Machinery Anomaly Detection using artificial neural networks and signature feature extraction

Mansour Zoubeirou A Mayaki, Michel Riveill

Machinery Anomaly Detection using artificial neural networks and signature feature extraction

Mansour Zoubeirou A Mayaki

Université Côte d'Azur
Inria, CNRS, Nice France
mansour.zoubeirou-a-mayaki@inria.fr

Michel Riveill

Université Côte d'Azur
CNRS, Inria, Nice France
michel.riveill@unice.fr

***Abstract*—Machine learning and artificial intelligence models are increasingly common in predictive maintenance due to their ability to automatically extract high-level features with less human intervention. These models have been shown to give good results in machinery or rotatory fault diagnosis. However, due to the complexity of vibration and audio signals used in fault diagnosis, some pre-processing is required before feeding the machine learning algorithm. Fast Fourier Transform (FFT) and the Hilbert transform (HT) envelope spectrum are mostly used in the literature. However, these frequency domain transforms are not very effective when applied to rotating systems (bearings) fault detection. Indeed, in these applications, fault signal patterns are usually very weak relative to background noise and other interference in the early damage stage. In this paper we propose to use signature to extract new features from sensor data and use these new feature to train machine learning models. The main idea is to use the extracted signature coefficients to create 2d images that are then fed to a deep neural network for classification. Our experimental results show that this method outperforms most state-of-the-art methods on eight (8) bearing fault diagnosis data sets and three (3) other time series classification data sets. For example, on the Case Western Reserve University (CWRU) data set, the proposed method accuracy ranges from 96.59 % to 100% accuracy. Moreover, the results show that this method is particularly well suited for high dimensional time series. The results also show that compared to Fast Fourier Transform (FFT), the signature method requires fewer data points to detect failure. This means that in a situation where the two methods have similar performances, the signature method detects failure faster than FFT. This is very important in fault detection and predictive maintenance, in particular where it is crucial to detect faults before they occur or get worse.**

***Index Terms*—Fault diagnosis, Anomaly detection, Predictive maintenance, Concept drift detection, Data streams, Signature, Machine learning**

I. Introduction

In automatic life monitoring system, we use signal processing techniques to analyze the raw signal collected from sensors installed on the machine or system under consideration. Vibration and audio signals are mostly used in fault diagnosis. Due to the complexity of these signals in fault diagnosis, mathematical transformations and signal processing techniques are widely used to process the raw signal. Signal processing refers to an ensemble of techniques and methods that transform original signals into useful features to accomplish fault diagnosis. These features should be independent of the normal machine operating conditions (variations of load and speed) and extraneous noise and be sensitive only to machinery faults [1].

Over the years, numerous signal analysis methods have been proposed to process sensor data such as bearing fault vibration signals [2]–[9]. In general, these methods are based on frequency domain analysis. The frequency domain refers to the analysis of signals with respect to frequency, rather than time. The most commonly used methods the Fast Fourier Transform (FFT) and the Hilbert transform (HT) envelope spectrum. The Fourier transform converts a time dependent signal into a complex valued sum or integral of sine waves of different frequencies, with amplitudes and phases, each of which represents a frequency component. The "spectrum" of frequency components is the frequency-domain representation of the signal. The main limitation of frequency domain analysis techniques in rotatory systems (bearings) fault detection is that fault feature signals are usually very weak relative to background noise and other interference in the early damage stage [10]. Thus, for rolling bearings fault diagnosis, the frequency domain analysis techniques lose efficacy.

In this paper we propose to use signature to extract new features from sensor data and use these new feature to train machine learning models. Signatures method is an operation that maps multi-dimensional paths to the sequence of their iterated integrals, where the sequence is equipped with some algebraic operations [11]. The main idea is to use the extracted signature coefficients to create 2d images that are then fed to a deep neural network for classification. Our results show that this method outperforms most of state-of-the-art methods on eight (8) bearing fault diagnosis data sets (bearing and gear box data sets) and three (3) other time series classification data sets. Our results also show that compared to Fast Fourier Transform (FFT) that is the most commonly used method, signature method requires fewer data points to detect failure. This means that in a situation where the two method have similar performances, signature method detects failure faster than FFT. This is very important in fault detection and predictive maintenance in particular where it is crucial to detect the fault before they occur or get worse.

Note that the focus of this paper is not to look for the best deep learning architecture but to compare signature based feature extraction method again other techniques such as Fast Fourier Transform (FFT). Therefore, this method can

be combined with any other machine learning model. The rest of the paper is outlined as follows. The **Related Works** section discusses the main techniques and studies (or articles) related to predictive maintenance and bearing fault detection. The third section is dedicated to the theoretical definition of signature and the description of our method. In section **Experimental Data sets**, we describe the data sets used for our experiments. The **hyperparameters optimization** section describes in detail the model's architecture, the performance metrics and the choice of signature order. The results are presented and discussed in section **Results and discussions**.

II. Related Works

In this section we discuss the state of the art and some of the most common machinery fault detection techniques. Vibration-based methods are the most popular and includes a wide range of techniques which have rapidly evolved during the last decades [2]–[9]. In the case of vibration-based methods, most references are related to bearing faults, followed by rotor/stator faults and gears [1].

A. Frequency domain analysis

Wang et al. [8] used short-time Fourier transform (STFT) to extract features from the vibrations signal and then converted these new features into two-dimensional (2D) images. The 2D images are then used to train VGG (Visual Geometry Group) neural network models for bearing fault classification. They applied their method to a subset of the Case Western Reserve University (CWRU) [12] and The society For machinery failure prevention technology (MFPT) data sets [13]. Before the feature extraction step, the raw vibration signals were segmented into segment of length of 900. Zhijian et al. [14] also used the MFPT Fault data sets in their paper. The authors proposed to convert the time-domain vibration signal into RGB image based on erosion operation (EOSTI) and then apply AlexNet Convolutional neural network on the obtained images to detect faults. The raw vibration signals were segmented into segment of length of 681 before applying the EOSTI transformation. The proposed method was tested on Coal washing machine dataset and the MFPT dataset.The raw vibration signals were segmented into segment of length of 681 before applying the EOSTI transformation. Their experimental results show that their method has good performance compared to other methods such as FCNN,Lenet-5, RBF-SVM, KELM, KNN, NB, RF, PNN. On the MFPT dataset, their model has a mean accuracy of 98.64%.

