2023 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)
June 28-30, 2023. Seattle, Washington, USA

Early Inner Race Fault Detection On A Ball Bearing Setup Using Histogram of Oriented Gradients and Wavelet Subselection

Cedric Van Heck, Jolan Wauters, Tom Staessens, Guillaume Creveceour and Ted Ooijevaar

Abstract— **Predictive maintenance is an industrial practice to detect component failure ahead in time and before major damage is done to the system. Bearings are susceptible to such damage phenomena and should be replaced before critical failure, therefore early detection proves important. For wide applicability, these detection methods should work on easily available sensor data and have limited computational complexity. This study presents a method for classifying the health of a bearing based on the machines vibrational sensor data and with additional focus on computational complexity. The process involves converting the signals to the frequency spectrum using continuous wavelet transforms, and identifying specific frequency ranges associated with damage phenomena in bearings. Two approaches were used: analyzing the wavelet responses and creating a scalogram image to locate relevant areas. The results obtained on a bearing monitoring data set, created within the Flanders AI Research program, were consistent for both approaches and identified a specific set of scales that resulted in reduced computational load whilst attaining high failure detection rates. Consolidation is achieved by repeating the procedure on two public data sets.**

I. Introduction

Bearings are one of the most ubiquitous components in industrial machines, especially in rotating parts such as electromechanical systems driven by motors. They allow for smooth relative motion (i.e. without much friction) and ensure that load forces and vibrations are diverted from the operating machine to the supporting structure [1]. These bearings however suffer from degradation phenomena such as fatigue, wear, fractures and more, ultimately hindering its proper functioning and therefore exposing the rest of the machine to excessive forces. Given that bearings are inexpensive compared to the total cost of a machine, timely detection and replacement of a degrading or damaged bearing is desirable.

Prognostic health management has been applied to bearings in a multitude of approaches. The most effective techniques ensure that operation should not be halted to evaluate the health state. Therefore methods primarily rely on sound, vibration, current and temperature sensor data. Conventional methods monitor these signals and analyze them using time-domain synchronous average analysis, envelope analysis, kurtosis and many more [2]. Another line of research focuses on the frequency domain representation, obtained by using the Fast Fourier Transform and revealing signal characteristics that were hidden or hard to notice in time-domain. This approach is particularly useful for signals with lots of noise and cyclic behaviour, such as rotating machinery. The frequency response is then combined with detection algorithms such as envelope detection analysis, frequency energy analysis and machine learning methodologies [3], [4]. A combination of both domains can be found by using time-frequency transformations such as the Short Time Frequency Transform [5], Continuous Wavelet Transform [6] and Hilbert-Huang Tranform [7], where frequency information is captured in function of time.

Another emergent research direction in health monitoring is the application of data-driven models. In [8], a nonlinear degradation model is proposed and updated using Expectation Maximization in order to predict remaining useful lifetime of a bearing setup directly from vibrational data. Other approaches combine the (time-)frequency domain with data-driven methodologies. For example, graph theory is combined with wavelet packet decomposition in [9] whilst in [10] adaptive wavelets are introduced and combined with autoencoders for fault diagnosis. Deep neural networks are also applied in a variety of forms to health monitoring, particularly Convolutional Network based Fault Diagnosis (CNFD), which has achieved state of the art performance in many benchmarks [11]. These type of networks are mostly known for image processing and hence often make use of a spectogram or scalogram image created from the frequency response. Another approach is using computer vision and feature extraction methods, such as Scale Invariant Feature Transform, Histogram of Oriented Gradients and Speeded-Up Robust Features [12], [13], which are local methods (i.e. unlike in CNNs, the location within the image is of importance). These features can be put together with a variety of machine learning methodologies such as regression methods and tree based models to obtain a classification methodology.

Given this abundance of possibilities, it remains a challenge to find the optimal combination of features and algorithms. Not only should these algorithms be compared in terms of accuracy, but also in terms of computational complexity, greatly influencing the applicability in practice. CNNs for example are known to achieve unseen accuracies, however at the same time often consist out of hundreds or thousands of parameters. This results in high training and evaluation times, therefore hindering the implementation in real-time monitoring systems with limited computational power such as programmable logic controllers. Opting for

C. Van Heck, J. Wauters, T. Staessens and G. Crevecoeur are with the Department of Electromechanical, Systems and Metal Engineering, Ghent University, B-9052 Ghent, Belgium e-mail: {cevheck.vanheck, jolan.wauters, tom.staessens, guillaume.crevecoeur}@ugent.be.

C. Van Heck, J. Wauters, T. Staessens and G. Crevecoeur are member of MIRO, Flanders Make, Belgium. T. Ooijevaar is a Senior Research Engineer at Flanders Make, Belgium.

978-1-6654-7633-1/23/$31.00 ©2023 IEEE

alternatives, such as the previously mentioned local feature extraction methods, can allow the usage of simpler models hence leading to lower computational burden. Additionally, not only the size of the model, but also the size of the feature set affects the computational complexity. This for a twofold of reasons. Firstly, training and validation of machine learning models, in almost all cases, scales proportionally to the feature dimension. Secondly, features are not always easily obtained and might involve preprocessing steps (e.g. the creation of a scalogram). Obviously in both cases, a reduced feature set size greatly ameliorates the algorithm's computational load and hence applicability. Feature selection is a crucial step in a machine learning pipeline and is also known to increase model performance and interpretability. A nice overview on possible selection methodologies is found in [14], ranging from statistical tests [15], [16] to genetic algorithms, to exploiting models with an inherent metric of feature importance.

