Compound Fault Detection of Rotating Machinery in Unobserved Conditions using Missing Data

Somaye Mohammadi[1*], Mehdi Behzad[2], Hesam Addin Arghand[3]

1- Assistant Professor, Mechanical Engineering Department, Sharif University of Technology, Tehran, Iran
2- Professor, Mechanical Engineering Department, Sharif University of Technology, Tehran, Iran
3- Assistant Professor, Mechanical Engineering Department, Faculty of Engineering, University of Zanjan, Zanjan, Iran

* Corresponding Author: somaye.mohammadi@sharif.edu

Highlights

- Introducing a model for simultaneous detection of gear-bearing-rotor faults
- Developing a fault detector insensitive to repetition, sensor position, and operating conditions
- Providing the detection model that can be used for missing data
- Proposing an improved algorithm for fault detection under new operating conditions

Abstract

In recent years, there has been a rise in the popularity of using data-driven artificial intelligence models for detecting faults in rotating machinery. The challenge lies in creating a model that can be used even when sensor data is not available and the operating conditions differ from those observed during development. This article addresses the issue of potential failures in gear, bearing, and shaft components and suggests two strategies - adjusting entry and cost functions - to address these challenges in developing a one-dimensional convolutional neural network model. These strategies enable the model to extract features from the input signal with minimal dependency on operating conditions. By analyzing the 2009 PHM (Prognostics and Health Management society) challenge competition dataset, the model achieved its highest accuracy by using the frequency spectrum of velocity and acceleration from vibrational signals. The model's average accuracy for signals recorded by any arbitrary sensor is 79.6%, even if some operating speeds were not observed during training. Incorporating a suggested penalty function based on p-value into the cost function increased accuracy by up to 13.6%. Consequently, implementing the proposed strategies in similar cases is highly recommended, as demonstrated by successful application in two industrial cases.

Keywords: Condition Monitoring, Fault Detection, Convolutional Neural Network, Gear Fault, Rolling Bearing, Rotor Faults, Missing Data.

1- Introduction

Health management of industrial machinery by diagnosing faults and evaluating their severity to prevent catastrophic failures is an indispensable part of nowadays industries' lifecycle. Three main challenges in fault detection of rotating machinery when solving real-life tasks are distinguishing between multiple faults that occurred simultaneously, dealing with missing (unknown or incomplete) input data, and detecting faults in unobserved conditions.

Rotating machinery consists of different parts like rotors (shafts), bearings, and sometimes gears. Each of them is susceptible to failure, and it is laborious to distinguish them if their failures occur concurrently. Most research has diagnosed a single part's faults assuming that the other parts are healthy. Given that reliable industrial dataset with compound faults is not usually available, studies have used laboratory dataset. Lyu et al., in 2019 [1], introduced the maximum correlated deconvolution based on quantum genetic algorithm to diagnosis compound planetary gear and bearings faults. Zhiyi et al., in 2020 [2] and Xin et al., in 2021 [3] detected rotor-bearing faults using infrared thermal images by convolutional neural network (CNN) for their simulation experimental platform. Xue et al. [4], used combined CNN and support vector machine (SVM) for diagnosis of rotor-bearing faults. Models trained on laboratory datasets have the potential to be generalized to industrial cases using transfer learning techniques. Among the available laboratory datasets, the most diverse one with simultaneous faults on the rotor, rolling bearings, and gears is the prognostics and health management (PHM) 2009 challenge competition dataset. It has been widely utilized as a reference dataset to demonstrate the effectiveness of newly developed models.

In vibration analysis, as a common tool in condition monitoring (CM) technique, measurement is generally done at two points in the vertical, horizontal, and axial directions. Some tests may not be performed due to the lack of this number of sensors and the inaccessibility to the appropriate locations to put the sensors. Moreover, failure of data collection equipment may cause some sensor data to be lost in some cases. All these cases make the input data, which is the input feature vector of fault detectors, have a variable size for each data. Therefore, the detector, which is actually a classifier, is expected to cope with missing data.

Classification techniques to handle missing values can be grouped into four different types of approaches. In the first approach, incomplete data are removed, while they may be informative. The second approach first imputes the missing values (completes them based on the available data) before classification. These methods are not usually successful in the test set, because of the separate imputation and classification steps. So, the third approach combines imputation and classification tasks by the methods like multitask learning and multiple imputation. The last approach forms an ensemble of a one-class classifier trained on each feature and a decision is made based on the active classifiers for each data using fuzzy logic, Gaussian mixture models, etc. [5]

In the field of fault diagnosis, a few research concentrated on missing data issues. Zhang and Dong, in 2014 [6], used a Bayesian-based approach to monitor continuous stirred-tank reactors using the sensed thermodynamic features. Zhang et al. [7] employed an expectation–maximization algorithm to handle missing data in fault diagnosis of a ball-and-tube system based on ultrasonic sensor data. Liu et al., in 2018 [8], imputed missing data in a chemical process platform using a deep learning method. Venkatasubramanian et al., in 2022 [9], managed denoising, missing data imputation, outlier discovery, and data fusion for the vibration data of the Case Western Reserve University bearing dataset using an ensemble network.

