



Comparative studies of deep learning neural network architectures in fault diagnosis of rubber vibration isolators

**Adnan, Z.¹, Chang, V. C.¹, Ho, J. H.^{1*}, Chai, A. B.¹, How, H. C.¹, Chai, T. Y.²,
Kamaruddin, S.³**

¹ Department of Mechanical, Materials and Manufacturing Engineering, Faculty of Science and Engineering, University of Nottingham Malaysia, Selangor, Malaysia.

² Department of Computer Science, Faculty of Information and Communication Technology, Universiti Tunku Abdul Rahman, Kampar, Perak, Malaysia.

³ Technology and Engineering Division, Malaysian Rubber Board, Kuala Lumpur, Malaysia.

*Corresponding author email: JeeHou.Ho@nottingham.edu.my

Received: 16/07/2024

Revised: 25/01/2025

Accepted: 19/02/2025

Abstract

Automating fault diagnosis of machine components is crucial as it prevents unexpected downtime of a system that affects the operation and safety of the users. Deep learning architectures such as convolutional neural network (CNN) and long short-term memory network (LSTM) have been proven as prominent in training of sequential data due to their

robustness in classifying time series sequences and achieving state-of-the-art performance for effective fault diagnosis in structural health monitoring (SHM) systems. In this study, hybrid CNN-LSTM and U-Net (a CNN-based model arranged in U-shaped architecture), are employed to detect different levels of cracks in rubber vibration isolators. Cracks were induced at the interface between the steel and rubber to simulate a faulty scenario similar to a mechanical failure in industrial practice. The vibration of experimental platform supported by rubber isolators was induced by a motor driving an eccentric disk with varying speeds. Results revealed that the proposed U-Net architecture could achieve the best overall accuracy with decent computational time for training and classification. In addition, influence of data segmentation on classification accuracy, often overlooked in literature, was also investigated in this work. Findings showed that cleaner raw signals could be less prone to classification accuracy fluctuations.

Keywords: fault diagnosis; rubber vibration isolator; neural network; U-Net; Hybrid CNN-LSTM

1. Introduction

In recent years, there have been several fatal incidents due to structural damage raising alerts of infrastructure safety. Extreme events like terrorist attacks, vehicle impacts, and explosions often cause local damage to building structures and pose a serious threat upon one or more vertical load-bearing components failure, which results in progressive collapse of the major part or the whole structure (Adam et al., 2018). Rubber bearings are the most crucial part in a based isolated structure that shields structures from unwanted vibrations with its high

damping and stiffness (Zeng et al., 2023). However, the complex viscoelastic behavior of rubber poses challenges in precisely defining the dynamic characteristics which leads to uncertainties in practical applications (Gil-Negrete et al., 2006). Underestimating and overestimating the rubber mount remaining lifespan could lead to avoidable waste and risk of fatality injury, respectively. Thus, numerous studies have focused on mechanical aspects including fatigue, ultimate tensile strength, and yield strength of the rubber component to gain deeper insights into the anti-vibration properties (Belkhiria et al., 2020). Yet, the industry's traditional approaches frequently hinge on expensive and heavy experimental procedures. Lapčík et al. investigated the variation of dynamic stiffness and damping properties of rubber through servo-hydraulic systems at specific frequencies (Lapčík et al., 2001). Through this method, the researchers observed an increment of dynamic stiffness of materials along increasing static load and frequency ranged between 0.03 to 1 MPa and 10 to 100Hz respectively. With the advancement of artificial intelligence techniques, it is possible to predict the operating conditions of rubber materials without measuring their dynamic properties using scientific instruments. Hence, this study aims to conduct fault diagnosis in a rubber vibration isolator using a machine learning technique that reliant on vibration signals.

Deep learning, a subset of machine learning, holds significant prominence in structural health monitoring (SHM) applications primarily due to its ability to make predictions based on past behaviors learned. Numerous published studies have explored machine fault diagnosis employing intelligent diagnostic models like support vector machines (Hu et al., 2007), the k-nearest neighbours algorithm (Moosavian et al., 2013), and genetic algorithms (Tse et al., 2004). These traditional methods, with their straightforward and shallow structures, remain fundamental in fault diagnosis approaches. Nevertheless, these models struggled to effectively identify nonlinear relationships within vibration signals, contributing to their shortcomings.

Another significant limitation in employing the classical machine learning approach is the need to redesign pre-processing methods for distinct fault characteristics (Chen et al., 2021).

