Mechanical Systems and Signal Processing 144 (2020) 106861

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Multisensor data fusion for gearbox fault diagnosis using 2-D convolutional neural network and motor current signature analysis

Moslem Azamfar [a,*], Jaskaran Singh [a], Inaki Bravo-Imaz [b], Jay Lee [a]

[a] NSF I|UCRC for Intelligent Maintenance System (IMS), University of Cincinnatti, Cincinnatti, Ohio, USA
[b] Intelligent Information System, IK4-TEKNIKER, Eibar 20600, Spain

ARTICLE INFO

Article history:
Received 27 July 2019
Received in revised form 21 February 2020
Accepted 29 March 2020
Available online 6 April 2020

Keywords:
Gearbox faults
Motor current signature analysis
Deep learning
Sensor fusion
Fault detection and diagnosis
Support vector machine
k-Nearest neighbors
Linear discriminant analysis
Naïve Bayes
Decision Tree

ABSTRACT

Gearboxes are widely used in rotating machinery and various industrial applications for transmission of power and torque. They operate for prolong hours and under different working conditions which may increase their probability of failure. Sudden failure of a gearbox may lead to significant downtime and increase maintenance costs. In industrial applications, usually fault detection and diagnosis techniques based on vibration signal are used for monitoring the health condition of gearboxes. In most of these techniques, time and frequency domain features are manually extracted from a vibration sensor and used for fault detection and diagnosis. In this research, a fault diagnosis methodology based on motor current signature analysis is proposed. The acquired data from multiple current sensors are fused by a novel 2-D convolutional neural network architecture and used for classification purpose directly without any need for manual feature extraction. Performance of the proposed method has been evaluated on the motor current data obtained from an industrial gearbox test rig in various health condition and with different working speeds. In comparison with classical machine learning (ML) algorithms, the presented methodology exhibits the best classification performance for gearbox fault detection and diagnosis.

1. Introduction

Gearboxes are frequently used in mechanical systems for power transmission between shafts and their reliability and high performance is critical for many industrial applications. Due to faults in important gears, serious financial and safety consequences may happen. Fault detection and diagnosis in the early stage may prevent unplanned downtime, expensive repair, damage to operational performance, and even personal casualties [1]. Therefore, it is essential to accurately detect early faults that may occur in gearboxes and take maintenance actions accordingly.

Health condition of a gearbox can be reflected in different measurements such as vibratory, electrical, thermal, acoustic, and oil-based data. Among them, vibration signal has been mostly employed because any fault in the gearbox changes the normal spectrum of the signal. Effectiveness of vibration signal in Fault Detection and Diagnosis (FDD) of gearboxes has been

* Corresponding author.
E-mail addresses: azamfamm@mail.uc.edu (M. Azamfar), singh2jn@ucmail.uc.edu (J. Singh), inaki.bravo@tekniker.es (I. Bravo-Imaz), lj2@ucmail.uc.edu (J. Lee).

https://doi.org/10.1016/j.ymssp.2020.106861
0888-3270/

studied in many references [2–5]. Sun et al. [2] proposed a customized multiwavelets approach for planetary gearbox fault diagnosis and reported satisfactory detection of incipient pitting faults. In [3] a diagnosis method is proposed to handle incomplete information due to sensor malfunction, communication latency, etc. which could significantly impact the fault diagnosis study on a gearbox. Qian et al. [4] proposed a fault diagnosis method based on vibration data to address the shift variant problem. Singh et al. [5] proposed a cross domain fault diagnosis method based on raw vibration data and for gearbox operating under variable speeds.

However, most of the current studies on gearbox fault diagnosis require fault features to be extracted manually based on domain knowledge and expertise about the system under consideration. In addition, majority of the methodologies don't consider the integration of information across multiple measurements. Moreover, special training is required to properly install vibration sensors, they are expensive, the measurement signal is highly dependent on the position of sensors, and their output can easily be affected by interfering noise from the surrounding components.

Motor current signal is the input source of many electromechanical machines which can be easily accessed by using low-cost sensors. Consequently, Motor Current Signature Analysis (MCSA) has been successfully applied for FDD within electrical motors such as for broken rotor bars, air-gap eccentricity, winding, and motor bearing faults [6,7]. However, a limited study has been done on usefulness of MCSA in FDD of downstream mechanisms (e.g. gearboxes) in an electromechanical machine. Due to the impact of the supply line frequency, raw time domain data cannot be directly used for FDD [8]. However, studies conducted by [9,10] shows that mechanical faults might be detectable from the frequency spectrum of the signal based on the amplitude of the supply line frequency and certain fault characteristic sidebands. In [11], a methodology based on discrete wavelet decomposition is applied on the motor current signal obtained from a transient working condition for FDD on a gearbox. In [12], a methodology based on bispectrum analysis on MCSA has been proposed to detect wear faults in a spur gear. In [13], a method based on data envelopment and resonance-based sparse signal decomposition has been proposed for spur gear fault detection. A current space vector approach was developed by [14] for fault detection in tooth surface of a spur gear. These methods show the importance and usefulness of MCSA for FDD in gearboxes. However, further researches and studies are required to improve FDD results.

Among several FDD algorithms that have been proposed for gearboxes, Support Vector Machine (SVM) [15], k-Nearest Neighbors (kNN) [16], Discriminant Analysis (DA), Naïve Bayes (NB), Decision Tree (DT) [17] and Artificial Neural Network (ANN) [18] have been employed frequently. In these approaches, time or frequency domain features extracted from the measurement signal are used for classification. These methods require domain knowledge and expertise for feature extraction and usually have a limited capacity in data fusion from multiple resources [19].

Consequently, deep learning approaches are gaining popularity due to their automatic feature learning from raw data and high performance in fusing signals from multiple resources [20]. In the field of bearing FDD, Convolutional Neural Networks [21], Deep Belief Networks (DBN) [22], Autoencoder [23,24], Deep neural networks (DNN)[24], and Long Short Term Memory (LSTM) [25] have been employed and their performance has been evaluated in some references. In the case of gearbox FDD, some limited studies on using CNN [26] and DBN [27] have been conducted. In these references, vibration signature has been used for FDD. To the best of the authors' knowledge, there is no study on applying deep learning techniques on motor current signal for gearboxes FDD.

In order to improve classification accuracy, data from multiple sensors might be fused to create more representative information. In comparison with a single sensor data, multisensor signals always have the redundant and complementary information which could be used for a better fault diagnosis [28]. A multisensor feature-level fusion for bearing fault classification based on DBN-SAE concept and using vibration signal has been proposed in [28]. In [19] a feature-level fusion technique based on SAE and DNN has been proposed for rotating machinery. A sensor fusion method for rolling element bearing based on acoustic emission and acceleration signals and kNN classifier has been developed in [29]. Kumar et al. [30] proposed a feature level fusion for gearbox application where a series of hand crafted features are used in a fusion process to enhance the diagnosis performance. These methods mostly perform feature level data fusion on vibration and acoustic signals and use manual feature extraction which requires high-level domain knowledge and expertise. In the case of the current signal, manual feature extraction is more challenging and require more expertise. Therefore, it is necessary to explore multisensor automatic data fusion techniques that don't require manual feature engineering and can provide reliable and accurate classification results.

In this paper, a data-level multisensor fusion approach based on a 2D-CNN model is proposed for automatic feature learning from raw frequency spectrum obtained from current sensors. The fused features then are used in a softmax layer for classification. The proposed method requires no manual feature extraction and the experimental results show the superiority of this method compared to some of the well-established classification algorithms. The major contributions of this paper include the following:

(1) Effectiveness of MCSA in FDD of downstream mechanical anomalies in a two-stage gearbox running at multiple speeds is investigated and analytical models are presented for feature extraction from frequncy doamin data.
(2) A data-level multisensor fusion method based on 2D-CNN is proposed and its effectiveness in FDD of a two-stage gearbox is evaluated. In the proposed method, the information lost is minimized and the correlation between different sensors is preserved by applying different convolution filters on raw data stacked in a 2D matrix.
(3) Effectiveness of different network structure and tuning parameters in the proposed CNN-based FDD model is discussed.

(4) A comparison between automatic feature learning and the effectiveness of hand-crafted features has been carried out on the motor current signal collected from a gearbox prognostic simulator test rig.

The rest of the paper is organized as follows. In Section 2, the theoretical background for the identification of gear faults in the frequency spectrum of the motor current signal is provided. The proposed multisensor data-level fusion approach is discussed in Section 3. The proposed method is applied for FDD of a two-stage gearbox and a comparison between the achieved results and performance of other techniques is given in Section 4. Finally, the conclusions are drawn in Section 5.

2. Fault diagnosis using motor current signature analysis (MCSA)

The rotor in an electrical motor is designed to have a uniform air gap all around it with the stator of the motor. A mechanical anomaly within the downstream mechanical components driven by the attached motor is supposed to generate unexpected oscillatory motion of the gearbox shaft both rotationally and radially. Such a motion would have a direct influence on the air gap eccentricity which affects the air gap flux density. The stator will draw current according to the variation in the electromagnetic field. Since the variations in motor current are related to motion in downstream components; using motor current measurements seem to be a reasonable choice for fault detection and diagnosis in a gearbox.

In case of vibration measurements, gear faults and variations of gear contact stiffness will lead to torsional vibrations of the gearbox when it rotates and thus, the torque signature of the gearbox output shaft will contain the rotating and meshing frequency components [31].Therefore, a fault in the load part of the drive is seen from the induction machine by a variation of the load torque and these torque oscillations make the stator current modulated as well [32]. The load torque observed by an induction machine coupled to a gearbox consists of a mean value and a time-varying part [10]:

$$T_L(t) = T_0(t) + T_{osc}(t) \tag{1}$$

where T_0 is the average torque or torque when there is no torsional vibration and $T_{osc}(t)$ is the load torque oscillation. The $T_{osc}(t)$ part is related to (a) pinion and wheel rotations along with (b) mesh and mesh-related frequency components:

$$T_{osc}(t) = \textit{torque oscillation related to both gear pinion and wheel rotations} + \textit{mesh related frequency componenets}$$

$$T_{osc}(t) = \sum_{n=1}^{N} T_n.cos(\omega_{rn}t + \varphi_{rn}) + \sum_{p=1}^{P} T_{mesh,p}.cos\left(\omega_{mesh,p}t + \varphi_{mesh,p}\right) \tag{2}$$

where T_n and φ_n are the torque amplitude and phase of the oscillatory component due to the torsional vibration induced by the transmission error in the input and output wheels due to the shaft rotation; $T_{mesh,p}$ and $\varphi_{mesh,p}$ are the amplitude and phase of the oscillatory component due to the torsional vibration induced by stiffness variation of the gear teeth contact; f_n and $f_{mesh,p}$ are the input and output gear rotating and mesh frequencies, respectively. f_{rn} (n = 1 , 2, . . ., N; where N is the number of shafts in the gearbox) represents the input and output gear rotating frequencies, $f_{mesh,p}$ (p = 1 , 2, . . ., P, where P is the number of gear pairs) represents the gear meshing frequencies.

