Automate Quality Prediction in an End-of-Line Test of a Highly Variant Production of Geared Motors – Discussion of a Full Concept

Peter Wißbrock[1], David Pelkmann[2], and Björn Tölle[3]

[1]*Lenze SE, Aerzen, 31855, Germany*
peter.wissbrock@lenze.com

[2]*Bielefeld University of Applied Sciences – Center for Applied Data Science Gütersloh, Bielefeld, 33619, Germany*
david.pelkmann@fh-bielefeld.de

[3]*OWL University of Applied Sciences and Arts, Lemgo, 32657, Germany*
bjoerntoelle@gmx.de

Abstract

The diversity of variants rises continuously within production of complex and modular mechanical components such as geared motors. At the same time, the requirement for a good sound quality of the product increases. In a batch size one production, it is difficult for the assembly worker to consistently evaluate the product quality. Automation of the end-of-line test leads to several challenges. Common vibration and acoustic measurements are time-consuming and the features must be defined individually for every single product configuration. This paper presents a full concept for automation or semi-automation of end-of-line tests in a highly variant production of geared motors. It is shown, how acoustic measurement can be done in common industrial production and an overview is given on typically used machine learning methods and features for quality prediction of geared motors. Furthermore, a concept for dealing with the lack of labeled examples is provided to analyze historically unknown product configurations. Finally, it is discussed how to structure classifiers to capture all known and unknown faults. The end-of-line concept is a fundamental module for industry 4.0 and can be generalized to all modern industrial productions, where batch size one and a high diversity of variants are typical.

1. Introduction

Industrial geared motors are often complex and freely or modularly configured components. This leads to a high diversity of variants, which rises continuously. Additionally, the batch size decrease, so that often a batch size one production is required. Therefore, it is hard for the assembly worker to evaluate the product quality without an automated end-of-line test. Firstly, this paper describes the challenges of an end-of-line test in a highly variant production and presents how geared motors can be analyzed in general. This includes a discussion of usable sensor types, feature creation, and fault diagnosis, which requires two machine learning (ML) use cases. These are the anomaly detection, to identify whether the geared motor is healthy, and the fault classification, which classifies the type of anomaly. Furthermore, a solution is presented of how to augment the concept for an end-of-line test in a highly variant production. Therefore, further ML techniques are proposed, including clustering, data augmentation, and transfer learning.

2. Characteristics of the End-of-Line Tests

Constant product quality is important. Therefore an end-of-line test is used to evaluate, whether a product is of high quality or not. In a batch-size-one production with a high diversity of variants this test is often carried out by the assembly worker. During a functional test, this worker must hear and feel faults the product may have. This could be difficult in a rough industrial environment. Firstly, it is determined whether the quality criteria are met. For products, which do not pass the test a decision is made whether the product will be repaired or sorted out. But there is no time planned for extra evaluation and it is often hard for the worker to evaluate. This decision mainly depends on the type of fault, its strength, and its location considering the repair time, material supply, and urgency. The test result is tracked for quality management reasons, including the type of fault. This is important to avoid further faults of the same type as well as to detect serial faults. Additionally, some customers demand good sound quality and quantify it.

To guarantee consistent quality, parts of the test must be automated, especially the measurement and fault diagnosis. The according quantification is achieved by using data-based methods. To summarize, an automated end-of-line test must provide, whether the quality is met and if not which fault has occurred. Automation of the end-of-line test leads to several challenges described below.

2.1. Challenges in Measurement

Vibration measurements are used for fault diagnosis as state-of-the-art signals. Acoustic signals can often be used in the same way as vibration signals. To achieve cost optimization, the time to prepare a motor for measurements must be as short as possible, but common vibration and acoustic measurements take a lot of time. Several methods are available to connect a vibration sensor head. However, those connections have repeatability problems with different types of motors. Often no unique connection rule can be defined or it requires plenty of installation time. Furthermore, many fault diagnosis methods are based on high frequencies. The maximum transfer frequency depends on the vibration sensor connection and is typically limited between 1 kHz and 20 kHz. Equipment for higher frequencies is very expensive and it is mainly used for laboratories. Some contact-free, laser-based measuring techniques exist. But position-based techniques, such as triangulation have resolution problems in the higher frequency range (Soave, D'Elia, & Mucchi, 2020). Acoustic measurements are an obvious choice, since quality issues reported by a customer are often associated to sound quality. The authors found out that the main problem in using acoustic signals is the rough acoustic environment in a typical production hall. The shielding of the end-of-line test from the production environment by using a chamber is very complex and requires a lot of time during production.

2.2. Challenges in Fault Diagnosis

Common end-of-line tests work for minor or major series production with ML. Often literature assumes, that all examples are associated with the same type of motor. In addition, for different fault types, different methods are evaluated to be the best. (Gangsar & Tiwari, 2020) The fault diagnosis must be set up separately for each product and fault, which is not feasible considering a high diversity of variants. Each dataset of a new configuration starts without examples and slowly increases. This leads to small datasets for both, examples with and without faults. In complex systems not all types of faults are well known, so anomaly detection is required that detects both, known and unknown faults.