Consequently, the industry strives to discover a universally applicable solution suitable for addressing typical machine failures (Guo et al., 2018). Moreover, for effective prevention of actual component failures, the SHM system must accurately detect failure symptoms, providing users with ample time to troubleshoot or replace the component. Deep learning negates the requirement for manual feature engineering, displaying potential in precise classification tasks without expert input. Janssens et al. suggest that their convolutional neural network (CNN) method for fault detection in rotary machines is notably more straightforward than the traditional engineering-based approach, which relies on expertise in vibration analysis (Janssens et al., 2016). In that work, CNN and classical machine learning models were trained with raw vibration data to detect failure conditions such as outer raceway faults and different levels of lubricant degradation. The results justified the proposed method could achieve high classification accuracy with less domain expertise compared to classical models.

Meanwhile, Zhou et al. proposed a CNN model for the prediction of remaining useful life (RUL) and fault diagnosis of bearings in rotating machinery (Zhou et al., 2020). The CNN model was trained by time and frequency features of vibration signals extracted through Short-time Fourier Transform (STFT). Classification results using CNN were compared with other classical machine learning methods. It was concluded the proposed method could achieve accuracy of 99.45% in rolling bearing remaining useful life prediction. It could also better identify different failure modes with higher accuracy compared to conventional approaches. Furthermore, Hoang et al. in their multi-model for concrete properties prediction study reported that machine learning capability was limited only for the trained dataset only, which cannot be applied comprehensively in various applications (Hoang et al., 2024).

On the other hand, long short-term memory (LSTM) network is known for training sequential data, which is ideal for time-based vibration signals. Guo et al. proposed a pyramid LSTM network for machine tool condition monitoring (Guo et al., 2022). The network is constructed based on multi-layer frequency spectrum of cutting signals and it could achieve good results under unknown tools and milling parameters. Meanwhile, Ma and Mao proposed a convolution-based long short-term memory network (CLSTM) for RUL prediction on bearings (Ma and Mao, 2021). The proposed model performs convolution operation in input-to-state and state-to-state transitions of LSTM layers and learns both time-frequency as well as temporal information of signals. Results showed that the model outperformed other deep learning algorithms in the field. Other studies also suggested hybrid CNN-LSTM models could perform well on time series data, as seen in LSTM-FCN networks (fully convolutional neural network) (Karim et al., 2017; Karim et al., 2019; Karim et al., 2019), 2D CNN-LSTM network (Wang et al., 2023) and CNN-LSTM with skip connections (Wahid et al., 20).

U-Net architecture, originally developed for image segmentation by Ronneberger et al. (Ronneberger et al., 2015) based on the fully convolutional network proposed by Long et al. (Long et al., 2015), is expanding its field of applications to analogue signal processing and classification. Due to small training set and high segmentation accuracy offered, there is a booming development of U-Net research works for various applications in signal and vibration analysis. For example, Zhang et al. developed a method called Threshold Acquisition U-Net (TA-UNet), enhancing feature extraction which learns adaptive thresholds effectively to mitigate noise interference and improve multi-scale extraction (Zhang et al., 2023). In early fault diagnosis, TA-UNet is used in a two-step process involving model training with noise-added vibration signals and utilizing the trained model to extract fault features for diagnosis of rolling bearings. The efficacy of TA-UNet's feature extraction capability was demonstrated via denoising simulated rolling bearing signals with successful early fault diagnosis using open-

source datasets. In another work, U-net++ was proposed for gearbox fault diagnosis using vibration signals in time-frequency domain (Zhang and Chen, 2023). Variants of U-Net architecture have been proposed for denoising applications as well, such as elevator vibration signals (Xie et al., 2024), seismic (Zhao et al., 2023; Wu and Stewart, 2023) and structural vibration data (Shen et al., 2023).

Furthermore, Koszewski et al. introduces a music mixing technique that automatically blends distinct raw recordings with high-quality outcomes across various music genres, showcasing a novel deep model rooted in 1D Wave-U-Net autoencoders and trained on a customized database, subsequently demonstrating that mixes generated through this approach are comparable in quality to professionally prepared mixes through objective evaluations and listener tests (Koszewski et al., 2023). Despite the capability of convolutional encoder-decoder architectures to capture overall patterns in the context of separating singing voices, there are several challenges related to the uncertainty about the loss of local details during compression and the requirement of using large substantial training datasets (Jansson et al., 2017). Time-series data employed U-Net (which is called U-Time) has been proven as a better alternative than the prominent technique combining the convolutional and recurrent neural network (Perslev et al., 2019). The major strength of feed-forward system owned by U-Time has overcome tuning and optimization difficulty issues of recurrent model. Without the need to optimize the hyperparameter or architecture, this model is more robust in the training process.