Since the stiffness of a rotor in rotating machinery varies in different positions, imputing the data based on the available data in different positions is associated with many uncertainties. Therefore, the fourth approach in dealing with missing data is in the spotlight. It is intended to develop an artificial intelligent (AI) model that can comment on the machine's state based on the available data, similar to CM experts but with a more systematic view. Among AI models, a common solution to develop a model applicable to unseen conditions is to use transfer learning (TL) methods. PHM 2009 dataset with compound rotor-bearing-gear faults has been repeatedly used to evaluate TL models based on adaptive CNN [10], adversarial CNN [11, 12], end-to-end CNN [13], combined multi-layer perception (MLP)-CNN [14], etc. when facing unobserved conditions.

This research aims to develop an AI-based model insensitive to operating and measurement conditions for detecting compound faults in rotating machinery. In this regard, the PHM 2009 dataset is used and therefore the model's input is vibration signals. Previous publications have dealt with two simultaneous defects in machine parts, but this study studies three simultaneous rotor-bearing-gear faults. Notably, while addressing the issue of missing data, existing research has largely overlooked the specific realm of gear defects, a gap that is squarely addressed in this article. In this research, "missing data" pertains to the unavailability of sensor data from inaccessible locations. Given that an AI model with a fixed number of input features is unable to generate output when facing missing data, it is imperative to develop a model that is unaffected by sensor location. The robustness of the model is achieved by equipping it with diverse data without any information about the operating conditions, such as speed, loading, sensor location, and gear type (spur or helical). Thus, the model attempts to classify inputs irrespective of the location of measurement and the prevailing operating conditions. As an additional innovation, this research modifies the loss function of the CNN to compel it to extract features insensitive to varying operating conditions when faced with missing data or unobserved operational scenarios, thereby enhancing the model's accuracy under such conditions.

The rest of the research is organized as follows: The next section briefly introduces the PHM 2009 dataset. Section 3 describes the architecture of CNN models for detecting rotor, bearing, and gear faults and selects the best input. Then, it explains how this research deals with missing data by adjusting the model's inputs. Section 4 proposes a modified CNN model to extract insensitive features and train a model for unobserved conditions. Section 5 reports the potential of the proposed model in fault detection of two industrial cases. The summary, conclusion, and future interests are given at the end.

2- Dataset

PHM 2009 challenge dataset is related to a two-stage laboratory gearbox with four gears, shown in Fig. 1. Its vibrations have been measured by two accelerometers on both sides of the box with a sampling rate of 66.7 kHz. Two types of simple and helical gears with a reduction ratio of 5:1 have been tested with, respectively, 8 and 6 sets of each in different states of health and failure on gears, bearings, and shafts according to Table 1. The faults analyzed in this dataset include a chipped gear tooth, eccentric gear on its shaft, broken tooth, damaged inner race/outer race/ ball bearings, imbalanced shaft, and bent shaft [15]. The tests have been carried out twice at five speeds from 30 Hz to 50 Hz with a step of 5 Hz for low and high load conditions. The total number of data is 560 (5 speeds * 2 loads * 2 repetitions * 2 sensors * (8 sets of spur gears + 6 sets of helical gears)).

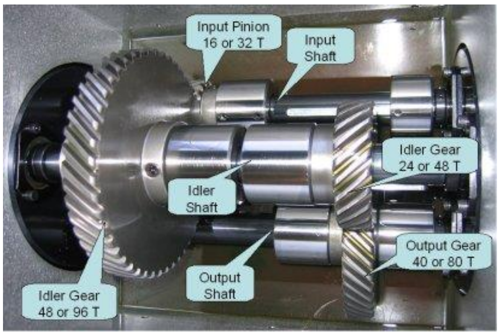

*Fig. 1 PHM 2009 data challenge **gearbox*** [15]

Fig. 2 depicts the number of healthy and defective PHM 2009 data with a single fault on one element and combined faults on several elements. It can be seen that there is no data in two cases including the failure of only bearings and the simultaneous failure of gears and shafts. The purpose of the model introduced in the next section is to detect the healthy or unhealthy state of each element (gear, bearing, and shaft).