$$\begin{aligned} &f_{r1} = \tfrac{\omega_{r1}}{2\pi}, f_{r2} = \tfrac{\omega_{r2}}{2\pi} \cdots\cdots f_{rn} = \tfrac{\omega_{rn}}{2\pi} \\ &f_{mesh,1} = \tfrac{\omega_{mesh,1}}{2\pi}, f_{mesh,2} = \tfrac{\omega_{mesh,2}}{2\pi} \cdots\cdots f_{mesh,p} = \tfrac{\omega_{mesh,p}}{2\pi} \\ &f_{mesh,1} = N_{r1}f_{r1} = N_{r2}f_{r2}; f_{mesh,2} = N_{r3}f_{r2} = N_{r4}f_{r3} \cdots\cdots f_{mesh,p} = N_{r(2P-1)}f_{r(P)} = N_{r(2P)}f_{r(P+1)} \end{aligned} \tag{3}$$

where f_{r1}, f_{r2} and f_{rn} represents the rotational frequency of shaft 1, 2 and N; $f_{mesh,1}, f_{mesh,2}$and $f_{mesh,P}$ represents the meshing frequency between gear pairs 1, 2 and P. Therefore Eq. (1) can be written as:

$$T_L(t) = T_0(t) + \sum_{n=1}^{N} T_n.cos(\omega_{rn}t + \varphi_{rn}) + \sum_{p=1}^{P} T_{mesh,p}.cos\left(\omega_{mesh,p}t + \varphi_{mesh,p}\right) \tag{4}$$

Using Eq. (1) and knowing that in the steady-state condition the average electromagnetic torque is equal to T_0, the machine mechanical equation relates the torque oscillation to the motor speed as follows [33]:

$$\sum T(t) = T_e(t) - T_L(t) = J\frac{d\omega_r}{dt} = J\frac{d^2\theta_r}{dt^2} \tag{5}$$

where T_e is the electromagnetic torque produced by the synchronous machine; J is the total inertia of the machine and the load and θ_r is the mechanical rotor position. Double Integration of Eq. (5) provides:

$$\theta_r(t) = \int_{t_o}^{t} \omega_r(\tau)\tau = \frac{1}{p}\chi(t) + \omega_{ro}t + \theta_{ro} \tag{6}$$

where

$$\chi(t) = \sum_{n=1}^{N} \beta_{rn}.cos(\omega_{rn}t + \varphi_{rn}) + \sum_{p=1}^{P} \beta_{mesh,p}.cos\left(\omega_{mesh,p}t + \varphi_{mesh,p}\right)$$
$$\beta_{rn} = \frac{pT_n}{J{\omega_{rn}}^2}, \beta_{mesh,p} = \frac{pT_{mesh,p}}{J{\omega_{mesh,p}}^2}, \omega_{ro} = \frac{(1-s)}{p}\omega_s \tag{7}$$

The stator current in an arbitrary phase can be expressed in a general form [10,33]:

$$\begin{aligned} I(t) &= I_{st}sin(\omega_s t + \varnothing_s) + I_{rt}sin(\omega_s t + \chi(t) + \varnothing_r) \\ &+ \left[\sum_{n=1}^{N} \omega_{rn}.\beta_{rn}.\left\{I_{rn}cos((\omega_s + \omega_{rn})t + \chi(t) + \varnothing_{rn} + \varnothing_r) - {I_{rn}}'cos((\omega_s - \omega_{rn})t + \chi(t) - \varnothing_{rn} + \varnothing_r)\right\}\right. \\ &\left.+ \sum_{p=1}^{P} \omega_{mesh,p}\beta_{mesh,p}.\left\{I_{mesh,p}cos\left((\omega_s + \omega_{mesh,p})t + \chi(t) + \varphi_{mesh,p} + \varnothing_r\right) - {I_{mesh,p}}'cos\left((\omega_s - \omega_{mesh,p})t + \chi(t) - \varphi_{mesh,p} + \varnothing_r\right)\right\}\right] \end{aligned} \tag{8}$$

where the first term results from the stator magneto-motive force (MMF) and it is not modulated and the second term is a direct consequence of the rotor MMF shows the phase modulation due to the considered load torque oscillations. Since the first term is dependent on the supply frequency, it can be ignored. Generally, for physically reasonable values of J, T_n, $T_{mesh,p}$, ω_{rn}, and $\omega_{mesh,p}$ the approximation of $\beta_{rn} \ll 1$ and $\beta_{mesh,p} \ll 1$ holds in most cases. Also, since $\beta_{r1} > \beta_{r2} > ..\beta_{rn} \gg \beta_{mesh,p}$ as $\omega_{mesh,p} \gg \omega_{rn} > ..\omega_{r2} > \omega_{r1}$. This would suppress the contributions of the mesh and mesh related frequencies $f_{mesh,1}.f_{mesh,2}\cdots f_{mesh,p}$ and renders them insignificant for gearbox fault diagnosis. Therefore, the phase modulation part of the second term can be written as per equation Eq. (7):

$$\begin{aligned} &I(t) \approx I_{rt}sin(\omega_s t + \varnothing_r)\chi(t) \\ &I(t) \approx I_{rt}\sum_{k=1}^{+\infty}\frac{\beta_{rn}}{2}sin((\omega_s - \omega_{rn})t + \varnothing_r) + I_{rt}\sum_{k=1}^{+\infty}\frac{\beta_{rn}}{2}sin((\omega_s + \omega_{rn})t + \varnothing_r) + I_{rt}\sum_{k=1}^{+\infty}\frac{\beta_{mesh,p}}{2}sin((\omega_s - \omega_{rn})t + \varnothing_r) \\ &\quad + I_{rt}\sum_{k=1}^{+\infty}\frac{\beta_{mesh,p}}{2}sin((\omega_s + \omega_{rn})t + \varnothing_r) \end{aligned} \tag{9}$$

Thus, the characteristic frequencies of gearbox torsional vibrations appear as sidebands $f_{s\_gearbox}$ across the fundamental frequency of the stator current of the electric machine and can be expressed as follows:

$$f_{s\_gearbox} = f_s(t) \pm m \times f_{rn}(t) \quad and \quad f_{s\_gearbox} = f_s(t) \pm m \times f_{mesh,p} \quad \text{with} \quad m = 1, 2, 3, \cdots \tag{10}$$

As mentioned above, $\beta_{mesh,p} \ll 1$ suppress the contributions of the mesh and mesh-related frequencies $f_{mesh,1}.f_{mesh,2}\cdots f_{mesh,p}$ and renders them insignificant for gearbox fault diagnosis. In addition, it is worth noting that torsional vibrations occur even for the case of a healthy gearbox. However, the amplitude of the excited sidebands would be insignificant in comparison to a gearbox with a fault. Moreover, a fault may even excite additional sidebands across the fundamental frequency of the stator current.

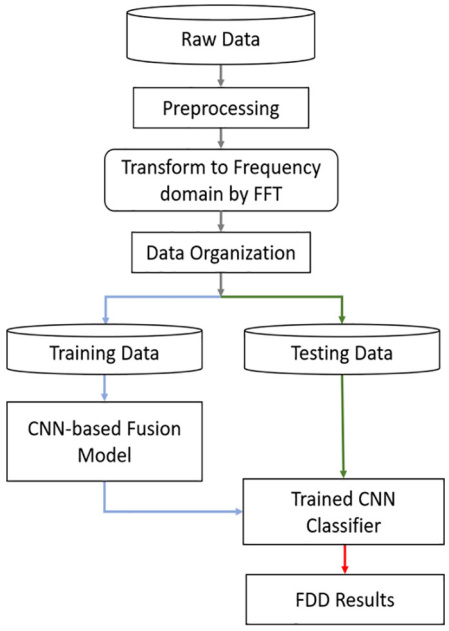


Fig. 1. The flowchart of the proposed FDD approach.

3. Multisensor fault diagnosis based on CNN

Fig. 1 shows the flowchart of the proposed FDD approach. Multiple measurement signals are collected from the electric motor running a gearbox. The input signals are preprocessed for removing noise and outliers and then transformed to the frequency domain by FFT technique and then the frequency domain data is divided into training and testing portions. For each portion, frequency spectrum of each sensor forms a set of one-dimensional data that are stacked row by row to generate a 2-D matrix. A matrix calculated from training dataset is used for training the proposed CNN model. The testing dataset then is used to evaluate the performance of the trained model. No manual feature extraction is required and the representative features are extracted automatically during the training process are used for classification.

In this study, an intelligent fault diagnosis method for rotating machinery is proposed based on CNN with the raw frequency data from multiple current sensors being used as an input to the FDD model. The proposed method is able to learn features automatically from the raw FFT signal and no hand-crafted features are required to perform classification on different health conditions. By combining the raw frequency spectrum from multiple sensors into a 2-D matrix as the input to the CNN model, a multisensor fusion structure is created. The main advantages of the proposed fusion method are summarized below:

(1) Data and feature level fusion: Data from multiple sensors are stacked row by row to form a 2D input matrix. As the convolutional filters move across the input matrix, information from multiple sensors are fused together. Therefore, the lost of information has been minimized in the fusion process by applying the convolution filter on raw data. In addition, as the network is trained, features extracted from multiple sensors are convoluted constantly in different layers of the network to generate a higher-level representation of features which then is used in the softmax layer for classification. In this process information extracted from multiple sensors are further fused to generate more meaningful representation of the input data.
(2) Preserving the dependency between different sensors: Stacking data from multiple sensors in a row by row format would preserve the time dependency between different sensors which could be very valuable for discovering the correlation between different sensors and thus extraction of more information for fault diagnosis.

The proposed CNN model is shown in Fig. 2. It consists of multiple convolutional and pooling layers followed with multiple fully connected layers and finally, a softmax layer is used to create the output classes. The convolutional and pooling layers capture the representative features from the raw frequency data without any need for manual feature extraction. Network's weight matrix and bias are the learnable parameters of the CNN structure and stochastic gradient descent (SGD) is used to update these parameters during the training process. The proposed multi-sensor fusion structure could be implemented with different numbers of convolutional, pooling, and fully connected layers. In this research, we have studied the effectiveness of different configurations and tuning parameters on the classification performance. Finally, as the model is trained, it can be tested on the new data samples.

