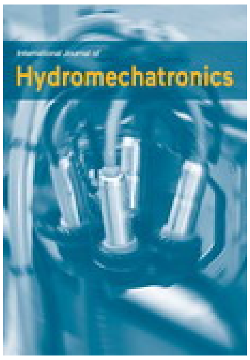


International Journal of Hydromechatronics

ISSN online: 2515-0472 - ISSN print: 2515-0464
https://www.inderscience.com/ijhm

One-shot transfer learning with limited data sample for bearing component fault diagnosis

Wei Ren Sia, Mohd Syahril Ramadhan Mohd Saufi, Muhammad Firdaus Bin Isham, Mohd Salman Leong

DOI: 10.1504/IJHM.2025.10070502

Article History:

Received:	21 November 2024
Last revised:	04 January 2025
Accepted:	26 February 2025
Published online:	15 April 2025

One-shot transfer learning with limited data sample for bearing component fault diagnosis

Wei Ren Sia,
Mohd Syahril Ramadhan Mohd Saufi*,
Muhammad Firdaus Bin Isham and
Mohd Salman Leong

Faculty of Mechanical Engineering,
Universiti Teknologi Malaysia,
81310 UTM Johor Bahru, Johor, Malaysia
and
Faculty of Mechanical Engineering,
Institute of Noise and Vibration,
Universiti Teknologi Malaysia,
54100 Kuala Lumpur,
Wilayah Persekutuan Kuala Lumpur, Malaysia
Email: siaren@graduate.utm.my
Email: mohdsyahrilramadhan@utm.my
Email: muhammadfirdausisham@utm.my
Email: salman.kl@utm.my
*Corresponding author

Abstract: Rolling element bearings are crucial components in the machine, so it is important to maintain the bearings' health. The classic deep learning model needs bulky quality data for the model to achieve high performance. However, it is difficult for the industry to obtain bulk-quality data due to machinery systems working in harsh conditions and the current one-shot learning model has limited capabilities in transfer learning. Thus, a one-shot learning with rhombus Siamese neural network (RSNN) is proposed for a small data size fault diagnosis. RSNN in this study focuses on a large number of classes with a small sample data size and transfer learning without pre-training the target data. The results proved that the one-shot RSNN has high prediction accuracy for the limited data fault detection and diagnosis (FDD) by achieving 90.63% performance with just four training data per class for the 64 classes for the CWRU dataset.

Keywords: one-shot learning; Siamese neural network; SNN; bearing component diagnosis; transfer learning; cross-domain analysis.

Reference to this paper should be made as follows: Sia, W.R., Saufi, M.S.R.M., Isham, M.F.B. and Leong, M.S. (2025) 'One-shot transfer learning with limited data sample for bearing component fault diagnosis', *Int. J. Hydromechatronics*, Vol. 8, No. 6, pp.1–29.

Biographical notes: Wei Ren Sia is currently pursuing PhD with Institute of Noise and Vibration, Faculty of Mechanical Engineering of Universiti Teknologi Malaysia. His research interest includes the machinery fault diagnosis, artificial intelligence and vibration analysis.

Mohd Syahril Ramadhan Mohd Saufi received PhD degree in Faculty of Mechanical Engineering from the Universiti Teknologi Malaysia, Malaysia. Currently, he works as a Senior Lecturer at Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, Skudai, Malaysia. His current research interests include vibration and acoustic emission analysis, machinery fault diagnosis and prognosis, and artificial intelligence system.

Muhammad Firdaus Bin Isham is a Senior Lecturer in the Faculty of Mechanical Engineering, Universiti Teknologi Malaysia. He has received his PhD in Mechanical Engineering, specialised in vibration from the Universiti Teknologi Malaysia. His research interests in rotating machinery fault diagnosis, machine learning, optimisation, and meta-heuristic. He has over five years of consultancy experience in machinery diagnostics, structural vibrations and building acoustics in Malaysia.

Mohd Salman Leong has more than 35 years professional engineering consulting experience and is acknowledged by the industry and government agencies as the leading authority in acoustics, noise, and vibration in the country. He has been involved in many of the mega-projects and high impact consulting and investigation projects in oil and gas, power generation, infrastructure and construction industries. He is currently a Professor and Principal Consultant, and the Founding Director of the Institute of Noise and Vibration, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia.

1 Introduction

In the current industrial system, the tolerance towards the degradation of performance and productivity is reduced especially in wind farms, aircraft engines, petrochemical production, metallurgical production and rotating machinery (Dai and Gao, 2013). Due to the demand and safety control, a reliable control system for the early detection of faults and failures is essential in industry. Thus, early fault detection and diagnosis (FDD) technology is essential to be developed to detect unexpected failures and degraded performance in the machinery (Ma et al., 2019). The early detection of the fault can highly reduce the economic loss from the maintenance, repair and replacement costs (Ma et al., 2019). FDD is used in the industry to monitor the overall system health, detect the malfunction of the system and determine the occurrence of failures in the machinery from the location and type of faults with high accuracy (Abid et al., 2020; Dai and Gao, 2013). It will be good guidance for the industrial experts or workers to know the time to replace the machinery components before a severe failure occurs in the machinery.

With the current growth of artificial intelligence (AI) technology, various types of AI tools have been developed and applied to FDD technology. Due to the diversity of AI technology, the expertise of all types of deep learning (DL)-based FDD technology is impractical for practitioners (Dai and Gao, 2013). Based on the needs of the large quantity of data in training DL models, it is almost impossible to achieve in the industry area. Firstly, the machinery in the industry area is not allowed to run until failure (Wang and Xu, 2021). This is because the lifetime of the machinery is well-extended by the electro-mechanical technology (Zhang et al., 2019a). The long degradation time of the machinery leads to difficulties in the fault data sample collection (Zhang et al., 2019a). Most of the time, the data collected from the machinery contain complex information including the noise and unwanted signals from other components of the machinery. The complex data obtained causes difficulties in distinguishing and extracting useful information (Fang and Wu, 2021). Not only that, but the machinery also usually works in complex conditions with the combination of the loads which gives the different traits in the data obtained. These constraints show that the limited data sample is the main obstacle in the development of the DL-based FDD technology as the current DL model needs the contribution of a high number and good quality of data samples in constructing a high-performance model (Zhao et al., 2022). Thus, the development of small samples or limited data DL-based FDD model is crucial to solving the problem (Wen et al., 2024).

Based on the needs of the current FDD system, the limited data FDD models are essential to be developed. However, in the current study and research, only a few examples are found that use the limited data FDD such as Siamese neural network (SNN) and few-shot learning constructed based on the DL models (Fang and Wu, 2021; Zhang et al., 2019a). The few-shot learning was developed and contributed by Yip and Sussman in 1997 and also Bromley et al. in 1993 in the signatures classification. The limited data DL-based FDD model is then developed by using Bayesian networks (Maas and Kemp, 2009) and using SNN to classify the data in the Omniglot (Koch et al., 2015). The SNN is used in various situations including the classification of the writing system of language, Omniglot (Koch et al., 2015; Malhotra, 2023), handwriting text (Roy et al., 2019), images (Atanbori and Rose, 2022), environmental audio (Honka, 2019), speech recognition (Zhang et al., 2019b) and signature (Bromley et al., 1993; Sharma et al., 2022).

With the application of the FDD technology, the machinery components need to determine the faults also essential. In most machinery components, the rotor, shaft, gears and bearing will be the critical parts that are found to have the highest rate of faults, especially in bearing components. The study of Choudhary et al. (2018) and Kumar et al. (2019) shows a summary survey from the Institution of Electrical and Electronics Engineers (IEEE), ASEA Brown Boveri (ABB) and Electric Power Research Institute (EPRI) about the rate of faults for the electric rotating components in the induction motors. Among the components, the bearing is the major failure component with an overall percentage of 41% in the IEEE study, 42% in the EPRI study and 51% in the ABB study. The criticality of the bearing components is also proved by de Azevedo et al. (2016) in the study of wind turbines. The bearing proved to have the longest downtime in wind turbine machinery (de Azevedo et al., 2016). The downtime and unscheduled maintenance due to unexpected machine failure will be the main challenges for the industrial systems as it will bring a major economic loss. Thus, the early detection of the faults in the bearing components is essential for the industry to prevent loss.

The few-shot and one-shot learning based on the SNN model in the FDD analysis has been done by several researchers. WDCNN-based few-shot learning is proposed by Zhang et al. (2019a) by using a wide first kernel in the basic structure of convolutional neural networks (CNN) to extract more information. Li et al. (2022) has proposed the SHNN to classify the faults of unmanned aerial vehicles to tackle a few data problems with the base model of CNN and long short-term memory (LSTM). Liu et al. (2023) proposed a convolutional neural network-bidirectional long short-term memory (CNN-BiLSTM) to tackle the imbalance data issues but the model has low performance in testing the data with a new dataset. Fang and Wu (2021) have proposed the ANS-net to tackle the noisy environment but has limited robustness in transfer learning without pre-training the model with new data. Zhao et al. (2022) proposed improved Siamese neural network (ISNN) in solving the small data issues. Based on the studies, the robustness of the model in cross-domain analysis or transfer learning without pre-train the model has limited performance. The contribution of this paper is to develop the rhombus Siamese neural network (RSNN) model in handling a large number of classes with limited sample and achieve the high performance of the cross-domain analysis without pretraining the model. The capabilities of RSNN are evaluated in different conditions, including the complex classification of the dataset, simple classification dataset with mixed severity of the fault in a single class, cross-load conditions, model testing with unseen data by using the fault data induced artificially and naturally. The importance of the data segmentation and support testing dataset is also discussed in this paper. The studied machinery components are mainly focused on the bearing component due to its criticality.

The paper consists of several sections. Section 2 discusses the proposed model of SNN model and the testing method. Section 3 discusses the data preparation method, while Section 4 and Section 5 discuss the result and analysis with the conclusion of the study.

2 The proposed model of one-shot learning

Few-shot learning has started to be widely applied in audio, image and text recognition (Maas and Kemp, 2009; Atanbori and Rose, 2022; Koch et al., 2015; Honka, 2019; Bromley et al., 1993; Malhotra, 2023; Sharma et al., 2022; Roy et al., 2019; Zhang et al., 2019b). One of the variations under the few-shot learning is the one-shot learning. The name of the learning is given based on the quantity of the training inputs for the model. N-way K-shot learning is the way that is used to determine the name of the model (Cao et al., 2020). The N-way is the number of the classes involved in the classification of the task while the K-shot is defined as the number of inputs for the training (Cao et al., 2020). The idea of few-shot learning is based on human thinking and recognising ability. In the learning process, humans can recognise the new thing by using just one or a few examples. By mimicking the human's ability, few-shot learning is purposely constructed to recognise the useful features based on the limited input information. It is known as supervised learning as the input information is given with the respective 'names'(classes), which is different from zero-shot learning (unsupervised learning).

