Comparison of Different Excitation Strategies for Fault Diagnosis of Belt Drives: Industrial Application Scenarios

Moritz Fehsenfeld[1] [a], Johannes Kühn[2], Zygimantas Ziaukas[1] [b] and Hans-Georg Jacob[1] [c]

[1]*Leibniz University Hannover, Institute of Mechatronic Systems, An der Universität 1, Garbsen, Germany*

[2]*Lenze SE, Hameln, Germany*

Keywords: Fault Diagnosis, Machine Learning, Industrial Application, Belt Drives, Mechatronics Systems.

Abstract: Machine learning (ML) has received a lot of attention in solving fault diagnosis (FD) tasks. As a result, more and more advanced machine learning algorithms have been developed to increase accuracy. But the system's excitation has likewise a high impact on the diagnosis performance and applicability. For this purpose, we describe different industrial application scenarios and the related set trajectory. They are divided into passive FD, where normal operation data serves as the input, and active FD, where an optimized excitation is injected. All scenarios are investigated concerning achievable accuracy and data requirement based on comprehensive measurements. We demonstrate that in active scenarios a high accuracy of 97.6 % combined with a small number of measurements are obtained by very basic algorithms like a one-nearest neighbor with Euclidean distance. In passive scenarios, where the FD task is generally harder, the demand for large datasets and more advanced ML methods increases. In this way, we illustrate how intelligent use of an optimized excitation strategy leads to feasible, reliable, and accurate fault diagnosis with a broad industrial application spectrum.

1 INTRODUCTION

Fault diagnosis (FD) has seen increasing attention in the last years. It has the potential to recognize faults at an early stage and guarantee optimal operation conditions. By this, downtime is decreased, maintenance costs are reduced and safe operation is ensured. Following this trend, many FD applications of components in electromechanical motion systems have been published in the past. Besides the motor as a major focus (Kande et al., 2017) other drive elements such as bearings (AlShorman et al., 2020) and gears (Sharma and Parey, 2016) are likewise subject to FD.

Belts are a popular drive solution with a wide variety of applications. A proper belt pretension is inevitable to operate with high efficiency and low wear. It is adjusted while commissioning but decreases during operation due to changing environments and wear. For that reason, continuous pretension monitoring ensures optimal working conditions. Nevertheless, FD applications targeting belt drives are rare. (Kang et al., 2018) predict belt cracks and (Hu et al., 2016)

---

[a] https://orcid.org/0000-0003-2639-7838

[b] https://orcid.org/0000-0001-9161-0709

[c] https://orcid.org/0000-0001-5605-9704

monitor belt oscillations. The majority of tension monitoring applications utilize external sensors like strain gauges (Musselman and Djurdjanovic, 2012; Bzinkowski et al., 2022) or optical lasers (Khazaee et al., 2017). These additional sensors are often undesired in practice given their extra costs and commissioning effort. (Picot et al., 2017) analyze motor current which does not necessitate extra equipment to discriminate four tension levels. Consequently, our proposed FD system to detect a faulty belt tension relies only on standard sensors.

Many contributions in the field of FD focus on the diagnosis methods. Recently, machine learning (ML) has attracted great attention. Procedures are categorized into conventional ML and deep learning (DL) approaches. Conventional ML follows a two-stage procedure. During feature engineering different signal processing techniques are applied to facilitate diagnosis. These features are handed over to a conventional classification method such as a support vector machine (Gangsar and Tiwari, 2017) or random forest (Toma et al., 2020). Meaningful features are crucial for successful FD, however, require high domain knowledge. DL approaches omit extensive feature extraction but follow an end-to-end approach where fea-

tures are learned during training. An outline is provided by (Thoppil et al., 2021).

Besides the choice of algorithm, the input data is even more important. A high information content about the considered task within the data is a basic prerequisite for successful FD. The selection of appropriate sensors has already been addressed. Motor current signature analysis (MCSA) and vibrational analysis are popular among other FD approaches for electrical drives. An overview of techniques is given by (Nandi et al., 2005). The restriction to standard sensor technology does not leave much choice. In this context, the motor's motion during diagnosis has not been considered yet. Generally, passive and active FD is distinguished with respect to the set motion. Passive FD takes normal operation data as input. The disadvantage is that potential faults can not be properly diagnosed because the information content is low. One reason is that high excitations are undesired during operation. Active FD overcomes this problem by injecting an additional excitation designed to maximize diagnosability. It is reported that active FD yields significantly better performance while it has the drawback of interfering with the system (Heirung and Mesbah, 2019). Nevertheless, real-world applications are rare. Therefore, this work provides an extensive overview of application scenarios for FD systems. Our main contributions are: (1) We describe four typical fault diagnosis scenarios of electromechanical motion systems without external sensors. (2) We emphasize the importance of the set trajectory selection regarding achievable accuracy and data requirements for three commonly applied FD algorithms. (3) We carry out a comprehensive investigation in all scenarios based on measurement data using the example of belt drives.

