uses image-text pairs available from online data crawling to train mappings between text and images. Learning directly from the text can also facilitate zero-shot classifiers from language descriptions. Elhoseiny et al. (2013) used text-based descriptions to create a zero-shot image classifier, with text features extracted through Term frequency–Inverse document frequency (Tf-Idf) followed by Clustered Latent Semantic Indexing. J. Lu et al. (2019) introduced ViLBERT, a Vision-and-Language version of BERT, that learns image recognition and language understanding in a two-stream model with interactions between image and text to improve performance compared to single-stream models. Y. Zhang et al. (2020) classified medical images by utilizing text-image pairs through contrastive visual representation learning (ConVIRT) to learn pairings between images and texts. Desai & Johnson (2020) introduced Virtex, which uses captions to enhance pre-training of an image recognition CNN. Sariyildiz et al. (2020) mask words in image-annotation pairs to create image-conditioned masked language modelling (ICMLM) for image classification.

In a recent publication, Radford et al. (2021a) at OpenAI presented CLIP, Contrastive Language–Image Pre-training, which popularised the term natural language supervision and

showed its efficacy for zero-shot classification. They used transformers (Vaswani et al., 2017) for both text and image encodings (Dosovitskiy et al., 2020), and a contrastive (Tian et al., 2020) BoW prediction objective to connect text label to image features in a vector quantised encoding space (van den Oord et al., 2018; Razavi et al., 2019). FILIP by Yao et al. (2021) uses a fine-grained word-patch image alignment to detect and classify objects based on text descriptions, obtaining finer level-alignment in image-text comprehension through unsupervised natural language supervision. C. Jia et al. (2021) scaled natural language supervision further by training directly on un-filtered images and annotations with over one billion image-text pairs. Z. Wang, Yu, Firat, & Cao (2021) introduced unsupervised data generation to synthesise labels for downstream tasks and thus achieve SOTA results on SuperGLUE (A. Wang et al., 2020).

In earlier models, Ramanathan et al. (2013) used natural language supervision to train a video event understanding model in 2013 through a rule-based BoW-like model, and Williams et al. (2018) used language as reward functions for training robots.

### 3.2. Technical Language Processing

The term Technical language processing was introduced in Dec 2020 by Brundage, Sexton, et al. (2021) in collaboration with the American National Institute of Standards and Technology, and concerns the application of NLP techniques and pipelines on technical language. The processing of technical language requires natural language processing methods with additional considerations related to the characteristics of technical language, which is characterised by a higher frequency of information-rich key-words, more abbreviations, and considerably less data than natural language. TLP can be used as a basis for TLS, but can also directly enhance CM practices by offering insights into key performance indicators from work order features (Sharp et al., 2021).

The challenges inherent in using free form text data from industrial contexts - namely data scarcity, a high density of important but (to the model) undefined abbreviations, and technical terms and concepts critical for maintenance context but not inherently defined by their context - are different enough from current NLP research to warrant its own key word in TLP. An ideal TLP model which performs as well as modern language models do on natural language would be able to answer free-form questions on the text dataset, understanding what parts of MWOs and annotations indicate fault class, severity or maintenance actions, and similar tasks currently only possible with human analysis. However, the aforementioned challenges make direct implementation of pre-trained language models difficult; Dima et al. (2021) describe the challenges in adapting natural language processing for technical text in detail. and warn against possible shortcomings

Table 4. Annotations associated with data from Figure 3.

Case	Months after fault detection	Annotation (translated from Swedish)
BPFO indication	4	BPFO Env low
BPFO	10	BPFO visible in mm/s as overtones high up in the spectrum between 1000 and 2000 Hz. WO written on BPFO
Feedback	12	Bearing replaced YYYYMMDD levels of BPFO low again

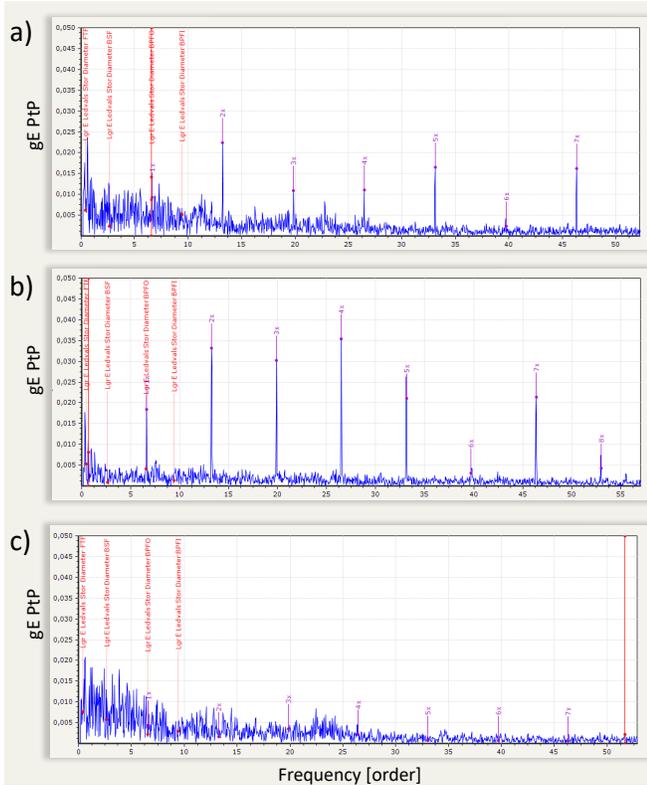


Figure 3. Order analysis results for a vibration signal at a) 4 months; b) 10 months; and c) 12 months after the first indication of a fault in a drying cylinder bearing of a paper machine. Included are also the corresponding text annotations written by experienced condition monitoring analysts employed at the factory. The annotations have been translated from Swedish to English to improve clarity. BPFO peaks are clearly visible in panel a) four months after the first indication of the bearing fault. After ten months, the amplitude of the BPFO peaks in panel b) have increased and a work order (WO) has been written by the analysts. Two months later the bearing has been replaced and no BPFO signature can be seen in panel c).

of implementing SOTA NLP models without considering the specific needs of the process or the people involved. A large black box model can lead to issues with model justifiability, scrutiny and bias, undermining confidence in the system.

### 3.2.1. Technical Language Processing Implementations

Implementations of word embedding models, among those language models, have seen some testing.

Nandyala et al. (2021), implemented five models for vector representation of technical text using an open source dataset describing 5,485 work orders for 5 excavators Hodkiewicz et al. (2017). To evaluate their results they relied on qualitative human evaluation in word and sentence similarities, as well as word cluster projections, as no obvious extrinsic evaluation tasks are available in the model. The authors also survey the literature on fields with challenges similar to those faces in technical language, and discovered similar problem formulations in finance, law, medicine and bio-medicine. In particular, the bio-medical community has developed public datasets for training and benchmarking of domain-specific NLP models.

Cadavid et al. (2020) used a French version of RoBERTa (Y. Liu et al., 2019) called CamemBERT to estimate language features such as duration and criticality of maintenance problems based on operator descriptions. They used equipment descriptions, importance and symptoms as input, and type of disturbance as criticality output (dominant or recessive) and maintenance workload (hours) as outputs for duration. Such input-output pairs allow for extrinsic evaluation, but also fine-tuning of model parameters. The results indicate that Tf-IDf considerably outperforms the base CamemBERT and almost as well as fine-tuned CamemBERT, which implies that the task, data or evaluation are insufficient to fully benefit from the representational capacities of large language models.

Brundage, Sharp, & Pavel (2021) show an association between signal values and expert annotations by generating a technical language dataset with the help of two technicians. One technician generated and monitored faults, followed by another technician writing annotations. The authors find a clear correlation between annotation contents and expert condition monitoring, which presents a strong case for language supervision. Lowenmark et al. (2022) investigate the effect of out-of-vocabulary technical terms on BERT and SentenceBERT performance annotation representations by substituting key terms with in-vocabulary natural language terms. The challenges of evaluation without labels or benchmark datasets were also discussed, and two methods to simulate extrinsic

metrics were suggested. The authors found that the clusterability as measured by k-means score, and the predictability of automatically assigned fault class labels, both improved with only a few key words substituted.