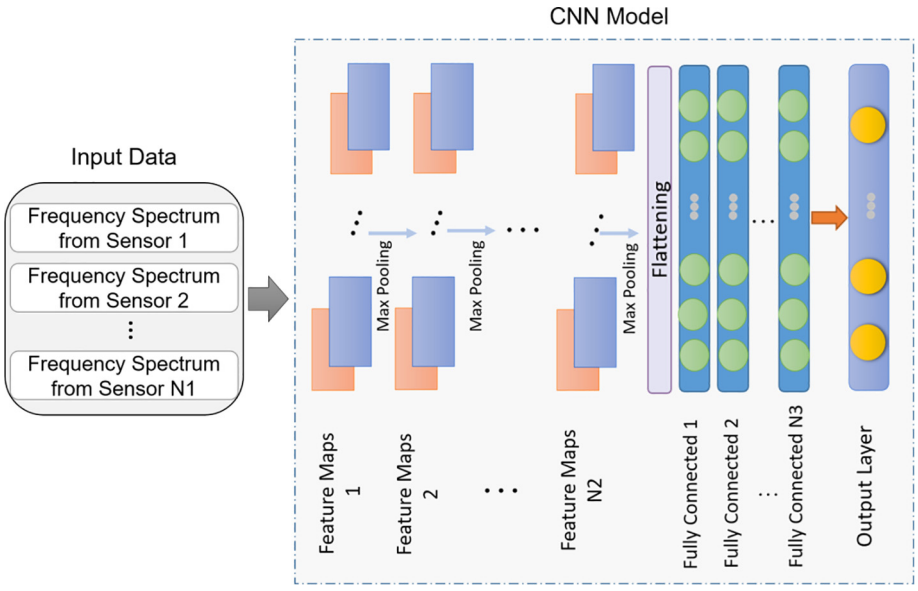


Fig. 2. The proposed CNN-based data fusion architecture where N1 = numbers of sensors, N2 = numbers of convolutional and pooling pairs, N3 = numbers of fully connected layers and the output layer is based on a softmax function.

4. Experimental evaluation

Having a large number of convolutional and pooling layers will create a deep structure for the deep learning model which usually leads to higher accuracy and more repeatable results. However, a deeper structure will increase the computation time, network complexity, and the difficulty to build and evaluate the effectiveness of different layers and parameters [34]. In this research, a flexible structure has been proposed where a stack of convolutional and pooling layers is followed with a number of fully connected layers. The best design for the network and tuning parameters might change for different applications. Here we have used an experimental dataset to evaluate the performance of different network and tuning parameters and finally, the configuration with the best performance is selected as the final model.

An experimental dataset has been used for evaluation of the proposed FDD method. Where the performance of the designed CNN-based fusion method in feature learning and classification is evaluated with different input type including raw frequency data and hand-crafted features. The proposed method is compared with several well-established methods including SVM, KNN, NB, DT, and DA. Fig. 3 shows the flowchart of all tested methods where the proposed method is marked in green.

4.1. Experimental setup and data description

The gearbox prognostic simulator from the Spectra Quest Company [35] has been used for this experiment, which is shown in Fig. 4. The test rig is composed of two confronted electrical motors, one acting as drive and the other as load, and two gearboxes, complemented with sensors and couplings. The drive and load motors have the same characteristics. Both are 10 Hp (7.35 kW), three-phase, induction motors having two pair of poles. The monitored electrical signals are the ones feeding the driving motor. In our configuration, the first gearbox after the drive motor is the one under test, in

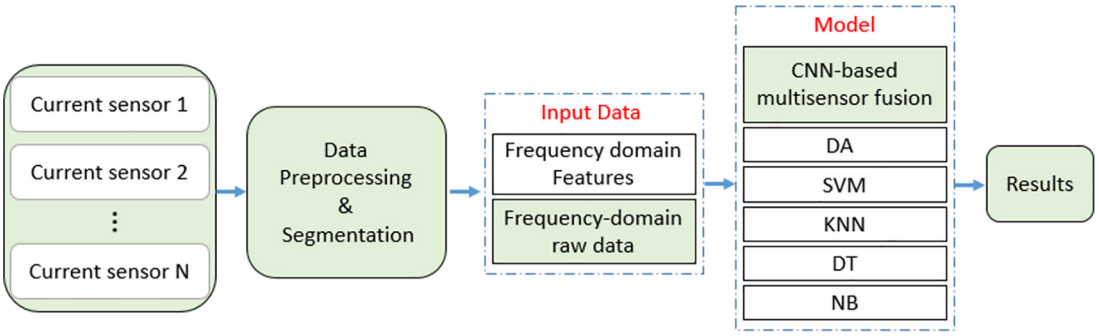


Fig. 3. The proposed FDD method is compared with classical ML strategies where the proposed workflow is shown in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

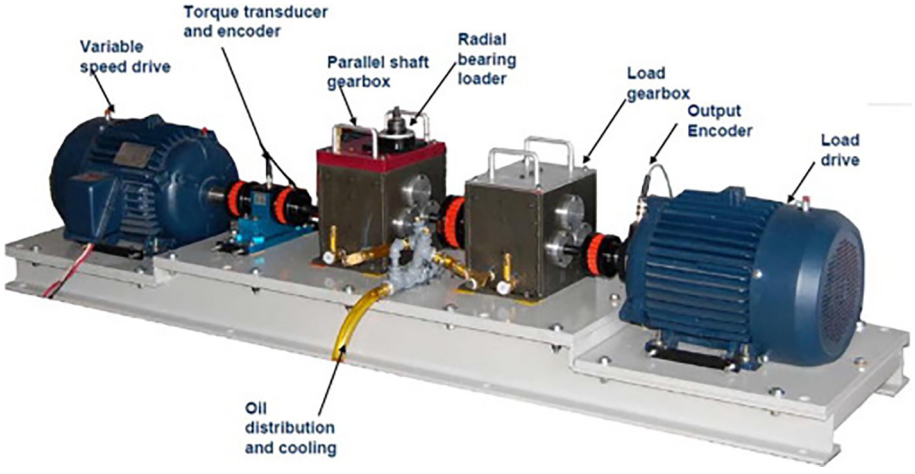


Fig. 4. The gearbox prognostic simulator test rig (Figure from http://spectraquest.com/, accessed on January 2016).

which gears in different health states are inserted. The second gearbox works as a reducer and loads motor protector in this set-up.

The monitored gearbox is composed of four spur gears (Fig. 5). The first gear, as it comes from the motor that drives the test bench, has 32 teeth. It is the one substituted by gears in different health state, leaving the rest unchanged. It is followed by a gear with 80 teeth. In the same axle, a gear with 48 teeth is found, connected to a gear with 64 teeth, resulting in a global transmission relationship of 3.33. Fig. 6 shows the seven health conditions examined in this paper.

The feeding currents are measured by LEM HTA 100 S [36] sensors. There are two sensors installed in the electrical cabinet, for the measurement of U and W electrical lines. The studied motor is an equilibrated motor, so U and W are equal signals, with a difference in phase of 120°. The data was recorded using a computer with a National Instruments acquisition card (NI 4472 series).

For this experiment, data is collected from different speeds at zero-load condition. The maximum speed that the test rig can reach is 1500 rpm. The minimum selected speed is 250 rpm. Two more intermediary speeds were selected, 500 and 1000 rpm. Each test condition is repeated 15 times to enable statistical robustness. Each repetition is independent to the rest as, between two repetitions, the speed is set to zero before the next test was launched. The pitting of the second gear was generated from intensive use in the gearbox. The eccentricity, the missing tooth, the chipped tooth, the low and high severity abrasion faults were intentionally produced by machining the gear. Table 1 summarized the collected data in this experiment.

4.2. Data preprocessing

There are different methods to detect and remove outliers from the signal, however, in our study we used a simple and robust method proposed in [37] for removing outliers from each data channel. Accordingly, an outlier is a value that is more than three scaled median absolute deviations (MAD) away from the median. The scaled MAD is defined as

$$MAD_S = C \times median(abs(A - median(A))) \tag{11}$$

where A is the input data and $C = \frac{1}{sqrt(2)\times erfcinv(3/2)}$ and erfcinv is the inverse complementary error function defined as inverfc (y)= {x : erfc(x) = y}[38]. In order to remove noise from the signal, a second-order Butterworth low-pass filter with cut-off frequency of 500 Hz was applied to the current signal after removing outliers.

Raw vibration data may have significant information on the components fault characteristics and therefore the extracted time domain features such as RMS, Kurtosis, and Peak to peak have been widely used in the mechanical components fault diagnosis. Due to the appearance of the fault characteristics in the raw vibration data, it is expected that deep learning-based methods such as CNN would provide satisfactory diagnosis results when raw data without any manual feature extraction or advanced data preprocessing is utilized. On the other hand, due to the high impact of supply line frequency (or the running frequency), the time domain features of the stator current signal would have less importance and hardly represent the failure characteristics of the downstream components. Therefore, unlike vibrations data, using raw motor current signal is not expected to provide satisfactory fault diagnosis results. Our analytical study presented in Section 2 clarifies that fault characteristics may appear as increase in the amplitudes of FFT in sidebands around the main frequency. A comparison between raw current data and frequency domain signals for all seven health conditions are given in Figs. 7 and 8 respectively. As

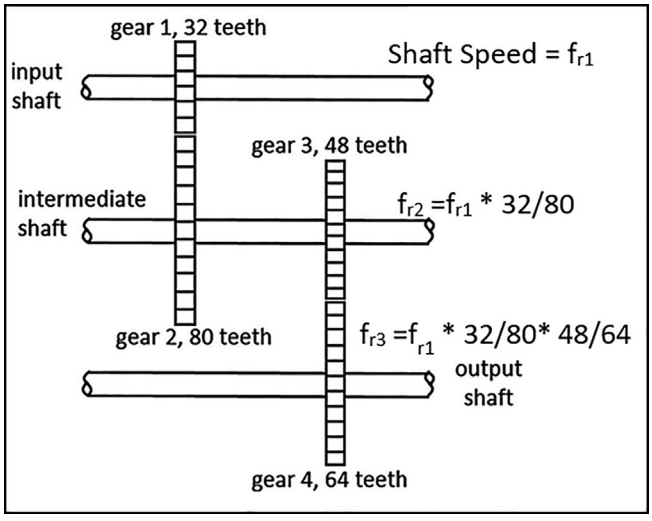


Fig. 5. Scheme of the gear disposition inside the monitored gearbox.

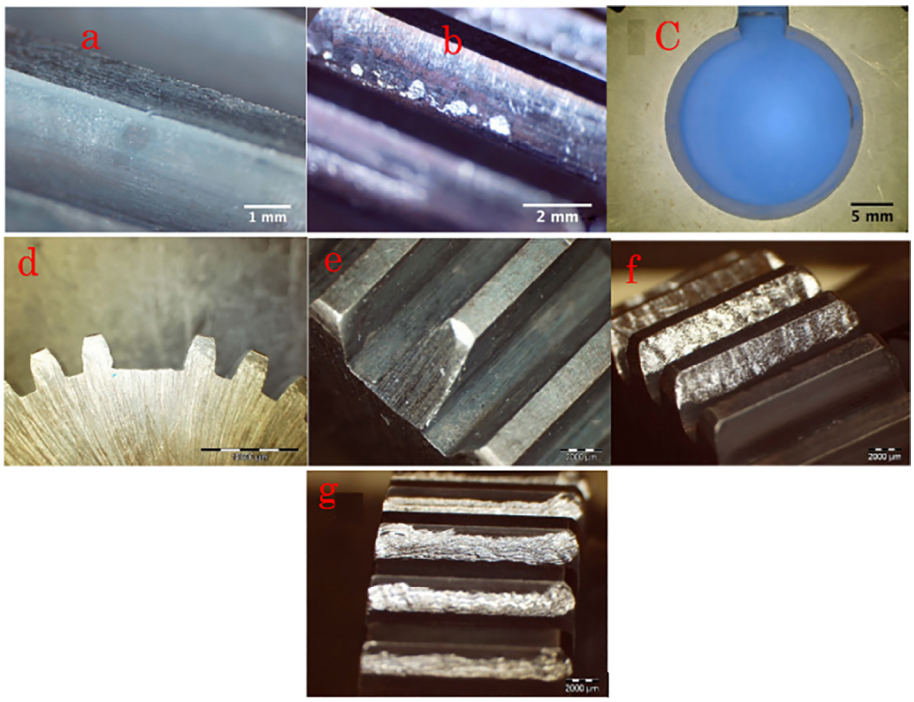


Fig. 6. Images representative of the health condition of the gears used in this experiment (a) healthy gear. (b) gear with pitting. (c) gear with eccentricity (d) missing tooth. (e) chipped tooth. (f) abrasion at tooth of gear (low severity). (g) abrasion at tooth of gear (high severity).

