Deep Learning Algorithm for Supervision Process in Production Using Acoustic Signal

Mahmood Safaei *a*,∗, Seyed Ahmad Soleymani *b*, Mitra Safaei *c*, Hassan Chizari *d* and Mehrbaksh Nilashi *e*

a Department of Computer Science, College of Engineering and Polymer Science, The University of Akron, Akron, 44325-3909, OH, USA
b 5GIC & 6GIC, Institute for Communication Systems (ICS), University of Surrey, Guildford, GU2 7XH, Surrey, UK
c Gottfried Wilhelm Leibniz Universität Hannover, Fakultät Electronic und Informatik, Hanover, 30060, Germany
d School of Computing & Engineering, University of Gloucestershire, The Park, Cheltenham, GL50 2RH, Gloucester, UK
e Centre for Global Sustainability Studies (CGSS), Universiti Sains Malaysia, Penang, 11800, Malaysia

∗Corresponding author
msafaei@uakron.edu (M. Safaei);
ORCID(s): 0000-0002-3924-6927 (M. Safaei)

Abstract

In an industrial environment, accurate fault diagnosis of machines is crucial to prevent shutdowns, failures, maintenance costs, and production downtime. Existing methods for system failure prevention are often unsatisfactory and expensive, prompting the need for alternative approaches. Acoustic signals have emerged as a new method for predicting machine component lifespan, but recognizing relevant features and distinguishing them from noise remains challenging. To address the aforementioned challenges, we present a comprehensive model that integrates various components to enhance the accuracy and effectiveness of machine process identification. The proposed model incorporates a deep learning algorithm, which enables the forecasting of machine operation based on acoustic signals. In addition, we employ a customized Continuous Wavelet Transformation (CWT) technique to convert the acoustic signals into CWT images, preserving vital information such as signal amplitude. This transformation allows for a more comprehensive analysis and representation of the acoustic data. Furthermore, a Convolutional Neural Network (CNN) is utilized as a powerful classifier to accurately classify and differentiate between different machine processes based on the extracted features from the CWT images. By combining these elements, our model provides a robust and efficient framework for machine process identification using acoustic signals. Testing our model on a dataset generated from the Institute for Manufacturing Technology and Machine Tools (IFW) for the Gildemeister machine (CTX420 linear), we achieve over 97% accuracy in discovering and early detecting emerging faults and machine processes based on acoustic signals.

Keywords
Deep learning
Acoustic
Production
Fault diagnosis

1. Introduction

In the realm of manufacturing, the maintenance of machines is essential to prevent breakdowns and extend their lifespan, often at considerable expense. Traditional approaches involved repairing machines only after they malfunctioned, resulting in costly production interruptions and potential collateral damage to other components [1]. Nowadays, a popular strategy involves establishing a life cycle for each machine part, replacing them when they reach the end of their predefined life. However, this approach may lead to the replacement of functioning parts, thereby increasing maintenance costs. To enable predictive maintenance, it is crucial to accurately determine whether a machine is operating properly. For instance, certain machines exhibit heightened vibrations when experiencing failures, while others, such as drilling machines, undergo abrupt acoustic or sound changes [2]. Hence, finding a solution to assess the working condition of target machines is of paramount importance in implementing an effective predictive maintenance process.

Nowadays, acoustic signals have emerged as a new method for predicting machine component lifespan. In [3], sparse Bayesian learning beamforming is used for the acoustic data captured by a ground microphone array for signal augmentation to detect wind turbine blade flaws. Then, the Short-Time Fourier Transform is carried out over the enhanced signals. For predicting tool-wear conditions, a convolutional bi-directional extended short-term memory network is proposed in [4]. To achieve this goal, Convolutional Neural Networks (CNN) layers are used to extract local features from the gathered raw signals, which are then inputted into long short-term memory neural networks to generate the necessary indicators. In [5], a combination of Continuous Wavelet Transform (CWT) and image conversion technology is employed to obtain vibration amplitude spectrum imaging features. Furthermore, a Convolutional Deep Belief Network (CDBN) using Gaussian distribution, referred to as CRWGCDN, is established to identify the distinctive features for the classification of bearing faults.

Moreover, fault diagnosis of machines is a key area of research in industrial maintenance and reliability engineering. Therefore, various studies are conducted in this field. In reference [6], a new technique for mechanical fault diagnosis is introduced using a combination of CNN and Extreme Learning Machines (ELMs). This method is divided into two stages: feature extraction utilizing CNN and fault classification using an ELM. CNN is trained to obtain features from the original vibration signals, which are then forwarded to an ELM for classification. The proposed approach is validated on a standard dataset of four distinct types of faults in a rolling bearing, demonstrating that it outperforms other current methods and showcasing its potential in practical mechanical fault diagnosis applications.

In [7], authors discussed the importance of fault diagnosis in the context of rotating machinery and how it can improve machine reliability and reduce maintenance costs. They highlighted the potential of deep learning-based methods for fault diagnosis in rotating machinery. Various deep learning techniques, including CNNs, RNNs, and Deep Belief Networks (DBNs), have shown promising results in diagnosing different types of faults, such as bearing faults, gearbox faults, and motor faults. The studies reviewed in this paper demonstrate that deep learning-based methods can achieve high diagnostic accuracy, often outperforming traditional machine learning methods and expert systems.

Authors in [8] reviewed fault diagnosis techniques for permanent magnet AC machines and drives, including model-based and data-driven methods for detecting faults like stator and rotor faults, bearing faults, and inverter faults. Their study highlighted that model-based

methods such as the extended Kalman filter and sliding mode observer are as effective for fault diagnosis, while data-driven methods like machine learning and signal processing are noted for their potential in detecting incipient faults. Their study concluded that fault diagnosis techniques have great potential for improving the reliability and performance of these machines and drives, and further research is needed to develop more robust and accurate diagnostic methods.

In [9], authors compared non-linear directional residual and machine learning-based methods for fault diagnosis in complex engineering systems, presenting case studies for a wind turbine gearbox and an internal combustion engine. Results show both methods are effective, with machine learning showing superior performance in some cases. The paper notes that the approach chosen depends on the specific application and a combination of methods may provide the best results. The study highlights the potential of data-driven fault diagnosis techniques for improving complex engineering systems' reliability and performance. In [10] the latest developments in deep learning techniques applied to the diagnosis of faults in rotating machinery using vibration signals is discussed. The study reviews various deep learning architectures, including CNNs and RNNs, and presents case studies showing their effectiveness in fault diagnosis. Results show that CNNs are particularly effective. They highlighted the potential of combining deep learning with other techniques, such as signal processing and feature extraction, to further improve diagnostic accuracy. A summary of the related works is illustrated in Table 1.