### 3.2.2. Technical Language Processing Embeddings

Language-based models require a mathematical representation of language. This is achieved through pre-processing and an embedding algorithm. The pre-processing step involves tokenisation, cleaning and spell-checking, stop-words removal, stemming/lemmatisation, and fundamental language analysis such as part of speech tagging and named entity recognition. The embedding algorithm can be as simple as one-hot encoding or a complex massive transformers based architecture.

Figure 3 and Table 4 illustrate an example of technical language annotations and condition monitoring signals from a craft liner production plant in northern Sweden. The Figure shows three different envelope-filtered measurements associated with the annotations shown in the Table. The first annotation indicates that there is a fault of class Ball-Pass Frequency Outer ring (BPFO) with a low severity, which is related to the low-intensity peaks at characteristic kinematically based order frequencies in the spectrum. The second annotation describes that the corresponding overtones have increased in magnitude and that a work order has been written. At that point the fault is estimated to be more severe and at the end of its RUL, so the component (bearing) has to be replaced. Finally, the third annotation informs that a bearing has been replaced and that the vibration levels are low, indicating a healthy component.

### 3.2.3. Challenges and Solutions

Pre-processing of technical language faces several difficulties, as use of technical language can vary even in the same field, and there is no uniformly defined list of stems/lemmas, stop-words or correct spellings. For instance, if a CM dataset contains faults of class "Ball-Pass Frequency Outer" (BPFO) and "Ball-Pass Frequency Inner" (BPFI), but one is considerably more common than the other, an automated spell-checker might assume that one is a spelling error. Likewise, there is no defined dictionary for stemming of technical words such as BPFO or BPFI, and reducing both words to "BPF" naturally loses critical information. Therefore it is necessary with a "human-in-the-loop" system until a level of language processing maturity which accurately covers the heterogeneous field of technical language is achieved. One dictionary of technical stop words has been produced (Sarica & Luo, 2021), though it is not necessarily the case that this list is accurate for industries besides those covered in the article.

Encoding technical language to vectors faces a major challenge in that many technical words specific to industries are

not in the vocabulary of NLP models trained on natural language. Addressing this directly with NLP methods is thus related to handling out of vocabulary (OOV) words. A common method to deal with OOV words, used in for instance BERT (Devlin et al., 2019) and GPT (Radford, 2018b; Radford et al., 2019b; Brown et al., 2020b), is to input subword encodings such as byte-pair encodings (BPE) (Gage, 1994; Sennrich et al., 2015) or WordPieces (Schuster & Nakajima, 2012; Y. Wu et al., 2016), rather than the words themselves as inputs to the model. Both models work by learning to maximise the coverage of words in the corpus using a typically fixed amount of subwords. Thus, common words are assigned one whole token, while uncommon words or word endings, such as the "ing"-suffix in for instance "running", might be assigned multiple tokens. The difference between BPE and WordPiece comes mainly from how the subwords are assigned, where BPE chooses the most frequent byte pair and WordPiece chooses the pair which maximises the likelihood of the training data. Other models try to learn to predict the meaning of an unknown word based on surrounding words, individual characters, or a combination of both (Lochter et al., 2020). Implementing an OOV solution which allows transfer learning of a pre-trained deep learning NLP encoder could potentiate more semantically accurate representations of technical language word embeddings, which in turn would improve the potential for TLS.

Another method to encode technical language is through human designed expert systems - essentially a set of rules describing the keywords for faults, actions, severities etc (Sexton et al., 2018). The annotation

*"High BPFO in env3. WO on bearing replacement"*

would thus be decomposed into

*class – BPFO; severity – high; detected in – env3;*

*action – WO replacement; action target – bearing.*

These keywords can then serve as targets for annotation prediction or language based supervision, acting as less noisy labels than learned embeddings for language representations. However, such a system is difficult to scale and vulnerable to new keywords being introduced, essentially requiring tailored engineering and maintenance for each unique industry. It is also vulnerable to oversights from the engineers of the expert system, for instance missing negations in statements, unforeseen keyword usage or a lack of context due to the removal of semantics.

### 3.3. Outline of Technical Language Supervision concepts and model

In the infant stage of TLP, classical NLP methods such as stop-word removal, lemmatisation, stemming and BoW analysis have been used. A potential improvement is to apply

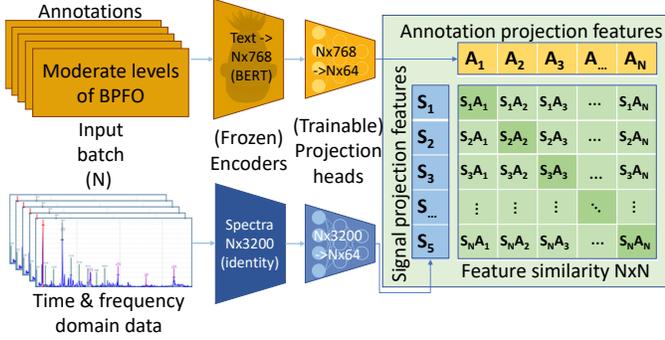


Figure 4. Example illustrating the pre-training step of a natural language supervision model. Annotations and time-frequency domain signal features are encoded, and the model is optimised to connect the correct text-feature pair in the batch of training examples, here marked with dark green colour, through contrastive learning.

more recent innovations in pre-processing and analysis, such as word embedding algorithms coupled with manual tagging of industry-specific technical language.

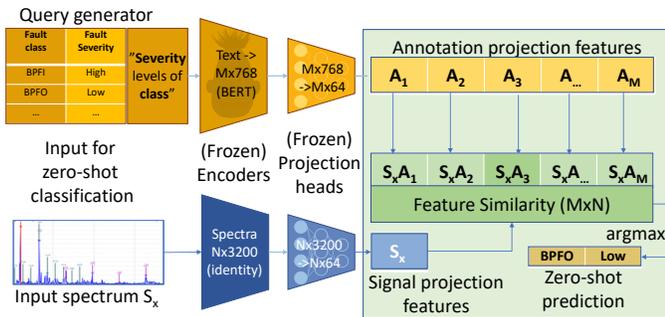


Figure 5. Example illustrating how inference can be generated with the natural-language supervised model outlined in Figure 4. Signal and language inputs are compared in the projection space learned during pre-training, and pairs with the highest feature similarity are used as outputs in either spectrum retrieval or zero-shot classification.

Figures 4 and 5 show a TLS model inspired by the CLIP model Radford et al. (2021a) describing natural language supervision. In the pre-training step, a technical language encoder and a fault diagnosis encoder are used to produce fault and text features. A mapping between fault and text encodings is learned through contrastive learning (Tian et al., 2020; Y. Zhang et al., 2020). In the inference phase, the same encoders are used, but additionally there exists a label query mechanism that maps an input signal to the annotation-based label that is closest to the query in the joint data and language embedding space.

In the case of IFD of rotating machinery, the input is typically sensor data in time-, frequency- and time-frequency-domains. IFD data encoding methods are described in section II, and typically consist of variations of CNNs. Recently, the Transformer (Vaswani et al., 2017), an architecture introduced to

model long-range dependencies and training inefficiencies in NLP, has been successfully used for image recognition without any convolutions in the model (Dosovitskiy et al., 2020; B. Wu et al., 2020; K. Han et al., 2021).

In order to train classification or regression models using language, and not just an annotation generator, a language based labelling method is required. Based on current state-of-the-art methods, some human intervention is required in this step to pre-define the label-space, so that annotations can be matched to the closest label semantically. In (Radford et al., 2021a), a BoW method is implemented to complete pre-defined sentence structures by inserting the correct term chosen from the bag. A similar model could be used in IFD, with more than one degree of freedom in the query to label both fault class and severity. Potentially, further degrees of freedom also enables labelling time-aspects of fault evolution. With a large text dataset and access to well defined labels in parallel with the annotations, a mapping between a more feature-rich encoding and the label space can be learned and implemented to produce labels in a weakly supervised manner for data-annotation pairs where labelled data are not available.

In the case of CM data, the volume and density of text data is low compared to web-crawl results for captioned images on the Internet, or extensively annotated datasets such as COCO (Lin et al., 2015). The language is also domain specific, and annotations are connected to time-frequency data recordings in the dataset, while the semantics of an annotation can be based on analysis of trends over many measurement recordings. This motivates the use of pre-trained models, in combination with feature-engineering and fine-tuning to adapt the model to the domain-specific terms used in process industry. Weak supervision will also be required to deal with unannotated faults, time-delays, a lack of annotations in healthy data, and noise in the annotations resulting from domain-specific language, spelling errors, and grounding noise due to subjective interpretations.