This paper proposes an iterative procedure for bearing health classification with additional attention for computational complexity, leveraging on feature selection. A dual approach is investigated where a continuous wavelet transform is both used directly and preprocessed into a scalogram image where Histogram of Oriented Gradients (HOG) is used to extract important regions. These features are combined with a Random Forest Classifier for health classification. Physical intuition about damage phenomena is leveraged as it is expected to capture such phenomena in a limited set of wavelet responses. Both approaches are used concurrently and confirm the existence of a limited subset of important frequencies by using feature selection methods. The proposed wavelet subselection technique, utilizing HOG, represents a novel methodology for enhancing computational efficiency and constitutes a significant contribution of this paper. The proposed procedure is applied to a new data set, created by Flanders Make [17] in the context of the Flanders AI Research (FLAIR [18]) program, and shows that inner bearing defects can be detected accurately with as little as two wavelet components. The results are verified on two publicly available data sets and corroborate the validity of the methodology.

II. Condition Monitoring on Bearings: Data Sets

Within this paper, the proposed method is validated on a multitude of data sets. First off, a data set has been created within the scope of the Flanders AI Research program (FLAIR) [18]. The data set, produced by Flanders Make [17], encompasses vibrational sensor data from a mixture of healthy and damaged bearings. Section II-A provides a more in depth overview of the data set. Section II-B handles two of the most renowned public data sets, on which the methodology will be verified. All of these data sets have in common that operational conditions are at quasi-static rpm and for the damaged bearings an indent has been inflicted prior to the start of the experiment. Therefore a discrete classification problem is presented, i.e. a bearing is either healthy or faulty, rather than a continuous health metric.

Fig. 1. Fleet of 4 bearing health monitoring setups at the Flanders Make lab [17]. A second identical test rig is present with 3 more setups.

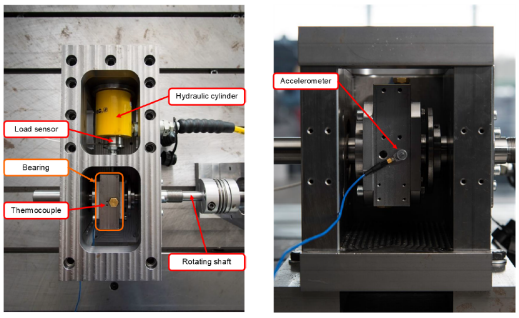

Fig. 2. Top view (Left) and front view (Right) on sensor locations within the experimental test setups.

A. FLAIR Data Set

With the aim of creating data suitable for bearing fault diagnostic and prognostic methods, seven identical setups have been created as seen in Fig. 1. On each of the separate setups, 10 bearings of type 6205-CTVH have been put into operation at a fixed rotational speed of 2000rpm with additional radial load of 9kN. Out of those 10 bearings, 7 were artificially damaged by introducing an indent at the inner race of $400\pm25\mu$m, leading to a total of 49 damaged bearing experiments and 21 healthy examples. Vibrational data is acquired at 50kHz by use of an accelerometer, as depicted in Fig. 2. Additional sensors are placed to monitor radial force and temperature inside the housing. Experiments were continued until one of the test stopping criteria, shown in Table I, was obtained.

TABLE I
Test Stopping Criteria

Bearing health	Stopping criteria
Healthy	Temperature stable for 15min and test duration of at least 2h
Unhealthy	30 min after measured acceleration exceeds 5g
	Acceleration exceeds 20g
	Temperature exceeds 70°C

B. Public Data Sets

For validation and benchmarking purposes, publicly available data sets have been examined. An overview of most known data sets on bearing health monitoring is found in Jiao et al. [11]. Considering the setting of the previously mentioned data set, i.e. fixed rotational speed, initial indent on the inner race and vibrational measurements, two data sets have been selected.

1) Case Western Reserve University: Case Western Reserve University (CWRU) provides a comparable test situation with bearing health monitoring at an operational speed between 1720 and 1797rpm [19]. Defects were introduced in the bearings through the use of electrical discharge machining. The diameters of the defects are of varying size (177.8, 355.6, 533.4, 711.2μm) and are introduced in distinct locations (inner raceway, rolling element and outer raceway). To ensure comparability to the FLAIR data set, we restricted ourselves to the inner raceway defects. The resulting data set contains 4 experiments of nominal bearing operation and 24 measurements of bearings corrupted with inner race faults.

2) Society for Machinery Failure Prevention Technology: Society for Machinery Failure Prevention Technology (MFPT) contributed another data set similar to those mentioned before [20]. Data is collected from the test rig operated at 1500rpm and subject to varying loads. Different fault have been induced and we focus on the inner bearing faults. This leads to three measurements of nominal (i.e healthy) operation, each with a duration of 6 seconds and 7 measurements of bearings with inner race defects, each lasting 3 seconds.