The remainder of this work is structured as follows: Section 2 gives an overview of three common machine learning algorithms used for fault diagnosis. Section 3 introduces the belt pretension monitoring as an example of FD commonly encountered in the automation industry. Different real-world application scenarios are described and the testbed used for gathering measurement data is presented. All scenarios are assessed on extensive datasets in section 4. Conclusions and further research directions are given in section 5.

2 FAULT DIAGNOSIS ALGORITHMS

ML-based FD is achieved by supervised classification of available sensor data $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ into healthy and faulty classes y. For this task, a lot of algorithms are proposed in the field of time series classification (TSC). We select three methods for benchmarking the classification accuracy in different application scenarios. te attempt was made to cover a broad range of algorithms from basic to advanced. This section gives insights into all methods.

2.1 One-nearest Neighbor with Distance Measure

The one-nearest neighbor (1-NN) with a distance measure is considered a simple baseline approach which is hard to beat by more advanced methods (Bagnall et al., 2016). It classifies the data without transformation. All data are simply stored and a distance measure assesses the similarity of a new sequence to all training samples. The closest sample's label is used as prediction following a one-nearest neighbor approach. Dynamic time warping (DTW) is an elastic distance measure that accounts for time shifts and has proven to be suitable for the classification of sequential data. Further details can be taken from (Bagnall et al., 2016). In this work, input data stems from time and frequency domain. If no data shift occurs, a simple Euclidean distance (ED) is likewise sufficient. A shift can be avoided e.g. in the frequency domain or by restricting the input data. For the sake of simplicity, we choose *1-NN ED* as the baseline algorithm.

2.2 Statistical Features

Conventional classification algorithms are not suitable for raw time series data. Therefore, features are extracted in the first step. Feature engineering aims to find features that are as informative as possible with regard to the target variable. This step is crucial for high accuracy. Statistical features that summarize certain characteristics of the underlying time series are a frequently chosen possibility (Fulcher and Jones, 2014). A collection of common statistical features applied in this work is given in Table 1. The advantage of simple feature functions such as mean $\mathcal{F}_1$, maximum $\mathcal{F}_5$ minimum $\mathcal{F}_6$, or energy $\mathcal{F}_{13}$ is a decent level of interpretability. The classifier's decision is thereby comprehensible promoting the general acceptance of machine learning in the industry.

A random forest classifier has proven to be effective in combination with statistical features (Fehsenfeld et al., 2020). It consists of multiple *Classification and Regression Trees* (CART) described by (Breiman, 2001). Decision trees are trained from the root to multiple leaves connected by nodes. At each node, the

Table 1: Overview of feature functions $\mathcal{F}_i$. In the time domain the sequence consist of a value x_i and a related time t_i at each step $\forall i \in \{1, \ldots, L\}$. Transformed by FFT into frequency domain the sequence comprises of an amplitude a_i and a related frequency f_i for $\forall i \in \{1, \ldots, K\}$.

Time domain	Frequency domain
$\mathcal{F}_1 = \frac{1}{L} \sum_{i=1}^{L} x_i$	$\mathcal{F}_9 = \frac{1}{K} \sum_{i=1}^{K} a_i$
$\mathcal{F}_2 = \frac{1}{L} \sum_{i=1}^{L} (x_i - \mathcal{F}_1)^2$	$\mathcal{F}_{10} = \frac{1}{K} \sum_{i=1}^{K} (a_i - \mathcal{F}_9)^2$
$\mathcal{F}_3 = \frac{\sum_{i=1}^{L}(x_i - \mathcal{F}_1)^3}{L \cdot \mathcal{F}_2^3}$	$F_{11} = \frac{\sum_{i=1}^{K}(a_i - \mathcal{F}_9)^3}{k \cdot \mathcal{F}_{10}^3}$
$\mathcal{F}_4 = \frac{\sum_{i=1}^{L}(x_i - \mathcal{F}_1)^4}{L \cdot \mathcal{F}_2^4}$	$\mathcal{F}_{12} = \frac{\sum_{i=1}^{K}(a_i - \mathcal{F}_9)^4}{K \cdot \mathcal{F}_{10}^4}$
$\mathcal{F}_5 = \min(x)$	$\mathcal{F}_{13} = \sum_{i=1}^{K} a_i^2$
$\mathcal{F}_6 = \max(x)$	$\mathcal{F}_{14} = \frac{\sum_{i=1}^{K} a_i \cdot f_i}{\sum_{i=1}^{K} a_i}$
$\mathcal{F}_7 = \frac{1}{L} \sum_{i=1}^{L} |x_i|$	
$\mathcal{F}_8 = \left(\frac{1}{L} \sum_{i=1}^{L} \sqrt{|x_i|} \right)^2$	

feature space is further split until regions exist where classification can be done. (Breiman, 2001) gives detailed information about the method. This approach is further referred to as *SF+RF*.