## 4. CASE STUDY

We implement a version of the architecture presented in Figures 4 and 5 using data from a craft paper production plant in northern Sweden, with spectrum and annotation embedding projection heads as trainable parameters through contrastive learning.

### 4.1. Data

The data used comes from six months of recorded data in two large paper mills producing Kraftliner in northern Sweden, and consists of annotated condition monitoring signals from assets, such as dryers, rollers and gearboxes etc. Figure 6 shows a schema of the data structure for each paper mill. Each paper mill forms a database. The database consists of multiple machine parts called assets, which occasionally have

associated annotations when faults have been detected and diagnosed. Each asset has multiple subassets consisting of different types of signals, from one or more sensors. Subassets can be two sensors mounted on the same asset but at opposite ends, or signals from the same sensor that have been transformed using filters such as enveloping. Subassets consist of multiple recordings in the form of time series and spectra measurements, which are the data used in this case study. A recording is one series of data, typically 6400 vibration measurement samples taken over 6.4 seconds, from which spectra with 3200 samples up to 500Hz are computed. Thus, for each annotation there is one associated asset, with multiple subassets, with multiple recordings.

In our dataset, we have 109 annotations with a total of 21090 associated recordings present in a span of ten days before and after the annotation date. Many annotations are identical, and of the 109 annotations there are 43 unique fault descriptions. As data scales up, the number of unique annotations will also increase, which is why a pre-trained language model is needed to ensure system scalability.

#### 4.2. Text Encoder

The text encoder part of our case study TLS model is seen in Figure 4, shown in orange at the top of the figure.

The annotations are embedded using a pretrained and frozen SentenceBERT (Reimers & Gurevych, 2019) model trained on Swedish corpora (Rekathati, 2021), which transforms every annotation to a 768-dimensional embedding vector. as shown in the first two boxes. SentenceBERT is based on BERT and RoBERTa, but is trained to specifically produce good sentence embeddings through siamese and triplet networks (Schroff et al., 2015). In the normal BERT model, each word is projected to a 768-dimensional embedding vector. For example, an annotation with ten words is embedded with dimensions  $10 \times 768$ . To use these embeddings for downstream tasks, it is common to pool them to  $1 \times 768$  then use a feed-forward neural network (FFN). Pooling can be accomplished by averaging each embedding, taking weighted max values, or by using the classification (CLS)-token, which is a final token added to the BERT model that effectively

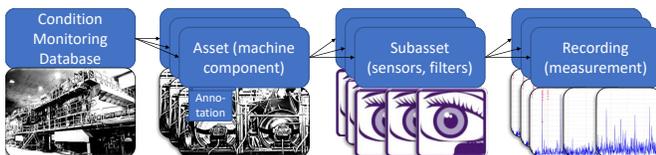


Figure 6. Schema of data structure in a condition monitoring database. The database consists of multiple machine parts called assets, which sometimes have associated annotations. Each asset has multiple subassets consisting of different types of signals, from one or more sensors. Each subasset consists of multiple recordings in the form of time series and spectra measurements.

becomes a learned pooling of the self-attention. SentenceBERT is a BERT-based model fine-tuned on the task of pooling word embeddings to sentence-embeddings, using corpora with similar and dissimilar sentences, and an objective function defined to minimise some distance measure, either softmax, cosine or euclidean between triplets, between similar sentence embeddings. Thus, annotations, which typically consist of one sentence, can be transformed directly to  $1 \times 768$  with a model specifically optimised for this task.

An FFN is then used to reduce the dimensions down to 64 to introduce trainable parameters and reduce the complexity of the dot-product in the contrastive learning step. The FFN is a simple two-layer network with one skip-connection going from 768 to 64 to 64, with a Gaussian error linear unit (GELU) activation function, a 10% dropout, and layer norm. The output of the FFN is then used as input for the contrastive learning step, seen in the rightmost two boxes of the figure

#### 4.3. Signal Encoder

As there are only 109 annotations it is challenging to optimise a network at the asset or subasset level. Therefore we propagate the labels down to the recordings level, where we have 21090 spectrum-annotation-pairs with 43 unique annotations. Thus, the same annotation at an asset will describe every spectra related to that asset, even if some spectra are void of fault features. However, as shown by C. Jia et al. (2021) and Z. Wang, Yu, Yu, et al. (2021), noisy text-image pairs can still converge to a general understanding through the weak supervision that is still present, and it is likely that the same will hold true when replacing images with sensor data.

We directly use the spectra as the fault features, which can be interpreted as the pre-trained model being a FFT and envelope filter of the raw time series. The spectra are projected from 3200 to 64 dimensions with the same reasoning and the same model setup as the annotation embeddings. This is shown in Figure 4 as the spectra encoder being empty, going from 3200 to 3200. As with the annotation embeddings, the resulting 64-dimensional vectors are then sent to the contrastive learning step, seen in the next blue box.

#### 4.4. Contrastive Learning

We train the data using contrastive loss to project positive pairs of signals close and negative pairs further away in a projection space, inspired by the methodology presented in (Radford et al., 2021b). Logits are computed through the dot product of the text and spectrum embeddings in a batch. The self-similarities of spectrum and text embeddings are then computed through the dot product with themselves. The targets, the "labels" for the contrastive loss, are then computed as the softmax of the averages of the self-similarities. The loss for the text-encoder and the spectrum-encoder are then computed separately through cross entropy loss of the logits

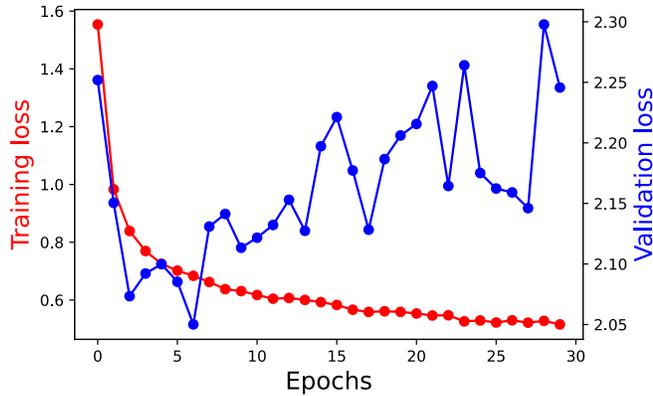


Figure 7. Training and validation contrastive loss for case study model.

and the targets. Finally, the model batch loss is defined as the mean of the spectrum and text losses. We trained the model for only three epochs, as the loss on the validation set quickly started deviating from the train loss, as seen in Figure 7.

#### 4.5. Zero-shot analysis

Finally, the pre-trained model is used to show which spectra in the dataset best correspond to fault queries through spectrum retrieval, and to predict fault classes based on an unlabelled spectra with a label query through zero-shot classification. More specifically, unlabelled spectra chosen from the dataset, and manually chosen label queries, are both used as inputs, while the highest dot product of the embeddings generates a prediction output.

Figure 8 shows spectrum retrieval using queries, described in Table 5, as inputs and receiving matching spectra as outputs. The queries are embedded using the pre-trained technical language supervision model, alongside all spectra in the training and validation set. The output spectra are those with the highest embedding dot product.

