Mechanical Systems and Signal Processing 99 (2018) 403–419

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Simulation-driven machine learning: Bearing fault classification

Cameron Sobie *, Carina Freitas, Mike Nicolai

Siemens Industry Software NV, Interleuvenlaan 68, 3001 Leuven, Belgium

ARTICLE INFO

Article history:
Received 16 January 2017
Received in revised form 16 June 2017
Accepted 19 June 2017
Available online 30 June 2017

Keywords:
Machine learning
Condition monitoring
Roller bearing
Fault detection
Prognostic health monitoring

ABSTRACT

Increasing the accuracy of mechanical fault detection has the potential to improve system safety and economic performance by minimizing scheduled maintenance and the probability of unexpected system failure. Advances in computational performance have enabled the application of machine learning algorithms across numerous applications including condition monitoring and failure detection. Past applications of machine learning to physical failure have relied explicitly on historical data, which limits the feasibility of this approach to in-service components with extended service histories. Furthermore, recorded failure data is often only valid for the specific circumstances and components for which it was collected. This work directly addresses these challenges for roller bearings with race faults by generating training data using information gained from high resolution simulations of roller bearing dynamics, which is used to train machine learning algorithms that are then validated against four experimental datasets. Several different machine learning methodologies are compared starting from well-established statistical feature-based methods to convolutional neural networks, and a novel application of dynamic time warping (DTW) to bearing fault classification is proposed as a robust, parameter free method for race fault detection.

1. Introduction

Rotating machines are ubiquitous in power production and transportation, among countless other applications. Downtime caused by failures of components such as bearing or gear faults directly reflects on the economic viability of large systems such as wind turbines [1] and industrial machinery [2]. Early warning of faults leading to probable failures can mitigate such risks as well as improve safety. The fatigue lives of common components in rotating systems can be described using widely accepted methods e.g. ISO standards; however, the lifetime of an individual component is inherently stochastic because of the nature of material variability, manufacturing defects, and loading regimes unique to each component. Therefore, lifetime predictions of these components are limited to probability distributions [3,4] that evolve differently for each physical realization of a component. Novel methods such as digital twins [5] aim to reduce this variability by explicitly accounting for the specific loading history on a structure. While this may become effective for larger structures such as aircraft, whose lifetimes are limited to tens of thousands of cycles between maintenance intervals, these methods are impractical for rotating machines with lifetimes from millions to billions of cycles and an early detection strategy is more feasible than a prediction strategy.

* Corresponding author.
E-mail address: cameron.sobie@siemens.com (C. Sobie).

http://dx.doi.org/10.1016/j.ymssp.2017.06.025
0888-3270/

Roller bearing failures constitute the majority cause of failures for industrial electric motors [6] and wind turbine gearboxes [7]; therefore, the early detection of their associated faults has the potential to significantly decrease unplanned downtime. The spatially constrained contact between the rollers and races combined with the time-varying nature of bearing loading drives fatigue damage. Such faults arise on the inner/outer races or roller elements when imperfections on the material surface propagate to form cracks. When these cracks merge, material is lost and pits are formed. This damage process is inherently multi-scale in nature, highly dependent on local material properties and structure, and influenced by multiple stochastic processes. While the microscopic spalling process is relatively well-understood [8], predicting the remaining time until spall formation requires detailed knowledge of the material microstructure and loading conditions; therefore, effort may be better spent on defect detection (and remaining useful life prediction) rather than predicting the formation of such faults. Rather than relying on recorded acceleration signals from defective bearings, physics-based modeling can provide insight into the vibration response of a faulted bearing. The simplest model for the excitation of a damaged bearing is a Dirac delta pulse train with a period associated to that of the defect as proposed by McFadden and Smith [9]. The physical fidelity of the model was improved with pulse modulation by the direction relative to the radial load on the bearing, and the bearing vibration response was assumed to be an under-damped oscillation. Following models proposed improvements such as waviness in the bearing race surfaces [10], introducing random fluctuations about the expected defect period [11], elastohydrodynamic film effects [12], and bearing race flexibility at higher modes and centrifugal load effects [13]. Finite element simulations have also been used to investigate the interactions between a roller element and a race defect [14,15].

Bearing fault detection is most often performed using linear acceleration or acoustic emission signals, and force sensors have been also shown to be effective [16,17]. Motor current signatures are also a proposed method for bearing fault detection [18–20] as bearing faults are a common cause for motor failure. Time-series data can be used to produce features that can differentiate between nominal and faulted bearings with measures such as root mean square (RMS), kurtosis, or skewness. Preprocessing with methods such as bandpassing, calculation of the envelope signal [21], minimum entropy deconvolution [22,17,23], and cepstral methods [17,23] have also been proposed. Frequency domain analysis is a more natural approach because of the periodic nature of rotating machines, and the fact that each bearing fault type (inner/outer race, roller element) has a characteristic frequency (assuming no slipping between the components). The power in the bearing vibration signal around these frequencies can be used to create features, in addition to other typical features such as spectral kurtosis. However, Fourier transform methods are inherently limited to steady state operation and are not well-suited to non-stationary operating conditions without additional processing. Modern time-frequency domain methods have been developed to address this challenge. These methods include the Hilbert-Huang transform [24,25], and discrete [26,27] and continuous [28] wavelet transforms (DWT/CWT). These techniques decompose the signal into a set of signals based on the frequency content. These transforms cannot be performed in closed form in general and the precise manifestation of a bearing fault in a signal processed with these transforms is difficult to predict, and therefore these methods are favored for numerical rather than analytical studies. Apart from transform methods, angle synchronous averaging (ASA) is a commonly used method to emphasize bearing defect signals and reduce noise under stationary and non-stationary regimes [29–33]. From the shaft speed, one can segment a time-series vibration signal into a series of intervals during which the bearing rotates over the angular defect period. Averaging the envelope signals over several of these intervals will isolate a response occurring at the averaging period length while reducing noise. This analysis method is resistant to varying shaft speeds and it includes the bearing geometry such that angle synchronous averages between different bearing geometries can be compared. These analysis techniques enable one to extract quantitative measures of the bearing vibrational signal characteristics; however, diagnosing a bearing fault using these measures, or combinations of them, presents a significant challenge to which machine learning algorithms have proven effective.

Numerous machine learning algorithms have been applied successfully in mechanical engineering applications and more specifically in the field of fault detection, condition monitoring, and prognostics. Contemporary condition monitoring methods using machine learning have required either commissioning experimental setups running under accelerated failure conditions until a fault is observed and measured, or the collection of in-service data (e.g. SCADA) [34] with which to train. At the system level, wind turbine performance and faults (for several applications such as power production prediction [35], pitch [36] and bearings faults [37]) have been heavily analyzed because production wind turbine units have integrated SCADA capabilities and years of data from a large number of turbines has been recorded. Within the domain of bearing diagnostics, experimental data-driven machine learning uses features such as time-series RMS, standard deviation, crest factor [38], and wavelet variance [39] among numerous other statistical measures. Combining all of the aforementioned and additional analysis methods, Rauber et al. [40] created a "feature pool" including time-series as well as frequency domain statistics and the energies of CWT decompositions to successfully classify bearing faults with several different classifiers. More recently, one-dimensional convolutional neural networks applied to the bearing acceleration frequency spectrum have been proposed as a more effective method of bearing fault detections than statistical feature analysis with the additional benefit of removing the need for expert knowledge associated with feature extraction [41]. This method shows promise, but in its proposed form is limited to stationary operating regimes as the input to the convolutional neural network is the power spectrum, which is inherently dependent on the operating speed and bearing configuration. In general, data-driven approaches are limited by the requirement of in-service data and there is little reason to expect that a classifier trained in one configuration would be applicable to a different bearing in a different application. Rather than relying on in-service data to train algorithms, simulation-driven machine learning overcomes this barrier by creating training data that can capture a wide range of operating conditions that can be later combined with in-service fault data applicable to its unique operating conditions as it

becomes available. This approach is far less common, with applications in cases such as the classification of internal combustion engine faults [42,43], model-based electrical drive fault diagnosis [44], and using neural networks trained using analytical models to detect building damage [45].

This work seeks to advance the automatic diagnosis of bearing race faults through two novel developments. First, simulated bearing dynamics data is used to train machine learning algorithms to detect failures in physical bearing measurements, and the methodology is validated using four experimental data sets including one industrial application and a measurement campaign performed using a SpectraQuest Machinery Fault Simulator. Second, two novel applications of machine learning algorithms to bearing faults are proposed: convolutional neural networks and nearest-neighbor dynamic time warping (NNDTW) [46] to angle synchronous averaged bearing acceleration signals. Their performance is compared to a range of traditional and neural network classifiers using common statistical features, and the accuracy of all algorithms between simulation and data-driven methodologies is compared.

2. Methodology

2.1. Procedure for analysis and classification of bearing faults

The task of simulation data driven bearing fault classification is accomplished in three stages. First, high resolution bearing dynamics simulations provide insight into the functional form of the envelope signal. This information is used to create longer time series signals incorporating stochastic effects, which are processed to produce features for machine learning as well as used directly without feature generation with CNNs and NNDTW. To validate this approach, experimental data is processed identically to produce angle synchronous averages of the bearing race acceleration envelope signal, which serves as validation data for the simulation-trained classifiers. For brevity, the classifiers trained using simulation data are denoted as simulation-driven, and those trained using experimental data are denoted as data-driven.