The one-shot learning of the RSNN proposed in this research is constructed based on CNNs. The model is named after the rhombus as it has the properties of being narrow in the beginning and the end of layers but wide in the centre layer. The parameters of the CNN base structure are listed in Table 1. The CNNs consist of five convolutional layers, four max pooling layers and global average layer and a fully connected layer. The common CNN applied in the SNNs in current years is the wide first kernel convolutional network as listed in Cui et al. (2021), Lee et al. (2024), Zhang et al. (2019a, 2022). The wide kernel is applied in the first kernel, but the size is decreased across the layer after the first layer. The base convolutional structure proposed in this study is in the rhombus shape. The first kernel is narrow in the beginning with increasing the kernel size until the third layer and decreasing the kernel size until the final layer. The structure proposed is effective in the analysis of using a large number of multiple classes. The large middle kernel can analyse and extract more information from the time series data when involving a large number of classes. The decrease at the end of the network is to extract the information based on the information of the middle large networks for ease of analysis. The base structure is then applied to the SNN structure.

Table 1 Parameters of base CNN structure for RSNN

No.	*Layer*	*Kernel size/pool size*	*Number of filter*	*Stride*	*Activation function*
1	Convolutional 1	8	10	2	ReLU
2	Max pooling 1	2	10	1	-
3	Convolutional 2	16	20	2	ReLU
4	Max pooling 2	2	20	1	-
5	Convolutional 3	32	40	2	ReLU
6	Max pooling 3	2	40	1	-
7	Convolutional 4	16	40	2	ReLU
8	Max pooling 4	2	40	1	-
9	Convolutional 5	8	40	2	ReLU
10	Global average pooling layer	-	40	-	-
11	Fully connected layer	64	1	-	Sigmoid

The RSNN structure applied includes pairs of data as input, subnetwork formed by using CNN, distance metrics, activation function and loss function as shown in Figure 1. The input data is the pair of data from the same classes or different classes to form a similar pair or dissimilar pair of data. The similar pair of data will be represented by 1 while the dissimilar pair will be represented by 0 as the similarity score will be indicated with the minimum of 0 and maximum of 1 in the prediction. Each data from the pairs will be passed through the single subnetwork with the identical weight and bias used. The results from the networks will be first concluded by using the distance metrics and passed to the fully connected network. The results from the fully connected network will be passed through the sigmoid curve and then the loss function. The loss will be calculated and updated for the next training. Different from the other types of neural networks that use numbers to indicate the classes, the proposed model only has two classes in the initial form, which are similar classes and dissimilar classes. The classes are then separated according to the argmax method in the comparison of the similarity score among the different classes.

The distance metrics used in the model are LI distance metrics, which are also known as the Manhattan Distance or Taxicab Distance, used to calculate the absolute difference between two vectors. The distance metric is first applied in the model to conclude the results of two subnetworks as shown in Figure 1. In the application of the distance metrics in 1D-SNN, The L1 distance metrics and L2 distance metrics (Euclidean distance) are almost similar methods as both of the methods applied the absolute value between two inputs. It is the easiest method to determine the similarity and difference between two inputs. In determining the distance metrics, the input pairs of the data are indicated by (x_1^i, x_2^i) can be similar and different, *f* indicates the CNNs constructed and d indicates the distance, d_f^2 calculated between the pairs of the data.

$$d_f^2\left(x_1^i, x_2^i\right) = \left\| f\left(x_1^i\right) - f\left(x_2^i\right) \right\| \tag{1}$$

After the distance values are obtained from the distance metrics, the results will need to determine the similarity of both data inputs. The value obtained from the distance metrics is then passed through the fully connected layer and sigmoid function to obtain the probability as shown in equation (2).

$$P\left(x_1^i, x_2^i\right) = sigm\left(FC\left(d_f^2\left(x_1^i, x_2^i\right)\right)\right) \tag{2}$$

where *P* is the probability obtained from the sigmoid function, *sigm* is the sigmoid function, *FC* is fully connected layer and $d_f^2(x_1^i, x_2^i)$ is the result of the distance metrics.

To increase the accuracy of the model, the probability obtained from the sigmoid function is passed to the cross-entropy loss function as shown in equation (3). It is applied with the adaptive moment estimation (Adam) optimiser to prevent the overfitting of results. The cross-entropy function is shown below in equation (3).

$$L\left(x_1^i, x_2^i, t^i\right) = \log P\left(x_1^i, x_2^i\right) + \left(1 + t^i\right) \log\left(P\left(x_1^i, x_2^i\right)\right) + \lambda^T \mid W \mid^2 \tag{3}$$

where t^i is the truth value to take the value between 0 and 1, $P(x_1^i, x_2^i)$ is the probability results from the sigmoid curve, *W* denotes the weight, *T* denotes the number of the training sample and λ denotes the hyperparameter.

The Argmax function is applied in the classification tasks to understand the competency of the model in the classification of the faults. This study only involves one-shot learning, so there is only one sample from each class from the comparison. The argmax function helps to select the maximum probability from the comparison of the single test data with all the samples in the support set as shown in equation (4). The maximum probability after the comparison indicates the predicted class of the fault.

$$C\left(\hat{x}, S\right) = \arg\max\left(P\left(\hat{x}, x_c\right)\right), x_c \in S \tag{4}$$

where $\hat{x}$ denotes the test sample, *S* denotes the support set, *C* denotes the predicted class and x_c denotes the sample from the support set. There are two types of model testing conducted which are the same support and training data (SSTr) testing method and the same support and testing data (SSTe) testing method. The same SSTr testing method selects the support dataset from the training dataset. At the same time, the SSTe testing method selects the support dataset from the testing dataset, known as the target domain. In both testing methods, only some of the data will be selected as the support dataset. SSTr is applied in both single-load testing and cross-load testing while the SSTe is applied in the cross-load testing only. Single-load testing uses the same group of datasets as training for the testing while cross-load testing uses a different group of data from the training dataset for testing purposes. The cross-load testing includes the cross workload of the dataset and the artificial fault for training while the natural fault for testing. The purpose of conducting the cross-load testing is to determine the model competency in determining the fault from the same machinery with different conditions while the application of the SSTe is to determine the model competency in classifying the fault of the machinery when the useful data is obtained from the machinery with the pre-trained model by using the dataset from the same machinery. This method can eliminate the time used in retraining the model when the new dataset is obtained from the machinery to determine the current machinery condition. SSTr testing method and SSTe testing method is shown in Figure 2.

Figure 1 RSNN structure with the convolutional network (CNN) structure (see online version for colours)

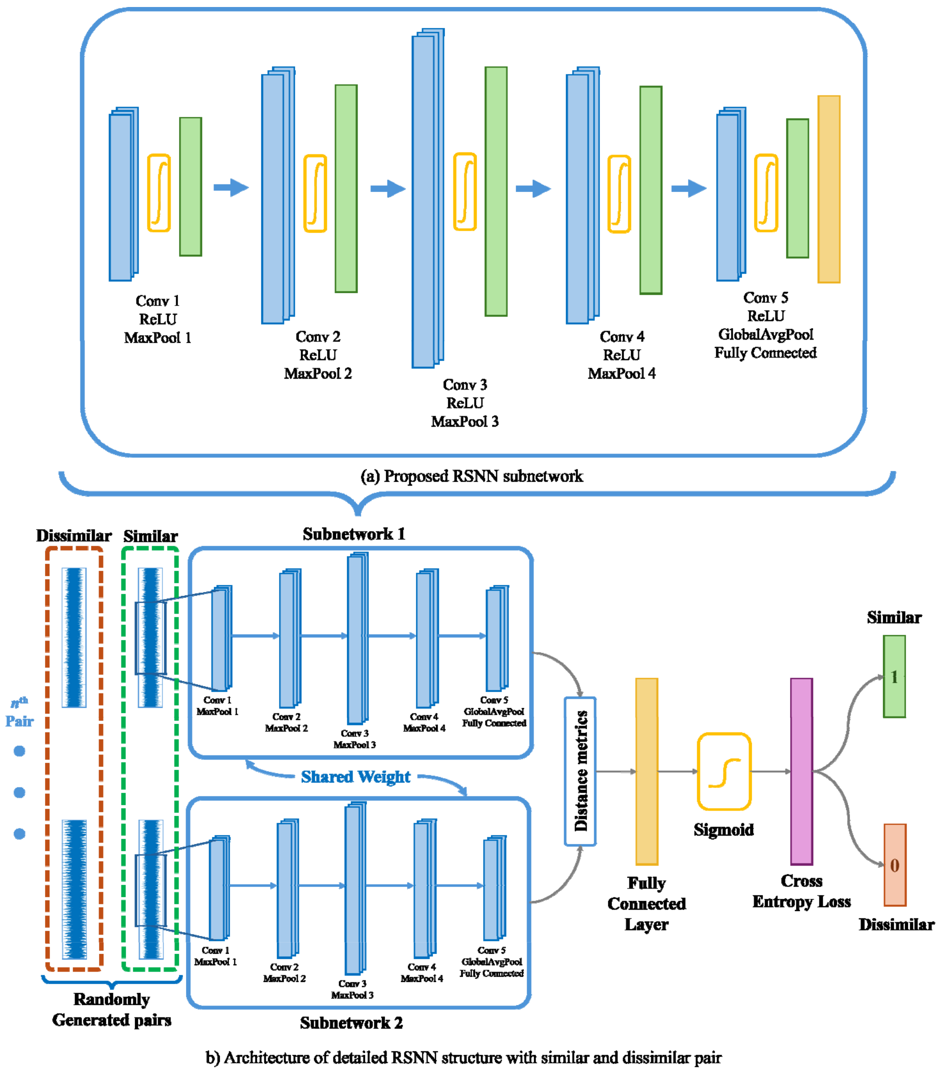
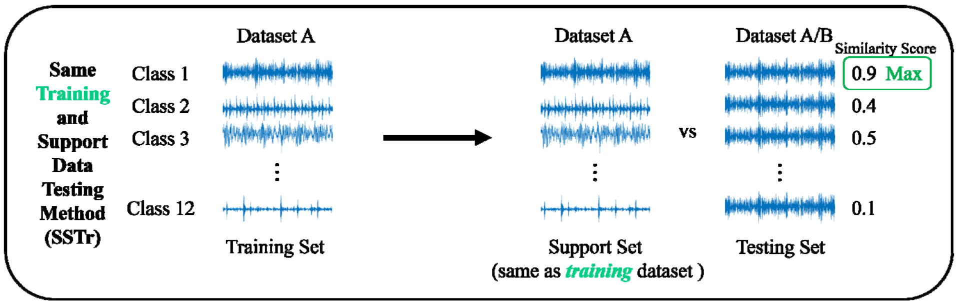
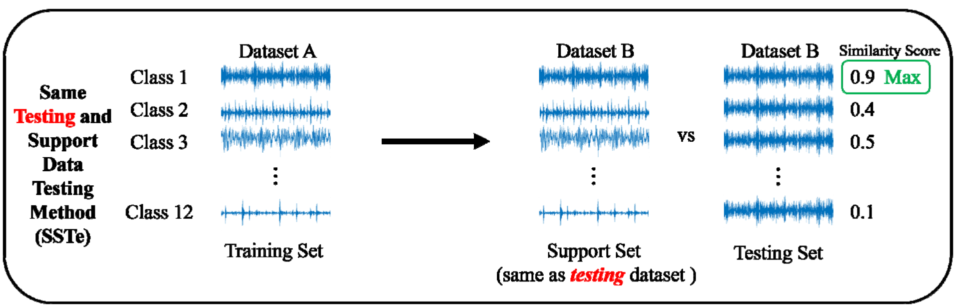