Figure 9 illustrates a zero-shot classification implementation of the technical language supervision framework, with queries also described in Table 5. Four examples of spectrum inputs are shown in overlapping pairs in the upper two parts of the figure. The corresponding annotations and axes for these spectra are colour coded and marked as S1-S4. The lower part of the figure illustrates zero-shot classification with five queries, where the inner product between the queries and the

Table 5. Query inputs for spectrum retrieval and zero-shot classification

Query ID	Query
Q1	"BPFO low levels"
Q2	"WO cable replacement"
Q3	"Replace sensor"
Q4	"DC FS"
Q5	"Breakdown"

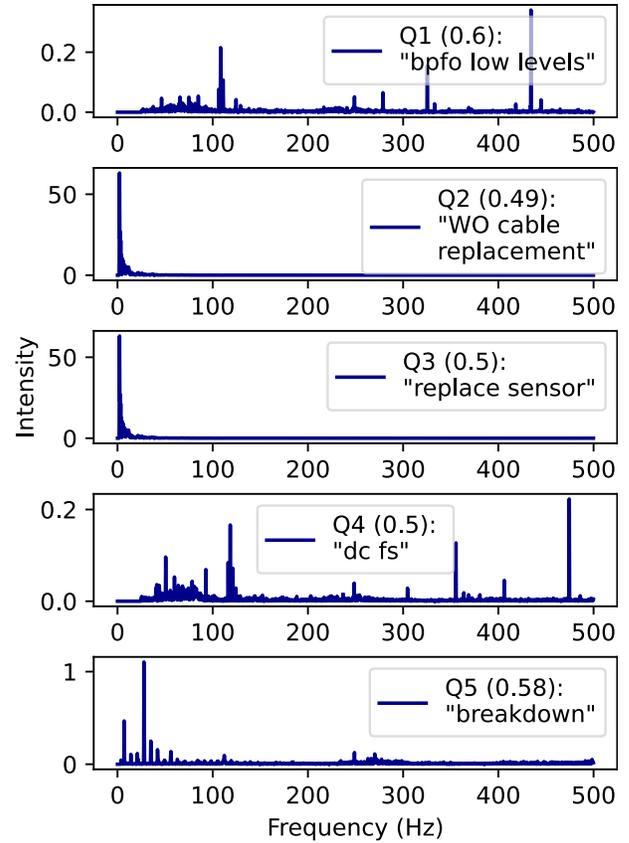


Figure 8. Spectrum retrieval using text queries to sample the top three spectra with the highest embedding dot products

spectra was computed. The inner product between a spectra and a query is represented directly over the query, with colours indicating which spectra the bar is related to.

#### 4.6. Results

##### 4.6.1. Spectrum retrieval

The results of the case study indicate that even with a limited amount of data and a relatively simple model with few hyperparameters, there are aspects of fault diagnosis learned without any labels. For instance, the top four spectra chosen in the spectrum retrieval task shown in Figure 8 are all examples of signals that correspond to their respective query; queries 1 and 4 retrieve spectra indicating bearing faults, with high frequency peaks likely corresponding to characteristic frequencies of bearings, while queries 2 and 3 both indicate cable or sensor faults, seen in the unnaturally high intensity close to zero, indicating a bias in the time series. Query 4 illustrates one interesting property of the correlations between language and signals, where "DC FS" means "drying group free side", which is a phrase commonly seen in conjunction with bearing fault detection or bearing replacement work orders. Query 5 was chosen to test the model where no clear correlations were to be expected, as there were very few oc-

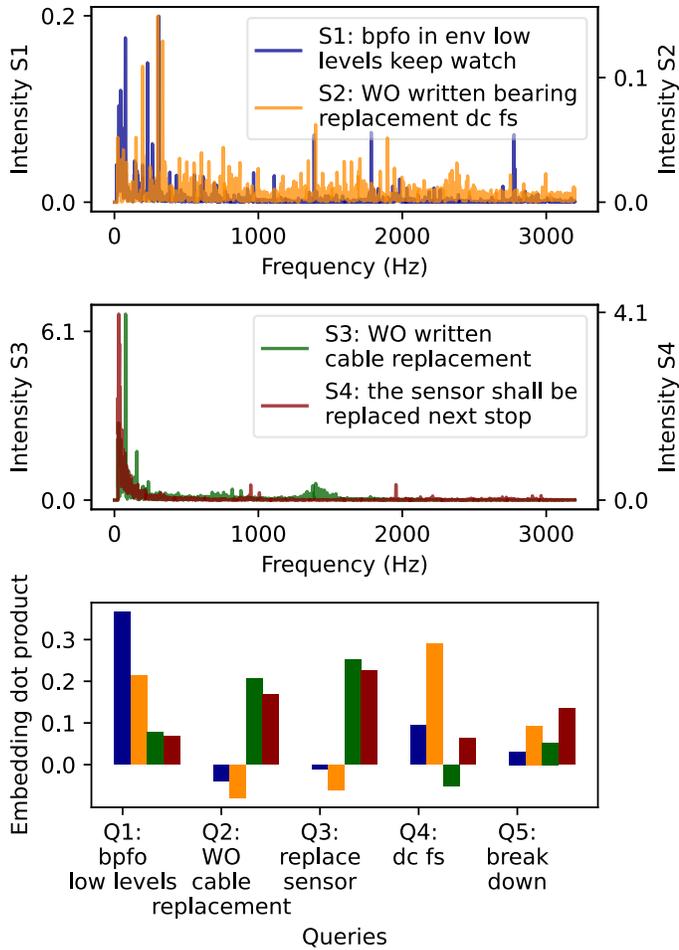


Figure 9. Zero-shot implementation on the case study data. Input spectra are shown in the upper two graphs, and input queries with matching inner products are shown in the lower graph.

currences of "breakdown" in the dataset and "breakdown" does not have one clear signal representation. However, upon consultation with an expert analyst, we learned that the model had picked up a correlation between spectra indicating looseness or play in a bearing, which apparently was the reason behind this breakdown. Thus, the analyst found this spectrum retrieval valuable and successful, indicating the need for close collaboration with experts even when developing self-supervised data-driven models.

#### 4.6.2. Zero-shot predictions

The zero-shot predictions shown in Figure 9 produce good results, where Q1 and Q4 correctly have a higher values for spectra 1 and 2 than for spectra 3 and 4, while Q2 and Q3 correctly correlate more with spectra 3 and 4. In particular, Q1, "BPFO low levels", correctly correlates significantly more with the spectra whose annotation reads "BPFO in env low levels keep watch", and Q4 likewise correctly correlates

much more with S2. Furthermore, both queries correlate more with S1 and S2 which are spectra that indicate bearing faults, despite the individual characteristics of each spectra being different. Q2 and Q3 both correlate strongly to the similar spectra S3 and S4, with Q3 correlating stronger with both spectra, and S3 stronger with each query. However, since cable and sensor faults in general show similar feature spaces, both queries are accurately mapped to similar spectra. Q5, "breakdown", correlates poorly with all chosen spectra, which is an accurate classification as none of the input spectra should indicate a breakdown.

In both the spectrum retrieval and the zero-shot prediction we used normalised text embedding and unnormalised spectrum embeddings before normalising the dot product, as opposed to the normalised spectrum embeddings used during training. Normalising the text embeddings had little impact on either task, but the spectrum retrieval was affected considerably by normalisation of the spectrum embeddings, producing better retrievals with higher values for BPFO-related annotations, but lower values and worse retrievals for cable and sensor faults, while zero-shot predictions were relatively unaffected.

## 4.7. Discussion

The technical language supervision model outlined and implemented in this case study is a basic adaptation of the model used in Radford et al. (2021b). It faces several challenges related to the application to technical language and condition monitoring signals, which are discussed in the following subsections. Table 6 summarises the tasks, challenges and proposed approaches for text encodings, signal encodings, contrastive learning and zero-shot classification.

### 4.7.1. Text encoder

The main challenge for the text encoder is to create good embeddings of technical language, as they are the basis for the potential of the contrastive learning step. As discussed in Section 3.2, this challenge is due to technical language being different from the natural language normally used to train language models, and technical language data scarcity. In this case study, we opted to use a pre-trained natural language model without any fine-tuning. Three approaches for improvements of technical language encodings are shown in the table, which can be summarised as using small-data industry specific solutions through technical language processing, discussed in Section 3.2; large data self-supervised pre-training solutions; and supervised fine-tuning, both discussed in Section 3.1.