For each of the above defined data sets, the goal is to classify the damaged bearings from the healthy ones. The next Section will handle in depth the used method, whilst the Section thereafter will handle the results. In this study, we will primarily analyze the results obtained from the FLAIR data set. This focus is warranted due to the FLAIR program serving as the context for the present research and due to the larger size of the FLAIR data set in comparison to the other data sets.

III. Methodology

The proposed methodology is presented in this section in a three-step process. Section III-A details the utilization of the Continuous Wavelet Transform (CWT) for the conversion of raw vibrational data into the time-frequency domain. Subsequently, Section III-B describes the conversion of these frequency responses into two feature sets that can be utilized in conjunction with various classification algorithms. Lastly, Section III-C addresses the feature reduction of both feature sets to attain a framework with reduced computational complexity. Fig. 5 shows an overview of the three steps combined in a single framework.

A. Continuous Wavelet Transform

The most accessible data obtained from bearings are the raw sound and vibration signals. Therefore, to ensure applicability, most health monitoring techniques start from either of these signals. Both signals are high-frequent and hard to interpret and are consequently preprocessed into more tangible features. Spectral analysis does exactly this as it summarizes a time-varying signal in terms of frequency-dependent quantities. The most prevalent method for doing so is the Fourier Transform, however, other methods exist as well such as the Hilbert-Huang Tranform [7], Continuous Wavelet Transform [6], Synchrosqueezing Transform [21] and many more. We opted for the use of the Continuous Wavelet Transform (CWT) because of its varying dilation, which allows to efficiently capture mixtures of high- and low frequency vibrations [22], as present in a damaged operational bearing. Additionally, CWT is known to be more sparse than the Fourier Transform [23] and will hence result in most of the information about the signal being stored in only a few coefficients.

The wavelet transform is a mathematical tool used to decompose a time-domain signal, denoted as $x(t)$, into a set of wavelets, $\psi(t)$. This transform is achieved through a convolution of the signal with the wavelets at various scales, where the varying scales allow for the wavelets to effectively stretch or compress in length. The result is an input-output mapping, where, at a certain time instance and for a defined wavelet scale a, the wavelet response, $\mathcal{C}$, is calculated by using (2). The structure of the wavelet $\psi(t)$ has yet to be defined and can take on many forms [24]. For CWT, the most commonly used form of the wavelet is a complex sine wave with Gaussian envelope, i.e. the Morlet wavelet [24].

$$\psi_{Morlet}(t) = e^{-\alpha t^2} e^{j2\pi f_c t} \tag{1}$$

And the transform is defined as:

$$\mathcal{C}(a,b) = \frac{1}{\sqrt{|a|}} \int x(t)\psi(t)^* (\frac{t-b}{a}) dt \tag{2}$$

With * denoting the complex conjugate and

- f_c the center frequency, taken fixed as $\frac{5}{2\pi}$Hz
- α the width of the Gaussian, taken fixed as 1/2
- a the scale of the wavelet
- b the time shift of the wavelet

B. Feature Engineering

In Machine Learning an important step in obtaining a well-performing model is finding a good set of features for the desired task.

1) Energy Response: A first step has been taken by reducing the raw vibration signals from the setups from Section II to their wavelet responses. Additionally, complete measurements are split up into separate samples with a length of $2s$. By doing so, the algorithm is enforced to be able to distinguish healthy and damaged bearings with a limited amount of information and this a substantial amount of time prior to critical failure. Furthermore it is important to note that a damage phenomenon will result in increased activity at a certain frequency. Therefore we limit ourselves to the energy response, being the squared magnitude of

the coefficients $|\mathcal{C}|^2$. Denoting $\mathcal{C}_a$ as the response for a wavelet with scale a, the response can be further reduced by averaging over the time-axis. A single response value is obtained for each wavelet, being $x_a = \frac{1}{N}\sum_0^N |\mathcal{C}_a|^2$, with N the amount of samples in 2 seconds of simulation (dependent on the used setup). Therefore, by focusing on a fixed set of wavelet scales, a one-dimensional feature vector $\mathbf{X}$ is obtained. This set of scales is initially chosen from 2 to 155, corresponding to a frequency range from 20.3kHz to 264Hz, comprising most bearing fault frequencies [25].

$$\mathbf{X} = [x_i | i = 2:155] \tag{3}$$

By combining the attained feature vector with known classification algorithms, in this case taken to be a Random Forest classifier [26], a first pipeline from signal to classification is obtained.

2) Histogram of Oriented Gradients (HOG): The same information can also be displayed by using a two-dimensional set of features. This is done by converting the time-dependent wavelet transform from (2) into a scalogram, i.e. a time-frequency plot where coefficients are shown using a heatmap. An example of such a scalogram is shown in Fig. 3.

This approach has received growing interest as a result of data-driven image recognition techniques, most prominently CNNs. Although great results in terms of accuracy have been obtained by using a CNN in combination with the scalograms, training and evaluation of the network proved to be time-consuming and therefore alternative processing of the image has been looked at with the goal of achieving real-time health monitoring. Alternative methods often do not look at pixel values directly but try to detect shapes or edges. One such alternative is HOG, mostly known from computer vision and object detection. The main disadvantages when compared to CNNs is that HOG is not translational- and rotational invariant. However, due to the nature of the current use case this forms no limitation. The algorithm is further explained using Fig. 4.