2.3 Minirocket

Minirocket (**MINI**mally **R**and**O**m **C**onvolutional **KE**rnel **T**ransform) proposed by (Dempster et al., 2021) is a fast classification algorithm which achieves state-of-the-art performance on a wide variety of TSC problems in the UCR time series classification archive (Dau et al., 2018). The time series is transformed by a large (by default $n_f = 10000$) number of random convolutional kernels. The feature space $\mathcal{K} = (k_1, \ldots k_{n_f})$ is obtained by calculating the proportion of positive values for each kernel transformation. Details on the kernels are reported in (Dempster et al., 2021). A linear ridge regression classifier is used as recommended by the authors. For this purpose, the linear regression

$$\hat{y} = \beta_0 + \beta_1 k_1 + \ldots + \beta_{n_f} k_{n_f} \quad (1)$$

is utilized, where $\hat{y}$ is the predicted target variable. The coefficients β_i are estimated based on available target variables y_i by

$$\hat{\beta} = \min_{\beta} \sum_{i=1}^{n_{obs}} \left(y_i - \beta_0 - \sum_{j=1}^{n_f} \beta_j k_{i,j} \right)^2 + \lambda \sum_{j=1}^{n_f} \beta_j^2, \quad (2)$$

where the first term is the residual sum of squares and the second term is a shrinkage penalty forcing β_j to be close to zero. λ is the tuning parameter trading off the two terms. (James et al., 2014)
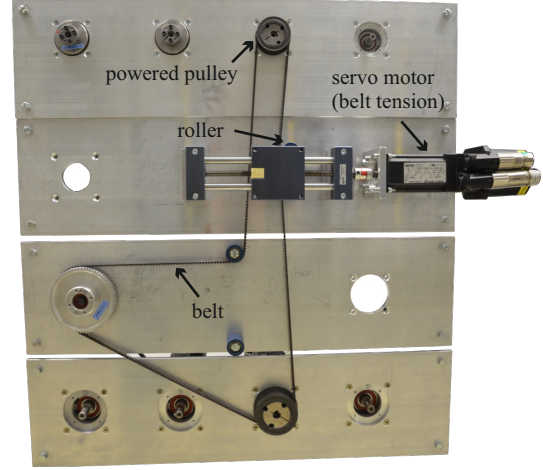

Figure 1: Belt drive used for evaluation of different application scenarios of belt pretension fault diagnosis.

3 FAULT DIAGNOSIS OF BELT DRIVES

Electromechanical motion systems are applied in many industrial fields. High reliability and efficiency of all components involved are vital for safe and efficient operation. Toothed belts are popular drive elements because they combine the capability of high acceleration, smooth running characteristics, and high precision for point-to-point motions (Perneder and Osborne, 2012). Diagnosing a belt drive's pretension serves as an FD example to introduce real-world application scenarios. At this point, it is highlighted that the selection of input data is not restricted to this case but should be considered for every FD system.

The validation of all scenarios is based on measurement data. Therefore, the experimental setup is introduced first. After that, we demonstrate how to realize active and passive FD in this example.

3.1 Experimental Setup

A belt drive with adjustable pretension is used to gather measurement data. It is depicted in Figure 1. There are four pulleys connected by a toothed belt with AT-5 profile and a total length of $l = 2\,\text{m}$. The upper pulley is powered by a servomotor with a rated torque of $M_0 = 1.2\,\text{Nm}$ and a rated power of $P_0 = 1.2\,\text{kW}$. The servomotor is equipped with standard sensors that are available in real-world applications:

- position φ_{act} (and derivatives $\dot{\varphi}_{\text{act}}$, $\ddot{\varphi}_{\text{act}}$),
- torque M_{act},
- temperature ϑ.

Consequently, broad applicability is ensured by involving only standard sensors.

Another servo motor adjusts the belt tension that moves a roller attached to a linear axis. The servomotor's position correlates with the belt pretension. A characteristic curve between the servomotor's position and the belt tension force F_{belt} is created which is only used to label the measurement data and not for prediction. The pretension force ranges from $F_{belt,min} = 40\,N$ to $F_{belt,max} = 200\,N$. Afterwards, it is discretized into $n_c = 5$ equally-spaced classes.

3.2 Application Scenarios

The diagnosis solely relies on the input data gathered from the machine regardless of the downstream algorithm. Obviously, the input data has a high impact on the diagnosis performance. In many real-world applications, there is a certain amount of freedom in designing the FD system which will be carved out in the section below. Set trajectories for passive and active scenarios are defined and measurements are gathered. Random samples of each class are shown in Figure 2 and discussed below.