Figure 2 Illustration of the one-shot fault analysis by using argmax function for (a) same support and training dataset (SSTr) testing method and the same support and testing dataset (SSTe) testing method (see online version for colours)



(a)



(b)

3 Experimental setup

Case Western Reserve University (CWRU) Dataset is the online database used for various studies to test the model capabilities. The experiment setup is constructed with a combination of a 2 HP motor at the left, a torque transducer or encoder at the middle, a dynamometer at the right and control electronics as shown in Figure 3 (CWRU, 2021a). The test bearings are used to support the motor shaft in the middle of the machinery (CWRU, 2021a). The test bearing has the artificial fault damaged by the electro-discharge machining as the single point fault with the diameters of 0.007 inch, 0.014 inch, 0.021 inch, 0.028 inch and 0.040 inch (CWRU, 2021a). There are two bearings used in the test including the SKF bearings with a fault scale from 0.007 inch to 0.021 inch and NTN equivalent bearings with the fault scale of 0.028 inch and 0.040 inch. The experiment is conducted in 4 workload conditions as listed in Table 2.

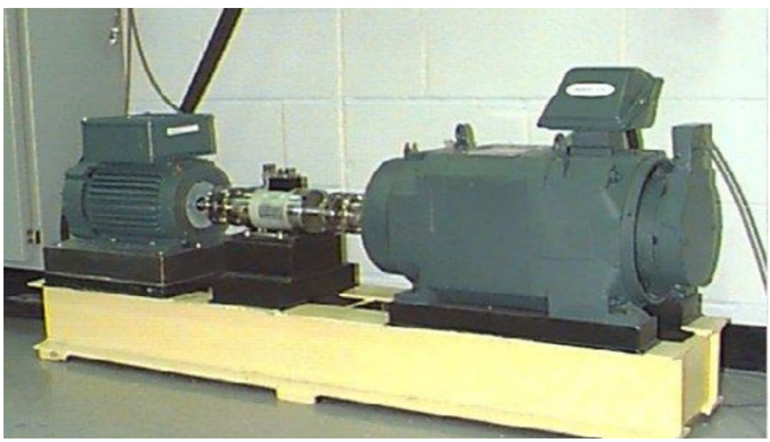
The bearing condition is measured by using the accelerometers to obtain the vibration data, which is attached to the housing of the magnetic bases (CWRU, 2021a). The accelerometers are installed at the 12 o'clock position at the both drive and fan ends of the motor housing and it also installed at the motor supporting base plate for some cases (CWRU, 2021a). The vibration signals are collected with the 16-channel DAT recorder and post-processed with MATLAB (CWRU, 2021a). The digital data is collected in both conditions of 12,000 samples per second and 48,000 samples per second for drive end-bearing faults (CWRU, 2021a). Since the outer race fault is the stationary fault, the placement of the fault relative to the load zone of the bearing has a direct impact on vibrational signals to the motor and the bearing system. Thus, the accelerometer is located at 3 o'clock (directly in the load zone), at 6 o'clock (orthogonal to the load zone) and at 12 o'clock to determine the significant effect of the detection for outer race fault for different locations of the sensor (CWRU, 2021a).

Table 2 Workload condition for CWRU experiments with approximate motor speed

Condition	*Motor load (HP)*	*Approx. motor speed (rpm)*
1	0	1797
2	1	1772
3	2	1750
4	3	1730

Source: CWRU (2021b)

Figure 3 Experiment setup for CWRU (see online version for colours)

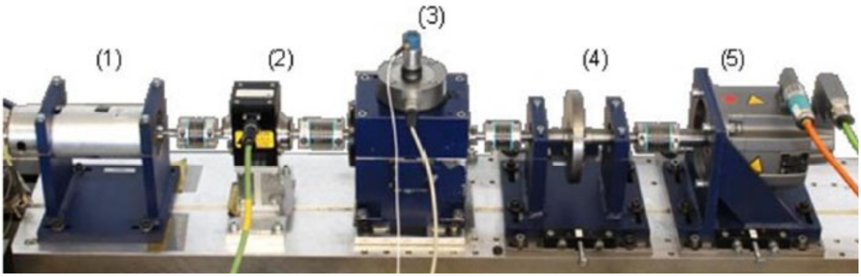


Source: CWRU (2021a)

The test rig for the KAt-Data Center from Paderborn University (PU) is shown in Figure 4. The test rig includes an electric motor (1), a torque-measurement shaft (2), a rolling bearing test module (3), a flywheel (4) and a load motor (5) (Lessmeier et al., 2016). The bearings prepared are assigned to the roiling bearing test module for the experiment's purpose. The roller bearing module is able to be adjusted to a constant 10 kN radial load before the experiment and the adapter allows the measurement of the inner housing (Lessmeier et al., 2016).

The experiment is conducted under different conditions to understand the effect of the different operation parameters towards the analysis (Lessmeier et al., 2016). The detail of the operation is listed in Table 3. Each of the measurements is taken for 4 seconds and with 20 repetitions for the range of the temperature between 45°C and 50°C (Lessmeier et al., 2016). There are 32 bearings used with the categories of 12 bearings with artificial damage, 14 bearings with natural damage and six healthy bearings (Lessmeier et al., 2016). In this study, not all the bearings will be used but only selected bearing conditions will be used.

Figure 4 Experiment setup for PU (see online version for colours)



Source: Lessmeier et al. (2016)

Table 3 Workload condition for PU experiments

No.	*Rotational speed (rpm)*	*Load torque (Nm)*	*Radial force (N)*	*Name of setting*
0	1,500	0.7	1,000	N15_M07_F10
1	900	0.7	1,000	N09_M07_F10
2	1,500	0.1	1,000	N15_M01_F10
3	1,500	0.7	400	N15_M07_F04

Source: Lessmeier et al. (2016)

3.1 Data preparation

After obtaining the data, the data preparation for the training and testing data is essential for the model to have the best performance and efficiency. The data preparation is separated into two parts which are data segmentation for the signal processing to be converted into images and the data segmentation from the raw vibrational signal data in time series.

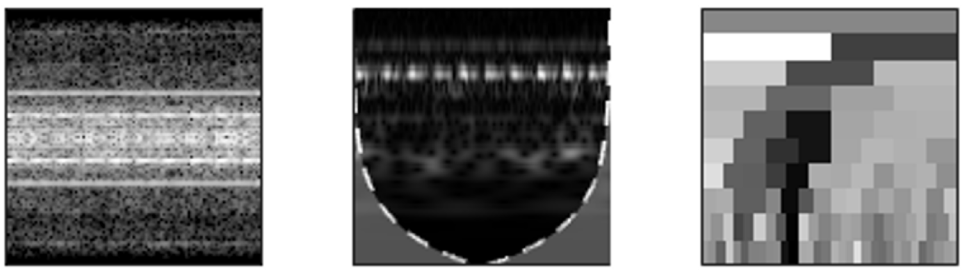
The image data converted from the signal processing is used in the preliminary analysis to determine the competency of the model. The vibrational signal is segmented according to the 10-revolution length to ensure enough information can be passed into the model for the best performance. The formula of one revolution length calculation is shown below as mentioned by Saufi et al. (2019).

$$One\ revolution\ length = \frac{Sampling\ rate}{(rpm / 60)} \tag{5}$$

The signal is segmented and converted into images by using the signal processing method of short-time Fourier transform (STFT), continuous wavelet transform (CWT) and Kurtogram with a size of 105 × 105 in width and height with a resolution of 64 dots per inch (dpi) in the black and white (B&W) images. The example of images converted from the signal processing is shown in Figure 5.

The raw vibrational data in the time series is used in the model evaluation to determine the model's competency in various conditions. In the model evaluation, the training and testing datasets are set according to the 200 to 100 data ratio. Thus, the total number of data that need to be obtained is 300 data. According to this ratio with the data length of 2,560 data points, the increment of the data is set based on the different total data lengths. For CWRU bearing dataset, the increment of the data length is 390 data points while for the PU-bearing dataset, the increment of the data length is 840 data points as shown in the Figure 6. After the data segmentation, the data is then separated according to the training dataset and the testing dataset. To ensure the data trained and tested is highly related, the data is separated simultaneously as shown in Figure 7 according to the ratio of 200:100. The experiment is conducted by using the hardware CPU of i5-10300H and the GPU of NVIDIA GTX1650.