As the target of fault classification is to identify the type of fault, it needs labeled data for all types. It can be assumed, that for each class there is a different number of labeled examples but sometimes none. Therefore, faults cannot be classified until they occur on this product often enough for model learning. In the target of fault classification, data behavior is much more complex with respect to different configurations, fault types, and intensities of faults. Additionally, the fault diagnosis must allow unidentified classes, which many classifiers do not support. It can be assumed that nowadays a single model cannot handle this complexity. When the number of labeled examples increases, one likes to update the classifier, but to do so it would be necessary to learn and evaluate all classes again. The result indicates to the worker, where to look first in order to find a possible fault. It would be helpful if the classifier gives a metric of how strongly the fault is associated with its class, for example a probability.

Many common ML methods work with features. It is often discussed, which are the best features to use for fault diagnosis. But it is rarely shown, how to handle multiple and perhaps unknown faults or changing environmental conditions like noise or vibration, which does not come from the test motor. Additionally, it is rarely shown in literature how to identify features, which describe motor faults in general and not only for a specific variant. Those research gaps make it difficult to decide, which features or if any should be used. In addition, a priori knowledge has to be provided automatically and is sometimes not available.

2.3. Definition of an End-of-Line Test Approach

A test system is required, which provides a reproducible test run. The test assumes steady operation of the motor or several cycles with steady operation, for example driving the motor at nominal speed for several seconds in both directions. During inspection of geared motors with a high gear ratio some errors are only present once every few seconds. The test should include more than two full cycles of the slowest shaft. The sensor measurement starts after the motor has reached nominal speed. The proposed concept works for nearly all types of vibration and acoustic sensors. Additionally, motor signals, such as current and velocity, are expected to work in principle (Jigyasu, Sharma, Mathew, & Chatterji, 2018).

In real-world problems labeled datasets are often not available before the data-based system is used. Then firstly only data is collected. Anomaly detection can be used to circumvent the problem of missing labeled datasets. This is referred to as semi-supervised learning and requires only data, which is labeled as normal. As mentioned in (Olivotti, Passlick, Axjonow, Eilers, & Breitner, 2018) all detected anomalies should be labeled by the worker, who is responsible for the product quality, so that labeled data for classification can be collected. If pseudo anomalies occur they will be labeled as well, so that the classification can handle them. Some faults are only present if the motor is under load. If it is tested without load, the recorded data has to be labeled by the service department.

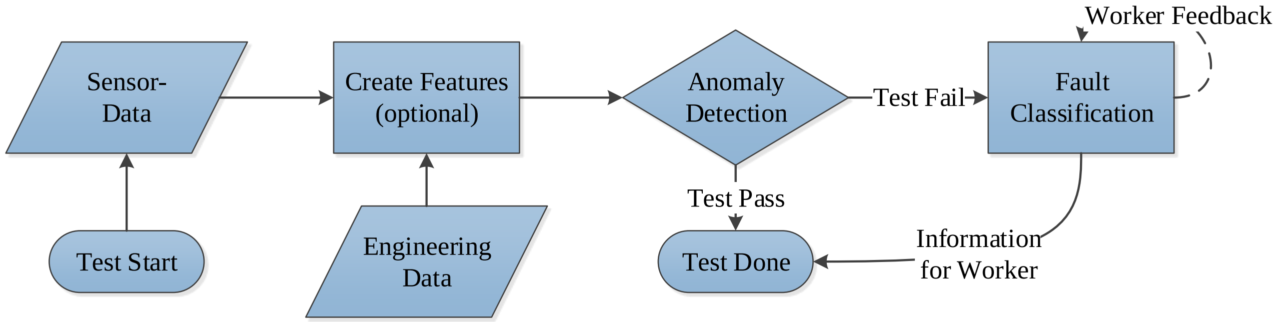

Figure 1. Fault diagnosis in an end-of-line test.

Figure 1 shows the proposed architecture with a pre-trained system. The creation of features is the first step after the measurement, which is optional. Then the anomaly detection indicates, whether the test is passed or not. If not, the fault classification indicates, which fault has occurred. Once a fault class is well trained, it indicates the true class to the worker, so only the unknown examples must be labeled. Often different names are feasible for the same fault so that the user interface should propose names to the worker.

3. Possible Solution Methods

In the next sections, it is discussed how the necessary steps for this end-of-line test can be set up with regard to the challenges mentioned above. Furthermore, the authors propose an augmentation of this concept, which enables it to be used in highly variant production.

3.1. Acoustic Measurement

Acoustic signals are a compelling alternative and sometimes reach better results than vibration signals for bearing (Grebenik, Zhang, Bingham, Srivastava, & others, 2016) or gear faults (Dhami, Pabla, & others, 2018). High-frequency areas can be used with nearly every microphone. To set up the motor for measurement it is only necessary to position it under the microphone. Microphones with directive efficiency can be used to reduce the problem of environmental sounds. Array microphones with calculation of the local sound are of rising interest in production environments (Benko, Tinta, & Mussiza, 2005). The amount of the directive efficiency is associated with the width of the array. In case of cardioid microphones, the directive efficiency depends on its length. Commonly used versions of those types work better within high-frequency areas. In case of using micro-electro-mechanical sensors, known as MEMS, array microphones may have the advantage of longer calibration periods. For both techniques, versions for laboratory and productive usage exist.