Fig. 1 illustrates the algorithm training and scoring methodologies used in this work as well as highlights the key novelties of this work: using simulation data for algorithm training (with experimental validation including one new dataset produced for this work), and the application of CNN and NNDTW to determine the presences of a bearing race fault from bearing acceleration signals. The SQ dataset denotes the novel dataset produced for this work.

The dashed horizontal line in Fig. 1 denotes the separation of simulation and experimental data. The same signal processing and feature extraction is performed regardless of the data source. The signal processing detailed below removes any dependence of bearing geometry and operating speed, and therefore any subset of the experimental data can be used as training data. The CWRU data is used as training data because it constitutes approximately two thirds of the experimental data as detailed in Table 2.

2.2. Angle synchronous averaging

The signal processing in this work is relatively simple and relies heavily on transforming time series data using angle synchronous averaging (ASA) and envelope functions, both of which are highly cited processing techniques for bearing analysis [29,47]. This processing is compatible with unsteady operating conditions, assuming that the shaft speed varies slowly

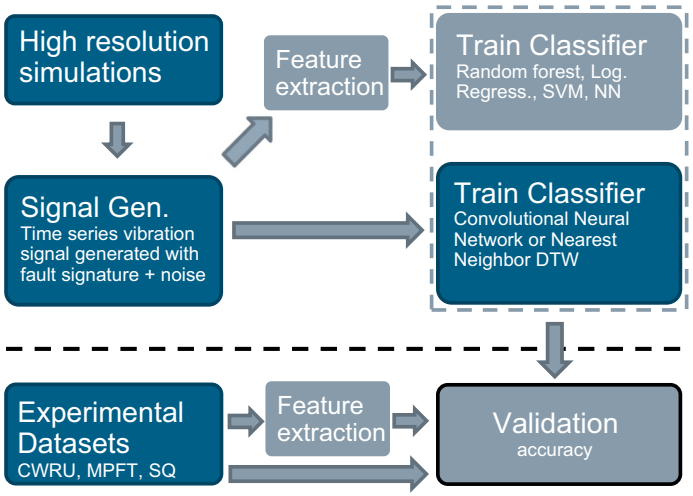


Fig. 1. Overview of the methodologies and novelties of this work (highlighted in blue). In the data driven case, the CWRU experimental data is used to train several algorithms that are scored against the MFPT and SQ datasets. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

compared to the defect period. This methodology has been shown to be effective in industrial applications, such as in the case of a wind turbine high-speed bearing fault detection [47].

Several methods exist to derive the envelope for a given signal such as squaring the signal and low-passing, taking the magnitude of the Hilbert spectrum, and secant based methods [48]. In this work, the magnitude of the Hilbert spectrum is used as it parameter-free. It is obtained by convolving the time signal $h(t)$ with the signal $1/\pi t$:

$$\tilde{h}(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} h(t)\frac{d\tau}{t-\tau}, \tag{1}$$

which is used to form the complex analytic signal:

$$h_a(t) = h(t) + i\tilde{h}(t) \tag{2}$$

The magnitude of Eq. (2) yields the envelope function of the function $h(t)$. In practice, the Hilbert spectrum is obtained using Fourier methods in which the negative half of the frequency spectrum is set to zero, and the magnitude of the inverse Fourier transform of this spectrum results in the envelope function.

Following the production of the accelerometer signal envelope, the synchronous average is taken over each characteristic defect signal period. This is defined in terms of angle of rotation for each defect by the outer/inner ball pass frequency, which can be derived from kinematic principles as Eqs. (3) and (4) with the assumption of no slip between the rollers and races.

$$\text{BPFO} = \frac{n_R}{2}\left(1 - \frac{D_R}{D_P}\right), \tag{3}$$

$$\text{BPFI} = \frac{n_R}{2}\left(1 + \frac{D_R}{D_P}\right), \tag{4}$$

where n_R is the number of rolling elements, D_R is roller diameter, D_P is the pitch diameter of the bearing, and the contact angle is zero. Eqs. (3) and (4) yield a factor that is multiplied to the shaft speed to obtain the defect frequency. In reality, bearing defects are often treated as pseudo-cyclostationary [30], indicating that the defect signal is produced every period with a random perturbation about the expected value.

Averaging numerous times over each period acts to diminish ergodic asynchronous noise at a rate of $1/\sqrt{n}$ where n is the number of averages [47], and the defect signals from any other defects present are also diminished. One synchronous signal is produced for each anticipated defect period so that a single series of acceleration data produces two ASA signals. The number of samples per synchronous period varies with the shaft speed and sampling rate, and therefore each synchronous period must be re-sampled to the same number of samples so that they can be averaged – in this case, each defect period is re-sampled to 128 samples a using a Fourier transform method. The number of samples was found to be a good balance between signal resolution and data size.

The number of periods used for the angle synchronous average plays an important role regarding the efficacy of the synchronous averaging method. Using too few periods does not average out noise or variations between pulses, whereas too many periods will accumulate too much deviation from cyclostationarity resulting in a less well-defined peak. In the course of this work, a range of 5–150 defect periods are used to create ASAs. The figures in this work use 25 defect periods, which is observed to be optimal or nearly so for many algorithms and simulated/experimental data. A comparison of the optimal algorithms is shown over a range of averaged defect periods at the end of this work for completeness.

The algorithms used here have no inherent limitation regarding the number of defects present – multiple faults are natively supported. All faults presented here occur at different frequencies, and averaging the signal over each defect's characteristic periodicity isolates each defect. Single defect types with multiple realizations present in the system can also be detected if the algorithm training includes such phenomena.

2.3. Data normalization

Energy-based bearing fault identification methods are commonly proposed in literature and used in industry [21]. These methods have proven effective when in-service data is available from which thresholds can be derived by experts or machine learning. However, these solutions are limited to the machinery providing in-service data and even further, energy-dependent features are correlated to the operating speed of the machine and are thus not robust against non-stationary conditions (or speeds that are not covered in experimental data). A key point of investigation of this work is how data produced from one source, be it simulations or experiments, can be used in other applications so that condition monitoring is possible on new machines from the first day of deployment. Enabling this level of functionality and robustness is only feasible if the dependence on sensor model, calibration, method of fixation, and drift, and structure transmission characteristics are removed. As a result, the envelope function is normalized to unit variance and zero mean everywhere in this work.

2.4. Machine learning

Machine learning has reached maturity with wide-spread usage in numerous commercial and open source packages offering optimized implementations. In this work, scikit-learn [49], keras [50] and FastDTW [51] are used for all machine learning applications, which allows for data handling and preprocessing using the Python scripting language.

Accuracy is one important measure of machine learning algorithm effectiveness; however, additional properties for binary classification such as precision and recall are equally important, particularly within the domain of condition monitoring. Precision is defined as the ratio of true positives to total samples classified as positive, and recall is defined as the ratio of true positives labeled as such to the total number of true positives. False positives (fault detections) manifest as low precision, and are undesirable because they may result in undue maintenance or a decrease in trust in the system by human operators. Low recall, indicating that faulty bearings are erroneously classified as nominal, is also clearly a negative attribute but may arise with high levels of noise or defect pulse error. With some algorithms such as logistic regression and support vector machines, the hyperparameters can be used to favor precision or recall, and the choice to favor specific metrics are domain and case specific. In this work, the maximum accuracy is taken as the value to optimize. For completeness, the correlation of precision and recall to classifier accuracy is investigated in Section 4.4.

2.4.1. Statistical feature-based

In this work, fault classification using statistical features is performed with the following algorithms:

- Logistic regression (LR) [52]
- K-nearest neighbors (KNN) [53]
- Random forests (RF) [54]
- Support vector machines (Radial basis function kernel) (SVM)[55]
- MLP Neural networks (NN) [56]

These algorithms are well known and extensively used in machine learning and more specifically in automatic bearing fault detection, and the details of their theoretical foundations, implementations and applications are widely available in the literature. The citations above provide examples of applications within the general domain of fault detection. Data-driven methods can achieve bearing race fault classification accuracy of well above 90% – in this work, we extend this knowledge to the domain of simulation-driven machine learning. It is therefore important to emphasize that in addition to the training, all parameters of the machine learning algorithms are optimized strictly on simulation generated data.