Figure 5 Images converted from signal processing for (a) STFT, (b) wavelet transform and (c) Kurtogram



(a) (b) (c)

Figure 6 Manual data segmentation for 2,560 data points with the increment length of 390 data points for CWRU bearing dataset and 840 data points for PU bearing dataset (see online version for colours)

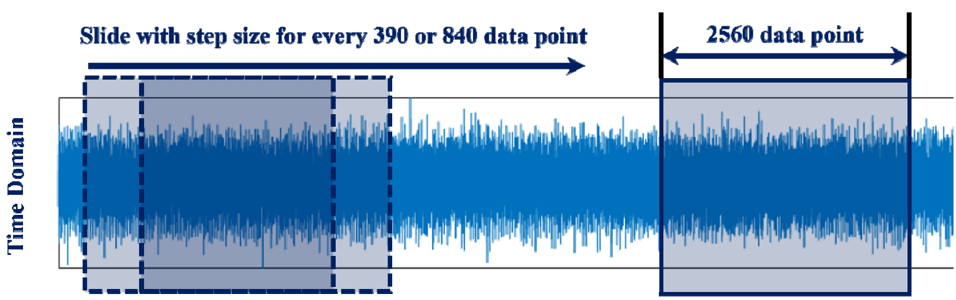


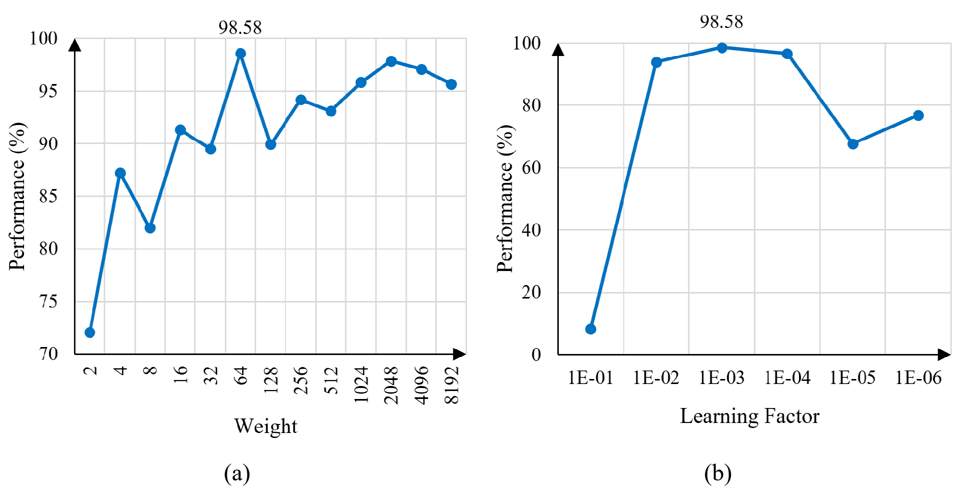
Figure 7 Data separation for the training data and the testing data (see online version for colours)

4 Result and discussion

This section consists of the preliminary analysis and the experimental analysis. The preliminary analysis is an initial analysis of the SNN in image classification. The signal is converted into images based on the signal processing method and test the model based on the similarity score to evaluate the performance. The model used in the preliminary analysis is an image classification model while the experimental analysis is the one-dimensional signal data classification model. The preliminary analysis used the simple SNN model to classify the fault based on the time-frequency images as the SNN is initially designed for the image classification. However, the conversion of the time-frequency images from raw signal data can be controlled to give quality data for the model to recognise useful information easily. Thus, the 1D-SNN proposed and constructed in the experimental analysis aims to minimise human intervention (applied signal processing method), ease or shorten the FDD process and tackle the problem of limited data analysis.

The model constructed will be modified according to the CWRU dataset for the hyperparameter including the learning rate and weight. The learning rate is adjusted according to the negative power for 10, 10^{-n}, while the weight is adjusted according to the power for 2, 2^n as shown in Figure 8. The final selection of weight and learning rate is 64 and 0.001 with the highest performance of 98.58% for both cases. After the tuning process, the model will be used in the performance evaluation based on the CWRU dataset and PU dataset and compared with the support vector machine (SVM), benchmark one-dimensional convolutional neural network (1D-CNN) with the structure as shown in Table 1, one-dimensional convolutional neural network long short-term memory (1D-CNNLSTM) and SNN method proposed in the previous study of SNN based on WDCNN (Zhang et al., 2019a) and SHNN (Li et al., 2022). Each of the experiments will be repeated five times to get the average data and avoid performance bias in the single testing results. The presentation results are the average accuracy of the five repeated results.

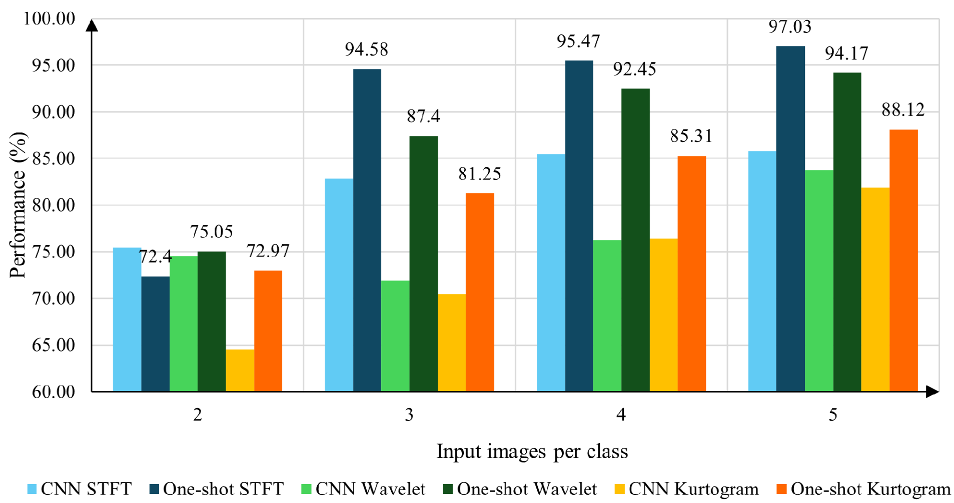
Figure 8 Graph of performance against (a) weight and (b) learning factor (see online version for colours)



4.1 Analysis with time-frequency image as input (CWRU dataset)

Figure 9 shows the graph of performance against training data per class for 64 classes with the overall performance of the one-shot learning being better than the CNN model starting from two training images per class. With the 64 class, the images are separated according to the fault severity, horsepower and the fault condition. Due to the clear classification of the images with the different conditions, one-shot learning has the capability to recognise the similarities and differences among the images according to the classes. Satisfactory performance is generally obtained for all types of signal processing in the one-shot learning model. However, for the CNN model, the enormous number of classes increases the difficulties of the model training to differentiate all the images due to the lack of distance metrics as the SNN model of one-shot learning. The model needs to recognise all the patterns of images for it to classify all classes' images. Thus, the maximum performance reached for the CNN model is only 85.78% for the STFT while the one-shot learning achieves 97.03% for the STFT by using the 5-training data per class. STFT images still reach the maximum performance while the Kurtogram still reaches the worst. It still faces the situation of the model having difficulties in recognising the similarity among the images.

Figure 9 Graph of performance against training data per class for 64 classes (see online version for colours)



From the results of the 64-class analysis, the classical one-shot learning model shows its competency in the fault classification with the signal processing type STFT, Wavelet and Kurtogram. Based on the results obtained, the one-dimensional SNN of one-shot learning can be developed for the analysis. In the classical analysis, the model is not mature enough. The minibatch function is not good enough and leads to a low minibatch number with a lower time in the same iteration number. The epoch number is absent in the training process and leads to the unknown total training number for all images from the training data. The model should be improved in the 1D-SNN to have stable performance in both training and testing.

Moreover, the images based SNN is not further studied due to:

1 its complexity in determining the suitable signal processing method for different signals

2 professional knowledge and experience are needed in signal processing method selection

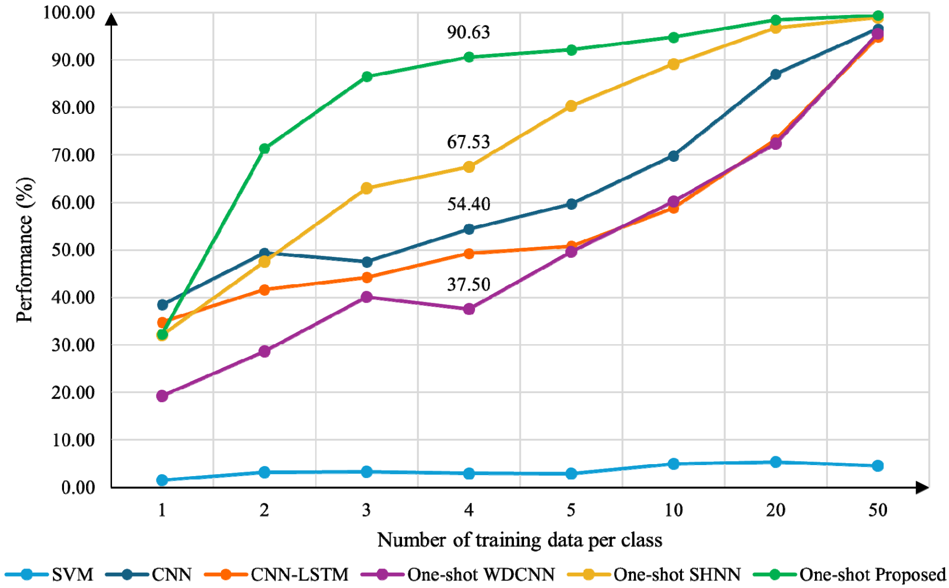
3 severe time-intensive training and testing process based on the SNN structure and images input.

Since the different signal has different properties due to environmental factors, different kinds of signal processing methods are needed for the different conditions. In addition, the application of the signal processing method needs professional knowledge to understand the properties of the signal processing for it to be applied in different conditions, such as STFT is suitable for fast analysis while CWT is suitable for the analysis and extraction of frequency information and Kurtogram is widely applied in high impact signals. Before the selection of the signal processing method, these factors should be considered by the researchers in determining the best method. If the performance of the raw data time series input is similar to or higher than the image input, the human intervention in the deep learning process can be minimised ideally. Moreover, the image input in the training process can be extremely high compared to the raw data time series input. With the same iteration number of 1,000, the average time for training one model of images model in this study is about 5.5 minutes while the average time for raw data time series model is about 1.5 minutes. The application of the image input is not necessary if the raw data time series input has similar or better performance.

4.2 Analysis with raw time domain data as input (CWRU dataset)

Figure 10 shows the graph of performance against training data number per class for 64 classes for a maximum performance of one-shot learning of 99.32% and a minimum performance of 32.22% for one-shot learning. In the graph, one-shot learning only has low performance with a maximum range of 6.22% compared with CNN while it is at least 20% higher performance when compared with other models for the number of training inputs per class from 2 to 10. More than that, the proposed one-shot learning model has a big performance difference when compared to the WDCNN (Zhang et al., 2019a) and SHNN (Li et al., 2022) especially with the four training data per class. The proposed model can achieve a performance of more than 90% while the WDCNN and SHNN are still below 70%. To identify similarities among the data, the 1D-SNN model requires more than one data in the comparison. The model is unable to obtain enough useful information when using just one data per class to distinguish similar data. However, when the training input data per class increased to 2, the performance started to boost up to 71.33% compared to CNN performance of 49.34% and 41.65% for the CNN-LSTM. From the results obtained, one-shot learning is applicable in differentiating the data in a large number of classes with limited data provided from different fault severities, load and fault conditions. When the data per class trained started to exceed 3, all the one-shot model performances exceeded 85% and nearly 100% after 20 data trained per class. One-shot learning proved that it only needs limited data to differentiate a large amount of data with different classes. From the result in Figure 10, SVM shows that it is not effective in the analysis with the raw data input, thus SVM will not be further studied and compared about its capabilities.