In this work, two deep learning models were proposed, namely hybrid CNN-LSTM network (with sequential and parallel connections) and U-Net model, that feed different segmentations of raw vibration data to train the models to detect cracks in rubber vibration isolators. U-Net architecture is considered in this study due to its capability of configuring deeper feature extraction layers. Meanwhile, hybrid CNN-LSTM network utilizes the strength of CNN to extract features of the input and consequently LSTM layer could study the patterns

more efficiently (Li et al., 2019). The network architectures are described in Section 2. In Section 3, a series of experiments were conducted with different fault modes and the results were analyzed by varying signal segmentation length. In section 4, the results were discussed where it was justified that the proposed methodology could achieve a high classification accuracy with less domain expertise in terms of configuring the neural network and hyperparameter tuning. Section 5 concludes this work which highlights that it is possible to adopt a deep learning approach to identify the severity of crack level in rubber isolators. Another contribution of this study is the investigation of the performance of the proposed neural network model by varying the segmentation length of the vibration signals.

2. Neural networks architectures

2.1 Convolutional versus recurrent neural networks

2.1.1 Convolutional Neural Network (CNN)

A convolutional neural network (CNN) is a feedforward deep learning model that is popular for its diagnostic and predictive technique of processing time-series data and image data. The mathematical theory behind is a linear operation known as convolution. A convolutional network convolves its input by general matrix multiplication in its layer and outputs it to the next layer. CNN has 2 unique benefits compared to other neural networks. The first one is the sparsity in the connection, it could reduce the number of parameters and training time by connecting each kernel with its local patch of previous features. The second feature is weight sharing among the hidden layers, this further reduces the learnable parameters and the complexity of the network. In contrast to the common 2D CNN, 1D CNN convolves over the time dimension instead of the spatial dimension.

2.1.2 Long Short-Term Memory (LSTM)

A Recurrent Neural Network (RNN) is one of the earliest feedback neural networks where its simplest design is having recurrent connections between hidden units to produce an output at each time step. It has an internal state (memory) to process variable-length sequences of inputs. Following more research studies, a Long Short-term Memory Neural Network (LSTM) is created which is also known as the “gated recurrent neural network” because it has similar architecture compared to RNN. RNN might encounter gradient vanishing and exploding issues in a complex scenario or it requires longer training time due to the matrix multiplication mechanism. Hence, gated units such as input, forget and output gate functions are introduced to create a self-loop condition to avoid very large or zero gradients (Hochreiter and Schmidhuber, 1997). The LSTM layer has introduced an intermediate feature known as cell state to make a self-loop condition. This feature allows the backpropagation to perform element-wise multiplication instead of matrix multiplication. The cell state is regulated by the Tanh function to decide whether it should be carried to the next memory cell. These features greatly reduce the learning time of the neural network and prevent the gradients from vanishing and exploding.

2.1.3 Hybrid CNN and LSTM network

By incorporating a hybrid CNN-LSTM network, the CNN module could enhance LSTM performance by extracting time-invariant features of the input. Meanwhile, the LSTM module could accurately capture the long-term dependency of the signal by studying the patterns in the past (Li et al., 2019). The hybrid network architecture could connect in a series or parallel manner. In a sequential setup, the LSTM layer is connected to the last layer of the CNN architecture before passing through the fully connected layer. In a parallel setup, the LSTM architecture is connected to the CNN architecture via a concatenation layer to combine the inputs into a mixed output.

Bidirectional LSTM (BiLSTM) is used to ensure the layer could capture the temporal dependencies between two directions instead of the conventional approach which is only the positive direction. This approach allows the network to learn from the complete time series at each time step. For the output layers, the fully connected layer combines all the features learned by every neuron in the previous layer to classify the signals. Hence, the output of this layer is equal to the number of classes of the data set. The softmax layer computes a probability distribution by using the softmax transfer function, it will then pass through the cross-entropy loss function in the classification layer to identify the training accuracy of the classifier by predicting the correct classes using current weights and biases.