Table 1
Summary of the collected data from the experiment.

Class labels	Gear Condition	Input Speed (rpm)	Sampling rate (Hz)	Load
0	Healthy	250/500/1000/1500	50,000	Zero
1	Pitting	250/500/1000/1500	50,000	Zero
2	Eccentricity	250/500/1000/1500	50,000	Zero
3	Missing Tooth	250/500/1000/1500	50,000	Zero
4	Chipped Tooth	250/500/1000/1500	50,000	Zero
5	Low Severity Abrasion	250/500/1000/1500	50,000	Zero
6	High Severity Abrasion	250/500/1000/1500	50,000	Zero

depicted, due to the presence of running frequency, the fault characteristics cannot be detected in the raw time domain data. However, there are clear differences in the FFT spectrums of different signals. Thus, after preparing samples, FFT of the signal is obtained and directly used in the CNN model for automatic feature learning and fault diagnosis.

In many deep leading methods, the size of the training dataset might be artificially increased to improve the classification accuracy. In this study, after removing outlier and noise from the signal, a data augmentation technique [1] is being used to increase the number of training samples. This approach has been explained in Fig. 9 wherein a fixed window size of the training data is chosen as the first training sample and the second training sample is created by shifting the fixed window across the data. This process is repeated for many times to create a sufficient amount of training samples. For example, a current signal with a length of 10,000 points can provide 8001 training samples if the window size is 2000 and shift size is 1 point. This data augmentation method has been applied on raw time domain data and then for each obtained window, the frequency domain transformation has been extracted by the FFT technique. In order to get a high-resolution frequency domain signal, the window size has been chosen 100,000 points.

4.3. Results and discussion

In this section, two experimental studies have been conducted. In the first experiment, only raw frequency domain signals are used for classification of 7 different working conditions discussed in Section 4.1. In the second experiment, hand-crafted features from frequency domain data are extracted by the methodology presented in Section 2. Then the extracted features are used in different classification techniques and their performance are compared. For each experiment, a comprehensive study on the performance of the proposed method and some of the well stablished classification techniques is presented. The effect of the training size and working speed on the classification performance has been studied in each experiment.

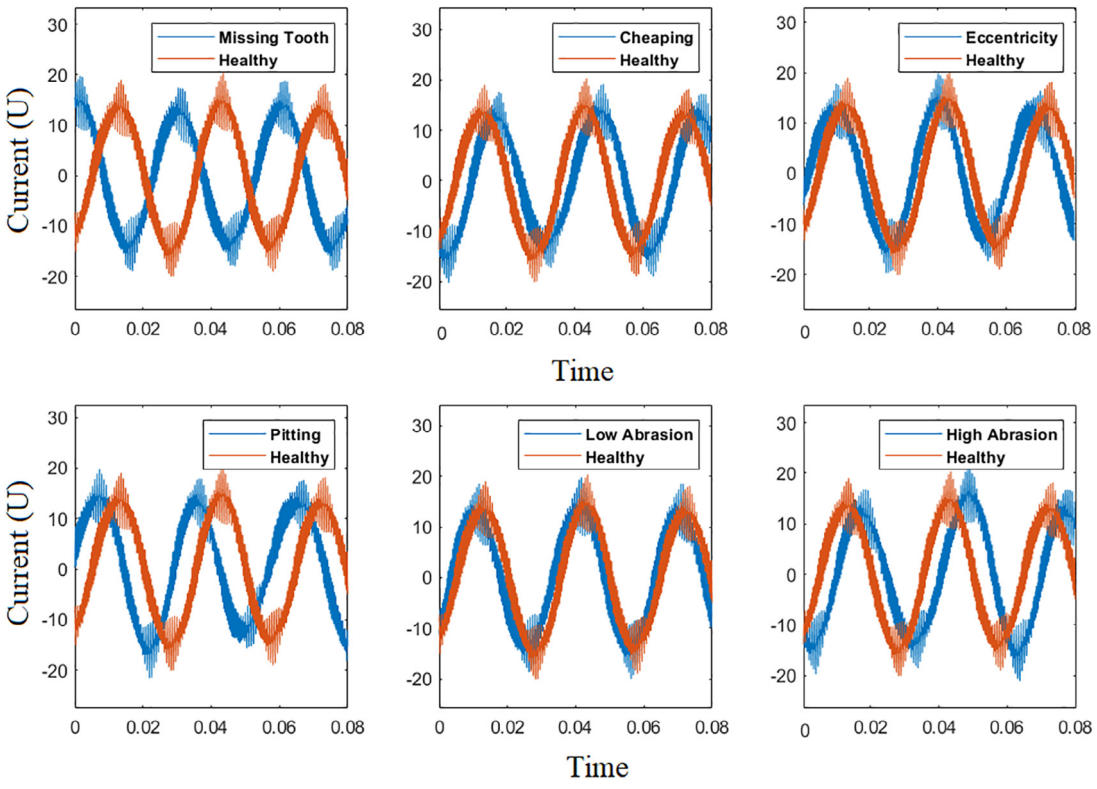


Fig. 7. comparison between raw current signals (Phase U) collected from a gearbox in seven health conditions including Healthy, Pitting, Chipped, Eccentric, Missing, Abrasion High, and Abrasion Low.

4.3.1. Experiment 1: Automatic feature learning for classification

In this section, raw frequency domain data is used for the evaluation of the proposed CNN-based data fusion method. Different structure for the CNN network along with their key parameters, include the size of filters, the size of data segments, the number of filters of convolutional layer and the number of nodes in fully-connected layers, are investigated. For each test, 20 trials are performed to avoid particularity and contingency in the fault diagnosis results and then the average and standard deviation across these 20 runs are calculated. This study is conducted in Python 3.6.5, 64 bit running on Windows 10 with RTX 2080 GPU. Keras public library is used for deep learning analysis. The achieved results of CNN with various configurations and tuning parameters are shown in Fig. 10 where Table 2 summarizes different studied configurations and their tuning parameters. As depicted in Fig. 10, with an increase in the number of layers the overall FDD performance improves. However, adding more layers and increasing the number of neurons in the fully connected layer may lead to overfitting and drastically increase the training time. Considering a trade-off between different factors, the configuration consisted of two pairs of convolutional and pooling layers with two fully-connected layers followed with a softmax layer is selected as the final model (model #3) The parameters of the selected model is given in Table 3.

Different classical ML methods such as Support vector machine (SVM), k-Nearest Neighbors (kNN), Discriminant Analysis (DA), Naïve Bayes (NB), and Decision Tree (DT) are used for comparison. Each ML method has a set of hyparameters that need to be carefully optimized to provide the best calssification results. In the optimization process, an objective function is designed and a group of hyperparameters are obtained such that the objective function is minimized.

In our study, Matlab Statistics and Machine Learning Toolbox™ was used to automatically optimize the hyperparameters of different methods. A list of hyperparameters that can be tuned for each model is given in Table 4 and for finding the most optimal selection of hyperparameters, bayesian optimization has been utilized. This method internally maintains a Gaussian process model of the objective function and evaluates its performance through the training process. In order to simplify the optimization problem, a set of Fit functions are applied to find hyperparameters that minimize k-fold cross-validation loss automatically. In the k-fold cross validation process, the data is randomly partitioned into k sets and a model is trained on 'k-1' sets and cross validated on the remaining one set of data. The overall optimization process can be summarized below:

a) Partitioning the data into training and test dataset. A model is trained and optimized on the training dataset and then evaluated on the test dataset.
b) Selection of hyperparameters optimization options for the Fit functions. Table 5 summarizes the selected parameters in our study.