The original m × m scalogram image is rescaled to a n × n image and divided into cells of c × c pixels. For each pixel in this cell, the pixel gradients (relative to surrounding pixels) are calculated resulting in a gradient magnitude g and direction θ.

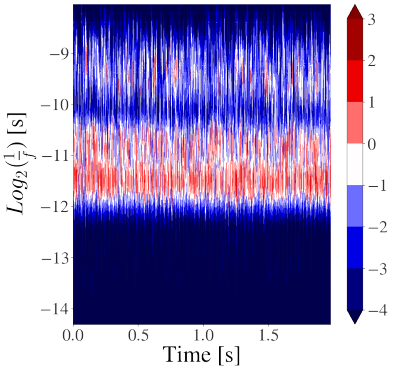

Fig. 3. Example of a scalogram on a single 2 second sample from the FLAIR data set.

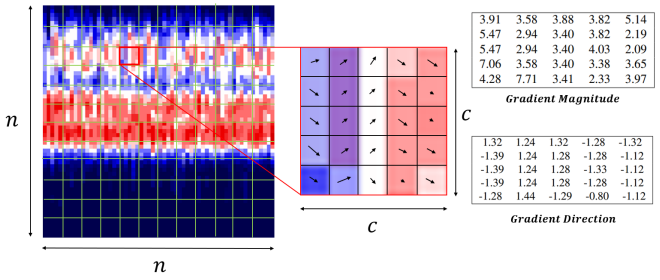

Fig. 4. Histogram of Oriented Gradients applied to a scalogram (adapted from [27]). Left part shows the original image. Middle part and right part show the gradients in a certain cell both visually as in numbers.

These gradient-orientation couples (g, θ) are grouped in a single histogram (dividing the possible gradient directions in d bins) by proportionally adding each gradient magnitude to the two directional bins most close to the gradient direction (h_i and h_j).

$$\begin{cases} h_i = g\frac{\theta-\theta_j}{\theta_i-\theta_j} \\ h_j = g\frac{\theta-\theta_i}{\theta_j-\theta_i} \end{cases} \tag{4}$$

Good practice adds the usage of unsigned gradients and histogram normalization (i.e. normalizing the histograms over blocks of b × b cells). For a hands-on example and implementation we refer to [27].

Concatenating the normalized histograms results in a single vector $\mathbf{X}_{HOG}$ with length $(\frac{n}{c} - (b-1))^2 b^2 d$ whilst directly using the pixels of an m × m RGB image would result in $3m^2$ feature values. Hence, by appropriate choice of the variables in the HOG methodology, the feature vector is greatly reduced in dimension and consequently more suitable to be combined with most traditional classification algorithms. Again we opted for Random Forest, analogous to the previous method. The parameters are taken as follows:

- n = 60
- c = 5
- d = 8
- b = 2
- m = 100

Resulting in a reduction in dimension from 10000 features to 3872 and a more informative feature set than directly using the pixel values.

C. Feature Reduction

Although using a full set of informative features leads to optimal capability of a classification model to capture underlying relations, it is not often desired due to a number of reasons. Firstly, using all of the features leads to increased computational load. Not only does the model need to learn more parameters and relationships during both training and evaluation, it also needs to create additional features. In this case the latter will result in more wavelets being looked at and convolved against the time signal, being quite an expensive operation. Secondly, reducing the amount of features often leads to a more simple model, therefore being less prone to overfitting.

In order to make the subselection of features, two methods have been employed. A first method uses statistical tests such as χ^2 [15] and ANOVA F-values [16] to evaluate the dependence between a feature and the classification label. One could select the k features with highest correlation to the output, and evaluate the performance with such a subselection. A popular alternative is using a classification model with an inherent metric of feature importance. A common example is using the coefficients in linear- and logistic regression, however many alternatives exist. A similar concept of feature importance exists with Random Forests, where features are ranked according to their Mean Decrease Impurity importance (MDI) [28].

Given the current understanding of bearing classification, it is acknowledged that a bearing fault often occurs at specific frequencies. These frequencies are mostly a function of the operated bearing (amount of bearing balls, bearing diameter etc.) and the rotational speed. In an industrial setting one can often assume that these are constant or limited to a small set of operating conditions. Under this assumption, it becomes valuable to focus on those frequencies where the bearing fault occurs, rather than the complete frequency spectrum. For both of the approaches stated in Section III-B such a reduction is possible. When investigating the features selected by the selection techniques, it is possible to trace back those features to a corresponding wavelet scale and pseudo-frequency (i.e. frequency of a wavelet [29]). For the first method this is straightforward as each feature in $\mathbf{X}$ directly relates to the selected wavelet coefficient. A similar deduction can be made on the HOG feature set, however with some additional effort. The feature in the concatenated array $\mathbf{X}_{HOG}$ can be traced back to the row of the c $\times$ c cell in the scalogram by using formulas (5)-(7), with $\{\mathbf{X}_{HOG}, C, B, R\}_i$ being respectively the i^th feature, cell, block and row.