Table 1 A summary of the related works.

Related works	Methods	Data	Findings	Limitation
[11]	Metric adversarial domain adaptation (MADA) approach	The proposed method is evaluated using the bearing dataset from the PRONOSTIA platform, which was collected for the IEEE PHM Challenge 2012.	MADA exhibits superior predictive performance and achieves reduced prediction errors compared to other methods	For predicting the remaining useful life, all possible necessary features were not extracted. Moreover, in terms of generalizability, obtained results may not be easily transferable to different domains or scenarios.
[12]	Short-time Fourier Transform (STFT) and CNN	To evaluate the efficacy of the proposed method, experiments are conducted using the bearing datasets from the Case Western Reserve University and the Machine Failure Prevention Technology Society.	The outcomes indicate that the proposed approach surpasses other comparative methods, achieving identification accuracies of 100% and 99.96% for CWRU and MFPT datasets, respectively.	The proposed method's performance in the study may be limited. It does not incorporate a comparison with advanced methods frequently used in fault diagnosis.
[13]	Improved Fray Wolf Algorithm (IGWO) and	Experimental data were collected from Western Reserve University.	The experimental findings demonstrate that the proposed methods effectively	The study proposed IGWO and SVM for fault diagnosis. Nevertheless, the

	Support Vector Machine (SVM)		enable fault diagnosis of rolling bearings. The average accuracy rate for fault diagnosis reaches 98.875%.	specific parameters and techniques used for the model optimization procedure did not extensively discuss.
[14]	Variational Mode Decomposition (VMD) and SVM	The data set of Case Western Reserve University.	The proposed parameter-optimized variational mode decomposition and support vector machine model in this study achieves and average accuracy rate of approximately 5% higher than other algorithms.	The study did not provide a comprehensive assessment or comparison with other present fault diagnosis techniques, which limits the capability to verify the efficiency and advantage of the suggested method.
[15]	Marine Predator Algorithm (MPA) and SVM	Case West Reserve University for their bearing datasets.	The attained accuracy of bearing fault diagnosis is 97.67%.	The sample size of this study is not effectively illustrated, which might affect the statistical consequence and the findings' reliability.
[16]	Multiclass Fuzzy Support Matrix Machine (MFSMM)	Experimental data were collected from Anhui University of Technology and Hunan University.	The experimental results demonstrate that the proposed methods exhibits a strong classification performance for roller-bearing fault diagnosis.	This study did not provide details information for the dataset testing which limits the generalizability of the method to the real-world environment.
[17]	K-Nearest Neighbor (KNN)	6205 deep groove ball bearing.	The test data classification accuracy stands at 83.3%.	The use of KNN for fault diagnosis might have constraints such as sensitivity to the choice of K value, and potential bias towards the majority class in imbalanced datasets.
[18]	SVM and Fast Fourier Transform (FFT)	Data was taken from Center Case Western Reserve University.	The combined SVM-FFT approach can be employed to diagnose broken bearings with exceptional accuracy, exceeding 99%.	The study integrated the SVM classifier with the FFT for feature extraction. Nevertheless, the

				limitations of FFT in capturing certain fault characteristics or exploring alternative feature extraction techniques did not discuss.

However, in the existing works recognizing relevant features and distinguishing them from noise remains challenging. To address these challenges, we present a comprehensive model that integrates various components to enhance the accuracy and effectiveness of machine process identification. The proposed model integrates a deep learning algorithm to enable accurate machine operation forecasting using acoustic signals. To enhance the analysis of the acoustic data, we apply a customized CWT technique, which preserves crucial information like signal amplitude by converting the signals into CWT images. This transformation enables a more comprehensive analysis and representation of the acoustic data. Moreover, we leverage the capabilities of a CNN as a robust classifier to precisely classify and distinguish between various machine processes by extracting features from the CWT images. By combining these elements (see Figure 1), our model provides a robust and efficient framework for machine process identification using acoustic signals. Therefore, the contributions of this study can be summarized as follows:

Figure 1 Proposed methodology for converting acoustic data to image and machine process classification.

- We employ a deep learning algorithm that enables accurate forecasting of machine operation based on acoustic signals. By leveraging the power of deep learning, the algorithm is capable of learning complex patterns and relationships within the acoustic data, allowing it to make reliable predictions about the machine's future performance.
- We employ a customized CWT technique to convert the acoustic signals into CWT images, preserving vital information such as signal amplitude. This transformation allows for a more comprehensive analysis and representation of the acoustic data.
- We utilize CNN as a powerful classifier to accurately classify and differentiate between different machine processes based on the extracted features from the CWT images.

Table 2 presents a comprehensive list of abbreviations used in this study, providing a convenient reference for readers to quickly understand and interpret the various abbreviations and acronyms used throughout the paper. The remainder of the paper is structured as follows: In Section 2, we present the proposed model. Section 3 explains the data collection process from an actual machine. In Section 4, we present a feature extraction. In Section 5 implementation of the proposed model is provided. Theoretical and practical implications are elaborated in Section 6. Finally, a conclusion is drawn in Section 7.

Table 2 List of abbreviations used in this study.