Saucedo et al. [9] used fast Fourier transform (FFT) and Power Spectral Density (PSD) to analyse the vibration and current for the diagnosis of different levels of uniform wear in a gearbox. The vibration analysis is done by means of a spectral analysis and the use of a theoretical model to extract some frequency components. Thus, the acquired vibration data are analysed with the FFT and the acquired motor current data are analysed with the PSD. The method consists of analyzing the increase in amplitude and the spectral modulation in order to quantify the uniform wear level in the gearbox and the bearing defect presence. By using these two signals (vibration and current), the author expect to have a more reliable diagnosis and detection of multiple faults. Sachan et al. [6] proposed a two level de-noising algorithm based on zero frequency filter and wavelet transform for early detection of bearing faults. They then detected the bearing fault using wavelet transform and zero frequency filter. They worked on two data sets: Intelligent maintenance System (IMS) bearing dataset [15] is considered for testing the algorithm with naturally grown faults and parts of the CWRU bearing dataset is considered for testing the algorithm with seeded faults.

B. Multidimensional feature extraction

Li et al. [2] combined support vector machine (SVM) single feature evaluation, correlation analysis and principal component analysis-weighted load evaluation (PCA-WLE) to select relevant features for bearing fault detection. They used parts of the CWRU bearing fault data set. The proposed method provides very good results but the feature extraction and the feature selection process is long and can be difficult to apply in real word application specially in embedded use case. Their model has 10 classes: Normal0 and 9 faults states. Chen et al. [3] proposed to use stacked denoising autoencoder (SDA) models for health state identifications for signals containing ambient noise and working condition fluctuations. They compared the performance of SDA models with that of state of the art unsupervised methods like SVM, random forest and deep autoencoder. The authors used the 48k drive end bearing data from the CWRU bearing data center as test data. Their results show that SDA based models outperform and are more stable than other models for all noise levels. Chen et al. [5] also used parts of the CWRU bearing data sets for imbalanced class classification tasks. The authors proposed a new version of the SMOTE algorithm to deal with the class imbalance problem. They are datasets consisting of normal data and abnormal data at rpm of 1797, 1772, and 1750, respectively.

David et al. [7] proposed to use a two step state detection method based on feature extracted using envelope analysis and time domain. At the first step, they used Chebyshev's inequality upper bound to classify samples as outer fault or not based on the envelope features. If the sample has not been classified as fault, they use the time domain features and KNN to classify it into another categories (inner fault, cage fault, ball fault or healthy). Piltan et al. [16] proposed a method based on a fuzzy orthonormal-ARX adaptive fuzzy logic-structure feedback observer. They used the CWRU dataset to test their method. The average fault diagnosis accuracy for the proposed method is about 97.5% on this dataset in the presence of load variation.

C. Sound based fault diagnosis

sound-based fault diagnosis using microphones is an emerging field with a great potential in fault diagnosis since microphones are noninvasive sensors [1], [17]–[19]. Most references in audio-based fault diagnosis are related to combustion engine

faults and the faults occur in motors, pumps, fans, helicopters, in bearing, gear and rotor test-rigs. Purohit et al. [17] used auto encoders to detect industrial machine malfunction using sounds. They presented a new dataset of industrial machine sounds for malfunctioning industrial machine investigation and inspection called MIMII dataset. The authors recorded normal sounds for different types of industrial machines (i.e., valves, pumps, fans, and slide rails), and to resemble a real-life scenario, various anomalous sounds were also recorded (e.g., contamination, leakage, rotating unbalance, and rail damage). They trained the auto encoders models on the normal classes. So for each machine type and model ID, they trained an auto encoder on the normal sounds and tested it on the abnormal sounds. Liu et al [18] used a spectral-temporal fusion based self-supervised method to model the feature of the normal sound. They combined the features extracted from short-time Fourier transform (STFT) with a CNN-based network (TgramNet) to extract the temporal feature from the raw sound signal. By combining the temporal high level features from a CNN network and spectral features from the log-Mel spectrogram, the proposed method exploits complementary spectral-temporal information from the normal sound via the fused features, and results in more stable detection performance of amongst different machines [18].

This review shows that the most commonly used methods for feature extraction in fault detection are frequency domain analysis such as the Fast Fourier Transform (FFT) and the Hilbert transform (HT) envelope spectrum. The Fourier transform converts a time dependant signal into a complex valued sum or integral of sine waves of different frequencies, with amplitudes and phases, each of which represents a frequency component. The "spectrum" of frequency components is the frequency-domain representation of the signal. The main limitation of frequency domain analysis techniques in rotatory systems (bearings) fault detection is that fault feature signals are usually very weak relative to background noise and other interference in the early damage stage [10]. Thus, for rolling bearings fault diagnosis, the frequency domain analysis techniques lose efficacy. To overcome some of the limitation of frequency domain analysis, time–frequency domain approaches such as short-time Fourier transform, wavelet transform, and Hilbert Huang transform have been used by some authors, but these techniques also suffer from background noise.

III. The proposed approach

A. Signature feature extraction

In this paper we propose to use signature feature extraction to transform the raw signals before applying machine learning methods. The signature of a path was first studied by K.T Chen in his studies on path integrals [20]. Signatures method is an operation that maps multi-dimensional paths to the sequence of their iterated integrals, where the sequence is equipped with some algebraic operations [11]. The idea of using signature to describe data streams comes from some properties of differentiable functions. A function is said to be differentiable precisely when the chord is a good local summary of the function over some small subset of its domain [11]. The effect of smooth path can then be well approximated by its piecewise chordal approximation. Since then signature has been used in many machine learning tasks such as time series analysis and images processing [11], [21]–[25]. Signatures have been proven to be an effective way of describing a data stream as they provide a non-parametric way to fully describe the measure on the stream space [21]. Yang et al. [25] used signature feature mapping combined with a recurrent neural network for handwriting recognition. Their results show that signature features are very useful and effective in handwriting recognition tasks. Gyurk0 et al. [11] proposed to use signature features to classify financial time series. They used a LASSO-based regression method combined with a Kolmogorov-Smirnov distance of the distributions. Their results suggest that signature method has great potential in extracting information from noisy market data and classify them with a degree of high accuracy. Fermanian et al. [26] in his paper discussed some interesting properties of signature and its feature mapping. The authors reviewed potential embeddings of signature and their predictive performance and come to the conclusion that the lead-lag embedding outperforms other embeddings, over different datasets and algorithms. They also showed that with a good embedding, the signature combined with a simple algorithm, such as a random forest classifier, obtains results comparable to state-of-the-art approaches in different application areas, while remaining a generic approach and computationally simple [26].