Rectified Linear Unit (ReLU) function is used in CNN where it allows the model to perform better and easier optimization that uses gradient-based methods. It applies a threshold operation to each element of the input and set any value less than zero to zero, which means it will not saturate even if the input values are large. CNN also use global average pooling for 1D data, unlike the maximum and minimum pooling average pooling performs down-sampling by outputting the average time series data. This is to avoid discarding the peak amplitudes in the vibration signal.

Figure 1 shows the proposed hybrid architecture of 1D CNN and LSTM used for time series classification. The sequence input layer is created in the first layer where the signal data is input into the network and then separated into multiple segments based on the particular sizes. At this stage, for the sequential setup (indicated by pink arrows), the input is transferred into convolution 1d layer followed by biLSTM layer before finalized to fully connected layer. Meanwhile, the parallel setup (indicated by blue arrows) transfers the input data to both 1D CNN and biLSTM layers simultaneously which then concatenates the outputs form both layers before classification. During training progress, errors due to the wrong classification is common in the first ten epochs. The optimizer will then evaluate a gradient using the entire

training set and minimize the loss function via backpropagation. Thus, the weights and biases are updated to compute a better classification. The same process is repeated until the training accuracy is converged.

A possible variant of the sequential hybrid CNN-LSTM network is to connect the LSTM block ahead of the CNN module (i.e. LSTM-CNN). However, such setup recorded poorer performance than the CNN-LSTM configuration in terms of both accuracy and training period (Aksan et al., 2023). This is because of the nature of CNN that offers faster computations as it is performed in parallel compared to LSTM that processed input sequentially. So, the LSTM-CNN network pathway requires the entire input signal to be analyzed by LSTM first instead of CNN, which negatively affects computational processing efficiency (Alshingiti et al., 2023).

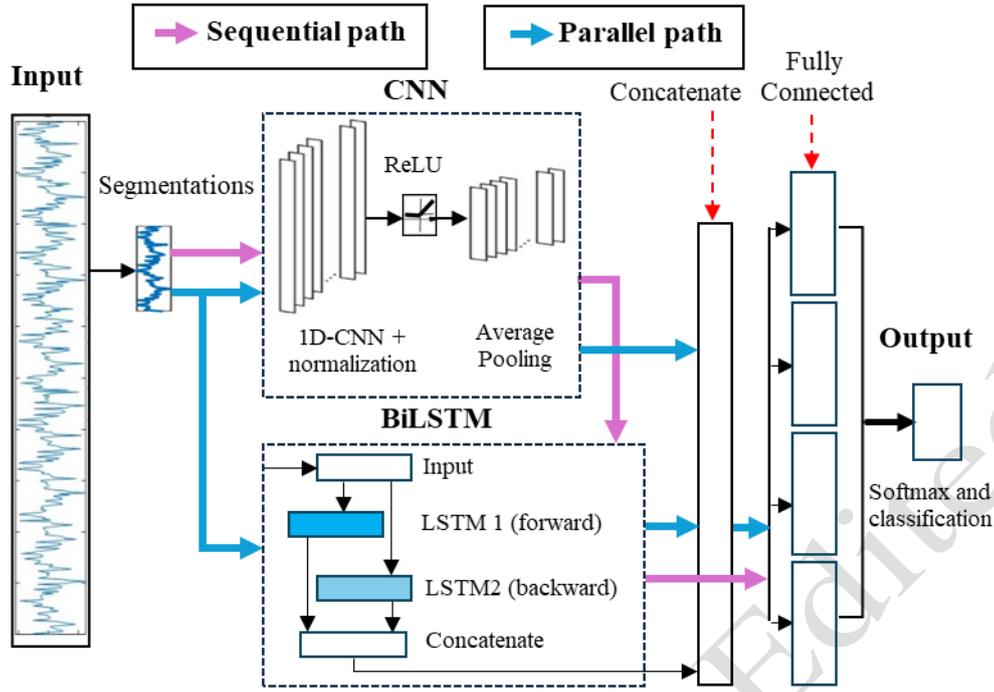


Figure 1: Neural network architecture of the proposed hybrid CNN-LSTM model.