Acronyms	Description
CWT	Continuous Wavelet Transformation
CNN	Convolutional Neural Network
PHM	Prognosis and Health Management
DL	Deep Learning
RNN	Recurrent Neural Network
SST	Synchro Squeezing Transform
CDBN	Convolutional Deep Belief Network
ELMs	Extreme Learning Machines
DBNs	Deep Belief Networks
RUL	Remaining Useful Life
IWF	Institut für Werkzeugmaschinen und Fertigungstechnik
TP	True Positives
TN	True Negatives
FP	False Positives
FN	False Negatives
DNN	Deep Neural Network
SSL	Single Scale-Low
SSH	Single Scale-High
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
MAPE	Mean Absolute Percentage Error
MADA	Metric Adversarial Domain Adaption
STFT	Short-time Fourier Transform
IGWO	Improved Gray Wolf Algorithm
VMD	Variational Mode Decomposition
SVM	Support Vector Machine
MPA	Marine Predator Algorithm
MFSMM	Multiclass Fuzzy Support Matrix Machine
KNN	K-Nearest Neighbor
FFT	Fast Fourier Transform

2. Proposed Model

The present study puts forth a fault diagnosis model that utilizes CWT and CNN. CWT, which is comparable to Fourier Transform, calculates the correlation between a signal and an analyzing function through inner products. However, wavelets are mathematical functions that allow cutting data into different frequency parts to survey each part. When comparing wavelets and Fourier methods, wavelets have advantages in analyzing physical conditions, particularly in situations with sharp spikes and discontinuities. A physical notion of scale in acoustic and wavelet is called scaling or dilation, related to the signal's ability to compress and stretch. We used CWT to compare the compressed and shifted signal with the wavelet. By examining the signal and the wavelet at various positions and scales, it is feasible to generate a two-variable function. Consequently, the 1-dimensional signal can be expressed in the form of 2-dimensional excessive data. If the wavelet is deemed as a complex value, then the CWT is classified as a complex-valued function that has two parameters scale and position.

In mathematics, the CWT is an official tool that represents a signal by continuously varying the transform and scaling parameters of the wavelet. The signal is treated as a time-scale

plane by the CWT, where the wavelet basis function a finite-sized zero average function with a square-integrable area is scaled and translated. The CWT of a function $f(t)$ at a scale a and translational value b is expressed by the following integral:

$$C(a,b;f(t),\psi(t))(a,b) = \int_{-\infty}^{\infty} f(t)\frac{1}{a}\psi^{*}\left(\frac{t-b}{a}\right)dt \tag{1}$$

In Equation 1,* represents the complex conjugate, Ψ is known as the continuous function or mother wavelet which includes both the time domain and the frequency domain. The parameter b is defined as the moving element that transmits the wavelet over the time axis, while a represents the scale element that expands and contracts the wavelets, and it always is greater than zero ($a > 0$). Consequently, a long wavelet is received from the lower-frequency band, whereas a short wavelet is received from the high-frequency band. If the signal is real-valued, the CWT is a real-valued function of scale and position. To obtain the CWT coefficients $C(a, b)$, it is necessary to continuously vary the values of the parameter a and the position of b. In conclusion, the general correspondence between scale and frequency shows in Figure 2, which means:

- Small scale a ⇒ Compressed wavelet ⇒ Rapidly changing details ⇒ High frequency ω.
- Long scale a ⇒ Stretched wavelet ⇒ Slowly changing, coarse features ⇒ Low frequency ω.

Figure 2 Long and small scales.

To simplify notation, the relationship between the CWT coefficient, function, and analyzing wavelet has been omitted. As a result, the CWT provides advantages for analyzing signals in both the time and frequency domains. Furthermore, signals that exist in the time-frequency domain can be represented as a two dimensional image. Consequently, wavelet analysis facilitates the representation of images and signals in the time-frequency domain.

The CNN, which is extensively utilized in various fields like speech recognition, natural language processing, and computer vision processing, is a neural network structure that can evaluate an original image directly without complicated processing. By introducing three essential concepts, namely sparse interaction, parameter sharing, and equivariant representation, CNN operations have improved machine learning systems. The three critical layers of a CNN consist of the convolutional layer, the pooling layer, and the fully connected layer. In the convoluted convolutional layer, the input from the previous layer is processed using multiple kernels and sent to the activation function to create a feature map. Eventually, the convoluted results from convolutional layers pass into the activation function equation.

Here are the two most commonly used activation functions: the first function, Equation 2, is a rectified linear unit (*ReLU*).

$$ReLU\ (x) = \max(0, x) \tag{2}$$

One of the most frequently utilized sub-sampling layers, which follows the convolutional layer, is max-pooling. Sub-sampling layers create low-resolution maps by removing the essential features. Figure 3 demonstrates that the max-pooling layer, which is extensively used, operates by selecting the highest value from each designated region.

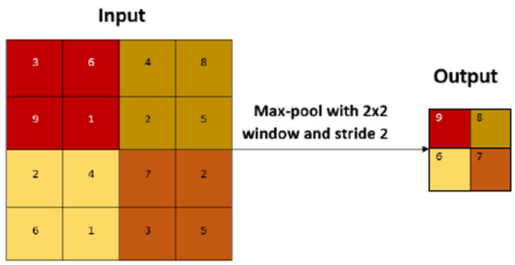

Figure 3 max-pooling transformation.

The flattened layer is created by merging all feature maps into a 1-D vector. Subsequently, in the fully connected layer, all neurons from both layers are connected in a manner similar to that of a conventional multilayer neural network. For example, bellow, the fully connected output is shown in function Equation 3:

$$O = f\left(\sum_{j=1}^{d} \chi_j^F \omega_j + b\right) \tag{3}$$

Where O as output amount, χ_j^F is jth neuron in FC layer, ω_j is the weight matching to O, at the end χ_j^F, b_i is bias according to O, f is known as a sigmoid function. A technique for predicting the Remaining Useful Life (RUL) using an image-based CNN has been introduced. This approach suggests 12 CNN architectures, including the conventional CNN and LetNet-5, and measures their performance against each other. Figure 4 presents the recommended CWT model. The proposed model works with 2-D images, which converted the CWT coefficients. Feature extraction is done in four convolutional, four ReLU, and four pooling layers. The convolutional layers and four average max-pooling layers have filter sizes of 3 × 3 and 2 × 2, respectively, with two strides for each layer. The four convolutional layers have 32, 64, 128, and 256 channel sizes. In this model, a CNN classification is achieved by connecting two Fully Connected layers, the last of which is connected to an output neuron. The resulting output is fed through a softmax function to produce four classes. To train the model, a loss function (Equation 4) is employed to minimize its error.