*Table 1 Unhealthy elements of PHM 2009 **gearbox*** [15]

Element	Case													
	Spur Gears Set								Helical Gears Set					
	1	2	3	4	5	6	7	8	1	2	3	4	5	6
Gears		✓	✓	✓	✓	✓				✓	✓		✓	
Bearings				✓	✓	✓	✓	✓			✓	✓	✓	
Shafts						✓	✓	✓			✓	✓		✓

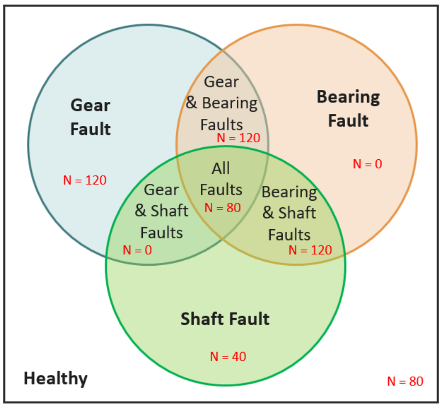

Fig. 2 Number of PHM 2009 data with no, single, and compound faults on gears, bearings, and shafts

3- Entry Adjustment Strategy

This section focuses on developing a model as insensitive to the operating conditions and missing data as possible. The strategy is to adjust the model's inputs and force it to provide a robust model. First, the model architecture is introduced, then the trick used to make the model applicable to missing data cases is described. This entry adjustment strategy can also help make the model insensitive to the operating conditions. Finally, the results are given, and the inputs are changed to increase the accuracy.

3-1- Model Architecture

Due to the complexity of extracting the appropriate features from the recorded signals to detect the defects of various elements, feature extraction from the signals has been left to the intelligent model in this research. In this regard, an1D-CNN model with the architecture shown in Fig. 3 is employed. In convolution layers, features are extracted from the input acceleration signals, and then the extracted features extracted by the intelligent model are used to classify the data by a fully-connected network with a hidden layer. This classification model determines whether the element is healthy or damaged.

The equations governing the convolutional and pooling layers can be expressed as follows:

$$OC(i,j) = I * C = \sum_m \sum_n I(i+m-1, j+n-1) C(m,n) + B \quad (1)$$

$$OP(i,j) = A \max_{m,n=1,\dots,k} \left\{ OC\begin{pmatrix} k(i-1)+m, \\ k(j-1)+n \end{pmatrix} \right\} + B \quad (2)$$

in which, *OC*, *OP*, *I*, *C*, *A*, and *B* respectively output of convolutional layer, output of the max pooling layer, input signal, convolutional filter, unknown multiplier, and unknown bias. Moreover, the equation representing the fully-connected network is as follows:

$$y = f_3\left(f_2\left(f_1\left(OF \times U\right) \times V\right) \times W\right) \quad (3)$$

in which, *OF* is the output of the convolution layers, and *U*, *V*, and *W* are unknown weights between flatten layer and fully-connected input layer, fully-connected input layer and hidden layer, hidden layer and output layer, respectively. Also, f_1, f_2, and f_3 are nonlinear activation functions of the fully-connected layers. Further information on how to write the mentioned equations can be find in [16] and [17]. It must be noted that all foundational codes of this research have been developed utilizing functions available in MATLAB software. The architecture and structure of the models have been defined using appropriate coding techniques.

Three CNN models are trained for fault detection in gears, rolling bearings, and shafts. Each model is trained three times, each time with one of the three inputs of the frequency spectrum, envelope spectrum, and cepstrum, so that the appropriate input is selected. The investigated frequency range is from 4 to 4800 Hz with a step of 1 Hz.

The input data for training each model are randomly divided ten times into three categories of training, validation, and testing in order to evaluate the sensitivity of the model to the available training data. Therefore, each model must be trained a total of 30 times (ten batches of random data division for each of the three different inputs).

It should be noted that oversampling is also used for training and validation data. So that the number of available data with the same labels (healthy or unhealthy) is equal and the model is not biased towards the data of a class with more data.

Selecting an appropriate number of Convolutional and Pooling layers is crucial to ensure optimal learning capacity within the model. Too few layers may hinder learning while an excess may lead to overfitting. In this study, overfitting has been prevented by leveraging the validation dataset. The stopping criteria have been determined based on the analysis of errors observed in both the training and validation datasets. Subsequently, alterations have been made to the number of layers and the extracted features for three distinct inputs, encompassing spectrum, envelope, and cepstrum data. A comprehensive evaluation has been conducted through the generation of boxplots illustrating the error distribution across multiple runs, in relation to the number of extracted features from each input as see in Fig. 4. The outcomes, encompassing training and validation data, revealed that setting the number of extracted features to 30 yielded the highest accuracy across all inputs, exhibiting minimal variance across different iterations. Therefore, this configuration was deemed optimal for the architecture, as a start to apply the proposed methods on in this research.