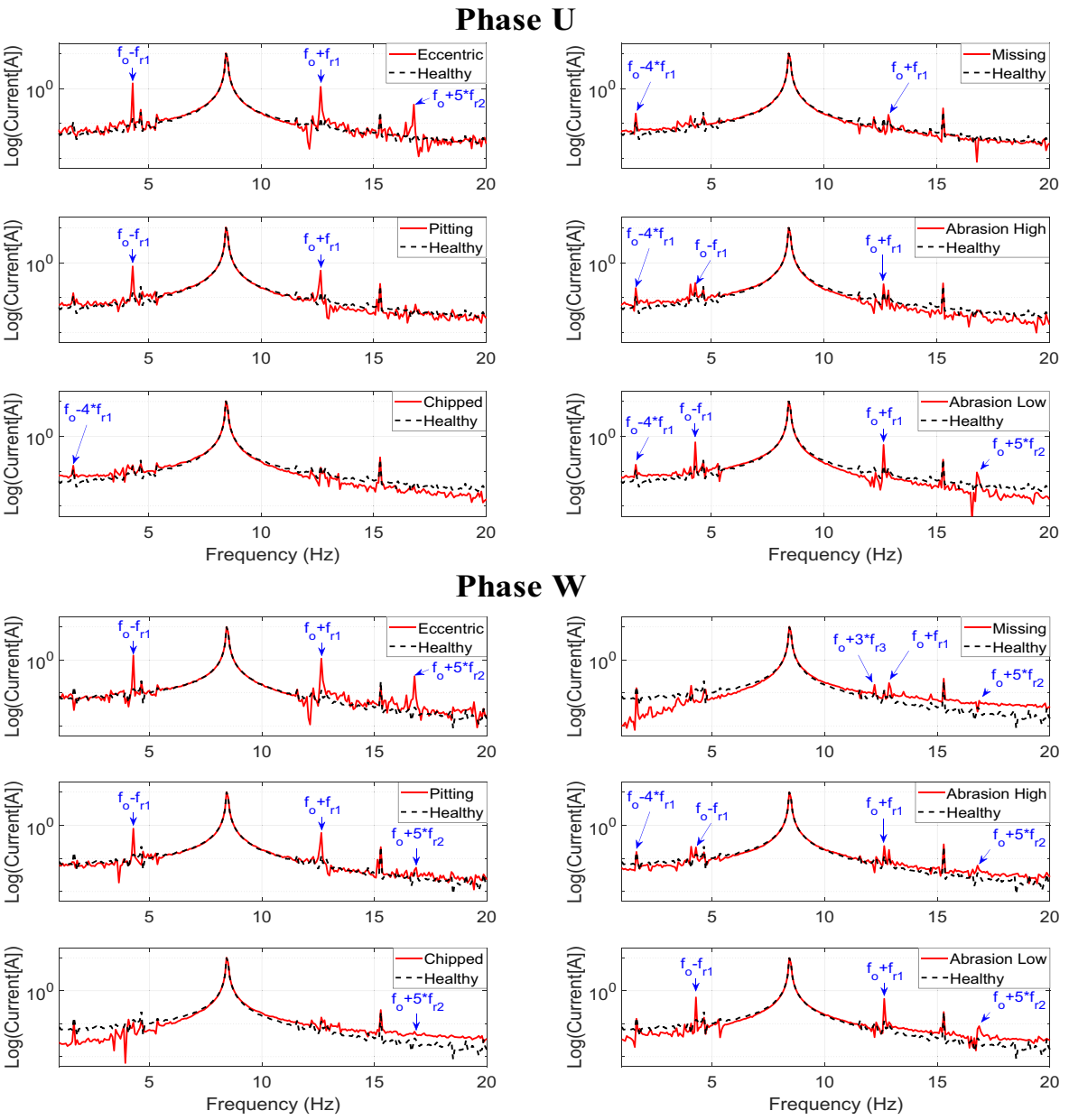


Fig. 8. A comparison between spectra of the motor current signatures obtained from the gearbox in seven steady state conditions including Healthy, Pitting, Chipped, Eccentric, Missing, Abrasion High, and Abrasion Low.

c) Selection of the hyperparameters to be optimized. In this study, all hyperparameters presented in Table 4 were considered during the model optimization process.
d) Train the optimization model and evaluate the results. If required, the whole process is repeated with new settings to guarantee the suitability of results.

Fig. 11 shows the classification error function during the optimization process on the training dataset with 10-fold cross validation. As depicted, different hyperparameters are used during the training process and for each selected values the classification error is obtained and minimized through the optimization process and finally the most optimum results are reported. The trained model finally is used on a new dataset (test dataset) and the achieved results are reported.

Comparison between the results of ML methods and the CNN based methods is given in Table 6. Accordingly, CNN-based fault diagnosis methods outperform all classical ML methods when raw frequency domain data are used as an input without manual feature extraction. The results also clearly show the superiority of the proposed CNN based fusion method in terms

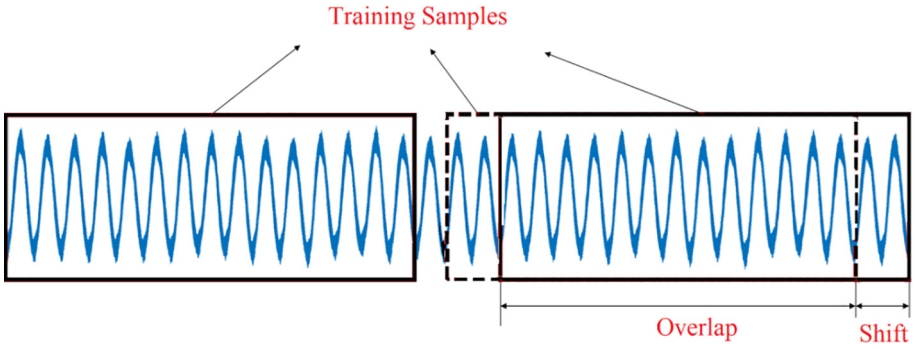


Fig. 9. Data augment with overlap.

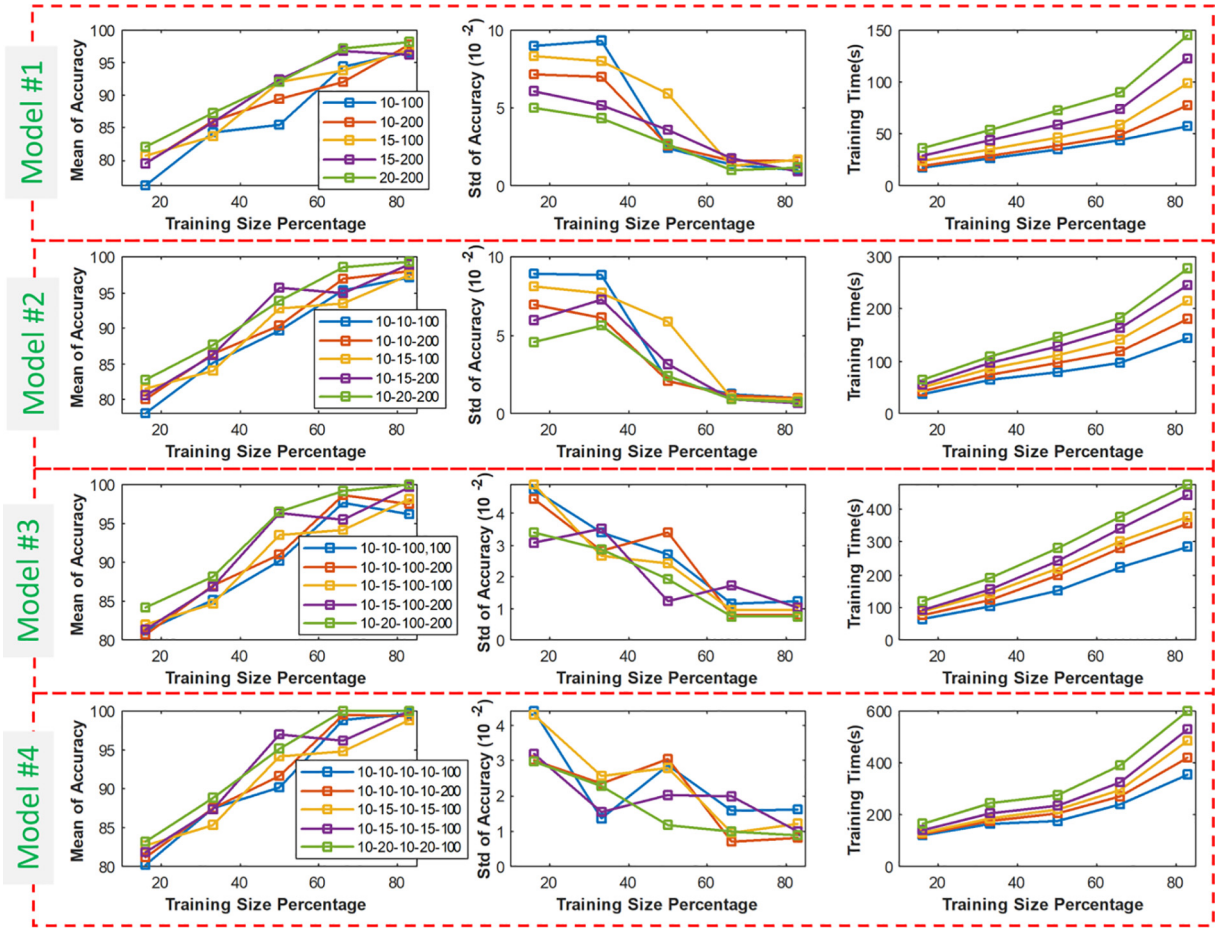


Fig. 10. Results of the proposed CNN structures with different configurations and tuning parameters. The tested configuration and parameters are summarized in Table 2 Legend of each figure has the same format as given for tuning parameters in Table 3 (E.g. in model #1, NF Conv#1– NN Full#1 is given as (10–100)/(10–200)/(15–100)/(15–200)/(20–200)).

of repeatability of the results (low standard deviation), sensitivity to training size, and classification accuracy. The achieved results also indicate that using data from each sensor individually in a typical CNN structure also leads to higher performance than all other classical ML methods. Among different classical methods, SVM exhibits better overall performance and have the highest classification accuracy in comparison to NB, DA, DT, and kNN respectively.

In order to evaluate the performance of the proposed fusion method in different working speeds, a study has been performed at zero-load condition and with 5 different running speeds. The achieved results for the proposed strategy and other

Table 2
Different structures tested with several tuning parameters. The studied 'Tuning parameters' are given as a legend for each model in Fig. 10.

Model	Description	Tuning Parameters
#1	Conv #1 + Pool #1 + Full #1 + Soft #1	NF Conv#1– NN Full#1
#2	Conv #1 + Pool #1 + Conv #2 + Pool #2 + Full #1 + Soft #1	NF Conv#1– NF Conv#2- NN Full#1
#3	Conv #1 + Pool #1 + Conv #2 + Pool #2 + Full #1 + Full #2 + Soft #1	NF Conv#1– NF Conv#2- NN Full#1- NN Full#1
#4	Conv #1 + Conv #2 + Pool #1 + Conv #3 + Conv #4 + Pool #2 + Full #1 + Soft #1	NF Conv#1– NF Conv#2– NF Conv#3 – NF Conv#4 – NN Full#1

Where Conv = Convolutional layer, Pool = Pooling layer, Full = Fully connected layer, Soft = Softmax layer, NF = Number of Filters, and NN = Number of Nodes. Therefore NF Conv means Number of Filters in a fully connected layer and NN Full means the number of nodes in a fully connected layer.

Table 3
Parameters of the selected CNN model.

Layers Type	Model Description
Convolution #1	NF = 10, SF = [2,15,1], S = 1, DR = 0.4, P = Valid
Pooling #1	PS = 8, S = 1
Convolution #2	NF = 20, SF = [1,8,10], S = 1, DR = 0.4, P = Valid
Pooling #2	PS = 4, S = 1
Fully connected #1	NN = 100, Activation = 'ReLU'
Fully connected# 2	NN = 200, Activation = 'ReLU'
Output (Fully connected# 3)	7 class output. Activation = Softmax

NF: Number of Filters, SF: Size of Filter, DR: Dropout Ratio, S: Stride, P: Padding, PS: Pooling Size, NN: Number of Nodes. Maximum epoch is 100 and the learning rate is set to 0.03.

Table 4
Hyperparameters tunable for each classification method.

Classifier	Hyperparameters
Decision Tree	• Maximum number of splits – search among integers log-scaled in the range [1,max(2,n–1)], where n is the number of observations. • Split criterion - search among Gini's diversity index, Twoing rule, and Maximum deviance reduction.
Discriminant Analysis	• Discriminant Type: search among Linear, Quadratic, Diagonal Linear, and Diagonal Quadratic.
Naive Bayes	• Distribution Type – Gaussian and Kernel. • Kernel type – search among Gaussioan, nBox, Epanechnikov, and Triangle.
SVM	• Kernel function – search among Gaussian, Linear, Quadratic, and Cubic. • Box constraint level – search among positive values log-scaled in the range [0.001,1000]. • Kernel scale – search among positive values log-scaled in the range [0.001,1000]. • Multiclass method – search between One-vs-One and One-vs-All. • Standardize data – search between true and false.
KNN	• Number of neighbors – search among integers log-scaled in the range [1,max(2,round(n/2))], where n is the number of observations. • Distance metric - search among Euclidean, City block, Chebyshev, Minkowski (cubic), Mahalanobis, Cosine, Correlation, Spearman, Hamming, Jaccard. • Distance weight – search among Equal, Inverse, and Squared inverse. • Standardize – search between true and false.