Figure 10 Graph of performance against training data number per class for 64 classes (see online version for colours)



4.3 Analysis with additive white Gaussian noise

To evaluate the performance of the model in noisy environments, the additive white Gaussian noise is added to the signal with the signal-to-noise ratio (SNR) from 8 to 10 in decibels (dB). The noise of the signal increased with the decreasing SNR values. Since the 64 class is too complex for the case, this analysis used the 0HP as the basis of analysis. The dataset consists of 12 classes with four fault severities for ball fault and inner race fault, three fault severities for outer race fault and one normal condition. In this case, the data is tested by using the model trained by the original signal data and with the added white Gaussian noise data. This method can reduce the time used to retrain the model and evaluate its performance in noise conditions. The formula for the SNR is shown below.

$$SNR_{dB} = 10 \times \log_{10} \frac{P_{signal}}{P_{noise}} \tag{6}$$

This analysis only involved three models, which are one-shot WDCNN (Zhang et al., 2019a), one-shot SHNN (Li et al., 2022) and the one-shot proposed. The performance of the three models is shown in Table 4 according to the number of training data per class (NTC). Table 4 is applied with the colour contour with the red colour indicating the low-performance region while the green colour indicates the high-performance region. Based on the performance, the one-shot proposed still have a very high performance compared to the one-shot WDCNN (Zhang et al., 2019a), and One-shot SHNN (Li et al., 2022) by using very small data. The model proved that it is still robust under the noise condition.

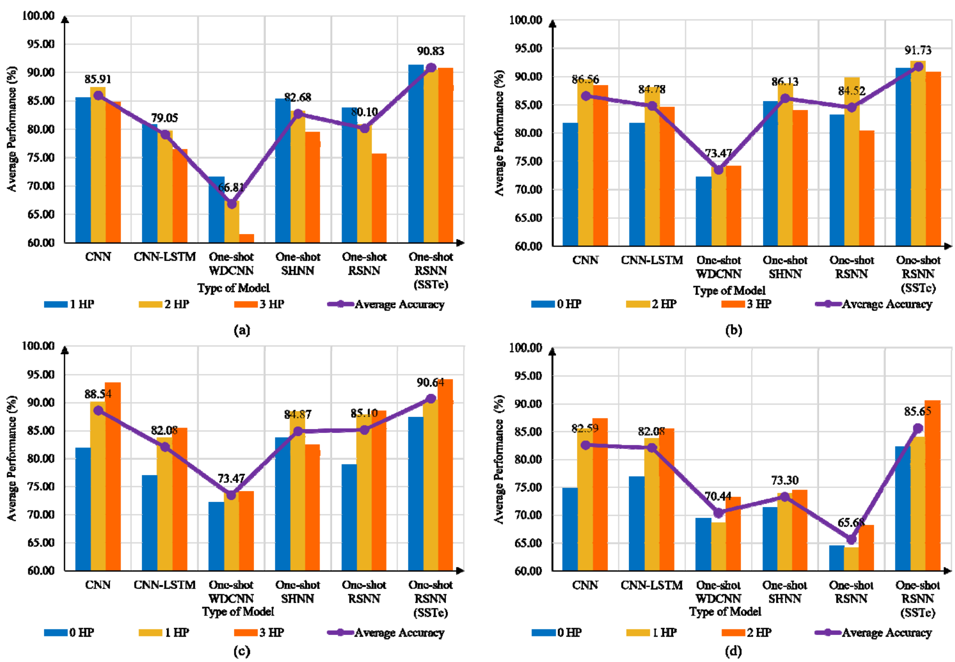
Table 4 Performance of the model under various noise conditions (see online version for colours)

Model	*NTC*	*SNR (dB)*		
		8	*9*	*10*
One-shot WDCNN	1	63.93	60.45	57.88
	2	75.6	75.68	73.7
	3	79.9	85.75	85.52
	4	80.32	78.75	81.82
	5	82.33	80.82	79.73
	10	84.83	85.13	82.47
	20	84.63	83.02	85.65
	50	91.78	90.52	90.92
One-shot SHNN	1	78.58	79.92	78.95
	2	82.05	81.43	78.47
	3	84.75	82.85	84.45
	4	83.23	79.52	82.12
	5	88.77	82.82	88.22
	10	88.5	87.42	84.73
	20	85.4	87.82	85.88
	50	80.85	83.25	82.73
One-shot RSNN	1	86.32	87.85	84.65
	2	85.05	83.03	85.22
	3	88.4	86.35	88.1
	4	87.08	84.67	87.2
	5	88.67	88.87	90.28
	10	90.13	87.97	90.73
	20	92.83	91.02	91.03
	50	88.13	88.93	91.52

4.4 *Cross-machine analysis of the CWRU dataset*

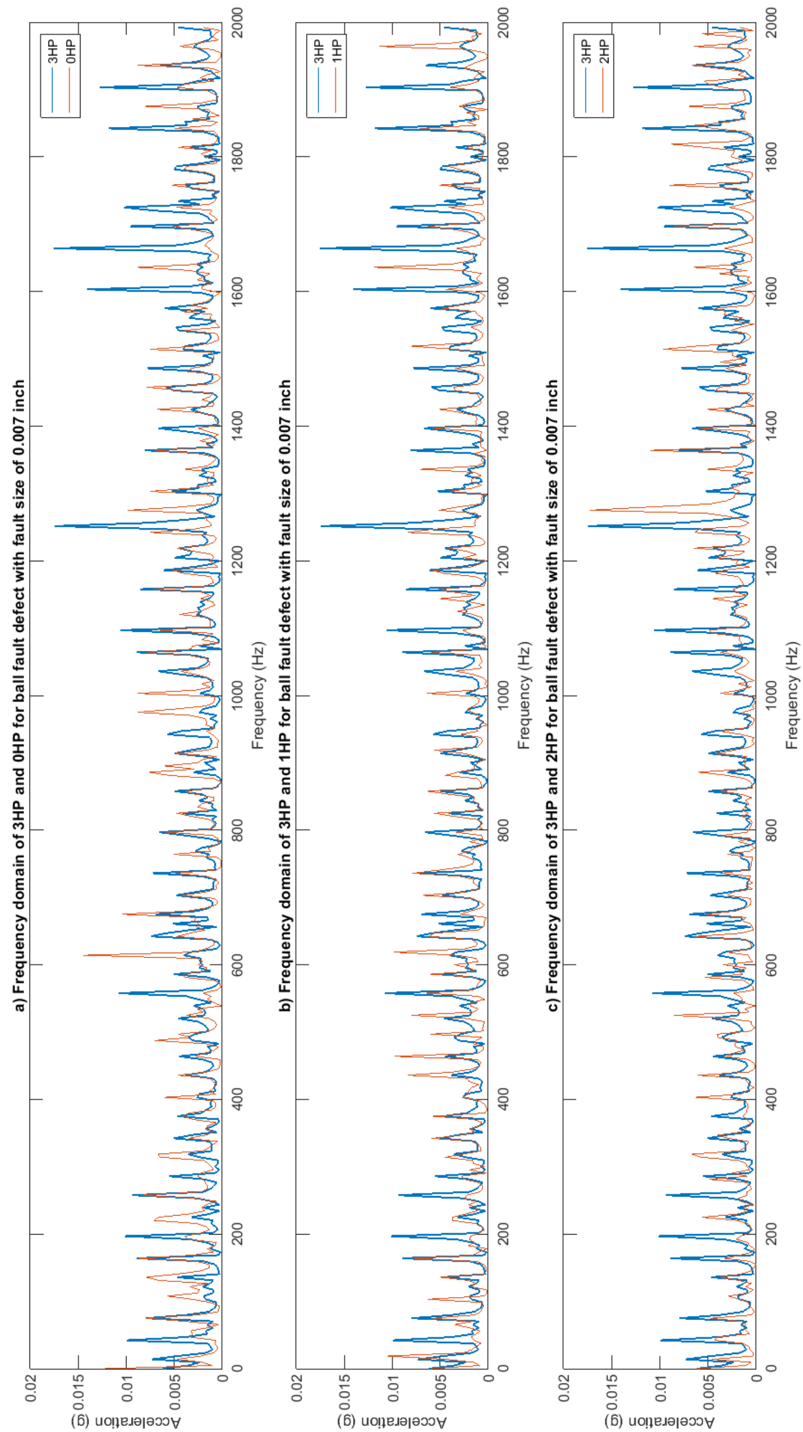
In this analysis, the training dataset and the testing dataset come from two sets of data with different load conditions. There are four sets of data consisting of 0 HP, 1 HP, 2 HP and 3 HP. Every dataset is used as the training dataset for the model to test the other three sets of the data. For example, if 0 HP is used as the training dataset, the 1 HP, 2 HP and 3 HP are used as the testing dataset. This analysis is constructed to determine the compatibility of a pre-trained model to be used for various speed and load conditions. This analysis is different from transfer learning as it does not have the process of training the model by using the testing dataset. It is more feasible to be applied in the real situation. Figure 11 shows the graph of the average performance against the type of models by using 0 HP to 3 HP as the training dataset to evaluate the probability of different loads can be analysed by using one load condition dataset. The legend of the graph in Figure 11 is the target testing dataset with the average accuracy of the three-target testing dataset. From the graph representation in Figure 11, the one-shot (SSTe) method has the highest performance overall while CNN has the second highest performance with the SHNN, and the proposed one-shot learning method has a similar performance except for the 3HP as the training dataset. The WDCNN has the overall lowest performance. The reason the one-shot (SSTe) method has a higher performance as it is used to measure the competency of the model in distinguishing the similarities and differences among the test data. From the results obtained from Figure 11, it shows that the constructed model is able to recognise the similarities and differences of the test data although the pre-trained model used another set of data in the training.

Figure 11 Graph of average performance against type of model (a) 0 HP as the training dataset, (b) 1 HP as the training dataset, (c) 2 HP as the training dataset and (d) 3 HP as the training dataset (see online version for colours)



Moreover, the graph also shows that the testing dataset with a similar load to the training dataset has higher performance compared with other testing datasets. For example, in Figure 11(b), the 2HP testing dataset has the highest performance compared to 0 HP and 3 HP as the testing dataset. The same situation occurs in different loads used as the training dataset except for the CNN model in the 0HP as the training dataset. In obtaining the dataset, the load condition affects the speed condition. It means that the 0 HP has the most similar speed condition with 1HP. The similar speed condition gives similar data characteristics as shown in Figure 12 when the frequency domain is the reference of the analysis. The 2 HP and 3 HP in Figure 12 have the most similar characteristics when compared to the other kinds of data. The similar characteristics in the data make it easier for the model to classify the data in extracting the information which leads to higher performance.

