Check for updates

Article

Journal of Vibration and Control
2022, Vol. 0(0) 1–10
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/10775463221091601
journals.sagepub.com/home/jvc
SAGE

A hierarchical deep learning framework for combined rolling bearing fault localization and identification with data fusion

Mingxuan Liang[1] and Kai Zhou[2]

Abstract

Fault diagnosis of rolling bearings becomes an important research subject, where the data-driven deep learning-based techniques have been extensively exploited. While the state-of-the-art research has shown the substantial progresses in bearing fault diagnosis, they mostly were implemented upon the hypothesis that the location of bearing prone to failure already is known. Nevertheless, in actual practice many rolling bearings are installed in a complex machinery system, any of which is likely subject to fault. As such, fault diagnosis essentially is a process to achieve both fault localization and identification, which results in many fault scenarios to be handled. This will significantly degrade the fault diagnosis performance using conventional deep learning analysis. In this research, we aim to develop a new deep learning framework to address abovementioned challenge. We particularly design a hierarchical deep learning framework consisting of multiple sequentially deployed deep learning models built upon the transfer learning. This can improve the learning adequacy for a high-dimensional problem with many fault scenarios involved even under limited dataset, thereby enhancing the fault diagnosis performance. Without the prior knowledge regarding the fault location, this methodology is greatly favored by the sensor/data fusion which takes full advantage of the enriched pivot fault-related features in the measurements acquired from different accelerometers. Systematic case studies using the publicly accessible experimental rolling bearing dataset are carried out to validate this new methodology.

Keywords

Rolling bearing, deep learning, transfer learning, fault localization and identification, data fusion

1. Introduction

Due to the extensive usage of rolling bearings in the manufacturing, aerospace, and energy and power industries, quality control and health monitoring of the rolling bearings for reliable operation and operation safety recently have attracted significantly growing attention from various communities. A wide range of studies have dedicated on advancing the bearing fault diagnosis capacity, where the machine learning approaches play an important role (Dong et al., 2021; Duan et al., 2021; Zhou et al., 2017). The underlying idea of these approaches is to elucidate the intrinsic correlation between the faults and different types of measurements, such as vibration, acoustic emission and eddy current and so on (Aasi et al., 2021; Ben Ali et al., 2015; Chen et al., 2016; De Moura et al., 2011; Jiang et al., 2019; Pandya et al., 2013; Tabatabaei et al., 2020;). Among them, vibration signals are most widely used for bearing fault diagnosis because of the low instrumentation cost and sufficient fault-related signatures contained (Ben Ali et al., 2015; Chen et al., 2016; De Moura et al., 2011; Liang and Zhou, 2021). Combined with the well-established signal processing techniques, such as wavelet transform, various machine learning approaches have become fully capable of analyzing the vibration data collected in the machinery system and then performing the fault pattern recognition. For instance, Zhang et al. (2016) used a novel classifier ensemble based on the lifting wavelet packet transforms and sample entropy to improve the fault detection accuracy for rolling bearings. Song et al. (2018) performed comprehensive feature extraction of raw signals by combining

[1]College of Mechanical and Electrical Engineering, China Jiliang University, China
[2]Department of Mechanical Engineering-Engineering Mechanics, Michigan Technological University, Houghton, MI, USA

Received: 10 October 2021; revised: 15 February 2022; accepted: 16 March 2022

Corresponding author:
Kai Zhou, Department of Mechanical Engineering-Engineering Mechanics, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931-1295, USA.
Email: kzhou@mtu.edu

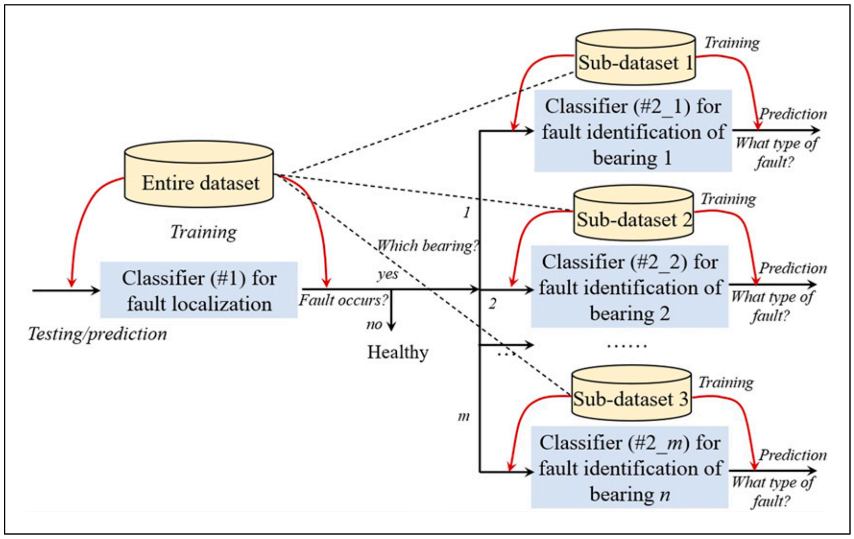

Figure 1. Implementation structure of hierarchical deep learning framework.

statistical filter (SF), wavelet package transform (WPT) and moving-peak-hold method (M-PH), upon which the decision tree was employed to achieve the bearing fault diagnosis. Rohani Bastami et al. (2019) employed the wavelet packet decomposition and neural network to monitor the conditions of rolling bearings and predict their remaining useful life (RUL). Inturi et al. (2019) conducted the fault diagnosis of bearing in the wind turbine gearbox operating under nonstationary loads by using the discrete wavelet transform (DWT)-based support vector machine. While the tremendous progresses have been achieved, the bearing fault diagnosis performance is hindered by the feature selection required in the signal processing procedure, which is highly dependent on the empirical experience and judgment. Moreover, the entire fault diagnosis process cannot be streamlined because of the manual feature selection.