Table 5
Hyperparameters optimization options for the Fit functions.

Field Name	Selected Value
Optimizer	Bayesian optimization
Acquisition Function Name	Expected-improvement-per-second-plus
Maximum Objective Evaluations	30
Number of Grid Divisions	10
Cross Validation	k-fold

methods are given in Table 7. As is shown, the proposed CNN-based fusion method significantly outperforms other methods in all tested running speeds. CNN based classification techniques that just use one sensor as the input also show very high performance and the highest accuracy after the proposed fusion technique.

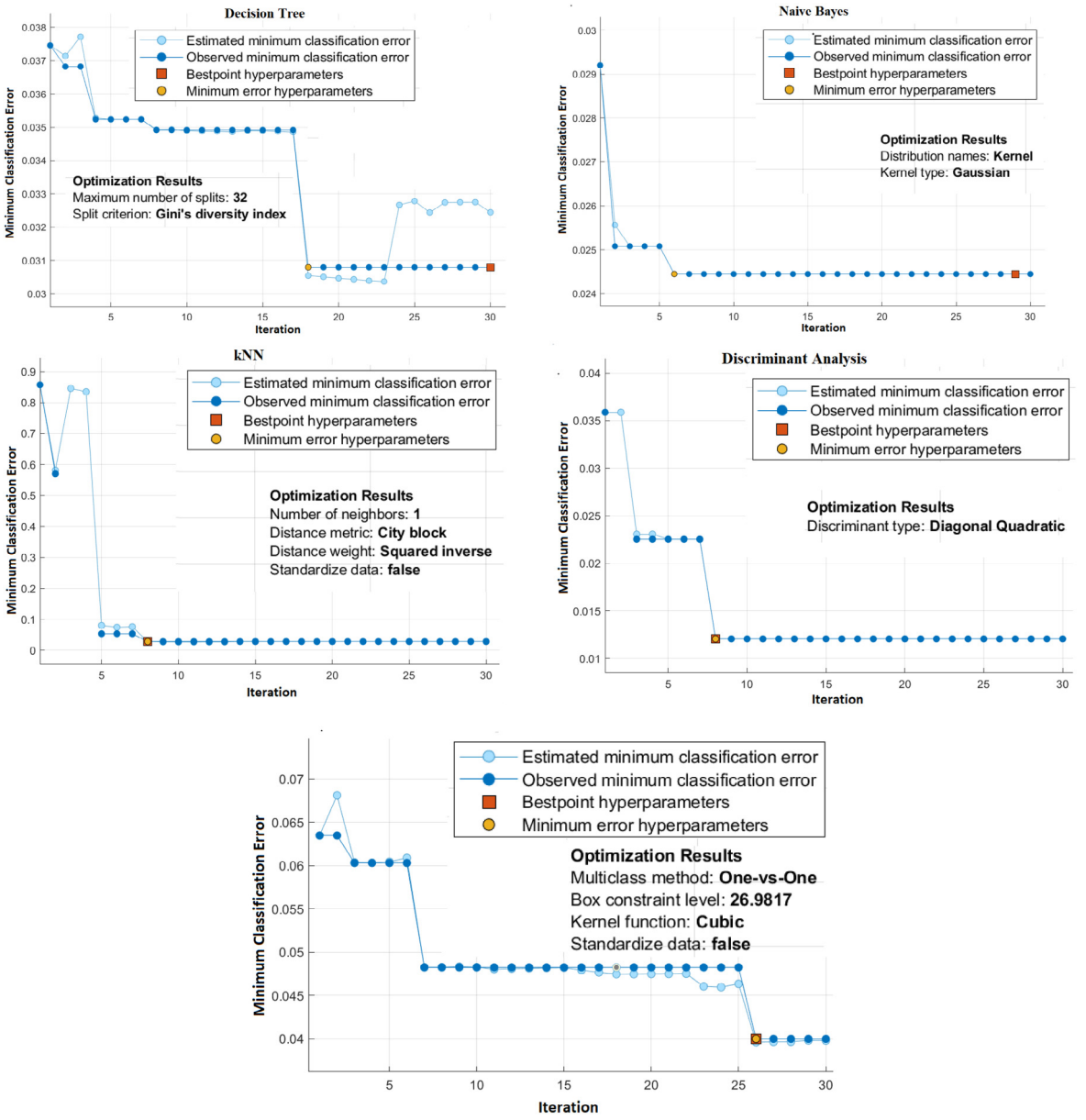


Fig. 11. Optimization process for different ML methods with Matlab Statistics and Machine Learning Toolbox™.

Table 6
Testing accuracy of different classification methods with different training size obtained from raw FFT data for speed 1000 RPM.

Method	Training Size				
	16.66%	33.33%	50%	66.66%	83.33%
kNN (U/W/UW)	61.77/60.36/61.20	67.45/62.78/65.57	76.05/68.58/69.91	77.07/74.99/75.36	78.40/76.46/75.87
NB (U/W/UW)	71.89/63.88/69.58	77.85/61.58/70.62	88.50/76.70/81.61	89.94/87.11/87.30	90.03/83.48/86.56
DT (U/W/UW)	57.76/57.32/57.07	63.48/57.97/58.69	87.27/86.45/86.29	88.38/82.90/85.65	89.68/83.71/84.80
SVM (U/W/UW)	73.78/65.34/72.86	78.83/68.58/76.59	88.74/88.11/87.79	89.61/86.14/88.97	90.08/87.95/89.23
DA (U/W/UW)	73.74/69.92/70.38	78.89/70.23/73.12	88.82/86.10/85.03	89.48/88.49/82.20	90.01/83.98/84.69
CNN (U)	79.56 ± 0.033	83.01 ± 0.022	92.82 ± 0.041	97.77 ± 0.002	98.82 ± 0.002
CNN (W)	80.12 ± 0.028	85.82 ± 0.037	95.54 ± 0.009	98.11 ± 0.003	99.21 ± 0.004
CNN-Proposed (UW)	**82.23 ± 0.018**	**85.89 ± 0.014**	**96.64 ± 0.011**	**98.23 ± 0.002**	**100 ± 0.0**

The format of the results for CNN is: average testing accuracy(%) ± standard deviation

Table 7
Performance of classification methods in different working speed with raw frequency data as inputs to the models.

Method	Speed			
	250 RPM	500 RPM	1000 RPM	1500 RPM
kNN (U/W/UW)	79.70/75.50/73.24	74.58/76.91/77.12	78.40/76.46/75.87	69.20/66.84/68.37
NB (U/W/UW)	84.00/79.06/90.42	91.18/80.44/86.14	90.03/83.48/86.56	77.84/76.43/81.55
DT (U/W/UW)	91.30/82.53/86.00	89.76/79.83/83.26	89.68/83.71/84.80	81.44/75.94/71.51
SVM (U/W/UW)	84.12/87.60/89.85	95.42/89.45/93.61	90.08/87.95/89.23	81.51/81.61/83.98
DA (U/W/UW)	86.20/85.22/87.15	88.42/86.17/80.65	90.01/83.98/84.69	77.86/77.72/76.54
CNN (U)	98.63 ± 0.04	97.22 ± 0.016	98.82 ± 0.002	93.54 ± 0.032
CNN (W)	95.71 ± 0.02	98.02 ± 0.021	99.21 ± 0.004	94.43 ± 0.074
CNN-Proposed (UW)	**98.83 ± 0.015**	**98.79 ± 0.005**	**100 ± 0.00**	**94.58 ± 0.029**

The format of the results for CNN is: average testing accuracy(%) ± standard deviation

4.3.2. Experiment 2: Hand-crafted features for classification

As depicted in Fig. 3, raw FFT signals and frequency domain hand-crafted features both can be used as input to the classification methods. In this section, the effectiveness of manual feature extraction on different classification methods is studied. Based on the analysis conducted in Section 2 and Eq. (10), the spectral location of the supply line frequency along with the sidebands which would be excited due to a mechanical fault within the gearbox is given in Table 8. In order to comprehensively capture the diagnostic information, the first five sidebands corresponding to the shaft rotational speeds (3 shafts: two-stage gearbox) have been considered. In overall, 26 frequency domain features are extracted from every FFT acquired from an individual current sensor. A comparison between the FFT of different classes of data is shown in Fig. 8. Clearly, amplitudes of signals at identified sidebands indicate differences in two classes of data and therefore could be used in classification models. Extraction of these sidebands require high-level domain knowledge and expertise for understanding the physical characteristic of the system thereby compromises the application of methods that require manual feature extraction.

The motor current signatures of the gearbox (Phase U and Phase W) in seven conditions for steady state at an input shaft operating speed of 250 rpm, within the frequency range 0–20 Hz is shown in Fig. 8. Because the motor in the gearbox prognostic simulator test rig has two pair of poles, in order to convert the rotation speeds to frequency, the speeds must be multiplied by two, apart from being translated into Hertz. Therefore, the spectrums exhibit the main peak at 8.33 Hz i.e. at the fundamental line frequency (Number of poles * Rotational Speed (Hz) = 2*4.16 Hz) in addition to many sidebands which are expected to be related to the different shaft rotational speeds. In the experimental study, Gear 1 (on the input shaft) is substituted with gears with different health conditions and therefore fault impact on the input and intermediate shafts spectrums are expected to be more clearly detectable compared to the fault signature modulated on the output shaft because of their proximity to the faulty gear.

It must be noted, that even in case of the frequency spectrum of healthy signals, there are some sidebands with appreciable amplitude at specific fault locations that have been identified in Table 8. The presence of these sidebands with appreciable amplitude highlights the fact that how sensitive the motor current signals are to the changes in the gear transmission condition and demonstrates the frequency modulation caused by the gear transmission. However, presence of a specific fault causes a significant increase in the amplitude of these sidebands (relative to the healthy gear) along with the generation of more sidebands. It needs to be observed that the amplitude and location of sidebands (associated with fault) is not consistent in spectrums of both phases of motor current (Phase U and Phase W). This experimental observation confirms that different

Table 8
Calculated fundamental sidebands and their harmonics for different running speeds. The amplitudes of FFT at these frequencies are used as input features for classification methods.