$$C_i = \left\lceil \frac{\mathbf{X}_{HOG,i} + 1}{d} \right\rceil, \tag{5}$$

$$B_i = \left\lceil \frac{C_i}{b^2} \right\rceil, \tag{6}$$

$$R_i = \left\lceil \frac{B_i}{(\frac{n}{c} - (b-1))^2} \right\rceil + ([C_i - (B_i - 1)b^2] > 2) \tag{7}$$

Looking back at Figure 3, one can translate the cell row to a range on the y-axis, $[y_{min}, y_{max}]$, and thus the important frequency range is found using the following:

$$f_i = 2^{-y_i}, \qquad i \in [min, max] \tag{8}$$

These frequency ranges overlap with the frequencies corresponding to the used wavelets. As a consequence, the selected HOG features can be related to a corresponding frequency and as such, a subset of wavelets can be used to create reduced future sets $\mathbf{X}_{red}$ and $\mathbf{X}_{HOG,red}$. An overview of the complete process from vibrational signals to classification is seen in Fig. 5.

IV. Results

The proposed methodology has been implemented and validated on all of the data sets presented in Section II. IV-B shows the results on the FLAIR data set in an extensive manner, in the Sections thereafter repeatability on public data sets is demonstrated. The evaluation is performed in a systematic manner as elaborated in the upcoming Section.

A. Evaluation Procedure

When investigating bearing health classification, accuracy is an often used initial measure of performance. However due to the high class imbalance in the data sets, accuracy does not suffice and other metrics such as sensitivity and specificity are taken into account. These values can be summarized in an Area Under Curve (AUC) value [30]. Cross-validation is performed to demonstrate repeatability of the results [31]. Since 2 second samples obtained from a single measurement run are highly correlated, they are aggregated in the same fold in order to avoid leakage in between folds [32].

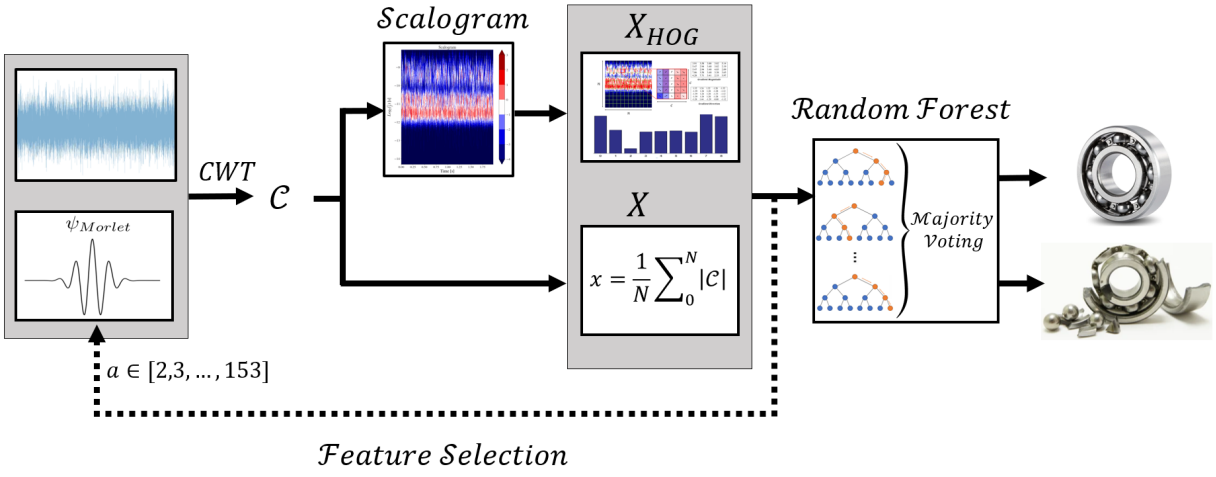

Fig. 5. Overview of the proposed methodology. Raw vibrational signals are transformed using CWT and further converted in two feature sets ($\mathbf{X}$, $\mathbf{X}_{HOG}$). Bearing health classification is performed using Random Forest classifiers and optional feature reduction, zooming in on specific frequencies, is added.

TABLE II

FLAIR DATA SET RESULTS USING FULL SET OF FEATURES

Features	Fold	Accuracy [%]	AUC [%]
$\mathbf{X}$	1	93.67 ± 1.64	96.72 ± 0.61
	2	90.75 ± 1.16	84.64 ± 2.82
	3	97.14 ± 0.09	99.39 ± 0.17
$\mathbf{X}_{HOG}$	1	96.67 ± 0.03	98.58 ± 0.05
	2	94.35 ± 0.04	97.16 ± 0.20
	3	97.40 ± 0.05	99.47 ± 0.03

The model used in both cases is a state-of-practice implementation of a Random Forest. Without much tuning effort, promising results are obtained, thus showing the viability of the proposed framework. A random forest consisting of 30 trees, each having a maximum depth of 10 splits, is trained with additional class reweighting using the Intel(R) extension for Scikit-learn for improved computational performance [33].

B. FLAIR Data Set

First results can be obtained without any feature subset selection. Evaluation is performed five times and the mean and standard deviation are depicted in Table II. From these results it is apparent that both methods obtain accurate bearing health predictions. Additionally, the results indicate that the more intensive feature engineering process of HOG leads to superior results.