2.4.2. Convolutional neural network

Convolutional neural networks were proposed as a new machine learning method better adapted to analyze high dimensional data and data with inherent structure (e.g. images or time series data) than classical neural networks [57]. Inspired by the biological structure of vision systems, LeCun and Bengio proposed CNNs in which the input neurons have receptive fields that are convolved across the entire across the input data, greatly reducing the number of parameters that would be associated to each pixel/time value connected to an input neuron in a classical MLP neural network. During algorithm training, the network learns filters that activate based on spatial/temporal locations of features in the sequence of data. To classify the data, the filters are flattened and each value in a filter is connected to an activation unit and the final layers of the network are a standard MLP. While CNNs have been highly popularized by their state-of-the-art performance for image classification [58], one-dimensional applications have been more limited, with examples including in time series classification [59] and text generation [60]. Numerous architectures were tested in the course of this work, and the optimal architecture is found to be very simple compared to those seen in other applications of machine learning. The network is composed of a single convolutional layer of 15 filters with a receptive field size of 10 and a stride of five. The filters are flattened, a 50% dropout is applied, and fed into a layer of 50 standard neurons, followed by another 50% dropout and a final single neuron indicating the fault state. Deepening or broadening the network did not improve cross-validation accuracy.

2.4.3. Nearest-neighbor dynamic time warping

Dynamic time warping was first proposed as a method to compare time-sequences with the goal of improving spoken word recognition [61]. Qualitatively, DTW is an elastic distance measure between two sequences of ordered numbers. An iterative process examines possible distortions of the sequences to minimize the total distance between them. In general, this is an $\mathcal{O}(n^2)$ operation in both time and memory. Numerous optimizations have been proposed, such as constraining the warping window size (bounding the largest possible distortions) [62] and lower bounding techniques [46]. Furthermore, numerical optimizations to nearly $\mathcal{O}(n)$ ($10^{-8}n^2 + 0.001n$ for a warping window of 100) in time and memory [51]. As a classification method, first nearest-neighbor dynamic time warping (NNDTW) has been extensively shown to be highly effective in tasks such as chromosome subsequence and time series motif recognition matching at the scale of billions and trillions of sequences [63]. While never before applied to the case of bearing faults, DTW has nevertheless been shown to be an effective tool for electric motor [64] and chemical process [65] fault diagnosis.

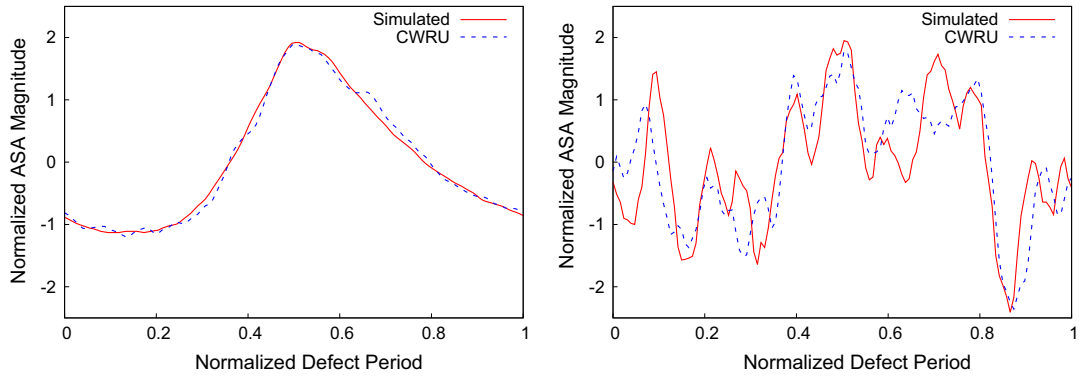


Fig. 2. Comparison of first nearest-neighbor DTW matches for ASAs with (2a) and without defects (2b). Because this method gives the exact training sample to which the test sample is matched, one can use annotated training data to infer properties of the experimental data source such as defect size and severity.

In this work, NNDTW is applied to bearing fault detection according to the following methodology. Angle synchronous averages are normalized to zero mean and unit variance, and the signal is permuted such that maximum value is in the center of the synchronous average. The ASAs derived from experiments are compared with simulation-generated ASAs, and the experimental signal is assigned the class of the nearest simulated signal. Fig. 2 compares the closest matches between simulated and experimental (CWRU) data found using NNDTW for a faulty and nominal bearing.

3. Data approaches

3.1. Simulation-driven

Bearing fault vibration signatures have been extensively studied from analytic and numerical perspectives [9,11–13,66,14,15]. McFadden and Smith proposed a first model approximating race faults as point defects and describing bearing dynamics using an under-damped equation of motion. The resulting exponentially decaying oscillatory behavior captured by this simplistic model is the fundamental signature of a bearing race fault; numerous further studies using models of increasing complexity, from a three degree of freedom bearing model using point masses, linear springs and linear damping proposed by Sassi et al.[12] to a 60 degree of freedom Bond graph model incorporating Hertzian contact and nonlinear damping [67], the core of the response remains the same. Further studies using 3D finite element analysis with advanced contact mechanics derive additional fine-scale information regarding bearing-race interactions [14], but the fundamental behavior of the system is nevertheless largely captured by the original model of McFadden and Smith regarding the specific behavior that is indicative of a bearing race fault. However, not all physical phenomena influencing the dynamic behavior of roller bearings have been incorporated into state of the art models. The aforementioned models predict cyclostationary signals, whereas defective bearings are considered to be pseudo-cyclostationary [30]. The stochastic effects arising due to deviations from the assumptions or simplifications under which these models were derived inherently excludes behavior such as rollers slipping relative to the races, rollers losing contact with one or both races, or vibrations of unloaded rolling elements. Therefore, the proposed methodology for detecting race faults using simulation-driven machine learning in this work separates bearing behavior into two components: the deterministic contribution, which is simulated using a simplified bearing model, and stochastic effects, which are incorporated by taking a wide range of parameter values with the goal of capturing a wide range of possible behavior. The role and selection of these parameters is detailed below.

The model we adopt in this work is the one-dimensional 3-DOF model of Sassi et al.[12], which is implemented in this work using Siemens LMS Imagine.Lab Amesim simulation software [68]. In this software, the system is represented using Bond graph theory; it can be equivalently described as a simple system of second-order ordinary differential equations as shown in Eq. (5). Siemens LMS Imagine.Lab Amesim is used to solve the system, which automatically selects between more than 20 different solvers including variable or fixed step, single or multi-step and non-stiff or stiff solvers as well as differential algebraic equation integration. The solver changes during a simulation depending on the evolution of the governing equations with the goal of minimizing error while optimizing runtime.

The equations of motion for a single roller interacting with the two races can be posed as follows:

$$M_{ij}\ddot{y}_j + C_{ij}\dot{y}_j + K_{ij}y_j = F_i \tag{5}$$

where $M_{ij}, K_{ij}, C_{ij}, y_j$, and F_i are the mass, stiffness, damping rate and positions, respectively, and are given by:

$$M_{ij} = \begin{bmatrix} M_{OR} & 0 & 0 \\ 0 & M_B & 0 \\ 0 & 0 & M_{IR} \end{bmatrix} \tag{6}$$

$$K_{ij} = \begin{bmatrix} K_{OR} + K_{OF} & -K_{OF} & 0 \\ -K_{OF} & K_{OF} + K_{IF} & -K_{IF} \\ 0 & -K_{IF} & K_{IR} + K_{IF} \end{bmatrix} \tag{7}$$

$$C_{ij} = \begin{bmatrix} C_{OF} & -C_{OF} & 0 \\ -C_{OF} & C_{OF} + C_{IF} & -C_{IF} \\ 0 & -C_{IF} & C_{IF} \end{bmatrix} \tag{8}$$

$$y_j = \begin{bmatrix} y_{OR} \\ y_B \\ y_{IR} \end{bmatrix}, \quad F_i = \begin{bmatrix} F_{OR} \\ F_B \\ F_{IR} \end{bmatrix}. \tag{9}$$

where subscripts OR, IR, OF, IF, B indicate properties relating to the outer race, inner race, outer fluid film, inner fluid film, and ball (roller) respectively. F_{OR}, F_B, and F_{IR} are the external forcing components acting on the system due to the presence of an outer race, ball, or inner race fault. Only outer and inner race faults are considered in this work, and thus $F_B = 0$ for all simulations.

To summarize the model, the first oscillatory modes of the inner and outer races constitute two of the degrees of freedom, and the third is the roller element. The elastohydrodynamic film was initially modeled with as a spring-damper whose parameters are formulated in [69,70] before a range of parameters was tested to explore the envelope of a typical dynamic bearing response. The roller element is assumed to be infinitely stiff, i.e. the vibrations of the roller element itself are neglected. A race defect is modeled either with a prescribed force representing the roller element entering and exiting the defect, or by directly representing the defect profile in the boundary conditions. Inner race and roller element defect vibrations are modulated by the load profile of the bearing as the load is time-varying (assuming an inner rotating shaft).

Regardless of the defect present, the oscillations of the outer race are analyzed because it is assumed that the inner race is rotating while the outer race is fixed in a pillow block, to which an accelerometer is attached. As previously mentioned, all studies in this work, the accelerometer signal (simulated or experimental) is normalized to zero mean and unit variance.