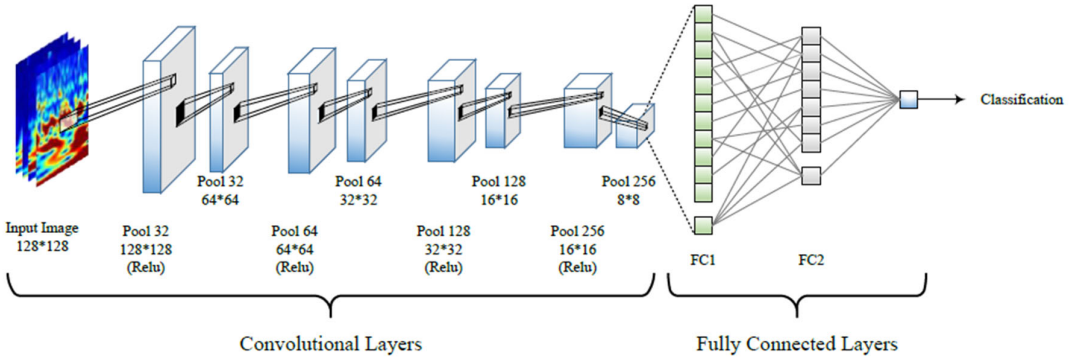

Figure 4 Proposed CNN model for classification process.

$$LOSS = \frac{1}{T}\sum_{t=0}^{T}\left(\Upsilon_t - \hat{\Upsilon}_t\right)^2 \tag{4}$$

To optimize the loss function, mini-batches are being used with the Adam optimizer. Additionally, normalization is applied to expedite the training process and address the issue of over-fitting in the model.

3. Data Collection

The dataset used in this study was generated by the Institut für Werkzeugmaschinen und Fertigungstechnik (IWF) in Germany. The data was collected from a single industrial machine called *CTX*420 from Gildemeister (Figure 5). This machine includes four different processes: Drilling, Turning, Milling, and Pocket Milling, respectively, which run one after the other, and data has been recorded of these processes. The dataset contained two different types of recording for each process. The first type records the normal process when the machine works correctly. The second type of recording is when the machine does not work correctly, which we call non-anomaly and anomaly data, respectively.

Figure 5 Fertigungstechnik und Werkzeugmaschinen.

The sampling rate for each acoustic signal is 12ms, and there are 82 samples for a second. The total data sample for the drilling process is 2500 seconds. Figure 6 displays collected data in time-frequency domain style. Moreover, Table 3 provides information regarding the

statistics of the dataset, including sampling time, mean, and standard deviation, during the drilling, turning, milling, and pocket milling processes.

Table 3 Data statistics.

Machine	#Data	Sampling time (ms)	Total sampling time (s)	Mean	SD
Drilling (B)	212796	12	2500	810.097	2826
Turning (D)	412864	12	4800	644.1412	816.64
Milling (T)	206121	12	3600	566.32	218.3923
Pocket milling (P)	317883	12	3500	398.1830	178.528

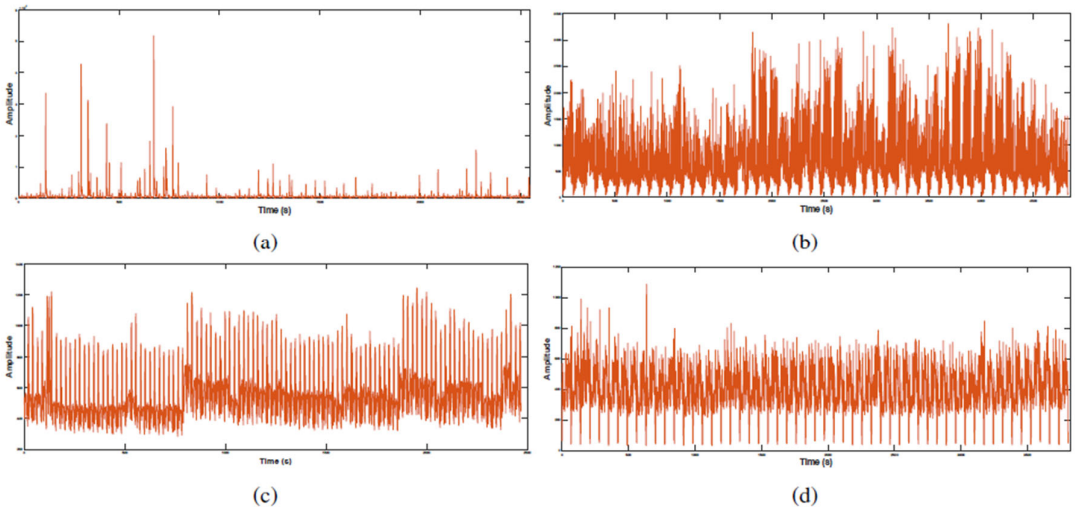

Figure 6 Raw acoustic signal of drilling (a), rotating (b), milling (c) and pocketing (d).

Figures 6a, 6b, 6c, 6d demonstrate raw acoustic signals which collected from rotating, milling, and pocketing respectively. The data has been recorded for the rotating procedure was 4800 seconds with the same position of microphones in the drilling procedure. This process has been repeated for the milling and pocketing process, and the result is 2500 seconds for the milling machine and about 3500 seconds of data for the pocketing machine.

4. Feature Extraction

Since acoustic signals are not linear and stationary, amplitude is one of the key factors in acoustic data. For instance, the amplitude for the milling procedure in our dataset is between 200 and 1200 dB (Figure 6c), but the amplitude for the pocketing process is almost between 0 to 1000 dB (Figure 6d).

In order to use the acoustic data with a deep learning algorithm, it is necessary to convert acoustic signals into a picture. This conversation is meaningful because the signal amplitude must be included in the converted images.

Therefore, in this work, $\mathcal{S}_r = \{s_{r1}, s_{r2}, s_{r3}, \ldots, s_{rn}\}$, $\mathcal{S}_p = \{s_{p1}, s_{p2}, s_{p3}, \ldots, s_{pn}\}$, $\mathcal{S}_d = \{s_{d1}, s_{d2}, s_{d3}, \ldots, s_{dn}\}$, and $\mathcal{S}_m = \{s_{m1}, s_{m2}, s_{m3}, \ldots, s_{mn}\}$ are considered as acoustic signals of rotating, pocketing, drilling, and milling process, respectively.

The CWT output is the converted images that represent I_r, I_p, I_d, and I_m. Here, $I_r = \{i_{r1}, i_{r2}, i_{r3}, \ldots, i_{rn}\}$, $I_p = \{i_{p1}, i_{p2}, i_{p3}, \ldots, i_{pn}\}$, $I_d = \{i_{d1}, i_{d2}, i_{d3}, \ldots, i_{dn}\}$ and $I_m = \{i_{m1}, i_{m2}, i_{m3}, \ldots, i_{mn}\}$ illustrate the input image of rolling, pocketing, drilling , and milling process for classifier algorithm.

$$L = \{l_r, l_p, l_m, l_d\} \tag{5}$$

L is considered as labeled the results.