3.2. Feature Creation

Feature creation for fault diagnosis in geared motors is well described in literature (Henriquez, Alonso, Ferrer, & Travieso, 2013), but not always easy to use. Both, features based on engineering data are discussed as well as those calculated without a priori knowledge. The features can be summarized in five areas. Their main advantages in fault diagnosis are shown in Table 1. The authors conclude to use a mix of more than one category of features to guarantee a robust system with a maximum of self-learning capability on different product types. Neural networks (NN) can be used for direct computation on raw signals, allowing the step of feature creation to be skipped. However, if one chooses not to use feature creation, the behavior of the overall system is similar to a system that uses artificially learned features. But it may have disadvantages in terms of the resources required.

1. The area of statistical features can generalize the behavior of sensor-data at a stationary working condition, so that every type of fault can be addressed without a priori knowledge of the faults and the motor. Often, kurtosis is used to describe if a signal is superimposed by impulses (Wei, Li, Xu, & Huang, 2019) and therefore directly linked to anomaly detection.
2. Features of the second area, the fault-associated features, can show the behavior of impulses. In literature, it is often discussed how to identify faults in that way. An overview of features for motor faults can be found in early literature (Finley, Hodowanec, & Holter, 1999). Most features in the second area describe fault frequencies and are thus dependent on the basic velocity. Discussing the behavior of the multiple of the basic frequency is called order analysis. If the signal is present in the order spectrum, it will be easier to compare different types of motors. However, in case of mechanical faults, it is necessary to separate the order spectrum for each shaft of a geared motor.
3. To address the sound quality itself, psychoacoustic features are discussed, which provide a basis for

simulating human hearing. Those features show better or nearly the same accuracy in classification as statistical features (Kane & Andhare, 2020) and thus can generalize the signal. Here the target of psychoacoustic is to generate features that match the worker's "ear"-evaluation as well as possible.

4. Model-based features from simulation models mostly require a priori knowledge of the motor and often special identification tools on the power train. On the other hand, they are mostly suitable to avoid data gaps in the discussion of different motor types. Modern converters, for example, can be used to identify the mechanical and electrical system behavior such as the electric equivalent circuit or the mass oscillator.
5. Dealing with features that describe the system behavior but without a priori knowledge leads to feature learning. Here a NN trains features with a good representation outperforming statistical features (Mao, He, Tang, & Li, 2018). Typical for those artificially learned features is the usage of autoencoders to train the dataset and to exclude the inner state (Mao., He, & Zuo, 2019), a technique, which is discussed in the next section of anomaly detection. Pretrained image processing NNs can be used to exclude likely better features when using time-frequency signals (Müller, Ritz, Illium, & Linnhoff-Popien, 2020).

Table 1. Characteristic of Features.

Area of Feature	**Characteristic**
Statistical features	Generalization without a priori knowledge
Fault associated features	Description of exactly one type of faults
Psychoacoustic features	Generalization of the sound quality
Model-based features	Description of system behavior
Artificially learned features	Learns system behavior

Depending on the type of feature, a signal transformation can be useful. Transformation in the frequency domain, like for example Fourier- or Hilbert transform, eliminate time-shifts. Bearing and gear faults can be shown more clearly by using an envelope spectrum. If the test is done in a non-stationary condition, it could be feasible to use signals in the time-frequency domain. Instead of an envelope spectrum, the Hilbert-Huang-transform can be used (Mao, Zhang, Tian, & Tang, 2020). For acoustic signals, Mel-frequency cepstral coefficients, known as MFCC are of rising interest to represent the behavior of signals in industrial environments (Benkedjouh, Chettibi, Saadouni, & Afroun, 2018). These methods transform and filter one time slot at a time like a short-time Fourier transform does. It could be useful to discuss the time-frequency domain in the case of measurements with a length of several seconds, too. Since these signals have three dimensions, the features must be evaluated appropriately.

3.3. One-Class-Classifier for Anomaly Detection

A complete overview of anomaly detection methods is out of scope of this work but can be found in literature (Chandola, Banerjee, & Kumar, 2009). To reduce the complexity of fault diagnosis, this work focusses on one-class classifiers and uses only one type of ML methods and one type of NNs. An ensemble of these one-class classifiers can be used for fault classification (Carino, et al., 2018), so that each type of fault will be classified by its model. Using one-class support vector machines (SVM) is common (Khan & Madden, 2014) and can be used with hand-crafted features in an end-of-line test (Leitner, Lagrange, & Endisch, 2016). But also a reduction of the raw data can be used as input (Fernández-Francos, Martínez-Rego, Fontenla-Romero, & Alonso-Betanzos, 2013). This method has the advantage of working with features that have already been calculated, so there is no need to store raw data for learning. Some literature shows, that neuronal networks can detect faults better than common ML algorithms like the SVM (Sun, Wyss, Steinecker, & Glocker, 2014). However, ML has advantages on small datasets (Esakimuthu Pandarakone, Mizuno, & Nakamura, 2019), which are often found on a highly variant production.

For the case of using NNs, there are specific surveys (Chalapathy & Chawla, 2019). Often autoencoders are used to calculate a difference between the measured signal and a feasible good one. During training, the autoencoder learns its inner state to reconstruct the same data that is given as input to the output. The inner layer of the network between encoder and decoder is chosen very small, so that the autoencoder must learn, what the most necessary information to reconstruct the input is. The difference between the true signal at the input and the reconstruction at the output is used as a metric for anomaly detection. Statistical features of this error can be used for thresholds. Since it has not been determined which fault is present, the error signal and the original data can be used for the following classification. The architecture of the NN can be selected according to the input signal used. For three-dimensional data in the time-frequency domain, convolutional layers are typical, while for time-domain signals long-short-term-memories named LSTM are used (Chalapathy & Chawla, 2019). Fully connected layers are not specified to an input type, but perform well, too (Principi, Rossetti, Squartini, & Piazza, 2019). In the end, time or frequency domain signals can be used as input vectors in all architectures (Huang, Chen, & Huang, 2019), (Zhang, Peng, Li, Chen, & Zhang, 2017). Additionally, deep architectures with more than one hidden layer can learn complex system behavior more correctly (Principi, Rossetti, Squartini, & Piazza, 2019).