3-2- Struggleing with Missing Data

The CNN structure introduced in the previous section is a classifier for fault detection. The approach of this research to provide a classifier insensitive to the sensor's location is described here. The sensor location-insensitive model can detect the fault even if data from only one sensor is available or if missing data is encountered. The approach is to import data from different sensors as independent inputs but with the same labels into the model in the training phase. Therefore, the model is trained to obtain the same output for both data, regardless of which sensor the data is from. In other words, a model that has not received any information about the location of the corresponding sensor is forced to classify the signals of both sensors into the same category. It must be noted that this proposed approach has the advantage that the the number of missing data holds no significance in fault detection.

3-3- Insensitive Model to Operating Conditions

To develop a model insensitive to operating conditions (speed and load), the same approach as considered in the previous section to deal with missing data can be used. In this way, no information about the speed and loading conditions is given to the model during the training phase. This model can adopt this ability to detect healthy or unhealthy data regardless of operating conditions.

3-4- Model with One-Channel Input

The results of three fault detector models for gears, bearings, and shafts are given in Table 2. Each model's results for three inputs, including spectrum, envelope, and cepstrum, are presented on training, validation, testing, and total data. It must be noted that by taking into account the number of data points within the three categories of training, validation, and testing, the accuracy of the model across all datasets are computed and reported as the total accuracy. The highest precision of the gear and bearing detectors is for the spectrum input by approximately 3.4% and 1.8% higher average accuracy than the other inputs. The shaft detector provides the highest accuracy with the cepstrum input for the entire data by an almost 0.6% higher accuracy than the other inputs. Therefore, the appropriate input identical for all detectors is the frequency spectrum of the vibrational acceleration signal.

All elements, including gears, bearings, and shafts, may be defective in real-condition monitoring problems. Therefore, it is necessary to combine the results of the three fault detectors. With this point of view, the hybrid model's accuracy is calculated in such a way that its outcomes are accepted only if the failure of all elements is correctly expressed.

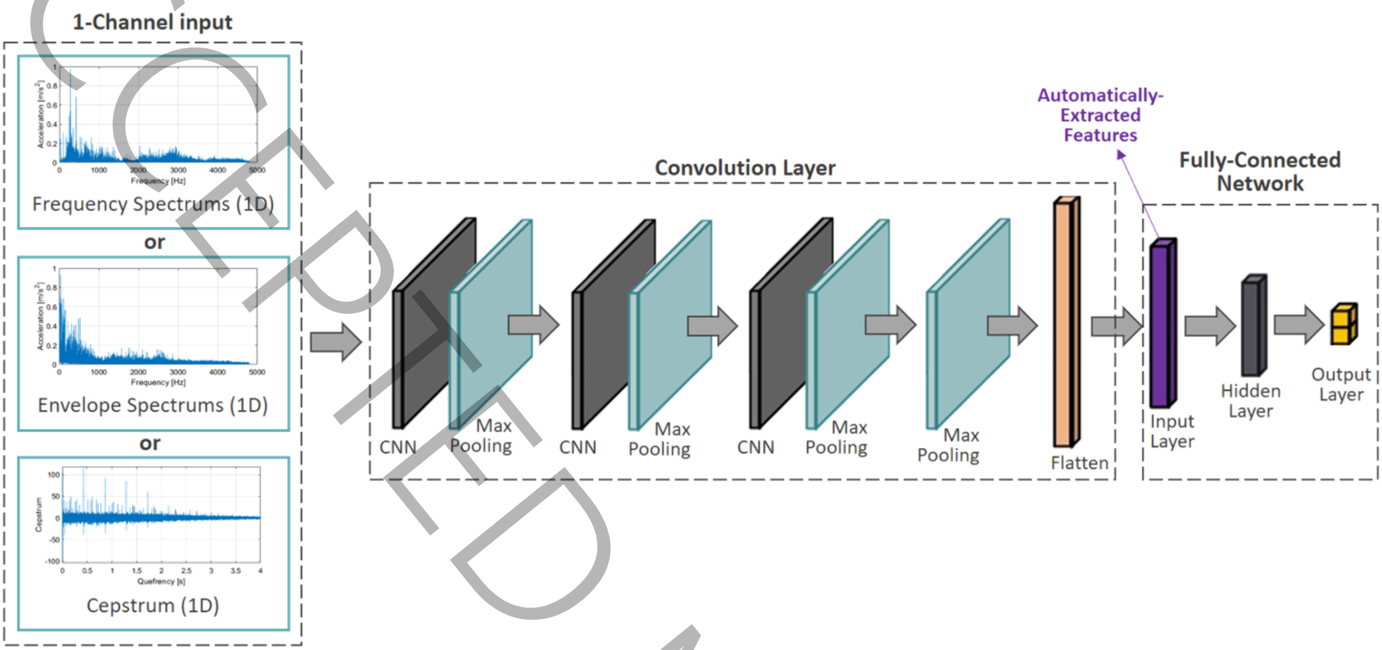

Fig. 3 1D-CNN architecture for fault detection of rotating machinery with one of three inputs including spectrum, envelope, and cepstrum