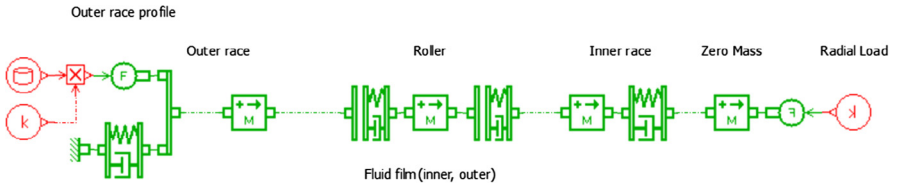
These simulations are performed under a high speed regime, resulting in transient defect-race interactions (impacts) that can be modeled using impulses. In low speed regimes such as the main bearing in a wind turbine (≈5–22 RPM) [71], the acceleration time-scales of the bearing motion and transient responses greatly differ, and the precise physical morphology of the defect will strongly influence the bearing acceleration response. Even at high speed regimes, a non-negligible defect length compared to the roller element size may produce a vibration signal depending on the defect surface morphology, before the vibration decays exponentially when the roller leaves the defect.

The vibration envelope is seen to evolve significantly (changing functional forms) under different loading conditions and parameters. An infinitesimal defect results in a typical under-damped oscillation with an exponentially decaying envelope, whereas a defect of a finite size causes a two-peaked response caused by the time passing between the roller element entering and exiting the defect and the resulting impacts with the races. This behavior has also been noted in significantly more complex Bond graph [67] and multibody dynamics models [15]. The damping of the fluid film governs the settling time for the under-damped response, which varies significantly depending on the parameters used. The fluid film properties are speed and load dependent [70,69]. Therefore, the dynamics are instead simulated over a range of values for the film properties (similarly for other component properties) to estimate the ranges of the envelope signal parameters (see Fig. 3).

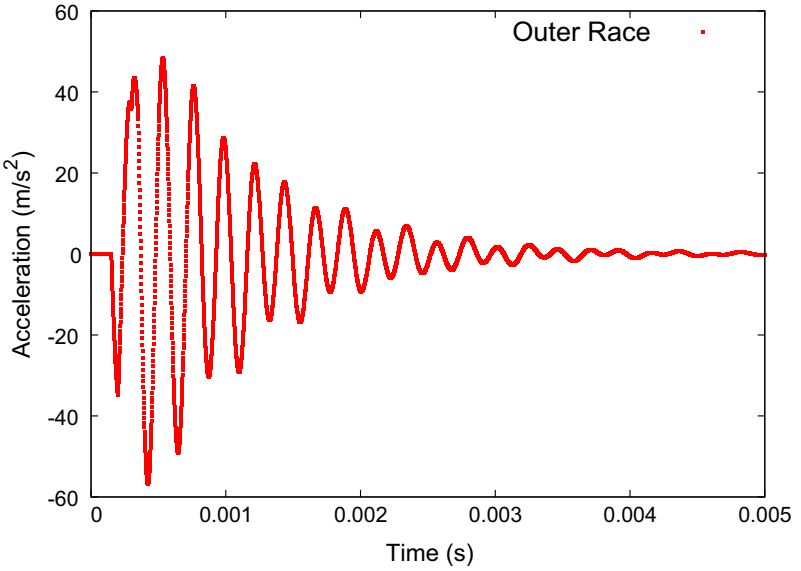
Deviations from idealized behavior such as slippage between the rollers and races cause the time between roller-fault interactions to be influenced by stochastic processes. These effects are included in the simulated time series data as follows: random fluctuations of rollers within the cage are assumed to cause Gaussian broadening in the distribution of defect periods and the retardation effect on the rollers caused by an azimuthal force in the roller-defect interaction for race defects is assumed to cause a phase advance governed by a half-normal distribution. For finite sized defects (which cause two distinct interaction roller-race events resulting in two exponential decays in the envelope signal), the size of the defect is taken as a random variable. Finally, the decay constant of the exponential decay that is governed by the fluid film damping is taken as a random variable, also informed by simulations. The values of these parameters and the interval at which the random parameters are generated are taken as shown in Table 1.

Noise is added to the simulated bearing dynamics to improve fidelity to a physical signal. Sassi et al. [12] added noise to match measure accelerometer RMS values; however, the frequency spectrum is not specified. In this work, pink noise ($1/f$ power spectrum) is selected to add to simulated signals as pink noise (rather than white noise) as it is ubiquitous in nature, electronics, machinery, and numerous other fields [72]. This power spectrum has also previously been selected for synthetic bearing signal generation [73]. The noise variance is constant for each simulation run, and for each run it is randomly selected such that the signal to noise ratio can range from 1 to 10.

The effectiveness of the proposed simulation-driven machine learning methodology regarding physical failure detection relies on the effective isolation of an identifiable defect response. For example, a bearing with diminished lubrication produces an acceleration response with increased noise relative to the nominal condition [74–76]. As a result, accuracy classification necessitates a complex model reproducing both the nominal and contaminated lubrication starved conditions such



(a) LMS Imagine.Lab Amesim model used to simulate bearing dynamics. This model simulates bearing excitation using a user-specified force input at the left of the figure. A zero-mass is necessary on the rightmost connection to ensure causality in the Bond graph is satisfied.



(b) Bearing outer race acceleration following roller passing an infinitesimal defect.

Fig. 3. 1D model and acceleration response for an infinitesimal defect. In the case of a finite defect, the spacing of the force impulses are determined by the width of the defect, and their magnitude by the defect depth. A wide range of values for the impulse spacing and magnitude are used to create training data.

Table 1
Model parameters taken as random variables during simulation. The parameter ranges are obtained by analyzing envelope function for typical physical parameters in the high resolution simulations.

Parameter	Range	Generated
Decay constant τ for exp. envelope	$2-16 \times 10^{-3}$ s	per sim.
Defect period error	0–15% of def. period	per pulse
Defect phase delay	0.01–1% of def. period	per pulse
Finite defect pulse delay	0–15% of τ	per sim.

that a clear differentiation can be made in training. The race defects in this work are relatively less sensitive to model accuracy because a clear transient response is expected with a dramatic departure from a nominal bearing acceleration.

3.2. Data-driven

All experimental data is considered here as an ensemble (except for the wind turbine data, considered in an independent section) and the accuracy scores reported are thus for the union of the CWRU, MFPT and SQ datasets.

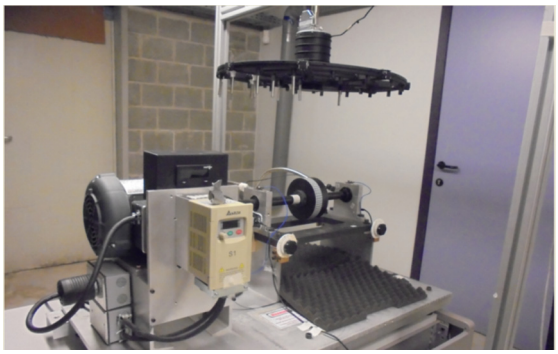


Fig. 4. SpectraQuest Machine Fault Simulator apparatus used to produce new experimental data for this work.

3.2.1. SpectraQuest machine fault simulator

The measurements for medium speed regime were performed in a fault machine simulator from SpectraQuest (depicted in Fig. 4). The machine is driven by a 3 HP Marathon Electric Three Phase AC motor controlled by a Delta VFD-S Inverter. Two different bearing conditions were measured (healthy and outer race fault) for speeds of 1260, 1500, 1860, and 2220 rpm. A 5 kg mass was added at the mid-point of the span between the shaft bearings to increase the radial force in the bearing and enhance the bearing-fault interaction. The bearings under analysis were ER-16 K. The accelerometer PCB 365A01 was located on the top of the housing of the faulty bearing and a microphone GRAS 40 PH was located close by this bearing (the microphone was not used in this work). This dataset will be denoted as the SQ dataset in this work.

3.2.2. Case Western Reserve University

The publicly available data from the Case Western Reserve University is produced using a 3 HP motor driving a dynamometer result in a power load between 0–3 HP. The data used in this work is strictly the "Drive end bearing," which is a 6205-2RS JEM SKF deep groove roller bearing. Faults with depths of 0.1778 mm, 0.3556 mm, and 0.5334 mm were created using electro-discharge machining. The fault is placed in-line with the load, at 90° relative to the load, and at 180° relative to the load. All orientations and loads are used in this work and no distinction is made between them.

The CWRU dataset has been used extensively as a data source for studies proposing novel signal processing, fault identification and machine learning studies for bearing fault detection. Following this extensive usage, Smith and Randall [77] performed a detailed benchmark study on this data comparing three fault detection methods for every data record in the dataset. In the course of this study, Smith and Randall observed atypical behavior in certain records such as random impulses and impulsive modulation at the shaft frequency, which the authors hypothesized as the result of mechanical looseness. This behavior is noteworthy as the CWRU dataset will be used in future sections as training data for the machine learning methods.

3.2.3. Society for machine failure prevention technology

The data made available by the MFPT uses a NICE bearing. The defect seeding processes is not described. Three measurements with a load of 1201 N on the bearing are provided for the baseline condition and an outer race fault, as well as seven measurements for both the outer and inner race faults over a range of 0–1334 N bearing load.

3.2.4. Wind turbine

Data from an industrial application is provided from a 2 MW commercial wind turbine. Six seconds of data per day is captured using a linear accelerometer on the high-speed shaft bearing. While no geometric or model information about the bearing is provided, the ball pass frequencies are published with the data.