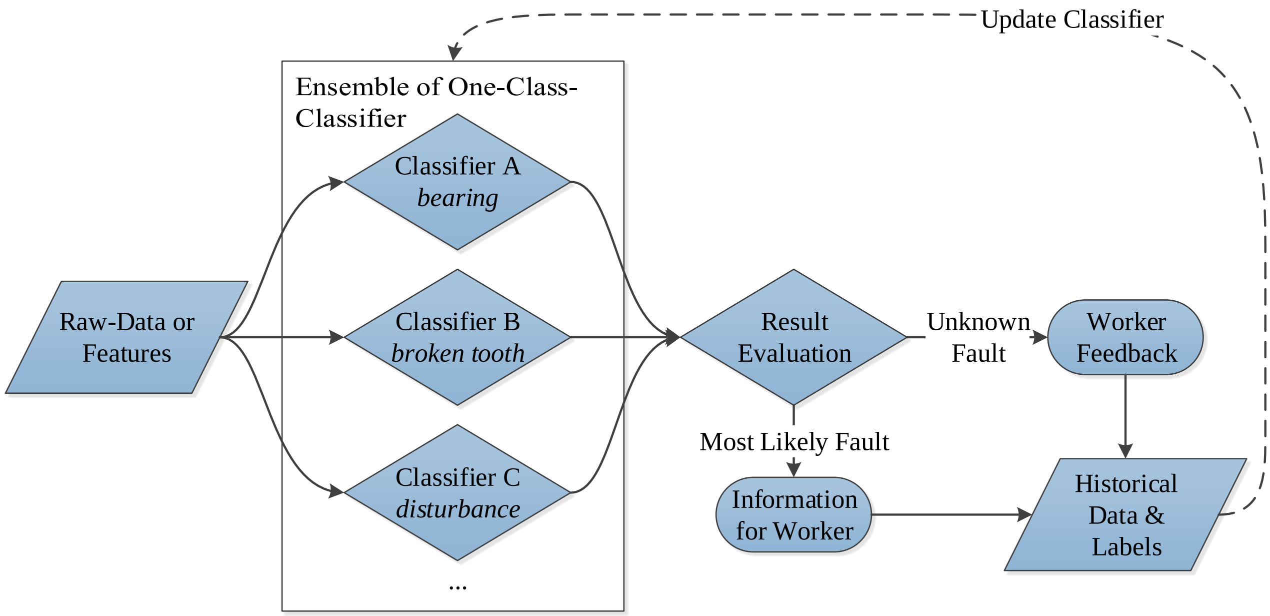

Figure 2. Ensemble of one-class-classifier for fault classification.

A high sampling rate results in a huge sample vector of raw data, which leads to complex computations of NNs (Sun, Wyss, Steinecker, & Glocker, 2014). Therefore, often only a short period of time is used (Shao, Jiang, Lin, & Li, 2018), (Huang, Chen, & Huang, 2019), but to be sure the data includes all possible faults, several of these periods need to be observed (Yang, et al., 2020). The use of a filter reduction in combination with a time-frequency transformation is also a viable way, for example in MFCC computation (Principi, Rossetti, Squartini, & Piazza, 2019).

3.4. Ensemble of One-Class-Classifier

The use of such an ensemble solves many of the issues mentioned above. While most literature discusses only one type of fault the concept shown in Figure 2 allows to identify multiple faults without explicit training of them. The features or even the raw data is sent to all classifiers, which estimate whether the data depends to its class or not. The results must be evaluated if only one fault was found or if many classifiers declare a fault and which fault is most likely. This result is given to the worker or if the class could not be classified he is asked to label it. A classifier can be trained and updated whenever enough data is historically stored without taking care of other classes. At least it is a good idea to include an additional classifier to detect acoustic disturbances from the environment, the acoustic measurement could not ignore. Another autoencoder could be trained for novelty detection, which solves the issue of unknown fault classes (Yang, et al., 2020). Choosing the right classifier for the ensemble is not easy, but it is guided (Krawczyk & Woźniak, 2014). As in the case of anomaly detection, the SVM and the autoencoders (Shao, Jiang, Lin, & Li, 2018) are common. A further advantage of this concept is, that for every class it is possible to choose the best method and there is no need to commit to one method that fits all faults. In the end, it would not be necessary to decide for one classifier at all, but using more than one classifier for the same type of fault makes an information fusion necessary to combine the results to a unique one.