3.2.1 Passive Fault Diagnosis

Passive FD takes only normal operation data as input to the algorithm. During normal operation, the servo motor fulfills specific tasks by following point-to-point motions. Jerk-limited trajectories (JLT) are the industrial standard for this purpose. The trajectory between the start φ_{start} and end position φ_{end} is designed in advance during path planning. Matching set position $\varphi_{set}(t)$ and set velocity $\dot{\varphi}_{set}(t)$ are calculated. A rectangular-shaped and thereby limited set jerk $\dddot{\varphi}_{set}(t)$ and the associated trapezoidal set acceleration is characteristic for JLT. The shape is defined by the maximum values of jerk $\dddot{\varphi}_{max}$, acceleration $\ddot{\varphi}_{max}$, velocity $\dot{\varphi}_{max}$ and the distance $d = \varphi_{end} - \varphi_{start}$. The outcome is a smooth trajectory between φ_{start} and x_{end} hereby limiting the excitation of oscillations initiated by the set trajectory.

In the case of passive FD, it is inevitable to analyze JLT whether a fault has occurred. Two scenarios have been chosen which are feasible in practice.

Random Test Trajectory. During operation a large variety of JLT can be the result to fulfill the designated tasks. Therefore, all kind of JLT must be analyzed to be capable of fault diagnosis at any time. In this scenario, a dataset is created where all characteristic

JLT parameters are chosen randomly and uniformly distributed as follows:

$$d \sim \mathcal{U}(100\,deg, 2000\,deg),$$
$$\dot{\varphi}_{max} \sim \mathcal{U}(1000\,deg/s, 7000\,deg/s),$$
$$\ddot{\varphi}_{max} \sim \mathcal{U}(1000\,deg/s^2, 6000\,deg/s^2),$$
$$\dddot{\varphi}_{max} \sim \mathcal{U}(6000\,deg/s^3, 60000\,deg/s^3). \quad (3)$$

Consequently, the input data of the FD algorithm has a high variety. Random sections of the trajectories are selected to cut all samples to equal length. In this manner, a sliding window with fixed size of $t_{win} = 1\,s$ is imitated. Examples are shown in Figure 2a. All set trajectories differ making it hard to recognize patterns between the classes with the naked eye.

Fixed Test Trajectory. During operation, a motion can occur regularly when accomplishing a repeating task e.g. moving a lifting station from the ground to the first level. By choosing a certain JLT the variety of the input data is eliminated. The characteristic pattern of the FD target is now potentially better visible because the input variety significantly drops. The drawback is that potential faults can only be detected during this typical motion. In the case of faulty belt tension, this drawback seems acceptable since the pretension reduces gradually due to wear. But a sudden loss remains possibly undetected.

There are many potential trajectories for this scenario. Two JLT are compared to evaluate the importance of trajectory selection. Trajectory T_1 is chosen to be comparatively fast, while trajectory T_2 has reduced velocity. The parameters are chosen as follows:

$$d_1 = 700\,deg, \qquad d_2 = 2500\,deg,$$
$$\dot{\varphi}_{max,1} = 7000\,deg/s, \qquad \dot{\varphi}_{max,2} = 1000\,deg/s,$$
$$\ddot{\varphi}_{max,1} = 50000\,deg/s, \qquad \ddot{\varphi}_{max,2} = 50000\,deg/s,$$
$$\dddot{\varphi}_{max,1} = 200000\,deg/s^3. \quad \dddot{\varphi}_{max,2} = 200000\,deg/s^3.$$

Trajectory T_2 is visualized in Figure 2b. The set trajectory is the same for all runs which can be seen in the rotational speed signal. No big differences are visible between the classes indicating that the controller is able to follow the set value. The torque acts as the control variable. Slight deviations between the classes can be detected. All findings apply for trajectory T_1 as well.