3.2.5. Data summary

Table 2 summarizes details of each experimental seeded fault dataset.

The wind turbine high-speed bearing fault data was collected from a commercial in-service 2 MW wind turbine. Six seconds of data was captured per day for 50 days. The shaft speed is approximately 1800RPM, but it is not constant during the data-capture process.

4. Data analysis

The capability to detect physical failures using a machine learning algorithm trained with simulation data has the potential to improve condition monitoring by removing significant hurdles associated to data-driven condition monitoring. The wider applicability and greater robustness of a simulation-driven method can only be tested by applying it to experimental

Table 2
Parameters of experimental datasets with seeded faults. The time column indicates the sum total length of all signals for the defect type and data source.

Dataset	Defect	Speed (rpm)	Time (s)
CWRU	Nom.	1700	140
	OR	1700	280
	IR	1700	120
MFPT	Nom.	1500	18
	OR	1500	39
	IR	1500	21
SQ	Nom.	1260–2220	122
	OR	1260–2220	122

datasets, and comparing its accuracy to the same algorithms trained and scored using data from distinct machines. All three experimental sets with seeded defects are considered together here, and the wind-turbine bearing data is elaborated in a following section.

4.1. Machine learning based on statistical features

A large number of potential statistical features for bearing fault detection have been proposed in the literature using statistics extracted from the time domain, the frequency domain, wavelet decomposition, HHT components, and more. In this work, the statistical features derived from the ASA are the skewness (Eq. (10)), kurtosis (Eq. (11)), crest factor (CF, Eq. (12)) and margin factor (MF, Eq. (13)), and from the ASA power spectrum, the skewness and kurtosis are used. Finally, including the energies of the first four discrete wavelet decompositions (using the Daubechies 2 wavelet) results in a total of twelve features.

$$\text{Skewness} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \bar{x}}{\sigma}\right)^3 \tag{10}$$

$$\text{Kurtosis} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \bar{x}}{\sigma}\right)^4 \tag{11}$$

$$\text{CF} = \frac{\max(|x_i|)}{\left(\frac{1}{N}\sum_{i=1}^{N}x_i^2\right)^{1/2}} \tag{12}$$

$$\text{MF} = \frac{\max(x_i)}{\left(\frac{1}{N}\sum_{i=1}^{N}\sqrt{|x_i|}\right)^2} \tag{13}$$

where x_i is the ASA value at ith sample value of the defect period. The energy of a signal at the J level decomposition is given by:

$$\|S\|^2 = \|A_j\|^2 + \sum_{j\leqslant J}\|D_j\|^2 \tag{14}$$

where A and D are the approximation and detail coefficients obtained following a discrete wavelet transform.

The extracted statistical features are normalized to zero mean and a maximum/minimum of ± 1. The range of tested and selected hyperparameters for each algorithm is summarized in Table 3.

Rauber et al. [40] showed that fault classification accuracy on the CWRU dataset from a data-driven approach can easily reach 99% for several classifiers. However, the industrial relevance of such a result is limited to cases where in-service fault data has been collected and this is a highly restrictive condition.

This work proposes simulation-driven machine learning as a methodology to produce data that is valid for a wide range of conditions (e.g. operating speed, signal-to-noise ratio, physical characteristics) so that it can be applied to bearings in machines without recorded data. A comparison using in-service data is only valid if one experimental dataset is used for training and another dataset from a different machine/configuration is used for scoring. Therefore the data-driven comparisons performed in this work use a classifier trained using the CWRU dataset and are scored on the MPFT and SQ datasets. Fig. 5 shows the performance of several classifiers using statistical features, with the accuracy reported for the optimal hyperparameters of each algorithm. Simulation cross-validation scores are also reported as in the case where no experimental or in-service data is available, the simulation cross-validation scores are the only metric by which one can select an algorithm and hyperparameters.

Table 3
Range of tested hyperparameters and the optimal selected values. The optimal values are found using a grid search and fivefold cross-validation strictly on simulation data.

Algorithm	Range	Optimal
LR	$C = [10^{-2}–10^{3}]$	$C = 10$
RF	[10–100] trees	30 trees
KNN	[10–100] neighbors	30 neighbors
SVM	$C = [10^{-2}–10^{3}], \Gamma = [10^{-1} - 10^{1}]$	$C = 1, \Gamma = 10$
NN	[5–50] neurons/layer, [1–3] layers	5, 30 neurons in 2 layers

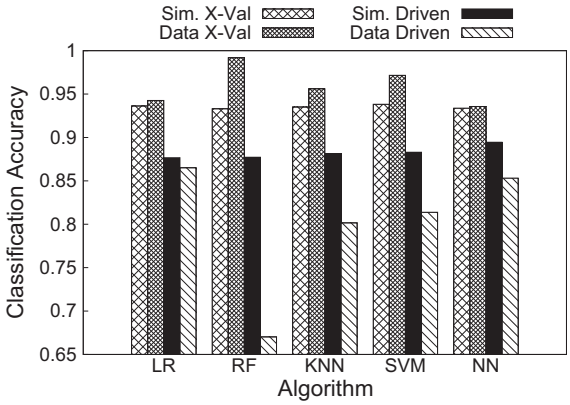


Fig. 5. Classification accuracy comparison for the statistical feature algorithms considered. The hyperparameters for each algorithm are shown in Table 3. Legend above figure: Sim. X-Val represents the accuracy achieved for a classifier trained and scored using purely simulation data; Sim. Driven is the score simulation trained classifiers scored on all seeded-fault experimental data; Data Driven takes CWRU data as training data and MFPT + SQ data for validation. In every case, simulation training data results in greater accuracy than experimental data.

Fig. 5 reveals several interesting key properties. Regarding cross-validation scores, simulation accuracy varies very little between algorithms and data trained algorithms show very high classification accuracy. This confirms that the statistical features selected can capture and very accurately predict defective bearing behavior when in-service data is available for a specific machine. In the simulation and data-driven cases, simulation-driven accuracy is similar whereas the limitations of a data-driven become clear: while data collected from one machine can provide excellent performance in diagnosing identical machines, it performs worse than simulation data when applied to unseen applications. The ability to simulate data representing a wide range of conditions provides a distinct advantage over experimental data when a classifier is used in a novel application. However, optimal classification accuracy is achieved using in-service data that captures the unique characteristics of the machine, its operation and the surrounding environment. Logistic regression is selected as the statistical classifier for future comparisons as it is nearly optimal in the simulation-driven case and optimal in the data-driven case. Its inherent resistance to overfitting and low computational requirements also makes it an attractive choice.

4.2. Convolutional neural network and NNDTW

The well-recognized success of CNNs for image classification has inspired greater investigation of this method in other application areas such as 1D signal analysis for condition monitoring. The previous proposed application of CNNs to identify faults in rotating machinery by Janssens et al. [41] was motivated by the drive to reduce the amount of so-called “expert” or domain-specific knowledge required to apply machine learning to fault detection. Janssens et al. proposed to analyze the power spectrum of accelerometers mounted to rotating machinery using a 1D CNN and to allow the CNN to learn optimal filters to detect faults, rather than using domain-knowledge to derive statistical measures. This method resulted in higher classification accuracy than a statistical feature-based classifier. However, entirely discarding domain knowledge leaves the method lacking robustness: if the machine speed is unsteady, or if the algorithm is applied to a different bearing geometry, there is can be no expectation of success. As shown in Eqs. (3) and (4), the characteristic defect frequencies change with the bearing geometry (number of rollers, pitch/roller diameter) and shaft speed. Including expert knowledge and applying angle synchronous averaging will greatly improve the robustness of a CNN-based methodology because it removes all dependence on bearing geometry, shaft speed, and very importantly, changes in shaft speed during data acquisition. Therefore, we claim that the analysis of bearing faults with 1D CNNs is significantly improved by including domain knowledge. Fig. 6 shows that a single set of angle-synchronous averaged training data simulated one speed and for one bearing geometry

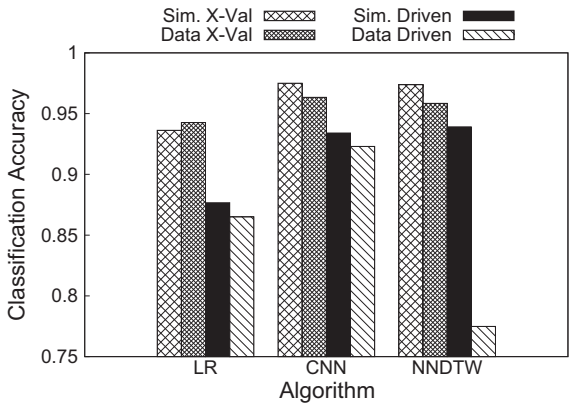


Fig. 6. Comparison of the optimal fault classification algorithms shows that both CNNs and NNDTW are superior to the optimal statistical feature based detection algorithm for simulation and data-driven methods.

can successfully classify race faults in three experimental datasets with different shaft speeds and bearing geometries – something that is impossible using the power spectrum.