To extract the features from raw acoustic signals, a customized Morlet-based continuous wavelet transform has been used. In fact, the wavelet function converts the acoustic signal generated by the machine into an image and includes signal amplitude. Figure 7 depicts the raw acoustic and contour plots of the wavelet spectrum produced through Morlet-based continuous wavelet transform for the rotating process under normal conditions. For instance, in Figures 7(e) and 7(g), there are 1 and 2 signals with higher amplitudes respectively which are visible clearly in Morlet-based CWT outputs (Figures 7(f) and 7(h)). Having amplitude peaks is not enough to achieve accurate predictions. Therefore, a colorful contour image of the wavelet spectrum has been generated to demonstrate energy distribution in acoustic signals. Areas with hotter colors, such as red, show high amplitude in acoustic signals, and color will be cooler when the energy in signals is decreased.

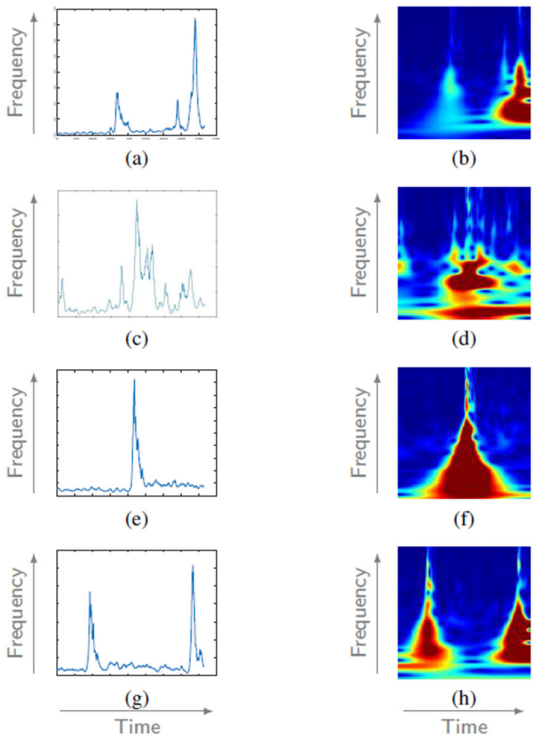

Figure 7 Time-frequency domain of acoustic signals of rotating raw signal in good condition, and contour plot of 40 seconds of the rotating process.

The accuracy of the proposed algorithm depends on the Morlet-based CWT output. Therefore, finding the best window size to generate the images is critical. To solve the windowing issue, a Monte Carlo technique has been used. Hence $W_s = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ which W_s represents the size of the window in seconds, and images have been generated for each time window, respectively. Accordingly, 12 sets of images have been used as input to run an initial experiment to find which time window can be used for the final algorithm. The results show 1, 2, 3, *and* 4 seconds are not suitable for the final simulation algorithm because the images did not contain acoustic signal features. Thus, the final set for windowing is as follows: $W_s = \{5, 6, 7, 8, 9, 10, 11, 12\}$. By considering window size, the next step is implementing a deep learning classification to identify the machine's process. Figure 8 shows images generated by Morlet-based CWT from the acoustic signal with a window size of 12 seconds.

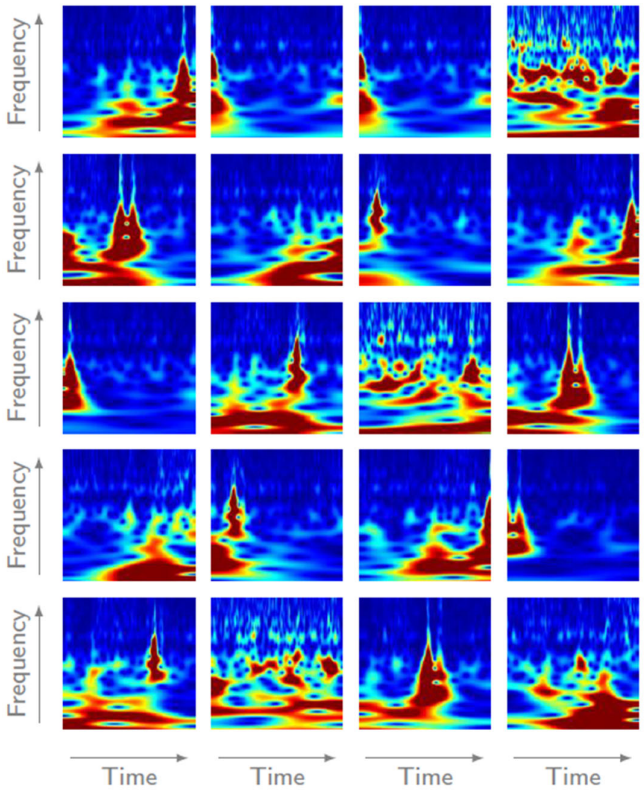

Figure 8 The wavelet power spectrum's contour plots during the pocketing process using a window size of 12 seconds.

5. Implementation of the proposed model

To implement the proposed model for classifying acoustic data from industrial machines using deep learning, all acoustic data is initially converted to images using CWT with several different time windows, in order to determine the best time window for predicting. Eight datasets are generated based on time windows of $W_s = \{5, 6, 7, 8, 9, 10, 11, 12\}$ seconds in length.

The classification algorithm utilizes the datasets that were created for both training and testing purposes. The training dataset is partitioned into three subsets, namely training,

validation, and testing, during the training phase. The trained algorithm is then tested on the test dataset to identify the classification accuracy and prevent over-fitting. Each simulation run is performed five times on each image dataset.

The experiment was conducted on a PC with an Intel core i9 CPU, 64 GB RAM, and a 2 TB SSD hard disk. MATLAB 2020 was used for programming and simulating the proposed algorithm. As mentioned in section 3, the data was recorded from *CTX*420 Gildemeister machine.