3.2.2 Active Fault Diagnosis

In an active FD scenario, input data is not restricted to JLT but every motion is possible. Therefore, the first task is designing an optimal input sequence leading to an easy diagnosis of the considered faults. In the
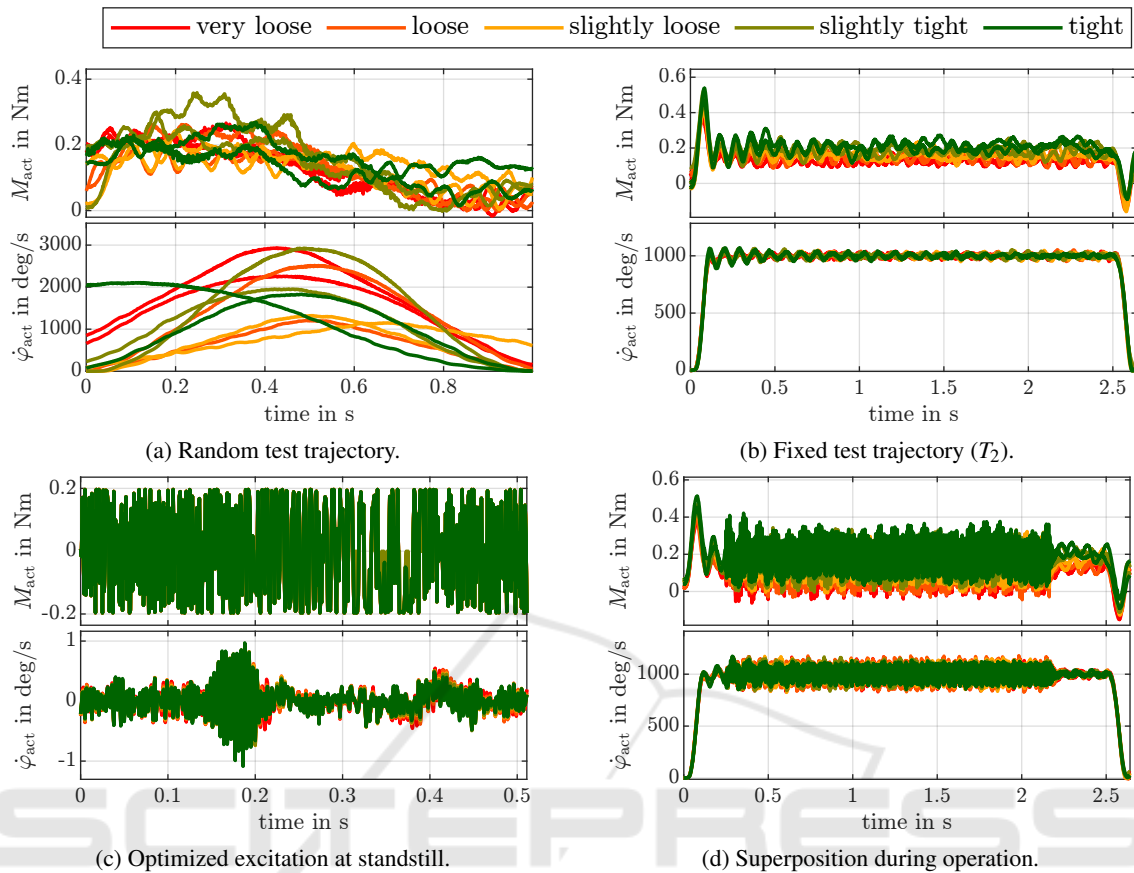
Figure 2: The actual torque M_{act} and the actual rotational speed $\dot{\varphi}_{\mathrm{act}}$ are shown for all scenarios in the time domain. Random observations for each class ($n_{\mathrm{c}} = 5$) are drawn.

case of belt drives, multi-frequency excitations (MFE) have proven successful (Fehsenfeld et al., 2020). It is a superposition of n_{freq} sine signals

$$M_{\mathrm{add}}(t) = A \cdot \sum_{i=0}^{n_{\mathrm{freq}}} \sin(2\pi f_i t + \phi_i), \qquad (4)$$

having different frequencies f_i and phases ϕ_i. All sine signals are phase-shifted to avoid extreme values making the signal as compact as possible. The each phase ϕ_i is chosen according to (Schroeder, 1970). The MFE can further be optimized by adjusting the amplitude A and the frequency content f_i.

Two different application scenarios are considered for active FD using an MFE as an auxiliary signal by adding it as torque offset M_{add}. First, the MFE is injected at standstill. In the other scenario, it is applied during operation.

Optimized MFE at Standstill. In practice, there might be natural stops of production overnight or during waiting periods until the next motion is triggered. In these situations, the test excitation can easily be

applied. Since the normal operation is not affected, constraints regarding the auxiliary signal are small. As long as the servo drive and the mechanics are not damaged, the impact is irrelevant.

A heuristic tuning of all MFE parameters is carried out in the design stage. The frequency content is observed to be optimal with $n_{\mathrm{freq}} = 256$ frequencies ranging from $f_{\mathrm{min}} = 1.9531\,\mathrm{Hz}$ to $f_{\mathrm{max}} = 500\,\mathrm{Hz}$. The length of the MFE is set to $t_{\mathrm{max}} = 0.512\,\mathrm{s}$ ensuring an integral multiple of the period duration of each frequency to avoid spectral leakage when transforming to frequency domain. The amplitude of all sines is set to $A = 0.015\,\mathrm{Nm}$. The optimized MFE is shown in Figure 2c as torque signal M_{act}. There are no differences between the belt classes since it is a set signal. The rotational speed $\dot{\varphi}_{\mathrm{act}}$ includes the system's response. In the time domain, only slight deviations are visible. A fast Fourier transform (FFT) is applied to obtain the frequency domain. A small but highly reproducible shift of characteristic frequencies across the classes can be observed in Figure 3.