In the subsequent experiment the same models are evaluated, however using only a limited set of features. These features have been determined through selection procedures previously described. The number of selected features, denoted as k, is varied in the experiment, and the performance of the models is calculated on all folds and averaged out. The outcome of this experiment, depicted in Figures 6 and 7, shows the results on feature sets $\mathbf{X}$ and $\mathbf{X}_{HOG}$ respectively. Similar to the previous experiment, the results shown in these figures are based on the average of 5 repetitions.

The results indicate that, even with a drastic reduction in the number of features, similar and even slightly higher

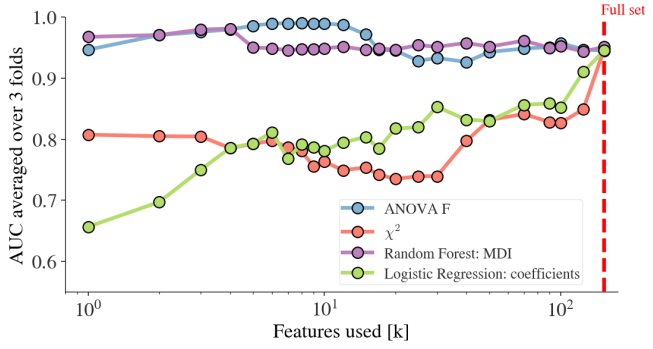

Fig. 6. Resulting model using a variable (k) amount of features out of the feature set $\mathbf{X}$, selected by a variety of methods.

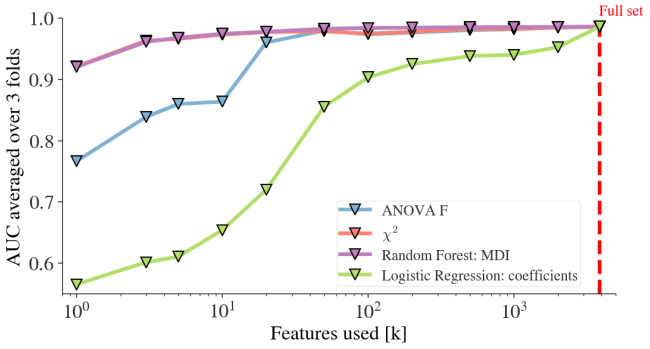

Fig. 7. Resulting model using a variable (k) amount of features out of the feature set $\mathbf{X}_{HOG}$, selected by a variety of methods.

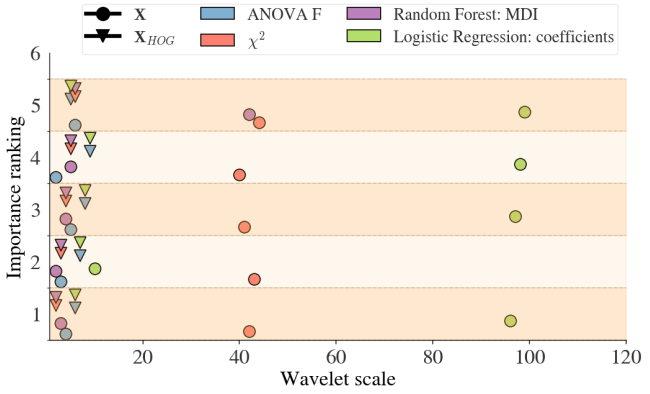

Fig. 8. Feature importance calculated back to 5 most important wavelet scales and shown for different feature sets and feature selection methods.

AUC values are obtained. When investigating those features selected by the selection techniques, it is possible to trace them back to a corresponding wavelet scale and frequency (See Section III-C). Using (5)-(8), the inverse transformation from selected HOG feature to wavelet scale is performed and Fig. 8 shows an overview of the five most important wavelets for each of the methods. Importance ranking is depicted on the y-axis, where importance ranking 1 indicates the most important wavelet scale. Relative height within the band is of no importance, but is added for visual clarity. The findings reveal that the top-performing feature selection methods consistently select the five lowest scales for both feature sets.

C. Computational Reduction

As mentioned in Section III-C, there is another major advantage to feature reduction. Apart from the increased performance and reduced model complexity, it is clear from Figures 6 and 7 that using as little as two wavelets results in increased model performance. As a result, for a reiteration of the model or for new incoming data, one could go to a smaller set of features. Applied to the FLAIR data set, this reduction, going from 153 wavelets to merely 2 (a = 2,3), reduced the total feature creation time from 94h to 18min. In case of the second feature set, which needs

scalogram creation, a similar gain is obtained, going from 375h to 8.6h. Calculations were performed on an 11th Gen Intel(R) Core(TM) i7-11700K @ 3.60GHz with 8 cores, dual threading and 32GB of RAM. For the first reduced feature set, $\mathbf{X}_{red} = [x_2, x_3]$, a subset of the complete $\mathbf{X}$ is obtained and hence performance has already been validated (see Fig. 6 with k = 2). However, by using a limited set of wavelets, the scalogram image changes, instead of being a subregion of the previously created image, resulting in new HOG features rather than a subset of the ones presented earlier. Therefore it deems necessary to evaluate the model performance on the new feature set $\mathbf{X}_{HOG,red}$. Table III shows that the proposed feature reduction indeed leads to a comparable model.