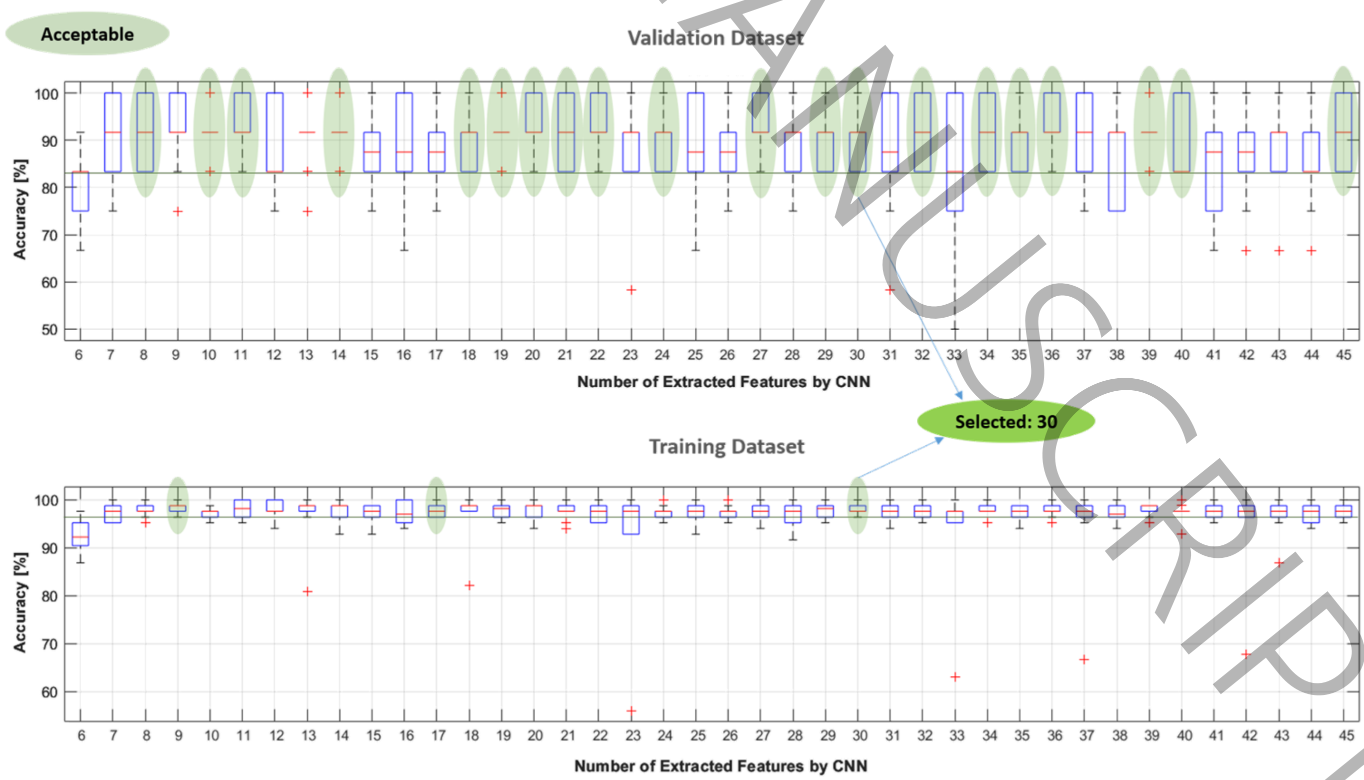

Fig. 4 Selecting the Number of Extreacted Features by the 1D-CNN for fault detection of rotating machinery with each of three inputs including spectrum, envelope, and cepstrum

Table 2 The accuracy (%) of the one-channel CNN in detecting gear, bearing, and shaft faults with different input

Fault Detector	Input: Spectrum				Input: Envelope				Input: Cepstrum			
	Train	Validation	Test	Total	Train	Validation	Test	Total	Train	Validation	Test	Total
Gear	100	94.5	93.1	97.9	98.6	85.6	81.9	93.6	100	88.0	84.6	95.4
Bearing	100	93.5	91.5	97.6	100	86.1	89.4	95.8	99.9	88.0	86.2	95.9
Shaft	99.3	82.9	77.7	92.9	99.6	79.2	76.1	92.3	98.9	83.3	78.7	93.4

Table 3 presents the accuracy of the combined model with spectrum input for different cases of element defects. It can be seen that the results get worse compared to Table 2 because the hybrid model gives wrong results if only one fault detector gives invalid result. The average accuracy of the model is approximately 90.9%.