While the possibilities to design an appropriate auxiliary signal are many, the biggest drawback is
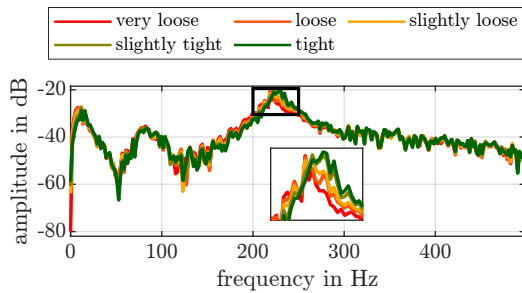
Figure 3: Frequency domain of rotational speed $\dot{\varphi}_{act}$ when excited by optimized MFE at standstill.

the fact that the operation must be interrupted. If no pauses in the process are possible or desired, the auxiliary signal must be injected during operation to make active FD viable in this case.

Optimized MFE during Operation. If the auxiliary signal is injected into a motion it is desirable to keep its impact at a minimum. Especially the amplitude A has a high effect and must be chosen as small as possible. This leads to a trade-off between performance and disturbance since a higher amplitude normally leads to better diagnosability. A heuristic tuning is done to balance this trade-off. The frequency content is found to be optimal with $n_{freq} = 4$ from $f_{min} = 100\,\mathrm{Hz}$ to $f_{max} = 300\,\mathrm{Hz}$. The amplitude is set to $A = 0.02\,\mathrm{Nm}$ and the total excitation length is $t_{max} = 1.921\,\mathrm{s}$. This MFE superposes the normal operation in the phase of constant velocity. The fixed trajectory T_2 is chosen for direct comparison to a passive scenario. The superposition is visible in both torque and rotational speed signals in Figure 2d starting at $t \approx 0.2\,\mathrm{s}$. The maximum speed error $\dot{\varphi}_{set} - \dot{\varphi}_{act}$ is increased by approximately $9\,\%$ due to superposition compared to the original motion. If no task is performed that requires high path accuracy this seems acceptable in many real-world applications.

4 EXPERIMENTAL RESULTS

For all scenarios training and test datasets are created. Both datasets are independently recorded. The training and test dataset size for the random trajectory scenario is $n_{obs,1} = 10000$, for all other scenarios it is $n_{obs,2} = 1000$.

All datasets are used to evaluate two different aspects of FD. The ability to reliably recognize faults is the key factor for successful FD. The classification accuracy is hence assessed in the first step. Furthermore, a big problem in real-world applications is that datasets are usually scarce because data gathering

and labeling are associated with large human effort. The number of required measurements is desirably as small as possible and therefore additionally investigated. An overview of all results on the test dataset is given in Table 2 and will be further discussed subsequently.

4.1 Achievable Accuracy

The classification results of all algorithms introduced in section 2 can be seen in Table 2. As expected, the accuracy in active FD scenarios is higher compared to passive FD scenarios. In both active scenarios, the accuracy is notably above $90\,\%$ accomplishing a near-perfect outcome. In a passive scenario, the accuracy drops significantly. When considering a fixed test trajectory the achievable performance depends on its selection. Trajectory 2 is more suitable for FD than trajectory 1 regardless of the algorithm. In the case of random trajectories by definition no selection is necessary. All kinds of trajectories are present whereby some are better suitable than others. The result is a maximum accuracy of approximately $90\,\%$ requiring a massive amount of measurements.

The comparison of selected FD algorithms yields expectable results. The state-of-the-art classifier *minirocket* outperforms *1-NN ED* and *SF+RF* in all scenarios. It can be concluded that the findings in both active scenarios do not differ much and basic algorithms perform equally well. In the passive scenarios where classification is more difficult *minirocket* shows its capabilities.

4.2 Data Requirements

The required amount of data is determined by gradually reducing the training dataset size. The results

Table 2: Maximum achievable test accuracy $\mathrm{Acc_{max}}$ (top) and minimum dataset size n_{min} (below). The belt tension is discretized into $n_c = 5$ classes.