Lower Sidebands	Speed (rpm)				Upper Sidebands	Speed (rpm)			
	250	500	1000	1500		250	500	1000	1500
f_0–f_{r3}	424.8	849.6	1699.8	2550	f_0+f_{r3}	574.8	1149.6	2299.8	3450
f_0–2*f_{r3}	349.8	699.6	1399.8	2100	f_0+2*f_{r3}	649.8	1299.6	2599.8	3900
f_0–3*f_{r3}	274.8	549.6	1099.8	1650	f_0+3*f_{r3}	724.8	1449.6	2899.8	4350
f_0–4*f_{r3}	199.8	399.6	799.8	1200	f_0+4*f_{r3}	799.8	1599.6	3199.8	4800
f_0–5*f_{r3}	124.8	249.6	499.8	750	f_0+5*f_{r3}	874.8	1749.6	3499.8	5250
f_0–f_{r2}	399.6	799.8	1599.6	2400	$f_0 + f_{r2}$	600	1200	2400	3600
f_0–2*f_{r3}	300	600	1200	1800	$f_0 + 2 * f_{r2}$	699.6	1399.8	2799.6	4200
$f_0 - 3 * f_{r2}$	199.8	400.2	799.8	1200	$f_0 + 3 * f_{r2}$	799.8	1599.6	3199.8	4800
$f_0 - 4 * f_{r2}$	99.6	199.8	400.2	600	$f_0 + 4 * f_{r2}$	900	1800	3600	5400
f_0	499.2	999.6	1999.2	3000	$f_0 + 5 * f_{r2}$	999.6	1999.8	3999.6	6000
$f_0 - f_{r1}$	249.6	499.8	999.6	1500	$f_0 + f_{r1}$	750	1500	3000	4500
$f_0 + 2 * f_{r1}$	999.6	1999.8	3998.4	6000	$f_0 + 3 * f_{r1}$	1248	2499	4998	7500
$f_0 + 4 * f_{r1}$	1497.6	2998.8	5997.6	9000	$f_0 + 5 * f_{r1}$	1747.2	3487.8	6997.2	10,500

sensors may contain different partial information about the same machine condition. Using information from only one of the sensors might result in providing low fault classification accuracy. Therefore, it makes logical sense to merge the features extracted from each of these sensors into one stream.

A comparison between the results of different methods when hand-crafted features are used as an input is given in Table 9. As is shown, compared to using raw frequency spectrum, when hand-crafted features are used the overall accuracy of classical ML methods increases significantly. SVM and DA show the highest overall accuracy for different training size and different combinations of input signals. Performance of the proposed method is still very good and in many cases outperforms most of the other classifiers. In all tested methods with an increase in the training size the overall performance improves. Performance of CNN method with only one sensor is still compatible with other methods.

The performance of each method in different working speeds with zero-load has been evaluated and the results are given in Table 10 As we were expecting, compared with using raw frequency spectrum, by using hand-crafted features the overall accuracy of all classical ML methods has increased. Classification accuracies based on CNN have been slightly reduced however, the results are still acceptable and even in many cases outperform the classical methods.

Fig. 12 shows the confusion matrix obtained from the proposed method for both experimental studies with raw frequency data and hand-crafted features. Description of data label and supportive information is given in Table 1. In both studies, testing data size is 732 samples. The actual labels of data are shown in the y-axis and the predicted labels are given in the x-axis. Overall prediction accuracy for each actual class is shown in the last column from right and overall prediction accuracy for each predicted class is shown in the first row from the bottom. As is shown, for the proposed method the results achieved with using raw frequency data is significantly better than using hand-crafted features.

4.3.3. Summary of the analytic study

Authors have performed, two experimental studies in order to evaluate the significance of their approach. In experiment 1, raw frequency domain signal is used for the fault diagnosis purpose. The achieved results highlight that the proposed method outperforms all of the classical ML methods in terms of classification accuracy, robustness to the training data size and variations in working speeds. In experiment 2, a set of hand crafted features are used as input to the classical ML methods and the proposed CNN-based fusion methodology. The results indicate that with careful selection of the hand-crafted features, even the classical ML methods can show satisfactory performance. However, the biggest challenge is in the feature engineering process which requires advanced domain knowledge about the system.

The overall comparison between two experiments is summarized in the Table 11. By comparing the results from the two experiments, we also observe that better diagnosis results are achieved by the proposed method when raw frequency domain data is utilized. The key advantage of the proposed method is that it eliminates the need for manual feature engineering and also outperforms the classical ML methods that require advanced feature engineering.

Table 9
Testing Accuracy of different classification methods with different training size obtained from handcraft features as input and for speed 1000 RPM.

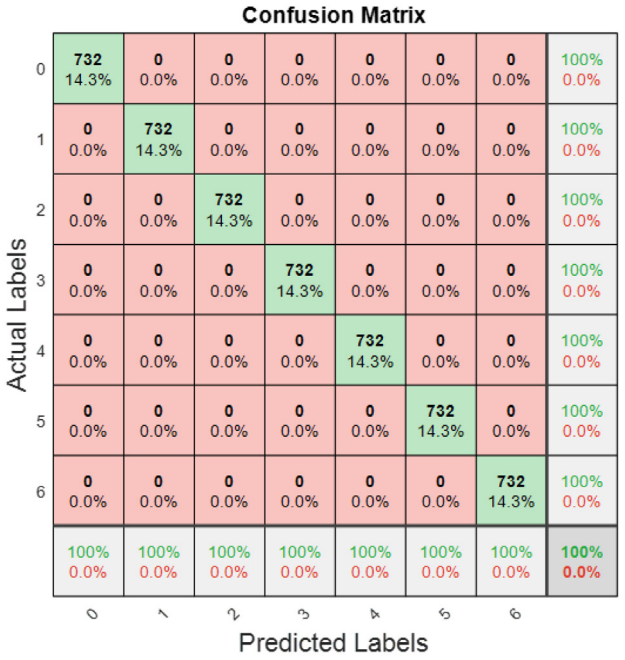
Method	Training Size				
	16.66%	33.33%	50%	66.66%	83.33%
kNN (U/W/UW)	78.15/77.06/78.08	81.43/80.09/83.64	89.39/87.58/89.17	96.76/95.66/96.23	95.12/97.61/96.78
NB (U/W/UW)	77.75/70.91/77.16	77.78/68.32/78.46	93.60/85.17/90.53	93.99/96.70/96.85	94.11/92.74/95.01
DT (U/W/UW)	62.66/63.55/63.33	64.45/64.40/65.18	89.90/95.97/95.80	94.40/92.09/95.23	91.87/92.86/94.12
SVM (U/W/UW)	**79.09**/72.60/78.65	85.37/76.13/85.01	92.51/**97.80**/97.40	95.03/95.66/94.33	97.64/97.70/**98.51**
DA (U/W/UW)	78.63/77.65/78.09	81.77/77.96/81.21	92.32/95.65/94.34	89.86/**98.12**/91.32	91.48/93.16/94.00
CNN (U)	78.51 ± 0.045	85.52 ± 0.033	94.15 ± 0.021	98.11 ± 0.024	97.62 ± 0.016
CNN (W)	77.11 ± 0.054	84.72 ± 0.042	95.13 ± 0.036	97.03 ± 0.023	96.45 ± 0.012
CNN-Proposed (UW)	78.63 ± 0.034	**85.76 ± 0.029**	95.58 ± 0.019	97.83 ± 0.013	98.40 ± 0.02

The format of the results for CNN is: average testing accuracy(%) ± standard deviation

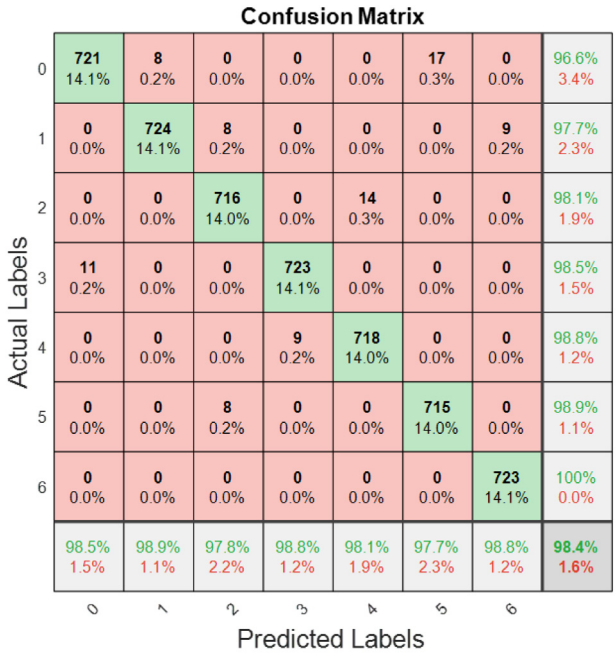
Table 10
Performance of classification methods in different working speed with hand-crafted frequency domain features as inputs to each method.

Method	Speed			
	250 RPM	500 RPM	1000 RPM	1500 RPM
kNN (U/W/UW)	95.97/96.83/97.71	93.74/94.76/95.41	95.12/97.61/96.78	89.73/87.28/90.59
NB (U/W/UW)	94.00/90.05/96.57	91.86/95.07/96.81	94.11/92.74/95.01	85.38/85.98/89.96
DT (U/W/UW)	87.58/87.23/92.95	92.13/89.87/95.03	91.87/92.86/94.12	83.63/93.18/91.09
SVM (U/W/UW)	96.95/97.02/**98.10**	**98.94**/96.57/98.45	97.64/97.70/**98.51**	90.84/91.06/92.91
DA (U/W/UW)	93.31/93.92/92.80	88.54/94.07/89.65	91.48/93.16/94.00	88.55/93.10/92.16
CNN (U)	95.34 ± 0.03	97.17 ± 0.04	97.62 ± 0.016	92.27 ± 0.043
CNN (W)	95.15 ± 0.07	96.95 ± 0.06	96.45 ± 0.012	93.14 ± 0.061
CNN-Proposed (UW)	96.27 ± 0.04	96.81 ± 0.012	98.40 ± 0.02	**93.48 ± 0.03**

The format of the results for CNN is: average testing accuracy(%) ± standard deviation



(a)



(b)

Fig. 12. The confusion matrix using the proposed method for speed 1000 rpm and using 83.3% of data for training - (a) using frequency domain raw data as an input (b) using hand-crafted features as an input.

5. Conclusion

This paper proposes a novel multisensor data fusion methodology based on 2-D convolutional neural network and for gearboxes FDD using MCSA. Multiple motor current sensors are used to monitor and diagnose the fault types in a gearbox under different working conditions. Based on the proposed framework, raw frequency domain data from multiple sensors

Table 11
A comparison between results achieved from experiments 1 and 2.

	Method	Speed			
		250 RPM	500 RPM	1000 RPM	1500 RPM
Experiment #2	kNN (U/W/UW)	95.97/96.83/97.71	93.74/94.76/95.41	95.12/97.61/96.78	89.73/87.28/90.59
	NB (U/W/UW)	94.00/90.05/96.57	91.86/95.07/96.81	94.11/92.74/95.01	85.38/85.98/89.96
	DT (U/W/UW)	87.58/87.23/92.95	92.13/89.87/95.03	91.87/92.86/94.12	83.63/93.18/91.09
	SVM (U/W/UW)	96.95/97.02/98.10	**98.94**/96.57/98.45	97.64/97.70/98.51	90.84/91.06/92.91
	DA (U/W/UW)	93.31/93.92/92.80	88.54/94.07/89.65	91.48/93.16/94.00	88.55/93.10/92.16
Experiment #1	CNN-Proposed (UW)	**98.83 ± 0.015**	**98.79 ± 0.005**	**100 ± 0.00**	**94.58 ± 0.029**

The format of the results for CNN is: average testing accuracy(%) ± standard deviation.

are stacked together to create a matrix as an input to the 2-D CNN architecture. Several design structure with different tuning parameters were tested to select the best structure of convolutional, pooling, fully connected, and softmax layers. Then the performance of the selected design was evaluated in two different experimental studies. Results achieved from the proposed method has been compared with some of the well-known ML methods. The proposed method exhibits the best performance in terms of classification accuracy, robustness to the training data size and variations in working speeds.