D. Public Data Sets

The same methodology and steps are applied on the public data sets proposed in Section II. All results pre- and post feature selection are presented in Table III, where again only two scales were allowed as subselection. The results prove to be repeatable on a variety of data sets and consistently show almost no decrease, and even an increase for feature set $\mathbf{X}$, in accuracy when moving to a reduced set of features. The more extensive feature selection method, $\mathbf{X}_{HOG}$ generally leads to better performance, however, since both methods ultimately converge to the same subset of wavelet scales, they ought to be used concurrently and affirmative to each other.

When both methods agree on the feature subset, a firm belief on the frequency at which the damage phenomenon takes place is obtained. This leads to a drastic reduction in feature space and computational complexity whilst providing the operator with improved insight on the damage phenomena.

V. Conclusion

In this work a systematic procedure was presented for classifying the health of a bearing based on the machines vibrational sensor data, whilst minimizing computational complexity of the framework. The conversion to the frequency spectrum of the signal is performed by continuous wavelet transforms and is successfully applied to bearing health classification. It is known that damage phenomena take place at certain frequency spectra. The proposed methodology exploits this knowledge by zooming in on those frequencies in a twofold of ways. The first one is directly from the wavelet responses whilst a second approach involves creation of a scalogram image and localizing important regions in the image. For the proposed use case, both methods agreed and provided a narrow set of relevant scales. Hereby a reduced feature set is created, leading to reduced computational complexity and model complexity.

Future work will look at the relation between the important frequencies and operational conditions (e.g. rotational speed). However, additional data is required as most publicly available data sets do not allow for such analysis.

TABLE III
Feature selection results and validation on public data sets

Set	Features	Fold	Accuracy [%]	AUC [%]
FLAIR	$\mathbf{X}$	1	93.67 ± 1.64	96.72 ± 0.61
		2	90.75 ± 1.16	84.64 ± 2.82
		3	97.14 ± 0.09	99.39 ± 0.17
	$\mathbf{X}_{HOG}$	1	96.67 ± 0.03	98.58 ± 0.05
		2	94.35 ± 0.04	97.16 ± 0.20
		3	97.40 ± 0.05	99.47 ± 0.03
	$\mathbf{X}_{red}$	1	94.37 ± 0.06	96.78 ± 0.21
		2	94.10 ± 0.07	97.39 ± 0.11
		3	97.30 ± 0.05	98.31 ± 0.07
	$\mathbf{X}_{HOG,red}$	1	94.45± 0.02	96.27 ± 0.02
		2	95.82 ± 0.03	97.81 ± 0.02
		3	96.82 ± 0.00	98.24 ± 0.03
CWRU	$\mathbf{X}$	1	100.00 ± 0.00	100.00 ± 0.00
		2	100.00 ± 0.00	100.00 ± 0.00
		3	100.00 ± 0.00	100.00 ± 0.00
	$\mathbf{X}_{HOG}$	1	100.00 ± 0.00	100.00 ± 0.00
		2	100.00 ± 0.00	100.00 ± 0.00
		3	100.00 ± 0.00	100.00 ± 0.00
	$\mathbf{X}_{red}$	1	100.00 ± 0.00	100.00 ± 0.00
		2	100.00 ± 0.00	100.00 ± 0.00
		3	100.00 ± 0.00	100.00 ± 0.00
	$\mathbf{X}_{HOG,red}$	1	100.00± 0.00	100.00 ± 0.00
		2	100.00 ± 0.00	100.00 ± 0.00
		3	100.00 ± 0.00	100.00 ± 0.00
MFPT	$\mathbf{X}$	1	100.00 ± 0.00	100.00 ± 0.00
		2	100.00 ± 0.00	100.00 ± 0.00
		3	78.33 ± 6.66	100.00 ± 0.00
	$\mathbf{X}_{HOG}$	1	98.67 ± 2.67	100.00 ± 0.00
		2	100.00 ± 0.00	100.00 ± 0.00
		3	100.00 ± 0.00	100.00 ± 0.00
	$\mathbf{X}_{red}$	1	100.00 ± 0.00	100.00 ± 0.00
		2	100.00 ± 0.00	100.00 ± 0.00
		3	100.00 ± 3.33	98.33 ± 0.00
	$\mathbf{X}_{HOG,red}$	1	93.33± 4.22	96.30 ± 3.31
		2	88.33 ± 4.08	95.56 ± 3.87
		3	98.33 ± 3.33	100.00 ± 0.00

Acknowledgment

This research received funding from the Flemish Government (AI Research Program). The data set of this research was provided within this program by Flanders Make.

References

[1] C. Chaudhari, B. Thakare, S. Patil, and S. Gunjal, "A study of bearing and its types," *International conference on recent trends in engineering science and management*, 2015.

[2] E. A. Sikora, "Detection of bearing damage by statistic vibration analysis (diagnosis using the excess, the concept of crest factor)," *International Conference on Mechanical Engineering, Automation and Control Systems (MEACS)*, pp. 1–5, 2015.

[3] G. Wu, T. Yan, G. Yang, H. Chai, and C. Cao, "A review on rolling bearing fault signal detection methods based on different sensors," *Sensors*, vol. 22, no. 21, 2022.