Table 3 Hybrid one-channel CNNs's accuracy with spectrum input in different defective cases

Defective Element			Total Accuracy (%)
Gear	Bearing	Shaft	
			87.9
✓			91.4
		✓	88.8
✓	✓		96.4
	✓	✓	90.8
✓	✓	✓	89.9

The outputs of hybrid models trained ten times with different inputs randomly divided into three categories of training, validation, and testing are different. To show the model's sensitivity to the initial data distribution, the results are presented in a box plot in Fig.5. This diagram shows the sensitivity of different combinations of defects in the elements. It can be seen that the sensitivity of the model to the division of the primary data, or in other words, the data available used for training is approximately between 8 and 15%.

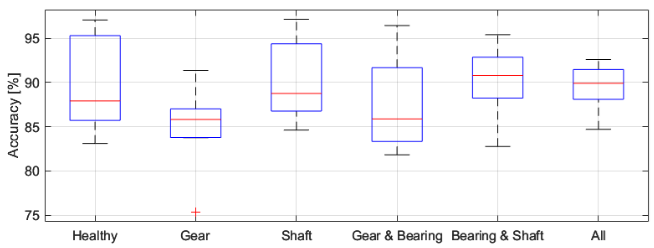

Fig. 5 The accuracies of the hybrid models with spectrum input trained with different data for different combinations of faults- All: gear, bearing, and shaft faults

In the training step, cases with different operating conditions have been entered into the model disregarding these variations. The aim was to develop an insensitive model to operating conditions in this way. Table 4 shows the detection accuracy at different conditions. It can be seen that the model is not sensitive to the sensor's location (drive end or non-drive end). Therefore, if the information for each sensor is unavailable (in the case of missing data), the model still responds with the same accuracy. The model's sensitivity to speed and loading is below 4.5% and is significant. As a result, if the data collection is done at a different rotor speed and load, this detector can still assess the failure status of the elements. The sensitivity to the type of gear (spur or helical) is relatively high and is equal to 10.7%. The following section attempts to increase accuracy by adding another channel to the model architecture.

Table 4 Hybrid model's accuracy with spectrum input in different conditions

Different Cases		Total Accuracy (%)
Loading	Low	86.4
	High	90.0
Input Shaft Speed [Hz]	30	87.1
	35	89.3
	40	89.7
	45	84.8
	50	91.5
Sensor Location	Input	88.9
	Output	88.0
Gear Type	Spur	93.0
	Helical	82.3

3-5- Model with Two-Channel Input

Depending on the speed of the output shaft (6-10 Hz), some faults may not be well represented in the acceleration signal. With this approach, it is possible to have the vibrational velocity signal as the input to a 1D-CNN model. Furthermore, another model can be developed by considering both velocity and acceleration as inputs of a two-channel CNN. The results of these newly trained models with spectrum input are presented in Table 5. For all three fault detectors, the accuracy with the vibrational velocity input is approximately 1.3% lower than the model with the vibrational acceleration input. Considering both inputs, the detection accuracy reaches 100%. Therefore, there is information in both signals that are necessary for the precise classification.

Although the accuracy has reached 100% with two channels, the noteworthy point is that the model had seen at least one data sample of different operating conditions in its input while training. The question is whether the prediction will still be accurate if the signals are input to the detector in unprecedented conditions. To answer this question, the model is trained by entering data related to some rotor speeds and then used for fault detection at the other speeds. The results of this study are presented in Table 6. It can be seen that the accuracy of the two-channel model with spectra input has decreased to 79.6%. The next section proposes a method to increase accuracy in such situations.

Table 5 The accuracy (%) of the one-channels and two-channel CNNs in detecting gear, bearing, and shaft faults with different input

Fault Detector	Input: Accelration				Input: Velocity				Inputs: Acceleration & Velocity			
	Train	Validation	Test	Total	Train	Validation	Test	Total	Train	Validation	Test	Total
Gear	100	94.5	93.1	97.9	100	90.3	86.7	96.5	100	100	100	100
Bearing	100	93.5	91.5	97.6	99.6	89.4	91.5	96.3	100	100	100	100
Shaft	99.3	82.9	77.7	92.9	99.2	80.6	76.6	91.8	100	100	100	100

Table 6 The accuracy of the two-channel CNN is evaluated in unobserved conditions while training

Trained Speeds [Hz]					Default Features			
30	35	40	45	50	Train	Validation	Test	Total
	✓		✓	✓	96.5	71.3	59.4	**79.6**
	✓	✓		✓	99.7	88.5	60.7	**83.7**
✓		✓		✓	79.5	73.8	58.5	**71.1**
✓		✓	✓		100	88.9	58.0	**82.8**
✓	✓		✓		98.4	77.8	57.1	**80.6**
Average (%)					94.8	80.1	58.7	**79.6**

4- Cost Function Adjustment Strategy

This section proposes a modification of the cost function of the 1D-CNN model to help convolution layers extract features that are as independent of speed as possible. After explaining the proposed model, the results are presented.