CRediT authorship contribution statement

Moslem Azamfar: Conceptualization, Methodology, Software, Visualization, Writing - original draft. **Jaskaran Singh:** Methodology, Investigation, Visualization. **Inaki Bravo-Imaz:** Methodology, Visualization, Investigation. **Jay Lee:** Methodology, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] W. Zhang, C. Li, G. Peng, Y. Chen, Z. Zhang, A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load, Mech. Syst. Signal Process. 100 (2018) 439–453, https://doi.org/10.1016/j.ymssp.2017.06.022.
[2] H. Sun, Y. Zi, Z. He, J. Yuan, X. Wang, L. Chen, Customized multiwavelets for planetary gearbox fault detection based on vibration sensor signals, Sensors. 13 (2013) 1183–1209, https://doi.org/10.3390/s130101183.
[3] J. Yu, Y. He, Planetary gearbox fault diagnosis based on data-driven valued characteristic multigranulation model with incomplete diagnostic information, J. Sound Vib. 429 (2018) 63–77, https://doi.org/10.1016/j.jsv.2018.05.020.
[4] W. Qian, S. Li, J. Wang, Z. An, X. Jiang, An intelligent fault diagnosis framework for raw vibration signals: Adaptive overlapping convolutional neural network, Meas. Sci. Technol. 29 (2018), https://doi.org/10.1088/1361-6501/aad101.
[5] J. Singh, M. Azamfar, A.N. Ainapure, J. Lee, Deep learning based cross domain adaptation for gearbox fault diagnosis under variable speed conditions, Meas. Sci. Technol. (2019), https://doi.org/10.1088/1361-6501/ab64aa.
[6] M.S.R. Krishna, K.S. Ravi, Fault diagnosis of induction motor using Motor Current Signature Analysis, in: Proc. IEEE Int. Conf. Circuit, Power Comput. Technol. ICCPCT 2013, 2013, pp. 180–186, https://doi.org/10.1109/ICCPCT.2013.6528849.
[7] J. Lee, M. Azamfar, J. Singh, A blockchain enabled cyber-physical system architecture for industry 4.0 manufacturing, Systems, Manuf. Lett. (2019), https://doi.org/10.1016/j.mfglet.2019.05.003.
[8] F. Li, X. Pang, Z. Yang, Motor current signal analysis using deep neural networks for planetary gear fault diagnosis, Meas. J. Int. Meas. Confed. 145 (2019) 45–54, https://doi.org/10.1016/j.measurement.2019.05.074.
[9] W. Thomson, On-line current monitoring to detect electrical and mechanical faults in three-phase induction motor drives, in: 2005: pp. 66–73. https://doi.org/10.1049/cp:19941096
[10] S.H. Kia, H. Henao, G.A. Capolino, Analytical and experimental study of gearbox mechanical effect on the induction machine stator current signature, IEEE Trans. Ind. Appl. 45 (2009) 1405–1415, https://doi.org/10.1109/TIA.2009.2023503.
[11] I. Bravo-Imaz, A. Arnaiz, H. Davari Ardakani, Z. Liu, J. Lee, A. García-Arribas, Motor current signature analysis for gearbox condition monitoring under transient speeds using wavelet analysis and dual-level time synchronous averaging, Mech. Syst. Signal Process. 94 (2017) 73–84, https://doi.org/10.1016/j.ymssp.2017.02.011.
[12] F. Gu, A.D. Ball, R. Zhang, H. Mansaf, T. Wang, Gear wear monitoring by modulation signal bispectrum based on motor current signal analysis, Mech. Syst. Signal Process. 94 (2017) 202–213, https://doi.org/10.1016/j.ymssp.2017.02.037.
[13] N. Chai, M. Yang, Q. Ni, D. Xu, Gear fault diagnosis based on dual parameter optimized resonance-based sparse signal decomposition of motor current, IEEE Trans. Ind. Appl. 54 (2018) 3782–3792, https://doi.org/10.1109/TIA.2018.2821099.
[14] S.H. Kia, H. Henao, G.-A. Capolino, Gear tooth surface damage fault detection using induction machine stator current space vector analysis, IEEE Trans. Ind. Electron. 62 (2015) 1866–1878, https://doi.org/10.1109/TIE.2014.2360068.
[15] Chuan Lia, René-Vinicio Sanchez, Grover Zurita, Mariela Cerrada, Diego Cabrera, Rafael E. Vásquezc, Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis, Neurocomputing 168 (2015) 119–127, https://doi.org/10.1016/j.neucom.2015.06.008.
[16] D. Wang, K-nearest neighbors based methods for identification of different gear crack levels under different motor speeds and loads: Revisited, Mech. Syst. Signal Process. 70–71 (2016) 201–208, https://doi.org/10.1016/j.ymssp.2015.10.007.
[17] S. Zhu, B. Jiao, Fault Diagnosis of Gearbox of Wind Turbine Based on Improved Decision Tree Algorithm, in: Atlantis-Press.Com, 2017, https://doi.org/10.2991/caai-17.2017.74.
[18] Y. Ren Wang, Q. Jin, G. Dong Sun, C. Fei Sun, Planetary gearbox fault feature learning using conditional variational neural networks under noise environment, Knowl.-Based Syst. 163 (2019) 438–449, https://doi.org/10.1016/j.knosys.2018.09.005.

[19] J. Liu, Y. Hu, Y. Wang, B. Wu, J. Fan, Z. Hu, An integrated multi-sensor fusion-based deep feature learning approach for rotating machinery diagnosis, Meas. Sci. Technol. 29 (2018), https://doi.org/10.1088/1361-6501/aaaca6.
[20] T. Yang, F. Zhai, J. Liu, M. Wang, H. Pen, Self-organized cyber physical power system blockchain architecture and protocol, Int. J. Distrib. Sens. Networks. 14 (2018), https://doi.org/10.1177/1550147718803311.
[21] Q. Jiang, F. Chang, B. Sheng, Bearing fault classification based on convolutional neural network in noise environment, IEEE Access. 7 (2019) 69795–69807, https://doi.org/10.1109/ACCESS.2019.2919126.
[22] H. Chen, J. Wang, B. Tang, K. Xiao, J. Li, An integrated approach to planetary gearbox fault diagnosis using deep belief networks, Meas. Sci. Technol. 28 (2017), https://doi.org/10.1088/1361-6501/aa50e7.
[23] G. Jiang, H. He, P. Xie, Y. Tang, Stacked multilevel-denoising autoencoders: a new representation learning approach for wind, Ieee Trans. Instrum. Meas. 66 (2017) 2391–2402, https://ieeexplore.ieee.org/abstract/document/7932980/ (accessed July 11, 2019).
[24] C. Shen, J. Shi, Y. Qi, X. Jiang, Z. Zhu, D. Wang, Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery, IEEE Access. 5 (2017) 15066–15079, https://doi.org/10.1109/access.2017.2728010.
[25] L. Cao, J. Wang, J. Lei, C. Liu, D. Jiang, Intelligent fault diagnosis of wind turbine gearbox based on long short-term memory networks, Elsevier. 133 (2019) 422–432, https://www.sciencedirect.com/science/article/pii/S0960148118312151 (accessed July 11, 2019).
[26] D.-F. Wang, Y. Guo, X. Wu, J. Na, G. Litak, Planetary-gearbox fault classification by convolutional neural network and recurrence plot, Appl. Sci. 10 (2020) 932, https://doi.org/10.3390/app10030932.
[27] Z. Chen, C. Li, R.V. Sánchez, Multi-layer neural network with deep belief network for gearbox fault diagnosis, J. Vibroengineering. 17 (2015) 2379–2392, https://www.jvejournals.com/article/16039 (accessed July 11, 2019).
[28] Z. Chen, W. Li, Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network, IEEE Trans. Instrum. Meas. 66 (2017) 1693–1702, https://doi.org/10.1109/TIM.2017.2669947.
[29] R.S. Gunerkar, A.K. Jalan, Classification of ball bearing faults using vibro-acoustic sensor data fusion, Exp. Tech. 43 (2019) 635–643, https://doi.org/10.1007/s40799-019-00324-0.
[30] T. Praveen Kumar, M. Saimurugan, R.B. Hari Haran, S. Siddharth, K.I. Ramachandran, A multi-sensor information fusion for fault diagnosis of a gearbox utilizing discrete wavelet features, Meas. Sci. Technol. 30 (2019), https://doi.org/10.1088/1361-6501/ab0737 085101.
[31] D. Lu, W. Qiao, X. Gong, Current-based gear fault detection for wind turbine gearboxes, IEEE Trans. Sustain. Energy. 8 (2017) 1453–1462, https://doi.org/10.1109/TSTE.2017.2690835.
[32] M. Blodt, M. Chabert, . . . J.R.-I.T. on, U. 2006, Mechanical load fault detection in induction motors by stator current time-frequency analysis, IEEE Trans. Ind. Appl. (2005) 1881–1888. https://doi.org/10.1109/IEMDC.2005.195977.
[33] Y. Li, Electrical signature analysis of synchronous motors under some mechanical anomalies, 2012. https://oaktrust.library.tamu.edu/handle/1969.1/156349 (accessed July 11, 2019).
[34] M.Z. Alom, T.M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M.S. Nasrin, B.C. Van Esesn, A.A.S. Awwal, V.K. Asari, The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches, ArXiv Prepr. ArXiv. 1803.01164 (2018). http://adsabs.harvard.edu/abs/2018arXiv180301164Z (accessed July 11, 2019).
[35] SpectraQuest Inc., Gearbox Prognostics Simulator, (n.d.). https://spectraquest.com/prognostics/details/gps/ (accessed July 11, 2019).
[36] Shop HTA 100-S - HTA 100-S by LEM | HTA 100-S, (n.d.). https://www.darrahelectric.com/products/hta-100-s (accessed July 26, 2019).
[37] P.J. Rousseeuw, C. Croux, Alternatives to the median absolute deviation, J. Am. Stat. Assoc. 88 (1993) 1273–1283, https://doi.org/10.1080/01621459.1993.10476408.
[38] M. Chiani, D. Dardari, Improved exponential bounds and approximation for the Q-function with application to average error probability computation, in: Conf. Rec./IEEE Glob. Telecommun. Conf., 2002, pp. 1399–1402, https://doi.org/10.1109/glocom.2002.1188428.