[4] E. Wescoat, L. Mears, J. Goodnough, and J. Sims, "Frequency energy analysis in detecting rolling bearing faults," *Procedia Manufacturing*, vol. 48, pp. 980–991, 2020.

[5] K. Gröchenig, *Foundations of Time–Frequency Analysis*, Birkhäuser Boston, MA, 2001.

[6] J. Morlet, G. Arens, E. Fourgeau, and D. Giard, "Wave propagation and sampling theory; Part I, Complex signal and scattering in multilayered media," *Geophysics*, vol. 47, no. 2, pp. 203–221, 1982.

[7] N. Huang and S. Shen, *Hilbert-Huang Transform and Its Applications*. World Scientific, Singapore, 2014.

[8] Z.-Q. Wang, C.-H. Hu, and H.-D. Fan, "Real-time remaining useful life prediction for a nonlinear degrading system in service: Application to bearing data," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 1, pp. 211–222, 2018.

[9] G. Lu, X. Wen, G. He, X. Yi, and P. Yan, "Early fault warning and identification in condition monitoring of bearing via wavelet packet decomposition coupled with graph," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 5, pp. 3155–3164, 2022.

[10] H. Shao, M. Xia, J. Wan, and C. W. de Silva, "Modified stacked autoencoder using adaptive morlet wavelet for intelligent fault diagnosis of rotating machinery," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 1, pp. 24–33, 2022.

[11] J. Jiao, M. Zhao, J. Lin, and K. Liang, "A comprehensive review on convolutional neural network in machine fault diagnosis," *Neurocomputing*, vol. 417, pp. 36–63, 2020.

[12] S. Routray, A. K. Ray, and C. Mishra, "Analysis of various image feature extraction methods against noisy image: Sift, surf and hog," *Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, pp. 1–5, 2017.

[13] Y. Kortli, J. Maher, A. Alfalou, and M. Atri, *Advanced Secure Optical Image Processing for Communications*, 2018, ch. A comparative study of CFs, LBP, HOG, SIFT, SURF, and BRIEF for security and face recognition.

[14] N. Pudjihartono, T. Fadason, A. Kempa-Liehr, and J. O'Sullivan, "A review of feature selection methods for machine learning-based disease risk prediction," *Frontiers in Bioinformatics*, vol. 2, p. 927312, 2022.

[15] K. Pearson, "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 50, no. 302, pp. 157–175, 1900.

[16] C. R. Rao, "R. a. fisher: The founder of modern statistics," *Statistical Science*, vol. 7, no. 1, pp. 34–48, 1992.

[17] "Flanders Make, Datasets for Condition Monitoring," https://www.flandersmake.be/en/datasets.

[18] "Flanders AI Research Program, Groundbreaking AI research enabling a meaningful impact on people, industry and society," https://www.flandersairesearch.be/en.

[19] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the case western reserve university data: A benchmark study," *Mechanical Systems and Signal Processing*, vol. 64-65, pp. 100–131, 2015.

[20] E. Bechhoefer, "Condition Based Maintenance Fault Database for Testing of Diagnostic and Prognostics Algorithms," https://www.mfpt.org/fault-data-sets/.

[21] Y. Wang, Y. Bai, X. Xia, Z. Niu, Y. Yang, J. He, and X. Li, "Comparison of synchrosqueezing transform to alternative methods for time-frequency analysis of tms-evoked eeg oscillations," *Biomedical Signal Processing and Control*, vol. 70, p. 102975, 2021.

[22] K. Griffiths, B. Hicks, P. Keogh, and D. Shires, "Wavelet analysis to decompose a vibration simulation signal to improve pre-distribution testing of packaging," *Mechanical Systems and Signal Processing*, vol. 76-77, pp. 780–795, 2016.

[23] S. Mallat, *A wavelet tour of signal processing the Sparse way*, 3rd ed. Academic Press Cambridge, Massachusetts, 2008.

[24] S. Scholl, "Fourier, gabor, morlet or wigner: Comparison of time-frequency transforms," *arXiv*, 2021.

[25] M. Kunli and W. Yunxin, "Fault diagnosis of rolling element bearing based on vibration frequency analysis," *Third International Conference on Measuring Technology and Mechatronics Automation*, vol. 2, pp. 198–201, 2011.

[26] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[27] S. Mallick, "Histogram of Oriented Gradients explained using OpenCV," https://learnopencv.com/histogram-of-oriented-gradients/.

[28] G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts, "Understanding variable importances in forests of randomized trees," *Advances in Neural Information Processing Systems*, vol. 26, 2013.

[29] D. Komorowski and S. Pietraszek, "The use of continuous wavelet transform based on the fast fourier transform in the analysis of multi-channel electrogastrography recordings," *Journal of Medical Systems*, vol. 40, no. 1, p. 10, 2015.

[30] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (roc) curve." *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

[31] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *International joint conference on Artificial Intelligence (IJCAI)*, vol. 14, 2001.

[32] R. Koti and G. V. S. Kumar, "A comprehensive study and comparison of various methods on data leakages," *International Journal of Advanced Computer Research*, vol. 8, pp. 627–632, 2017.

[33] "scikit-learn-intelex 2023.0.0: Intel(R) Extension for Scikit-learn," https://pypi.org/project/scikit-learn-intelex/.