5.1. Result

This study designed a novel deep learning approach to clearly deal with diagnosing different types of faults (bearing faults). A deep learning classifier was carried out to train, test, and analyze acoustic signals initiated from diverse machining process. The results are summarized in Table 4. As mentioned before, the simulation started with five seconds of windowing, and we obtained 77.49% accuracy after five rounds of training and testing. The accuracy result significantly improve from 77.49% to 86.97%, almost 10% increase, when the window time was increased from five to eight seconds. Table 4 reveals that there is a correlation between classification accuracy and the duration of the window; as the window time increases, the accuracy also increases; therefore, a data length of 12s can classify all four processes with 95.87% accuracy. This implies that the notable correlation was observed between the duration of the window and the classification accuracy, indicating that as the window time increased, there was a corresponding improvement in accuracy.

Table 4 Accuracy mean and standard deviation for five times ruining simulation for each windowing.

Dataset name	SD	Mean
5s Length	1.77	77.49
6s Length	2.30	82.85
7s Length	1.80	83.71
8s Length	2.56	86.97
9s Length	5.31	89.77
10s Length	0.41	91.51
11s Length	2.52	92.585
12s Length	4.59	95.875

Figure 9 displays the accuracy of our model at timestamps $W_s = \{5, 6, 7, 8, 9, 10, 11, 12\}$ for drilling, turning, milling and pocket milling. The accuracy is calculated using Equation 6.

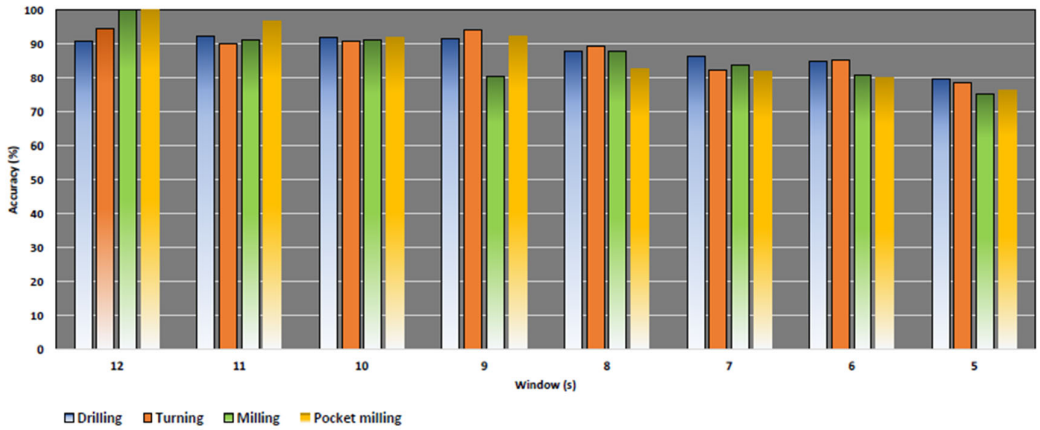

Figure 9 Accuracy comparison between different windowing in second, for drilling, turning, milling and pocket milling.

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

True positives and true negatives are denoted by TP and TN, respectively. FP indicates false positives, and FN represents false negatives. These metrics are commonly used to accurately measure sound detection classification systems.

As demonstrated in Figure 9, the accuracy of each process differs based on the corresponding window size. The variation is due to the differences in signal amplitude across the process. Our model achieves higher accuracy with a window size of 12, and accuracy decreases as the windowing time is reduced.

Figure 10a displays the confusion matrix for testing 12 seconds length images to identify classification accuracy for a single trial, which invalidated the experiment. According to the German language, the abbreviations B, D, T, and P are drilling, turning, milling, and pocket milling respectively, which are the four procedures of the CTX420 machine. The accuracy generally increased in the comparison between the two confusion matrices. However, in Figure 10b, when observing 9 seconds window time slides, drilling, and turning showed a slightly better result than in the 12 second windowing, likely due to random feature selection by the classification algorithm. Therefore, the experiment was repeated multiple times to prevent over-fitting and under-fitting. In order to achieve accurate classification results through deep learning methods, the number of images used to train the algorithm is a crucial factor. Increasing the number of images for training and testing can improve accuracy. However, this study suggests that the number of images alone is not the primary factor in classifying the acoustic data. Amplitude and signal energy must be considered in generating images. For example, the results show that a 12 second window time yields higher accuracy compared to other window times, despite requiring fewer images. Moreover, this study conducted F1 score accuracy based on Equation 7 to achieve an imbalance dataset accuracy. The precision and recall are considered the initial and second components of the F1-score which can also function as individual metrics within machine learning. Table 5 presents the accuracy, precision, recall, and F1-score results for 12 s acoustic classification.

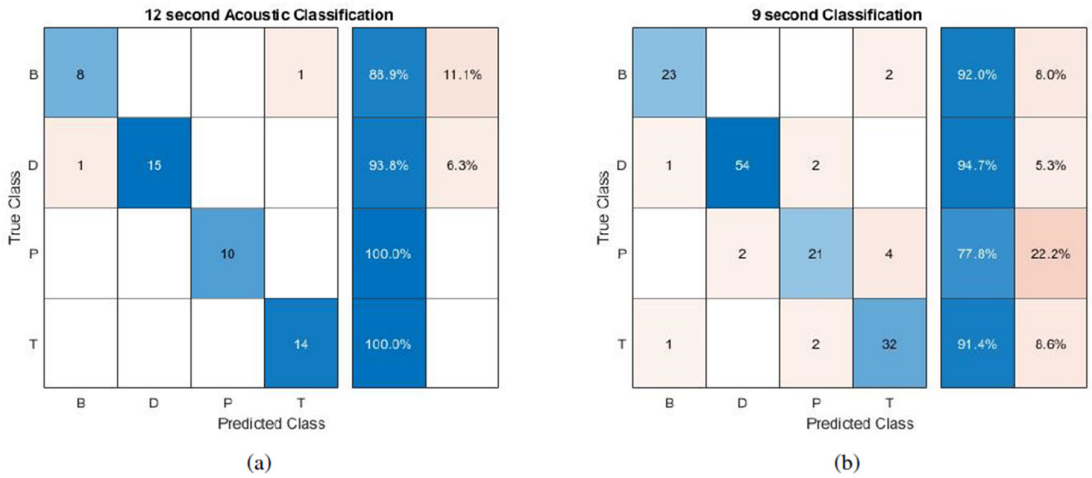

Figure 10 (a) and (b) are the confusion matrices for the 12 and 9-second acoustic classifications, respectively.

$$\text{F1-score} = 2 \times \frac{precision \times recall}{precision + recall} \tag{7}$$

Table 5 Results of F1-score for 12s acoustic classifications.