While statistical feature extraction attempts to encapsulate as much information as possible using carefully selected statistics and convolutional neural networks learn feature filters that capture the "shape" of a bearing acceleration signal associated to a fault, dynamic time warping provides a direct distance measure between signals with no need for feature extraction or algorithm training.

Fig. 6 shows that both the CNN and NNDTW classifiers exceed the performance of the best statistical feature based classifier in simulation-driven performance, and in the case of the CNN, data-driven applications as well. Data cross-validation accuracy is above the statistical classifier average but lower than random forests. NNDTW is particularly well suited to a simulation-driven condition monitoring approach because the initial data that can be supplied through simulations can be augmented online with in-service fault data as it is recorded, whereas all other classifiers require retraining. The outcome of NNDTW can also provide much greater insight into the fault compared to other methods, particularly in combination with the proposed simulation-driven methodology. Beyond the signal classification, all of the physical parameters associated to the nearest-neighbor match are available, which could give deeper insight into the defect nature and severity. Such insight cannot be gained from a CNN or statistical feature classifiers.

4.3. Convergence characteristics

As noted in a previous section, the number of synchronously averaged defect periods is an important parameter to determine during signal processing. Too few averages will have little noise reducing effect, while too many averages will broaden the expected defect peak due to deviations from the synchronous assumption and cause a decrease in accuracy. Fig. 7 reveals the sensitivity of the highest-scoring algorithms in this work to the number of averaged defect periods. The number of synchronously-averaged defect periods is varied the training and test data is increased, and the accuracy of each algorithm is reported.

All methods exhibit a general trend of increasing accuracy with the number of averaged periods until 50 averaged defect periods. This would indicate that the simulations can accurately capture the stochastic effects – otherwise, the accuracy would decrease as these missed effects accumulate with a large number of averaged periods. It is important to note that at 150 averaged defect periods, only several hundred data points can be created from the experimental data for scoring, whereas thousands are available at lower averaged periods (non-overlapping time samples are extracted from the experimental data, such that if an experimental sample contains 250 defect periods, only a single 150 defect period sample is extracted). The key knowledge to take from Fig. 7 is that for any algorithm, the performance varies little and is nearly optimal over the range of 25–50 averaged defect periods.

Estimating the necessary sample size to reach a performance target is the subject of active research [78,79] given that data-driven machine learning relies instrumentally on the availability of in-service data, and it is often difficult to obtain. The proposed simulation-driven methodology has no constraint on the amount of training data that can be produced; however, several parameters are only specified with a range and each simulation draws a value from a random distribution so a large number of samples may be necessary to cover the high-dimensional space spanned by these variables. The convergence characteristics for the highest performing classifiers and hyperparameters from the statistical, CNN, and NNDTW methods are investigated here for simulation-driven and data-driven machine learning. Here, the CWRU dataset is the training data, and the MFPT and SQ datasets act as test data. Twenty-five defect periods of averaging are used for all calculations. Fig. 8 shows the dependence of algorithm accuracy as a function of the number of training samples in the simulation and data-driven cases. The reported accuracy is an average of five randomly selected training sets for those with $\leqslant$3000 samples.

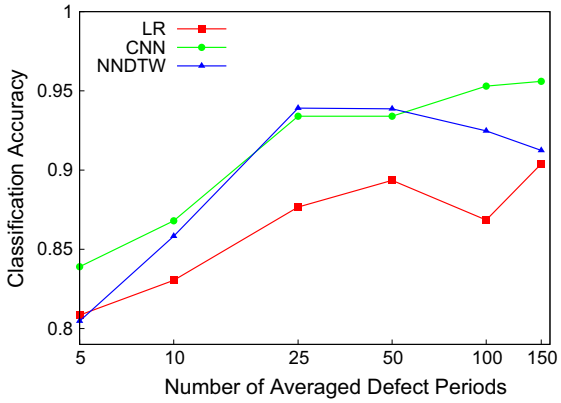


Fig. 7. Effect of the number of averaged defect periods on fault classification accuracy. A general trend of increasing accuracy with increasing synchronous averages is clear; however, the number of test samples at the highest numbers of averages may not be representative. The data shown represents the accuracy of simulation-driven classifiers scored against the CWRU, MFPT and SQ datasets.

Fig. 8 reveals that even for low numbers of training samples, simulation-driven algorithms achieve their asymptotic performance. The data-driven algorithms reach asymptotic performance more slowly, and in the case of CNNs, additional data may improve performance further but it is nevertheless the highest performing data-driven algorithm. In all cases, the simulation driven algorithms outperform their data-driven counterparts. These two trends indicate that a large amount of data is not necessary to detect faults in roller bearings, but what is of much greater importance is the variation in the data. Experimental data production in a laboratory setting is often limited to a select number of faults and conditions that cannot capture the variation seen in an industrial setting, whereas simulation-generated data can augment limited experimental data to improve the classifier's performance in novel applications. An ideal approach combines the two data sources starting with simulation data from deployment and augmenting in-service data as faults appear.

4.4. Precision and recall

As noted in Section 2.4, the classifiers' accuracy is taken as the key performance metric in this work. For completeness, and to ensure that the accuracy values are a reliable indicator of overall performance, the precision and recall of all classifiers is also investigated. Fig. 9 plots precision and recall for all statistical classifiers as well as the CNN and NNDTW classifiers as a function of accuracy.

As shown in Fig. 9, classifier precision exceeds the recall in most instances. The relatively lower accuracy of the data-driven classifiers can be directly attributed to low recall, which reflects the key challenge of data-driven fault detection; if the training data does not contain the same fault and operating characteristics as the signal to be classified, the classifier does not correctly identify the fault. Simulation data can provide training data over a wide range of conditions to enable better fault classification accuracy for novel applications to which in-service data can be added in the future.

4.5. Industrial application: wind turbine bearing fault

The previous sections have established not only the validity but the high degree of performance of simulation-driven machine learning to bearing race fault detection. However, the experimental datasets used to validate the method were generated using seeded faults, rather than faults caused by wear under normal loading conditions. Therefore to further establish the effectiveness of the proposed method, an in-service fault from an industrial application is considered. A crack was found in the inner race of a high-speed bearing of a wind turbine, and fifty days of vibration data preceding the crack discovery have been made available. The data from each of the fifty days is processed and classified using the algorithms. Each ASA acts as a single vote on the defect state, and Fig. 10 shows the average vote as a function of time. The best performing algorithm is that which correctly identifies the inner race fault with the highest confidence but also gives an indication to the presence of a fault as early as possible.

The probability of an outer race fault prediction is highly consistent and below 20%, which provides a baseline for a defect-free race to which the inner race predictions can be compared. The predicted inner race fault state is steady for 10 days, after which it climbs dramatically to nearly 100% probability of a defect by the final day. Comparing the efficacy of each method requires defining a threshold for anomalous behavior. Taking a 50% threshold results in a fault detected 22 days in advance for NNDTW, and 18 for CNN and LR. However, the initial deviation can be seen as early as 38 days in advance. All method clearly detect the fault, and simulation-driven methods consistently predict the defect state with higher accuracy than data-driven methods. Fig. 10 reinforces the value of the proposed simulation-driven methodology for fault

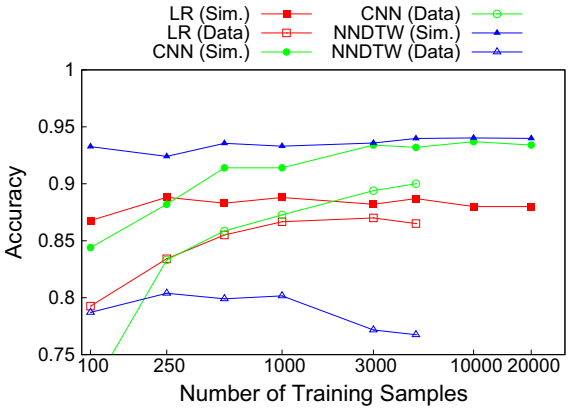


Fig. 8. Convergence characteristics of algorithm accuracy for simulation-driven (solid points) and data-driven (empty points) classifiers. Relatively little data is required to reach convergence, and it is clear that variation in the data is of greater importance than large amounts of data.