4-1- Proposed Cost Function

According to Fig. 3, the first part of the CNN model extracts 30 features from the input signal, listed in the flattened layer. The dependence of these features can be quickly evaluated using the p-value parameter. A higher value corresponds to a greater dependence of the speed on features. A p-value higher than 0.95 is generally significant and observed for all models trained.

For the models whose results are presented in Table 5, the p-value expressing the effect of speed on the 30 extracted features is calculated and shown in Fig. 5. Almost 40% of the calculated P-values are greater than 95%, indicating a significant dependence of the features on speed.

The following cost function with a penalty function on the p-value is proposed to extract features with negligible rotor speed dependence:

$$E = \sum_{n=1}^{N}(t_i - y_i)^2 + 10^{10}\sum_{f=1}^{F} g_f \qquad (4)$$

$$g_f = \begin{cases} 1 & p_{value}(feature(f), speed) \geq 0.95 \\ 0 & otherwise \end{cases} \qquad (5)$$

in which, E, N, F, t, y, and g_f, are the modified cost function of the 1D-CNN model, the number of training samples, the number of extracted features, the target state of the samples, the predicted state of the samples, and penalty function, respectively. The results of applying this cost function are given in the following section.

4-2- Model Architecture

The proposed cost function forces the CNN model to provide features with a p-value less than 0.95 in evaluating their sensitivity to speed. The accuracy of the model trained in this way is presented in Table 7. Compared to Table 6, the total accuracy for all cases has increased from 0.4% to 13.6%. In some cases, the accuracy of the training and validation data has been slightly reduced because the model was looking for speed-insensitive features, resulting in higher accuracy in the testing data. The average increment in accuracy using this method is 5%, which is a simple method to achieve higher accuracy in fault detection and is worth implementing.

Table 7 The accuracy of the model trained with modified cost function in unobserved conditions while training

Trained Speeds [Hz]					Default Features			
30	35	40	45	50	Train	Validation	Test	Total
	✓		✓	✓	95.8	79.5	58.1	**80.0**
	✓	✓		✓	100	98.7	63.8	**86.6**
✓		✓		✓	99.7	82.9	65.2	**84.7**
✓		✓	✓		99.0	82.4	63.0	**83.3**
✓	✓		✓		100	76.1	70.1	**85.7**
Average (%)					98.9	83.9	64.0	**84.1**

5- Industrial Cases

As stated in the Introduction, industrial data containing verified combined defects—meaning instances where equipment has undergone vibration data collection, troubleshooting in the condition monitoring process, and subsequent confirmation of identified defects upon equipment inspection—are seldom encountered. Due to the limited availability of industrial data pertaining to this scenario, laboratory PHM dataset has been utilized to develop the model. To demonstrate the efficacy of the proposed model on industrial cases, this section assesses the model's performance using two additional industrial datasets. It must be noted that before employing the proposed techniques in the article, the fault detection of these cases was unsuccessful based on the CNN model.

5-1- Pump Bearing Fault Detection

The first dataset pertains to a faulty bearing within a 37 kW pump at a power plant with a speed of 2965 rpm, wherein all components—namely, the inner race, outer race, and rolling element balls—exhibited damage. Fig. 6 illustreates the defective elements. It must be noted that this detection has been done based on the signals on only one side of the bearing in a single direction.

Fig. 6 Defective bearing detected by the proposed CNN model

5-2- Ball Mill Gearbox Fault Detection

The second industrial case involves the gearbox of a ball mill within a pelletizing factory with a motor speed of 995 rpm and gear ratio of 1:6.77. See the schematic in Fig. 7. The resulting outcomes from the proposed model accurately detect gear defects on the middle gearbox. It must be noted that diagnosed gear fault by the conidion monitoring team of this ball mill is worn gear.

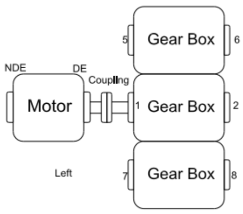

Fig. 7 Schematic of a ball mill driver with the defect on the middle gearbox detected by the proposed model.

It is praise to say that the defect identified by the condition monitoring team in the examined ball mill was the presence of a failure in the gear teeth, and the specific type of failure has not been determined. As is well known, accurately diagnosing the type of gear teeth failure through vibration analysis is a complex task in the industry. In many instances, the condition monitoring expert can only differentiate between gear tooth failure, gear misalignment, looseness, or lubrication issues. The model presented in this research has successfully identified failures associated with the gear. The diagnosis of the specific type of gear failure will be the focus of the authors' future research.