Deep learning recently has become a mainstream in machinery fault diagnosis research owing to its powerful capability in extracting features from massive raw data. As such, no pre-processing steps using signal processing are required and the fault diagnosis as a result can be automated to accommodate the real-time monitoring purpose (Chen et al., 2018; Pan et al., 2018; Sonkul et al., 2021). More importantly, because of the flexibility in architecture design, deep learning generally is tailored for a broad range of fault diagnosis applications than traditional machine learning approaches. The well-known deep learning models including but not limited to deep convolutional neural network (CNN), stacked autoencoder and deep belief network have been extensively employed in bearing fault diagnosis nowadays (Sun et al., 2019; Xu and Tse, 2019; Zhang et al., 2018). While the deep leaning methods indeed show the promising prospect, there still exist the challenges for practical fault diagnosis applications. To build the complex input-output mapping, deep learning model oftentimes is built with large scale, which is expected to be adequately trained upon the sufficient data samples. However, the labeled data collected from the real system usually are limited. As a result, the model overfitting likely occurs. To address the issue of data scarcity, the artificial data enrichment techniques, such as data overlap truncation and interpolation, and data augmentation were proposed (Li et al., 2020; Zhang et al., 2018). Apart from these solutions, the novel configuration/design of deep learning strategy also appears to be effective, leading to the establishment of some representative approaches, that is, transfer learning and semi-supervised learning (Verstraete et al., 2020; Wen et al., 2019). Transfer learning follows the concept of domain knowledge transfer where a model developed for a previous task can be repurposed for a new task. With the transfer learning integrated, the deep learning analysis can yield the satisfactory fault diagnosis performance even under the small-sized dataset. In comparison, the major strength of semi-supervised learning is to improve the fault diagnosis quality through fully exploiting the large amount of unlabeled data that can be inexpensively collected. According to the way to handle the unlabeled data, different types of approaches under semi-supervised learning framework, such as pseudo labeling, generative adversarial network (GAN) were developed and implemented (Arazo et al., 2020; Verstraete et al., 2020). The other notable challenge to be pointed out is that the actual unseen fault conditions are far beyond the known fault labels collected in training data for constructing the deep learning model. As such, it becomes critically important to develop an enhanced deep learning model with extended inference capability (Wu et al., 2021; Zhou and Tang, 2021).

It is worth mentioning that the bearing fault diagnosis in above literature review were performed upon the premise that the bearing with fault occurrence is known and the proper sensor placement for vibration measurement can be executed accordingly. Nevertheless, in actual practice when a complex machinery system is concerned, the fault likely occurs in any bearings in the system. Therefore, the simultaneous localization and identification of bearing faults bear practical significance, which however remains an open research area. In this context, we aim to accomplish the combined bearing fault localization and identification in this research. In addition to the fact that the fault can be induced at any bearings in a system, the fault is represented by different conditions/types, resulting in the numerous fault scenarios to be classified by the deep learning analysis. The resulting high output dimensionality will lead to the poor model predictive ability. To reduce the classification output dimensionality and thus enhance the decision-making reliability and accuracy, a hierarchical deep learning framework is particularly developed. Additionally, the fusion of data from different accelerometers is utilized, aiming at providing the spatially distributed vibration information to facilitate the fault diagnosis. In other words, the hierarchical transfer learning and data fusion are two key parameters of the proposed methodology. The remainder of this paper is organized as follows. In Sec. 2, the proposed methodology and its major components are outlined. Sec. 3 provides the case illustration for methodology validation using the public bearing fault database from Case Western Reserve University (CWRU) bearing data center, followed by the Concluding remarks that are summarized in Sec. 4.

2. Bearing fault diagnosis framework

In this section, the proposed hierarchical deep learning framework built upon the transfer learning for bearing fault diagnosis is succinctly introduced.

2.1. Hierarchical deep learning framework

The main motivation to adopt a hierarchical deep learning framework instead of conventional single deep learning model for bearing fault diagnosis is to reduce the output dimension of the classification analysis. This framework consists of multiple deep learning models that are hierarchically configured as illustrated in Figure 1. The classifier located in the first level tends to pinpoint the bearing with fault occurrence. Training this classifier needs to involve entire training dataset, in which the original known fault labels will be consolidated to ones in light of the bearing locations. It is noted that the fault is sparse in nature; the fault of single bearing thus is only concerned in this research. Therefore, each output of this classifier represents the respective bearing with fault occurrence. In the second level, a cluster of deep learning models are established to identify the specific fault conditions of bearings. The number of classifiers is identical to the number of outputs in the first-level classifier. The proper subset selection of entire training dataset is required to train each corresponding classifier in the second level. Once all integrated classifiers are well established, the testing/prediction can be performed in a sequential manner. Specifically, if the testing sample is determined as "Healthy" by the first-level classifier, no further action is required. Otherwise, the predicted result of the first-level classifier will point to associated second-level classifier for further fault condition identification. The two-level architecture is simply used for illustrating the underlying idea of this framework (Figure 1). This architecture can be generically extended into multiple levels according to the nature of problem to be investigated. For example, the fault condition to be identified in the second level may include the fault type and severity. In the case when the numbers of fault types and severities both are large, additional level can be incorporated to form a three-level hierarchical architecture to keep the appropriate output dimension that can be directly managed. Conversely, the hierarchical architecture will degrade to one level (i.e., single model) if the fault location is not concerned and the number of fault conditions is small.