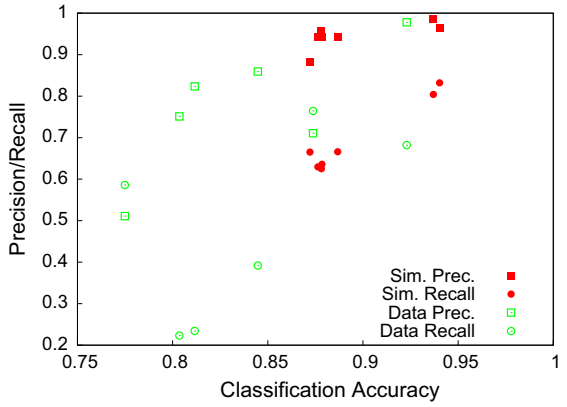
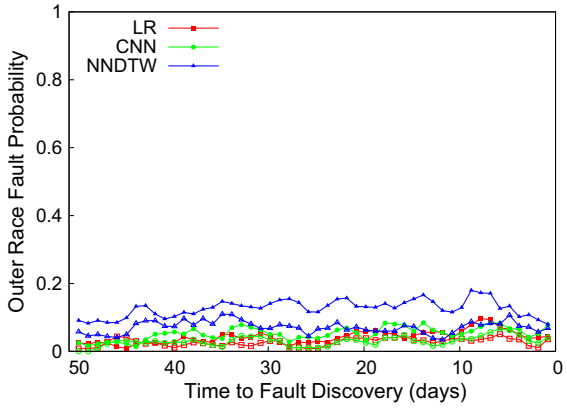
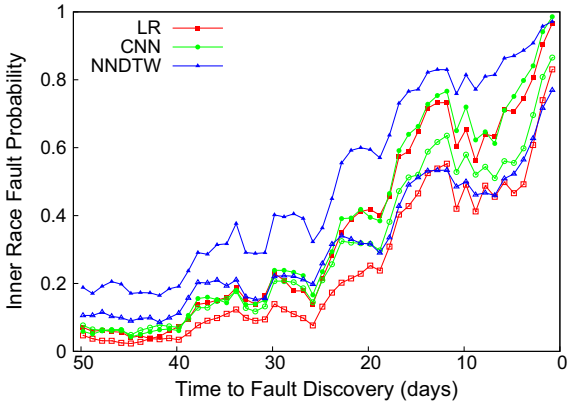


Fig. 9. Simulation-driven and data-driven classifier precision and recall as a function of accuracy. In nearly all cases, precision is notably higher than recall and the simulation-driven case significantly outperforms the data-driven case.



(a) Predicted probability for the presence of an outer race fault.



(b) Predicted probability for the presence of an inner race fault.

Fig. 10. Fault state of the wind turbine high speed bearing as predicted by simulation-driven (solid points) and data-driven (empty points) classifiers. The CNN and NNDTW methods both predict a fault earlier than using statistical features, and the simulation-driven classifiers perform at least as well as data-driven ones.

detection by showing that simulated training data results in equal or greater performance in novel applications compared to experimental data.

5. Conclusion

Data-driven machine learning has received significant attention in the field of condition monitoring and failure prognosis, but its dependence on in-service failure data has impeded widespread deployment in industry. In this work, we show that data generated using relatively simple bearing simulations augmented with stochastic effects can be used to train defect classifiers with high accuracy on four experimental datasets. The simulation-driven machine learning methodology enables condition monitoring from the first day of deployment and is valid independent of bearing configuration and operating condition using a single set of training data. Simulated data cannot replace experimental data derived from the exact conditions under which monitoring is being performed; however, it provides a stronger starting point for novel applications, to which in-service data can be added to further improve classifier performance. Two new applications of machine learning algorithms to fault identification proposed in this work that do not rely on statistical features, convolutional neural networks and NNDTW applied to ASAs, are shown to match or outperform statistical feature based classifiers. The novel application of NNDTW is of particular interest because new data can be included without algorithm re-training and it can give greater insight into the fault process as it identifies the most similar simulation. The proposed methodology can be applied to any (pseudo) cyclostationary defect in a rotating machine, and on a greater scale, simulation data has the potential to improve machine learning in industrial condition monitoring applications.

Acknowledgments

We would like to acknowledge the support of VLAIO and ITEA of the Reflexion project (#14035). We would like to acknowledge helpful discussions with Bram Cornelis. Additionally, the computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided in part by the VSC (Flemish Supercomputer Center), funded by Ghent University, the Hercules Foundation and the Flemish Government department EWI.

References

[1] W. Yang, P.J. Tavner, C.J. Crabtree, Y. Feng, Y. Qiu, Wind turbine condition monitoring: technical and commercial challenges, Wind Energy 17 (5) (2014) 673–693.
[2] A. Bey-Temsamani, A. Van Horenbeek, B. De Ketelaere, L. Pintelon, A. Bartic, Prognostics for optimal maintenance: industrial production capacity optimization using temperature condition monitoring, 2013.
[3] J. Lieblein, M. Zelen, Statistical investigation of the fatigue life of deep-groove ball bearings, J. Res. Nat. Bureau Stand. 57 (5) (1956) 273–316.
[4] T. Tallian, Weibull distribution of rolling contact fatigue life and deviations therefrom, ASLE Trans. 5 (1) (1962) 183–196.
[5] E.J. Tuegel, A.R. Ingraffea, T.G. Eason, S.M. Spottswood, Reengineering aircraft structural life prediction using a digital twin, Int. J. Aerosp. Eng. 2011 (2011).
[6] A.H. Bonnett, C. Yung, Increased efficiency versus increased reliability, IEEE Ind. Appl. Mag. 14 (1) (2008) 29–36.
[7] K. Fischer, F. Besnard, L. Bertling, Reliability-centered maintenance for wind turbines based on statistical analysis and practical experience, IEEE Trans. Energy Convers. 27 (1) (2012) 184–195.
[8] T. Slack, F. Sadeghi, Explicit finite element modeling of subsurface initiated spalling in rolling contacts, Tribol. Int. 43 (9) (2010) 1693–1702.
[9] P. McFadden, J. Smith, Model for the vibration produced by a single point defect in a rolling element bearing, J. Sound Vib. 96 (1) (1984) 69–82.
[10] N. Akturk, The effect of waviness on vibrations associated with ball bearings, J. Tribol. 121 (4) (1999) 667–677.
[11] D. Brie, Modelling of the spalled rolling element bearing vibration signal: an overview and some new results, Mech. Syst. Sign. Process. 14 (3) (2000) 353–369.
[12] S. Sassi, B. Badri, M. Thomas, A numerical model to predict damaged bearing vibrations, J. Vib. Contr. 13 (11) (2007) 1603–1628.
[13] M. Tadina, M. Boltežar, Improved model of a ball bearing for the simulation of vibration signals due to faults during run-up, J. Sound Vib. 330 (17) (2011) 4287–4301.
[14] Y. Shao, J. Liu, J. Ye, A new method to model a localized surface defect in a cylindrical roller-bearing dynamic simulation, Proc. Inst. Mech. Eng. Part J: J. Eng. Tribol. 228 (2) (2014) 140–159.
[15] A.M. Ahmadi, D. Petersen, C. Howard, A nonlinear dynamic vibration model of defective bearings – the importance of modelling the finite size of rolling elements, Mech. Syst. Sign. Process. 52 (2015) 309–326.
[16] J. Slavič, A. Brković, M. Boltežar, Typical bearing-fault rating using force measurements: application to real data, J. Vib. Contr. 17 (14) (2011) 2164–2174.
[17] C. Freitas, J. Cuenca, P. Morais, A. Ompusunggu, M. Sarrazin, K. Janssens, Comparison of vibration and acoustic measurements for detection of bearing defects, in: ISMA, 2016.
[18] R.R. Schoen, T.G. Habetler, F. Kamran, R. Bartfield, Motor bearing damage detection using stator current monitoring, IEEE Trans. Ind. Appl. 31 (6) (1995) 1274–1279.
[19] M.E.H. Benbouzid, M. Vieira, C. Theys, Induction motors' faults detection and localization using stator current advanced signal processing techniques, IEEE Trans. Power Electron. 14 (1) (1999) 14–22.
[20] M.E.H. Benbouzid, A review of induction motors signature analysis as a medium for faults detection, IEEE Trans. Ind. Electron. 47 (5) (2000) 984–993.
[21] N. Tandon, A. Choudhury, A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings, Tribol. Int. 32 (8) (1999) 469–480.
[22] N. Sawalhi, R. Randall, H. Endo, The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis, Mech. Syst. Sign. Process. 21 (6) (2007) 2616–2633.
[23] C. Freitas, P. Morais, J. Cuenca, A.P. Ompusunggu, M. Sarrazin, K. Janssens, Condition monitoring of bearings under medium and low rotational speed, in: European Workshop in Structural Health Monitoring, 2016.
[24] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 454, The Royal Society, 1998, pp. 903–995.