1) Paths in Euclidean space: We define a path $\mathbb{X}$ in $\mathbb{R}^d$ as a continuous function mapping from some interval $[a, b] \longrightarrow \mathbb{R}^d$ written as:

$$X : [a, b] \longrightarrow \mathbb{R}^d$$
$$t \longrightarrow \mathbb{X}(t) = \mathbb{X}_t$$

The d-dimensions path can be written as:

$$X : [a, b] \longrightarrow \mathbb{R}^d, \qquad \mathbb{X}_t = \{X_t^1, X_t^2, \dots X_t^d\}$$

2) Signature of paths: Let $X : T \longrightarrow \mathbb{R}^d$ be a d-dimensional path where T is a compact interval. The signature $S(X)$ of X over the time interval T is defined by [21]:

$$X_T = (1, X^1, \dots, X^m, \dots)$$

where $m \in \mathbb{N}^+$ an integer such that,

$$X^m = \int_{\substack{u_1 < \dots < u_m \\ u_1, \dots, u_m \in T}} \dots \int_{u_1} \otimes \dots \otimes dX_{u_m} \in E^{\otimes m} \quad (1)$$

The truncated signature of X of order q is denoted by $S^q(X)$ where

$$S^q(X) = (1, X^1, \dots, X^q)$$

For a d-dimensions path $X : [a, b] \longrightarrow \mathbb{R}^d$ where each $X^i : [a, b] \longrightarrow \mathbb{R}$, $i \in \{1, \dots, d\}$ is a real-valued path. For the

i-th coordinate of the path at time $t \in [a, b]$ let us define the quantity

$$S(X)^i_{a,t} = \int_{a<s<t} dX^i_s = X^i_t - X^i_0 \tag{2}$$

Now for any pair $i, j \in \{1, \ldots, d\}$, let us define the double-iterated integral

$$S(X)^{i,j}_{a,t} = \int_{a<s<t} S(X)^i_{a,s} dX^j_s = \int_{a<r<s<t} dX^i_r dX^j_s \tag{3}$$

Where $S(X)^i_{a,s}$ is given by (2). Notice that $S(X)^i_{a,\cdot} : [a, b] \longrightarrow \mathbb{R}$ is itself a real valued path.
Likewise for any triple $i, j, k \in \{1, \ldots, d\}$ we define the triple-iterated integral

$$S(X)^{i,j,k}_{a,t} = \int_{a<s<t} S(X)^{i,j}_{a,s} dX^k_s = \int_{a<q<r<s<t} dX^i_q dX^j_r dX^k_s \tag{4}$$

We can continue recursively, and for any integer $k \geq 1$ and collection of indexes $i_1, \ldots, i_k \in \{1, \ldots, d\}$, we define

$$S(X)^{i_1,\ldots,i_k}_{a,t} = \int_{a<s<t} S(X)^{i_1,\ldots,i_{k-1}}_{a,s} dX^{i_k}_s \tag{5}$$

As $S(X)^{i_1,\ldots,i_{k-1}}_{a,s}$ and $X^{i_k}_s$ are real-valued paths, $S(X)^{i_1,\ldots,i_k}_{a,t} : [a, b] \longrightarrow \mathbb{R}$ is also a real-valued path [22] . The equation (5) can be rewritten as follows

$$S(X)^{i_1,\ldots,i_k}_{a,t} = \int_{a<t_k<t} \ldots \int_{a<t_1<t_2} dX^{i_1}_{t_1} \ldots dX^{i_k}_{t_k} \tag{6}$$

The real number $S(X)^{i_1,\ldots,i_k}_{a,b}$ is called the k-fold iterated integral of X along the indexes $i_1, \ldots, i_k$.

The truncated signature of order q of a d-dimensions path $X : [a, b] \longrightarrow \mathbb{R}^d$ denoted by $\mathbb{S}(X)_{a,b}$ is the sequence of real numbers:

$$\mathbb{S}(X) = \left(1, \mathbb{S}(X)^1_{a,b}, \ldots, \mathbb{S}(X)^q_{a,b}\right) \tag{7}$$

where the first term is by convention equal to 1, and the superscripts run along the set of all multi-indexes

$$\mathbb{W} = \{(i_1, \ldots, i_k) | k \geq 1, i_1, \ldots, i_k \in \{1, \ldots, d\}\}$$

Signature has some interesting mathematical properties such as the **Invariance under time reparametrisations** It can be proven that for a given paths, their path integrals are invariant under a time reparametrization of both paths. As the signature of a d-dimensions path $\mathbb{S}(X)$ is a collection of iterated path integral of X, it follows from the above that the signature $S(X)_{a,b}$ is also invariant under time reparametrizations of X. **Shuffle product** Ree et al. [27] proved that the product of two terms $S(X)^{i_1,\ldots,i_k}_{a,t}$ and $S(X)^{j_1,\ldots,j_m}_{a,t}$ can always be expressed as a sum of another collection of terms of $S(X)_{a,b}$which only depends on the multi-indexes $(i_1, \ldots, i_k)$ and $(j_1, \ldots, j_m)$.

For more interesting properties of signatures, the reader can refer to [22], [26].

B. The proposed approach

As discussed in the preview subsection, signature transformation can be a powerful tool for machine learning problems due to its ability to extract characteristic features from data. The advantages of feature extraction with the signature method is that the signature is sensitive to the geometric shape of a path [22]. In this paper we propose to use signature to extract new features from sequential data and use these new feature to learn deep learning models. Like Yan et. [25] and Graham [28], we extracted signature features from the raw data and create images of these signature coefficients. These images are then fed to a deep neural network for classification. Our workflow is simple and is summarised in the following steps (see figure 1): (1) collect raw data from sensors and split the raw sequential data into sub times series according to a given window; (2)Transform each time series into continuous path using lead lag and compute the signature of order d of all the sub series. In this way, the time-dependent input is mapped into a time-independent finite set of coefficients. (3) reshape the signature coefficients into 2-dimension array or 2-d images, so they can be used as input for deep convolutional networks; (4) Use the obtained 2-d images as input for deep learning models.

IV. Experimental data sets

The proposed method was tested on 11 publicly available datasets and the results have been compared to other works that use these datasets II. These data sets include 8 bearing fault data sets and 3 multi-dimensional time series classification data sets. Figure 2 shows the types of bearing faults inside a rotatory system.