2.2. Transfer learning

The classifier used in this research is modeled as the deep learning neural network (i.e., convolutional neural network) because of its notable advantages highlighted earlier. It is worth pointing out that the architecture of the deep learning neural networks embedded needs to be carefully configured in order to ensure the adequate training for inherent input-output correlation characterization. Generally, such architecture configuration resorts to the trial-and-error tunning, which heavily relies on the empirical experience. Moreover, while the performance of deep learning neural network intuitively can be enhanced by increasing its scale, it is also profoundly correlated with the size of training data. To minimize the effort in model architecture tuning and meanwhile ensure the desired model performance given limited dataset, the transfer learning, the architecture of which has been well established and validated, will be used as the backbone of this framework.

When employing transfer learning, the first m layers of a well-trained network can be directly transferred to the target network. The rest of layers (M-m) in the target network are left untrained, which will be trained subsequently using the training dataset from the new task. Let the training datasets for the previous and new tasks be represented, respectively, as

$$\mathbf{D}_{pre} = \left[\mathbf{X}_{pre}, \mathbf{y}_{pre}\right] \ \mathbf{D}_{new} = \left[\mathbf{X}_{new}, \mathbf{y}_{new}\right] \tag{1}$$

where $\boldsymbol{x}$ and y are the vibration measurement and respective fault label in one sample, respectively. Let $\boldsymbol{\varphi}_{pre}$ be the parameters of deep learning neural network for previous task. The input-output mapping can be defined as

$$\mathbf{y}_{pre} = g(\mathbf{X}_{pre}, \boldsymbol{\varphi}_{pre}) \tag{2}$$

After training, the parameters then can be updated as

$$\boldsymbol{\varphi}'_{pre} = \underset{\boldsymbol{\varphi}_{pre}}{\arg\min}\left(\mathbf{y}_{pre} - \mathbf{y}'_{pre}\right) = \underset{\boldsymbol{\varphi}_{pre}}{\arg\min}\left(g(\mathbf{X}_{pre}, \boldsymbol{\varphi}_{pre}) - \mathbf{y}'_{pre}\right) \tag{3}$$

In the new task, the first m layers that are transferred from the previous task are frozen, and the rest parameters of the last (M-m) layers can be trained using the new training dataset

$$\begin{aligned} \boldsymbol{\varphi}'_{new}(1:M) &= \left[\boldsymbol{\varphi}'_{pre}(1:m), \boldsymbol{\varphi}'_{new}(m+1:M)\right] \\ &= \underset{\boldsymbol{\varphi}_{new}(m+1:M)}{\arg\min} \left(\mathbf{y}_{new} - \mathbf{y}'_{new}\right) \\ &= \underset{\boldsymbol{\varphi}_{new}(m+1:M)}{\arg\min} \left(g(\mathbf{X}_{new}, \boldsymbol{\varphi}_{new}) - \mathbf{y}'_{new}\right) \end{aligned} \tag{4}$$

Once the training for the new task is completed, the fault diagnosis can be performed over any given vibration time-series sample, expressed as

$$y_{test} = g\left(X_{test}, \left[\boldsymbol{\varphi}'_{pre}(1:m), \boldsymbol{\varphi}'_{new}(m+1:M)\right]\right) \tag{5}$$

Fortunately, many state-of-the-art deep learning architectures for transfer learning spanning across different application domains have been developed, validated, and openly shared (Krizhevsky et al., 2012; Rezende et al., 2017; Wen et al., 2019; Xia et al., 2017). In this research, we particularly adopt the AlexNet network as the classifier. It was originally proposed by Kizhevsky et al. (2012), and its architecture consists of five convolutional and three fully connected stages (Figure 2). In the figure, n is the number of nodes in softmax layer which denotes the number of fault classes in the problem. While different types of images will be used to train the deep learning neural network, the high-level abstraction can be extracted in a similar way through the convolutional stages of the network (Yosinski et al., 2014). Therefore, the first five convolutional stages can be directly transferred to other tasks. To enable the best model learning performance, we treat the fully connected stages to be retrained in the new task as hyperparameters and create three hyperparameter options, including (1) only last fully connected stage to be retrained; (2) last two fully connected stages to be retrained; (3) all three fully connected stages to be retrained. The best option needs to be finalized in subsequent analysis.

3. Fault diagnosis implementation on CRWU bearing fault data

In this section, the bearing fault diagnosis using the proposed framework will be practiced on the dataset from the Case Western Reserve University (CWRU) bearing data center (http://csegroups.case.edu/bearingdatacenter). The advantages of hierarchical deep learning framework and data fusion will be illustrated through the systematic investigation.

3.1. Experimental data acquisition and problem formulation

The experimental testbed to generate the bearing fault data is shown in Figure 3. The testbed consists of a dynamometer, a torque transducer, and an induction motor. Two

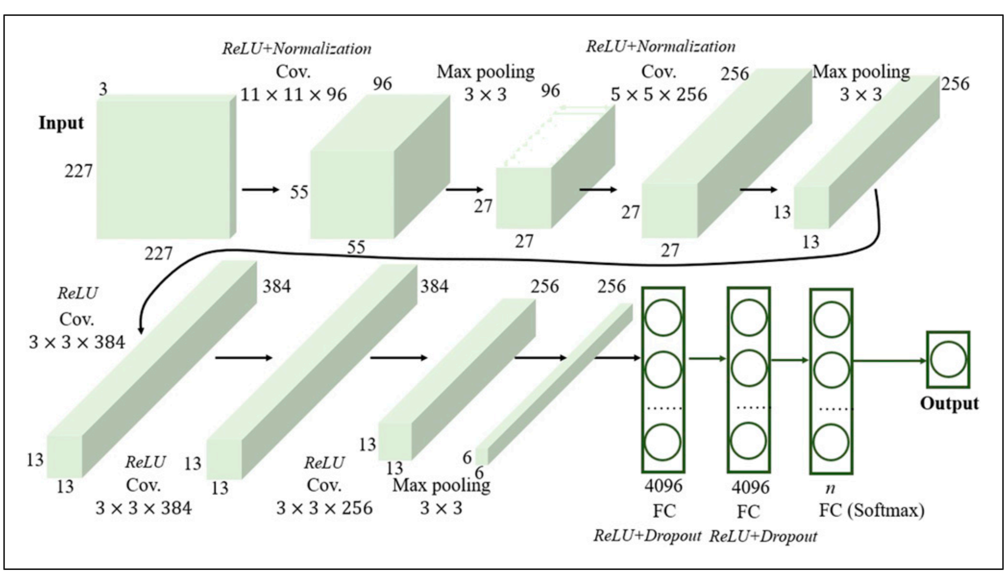

Figure 2. AlexNet network architecture.