Figure 12 Comparison of frequency domain among 3 HP in 0.007-inch ball fault defect with (a) 0 HP, (b) 1 HP and (c) 2HP with the data length of 2,560 sample points (see online version for colours)



4.5 Performance based on Paderborn dataset

The PU dataset applied in the 1D-SNN has two conditions, which are condition 1, test with the fault severity, fault condition and cause of faults as shown in Figure 13 while condition 2 is the test that only involves fault conditions as shown in Figure 16. The fault data used is obtained at the rotational speed of 1,500 rpm, 0.7 Nm load torque, and 1,000 radial force with the initial setting name of N15_M07_F10. The analysis conducted for the PU dataset aims to determine model competency in training using artificial fault but testing with natural fault. Each class of the data has 200 samples for the training dataset and 100 samples for the testing dataset.

Table 5 Description of the PU dataset condition 1

Bearing condition	*Fault severity*	*Training*	*Test*	
		Artificial	*Artificial*	*Natural*
Healthy	0	K004	K005	K005
OR	1	KA05	KA05	KA04
	2	KA06	KA06	KA16
IR	1	KI03	KI05	KI21
	2	KI07	KI08	KI18

Table 6 Description of the PU dataset condition 2

Bearing condition	*Training*	*Test*
Healthy	K002	K001
	K003	K005
	K004	K006
OR	KA01	KA04
	KA05	KA15
	KA07	KA30
IR	KI01	KI14
	KI05	KI21
	KI07	KI17

4.5.1 Condition 1 – test of fault severity with artificial and natural fault

Figure 13 shows the graph of performance against training data number per class for a single load while Figure 14 and Figure 15 show the graph of performance against training data number per class for the cross load of artificial fault and natural fault. In the single load analysis, the one-shot learning model has the best performance. CNN, CNN-LSTM and SHNN have similar performances in this condition. Starting from eight training data per class to ten training data per class, one-shot learning has a performance of almost 10% higher compared to CNN and CNN-LSTM but SHNN has similar results. However, the SHNN model has a decreasing performance with the increasing training data per class for 10 and 20. When the condition of training involves more class numbers, one-shot learning usually has the best performance compared to other models. One-shot learning shows its capabilities in learning and interpreting difficult tasks.

Figure 13 Graph of performance against training data number per class (condition 1 single load) (see online version for colours)

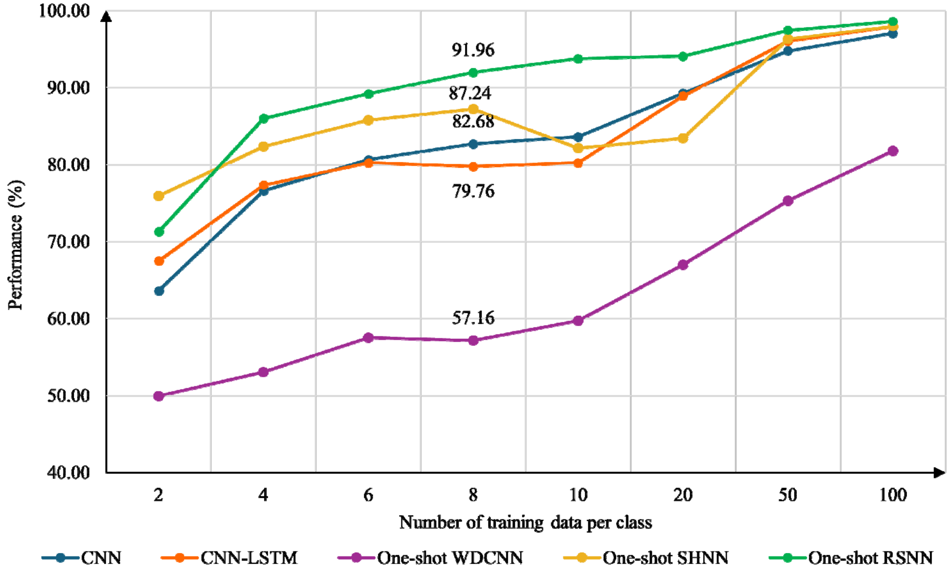
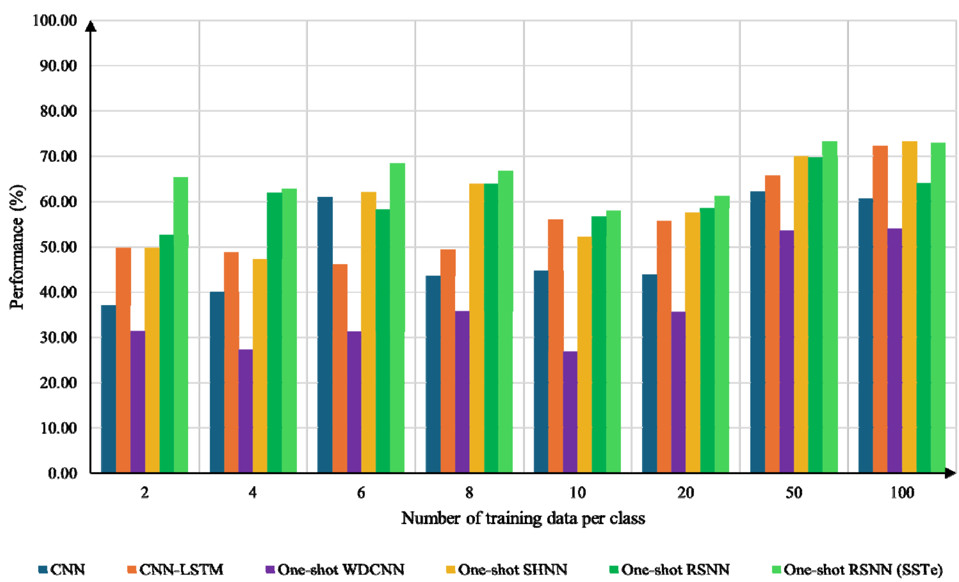
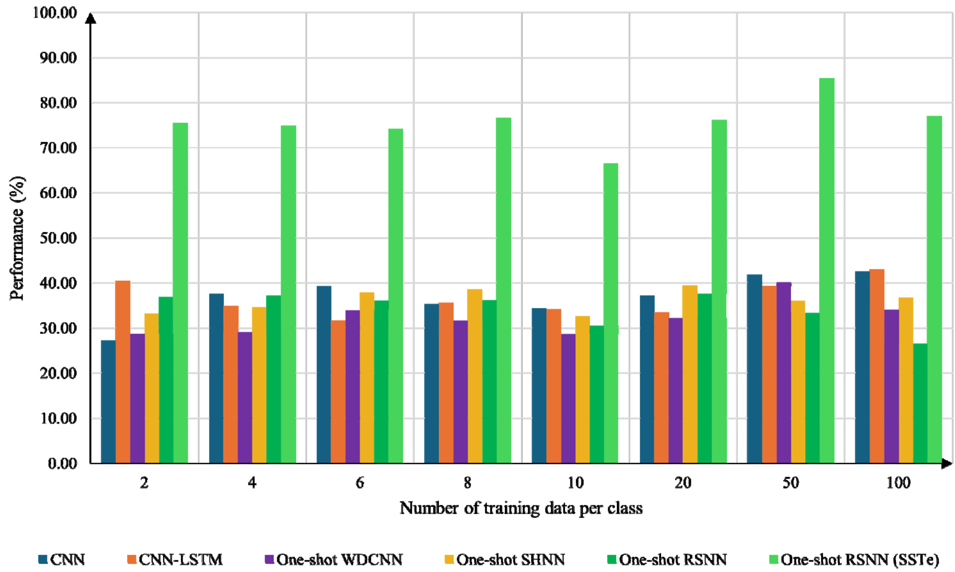


Figure 14 Graph of performance against training data number per class (condition 1 – cross load artificial fault test) (see online version for colours)



In the classification of artificial faults, the three model conditions cannot give good results due to the complexity and noise of the signal as shown in Figure 14. However, when using the target domain support dataset in natural fault detection, the one-shot learning model has a significantly better performance compared to the other two conditions. This indicates that the training dataset used in the model training is similar to the natural fault so the model can classify it accurately as shown in Figure 15.

Figure 15 Graph of performance against training data number per class (condition 1 – cross load natural fault test) (see online version for colours)



4.5.2 Condition 2 – test of combine class for fault condition

Figure 16 shows the graph of performance against training data number per class for a single load with only classes while Figure 17 shows the graph of performance against training data number per class for cross load. From the results shown in Figure 16, the CNN, CNN-LSTM, SHNN and the proposed one-shot learning model have similar performance with the increasing number of training data except for WDCNN cannot give good performance across the study. It indicates that both models face the same level of difficulty in classifying the data according to the classes. Both models need large data to achieve a good performance. One-shot learning is proved to have difficulties in classifying the different characteristics as one class due to the training method and the distance metrics used. For the cross-load analysis shown in Figure 17, both models do not give good results due to the different datasets used in the testing. The models face difficulties in extracting useful information from the segmented data.

4.6 Verification of the dataset

The CWRU dataset is a very clean bearing signal as it can show clear frequency results with just 2,560 sample points as shown in Figure 18 when compared to the full length of the signal data. The blue colour indicates the full length of the signal data while the red line shows the segmented 2,560 sample data points signal data. It shows the reason why the model can do the prediction and classification with extremely high performance with a low number of training data. It is hard to achieve for a normal dataset and a normal model. With this kind of data, the results are valid as the segmented data can give clear characteristics at the segmented length. This proves that DL can classify the faults without human intervention with the initial classification of the data.