A. Bearing fault data sets

First, the **Bearing Data Center Seeded Fault Test Datasets** are provided by Case Western Reserve University (CWRU) of Cleveland in the USA [12]. The data sets are collected from Motor bearings with seeded faults using electro-discharge machining (EDM). The experiments were conducted using a 2 hp Reliance Electric motor, and acceleration data was measured at locations near to and remote from the motor bearings. Vibration data was collected using accelerometers, which were attached to the housing with magnetic bases. The seeded faults were introduced ranging from 0.007 inches in diameter to 0.028 inches in diameter separately at the inner raceway, rolling element (i.e. ball) and outer raceway. The damaged bearings were then reinstalled into the test motor and vibration data was recorded for motor loads of 0 to 3 horsepower (motor speeds of 1797 to 1720 RPM). The data sets contain two types of fault : drive end bearing fault and fan end bearing fault. In addition to these faulty data, the normal condition of the motor has been recorded. In this paper we used the drive end bearing at the sampling rate of 12,000 samples/second for fault diameters ranging from 0.007 inches to 0.021 inches separately at the inner raceway, rolling element (i.e. ball) and outer raceway. The data sets have been separated into 4 different datasets according to the fault

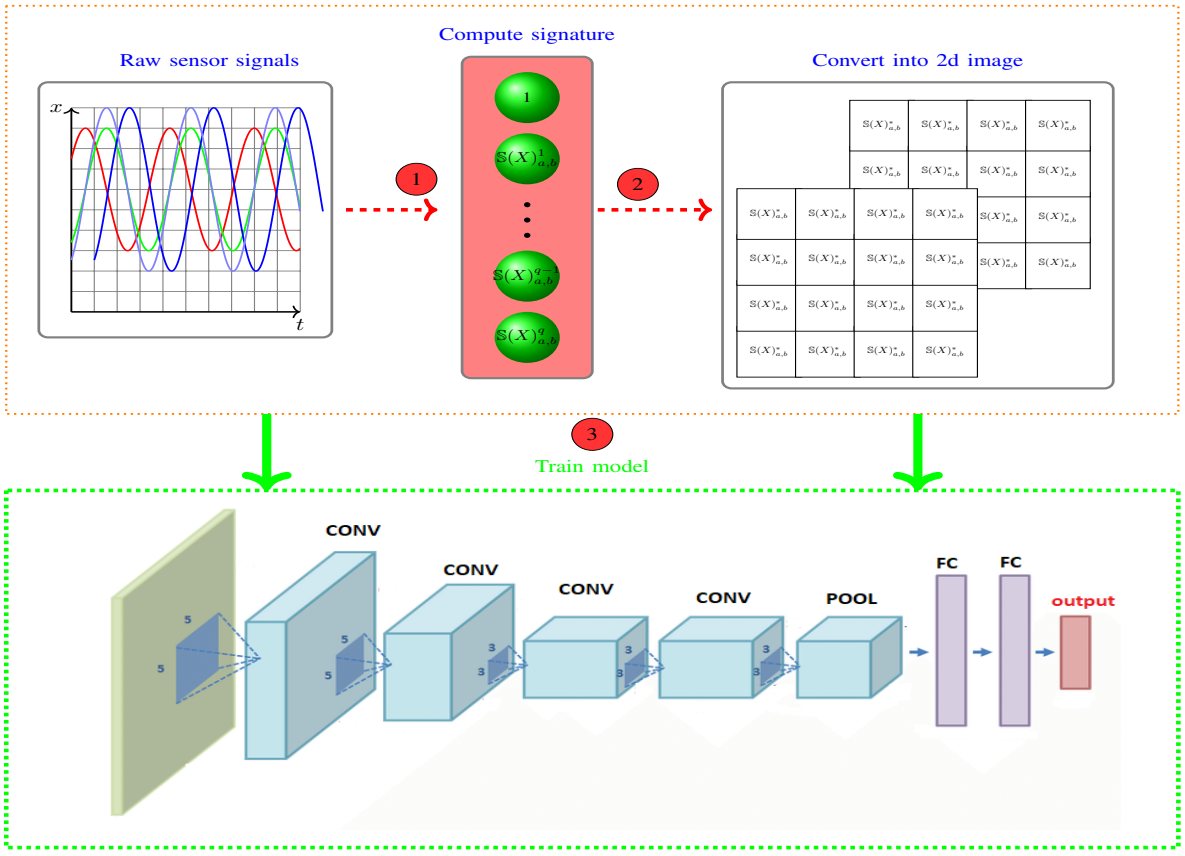


Fig. 1: Architecture and dataflow of the proposed approach. ① Compute signatures. ② Reshape signatures into 2d arrays. ③ Train a predictive model with 2 d data sets.

diameter and motor speed. In the first three dasets, we train the models to detect the bearing condition ((healthy, inner, ball, outer)) according to the fault diameter as shown in Table I. In the last dataset (dataset 4), the models are constructed to detect for a given motor speed the type of bearing fault (inner raceway, rolling element and outer raceway) and the diameter of damage. For this set up we have teen states or classes to predict:'normal','IF 0.007','BF 0.007','OF 0.007','IF 0.014','BF 0.014','OF 0.014','IF 0.021','BO 0.021','OF 0.021'.

The second dataset is the **Gearbox Fault Diagnosis** data [29]. This publicly available data set include the vibration data recorded by using SpectraQuest's Gearbox fault diagnostics simulator. Data set has been recorded using 4 vibration sensors placed in four different direction, and under variation of load from 0 to 90 percent. Two different scenario are included: Healthy condition and Broken Tooth Condition. There are 20 files in total, 10 for healthy gearbox and 10 from broken one. Each file corresponds to a given load from 0% to 90% in steps of 10% [29]. The merged data set contains 2.021.119 recordings and four features (a1, a2, a3, a4) corresponding to the value of the four vibration sensors. The broken condition represents 49.74% of the whole dataset, the dataset is then balanced.

The third data set called **MFPT Fault Data Sets** have been collected by the Society For Machinery Failure Prevention Technology (MFPT) [13]. The data set comprises data from a bearing test rig (nominal bearing data, an outer race fault at various loads, and inner race fault and various loads), and three real-world faults. The data set comprises 3 baseline conditions, 3 outer race fault conditions, 7 outer race fault conditions, 7 inner race fault conditions. In this paper we use normal signals (NO) with a load of 270 lbs, outer ring faults (OF 50) with a load of 50 lbs, outer ring faults (OF 100) with a load of 100 lbs, outer ring faults (OF 150) with a load of 150 lbs, outer ring faults (OF 200) with a load of 200 lbs, outer ring faults (OF 250) with a load of 250 lbs, inner ring faults (IF 50) with a load of 50 lbs, inner ring faults (IF 100) with a load of 100 lbs, inner ring faults (IF 150) with a load of 150 lbs, inner ring faults (IF 200) with a load of 200 lbs and inner ring faults (IF 250) with a load of 250 lbs as in [8], [14]. The input shaft rate is 25 Hz and sample rate of 97,656 sps for 6 seconds. Thus, there is about 3906 data points collected at each rotation of the shaft. The final training dataset contains 1.171.872 recordings divided into 13 classes. In training dataset, the labels are divided into five categories: NO, OF 50, OF 100, OF 150, OF 200, OF 250, IF 50, IF 100,

IF 150, IF 200 and IF 250.