Machine	Precision	Recall	F1-score
Drilling (B)	100%	88.9%	93.6%
Turning (D)	100%	93.8%	96.3%
Milling (T)	100%	100%	100%
Pocket Milling (P)	100%	100%	100%

5.2. Comparison Approach

The proposed method's efficacy is shown using multiple deep neural network implementations, which incorporate multi-scale feature extraction. To achieve this, various alternative approaches are explored and compared, including:

- Deep Neural Network (DNN): The basic deep learning technique used is a multi-layer perceptron, also known as a DNN. Like the proposed network, this approach employs a pair of Fully Connected (FC) layers, each with 128 neurons, to connect every input neuron to every output neuron. This strikes a balance between the method's complexity and efficiency, allowing it to extract sequence features effectively. Then, an additional two FC layers are employed, with 64 and 1 neuron, respectively, for RUL regression [19].
- Single Scale-Low (SSL): In order to showcase the advancement of the proposed multi-scale feature extraction technique, the researchers utilized the Single Scale-Low method. This method employs a single convolutional layer to procedure every data order without concatenating any features during the initial extraction phase. Consequently, the method solely utilizes low-scale features to estimate the Remaining Useful Life (RUL) of a system.
- Single Scale-High (SSH): The SSH approach and the SSL method share a common characteristic in that they both don't take advantage of the multi-scale approach. Rather, the SSH approach implements a distinctive strategy by utilizing three

feedforward convolutional layers to extract features from each data sequence. These convolutional layers enable the SSH approach to capture intricate and significant characteristics from the data. Moreover, the high-level features obtained are utilized immediately for further processing. By doing so, the SSH approach reduces the complexity and computation time required for the analysis of the input data [19].

To evaluate the performance of the proposed algorithm three performance metrics has been used as follows:

- Mean Absolute Percentage Error (MAPE): MAPE in Equation 8 is a metric that measures the average percentage difference between the predicted Remaining Useful Life (RUL) values (E_i) and the actual RUL values (L_i) for each sample. Its calculation can be expressed as follows:

$$MAPE = \frac{1}{N_s}\left(\sum_{i=1}^{N_s}\left|\frac{L_i - E_i}{L_i}\right|\right) \times 100 \tag{8}$$

- Root Mean Square Error (RMSE): According to Equation 9, RMSE quantifies the square root of the average squared difference between the predicted RUL values (E_i) and the actual RUL values (L_i) for each sample. The formula is as follows:

$$RMSE = \sqrt{\sum_{i=1}^{N_s}(L_i - E_i)^2/N_s} \tag{9}$$

- Mean Absolute Error (MAE): MAE in Equation 10 represents the average absolute difference between the predicted RUL values (E_i) and the actual RUL values (L_i) for each sample. It is calculated as:

$$MAE = \sum^{N_s}|L_i - E_i|/N_s \tag{10}$$

The RUL of the i_{th} sample is denoted as L_i, and the subsequent RUL evaluation attained by the DNN is represented by E_i. There are a total of N_s samples, and each epoch involves randomly dividing the training samples into several mini-batches, with 32 samples in each batch. These mini-batches are then fed into the network. The weights and biases of each layer in the network are optimized based on the mean loss value of each mini-batch. The training process continues for 200 epochs, the learning rate has been fixed at 0.0001, which is an essential factor that influences the speed of learning of the machine learning model from the training data.

The research conducted a comparison validation using the PHM 2012 Challenge dataset [20], which was acquired from the PRONOSTIA platform. The dataset consists of information

about the breakdown of seventeen bearings that were initially in good condition without any defects. The bearings went through a natural degradation process, eventually leading to their failure at the end of the tests. The vibration signals were recorded at a frequency of 25.6kHz and 2560 data points (0.1 s) were sampled every ten seconds. The experiment was stopped once the monitoring data amplitude exceeded 20g to prevent any further damage to the entire test bed.

Table 6 shows that the proposed method performs better than other methods in predicting errors in various testing situations. This superiority is demonstrated by metrics like MAE, RMSE, and MAPE. Although SSL and SSH methods are competitive in several instances, the proposed method achieves this performance by utilizing a multi-scale feature extraction technique and makes it more resilient across diverse tasks. These findings suggest that the DNN architecture proposed in this study is proficient in capturing features associated with machine degradation, and the multi-scale feature extraction technique can adeptly manage intricate data patterns and extract insightful features and enhances the testing data's ability to generalize. Thus, the results presented validate the effectiveness and superiority of the proposed method when compared to existing approaches, underscoring its potential for practical implementation in real-world scenarios.

Table 6 Comparisons of the numerical performance of different methods.

Testing Bearing	Proposed			DNN			SSL			SSH		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
Bearing 1-1	24.7	26.9	63.8	29.6	42.3	171.6	26.2	28.7	69.8	26.5	28.9	79.5
Bearing 1-2	17.3	20.8	55.1	28.7	33.2	118.1	20.8	23.1	59.4	23.6	27.9	60.1
Bearing 1-3	9.8	10.3	16.8	17.4	23	61.9	11.2	14.6	17.7	9.9	12.1	25.3
Bearing 1-4	20.5	22.4	71.9	32.5	35.8	98.6	22.8	26.6	78.5	26.9	26	75.1
Bearing 1-5	17	20.3	56.3	28.2	31.6	102.3	20.2	22.6	75.4	18.9	19.5	76.0

6. Theoretical and practical implications

The methodological implications of this study are significant. Firstly, this research represents an initial attempt to identify and evaluate the challenges associated with fault diagnosis for machine operation in a production environment. Secondly, we have designed a precise CWT technique for converting signal data into CWT images. Thus, this technique enables a more accurate conversion of acoustic data into images for utilization by a convolutional neural network. Thirdly, CNN has been employed as a powerful classification algorithm in this work to classify the different machine processes based on the extracted features from CWT images. By combining CWT and CNN, we achieve over 97% accuracy in discovering and early detecting emerging faults in machine processes based on acoustic signals. By doing so, this study offers a comprehensive model that integrates various components to enhance the accuracy and effectiveness of machine process identification. The proposed model integrates a deep learning algorithm with acoustic signals, enabling accurate machine operation forecasting.