### 4.7.2. Signal encoder

For the signal encoder, the main task is to produce good fault feature representations prior to the projection head, comparable to the language model step of the language encoder. The

Table 6. Different TLS tasks, challenges and approaches

Task	Challenges	Approaches
Encoding technical language.	Technical language different from natural language. Limited data availability.	Technical language processing, see 3.2. Self-supervised pre-training, see 3.1. Supervised fine-tuning, see 3.1.
Encoding fault features	Labelled industry data scarce. Lab features difficult to transfer. Non-linear evolution of fault severity. Fault severity levels industry specific.	Transfer Learning, see 2.3. Weak Supervision, see 2.4. Contrastive learning for fine-tuning, see 2.4.2 and 3.1.
Contrastive learning optimisation.	Faults appear and evolve over multiple recordings and signal types.	Sequential model projection heads, see 4.7.3. Data augmentation, see 4.7.3.
Evaluating zero-shot performance and implementation.	Novel task. No benchmark test set.	Industry expert analysis & Industry test deployment, see 4.7.4.

lack of labelled industry data sets, the difficulty of feature transfer and the non-linear and industry-specific properties of fault severity, are all challenges for this task. Approaches to overcome these challenges are discussed in Section two, but more specifically transfer learning, weak supervision and contrastive learning are viable approaches, with specific sections shown in Table 6

#### 4.7.3. Contrastive learning

The task of the contrastive learning part of the model is to force positive pairs to a similar projection space, while negative pairs are pushed away. The main challenge in this step is related to data properties, where annotations are too scarce to fully leverage the utility of scale that is shown in NLS (F. Chen et al., 2022a), and fault evolution too nonlinear for annotation propagation to accurately work as data augmentation. Furthermore, unlike in NLS prediction of image classes, TLS individual recordings are insufficient information to fully assess fault characteristics, akin to describing a movie from just one frame. Thus, multiple recordings must be considered to mimic human analysis in the contrastive learning step, which requires methods able to attend to sequential data such as recurrent neural networks or transformers, either as projection heads or integrated in the text and signal encoders.

Propagating annotation embeddings to each corresponding recording increases the size of the dataset, but also leads to inaccurate supervision from annotations on the recordings level, arising due to the inexactness between recordings level and asset level. For example, if a sensor is faulty at half of its measurements, but works for the other half, the model should ideally be trained only on the faulty signals. Likewise, BPFO is typically detected first in envelope spectra, thus resulting in BPFO annotations being associated with normal spectra where BPFO features have likely not appeared yet. The variety of input types in the spectra inputs is in itself an issue for optimisation, as the network will have to learn to project two very different signals to the same projection in the joint embedding space. However, knowledge of expected fault

behaviour with regards to annotation types could be leveraged to perform improved data augmentation and more accurately propagate annotations in time with annotation contents changing depending on fault type and time distance from true annotation.

#### 4.7.4. Zero-shot predictions

The main challenges with zero-shot classification in an industry environment is that it is a novel field and hard to evaluate without labelled test sets. The contrastive loss or accuracy during optimisation is relative to the model, and offers little insight into model performance at implementation. Therefore we use Figures 8 and 9 to illustrate model performance for two test scenarios. This evaluation requires prior fault diagnosis knowledge however, compared to the much simpler task of evaluating natural language supervision classification for image captioning. However, the efficacy of the model can also be evaluated by test deployment in industry, where feedback from industry experts evaluates whether the model works to improve current fault diagnosis practices or not.

Investigating the zero-shot predictions in Figure 9 showed that a spectrum containing BPFO features gave high inner products also for cable and sensor queries, and we speculated that this might be due to latent BPFO features occasionally seen in the cable and sensor-associated spectrum training data. This is an issue of incomplete supervision, which is further exacerbated if unannotated data is used during training, as the absence of fault annotations does not necessarily guarantee the absence of fault features, given that early faults might go undetected by current analysis. The issues of weak supervision can be addressed by adding data-specific solutions, by for instance limiting extraction dates to after the annotation, or adding pre-processing of annotations to manually handle "replaced"-like annotations as a different class. This issue might also be solveable by simply scaling up data, which has worked in natural language supervision as discussed in Section 4.3.

## 5. CONCLUSION

The fault descriptions and maintenance records commonly stored in modern process industry CM systems are unexploited sources of information for training IFD systems. Language present in CM datasets can be used for technical language based supervision of IFD models to facilitate automation of routine FD tasks and develop more accurate decision support for complex tasks (Ekström & Sandin, 2020). Since language-based labels are intrinsically uncertain, weakly supervised learning methods need to be developed, which can also support transfer learning of pretrained IFD models with labels extracted from language in industry datasets. Our experiments show that even with a basic TLS implementation, without custom signal processing or pre-trained fault diagnosis encoders, a joint embedding space for annotations and fault features can be learned and used for zero-shot classification.

Improvements in TLS can occur both through an enhancement of the TLP pipeline for technical language representations, or through augmented integration of IFD-based signal encoders. However, a major challenge for TLP and TLS research is the lack of realistic and open annotated industry data, which can be used for comparative studies and benchmarks. Furthermore, the assistance of industry experts was sometimes required to understand the annotation language and how annotations were motivated by signal features and the context. Thus, in this work the collaboration between industry and academia was key. Open access annotated datasets with clearly described features and valid benchmark tasks are needed to make this important direction of research more readily accessible.

## ACKNOWLEDGEMENT

This work is supported by the Strategic innovation program Process industrial IT and Automation (PiIA), a joint investment of Vinnova, Formas and the Swedish Energy Agency, reference number 2019-02533.

The analysis of the results was done with help from Håkan Sirkka, a condition monitoring analyst with experience of the industry data used.

We thank the members of the project reference group including Per-Erik Larsson, Kjell Lundberg, Håkan Sirkka and Peter Wikström, for valuable inputs.

KL thanks Prakash Chandra Chhipa for helpful discussions on contrastive learning and joint embedding spaces.

## REFERENCES

Aghdam, H. H., Gonzalez-Garcia, A., van de Weijer, J., & López, A. M. (2019). *Active learning for deep detection neural networks*.

- An, Z., Li, S., Xin, Y., Xu, K., & Ma, H. (2019). An intelligent fault diagnosis framework dealing with arbitrary length inputs under different working conditions. *Measurement Science and Technology*, 30(12).
- Anderson, P., He, X., Buehler, C., Teney, D., Johnson, M., Gould, S., & Zhang, L. (2018). *Bottom-up and top-down attention for image captioning and visual question answering*.
- Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Zitnick, C. L., & Parikh, D. (2015, December). Vqa: Visual question answering. In *Proceedings of the IEEE international conference on computer vision (iccv)*.
- Babu, G., Zhao, P., & Li, X.-L. (2016). Deep convolutional neural network based regression approach for estimation of remaining useful life. *Lecture notes in Computer Science (including subseries Lecture notes in Artificial Intelligence and Lecture notes in Bioinformatics)*, 9642, 214-228.
- Ben Ali, J., Chebel-Morello, B., Saidi, L., Malinowski, S., & Fnaiech, F. (2015). Accurate bearing remaining useful life prediction based on weibull distribution and artificial neural network. *Mechanical Systems and Signal Processing*, 56, 150-172.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... Amodei, D. (2020a). *Language models are few-shot learners*.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... Amodei, D. (2020b). *Language models are few-shot learners*.
- Brundage, M. P., Sexton, T., Hodkiewicz, M., Dima, A., & Lukens, S. (2021). Technical language processing: Unlocking maintenance knowledge. *Manufacturing Letters*, 27, 42-46.
- Brundage, M. P., Sharp, M., & Pavel, R. (2021, Jun). Qualifying evaluations from human operators: Integrating sensor data with natural language logs. *PHM Society European Conference*. 6.
- Cadavid, J. P. U., Grabot, B., Lamouri, S., Pellerin, R., & Fortin, A. (2020). Valuing free-form text data from maintenance logs through transfer learning with camembert. *Enterprise Information Systems*, 0(0), 1-29.
- Cahill, J. (2021). *Improving subsurface models to reduce drilling uncertainty*.
- Cao, P., Zhang, S., & Tang, J. (2018a). Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning. *IEEE Access*, 6, 26241-26253.
- Cao, P., Zhang, S., & Tang, J. (2018b). Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning. *IEEE Access*, 6, 26241-26253.