Finally the **Triaxial Bearing Vibration Dataset** has been provied by Kumar et al. [30] for bearing fault detection usage. The dataset is collected using a MEMS based triaxial accelerometer and the National Instruments myRIO board. The data set includes triaxial vibration data of bearing of induction motor operated under different load conditions along the axes x, y, and z. The faulty conditions of bearings include inner race and outer race faults with differents diameters: (i) 0.7mm, (i) 0.9mm, (i) 1.1mm, (i) 1.3mm, (i) 1.5m, and (i) 1.7mm. The bearings with these fault severity levels were operated under three load conditions:100W,200W, 300W. Each set contains four columns: Time Stamp, X-axis, Y-axis, and Z-axis. The free axis columns contain vibration data of the motor along the x, y, and z axes. In this paper, we used only the files from the healthy bearing condition and three faults diameters: 0.7mm, 0.9mm, 1.1mm. The merged dataset contains 2.499.951 recordings. This dataset has been divide into 7 classes: Healthy, 0.7mm inner race fault, 0.9mm inner race fault, 1.1mm inner race fault, 0.7mm outer race fault, 0.9mm outer race fault, 1.1mm outer race fault.

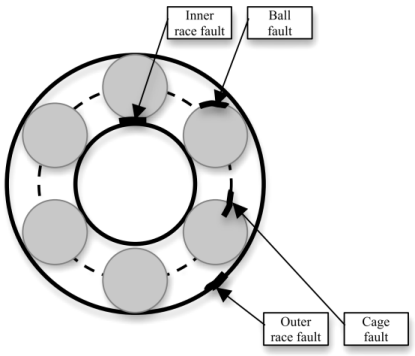


Fig. 2: Types of bearing faults [4]

B. Time series classification data sets

The **NSW** data set contains data from the Australian New South Wales electricity market [31]. In this market, prices are flexible and are affected by demand and supply of the market. The data set contains 45.312 instances and nine features dated from 7 May 1996 to 5 December 1998. The goal is to predict if the electricity price goes up or down each 30 minutes [32]. Concept drift is present due to seasonal weather changes which affects the electricity demand and its price. The **Insect stream data** was built by Souza et al. [33] for drift detection and classification task. The data set cointains 52848 samples and 34 features. The learning task is to correctly classify the insect by species. There are 6 type (class) of insects. The **Sensor Stream** data set contains information (temperature, humidity, light, and sensor voltage) collected from 54 sensors [34]. The data contains consecutive samples recorded over a 2 months period. The data sets contains 2.219.803 samples and 5 features. The learning task is to correctly identify the sensor ID (1 out of 54 sensors) based on the sensor data and the corresponding recording time. In this work we only used a subset of the whole data set. We selected the data from 5 sensors: b'1',b'2',b'3',b'4',b'6'. So in our case we only have 180.388 samples divided into 5 classes. The **earthquake** data set is taken from Northern California Earthquake Data Center [35]. The learning task is predicting whether a major event is about to occur based on the most recent readings in the surrounding area. The data set contains 8666 samples and each sample is an averaged reading for one hour. The original time series signal was split into 461 sub samples of length 512 for classification task.

TABLE I: Bearing Fault Data for varying fault diameter, motor load and motor speed

Dataset	type of fault	fault diameter	motor load	motor speed
Dataset 1	healthy	0.007	0,1,2,3	1797, 1772, 1750, 1730
	inner	0.007		
	ball	0.007		
	outer	0.007		
Dataset 2	healthy	0.014	0,1,2,3	1797, 1772, 1750, 1730
	inner	0.014		
	ball	0.014		
	outer	0.014		
Dataset 3	healthy	0.021	0,1,2,3	1797, 1772, 1750, 1730
	inner	0.021		
	ball	0.021		
	outer	0.021		
Dataset 4	healthy	0.007,0.014,0.021	0	1797
	inner		0	1797
	ball		0	1797
	outer		0	1797

TABLE II: Experimental data description.

Data set	**Samples**	**Features**	**number of classes**
Triaxial Bearing [30]	2.499.951	3	7
MFPT Data Sets [13]	2.343.744	1	13
Gear box data [29]	2.021.119	4	2
Sensor Stream [34]	2.219.803	5	54
Data set 1	1.707.738	2	4
Data set 2	1.707.105	2	4
Data set 3	1.708.120	2	4
Data set 4	1.341.856	2	10
NSW electricity [31]	45.312	9	2
Insect stream data [33]	52.848	34	6
Earthquakes [35]	236.032	1	2

C. Data segmentation

For each data set, we trained and tested the model by varying the window size and choose the value that maximizes the accuracy score. For the bearing data sets, the window values are chosen according to the motor speed and the sampling rate. For example for the CWRU data set, we set the motor speed to 1797 rpm, this means that there is approximately 30 rotation/second. As the sensor sampling frequency is set to 12000 samples/second, there is about 400 data points are collected in one rotation of the bearing. Therefore, in order to ensure that the length of a single sample can completely and accurately reflect that data distribution of the bearing vibration signals in this state, each 400 data points can be regarded as a small sample.

Notice that the lower the segmentation window the faster the model can detect faults. This is very important in predictive maintenance where it is crucial to detect the fault before they occur or get worse.

V. Experimental Design

A. CNN hyper-parameters

The models are trained using a k-folds cross-validation ($k = 5$ in our case). Each model is therefore trained and

validated on 5 different sub-samples. The final performance is computed on the test set. The final value of the performance metric(accuracy) is computed by taking the average of the five measurements obtained during the ten iterations of the cross-validation. We set the number of epochs to 200 for each model and used **early stopping** to avoid overfitting. The rectified linear unit (ReLU) activation function is used in neurons of the hidden layers. We also added between the hidden layers a Batch Normalization layer followed by a dropout layer. Our final CNN model has 16.586 Trainable parameters and the following architecture: 3 convolution layers with 16 filters each, an average pooling layer and a fully connected layer of 32 neurons.