2.2 U-Net architecture

In a U-Net, utilizing a fully convolutional neural network, a series of encoder layers is used to reduce the signal size by half while doubling the number of channels, creating a compact representation. In a deconvolutional network, a stack of convolutional layers where each layer halves the size of the signal but doubles the number of channels, encodes the signal into a small and deep representation. That encoding is then decoded to the original signal size by a stack of up-sampling layers (Jansson et al., 2017). This compressed version is then expanded back to the original signal size using decoder layers. As in U-Net with 4 layers (U-Net4) illustrated in Figure 2, the encoder contains a total of three downsampling stages. Each stage consists of two 1D-convolution operations (kernel size = 3, strides = 1, and same padding) and a max-pooling operation (pool size = 2). The output of each stage in the encoder is sent to the corresponding decoder layer for further convolutional operations.

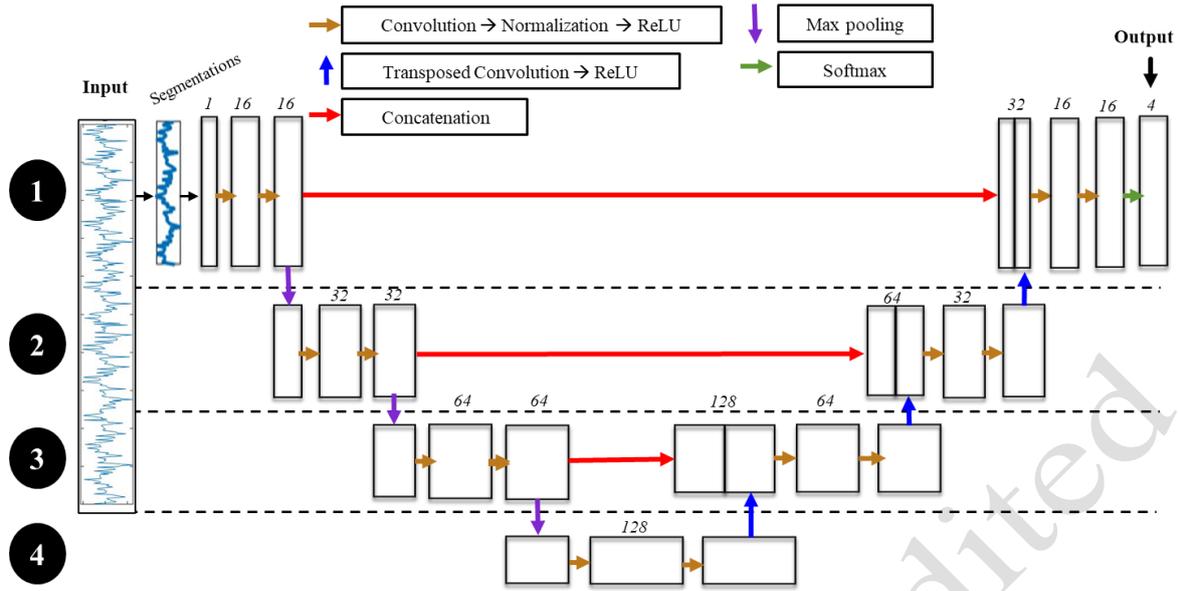


Figure 2: The proposed 1D U-Net4 architecture.

To better characterize the input data with different resolutions, the output of the lower layer of the decoder concatenates with the output of the upper-level layer in the encoder after an up-sampling process which recovers the feature map size. This concatenation process enables the model to perceive low-level and high-level features and prevents loss of spatial information. In some cases, the larger size of output from convolution at the encoder part will be central cropped to match the size of the same level at the decoder part (Ronneberger et al., 2015). However, this phenomenon is not observed in this work with the use of padding after every single convolution process, which maintains the signal data size to be equivalent before and after convoluted. These concatenated signals are then proceeded with convolution process followed by transposed convolution for up-sampling process. Lastly, the data were sent to the fully connected layer and the output classes are obtained with their own prediction accuracy.

Inspired by this network, U-Net was adopted as the backbone of deep neural network for the classification task. The original U-Net architecture has been modified to cater for 1D vibration signal instead of 2D image data by replacing the 2D convolution operations with 1D convolution, as expressed in Equation (1):

$$y(n) = x(n) * w(n) = \sum_{k=-m}^m x(k) * w(n - k) \quad (1)$$

where x is the input data sequence, y is the output of the convolution and w is the convolution kernel of size m . This kernel w slides over and integrates with the signal x to obtain feature maps. In this work, the U-Net is configured with 2 and 4 layers separately and named as U-Net 2 and U-Net 4. Same as the original U-Net network structure, encoder and decoder are the main components of the U-Net2 and U-Net4, where the encoder is designed to extract high-resolution features of the input data.