accelerometers are placed at drive end (DE) and fan end (FE), respectively, for vibration measurement. The tested bearing is deep groove ball bearing (6205-2RS JEM SKF), which is installed either at DE or FE of the induction motor to mimic the different fault locations. While various datasets are collected under different motor speeds and sampling frequencies, we specifically use the datasets under 1797 r/min motor speed and 12K sampling frequency to facilitate the methodology validation. The new dataset to be used is the combination of the two datasets corresponding to the bearing faults at DE and FE, respectively. For each bearing that is likely subject to fault, there are three fault types (i.e., inner ring (IR), ball element (BA), and outer ring (OR) faults), each of which has three severities (i.e., 0.007, 0.014, and 0.021 inches), yielding nine fault conditions. With the healthy condition, totally 19 ($9 \times 2 + 1$) fault labels/classes are involved for subsequent classification analysis. For each fault condition, the first 10-s vibration signals are utilized and segmented into 75 samples, each of which has 1600 data points to reflect the vibration in 4 revolutions. It is noted that there is no overlap between adjacent samples which avoids the data dependency in model training. The total number of segmented samples covering all fault locations and conditions hence are 1425 (75×19). The details of the dataset used in this research are summarized in Table 1.

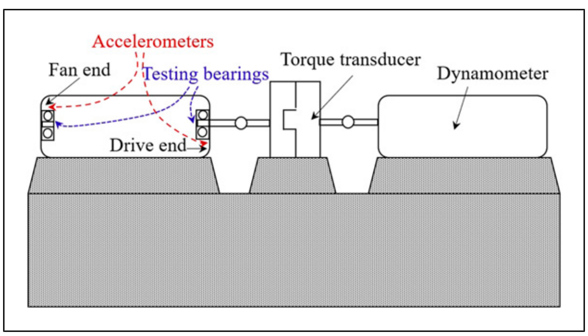

Figure 3. Experimental testbed.

To fully verify the effectiveness of the proposed methodology, several testing scenarios are formulated with details given in Table 2. We firstly examine the fault diagnosis performance among Scenarios 1, 2, and 3, in which the influential variable is the vibration measurement. The vibration data measured via single accelerometer at DE and FE are employed in Scenarios 1, 2, respectively, whereas the fusion of vibration data measured via both accelerometers is adopted in Scenario 3. The single deep learning model is used for all above three scenarios. The purpose of this investigation is to elucidate the effect of data fusion. We then look into the results of Scenarios 3 and 4, which both utilize the data fusion. In Scenario 4, the hierarchical deep learning framework is employed, and the result will be compared with that of Scenario 3. Through such scenario formulation and associated result comparison, the advantages of two key built-in features, that is, hierarchical deep learning architecture and data fusion in the proposed methodology can be highlighted.

Table 1. Overview of bearing fault data.

Bearing location	Fault condition	Fault severity/Inch	Dataset size	Class ID
NA	Healthy	No	75	1
	Inner Ring (IR) fault	0.007	75	2
Drive end		0.014	75	3
		0.021	75	4
	Ball Element (BA) fault	0.007	75	5
Drive end		0.014	75	6
		0.021	75	7
	Outer Ring (OR) fault	0.007	75	8
Drive end		0.014	75	9
		0.021	75	10
	Inner Ring (IR) fault	0.007	75	11
Fan end		0.014	75	12
		0.021	75	13
	Ball Element (BA) fault	0.007	75	14
Fan end		0.014	75	15
		0.021	75	16
	Outer Ring (OR) fault	0.007	75	17
Fan end		0.014	75	18
		0.021	75	19

3.2. Fault diagnosis performance examination and validation

As mentioned, each data sample contains 1600 data points (or data features), indicating a time series with 0.133 s. When carrying out the data fusion using two accelerometers, the number of original data features in each sample will be doubled, that is, 3200. In other words, the data fusion increases the number of features in each sample instead of the size of data samples. Figure 4 gives the examples of data fusion under different fault locations and conditions. Figures 4(a) and (b) indicate that the fault-induced vibration amplitude will become more significant using the accelerometer that is closer to the fault location.

Table 2. Scenarios formulated.

Scenario ID	Model type	Accelerometer placement
1	Single deep learning network	DE
2	Single deep learning network	FE
3	Single deep learning network	DE+FE
4	Hierarchical deep learning networks	DE+FE

The vibration amplitude discrepancy also depends on the fault type. It seems that the vibration amplitude variation due to OR fault is small despite accelerometer location (Figure 4(c)). For healthy condition, the vibrations measured at DE and FE become quite identical (Figure 4(d)). It is worth mentioning that the fundamental mechanism of machine learning is to learn the intrinsic correlation between the features and associated faults. The learning process becomes more adequate if the features contain more pivot signatures explicitly pointing to the faults. Increasing the features through data fusion essentially will enrich the pivot signatures, thereby enhancing the fault diagnosis performance.