Note the focus of this paper is not to look for the best deep learning architecture but to compare signature based feature extraction method again other techniques such as Fast Fourier Transform (FFT). The reader can combine our method with any other machine learning model if needed.

B. Choice of signature order

The order of the signature is a hyperparameter that need to be set by the user. For given path in $\mathbb{R}^d$, there are ^{p} coefficients of order p. Therefore the truncated signature at order q is therefore a vector of dimension:

$$W = \sum_{i=0}^{q} {}^{i} = \frac{{}^{q+1} - 1}{d - 1} \tag{8}$$

When $d > 1$, W the size of $\mathbb{S}(X)$ increases exponentially with q. For univariate sigal, $W = q + 1$. In this paper we want to convert the signature vector into a 2d images and use the images as input for our deep learning models. To chose the truncation order of the signature, we set the size of our 2d images to $48 \times 48 = 2\ 304$. So as $W = 2304$, we can compute for each data set, the value of the signature order. For a given value of W, using equation (8), the order of the signature is given by:

$$q^* = \frac{\log(W \times d + 1)}{\log(d + 1)} - 1 \tag{9}$$

VI. Results and Discussions

In order to evaluate the effectiveness of signature feature extraction method, we compared its performance to those of some state-of-the art approaches. We present the results of our method and compare them to those of the literature papers on the experimental data sets. We also compare this method to Fourier transform as it is the most commonly used method for predictive maintenance. In addition, we did some experiments to see in which case each method works well or not and their preprocessing time.

A. Performance results

In this subsection, we discuss the results of signature feature extraction and some state-of-the-art feature extraction techniques. The goal is not to find the machine learning model but to compare the performance of signature feature extraction to others feature extraction method used in the literature. The results from Table III and IV show that signature feature extraction method has better accuracy than other methods such as short-time Fourier transform (STFT), principal component analysis and fuzzy othonormal on most of the experimental data sets. On the CWRU bearing data sets (see Table III), the proposed signature feature extraction method combined with convolutional neural networks (CNN) outperforms all other methods. Signature reaches 100% of accuracy on three of four bearing data sets. Figure 3b shows the confusion matrix of the CNN model on the test data of the bearing data set. We can see that almost all teen classes have been correctly classified. FFT (98.33%) outperforms signature on the MFPT data set (95.89%). Note that we can not directly compare our results on the MFPT data set those of Wang et al. [8] and Zhijian et al. [14]because these authors used only a subset of 5 classes of the data set while we used 13 classes. The results from Table IV show that signature feature extraction method has better accuracy than FFT on all the other types of time series data sets. This table shows the ranges of the accuracy for varying window size. For the gear box data set, both FFT et signature reach 100% of accuracy. However, with signature we only need 100 data points as window size whereas for FFT the window size is set to 200 data points. On the triaxial data set, signature performance varies from 85.69 % to 99.37% when FFT performance lies between 84.93% and 97.77%.

TABLE III: Highest accuracy on bearing data sets

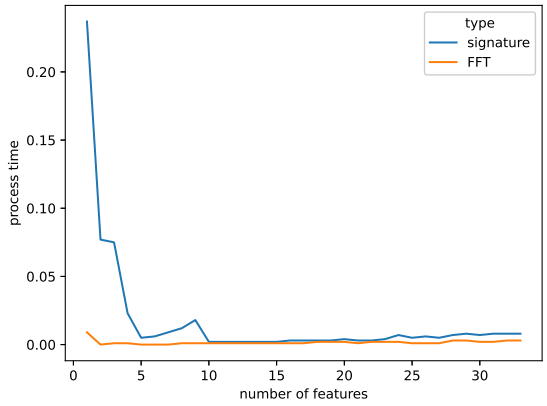
Data sets	Data set 1	Data set 2	Data set 3	Data set 4	MFPT
Signature + CNN (ours)	**100%**	**99.59** %	**100%**	**100 %**	95.89 %
STFT + VGG [8]	-	-	-	99.85%	**99.8%**
PCA-WLE+PSO-SVM [2]	-	-	-	99.61%	-
STFT+ DNN [36]	99.82%	-	-	-	-
SDA [3]	93.54%	-	-	-	-
EOSTI + AlexNet [14]	93.54%	-	-	-	98.46%
Fuzzy orthonormal-ARX+SVM [16]	95.8%	97.5 %	97.8 %	-	-

TABLE IV: Accuracy ranges on other data sets

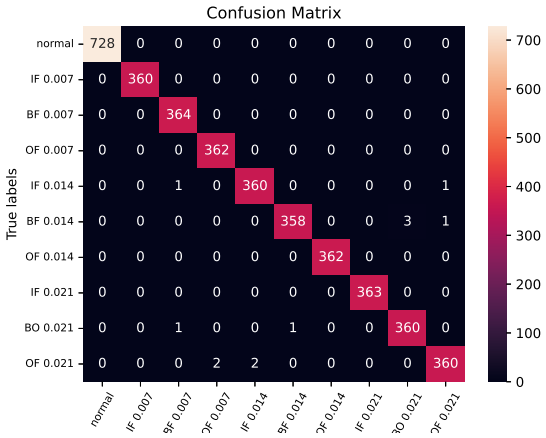
Data sets	**Metric**	Gear box data	Triaxial Bearing Vibration	Sensor Stream	NSW electricity	Insect	Earthquakes
Signature + CNN	*Max acc*	**100 %**	**99.37%**	**89.07 %**	**86.46** %	**75.23%**	**97.32**
	Min acc	99.27%	85.69 %	79.67%	83.45%	75.05%	39.78%
FFT + CNN	Max acc	**100%**	98.77%	80.93 %	71.43 %	32.43 %	74.64%
	Min acc	99.99%	84.93 %	77.27%	63.65%	31.82%	74.63%

B. Preprocessing time comparison

Figure 3a shows the execution times of fast Fourier transform (FFT) and signature method according to the number of features. To compare the execution times of the two methods, we set the size of the signature vector to $48 \times 48 = 2\ 304$ as we want to convert it into 2d image of size 48×48. As shown in equation (9) the size of the signature vector depend on the number of features and the order of the signature. For a given signature vector size, the order of the signature is an exponentially decreasing function of the number of features. Figure 3a shows that when the number of features is less than 5, FFT is way faster than signature processing for any window size. However, when the number of features is greater than 5, the two methods, have almost the same processing times. This analysis shows that for high dimensional time series, signature is very effective tool for features extraction for machine learning tasks.