Figure 16 Graph of performance against training data number per class (condition 2-cross load) (see online version for colours)

Figure 17 Graph of performance against training data number per class (condition 2 cross load) (see online version for colours)

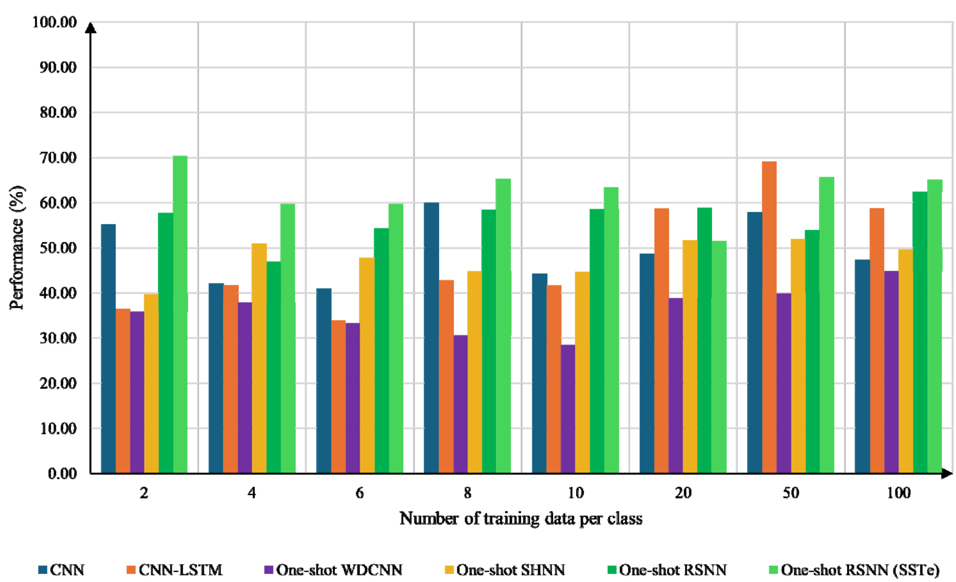


Figure 18 Frequency domain of full data length versus the 2,560 sample points length for the normal condition data with 0 HP in the CWRU dataset (see online version for colours)

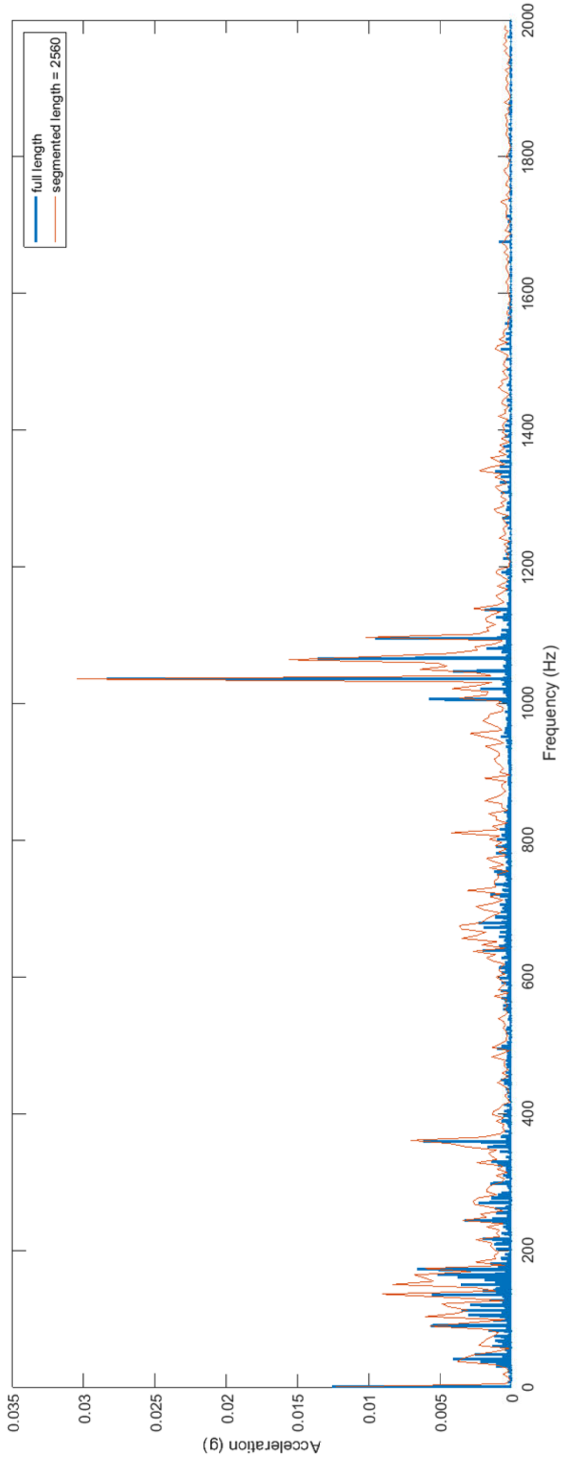


Figure 19 Frequency domain of full data length versus the 2,560 sample points length for the inner race fault of KI01 in PU dataset (see online version for colours)

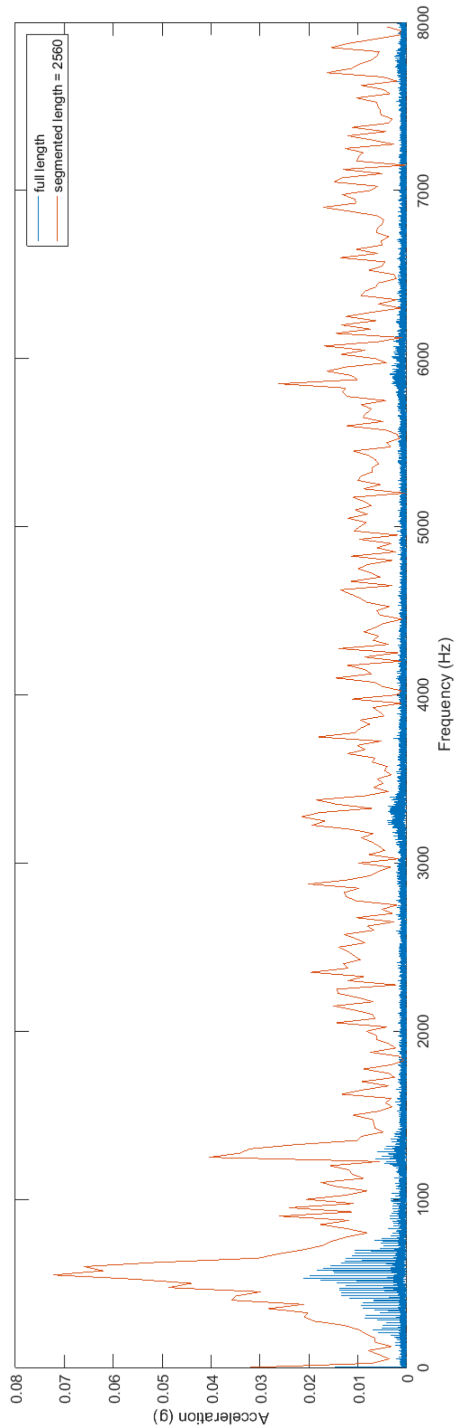


Figure 20 Comparison of frequency domain among KI01 and KI14 with the data length of 2,560 sample points for PU dataset (see online version for colours)

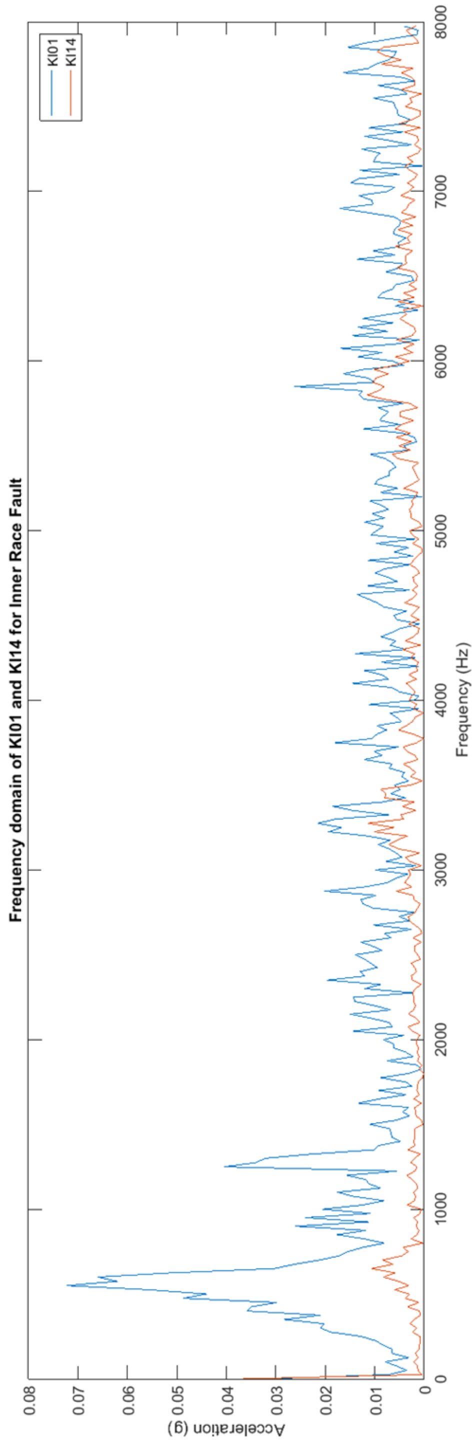
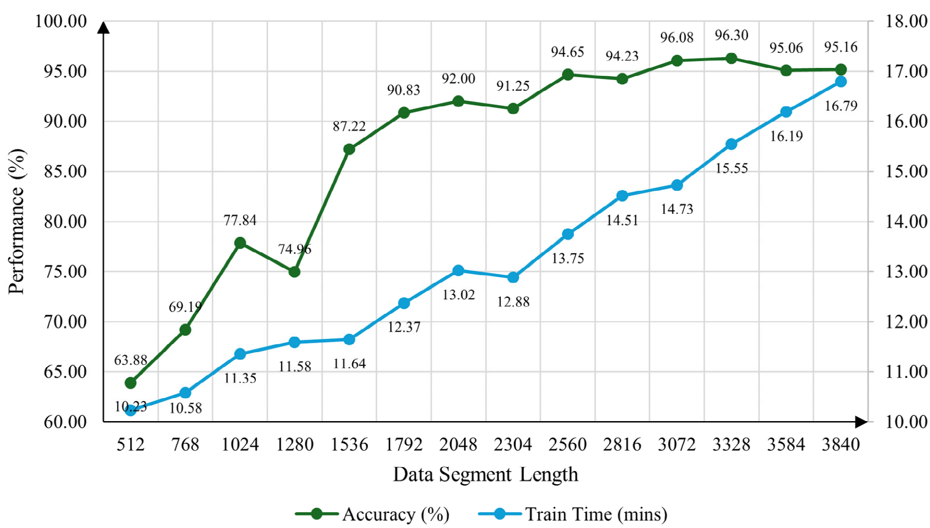


Figure 19 shows the frequency domain of full data length versus the 2,560 sample points length for the inner race fault of KI01 in the PU dataset while Figure 20 shows the comparison of frequency domain among KI01 (artificial fault) and KI14 (natural fault) with the data length of 2,560 sample points for the PU dataset. From the analysis of the PU dataset, the cross-load analysis is difficult to be done can be caused by the inappropriate segmented data range. The frequency domain obtained from the 2,560 sample points is not enough to give the correct frequency information for analysis purposes. This trait leads to good analysis in the single-load test but bad performance in the cross-load analysis. This characteristic can be observed in Figure 20 also. The totally different characteristics of the frequency domain will lead to different analyses in the model. It leads to the low performance of cross-load analysis for a model to recognise both inner race faults as the same class.

However, in the segmentation analysis based on CWRU data, the segment with 2,560 data points can give a good performance with 94.65% as shown in Figure 21 for the graph of performance and the train time against the data segment length. After the segmentation of the 2,560 data points, the increment of the results is not significant with the range of 1.65% of performance while the time to train the model keeps increasing from 13.75 mins for 2,560 data points to 16.79 mins for 3840 data points. The insignificance improvement with the increasing time and the data segmentation length shows that 2,560 data points are the best segmentation length to be used in the analysis. However, the different dataset has different suitable data segmentation lengths for the analysis. The one segmentation length of CWRU 0 HP data is 401 data points but for PU data is around 2,560 data points. The 2,560 data points used for CWRU data points include at least six revolution data points while for PU data is just one revolution of data points. Compared to the six revolutions of information, the one revolution of data points is not sufficient for the model to obtain useful information in the limited data condition.