6- Summary/ Conclusion

This research has focused on improving fault detection models built with CNN architectures. Two challenges in developing them are providing a model for missing input data and predicting unobserved cases. New conditions and missing data can be attributed to measurements at different loads and rotor speeds with a sensor placed in only one location. The topic focused on in this research has been the simultaneous fault detection of gears, bearings, and shafts using the PHM 2009 challenge dataset. However, the strategies introduced here can be applied to similar problems with different datasets. The proposed tricks have been adjusting the input and cost functions in the training phase. Three input vibrational acceleration signals, including spectrum, envelope, and cepstrum, have been studied, and the highest accuracy has been obtained using the first one with an accuracy of approximately 3.4%, 1.8%, and 0.1% higher than others in gear, bearing, and shaft fault detection, respectively. By adjusting the input, the models have provided approximately the same level of detection accuracy at various loads and operating speeds for different sensor locations. Considering velocity spectrum as inputs instead of acceleration have resulted in 1.3% less precision. Entering both signalw (acceleration and velocity) into the model has led to 100% accuracy. By training the model with data from only three rotor speeds, the accuracy has been reduced to 79.6%. Then, the cost function of the CNN model has been modified to provide speed-insensitive features based on statistical analysis using p-value, and average accuracy has reached to 84.1%. Cost function adjustment has resulted in an accuracy increase of up to 13.6% for the PHM dataset. The proposed model has been used for two industrial cases helped to detect the generated faults. Future interest is in increasing the accuracy using other techniques such as transfer learning, especially when dealing with more industrial cases.

References

[1] X. Lyu, Z. Hu, H. Zhou, Q. Wang, Application of improved MCKD method based on QGA in planetary gear compound fault diagnosis, Measurement, 139 (2019) 236-248.

[2] H. Zhiyi, S. Haidong, Z. Xiang, Y. Yu, C. Junsheng, An intelligent fault diagnosis method for rotor-bearing system using small labeled infrared thermal images and enhanced CNN transferred from CAE, Advanced Engineering Informatics, 46 (2020) 101150.

[3] L. Xin, S. Haidong, J. Hongkai, X. Jiawei, Modified Gaussian convolutional deep belief network and infrared thermal imaging for intelligent fault diagnosis of rotor-bearing system under time-varying speeds, Structural Health Monitoring, 21(2) (2022) 339-353.

[4] Y. Xue, D. Dou, J. Yang, Multi-fault diagnosis of rotating machinery based on deep convolution neural network and support vector machine, Measurement, 156 (2020) 107571.

[5] P.J. García-Laencina, J.L. Sancho-Gómez, A.R. Figueiras-Vidal, Machine learning techniques for solving classification problems with missing input data, in: Proceedings of the 12th World Multi-Conference on Systems, Cybernetics and Informatics, 2008, pp. 1-6.

[6] Z. Zhang, F. Dong, Fault detection and diagnosis for missing data systems with a three time-slice dynamic Bayesian network approach, Chemometrics and Intelligent Laboratory Systems, 138 (2014) 30-40.

[7] K. Zhang, R. Gonzalez, B. Huang, G. Ji, Expectation–maximization approach to fault diagnosis with missing data, IEEE Transactions on Industrial Electronics, 62(2) (2014) 1231-1240.

[8] W. Liu, D. Wei, F. Zhou, Fault diagnosis based on deep learning subject to missing data, in: 2018 Chinese control and decision conference (CCDC), IEEE, 2018, pp. 3972-3977.

[9] S. Venkatasubramanian, S. Raja, V. Sumanth, J.N. Dwivedi, J. Sathiaparkavi, S. Modak, M.L. Kejela, Fault diagnosis using data fusion with ensemble deep learning technique in IIoT, Mathematical Problems in Engineering, 2022(1) (2022) 1682874.

[10] B. Zhang, W. Li, X.-L. Li, S.-K. Ng, Intelligent fault diagnosis under varying working conditions based on domain adaptive convolutional neural networks, Ieee Access, 6 (2018) 66367-66384.

[11] T. Han, C. Liu, W. Yang, D. Jiang, A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults, Knowledge-based systems, 165 (2019) 474-487.

[12] J. Shao, Z. Huang, Y. Zhu, J. Zhu, D. Fang, Rotating machinery fault diagnosis by deep adversarial transfer learning based on subdomain adaptation, Advances in Mechanical Engineering, 13(8) (2021) 16878140211040226.

[13] T. Han, C. Liu, W. Yang, D. Jiang, Learning transferable features in deep convolutional neural networks for diagnosing unseen machine conditions, ISA transactions, 93 (2019) 341-353.

[14] X. Li, Y. Hu, M. Li, J. Zheng, Fault diagnostics between different type of components: A transfer learning approach, Applied Soft Computing, 86 (2020) 105950.

[15] https://phmsociety.org/data-analysis-competition/, in, 2024.

[16] M. Ozeki, T. Okatani, Understanding convolutional neural networks in terms of category-level attributes, in: Computer Vision--ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part II 12, Springer, 2015, pp. 362-375.

[17] C.M. Bishop, Pattern recognition and machine learning, Springer google schola, 2 (2006) 1122-1128.