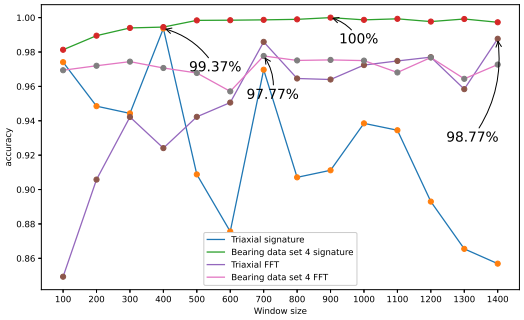
(a) Preprocessing time according to the number of features



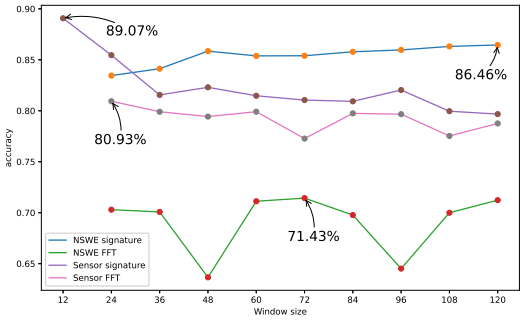
(b) Confusion matrix of on test data for bearing data set 4

C. Discussions

Our results show that signature based feature extraction is very powerful tool for machinery fault detection using machine learning. This methods outperforms state-of-the-art methods such as fast Fourier transform (FFT) or PCA. The results from figure 4 show that signature method require fewer data points to detect failure than FFT.We make the hypothesis that the lower the segmentation window the faster the model can detect faults. This is very important in fault detection where it is crucial to detect the fault before they occur or get worse. Figure 4a and 4b show the performance of signature and FFT for varying window sizes. For each window size, we did a 5-fold cross- validation and took the average accuracy of the model. The signature method has the highest accuracy for the bearing data set (100%) and the triaxial data set (99.37%). Figure 4a shows that for the bearing data set, signature outperforms FFT for any window size and reaches 100% of accuracy after two rotation of the bearing. Thus, on the bearing fault data set, signature accuracy ranges from 96.59 % to 100%. On the Triaxial data set, signature reaches the highest accuracy (99.37%) after only one rotation (400 points) whereas FFT reaches its highest accuracy (98.77%) after three rotation (1400 points). From figure 4b, we can see that signature outperforms FFT for any window size and detects the anomalies faster than FFT. Figure 3a shows that signature feature extraction takes more time than FFT when dealing with low dimensional data sets. Therefore, for high dimensional data sets, signature can be very effective method for machine learning feature extraction. We also did some experiment to see each method performance for varying sampling rate. The goal is find out which methods works better according to the data sampling rate. Figures 5a and 5b show the performances when we divide the sampling rate of the corresponding data sets by 2,4,6 ...16. These results show that when we reduce the sampling up to a factor of 10, FFT method outperforms signature method. This results suggest that if the data sampling rate is low, the FFT feature extraction may achieve better performance than signature.



(a) Varying window for bearing data sets

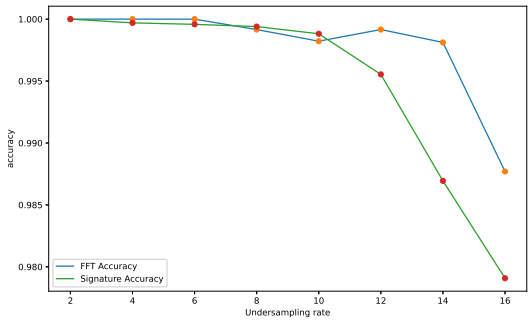


(b) Varying window for sensor and nswe data

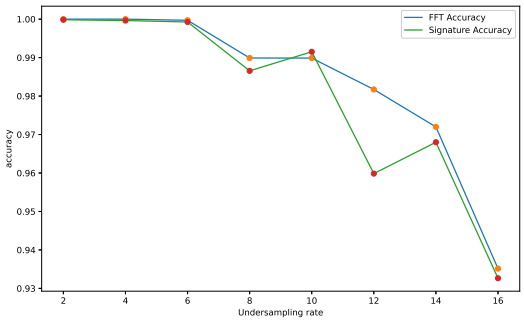
Fig. 4: FFT and signature performance for varying window size.

CONCLUSION

Machine learning and artificial intelligence methods are increasingly used in predictive maintenance due to their ability to extract automatically high-level features with less human intervention. Machine learning models have shown to give good results in machinery or rotatory fault diagnosis. However, due to the complexity of vibration and audio signals



(a) Varying sampling rate for bearing data set



(b) Varying sampling rate for gear box data set

primarily used in fault diagnosis, some pre-processing is required before feeding the machine learning algorithm. In the literature, most people use frequency domain transform such as the Fast Fourier Transform (FFT) and the Hilbert transform (HT) envelope spectrum. The main limitation of frequency domain analysis in rotatory systems (bearings) fault detection is that fault feature signals are usually very weak relative to background noise and other interference in the early damage stage. In this paper we propose to use signature to extract new features from sensor data and use these new feature to train machine learning models. The main idea is to use the extracted signature coefficients to create 2d images that are then fed to a deep neural network for classification. Our experimental results show that this method outperforms most of state-of-the-art methods on eight (8) bearing fault diagnosis data sets (bearing and gear box data sets) and three (3) other time series classification data sets. Our results also show that compared to Fast Fourier Transform (FFT), signature method requires fewer data points to detect failure. This means that in a situation where the methods have similar performances, signature method detects the failure faster than FFT. This is very important in fault detection and predictive maintenance in particular where it is crucial to detect the fault before they occur or get worse.

REFERENCES

[1] P. Henriquez, J. B. Alonso, M. A. Ferrer, and C. M. Travieso, "Review of automatic fault diagnosis systems using audio and vibration signals," *IEEE Transactions on systems, man, and cybernetics: Systems*, vol. 44, no. 5, pp. 642–652, 2013.

[2] Y. Li, W. Dai, and W. Zhang, "Bearing fault feature selection method based on weighted multidimensional feature fusion," *IEEE Access*, vol. 8, pp. 19 008–19 025, 2020.

[3] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Processing*, vol. 130, pp. 377–388, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168416301797

[4] M. Segla, S. Wang, and F. Wang, "Bearing fault diagnosis with an improved high frequency resonance technique," in *IEEE 10th International Conference on Industrial Informatics*, 2012, pp. 580–585.