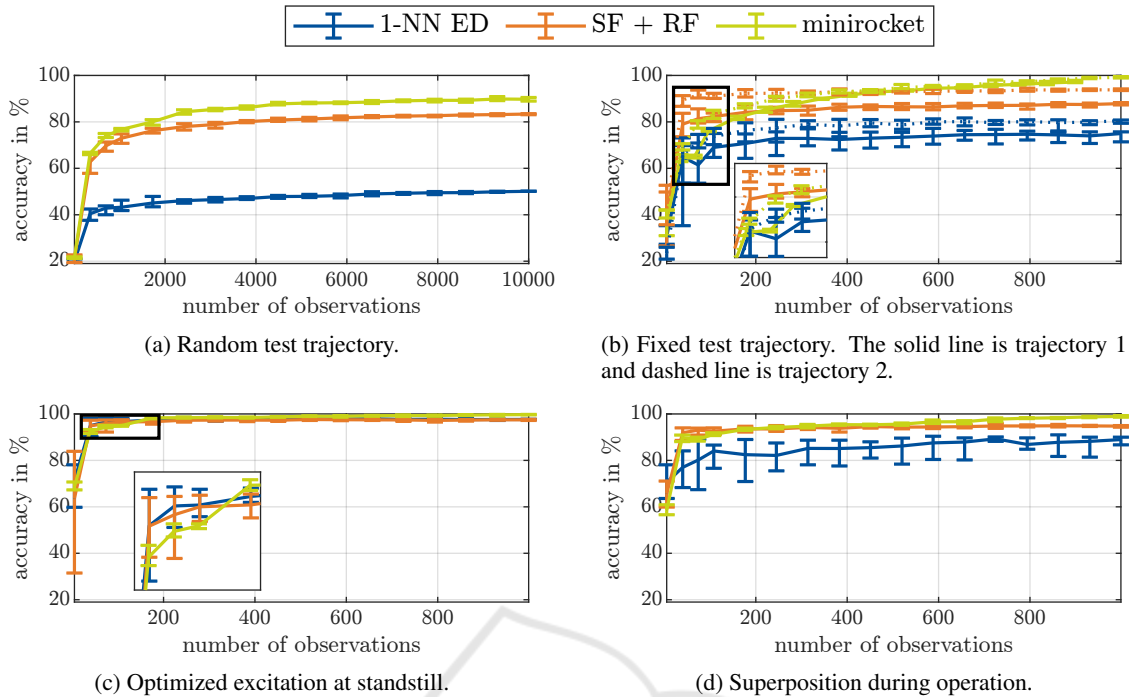| Scenario	1-NN ED	SF + RF	Minirocket
MFE	97.6 % **143**	98 % **143**	**99.7** % 484
Superposed traj.	89.8 % 725	95.6 % **310**	**99.2** % 502
Fixed test traj. 1	74.7 % **660**	87.8 % 690	**97.4** % 930
Fixed test traj. 2	80.5 % **660**	93.8 % 725	**99.1** % 930
Random traj.	50.1 % 8860	81.5 % **7860**	**89.7** % 7970

Figure 4: Test accuracy over number of training observations in all scenarios and FD algorithms. The training dataset is drawn five times independently from all training samples while test dataset is kept constant. Errorbars show the scattering.

for data requirements are depicted in Figure 4. In all cases, a convergence behavior towards the maximum achievable test accuracy is observed. It can be concluded that enough data was gathered. The required dataset size n_{min} is calculated as the smallest number of observations needed to achieve 99 % of maximum test accuracy in each scenario. The results are added to Table 2.

Again, active FD scenarios show advantageous properties. Small datasets are sufficient to reach a high accuracy regardless of the algorithm. Basic methods generally require fewer data while *minirocket* tends to require more data. It can be noticed in Figure 4c and Figure 4b that *SF+RF* even outperforms *minirocket* on small dataset sizes $n_{obs} < 200$. The superposition of a JLT has a positive effect compared to the original motion. The data requirement notably decreases and the accuracy improves in particular for *1-NN ED*, while for *SF+RF* and *minirocket* it is already on a high level.

In passive FD scenarios, the selection of the algorithm is of greater importance as large differences are observed. *1-NN ED* shows worst results followed by *SF+RF*, whereas *minirocket* yields a high accuracy for both fixed trajectories. A random trajectory scenario seems infeasible in practical applications because of the massive amount of data ($n_{obs} > 7500$) needed.

5 CONCLUSIONS AND FUTURE WORK

In this work, different FD application scenarios of electromechanical drive systems are described and examined. They are divided into passive scenarios which rely on jerk-limited trajectories and active scenarios where an MFE is injected. We demonstrate how an additional excitation leads to several advantages: By using an optimized MFE at standstill, even a very basic algorithm like *1-NN ED* achieves a high accuracy of 97.6 % while only $n_{obs} \approx 143$ number of training samples are required. Furthermore, the applicability of basic methods promotes secondary objectives like interpretability. In passive scenarios classification becomes more difficult. The accuracy drops below 90 % in many cases. In this case, the state-of-the-art classifier *minirocket* stands out with high accuracy, but requiring a significantly increased amount of data ($n_{obs} > 800$). A random trajectory scenario seems practically infeasible due to low accuracy (Acc < 90 %) and massive data requirement ($n_{obs} > 7500$). As a result, active FD is preferred over passive FD. The intelligent use of optimized excitations leads to feasible, reliable, and accurate fault diagnosis in a broad application spectrum.

Active fault diagnosis is lacking practical examples. The design of appropriate auxiliary signals for a broad range of FD applications is still open. Further research in this direction has the potential to enable safe and economical FD in real-world applications.