Each time-series sample essentially is a 1-D array. It is converted to a 2-D image which will be further resized to $227 \times 227 \times 3$ to be consistent with the size of input layer of AlexNet network. Recall that each AlexNet network is integrated with transfer learning, and three retraining options are available for model establishment. We examine the training performance of all options and find that the option 1 tops the rest. Unless otherwise specified, the results throughout the manuscript are analyzed through the AlexNet network with transfer learning hyperparameters shown in option 1. The Adam optimizer built upon the stochastic gradient descent (SGD) algorithm is adopted to

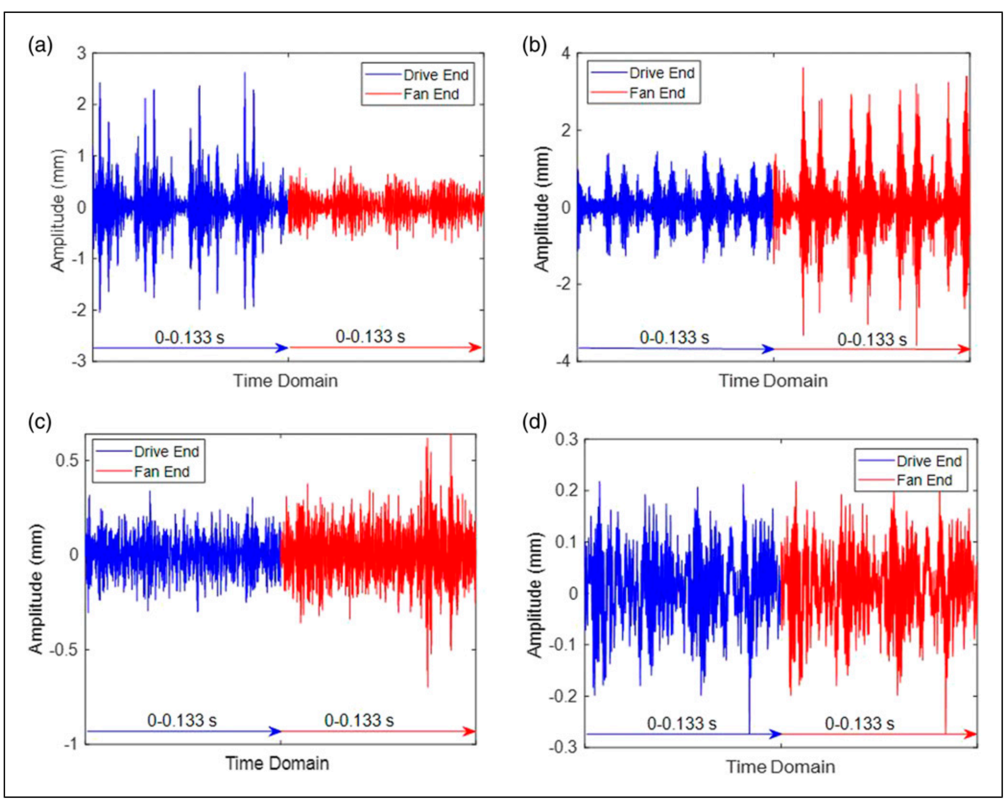

Figure 4. Illustration of data fusion (single sample) (a) 0.021 IR fault of bearing at DE; (b) 0.021 OR fault of bearing at FE; (c) 0.021 BA fault of bearing at FE; (d) Healthy condition. (Note: each sample lasts 0.133 s according to the sampling frequency and sample segmentation mentioned above).

direct the backpropagation training (Zhou et al., 2021). The epoch and batch sizes are set as 10 and 5, respectively, which are proven to enable the adequate training through observing the training and validation loss trends with respect to epoch. The 90% of entire 1425 data samples and the rest are used for training and testing, respectively. It is noted that the class balance also is maintained in above data split using so called stratification. The classification accuracy metric defined as the ratio of number of correct predictions to the total number is used to assess the performance, as given below

$$T_{acc}(\%) = \frac{N_{cor}}{N_{tot}} \times 100 \tag{6}$$

where N_{cor} denotes the number of samples that are correctly identified, and N_{tot} denotes the total number of samples.

Under the operating set-up defined above, the classification-based fault diagnosis analyses of Scenarios 1, 2 and 3 are carried out, and the overall accuracy are computed as 84.21%, 83.46%, and 89.47%, respectively. Apparently, the data fusion plays a positive role in enhancing the fault diagnosis accuracy. To dig the result more deeply, the misclassifications take place in these scenarios are further investigated as shown in Figure 5, where each black line denotes the mapping between the actual/true class and wrongly predicted class of certain sample. Overall, the distribution of classes for misclassified samples varies with respect to the scenario. In Scenario 1, the samples of BA faults of bearing at FE (i.e., Classes 15 and 16) being misclassified to BA fault of bearing at DE (i.e., Class 6) play a major contribution (Figure 5(a)). The reason may lie in the lack of sensitivity of bearing faults at FE with respect to the data measured at DE. In Scenario 2, BA faults of bearing at FE (i.e., Classes 14, 15 and 16) are interfered by different severities (Figure 5(b)). The misclassification however doesn't occur across different bearings. Similar observation also can be captured for OR bearing fault at FE (i.e., Class 18). According to the distribution of misclassified samples, the faults of bearing at FE are discriminated more difficultly than that at DE. While in Scenario 3 most misclassifications still are found for the faults of bearing at FE (Figure 5(c)), the data fusion indeed can significantly reduce the misclassification occurrence as compared with both Scenarios 1 and 2.