3. Methodology

3.1 Vibration platform setup

Figure 3 (a) shows the experimental setup with an enlarged view of the accelerometer (3 axis ADXL335) mounted on a lower wooden platform. Figure 3 (b) shows the schematic diagram of the experimental setup. The accelerometer is placed close to the faulty rubber vibration isolator.

As shown in Figure 3(a), the experimental setup consists of two wooden platforms stacked on top of each other with four rubber vibration isolators located at the four corners of the intermediate layer. A 12V DC motor (RS PRO 321-3192), eccentric mass and encoder (KÜBLER 8.KIS40.1362.1024) are assembled (Figure 3(c)) and surrounded by the acrylic plates at the sides and aluminium plate on the top for safety precaution. Encoder (Rotary Encoder 500 PPR) was mounted on the shaft to measure the rotational speed of the motor. Two motor speeds of 880rpm and 1160rpm were investigated in this study. The signals were

recorded by the accelerometer and then transferred to a local desktop via Raspberry Pi 4 board (R.Pi) model B.

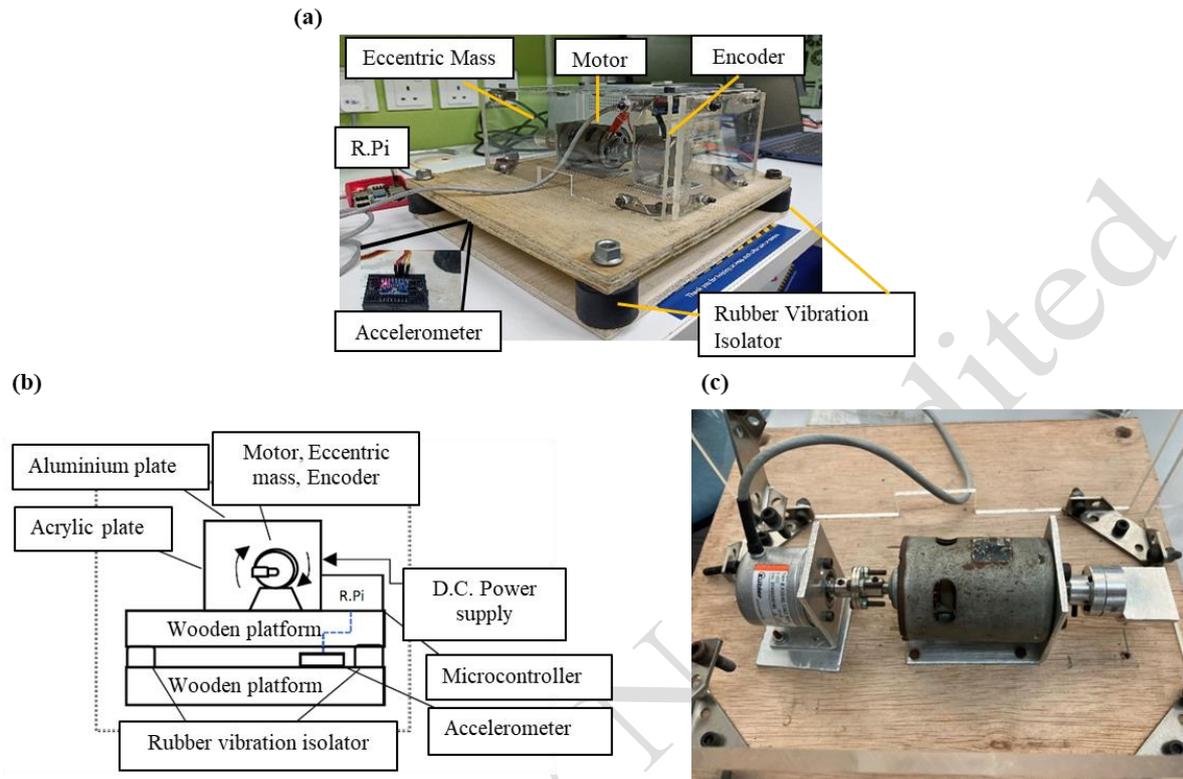


Figure 3: (a) The vibration test rig with the enlarged view of the accelerometer; (b) the schematic diagram of the test rig and (c) the DC motor, motor couplings and eccentric disc.