ACKNOWLEDGEMENTS

REFERENCES

AlShorman, O., Irfan, M., Saad, N., Zhen, D., Haider, N., Glowacz, A., and AlShorman, A. (2020). A Review of Artificial Intelligence Methods for Condition Monitoring and Fault Diagnosis of Rolling Element Bearings for Induction Motor.

Bagnall, A., Lines, J., Bostrom, A., Large, J., and Keogh, E. (2016). The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery*.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Bzinkowski, D., Ryba, T., Siemiatkowski, Z., and Rucki, M. (2022). Real-time monitoring of the rubber belt tension in an industrial conveyor. *Reports in Mechanical Engineering*, 3(1):1–10.

Dau, H. A., Keogh, E., Kamgar, K., Yeh, C.-C. M., Zhu, Y., Gharghabi, S., Ratanamahatana, C. A., Yanping, Hu, B., Begum, N., Bagnall, A., Mueen, A., Batista, G., and Hexagon-ML (2018). The ucr time series classification archive. URL:https://www.cs.ucr.edu/~eamonn/time_series_data_2018/.

Dempster, A., Schmidt, D. F., and Webb, G. I. (2021). MINIROCKET: A Very Fast (Almost) Deterministic Transform for Time Series Classification. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 248–257. arXiv: 2012.08791.

Fehsenfeld, M., Kühn, J., Wielitzka, M., and Ortmaier, T. (2020). Tension Monitoring of Toothed Belt Drives Using Interval-Based Spectral Features. *IFAC-PapersOnLine*, 53(2):738–743.

Fulcher, B. D. and Jones, N. S. (2014). Highly comparative feature-based time-series classification. *IEEE Transactions in Knowledge and Data Engineering*, pages 3026–3037.

Gangsar, P. and Tiwari, R. (2017). Comparative investigation of vibration and current monitoring for prediction of mechanical and electrical faults in induction motor based on multiclass-support vector machine algorithms. *Mechanical Systems and Signal Processing*, 94:464–481.

Heirung, T. A. N. and Mesbah, A. (2019). Input design for active fault diagnosis. *Annual Reviews in Control*, 47:35–50.

Hu, Y., Yan, Y., Wang, L., and Qian, X. (2016). Non-contact vibration monitoring of power transmission belts through electrostatic sensing. *IEEE Sensors Journal*, 16(10):3541–3550.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2014). *An Introduction to Statistical Learning: With Applications in R*. Springer, New York, NY, 1 edition.

Kande, M., Isaksson, A., Thottappillil, R., and Taylor, N. (2017). Rotating electrical machine condition monitoring automation—a review. *Machines*, 5(4):24.

Kang, T., Yang, C., Park, Y., Hyun, D., Lee, S. B., and Teska, M. (2018). Electrical monitoring of mechanical defects in induction motor-driven v-belt–pulley speed reduction couplings. *IEEE Transactions on Industry Applications*, 54(3):2255–2264.

Khazaee, M., Banakar, A., Ghobadian, B., Mirsalim, M. A., Minaei, S., and Jafari, S. M. (2017). Detection of inappropriate working conditions for the timing belt in internal-combustion engines using vibration signals and data mining. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 231(3):418–432.

Musselman, M. and Djurdjanovic, D. (2012). Tension monitoring in a belt-driven automated material handling system. *CIRP Journal of Manufacturing Science and Technology*, 5(1):67 – 76.

Nandi, S., Toliyat, H. A., and Li, X. (2005). Condition monitoring and fault diagnosis of electrical motors—a review. *IEEE Transactions on Energy Conversion*, 20(4):719–729.

Perneder, R. and Osborne, I. (2012). *Handbook Timing Belts*. Springer Berlin Heidelberg.

Picot, A., Fournier, E., Régnier, J., TientcheuYamdeu, M., Andréjak, J., and Maussion, P. (2017). Statistic-based method to monitor belt transmission looseness through motor phase currents. *IEEE Transactions on Industrial Informatics*, 13(3):1332–1340.

Schroeder, M. (1970). Synthesis of low-peak-factor signals and binary sequences with low autocorrelation (corresp.). *IEEE Transactions on Information Theory*, 16(1):85–89.

Sharma, V. and Parey, A. (2016). A Review of Gear Fault Diagnosis Using Various Condition Indicators. *Procedia Engineering*, 144:253–263.

Thoppil, N. M., Vasu, V., and Rao, C. S. P. (2021). Deep Learning Algorithms for Machinery Health Prognostics Using Time-Series Data: A Review. *Journal of Vibration Engineering & Technologies*.

Toma, R. N., Prosvirin, A. E., and Kim, J.-M. (2020). Bearing Fault Diagnosis of Induction Motors Using a Genetic Algorithm and Machine Learning Classifiers. *Sensors*, 20(7):1884.