Considering the randomness existed in training and testing data split, and backpropagation optimization for model training, the results of above scenarios need to be examined in a more robust manner. This generally can be realized by cross-validation analysis. It is well known that k-fold cross validation is a popular cross-validation method

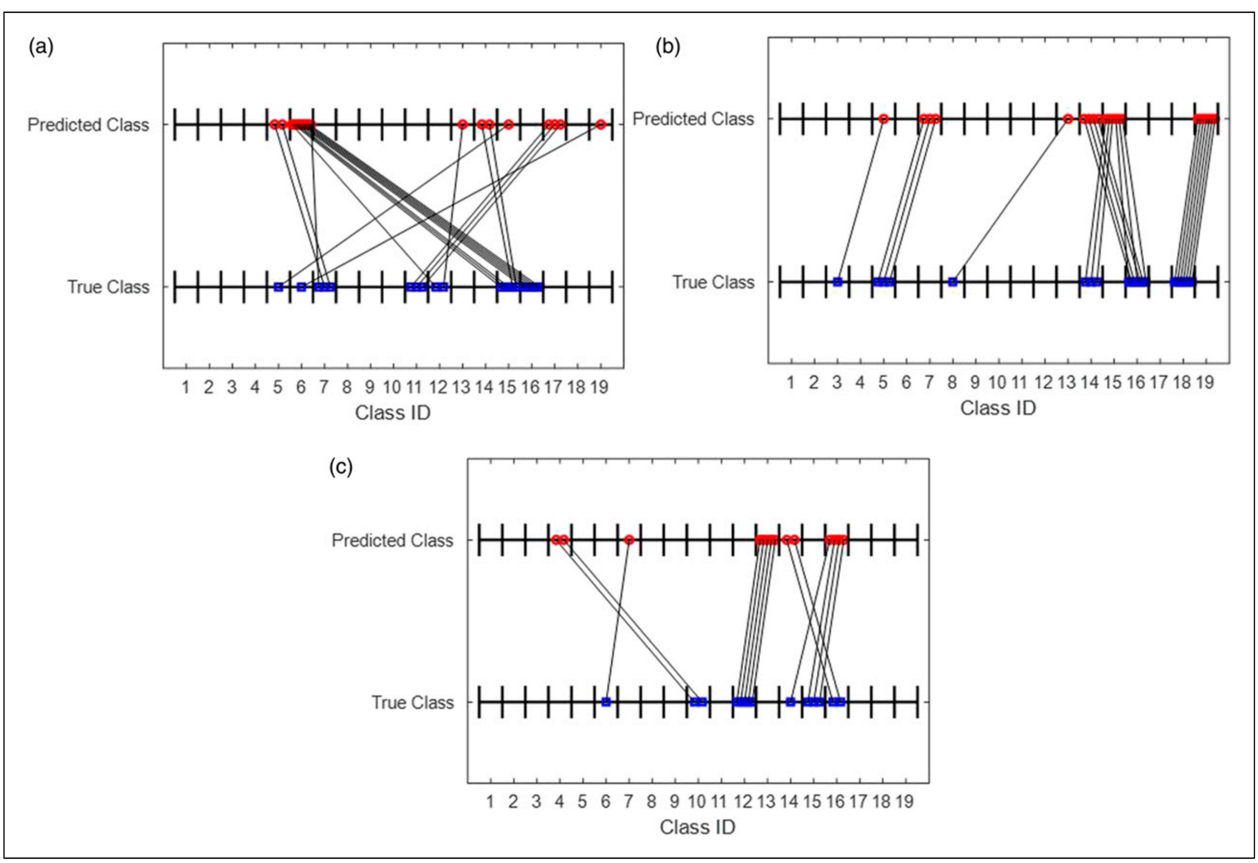

Figure 5. Misclassifications over testing space (a) Scenario 1; (b) Scenario 2; (c) Scenario 3.

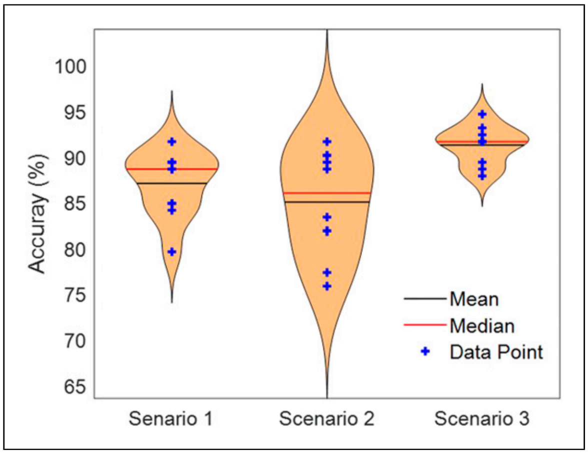

Figure 6. Cross-validation result versus the scenario.

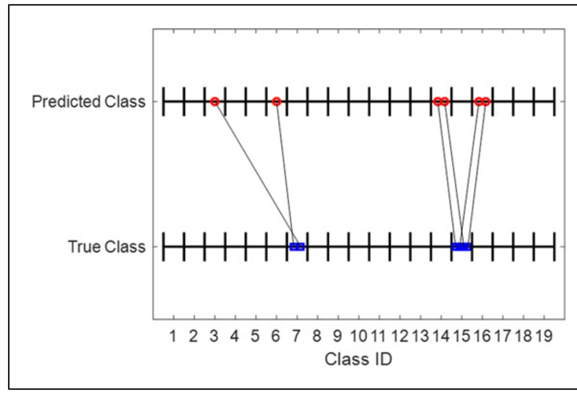

Figure 7. Misclassifications over testing space in Scenario 4.