3.5. Diversity of Variants

To augment the concept for production lines with a high diversity of variants, all steps must be adapted. As already mentioned, one instance of the fault diagnosis methods is needed for each product configuration. Once, all the labeled data is available to train the models, an enterprise resource planning system (ERP) can be used to hold and match configurations and models, as shown in green in Figure 3. To use the fault diagnosis methods discussed before, the models must be relearned or the existing model can be updated, which is referred to as transfer learning. To do so, all sensor-data during the test and its worker's feedback must first be saved as historical data according to the configuration. And, if already calculated, the feedback of the fault diagnosis must additionally be stored. If one wants to reduce the storage space and does not need to use raw data in NNs, only the calculated features will be stored historically. Then the data that match the product configuration must be selected and the model must be trained. The steps for storing and using historical data are marked for cloud platform in yellow in Figure 3. If only one production line exists, those steps can be done on the tester itself. And if only one manufacturing location exists, those steps can be done in the local cloud.

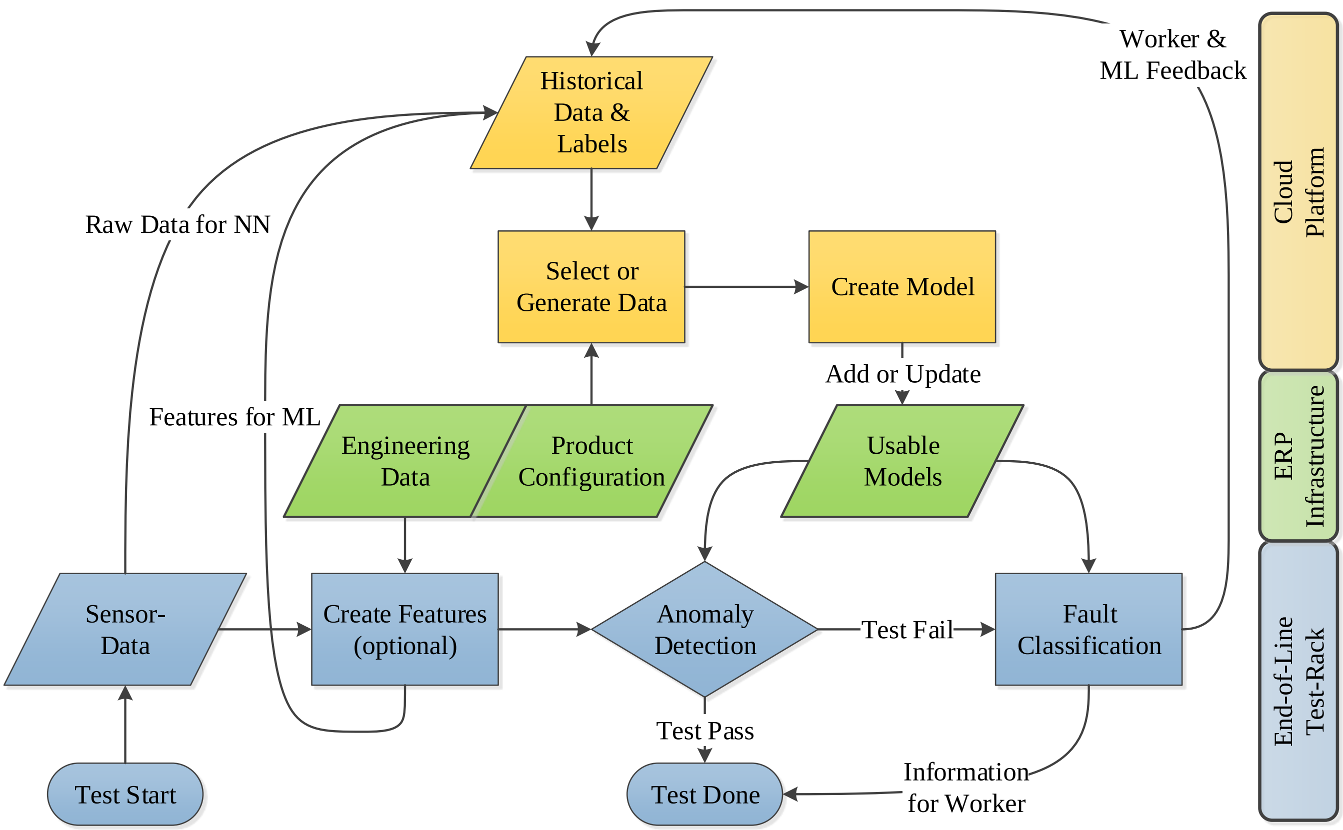

Figure 3. Fault diagnosis in a highly variant production.

Using data from all around the world in the cloud allows increasing the availability of examples. If data drifts over time or between assembly lines exists, they may be solved by using adaptation techniques, likewise (Lin, Deng, Kuo, & Chen, 2019).

Up to this point, some products cannot be tested. Those products have not been produced often enough for model learning. Several approaches can be used to solve this problem. Data augmentation can help to increase the number of examples, if already some exist. If not, it would be helpful to identify configurations that are almost similar to the unknown one and can be used for training as well. This could be done model-based or data-based. If not enough a priori knowledge exists, clustering methods can help to identify groups of similar configurations. Often similar configurations do not exist to train robust fault diagnosis. Then artificial intelligence is needed, that can calculate how the raw data or the features will be on the new configuration. This intelligence could be a NN with the ability to generate data, which is referred to as generative nets.

After data for model creation of a specific configuration is provided, two approaches can be used to get the new model for this specific configuration. Just train a new model from scratch or adapt the existing model to solve the same problem on another data domain. This is called domain adaptation, which is a transfer learning problem and has the advantages of faster learning and the need for fewer examples.