[25] Y. Yang, D. Yu, J. Cheng, A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM, Measurement 40 (9) (2007) 943–950.
[26] S. Prabhakar, A. Mohanty, A. Sekhar, Application of discrete wavelet transform for detection of ball bearing race faults, Tribol. Int. 35 (12) (2002) 793–800.
[27] C. Castejon, O. Lara, J. Garcia-Prada, Automated diagnosis of rolling bearings using MRA and neural networks, Mech. Syst. Sign. Process. 24 (1) (2010) 289–299.
[28] Q. Sun, Y. Tang, Singularity analysis using continuous wavelet transform for bearing fault diagnosis, Mech. Syst. Sign. Process. 16 (6) (2002) 1025–1041.
[29] P. McFadden, M. Toozhy, Application of synchronous averaging to vibration monitoring of rolling element bearings, Mech. Syst. Sign. Process. 14 (6) (2000) 891–906.
[30] R.B. Randall, J. Antoni, S. Chobsaard, The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals, Mech. Syst. Sign. Process. 15 (5) (2001) 945–962.
[31] K. Christian, N. Mureithi, A. Lakis, M. Thomas, On the use of time synchronous averaging, independent component analysis and support vector machines for bearing fault diagnosis, in: First International Conference on Industrial Risk Engineering, 2007.
[32] D. Siegel, H. Al-Atat, V. Shauche, L. Liao, J. Snyder, J. Lee, Novel method for rolling element bearing health assessment: a tachometer-less synchronously averaged envelope feature extraction technique, Mech. Syst. Sign. Process. 29 (2012) 362–376.
[33] D. Abboud, M. Eltabach, J. Antoni, S. Sieg-Zieba, On the extraction of rolling-element bearing fault signature in speed-varying conditions, in: Proceedings of the Eleventh International Conference on Condition Monitoring and. Machinery Failure Prevention Technologies, 2014, pp. 10–12.
[34] M. Sharif, R. Grosvenor, Process plant condition monitoring and fault diagnosis, Proc. Inst. Mech. Eng. Part E: J. Process Mech. Eng. 212 (1) (1998) 13–30.
[35] A. Kusiak, H. Zheng, Z. Song, Short-term prediction of wind farm power: a data mining approach, IEEE Trans. Energy Convers. 24 (1) (2009) 125–136.
[36] A. Kusiak, A. Verma, A data-driven approach for monitoring blade pitch faults in wind turbines, IEEE Trans. Sustain. Energy 2 (1) (2011) 87–96.
[37] A. Widodo, E.Y. Kim, J.-D. Son, B.-S. Yang, A.C. Tan, D.-S. Gu, B.-K. Choi, J. Mathew, Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine, Expert Syst. Appl. 36 (3) (2009) 7252–7261.
[38] F. Camci, K. Medjaher, N. Zerhouni, P. Nectoux, Feature evaluation for effective bearing prognostics, Qual. Reliab. Eng. Int. 29 (4) (2013) 477–486.
[39] A. Ziaja, I. Antoniadou, T. Barszcz, W.J. Staszewski, K. Worden, Fault detection in rolling element bearings using wavelet-based variance analysis and novelty detection, J. Vib. Contr. 22 (2) (2016) 396–411.
[40] T.W. Rauber, F. de Assis Boldt, F.M. Varejão, Heterogeneous feature models and feature selection applied to bearing fault diagnosis, IEEE Trans. Ind. Electron. 62 (1) (2015) 637–646.
[41] O. Janssens, V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufier, S. Verstockt, R. Van de Walle, S. Van Hoecke, Convolutional neural network based fault detection for rotating machinery, J. Sound Vib. 377 (2016) 331–345.
[42] J. Chen, R. Randall, B. Peeters, W. Desmet, H. Van der Auweraer, Artificial neural network based fault diagnosis of ic engines, Key Engineering Materials, vol. 518, Trans Tech Publ, 2012, pp. 47–56.
[43] J. Chen, Internal Combustion Engine Diagnostics using Vibration Simulation Ph.D. thesis, The University of New South Wales, 2013.
[44] Y.L. Murphey, M.A. Masrur, Z. Chen, B. Zhang, Model-based fault diagnosis in electric drives using machine learning, IEEE/ASME Trans. Mechatron. 11 (3) (2006) 290–303.
[45] M. Elkordy, K. Chang, G. Lee, Neural networks trained by analytically simulated damage states, J. Comput. Civ. Eng. 7 (2) (1993) 130–145.
[46] E. Keogh, C.A. Ratanamahatana, Exact indexing of dynamic time warping, Knowl. Inform. Syst. 7 (3) (2005) 358–386.
[47] E. Bechhoefer, M. Kingsley, A review of time synchronous average algorithms, in: Annual Conference of the Prognostics and Health Management Society, San Diego, CA, September 2009, pp. 24–33.
[48] J.F. Monahan, Numerical Methods of Statistics, Cambridge University Press, 2011.
[49] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[50] F. Chollet, Keras. <https://github.com/fchollet/keras>, 2015.
[51] S. Salvador, P. Chan, Toward accurate dynamic time warping in linear time and space, Intell. Data Anal. 11 (5) (2007) 561–580.
[52] J. Yan, J. Lee, M. Koc, Predictive algorithm for machine degradation detection using logistic regression, in: Fifth International Conference on Managing Innovations in Manufacturing (MIM), 2002, pp. 172–178.
[53] C. Mechefske, J. Mathew, Fault detection and diagnosis in low speed rolling element bearings Part II: The use of nearest neighbour classification, Mech. Syst. Sign. Process. 6 (4) (1992) 309–316.
[54] A. Sharma, V. Sugumaran, S.B. Devasenapati, Misfire detection in an ic engine using vibration signal and decision tree algorithms, Measurement 50 (2014) 370–380.
[55] L. Jack, A. Nandi, Support vector machines for detection and characterization of rolling element bearing faults, Proc. Inst. Mech. Eng. Part C: J. Mech. Eng. Sci. 215 (9) (2001) 1065–1074.
[56] I. Alguindigue, R.E. Uhrig, Vibration monitoring with artificial neural networks, in: Proceedings of SMORN-VI, Gatlinburg 59.
[57] Y. LeCun, Y. Bengio, Convolutional networks for images, speech, and time series, Handbook Brain Theory Neural Netw. 3361 (10) (1995) 1995.
[58] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
[59] Y. Zheng, Q. Liu, E. Chen, Y. Ge, J.L. Zhao, Time series classification using multi-channels deep convolutional neural networks, in: International Conference on Web-Age Information Management, Springer, 2014, pp. 298–310.
[60] N. Kalchbrenner, E. Grefenstette, P. Blunsom, A convolutional neural network for modelling sentences. Available from: <arXiv:1404.2188>.
[61] H. Sakoe, S. Chiba, Dynamic programming algorithm optimization for spoken word recognition, IEEE Trans. Acoust. Speech Sign. Process. 26 (1) (1978) 43–49.
[62] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, E. Keogh, Querying and mining of time series data: experimental comparison of representations and distance measures, Proc. VLDB Endowment 1 (2) (2008) 1542–1552.
[63] T. Rakthanmanon, B. Campana, A. Mueen, G. Batista, B. Westover, Q. Zhu, J. Zakaria, E. Keogh, Searching and mining trillions of time series subsequences under dynamic time warping, in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2012, pp. 262–270.
[64] D. Zhen, T. Wang, F. Gu, A. Ball, Fault diagnosis of motor drives using stator current signal analysis based on dynamic time warping, Mech. Syst. Sign. Process. 34 (1) (2013) 191–202.
[65] Y. Dai, J. Zhao, Fault diagnosis of batch chemical processes using a dynamic time warping (DTW)-based artificial immune system, Ind. Eng. Chem. Res. 50 (8) (2011) 4534–4544.
[66] F. Cong, J. Chen, G. Dong, M. Pecht, Vibration model of rolling element bearings in a rotor-bearing system for fault diagnosis, J. Sound Vib. 332 (8) (2013) 2081–2097.
[67] M. Nakhaeinejad, Fault detection and model-based diagnostics in nonlinear dynamic systems.
[68] Siemens LMS Imagine.Lab Amesim (version 15). <https://www.plm.automation.siemens.com/en_us/products/lms/imagine-lab/amesim/>, 2016.
[69] Y.H. Wijnant, Contact Dynamics in the Field of Elastohydrodynamic Lubrication, Universiteit Twente, 1998.
[70] J.A. Wensing, On the Dynamics of Ball Bearings, University of Twente, 1998.
[71] A. Ragheb, M. Ragheb, Wind turbine gearbox technologies, in: 2010 1st International Nuclear & Renewable Energy Conference (INREC), IEEE, 2010, pp. 1–8.
[72] P. Bak, C. Tang, K. Wiesenfeld, Self-organized criticality: an explanation of the 1/f noise, Phys. Rev. Lett. 59 (4) (1987) 381.

[73] B. Chen, Z. Yan, W. Chen, Defect detection for wheel-bearings with time-spectral kurtosis and entropy, Entropy 16 (1) (2014) 607–626.
[74] J. Miettinen, P. Andersson, Acoustic emission of rolling bearings lubricated with contaminated grease, Tribol. Int. 33 (11) (2000) 777–787.
[75] M. Maru, R. Castillo, L. Padovese, Study of solid contamination in ball bearings through vibration and wear analyses, Tribol. Int. 40 (3) (2007) 433–440.
[76] N. Tandon, G. Yadava, K. Ramakrishna, A comparison of some condition monitoring techniques for the detection of defect in induction motor ball bearings, Mech. Syst. Sign. Process. 21 (1) (2007) 244–256.
[77] W.A. Smith, R.B. Randall, Rolling element bearing diagnostics using the case western reserve university data: a benchmark study, Mech. Syst. Sign. Process. 64 (2015) 100–131.
[78] P. Blamire, The influence of relative sample size in training artificial neural networks, Int. J. Rem. Sens. 17 (1) (1996) 223–230.
[79] R.L. Figueroa, Q. Zeng-Treitler, S. Kandula, L.H. Ngo, Predicting sample size required for classification performance, BMC Med. Inform. Decis. Mak. 12 (1) (2012) 1.