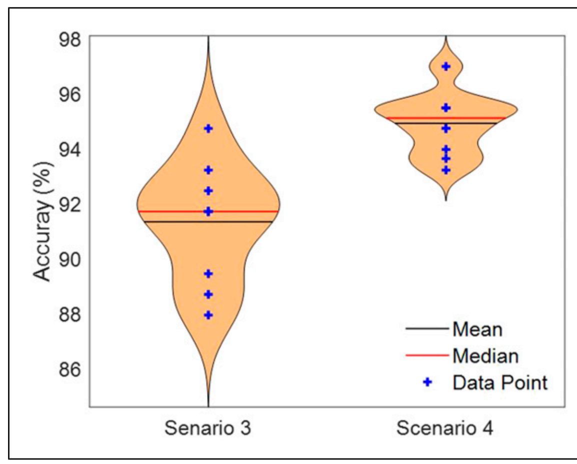

Figure 8. Cross-validation result versus the scenario.

that has been extensively adopted. However, we employ repeated random subsampling cross validation also known as Monte Carlo cross validation in this research because it is more convenient in implementation and is especially suitable for relatively small-sized dataset (Raschka, 2018). To ensure the class balance in classification analysis, stratification is performed during cross-validation analysis. 10 emulations for each scenario are carried out, and each of emulations follows the same training and testing data split ratio used above. The testing accuracy of 10 emulations for all scenarios are compared using violin plots (Figure 6). It is consistently found that the Scenario 3 has the best performance while possessing good robustness, illustrating the feasibility of data fusion for bearing fault diagnosis. Moreover, the performance of Scenario 2 is slightly inferior to that of Scenario 1 because of the worse robustness (i.e., wider distribution). The reason may be associated with the more useful fault-related features that are contained in the measurement from accelerometer at DE than at FE.

Taking full advantage of both the developed hierarchical deep learning framework and data fusion, we then perform the fault diagnosis analysis of Scenario 4 following the similar procedure shown above and compare the result with that of Scenario 3. The testing accuracy of Scenario 4 obtained is 95.49%, which is considerably improved as compared with the accuracy of Scenario 3. The misclassification details also are examined (Figure 7), showing that the misclassified samples are dramatically reduced, especially for the samples corresponding to the faults of bearing at FE. Cross-validation analysis is subsequently provided for systematic validation. The cross-validation results of 10 emulations for both Scenarios 3 and 4 are compared in Figure 8. As can be seen clearly, the accuracy of all emulations for Scenario 4 is above 90%, indicating excellent robustness. Overall, the accuracy level of Scenario 4 is much higher than that of other scenarios, which readily verifies the effectiveness of proposed methodology.

4. Conclusion

In this research, a new deep learning framework is proposed to realize the combined bearing fault localization and identification. This methodology can be tailored to the practical fault diagnosis of bearings in a complex machinery system where the fault likely occurs at any of the bearings. Many fault conditions together with possible fault locations may result in the excessive fault scenarios to be differentiated/identified in the classification analysis. The conventional single deep learning model usually lacks the capability to construct the clear boundaries for these fault scenarios with high dimensionality. To address this issue, a new deep learning framework is particularly designed with hierarchical architecture, encompassing multiple deep learning models in a sequential deployment. The transfer learning is integrated into the deep learning models to direct fault diagnosis analysis, which can not only minimize the effort for tuning network configuration, but also avoid the overfitting given limited dataset. Moreover, the data fusion is adopted to facilitate the adequate learning of intrinsic correlation between the features in the measurement and associated faults, which further improves the fault diagnosis reliability. Inspired by the data fusion concept, the vision-based sensing technique that can lead to the full-field measurement will become one of focal points in our future fault diagnosis research. The effectiveness of proposed framework is verified through carrying out the fault diagnosis implementation on the publicly accessible rolling bearing fault dataset. The results clearly illustrate the strength of this framework as compared with the conventional deep learning method without data fusion.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is supported by National Science Foundation under grant CMMI–2138522.

ORCID iD

Kai Zhou https://orcid.org/0000-0003-2201-6065

References

Aasi A, Tabatabaei R, Aasi E, et al. (2021) Experimental investigation on time-domain features in the diagnosis of rolling element bearings by acoustic emission. *Journal of Vibration and Control*. DOI: 10.1177/10775463211016130.

Arazo E, Ortego D, Albert P, et al. (2020) Pseudo-labeling and confirmation bias in deep semi-Supervised learning. In: 2020 International Joint Conference on Neural Networks (IJCNN): IEEE, pp. 1–8. DOI: 10.1109/IJCNN48605.2020.9207304.

Ben Ali J, Fnaiech N, Saidi L, et al. (2015) Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Applied Acoustics* 89: 16–27. DOI: 10.1016/j.apacoust.2014.08.016.

Chen J, Pan J, Li Z, et al. (2016) Generator bearing fault diagnosis for wind turbine via empirical wavelet transform using measured vibration signals. *Renewable Energy* 89: 80–92. DOI: 10.1016/j.renene.2015.12.010.

Chen Y, Peng G, Xie C, et al. (2018) ACDIN: Bridging the gap between artificial and real bearing damages for bearing fault diagnosis. *Neurocomputing* 294: 61–71. DOI: 10.1016/j.neucom.2018.03.014.

de Moura EP, Souto CR, Silva AA, et al. (2011) Evaluation of principal component analysis and neural network performance for bearing fault diagnosis from vibration signal processed by RS and DF analyses. *Mechanical Systems and Signal Processing* 25(5): 1765–1772. DOI: 10.1016/j.ymssp.2010.11.021.

Dong W, Zhang S, Jiang A, et al. (2021) Intelligent fault diagnosis of rolling bearings based on refined composite multi-scale dispersion q-complexity and adaptive whale algorithm-extreme learning machine. *Measurement* 176: 108977. DOI: 10.1016/j.measurement.2021.108977.

Duan J, Shi T, Zhou H, et al. (2021) A novel ResNet-based model structure and its applications in machine health monitoring. *Journal of Vibration and Control* 27(9–10): 1036–1050. DOI: 10.1177/1077546320936506.

Inturi V, Sabareesh G, Supradeepan K, et al. (2019) Integrated condition monitoring scheme for bearing fault diagnosis of a wind turbine gearbox. *Journal of Vibration and Control* 25(12): 1852–1865. DOI: 10.1177/1077546319841495.