3.6. Similarity Clustering

Identifying groups of configurations with similar data behavior is referred to as clustering. Common ML methods use the features already discussed. Depending on the method it is required either to know the number of clusters, which is very difficult in this problem, or the required density in data representation. To be sure the fault diagnosis performs well on the combined data one will require a density in the cluster that is larger than the density in the original class of target configuration. To use those methods with raw data or its frequency representation several techniques can be compared with and without scaling to a lower-dimensional space (Hennig, Grafinger, Gerhard, Dumss, & Rosenberger, 2020). If one likes to use NNs for clustering, self-organizing maps, known as SOM can be used (Germen, Başaran, & Fidan, 2014). Clustering using a SOM can help to label not fully labeled datasets for classification. This occurs because the worker cannot always find proper labels for all detected anomalies (Oh, Jung, Jeon, & Youn, 2017). In the end, the topic of clustering and its advantage in fault diagnosis is much more complex and needs to be considered more closely.

3.7. Data Augmentation

In ML it is common to use mechanisms to increase the number of examples, named data augmentation. A simple but useful method is to simply copy the existing data and add some noise. Another common technique is to calculate the distance between two similar examples and return something in between as a new sample. Both methods work well on features but are critical to use on raw data in case of time shift and other data behaviors. Different data augmentation techniques for time series lead to different performance losses in classification (Li, Zhang, Ding, & Sun, 2020). In the last years a new type of NN, the generative net was introduced to learn the data behavior and create new examples.

Generative Nets

In general, generative NNs have the target to generate continuously changing data but are always related to a specific real-world dataset. Another NN is used to evaluate if an example is original or generated, called a discriminator. If the discriminator cannot find it out, the generative net is well trained. This concept is known as a generative advertising network (GAN). In literature, GANs are often used to increase a dataset for classification (Shao, Wang, & Yan, 2019) or are part of transfer learning approaches as considered below (Schockaert & Hoyez, 2020). In principle, GANs could be useful to be trained by various configurations and be able to provide signals or features for a specific and new configuration. As far as the authors are aware, this is not described in literature and it is necessary to discuss detailed architecture and requirements for the input dataset. First examples in artificially acoustic data generation are developed for realistic music generation (Dieleman, Oord, & Simonyan, 2018).

3.8. Domain Adaptation

Domain adaptation in general is an area of transfer learning. Here the method which can solve a specific task on a specific dataset called the source domain is retrained to solve the same task for another dataset, the target domain. Various types of domain adaptation for fault diagnosis methods can be found in the literature, including those described previously. Domain adaptation for fault diagnosis can be divided into categories for the motivation and settings for the data problem (Zheng, et al., 2019). This paper discusses the motivation of leveraging from different but related machines. But also the motivations of different fault degrees or incomplete information on a dataset like incomplete labels could be helpful in the case of an end-of-line test. The domain adaptation approaches can influence the fault diagnosis at several instances, all used in machinery diagnosis. One possibility is to retrain the classifier itself (Jiao, Zhao, Lin, & Ding, 2019), another is to use statistical information of the target data to adapt parameters inside (Zhang, Peng, Li, Chen, & Zhang, 2017). In contrast, the features can also be transformed instead (Mao, Zhang, Tian, & Tang, 2020).

Often there are no target domain examples available for anomaly detection, so for adaptation, a solution of data augmentation is necessary. Fault examples are rare, but the classification is only done after an anomaly is detected. It can be assumed that mostly enough examples without faults are historically stored or created by one of the above-mentioned methods. So in this setting only samples labeled as normal are available. For the motivation of related machines, an autoencoder can be used. First, features for the given data are learned and then a generative net learns a transformation from data with normal behavior to data with fault behavior (Michau & Fink, 2019). This data is used to train the fault diagnosis as a one-class classification. For limited labeled datasets a NN trained on simulation data can be adapted by adding a transfer layer performing well on real-world data (Xu, Sun, Liu, & Zheng, 2019).

For the motivation of leveraging from different but related machines in total nine papers are referred, but for the motivation of different working conditions there are 44 references (Zheng, et al., 2019). This is to be expected, since such experimental data are much easier to obtain. One can assume, that different working conditions like higher torque and a motor configuration with higher related torque could be discussed by the same domain adaption methods.

4. Next Steps of Research

Many steps of the proposed concept are well described in literature but some open points exist. The robustness against environmental sounds is considered for a specific combination of sensors, optional features, and anomaly detection. Therefore, datasets with and without acoustic disturbances are recorded with typical sounds of an industrial environment. These include, but are not limited to, pneumatic and electric screwdrivers, hammer blows, ventilation systems, logistic vehicles, and speech. The target is then to find the most robust overall system. Additionally, methods aimed to detect these sounds can improve this system by suggesting whether or not the motor should be retested.

In the next step it is considered, which features or methods are most variant-independent and can therefore be used on smaller configuration differences without domain adaptation. Therefore, datasets are collected with smaller and wider configuration differences. At the same time, it is observed, which configuration differences would cause the most trouble. Based on this expertise, clustering methods are trained with the target of finding configurations, which can be tested with the same models. Therefore, the question to be answered is how to describe the similarity of configurations. Those experiments require extensive data collections, which can only be collected by a test system that is already included at the assembly line.

The same methods are also used to find similar datasets, which can be a good starting point for transfer learning. Domain adaptation promises high advances in transferring from a well-known configuration to an unknown one. The literature describes several adaptation methods used in condition monitoring, but not transferring between variants. It must be considered, which domain adaptation approach is best for an automated test system. However, this may not be the approach that is best for two specific configurations. Further work will help to close this research gap.