[5] L. Chen, P. Dong, W. Su, and Y. Zhang, "Improving classification of imbalanced datasets based on km++ smote algorithm," in *2019 2nd International Conference on Safety Produce Informatization (IICSPI)*, 2019, pp. 300–306.

[6] S. Sachan, S. Shukla, and S. K. Singh, "Two level de-noising algorithm for early detection of bearing fault using wavelet transform and zero frequency filter," *Tribology International*, vol. 143, p. 106088, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301679X19306024

[7] D. He, R. Li, and J. Zhu, "Plastic bearing fault diagnosis based on a two-step data mining approach," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 8, pp. 3429–3440, 2013.

[8] B. Wang, G. Feng, D. Huo, and Y. Kang, "A bearing fault diagnosis method based on spectrum map information fusion and convolutional neural network," *Processes*, vol. 10, no. 7, p. 1426, 2022.

[9] J. J. Saucedo-Dorantes, M. Delgado-Prieto, J. A. Ortega-Redondo, R. A. Osornio-Rios, and R. d. J. Romero-Troncoso, "Multiple-fault detection methodology based on vibration and current analysis applied to bearings in induction motors and gearboxes on the kinematic chain," *Shock and Vibration*, vol. 2016, 2016.

[10] B. Pang, G. Tang, T. Tian, and C. Zhou, "Rolling bearing fault diagnosis based on an improved htt transform," *Sensors*, vol. 18, no. 4, p. 1203, 2018.

[11] L. G. Gyurkó, T. Lyons, M. Kontkowski, and J. Field, "Extracting information from the signature of a financial data stream," *arXiv preprint arXiv:1307.7244*, 2013.

[12] Bearing data center seeded fault test data. Case School of Engineering. [Online]. Available: https://engineering.case.edu/bearingdatacenter

[13] Condition based maintenance fault database for testing of diagnostic and prognostics algorithms. MFPT. [Online]. Available: https://www.mfpt.org/fault-data-sets/

[14] Z. Wang, W. Zhao, W. Du, N. Li, and J. Wang, "Data-driven fault diagnosis method based on the conversion of erosion operation signals into images and convolutional neural network," *Process Safety and Environmental Protection*, vol. 149, pp. 591–601, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957582021001300

[15] J. Lee, H. Qiu, G. Yu, and J. Lin, "Bearing data set, nasa ames prognostics data repository," *Rexnord Technical Services, IMS, University of Cincinnati*, 2007.

[16] F. Piltan and J.-M. Kim, "Hybrid fault diagnosis of bearings: Adaptive fuzzy orthonormal-arx robust feedback observer," *Applied Sciences*, vol. 10, no. 10, p. 3587, 2020.

[17] H. Purohit, R. Tanabe, K. Ichige, T. Endo, Y. Nikaido, K. Suefusa, and Y. Kawaguchi, "Mimii dataset: Sound dataset for malfunctioning industrial machine investigation and inspection," 2019. [Online]. Available: https://arxiv.org/abs/1909.09347

[18] Y. Liu, J. Guan, Q. Zhu, and W. Wang, "Anomalous sound detection using spectral-temporal information fusion," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 816–820.

[19] Y. Koizumi, Y. Kawaguchi, K. Imoto, T. Nakamura, Y. Nikaido, R. Tanabe, H. Purohit, K. Suefusa, T. Endo, M. Yasuda *et al.*, "Description and discussion on dcase2020 challenge task2: Unsupervised anomalous sound detection for machine condition monitoring," *arXiv preprint arXiv:2006.05822*, 2020.

[20] K.-T. Chen, "Integration of paths, geometric invariants and a generalized baker-hausdorff formula," *Annals of Mathematics*, pp. 163–178, 1957.

[21] D. Levin, T. Lyons, and H. Ni, "Learning from the past, predicting the statistics for the future, learning an evolving system," *arXiv preprint arXiv:1309.0260*, 2013.
[22] I. Chevyrev and A. Kormilitzin, "A primer on the signature method in machine learning," *arXiv preprint arXiv:1603.03788*, 2016.
[23] A statistical point of view on signatures. Sorbonne Université. [Online]. Available: https://gdrtrag.math.cnrs.fr/cirm2021/slides/fermanian.pdf
[24] I. Perez. Using a signature-based machine learning model to analyse a psychiatric stream of data.
[25] W. Yang, L. Jin, and M. Liu, "Deepwriterid: An end-to-end online text-independent writer identification system," *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 45–53, 2016.
[26] A. Fermanian, "Learning time-dependent data with the signature transform," Ph.D. dissertation, Sorbonne Université, 2021.
[27] R. Ree, "Lie elements and an algebra associated with shuffles," *Annals of Mathematics*, pp. 210–220, 1958.
[28] B. Graham, "Sparse arrays of signatures for online character recognition," *arXiv preprint arXiv:1308.0371*, 2013.
[29] Y. Pandya, "Gearbox fault diagnosis data," 06 2018. [Online]. Available: https://data.openei.org/submissions/623
[30] D. Kumar, S. Mehran, M. Z. Shaikh, M. Hussain, B. S. Chowdhry, and T. Hussain, "Triaxial bearing vibration dataset of induction motor under varying load conditions," *Data in Brief*, p. 108315, 2022.
[31] J. Gama, P. Medas, G. Castillo, and P. Rodrigues, "Learning with drift detection," in *Brazilian symposium on artificial intelligence*. Springer, 2004, pp. 286–295.
[32] M. Zoubeirou a Mayaki and M. Riveill, "Autoregressive based Drift Detection Method," in *IEEE WCCI 2022 - IEEE world congress on computational intelligenceWORLD CONGRESS ON COMPUTATIONAL INTELLIGENCE*, Padoue, Italy, Jul. 2022. [Online]. Available: https://hal.archives-ouvertes.fr/hal-03740180
[33] V. Souza, D. M. dos Reis, A. G. Maletzke, and G. E. Batista, "Challenges in benchmarking stream learning algorithms with real-world data," *Data Mining and Knowledge Discovery*, vol. 34, no. 6, pp. 1805–1858, 2020.
[34] X. Zhu. Stream data mining repository.
[35] A. Bagnall, J. Lines, J. Hills, and A. Bostrom, "Time-series classification with cote: The collective of transformation-based ensembles."
[36] Y. Yang, P. Fu, and Y. He, "Bearing fault automatic classification based on deep learning," *IEEE Access*, vol. 6, pp. 71 540–71 554, 2018.