- Carden, E. P., & Fanning, P. (2004). Vibration based condition monitoring: A review. *Structural Health Monitoring*, 3(4), 355-377.
- Case western reserve university bearing data center website. (n.d.). <https://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website>.
- Cerrada, M., Sánchez, R.-V., Li, C., Pacheco, F., Cabrera, D., de Oliveira, J. V., & Vásquez, R. E. (2018). A review on data-driven fault severity assessment in rolling bearings. *Mechanical Systems and Signal Processing*, 99, 169 - 196.
- Chen, F., Zhang, D., Han, M., Chen, X., Shi, J., Xu, S., & Xu, B. (2022a). *Vlp: A survey on vision-language pre-training*. arXiv.
- Chen, F., Zhang, D., Han, M., Chen, X., Shi, J., Xu, S., & Xu, B. (2022b). *Vlp: A survey on vision-language pre-training*. arXiv.
- Chen, X., Zhang, B., & Gao, D. (2021). Bearing fault diagnosis base on multi-scale cnn and lstm model. *Journal of Intelligent Manufacturing*, 32(4), 971-987.
- Chen, Z., Gryllias, K., & Li, W. (2020). Intelligent fault diagnosis for rotary machinery using transferable convolutional neural network. *IEEE Transactions on Industrial Informatics*, 16(1), 339-349.
- chung Fu, T. (2011). A review on time series data mining. *Engineering Applications of Artificial Intelligence*, 24(1), 164-181.
- Condition based maintenance fault database for testing of diagnostic and prognostics algorithms. (n.d.). <https://www.mfpt.org/fault-data-sets/>.
- del Campo, S. M., & Sandin, F. (2017). Online feature learning for condition monitoring of rotating machinery. *Engineering Applications of Artificial Intelligence*, 64, 187 - 196.
- Desai, K., & Johnson, J. (2020). *Virtex: Learning visual representations from textual annotations*.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *Bert: Pre-training of deep bidirectional transformers for language understanding*.
- Dietterich, T. G., Lathrop, R. H., & Lozano-Pérez, T. (1997). Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, 89(1), 31-71.
- Dima, A., Lukens, S., Hodkiewicz, M., Sexton, T., & Brundage, M. P. (2021). Adapting natural language processing for technical text. *Applied AI Letters*, 2(3), e33.
- Ding, S., Qu, S., Xi, Y., & Wan, S. (2020). Stimulus-driven and concept-driven analysis for image caption generation. *Neurocomputing*, 398, 520-530.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... Houlsby, N. (2020). *An image is worth 16x16 words: Transformers for image recognition at scale*.
- Ekström, K., & Sandin, F. (2020). *Fault severity estimation using weak supervision with language based labels and condition monitoring data*.
- Elhoseiny, M., Saleh, B., & Elgammal, A. (2013). Write a classifier: Zero-shot learning using purely textual descriptions. In *2013 IEEE International Conference on Computer Vision* (p. 2584-2591).
- Emerson. (2021). *Featured technologies/machine learning*.
- Feng, L., & Zhao, C. (2021). Fault description based attribute transfer for zero-sample industrial fault diagnosis. *IEEE Transactions on Industrial Informatics*, 17(3), 1852-1862.
- Fink, O., Wang, Q., Svensén, M., Dersin, P., Lee, W.-J., & Ducoffe, M. (2020). Potential, challenges and future directions for deep learning in prognostics and health management applications. *Engineering Applications of Artificial Intelligence*, 92, 103678.
- Gage, P. (1994). A new algorithm for data compression. *The C Users Journal archive*, 12, 23-38.
- Gao, Y., Gao, L., Li, X., & Zheng, Y. (2020). A zero-shot learning method for fault diagnosis under unknown working loads. *Journal of Intelligent Manufacturing*, 31(4), 899-909.
- Guo, L., Lei, Y., Xing, S., Yan, T., & Li, N. (2019). Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data. *IEEE Transactions on Industrial Electronics*, 66(9), 7316-7325.
- Guo, L., Li, N., Jia, F., Lei, Y., & Lin, J. (2017). A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing*, 240, 98-109.
- Haidong, S., Hongkai, J., Xingqiu, L., & Shuaipeng, W. (2018). Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine. *Knowledge-Based Systems*, 140, 1-14.
- Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., ... Tao, D. (2021). *A survey on visual transformer*.
- Han, T., Liu, C., Yang, W., & Jiang, D. (2019). A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults. *Knowledge-Based Systems*, 165, 474-487.
- Han, T., Liu, C., Yang, W., & Jiang, D. (2020). Deep transfer network with joint distribution adaptation: A new intelligent fault diagnosis framework for industry application. *ISA Transactions*, 97, 269 - 281.
- Hancock, B., Varma, P., Wang, S., Bringmann, M., Liang, P., & Ré, C. (2018). *Training classifiers with natural language explanations*.

- He, S., Liao, W., Tavakoli, H. R., Yang, M., Rosenhahn, B., & Pugeault, N. (2020). *Image captioning through image transformer*.
- He, Z., Shao, H., Zhong, X., & Zhao, X. (2020). Ensemble transfer cnns driven by multi-channel signals for fault diagnosis of rotating machinery cross working conditions. *Knowledge-Based Systems*, 207.
- Helbing, G., & Ritter, M. (2018). Deep learning for fault detection in wind turbines. *Renewable and Sustainable Energy Reviews*, 98, 189 - 198.
- Hoang, D.-T., & Kang, H.-J. (2019). A survey on deep learning based bearing fault diagnosis. *Neurocomputing*, 335, 327-335.
- Hodkiewicz, M. R., Batsioudis, Z., Radomiljac, T., & Ho, M. T. (2017). Why autonomous assets are good for reliability – the impact of ‘operator-related component’ failures on heavy mobile equipment reliability..
- Hoffmann, R., Zhang, C., Ling, X., Zettlemoyer, L., & Weld, D. S. (2011, June). Knowledge-based weak supervision for information extraction of overlapping relations. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies* (pp. 541–550). Portland, Oregon, USA: Association for Computational Linguistics.
- Jardine, A., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), 1483-1510.
- Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), 1483 - 1510.
- Jia, C., Yang, Y., Xia, Y., Chen, Y.-T., Parekh, Z., Pham, H., ... Duerig, T. (2021). *Scaling up visual and vision-language representation learning with noisy text supervision*.
- Jia, F., Lei, Y., Lin, J., Zhou, X., & Lu, N. (2016). Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems and Signal Processing*, 72-73, 303-315.
- Jia, F., Lei, Y., Lu, N., & Xing, S. (2018). Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization. *Mechanical Systems and Signal Processing*, 110, 349-367.
- Jian, C., Yang, K., & Ao, Y. (2021). Industrial fault diagnosis based on active learning and semi-supervised learning using small training set. *Engineering Applications of Artificial Intelligence*, 104, 104365.
- Jiang, G., Xie, P., He, H., & Yan, J. (2018). Wind turbine fault detection using a denoising autoencoder with temporal information. *IEEE/ASME Transactions on Mechatronics*, 23(1), 89-100.
- Khan, S., & Yairi, T. (2018). A review on the application of deep learning in system health management. *Mechanical Systems and Signal Processing*, 107, 241-265.
- Kothamasu, R., Huang, S. H., & VerDuin, W. H. (2006, Jul 01). System health monitoring and prognostics — a review of current paradigms and practices. *The International Journal of Advanced Manufacturing Technology*, 28(9), 1012-1024.
- Labutov, I., Yang, B., & Mitchell, T. (2019). *Learning to learn semantic parsers from natural language supervision*.
- Lei, Y., Jia, F., Lin, J., Xing, S., & Ding, S. (2016). An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. *IEEE Transactions on Industrial Electronics*, 63(5), 3137-3147.
- Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to rul prediction. *Mechanical Systems and Signal Processing*, 104, 799-834.
- Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing*, 138, 106587.
- Li, C., Zhang, S., Qin, Y., & Estupinan, E. (2020). A systematic review of deep transfer learning for machinery fault diagnosis. *Neurocomputing*, 407, 121 - 135.
- Li, J., Wong, Y., Zhao, Q., & Kankanhalli, M. (2019). Learning to learn from noisy labeled data. In (Vol. 2019-June, p. 5046-5054).
- Li, X., Ding, Q., & Sun, J.-Q. (2018). Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Engineering and System Safety*, 172, 1-11.
- Li, X., Li, X., & Ma, H. (2020). Deep representation clustering-based fault diagnosis method with unsupervised data applied to rotating machinery. *Mechanical Systems and Signal Processing*, 143, 106825.
- Li, X., Zhang, W., Ding, Q., & Li, X. (2020). Diagnosing rotating machines with weakly supervised data using deep transfer learning. *IEEE Transactions on Industrial Informatics*, 16(3), 1688-1697.
- Li, X., Zhang, W., Ding, Q., & Li, X. (2020). Diagnosing rotating machines with weakly supervised data using deep transfer learning. *IEEE Transactions on Industrial Informatics*, 16(3), 1688-1697.
- Li, X., Zhang, W., Xu, N.-X., & Ding, Q. (2020). Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places. *IEEE Transactions on Industrial Electronics*, 67(8), 6785-6794.