5. Conclusion

End-of-line tests are a fundamental element of modern production lines, but common systems are not applicable for every use case. To set up such a system in a highly variant production with batch size one several steps need to be discussed. Acoustic arrays or microphones with directional efficiency fulfill both, the requirements of a flexible production and the data quality at the same time. For fault diagnosis different combinations of features and ML tools are possible. For a specific manufacturer, the best overall system need to be evaluated individually. For that purpose a guideline is given. Literature shows that ML techniques can benefit both, anomaly detection and fault classification. One-class SVMs and autoencoders are pointed out as fundamental techniques for fault diagnosis in an end-of-line test.

This paper proposes a concept for handling high diversity of variants in end-of-line tests, which is rarely described in literature. An information and analysis infrastructure is necessary to bring the ERP System and the cloud solution together with the test rack. Given the configuration of the products, clustering techniques can help to find those configurations that can be tested with the same model. To train those models for small datasets, augmentation techniques can improve the solution. For all mentioned application scenarios, common and easily manageable features and ML methods can be used. In comparison using NNs can improve the results and open the door for transfer learning.

Acknowledgement

Parts of this article were supported by the Ministry of Economic Affairs, Innovation, Digitization and Energy of North Rhine Westphalia through the innovation project Machine Learning for the Production and its Products (ML4Pro²) of the excellence cluster itsOWL (Intelligente Technische Systeme OstWestfalenLippe).

References

Benkedjouh, T., Chettibi, T., Saadouni, Y., & Afroun, M. (2018). Gearbox fault diagnosis based on mel-frequency cepstral coefficients and support vector machine. *IFIP International Conference on Computational Intelligence and Its Applications*, (S. 220–231).

Benko, U., Tinta, D., & Mussiza, B. (2005). *Using Microphone Array for Fault Detection Constant Directivity Beamforming.*

Carino, J. A., Delgado-Prieto, M., Iglesias, J. A., Sanchis, A., Zurita, D., Millan, M., . . . Romero-Troncoso, R. (2018). Fault detection and identification methodology under an incremental learning framework applied to industrial machinery. *IEEE access, 6*, 49755–49766.

Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407*.

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR), 41*, 1–58.

Dhami, S. S., Pabla, B. S., & others. (2018). Gear fault classification using Vibration and Acoustic Sensor Fusion: A Case Study. *2018 Condition Monitoring and Diagnosis (CMD)*, (S. 1–6).

Dieleman, S., Oord, A. v., & Simonyan, K. (2018). The challenge of realistic music generation: modelling raw audio at scale. *arXiv preprint arXiv:1806.10474*.

Esakimuthu Pandarakone, S., Mizuno, Y., & Nakamura, H. (2019). A comparative study between machine learning algorithm and artificial intelligence neural network in detecting minor bearing fault of induction motors. *Energies, 12*, 2105.

Fernández-Francos, D., Martínez-Rego, D., Fontenla-Romero, O., & Alonso-Betanzos, A. (2013). Automatic bearing fault diagnosis based on one-class ν-SVM. *Computers & Industrial Engineering, 64*, 357–365.

Finley, W. R., Hodowanec, M. M., & Holter, W. G. (1999). An analytical approach to solving motor vibration problems. *Industry Applications Society 46th Annual Petroleum and Chemical Technical Conference (Cat. No. 99CH37000)*, (S. 217–232).

Gangsar, P., & Tiwari, R. (2020). Signal based condition monitoring techniques for fault detection and diagnosis of induction motors: A state-of-the-art review. *Mechanical Systems and Signal Processing, 144*, 106908.

Germen, E., Başaran, M., & Fidan, M. (2014). Sound based induction motor fault diagnosis using Kohonen self-organizing map. *Mechanical Systems and Signal Processing, 46*, 45–58.

Grebenik, J., Zhang, Y., Bingham, C., Srivastava, S., & others. (2016). Roller element bearing acoustic fault detection using smartphone and consumer microphones.

Hennig, M., Grafinger, M., Gerhard, D., Dumss, S., & Rosenberger, P. (2020). Comparison of Time Series Clustering Algorithms for Machine State Detection. *Procedia CIRP*, *93*, 1352–1357.

Henriquez, P., Alonso, J. B., Ferrer, M. A., & Travieso, C. M. (2013). Review of automatic fault diagnosis systems using audio and vibration signals. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *44*, 642–652.

Huang, Y., Chen, C.-H., & Huang, C.-J. (2019). Motor fault detection and feature extraction using rnn-based variational autoencoder. *IEEE Access, 7,* 139086–139096.

Jiao, J., Zhao, M., Lin, J., & Ding, C. (2019). Classifier Inconsistency-Based Domain Adaptation Network for Partial Transfer Intelligent Diagnosis. *IEEE Transactions on Industrial Informatics, 16*, 5965–5974.

Jigyasu, R., Sharma, A., Mathew, L., & Chatterji, S. (2018). A review of condition monitoring and fault diagnosis methods for induction motor. *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, (S. 1713–1721).

Kane, P. V., & Andhare, A. B. (2020). Critical evaluation and comparison of psychoacoustics, acoustics and vibration features for gear fault correlation and classification. *Measurement*, *154*, 107495.