Figure 21 Graph of performance and the train time against the data segment length for RSNN (see online version for colours)



Next, if we analyse the results based on the cross-load analysis, one-shot learning has a large difference in performance compared to CNN. SNN can classify the fault condition with its severity very accurately in a cross-load condition with the target domain support dataset in the comparison of similarity and dissimilarity. The cross-load analysis shows that the selection of the reference data in the support set is very important to determine the model performance. The good reference data in the support set can give the model a good performance of prediction. Other than that, the study of the cross-load analysis is to determine the probability of the single pre-trained DL model applied in another dataset of fault diagnosis. Due to the lack of well-segmented length of the data, the model faces difficulties in determining the cross-load situation for the artificial fault in condition 1 and condition 2 but has a significantly excellent result in the cross-load natural fault test in condition 1. The testing dataset used in the natural fault is a new dataset. However, the features extracted in the model training are similar to the testing dataset, so the testing dataset possesses a high dissimilarity in every fault feature. Thus, it gives a good performance in fault prediction. This proves that one-shot learning is capable of being applied in cross-load analysis with a good support set to predict and classify the fault. One-shot learning has also proved that it has good competency in the classification of the data with very limited data samples to achieve the results, especially with the clear classification of faults condition and severity.

5 Conclusions

In conclusion, the current stage of the one-shot learning model based on the RSNN model is able to give a good analysis of FDD for various conditions, especially in the clear division and classification of the dataset. RSNN with the designed structure work well with the cross domain and transfer learning tasks without pre-train the model. However, the one-shot learning model has a similar performance with CNN and CNN-LSTM models in the mixed-severity of the dataset. The analysis proves that the SNN model is more capable in the classification of the clear classification of condition and severity for the dataset. By using two types of datasets in the analysis, the analysis shows that the segmentation data length is important for the model to extract useful information. The suitable data segmentation length helps the model achieve good performance with the very limited data sample. In the overall analysis, by using a small data size, the one-shot learning model possesses an excellent prediction result in single-load conditions and cross-load conditions with good support set data.

Acknowledgements

The work was supported by the Fundamental Research Grant Scheme (FRGS) from the Ministry of Higher Education (MOHE) Malaysia, Grant no: FRGS/1/2023/TK02/UTM/02/11. This work was also supported in part by the Institute of Noise and Vibration Universiti Teknologi Malaysia (UTM) under the Higher Institution Centre of Excellence (HICoE) Grant Scheme under Grant R.K130000.7809.4J226.

References

Abid, A., Khan, M.T. and Iqbal, J. (2020) 'A review on fault detection and diagnosis techniques: basics and beyond', *Artificial Intelligence Review*, Vol. 54, No. 5, pp.3639–3664, https://doi.org/10.1007/s10462-020-09934-2.

Atanbori, J. and Rose, S. (2022) 'MergedNET: a simple approach for one-shot learning in Siamese networks based on similarity layers', *Neurocomputing*, Vol. 509, pp.1–10, https://doi.org/10.1016/j.neucom.2022.08.070.

Bromley, J., Guyon, I., LeCun, Y., Shah, R., et al. (1993) 'Signature verification using a 'Siamese' time delay neural network', *Advances in Neural Information Processing Systems*, No. 6, pp.737–744.

Cao, T., Law, M.T. and Fidler, S. (2020) *A Theoretical Analysis of the Number of Shots in Few-Shot Learning*, arXiv, Cornell University, https://doi.org/10.48550/arxiv.1909.11722.

Choudhary, A., Goyal, D., Shimi, S.L. and Akula, A (2018) 'Condition monitoring and fault diagnosis of induction motors: a review', *Archives of Computational Methods in Engineering*, Vol. 26, No. 4, pp.1221–1238, https://doi.org/10.1007/s11831-018-9286-z.

Cui, Z., Kong, X. and Hao, P. (2021) 'Few-shot learning for rolling bearing fault diagnosis based on residual convolutional neural network', *2021 4th International Conference on Artificial Intelligence and Big Data (ICAIBD)*.

Case Western Reserve University (CWRU) (2021a) *Apparatus & Procedures* | Case School of Engineering | Case Western [online] https://engineering.case.edu/bearingdatacenter/apparatus-and-procedures (accessed 20 June 2024).

Case Western Reserve University (CWRU) (2021b) *Normal Baseline Data* | Case School of Engineering | Case Western. Case School of Engineering [online] https://engineering.case.edu/bearingdatacenter/normal-baseline-data (accessed 20 June 2024).

Dai, X. and Gao, Z. (2013) 'From model, signal to knowledge: a data-driven perspective of fault detection and diagnosis', *IEEE Transactions on Industrial Informatics*, Vol. 9, No. 4, pp.2226–2238, https://doi.org/10.1109/tii.2013.2243743.

de Azevedo, H.D.M., Araújo, A.M. and Bouchonneau, N. (2016) 'A review of wind turbine bearing condition monitoring: state of the art and challenges', *Renewable and Sustainable Energy Reviews*, Vol. 56, pp.368–379, https://doi.org/10.1016/j.rser.2015.11.032.

Fang, Q. and Wu, D. (2021) 'ANS-net: anti-noise Siamese network for bearing fault diagnosis with a few data', *Nonlinear Dynamics*, Vol. 104, No. 3, pp.2497–2514, https://doi.org/10.1007/s11071-021-06393-4.

Honka, T. (2019) *One-Shot Learning with Siamese Networks for Environmental Audio Tampere*, Tampere University, Tampere, Finland.

Koch, G., Zemel, R. and Salakhutdinov, R. (2015) 'Siamese neural networks for one-shot image recognition', *ICML Deep Learning Workshop*, Vol. 2, No. 1 [online] https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf (accessed 20 June 2024).

Kumar, S., Mukherjee, D., Guchhait, P.K., Banerjee, R., Srivastava, A.K., Vishwakarma, D.N. and Saket, R.K. (2019) 'A comprehensive review of condition based prognostic maintenance (CBPM) for induction motor', *IEEE Access*, Vol. 7, pp.90690–90704, https://doi.org/10.1109/access.2019.2926527.

Lee, I., Cho, H., Kim, K., Lee, H., Kim, J. and Paik, J. (2024) 'Fault diagnosis of wind turbine bearings using Siamese networks', *2024 IEEE 19th Conference on Industrial Electronics and Applications (ICIEA)*.

Lessmeier, C., Kimotho, J.K., Zimmer, D. and Sextro, W. (2016) 'Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: a benchmark data set for data-driven classification', *European Conference of the Prognostics and Health Management Society*, 1 No. 31, https://doi.org/10.36001/phme.2016.v3i1.1577.

Li, C., Li, S., Zhang, A., Yang, L., Zio, E., Pecht, M. and Gryllias, K. (2022) 'A Siamese hybrid neural network framework for few-shot fault diagnosis of fixed-wing unmanned aerial vehicles', *Journal of Computational Design and Engineering*, Vol. 9, No. 4, pp.1511–1524, https://doi.org/10.1093/jcde/qwac070.

Liu, X., Chen, G., Wang, H. and Wei, X. (2023) 'A Siamese CNN-BiLSTM-based method for unbalance few-shot fault diagnosis of rolling bearings, *Measurement and Control*, Vol. 57, No. 5, pp.551–565, https://doi.org/10.1177/00202940231212146.

Ma, P., Zhang, H., Fan, W. and Wang, C. (2019) 'Early fault diagnosis of bearing based on frequency band extraction and improved tunable Q-factor wavelet transform, *Measurement*, Vol. 137, pp.189–202, https://doi.org/10.1016/j.measurement.2019.01.036.

Maas, A. and Kemp, C. (2009) 'One-shot learning with Bayesian networks', *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 31, No. 31, https://doi.org/10.1184/r1/6617375.v1.

Malhotra, A. (2023) 'Single-shot image recognition using Siamese neural networks', *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*.

Roy, S.K., Harandi, M., Nock, R. and Hartley, R. (2019) 'Siamese networks the tale of two manifolds', *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp.3046–3055, https://doi.org/10.1109/iccv.2019.00314.

Saufi, S.R., Ahmad, Z.A.B., Leong, M.S. and Lim, M.H. (2019) 'Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: a review', *IEEE Access*, Vol. 7, pp.122644–122662, https://doi.org/10.1109/access.2019.2938227.

Sharma, N., Gupta, S., Mohamed, H.G., Anand, D., Mazón, J.L.V., Gupta, D. and Goyal, N. (2022) 'Siamese convolutional neural network-based twin structure model for independent offline signature verification', *Sustainability*, Vol. 14, No. 18, https://doi.org/10.3390/su141811484.

Wang, C. and Xu, Z. (2021) 'An intelligent fault diagnosis model based on deep neural network for few-shot fault diagnosis', *Neurocomputing*, Vol. 456, pp.550–562, https://doi.org/10.1016/j.neucom.2020.11.070.

Wen, C., Xue, Y., Liu, W., Chen, G. and Liu, X. (2024) 'Bearing fault diagnosis via fusing small samples and training multi-state Siamese neural networks', *Neurocomputing*, Vol. 576, https://doi.org/10.1016/j.neucom.2024.127355.

Yip, K. and Sussman, G.J. (1997) 'Sparse representations for fast, one-shot learning', *National Conference on Artificial Intelligence* [online] https://dblp.uni-trier.de/db/conf/aaai/aaai97.html#YipS97 (accessed 20 June 2024).

Zhang, A., Li, S., Cui, Y., Yang, W., Dong, R. and Hu, J. (2019a) 'Limited data rolling bearing fault diagnosis with few-shot learning', *IEEE Access*, Vol. 7, pp.110895–110904, https://doi.org/10.1109/access.2019.2934233.

Zhang, Y., Pardo, B. and Duan, Z. (2019b) 'Siamese style convolutional neural networks for sound search by vocal imitation', *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 27, No. 2, pp.429–441, https://doi.org/10.1109/taslp.2018.2868428.

Zhang, Y., Li, S., Zhang, A., Li, C. and Qiu, L. (2022) 'A novel bearing fault diagnosis method based on few-shot transfer learning across different datasets', *Entropy (Basel)*, Vol. 24, No. 9, https://doi.org/10.3390/e24091295.

Zhao, X., Ma, M. and Shao, F. (2022) 'Bearing fault diagnosis method based on improved Siamese neural network with small sample', *Journal of Cloud Computing*, Vol. 11, No. 1, https://doi.org/10.1186/s13677-022-00350-1.