- Li, Y., Lin, T., Yi, K., Bear, D. M., Yamins, D. L. K., Wu, J., ... Torralba, A. (2020). *Visual grounding of learned physical models*.
- Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., ... Dollár, P. (2015). *Microsoft coco: Common objects in context*.
- Liu, H., Liu, C., & Huang, Y. (2011). Adaptive feature extraction using sparse coding for machinery fault diagnosis. *Mechanical Systems and Signal Processing*, 25(2), 558 - 574.
- Liu, H., Zhou, J., Zheng, Y., Jiang, W., & Zhang, Y. (2018). Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA Transactions*, 77, 167-178.
- Liu, R., Yang, B., Zio, E., & Chen, X. (2018). Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, 108, 33-47.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... Stoyanov, V. (2019). *Roberta: A robustly optimized bert pretraining approach*.
- Lochter, J. V., Silva, R. M., & Almeida, T. A. (2020). *Deep learning models for representing out-of-vocabulary words*.
- Lowenmark, K., Taal, C., Nivre, J., Liwicki, M., & Sandin, F. (2022). Processing of condition monitoring annotations with bert and technical language substitution: A case study. *Proceedings of the 7th European Conference of the Prognostics and Health Management Society 2022*, 306-314.
- Lu, C., Wang, Z.-Y., Qin, W.-L., & Ma, J. (2017). Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Processing*, 130, 377-388.
- Lu, J., Batra, D., Parikh, D., & Lee, S. (2019). *Vilbert: Pre-training task-agnostic visiolinguistic representations for vision-and-language tasks*.
- Lu, X., Wang, B., Zheng, X., & Li, X. (2018). Exploring models and data for remote sensing image caption generation. *IEEE Transactions on Geoscience and Remote Sensing*, 56(4), 2183-2195.
- Mao, J., Gan, C., Kohli, P., Tenenbaum, J. B., & Wu, J. (2019). *The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision*.
- Microsoft. (2020). *Turing-nlg: A 17-billion paramater language model by microsoft*.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). *Distributed representations of words and phrases and their compositionality*.
- Monitron, A. (n.d.). *Detect abnormal machine behavior and enable predictive maintenance*.
- Murty, S., Koh, P. W., & Liang, P. (2020). *Expbert: Representation engineering with natural language explanations*.
- Nan, C., Khan, F., & Iqbal, M. T. (2008). Real-time fault diagnosis using knowledge-based expert system. *Process Safety and Environmental Protection*, 86(1), 55-71.
- Nandyala, A., Lukens, S., Rathod, S., & Agarwal. (2021, Jun). Evaluating word representations in a technical language processing pipeline. *PHM Society European Conference*. 6.
- Nasa prognostic data repository. (n.d.). <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/>.
- Pan, J., Zi, Y., Chen, J., Zhou, Z., & Wang, B. (2018). Liftingnet: A novel deep learning network with layerwise feature learning from noisy mechanical data for fault classification. *IEEE Transactions on Industrial Electronics*, 65(6), 4973-4982.
- Papayan, V., Romano, Y., Sulam, J., & Elad, M. (2018). Theoretical foundations of deep learning via sparse representations: A multilayer sparse model and its connection to convolutional neural networks. *IEEE Signal Processing Magazine*, 35(4), 72-89.
- PdM. (2021). *Pdm services vibration analysis monitoring*.
- Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). *Deep contextualized word representations*.
- ProcessIT. (2018). *Processit.eu european roadmap for process industrial automation. second version.*, 3(3).
- Qiao, M., Yan, S., Tang, X., & Xu, C. (2020). Deep convolutional and lstm recurrent neural networks for rolling bearing fault diagnosis under strong noises and variable loads. *IEEE Access*, 8, 66257-66269.
- Qin, S. (2012). Survey on data-driven industrial process monitoring and diagnosis. *Annual Reviews in Control*, 36(2), 220-234.
- Radford, A. (2018a). *Improving language understanding by generative pre-training..*
- Radford, A. (2018b). *Improving language understanding by generative pre-training..*
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... Sutskever, I. (2021a). *Learning transferable visual models from natural language supervision*.
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... Sutskever, I. (2021b). *Learning transferable visual models from natural language supervision*.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019a). *Language models are unsupervised multitask learners..*
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019b). *Language models are unsupervised multitask learners..*

- Rai, A., & Upadhyay, S. (2016). A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribology International*, 96, 289-306.
- Ramanathan, V., Liang, P., & Fei-Fei, L. (2013). Video event understanding using natural language descriptions. In *2013 IEEE International Conference on Computer Vision* (p. 905-912).
- Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., ... Sutskever, I. (2021). *Zero-shot text-to-image generation*.
- Randall, R. B., & Antoni, J. (2011). Rolling element bearing diagnostics—a tutorial. *Mechanical Systems and Signal Processing*, 25(2), 485 - 520.
- Razavi, A., van den Oord, A., & Vinyals, O. (2019). *Generating diverse high-fidelity images with vq-vae-2*.
- Razavi-Far, R., Hallaji, E., Farajzadeh-Zanjani, M., & Saif, M. (2019). A semi-supervised diagnostic framework based on the surface estimation of faulty distributions. *IEEE Transactions on Industrial Informatics*, 15(3), 1277-1286.
- Reimers, N., & Gurevych, I. (2019, November). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *EMNLP-IJCNLP 2019* (pp. 3982–3992). Hong Kong, China: Association for Computational Linguistics.
- Rekathati, F. (2021). The KBLab blog: Introducing a Swedish sentence transformer.
- Ren, Z., Zhu, Y., Yan, K., Chen, K., Kang, W., Yue, Y., & Gao, D. (2020). A novel model with the ability of few-shot learning and quick updating for intelligent fault diagnosis. *Mechanical Systems and Signal Processing*, 138.
- Sarica, S., & Luo, J. (2021, Aug). Stopwords in technical language processing. *PLOS ONE*, 16(8), e0254937.
- Sariyildiz, M. B., Perez, J., & Larlus, D. (2020). *Learning visual representations with caption annotations*.
- Schroff, F., Kalenichenko, D., & Philbin, J. (2015, Jun). Facenet: A unified embedding for face recognition and clustering. *CVPR 2015*.
- Schuster, M., & Nakajima, K. (2012). Japanese and korean voice search. In *ICASSP 2012* (p. 5149-5152).
- Schwendemann, S., Amjad, Z., & Sikora, A. (2021). Bearing fault diagnosis with intermediate domain based layered maximum mean discrepancy: A new transfer learning approach. *Engineering Applications of Artificial Intelligence*, 105, 104415.
- Sennrich, R., Haddow, B., & Birch, A. (2015). Neural machine translation of rare words with subword units. *arXiv preprint arXiv:1508.07909*.
- Sexton, T., Brundage, M., Hodkiewicz, M., & Smoker, T. (2018, 2018-09-24). Benchmarking for keyword extraction methodologies in maintenance work orders. 2018 Annual Conference of the Prognostics and Health Management Society, Philadelphia, PA.
- Shao, H., Xia, M., Han, G., Zhang, Y., & Wan, J. (2021). Intelligent fault diagnosis of rotor-bearing system under varying working conditions with modified transfer convolutional neural network and thermal images. *IEEE Transactions on Industrial Informatics*, 17(5), 3488-3496.
- Shao, S., McAleer, S., Yan, R., & Baldi, P. (2019). Highly accurate machine fault diagnosis using deep transfer learning. *IEEE Transactions on Industrial Informatics*, 15(4), 2446-2455.
- Sharma, V., & Parey, A. (2016). A review of gear fault diagnosis using various condition indicators. In (Vol. 144, p. 253-263).
- Sharp, M., Brundage, M., Sexton, T., & Madhusudanan, F. (2021, 2021-04-22 04:04:00). Discovering critical KPI factors from natural language in maintenance work orders. , 3(3).
- Shin, J.-H., & Jun, H.-B. (2015). On condition based maintenance policy. *Journal of Computational Design and Engineering*, 2(2), 119 - 127.
- Simon, J. (n.d.). *Amazon monitron, a simple and cost-effective service enabling predictive maintenance*.
- SKF. (n.d.). *Skf enlight ai*.
- SKF. (2022). *Skf @ptitude observer user manual*.
- SKF, A., & Kommunikation, S. (2020, March). *Skf annual report 2020*. <https://investors.skf.com/sites/default/files/pr/202103032688-1.pdf>.
- Smith, W., & Randall, R. (2015). Rolling element bearing diagnostics using the case western reserve university data: A benchmark study. *Mechanical Systems and Signal Processing*, 64-65, 100-131.
- Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., ... Nenadic, G. (2019). Machine learning methods for wind turbine condition monitoring: A review. *Renewable Energy*, 133, 620-635.
- Stief, A., Ottewill, J., Baranowski, J., & Orkisz, M. (2019). A pca and two-stage bayesian sensor fusion approach for diagnosing electrical and mechanical faults in induction motors. *IEEE Transactions on Industrial Electronics*, 66(12), 9510-9520.
- Tanaka, D., Ikami, D., Yamasaki, T., & Aizawa, K. (2018). Joint optimization framework for learning with noisy labels. In (p. 5552-5560).
- Tian, Y., Krishnan, D., & Isola, P. (2020). *Contrastive multi-view coding*.
- van den Oord, A., Vinyals, O., & Kavukcuoglu, K. (2018). *Neural discrete representation learning*.