Khan, S. S., & Madden, M. G. (2014). One-class classification: taxonomy of study and review of techniques. *The Knowledge Engineering Review, 29*, 345–374.

Krawczyk, B., & Woźniak, M. (2014). Diversity measures for one-class classifier ensembles. *Neurocomputing, 126*, 36–44.

Leitner, L., Lagrange, A., & Endisch, C. (2016). End-of-line fault detection for combustion engines using one-class classification. *2016 IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*, 207–213.

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

Lin, C.-C., Deng, D.-J., Kuo, C.-H., & Chen, L. (2019). Concept drift detection and adaption in big imbalance industrial IoT data using an ensemble learning method of offline classifiers. *IEEE Access, 7*, 56198–56207.

Mao, W., He, J., Tang, J., & Li, Y. (2018). Predicting remaining useful life of rolling bearings based on deep feature representation and long short-term memory neural network. *Advances in Mechanical Engineering, 10*, 1687814018817184.

Mao, W., Zhang, D., Tian, S., & Tang, J. (2020). Robust Detection of Bearing Early Fault Based on Deep Transfer Learning. *Electronics, 9*, 323.

Mao., W., He, J., & Zuo, M. J. (2019). Predicting remaining useful life of rolling bearings based on deep feature representation and transfer learning. *IEEE Transactions on Instrumentation and Measurement, 69*, 1594–1608.

Michau, G., & Fink, O. (2019). Domain adaptation for one-class classification: monitoring the health of critical systems under limited information. *International Journal of Prognostics and Health Management, 10*, 11.

Müller, R., Ritz, F., Illium, S., & Linnhoff-Popien, C. (2020). Acoustic anomaly detection for machine sounds based on image transfer learning. *arXiv preprint arXiv:2006.03429*.

Oh, H., Jung, J. H., Jeon, B. C., & Youn, B. D. (2017). Scalable and unsupervised feature engineering using vibration-imaging and deep learning for rotor system diagnosis. *IEEE Transactions on Industrial Electronics*, *65*, 3539–3549.

Olivotti, D., Passlick, J., Axjonow, A., Eilers, D., & Breitner, M. H. (2018). Combining machine learning and domain experience: a hybrid-learning monitor approach for industrial machines. *International Conference on Exploring Service Science*, (S. 261–273).

Principi, E., Rossetti, D., Squartini, S., & Piazza, F. (2019). Unsupervised electric motor fault detection by using deep autoencoders. *IEEE/CAA Journal of Automatica Sinica, 6*, 441–451.

Schockaert, C., & Hoyez, H. (2020). MTS-CycleGAN: An Adversarial-based Deep Mapping Learning Network for Multivariate Time Series Domain Adaptation Applied to the Ironmaking Industry. *arXiv preprint arXiv:2007.07518*.

Shao, H., Jiang, H., Lin, Y., & Li, X. (2018). A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders. *Mechanical Systems and Signal Processing, 102*, 278–297.

Shao, S., Wang, P., & Yan, R. (2019). Generative adversarial networks for data augmentation in machine fault diagnosis. *Computers in Industry, 106*, 85–93.

Soave, E., D'Elia, G., & Mucchi, E. (2020). A laser triangulation sensor for vibrational structural analysis and diagnostics. *Measurement and Control, 53*, 73–82.

Sun, J., Wyss, R., Steinecker, A., & Glocker, P. (2014). Automated fault detection using deep belief networks for the quality inspection of electromotors. *tm-Technisches Messen, 81*, 255–263.

Wei, Y., Li, Y., Xu, M., & Huang, W. (2019). A review of early fault diagnosis approaches and their applications in rotating machinery. *Entropy, 21*, 409.

Xu, Y., Sun, Y., Liu, X., & Zheng, Y. (2019). A digital-twin-assisted fault diagnosis using deep transfer learning. *IEEE Access, 7*, 19990–19999.

Yang, Z., Gjorgjevikj, D., Long, J.-Y., Zi, Y.-Y., Zhang, S.-H., & Li, C. (2020). Sparse Autoencoder-Based Multi-Head Deep Neural Networks for Machinery Fault Diagnostics With Novelty Detection.

Zhang, W., Peng, G., Li, C., Chen, Y., & Zhang, Z. (2017). A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors*, *17*, 425.

Zheng, H., Wang, R., Yang, Y., Yin, J., Li, Y., Li, Y., & Xu, M. (2019). Cross-domain fault diagnosis using knowledge transfer strategy: A review. *IEEE Access*, *7*, 129260–129290.

BIOGRAPHIES

Peter Wissbrock is currently working as a researcher at the Innovation Department of Lenze SE. He obtained his BSc and MSc in electrical engineering at the OWL University of Applied Sciences and Arts, Germany. He is pursuing Ph.D. from Leibniz University Hannover, Germany in field of industrial analytics. He works in domain of drive train fault diagnosis, machinery anomlay detection and data acquisition infrastructure.

David Pelkmann is a research assistant in the area of machine learning for industrial applications at the Bielefeld University of Applied Sciences, Germany. There he received his B.Eng. in engineering informatics and is currently studying for his M.Eng. in digital technologies. His research interests in the industrial environment are acoustic quality and condition assessment of machinery and document retrieval systems.

Björn Tölle obtained his BSc and MSc in mechanical engineering at the OWL University of Applied Sciences and Arts, Germany. The main research focus is using machine learning methods in order to detect faults of motors by the means of audio signals.