Jiang K, Li L, Han L, et al. (2019) In-process quality inspection of rolling element bearings based on the measurement of microelastic deformation of outer ring. *Shock and Vibration* 2019: 1–12. DOI: 10.1155/2019/5656143.

Krizhevsky A, Sutskever I and Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, et al. (eds), *Advances in Neural Information Processing Systems*. Curran Associates, Inc. Available at: https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

Li X, Zhang W, Ding Q, et al. (2020) Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation. *Journal of Intelligent Manufacturing* 31(2): 433–452. DOI: 10.1007/s10845-018-1456-1.

Liang M and Zhou K (2021) Probabilistic bearing fault diagnosis using Gaussian process with tailored feature extraction. *The International Journal of Advanced Manufacturing Technology* 119(3–4): 2059–2076. DOI: 10.1007/s00170-021-08392-6.

Pan H, He X, Tang S, et al. (2018) An improved bearing fault diagnosis method using one-dimensional CNN and LSTM. *Strojniški Vestnik - Journal of Mechanical Engineering* 64(7–8). DOI: 10.5545/sv-jme.2018.5249.

Pandya DH, Upadhyay SH and Harsha SP (2013) Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN. *Expert Systems with Applications* 40(10): 4137–4145. DOI: 10.1016/j.eswa.2013.01.033.

Raschka S (2018) Model Evaluation, Model selection, and algorithm selection in machine learning. Available at: http://arxiv.org/abs/1811.12808.

Rezende E, Ruppert G, Carvalho T, et al. (2017) Malicious software classification using transfer learning of ResNet-50 Deep Neural Network. In: 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA): IEEE, pp. 1011–1014. DOI: 10.1109/ICMLA.2017.00-19.

Rohani Bastami A, Aasi A and Arghand HA (2019) Estimation of remaining useful life of rolling element bearings using wavelet packet decomposition and artificial neural network. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering* 43(S1): 233–245. DOI: 10.1007/s40998-018-0108-y.

Song L, Wang H and Chen P (2018) Vibration-based intelligent fault diagnosis for roller bearings in low-speed rotating machinery. *IEEE Transactions on Instrumentation and Measurement* 67(8): 1887–1899. DOI: 10.1109/TIM.2018.2806984.

Sonkul NA, Dhage GS and Vyas NS (2021) Single and multi-label fault classification in rotors from unprocessed multi-sensor data through deep and parallel CNN architectures. *Expert Systems with Applications* 185: 115565. DOI: 10.1016/j.eswa.2021.115565.

Sun M, Wang H, Liu P, et al. (2019) A sparse stacked denoising autoencoder with optimized transfer learning applied to the fault diagnosis of rolling bearings. *Measurement* 146: 305–314. DOI: 10.1016/j.measurement.2019.06.029.

Tabatabaei R, Aasi A, Jafari SM, et al. (2020) Experimental investigation of the diagnosis of angular contact ball bearings using acoustic emission method and empirical mode decomposition. *Advances in Tribology* 2020: 1–14. DOI: 10.1155/2020/8231752.

Verstraete DB, Droguett EL, Meruane V, et al. (2020) Deep semi-supervised generative adversarial fault diagnostics of rolling element bearings. *Structural Health Monitoring* 19(2): 390–411. DOI: 10.1177/1475921719850576.

Wen L, Gao L and Li X (2019) A new deep transfer learning based on sparse auto-encoder for fault diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 49(1): 136–144. DOI: 10.1109/TSMC.2017.2754287.

Wen L, Li X, Li X, et al. (2019) A New Transfer Learning Based on VGG-19 Network for Fault Diagnosis. In: IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, pp. 205–209. DOI: 10.1109/CSCWD.2019.8791884.

Wu Y, Jin W, Li Y, et al. (2021) Detecting unexpected faults of high-speed train bogie based on Bayesian deep learning. *IEEE Transactions on Vehicular Technology* 70(1): 158–172. DOI: 10.1109/TVT.2020.3048027.

Xia X, Bing Nan C and Nan B (2017) Inception-v3 for flower classification. In: 2017 2nd International Conference on Image, Vision and Computing (ICIVC). IEEE, pp. 783–787. DOI: 10.1109/ICIVC.2017.7984661.

Xu F and Tse PW (2019) Combined deep belief network in deep learning with affinity propagation clustering algorithm for roller bearings fault diagnosis without data label. *Journal of Vibration and Control* 25(2): 473–482. DOI: 10.1177/1077546318783886.

Yosinski J, Clune J, Bengio Y, et al. (2014) How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems* 4(January): 3320–3328.

Zhang L, Zhang L, Hu J, et al. (2016) Bearing fault diagnosis using a novel classifier ensemble based on lifting wavelet packet transforms and sample entropy. *Shock and Vibration* 2016: 1–13. DOI: 10.1155/2016/4805383.

Zhang W, Li C, Peng G, et al. (2018) A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mechanical Systems and Signal Processing* 100: 439–453. DOI: 10.1016/j.ymssp.2017.06.022.

Zhou H, Shi T, Liao G, et al. (2017) Using supervised kernel entropy component analysis for fault diagnosis of rolling bearings. *Journal of Vibration and Control* 23(13): 2167–2178. DOI: 10.1177/1077546315608724.

Zhou K and Tang J (2021) Harnessing fuzzy neural network for gear fault diagnosis with limited data labels. *The International Journal of Advanced Manufacturing Technology* 115(4): 1005–1019. DOI: 10.1007/s00170-021-07253-6.

Zhou K., Sun H., Enos R, et al. (2021) Harnessing deep learning for physics-informed prediction of composite strength with microstructural uncertainties. *Computational Materials Science* 197: 110663. DOI: 10.1016/j.commatsci.2021.110663.