The findings of this study also have significant implications for several industries. For instance, these findings can assist in developing a more accurate and reliable system for fault diagnosis in machine production. Therefore, early fault diagnosis will aid in preventing further damage and decreasing the interruption of the machine. Furthermore, the time and cost savings are another significant aspect of the proposed method, CNN and CWT, which industries can employ for early fault diagnosis of machines. This leads to enhancing productivity, timely intervention, and saving costs by preventing machine breakdowns.

7. Conclusion

In this study, a deep learning-based approach has proposed for accurate machine operation forecasting and analysis based on acoustic data. The model has achieved over 97% accuracy in detecting emerging faults and identifying machine processes by utilizing customized CWT and CNN. The dataset used in this study is obtained from a machine, namely the *CTX*420 Gildemeister machine. In future, we plan to focus on incorporating larger datasets with diverse machines to improve generalizability, exploring unsupervised machine learning algorithms for predicting machine failure, and incorporating multiple data sources like vibration or temperature to enhance the model's accuracy and applicability.

References

[1] Alberto Rivas, Jesús M Fraile, Pablo Chamoso, Alfonso González-Briones, Inés Sittón, and Jua nM Corchado. A predictive maintenance model using recurrent neural networks. In *14th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2019) Seville, Spain, May 13–15, 2019, Proceedings 14*, pages 261–270. Springer, 2020.

[2] Weihua Li, Ruyi Huang, Jipu Li, Yixiao Liao, Zhuyun Chen, Guolin He, Ruqiang Yan, and Konstantinos Gryllias. A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges. *Mechanical Systems and Signal Processing*, 167:108487, 2022.

[3] Xiang Pan, Zefeng Cheng, Zheng Zheng, and Yehui Zhang. Sparse bayesian learning beamforming combined with short-time fourier transform for fault detection of wind turbine blades. *The Journal of the Acoustical Society of America*, 145(3):1802–1802, 2019.

[4] Rui Zhao, Ruqiang Yan, JinjiangWang, and Kezhi Mao. Learning to monitor machine health with convolutional bi-directional lstm networks. *Sensors*, 17(2):273, 2017.

[5] Huimin Zhao, Jie Liu, Huayue Chen, Jie Chen, Yang Li, Junjie Xu, and Wu Deng. Intelligent diagnosis using continuous wavelet transform and gauss convolutional deep belief network. *IEEE Transactions on Reliability*, 2022.

[6] Zhuyun Chen, Konstantinos Gryllias, and Weihua Li. Mechanical fault diagnosis using convolutional neural networks and extreme learning machine. *Mechanical systems and signal processing*, 133:106272, 2019.

[7] Shengnan Tang, Shouqi Yuan, and Yong Zhu. Deep learning-based intelligent fault diagnosis methods toward rotating machinery. *Ieee Access*, 8:9335–9346, 2019.

[8] Seungdeog Choi, Moinul Shahidul Haque, Md Tawhid Bin Tarek, Vamsi Mulpuri, Yao Duan, Sanjoy Das, Vijay Garg, Dan M Ionel, M Abul Masrur, Behrooz Mirafzal, et al. Fault diagnosis techniques for permanent magnet ac machine and drives—a review of current state of the art. *IEEE Transactions on Transportation Electrification*, 4(2):444–463, 2018.

[9] Nicholas Cartocci, Marcello R Napolitano, Francesco Crocetti, Gabriele Costante, Paolo Valigi, and Mario L Fravolini. Data-driven fault diagnosis techniques: non-linear directional residual vs. machine-learning-based methods. *Sensors*, 22(7):2635, 2022.
[10] Bayu Adhi Tama, Malinda Vania, Seungchul Lee, and Sunghoon Lim. Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals. *Artificial Intelligence Review*, pages 1–43, 2022.
[11] Jichao Zhuang, Minping Jia, and Xiaoli Zhao. An adversarial transfer network with supervised metric for remaining useful life prediction of rolling bearing under multiple working conditions. *Reliability Engineering & System Safety*, 225:108599, 2022.
[12] Qi Zhang and Linfeng Deng. An intelligent fault diagnosis method of rolling bearings based on short-time fourier transform and convolutional neural network. *Journal of Failure Analysis and Prevention*, pages 1–17, 2023.
[13] Lin Li, Weilun Meng, Xiaodong Liu, and Jiyou Fei. Research on rolling bearing fault diagnosis based on variational modal decomposition parameter optimization and an improved support vector machine. *Electronics*, 12(6):1290, 2023.
[14] Yumin Wang, Chao Jiang, and Li Su. Fault diagnosis method of rolling bearing based on variational mode decomposition algorithm of parameter optimization and support vector machine. In *Proceedings of 2022 Chinese Intelligent Systems Conference: Volume II*, pages 760–776. Springer, 2022.
[15] Mingxiu Yi, Chengjiang Zhou, Limiao Yang, Jintao Yang, Tong Tang, Yunhua Jia, and Xuyi Yuan. Bearing fault diagnosis method based on rcmfde-splr and ocean predator algorithm optimizing support vector machine. *Entropy*, 24(11):1696, 2022.
[16] Haiyang Pan, Haifeng Xu, Jinde Zheng, Jin Su, and Jinyu Tong. Multi-class fuzzy support matrix machine for classification in roller bearing fault diagnosis. *Advanced Engineering Informatics*, 51:101445, 2022.
[17] HS Kumar and Gururaj Upadhyaya. Fault diagnosis of rolling element bearing using continuous wavelet transform and k-nearest neighbour. *Materials Today: Proceedings*, 2023.
[18] Pramudyana Agus Harlianto, Noor Akhmad Setiawan, and Teguh Bharata Adji. Combining support vector machine–fast fourier transform (svm–fft) for improving accuracy on broken bearing diagnosis. In *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, pages 576–581. IEEE, 2022.
[19] Xiang Li, Wei Zhang, and Qian Ding. Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. *Reliability engineering & system safety*, 182:208–218, 2019.
[20] Patrick Nectoux, Rafael Gouriveau, Kamal Medjaher, Emmanuel Ramasso, Brigitte Chebel-Morello, Noureddine Zerhouni, and Christophe Varnier. Pronostia: An experimental platform for bearings accelerated degradation tests. In *IEEE International Conference on Prognostics and Health Management, PHM'12.*, pages 1–8. IEEE Catalog Number: CPF12PHM-CDR, 2012.