- van Engelen, J. E., & Hoos, H. H. (2020, Feb 01). A survey on semi-supervised learning. *Machine Learning*, 109(2), 373-440.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). *Attention is all you need*.
- Wang, A., Pruksachatkun, Y., Nangia, N., Singh, A., Michael, J., Hill, F., ... Bowman, S. R. (2020). *Superglue: A stickier benchmark for general-purpose language understanding systems*.
- Wang, D., Tsui, K.-L., & Miao, Q. (2017). Prognostics and health management: A review of vibration based bearing and gear health indicators. *IEEE Access*, 6, 665-676.
- Wang, Q., Michau, G., & Fink, O. (2019). Domain adaptive transfer learning for fault diagnosis. In (p. 279-285).
- Wang, Y., Yao, Q., Kwok, J., & Ni, L. M. (2020). *Generalizing from a few examples: A survey on few-shot learning*.
- Wang, Z., Yu, A. W., Firat, O., & Cao, Y. (2021). *Towards zero-label language learning*.
- Wang, Z., Yu, J., Yu, A. W., Dai, Z., Tsvetkov, Y., & Cao, Y. (2021). *Simvlm: Simple visual language model pretraining with weak supervision*. arXiv.
- Wen, L., Gao, L., & Li, X. (2019). A new deep transfer learning based on sparse auto-encoder for fault diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(1), 136-144.
- Wen, L., Li, X., & Gao, L. (2020). A transfer convolutional neural network for fault diagnosis based on resnet-50. *Neural Computing and Applications*, 32(10), 6111-6124.
- Williams, E. C., Gopalan, N., Rhee, M., & Tellex, S. (2018). Learning to parse natural language to grounded reward functions with weak supervision. In *2018 IEEE international conference on robotics and automation (icra)* (p. 4430-4436).
- Wu, B., Xu, C., Dai, X., Wan, A., Zhang, P., Yan, Z., ... Vajda, P. (2020). *Visual transformers: Token-based image representation and processing for computer vision*.
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., ... Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *ArXiv, abs/1609.08144*.
- Xu, G., Liu, M., Jiang, Z., Shen, W., & Huang, C. (2020). Online fault diagnosis method based on transfer convolutional neural networks. *IEEE Transactions on Instrumentation and Measurement*, 69(2), 509-520.
- Xu, Y., Sun, Y., Liu, X., & Zheng, Y. (2019). A digital-twin-assisted fault diagnosis using deep transfer learning. *IEEE Access*, 7, 19990-19999.
- Yan, X., She, D., Xu, Y., & Jia, M. (2021). Deep regularized variational autoencoder for intelligent fault diagnosis of rotor bearing system within entire life-cycle process. *Knowledge-Based Systems*, 226, 107142.
- Yang, B., Lei, Y., Jia, F., & Xing, S. (2019). An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. *Mechanical Systems and Signal Processing*, 122, 692 - 706.
- Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. V. (2020). *Xlnet: Generalized autoregressive pretraining for language understanding*.
- Yao, L., Huang, R., Hou, L., Lu, G., Niu, M., Xu, H., ... Xu, C. (2021). *Filip: Fine-grained interactive language-image pre-training*.
- Yiakopoulos, C., Gryllias, K., & Antoniadis, I. (2011). Rolling element bearing fault detection in industrial environments based on a k-means clustering approach. *Expert Systems with Applications*, 38(3), 2888 - 2911.
- Yin, S., Ding, S., Xie, X., & Luo, H. (2014). A review on basic data-driven approaches for industrial process monitoring. *IEEE Transactions on Industrial Electronics*, 61(11), 6418-6428.
- Yu, K., Fu, Q., Ma, H., Lin, T., & Li, X. (2021, 07). Simulation data driven weakly supervised adversarial domain adaptation approach for intelligent cross-machine fault diagnosis. *Structural Health Monitoring*, 20.
- Yu, K., Lin, T. R., Ma, H., Li, X., & Li, X. (2021). A multi-stage semi-supervised learning approach for intelligent fault diagnosis of rolling bearing using data augmentation and metric learning. *Mechanical Systems and Signal Processing*, 146, 107043.
- Zakir Hossain, M., Sohel, F., Shiratuddin, M., & Laga, H. (2019). A comprehensive survey of deep learning for image captioning. *ACM Computing Surveys*, 51(6).
- Zeng, D., Liu, K., Chen, Y., & Zhao, J. (2015, September). Distant supervision for relation extraction via piecewise convolutional neural networks. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 1753-1762). Lisbon, Portugal: Association for Computational Linguistics.
- Zhai, X., Oliver, A., Kolesnikov, A., & Beyer, L. (2019). *S4l: Self-supervised semi-supervised learning*.
- Zhang, A., Li, S., Cui, Y., Yang, W., Dong, R., & Hu, J. (2019). Limited data rolling bearing fault diagnosis with few-shot learning. *IEEE Access*, 7, 110895-110904.
- Zhang, A., Wang, H., Li, S., Cui, Y., Liu, Z., Yang, G., & Hu, J. (2018). Transfer learning with deep recurrent neural networks for remaining useful life estimation. *Applied Sciences (Switzerland)*, 8(12).
- Zhang, D., Qian, L., Mao, B., Huang, C., Huang, B., & Si, Y. (2018). A data-driven design for fault detection of wind turbines using random forests and xgboost. *IEEE Access*, 6, 21020-21031.

- Zhang, H., Zhang, Q., Liu, J., & Guo, H. (2018). Fault detection and repairing for intelligent connected vehicles based on dynamic bayesian network model. *IEEE Internet of Things Journal*, 5(4), 2431-2440.
- Zhang, Q., Lu, J., & Jin, Y. (2021, Feb 01). Artificial intelligence in recommender systems. *Complex & Intelligent Systems*, 7(1), 439-457.
- Zhang, S., Ye, F., Wang, B., & Habetler, T. G. (2019). *Semi-supervised learning of bearing anomaly detection via deep variational autoencoders*.
- Zhang, S., Zhang, S., Wang, B., & Habetler, T. G. (2020). Deep learning algorithms for bearing fault diagnostics—a comprehensive review. *IEEE Access*, 8, 29857-29881.
- Zhang, T., Chen, J., Li, F., Zhang, K., Lv, H., He, S., & Xu, E. (2021). Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions. *ISA Transactions*.
- Zhang, Y., Jiang, H., Miura, Y., Manning, C. D., & Langlotz, C. P. (2020). *Contrastive learning of medical visual representations from paired images and text*.
- Zhang, Z., Wu, Q., Wang, Y., & Chen, F. (2019). High-quality image captioning with fine-grained and semantic-guided visual attention. *IEEE Transactions on Multimedia*, 21(7), 1681-1693.
- Zhao, K., Jiang, H., Wu, Z., & Lu, T. (2020). A novel transfer learning fault diagnosis method based on manifold embedded distribution alignment with a little labeled data. *Journal of Intelligent Manufacturing*.
- Zhong, S.-S., Fu, S., & Lin, L. (2019). A novel gas turbine fault diagnosis method based on transfer learning with cnn. *Measurement: Journal of the International Measurement Confederation*, 137, 435-453.
- Zhou, D., He, J., Yang, H., & Fan, W. (2018). Sparc: Self-paced network representation for few-shot rare category characterization. In (p. 2807-2816).
- Zhou, Z.-H. (2017, 08). A brief introduction to weakly supervised learning. *National Science Review*, 5(1), 44-53.