Rubber vibration isolators were used as the damper in this vibration system. A paper cutter is used to induce cracks at the interface between the steel plate and rubber material at a depth of 30mm. This method was intended to mimic a mechanical failure similar to the industrial practice. The location of crack initiation and propagation closely matches the findings from Luo based on finite element analysis and experimental cyclical fatigue test (Luo, 2021).

4 scenarios of vibration signals were acquired. In the faulty scenarios, only one rubber isolator will be replaced in the vibration platform. 1) Healthy scenario (H): contain undamaged rubbers. 2) Faulty scenario (F): contains rubber with one cut. 3) Faulty mildly destroyed

scenario (FM): contains rubber with one cut at the top and another at the bottom interface. 4) Faulty destroyed (FD): contains rubber with multiple severe deep cuts. Steady-state vibration signals were recorded. Figure 4 shows photographs of various crack scenarios. The same setup was run ten times, each trial containing 1 minute, with a sampling rate of 1kHz. Thus, there are 80 vibration signals (4 scenarios x 10 dataset x 2 motor speeds) each of them has a signal length of 60k.

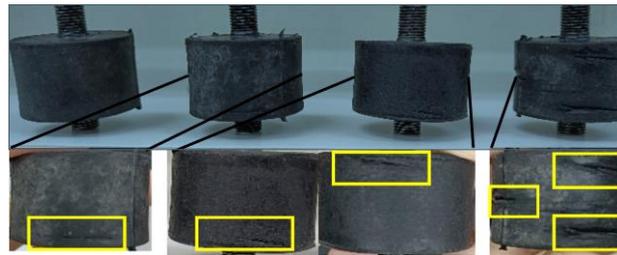


Figure 4: (from the left) Healthy, Faulty, Faulty mildly destroyed, Faulty destroyed, the enlarged views of the faulty scenarios. The cracks are highlighted in the yellow boxes.

3.3 Neural network configurations

Of all 10 datasets for each setup, 8 datasets are divided into 6 training and 2 validation datasets. Another 2 datasets are used for testing. The total signal length for training samples are 1.44M (4 scenarios x 6 training datasets x 60k signal length per trial). Each training set contains a segmented data length of 100. The amount of training data used in this study is comparable to some related studies, such as (Chen et al., 2021) and (Zhang et al., 2023). Therefore, there are a total of 14400 training sets. A 128 mini-batch size of over the training set of 14400 is used. Hence, the number of iterations per epoch is 112. The max epoch used is 60. But the training progress is stopped once it is converged up to 0.01 loss threshold, as this stopping criterion avoids overfitting which may reduce the classification accuracy. Also, the training

progress will be stopped manually if the training accuracy does not show noticeable changes within 10 epochs. These tuning settings are summarized in Table 1.

Table 1: Neural network training hyperparameter

Hyperparameter	Settings
Optimizer	ADAM
Max epochs	60
Validation frequency	10
Mini batch size	128
Initial learning rate	0.01
Loss threshold	0.01

In the vibration platform, there might be random noises generated by the system which will reduce the classification accuracy as there could be interference in the feature extraction inside the neural network. Hence, the optimizer chosen was Adam (adaptive moment estimation). It is an optimization algorithm that can adjust the learning rate of each parameter by estimating the first moment and the second moment of the gradient. This feature allows it to improve the training and convergence speed for vibrational signals.

In this work, the proposed 1D-UNet contains 1.8×10^5 parameters as detailed in Table 2. The number of parameters are contributed by the number of filters applied that produced an equivalent number of channels output through the convolution followed by batch normalization processes. For instance, at layer number 4, the first convolution (Conv_1) produced 3 (filter size of 3×1) \times 64 (weight of each channel from the previous layers) \times 128 (weight of each channel of the current layers) + 128 (bias) + 128 (normalization offset) + 128 (normalization

scale). As a result, there are a total of 24960 learnable parameters at this layer, which then summed up until the output layer.

As such, different amount of layers in U-Net gives a huge gap in the total number of parameters learned, as U-Net with 2 layers (U-Net 2) produces only 9.7×10^3 parameters, incredibly smaller than U-Net 4. Furthermore, it is seen from the output dimension that the size of the data signal getting reduced by half after interlayer process, which is due to the max-pooling applied at each layer for the encoding part of the U-Net. Conversely, the compressed signal data at layer 4 were doubled back via transposed convolution at each interlayer for decoding part.

Table 2: Learnable parameters and output dimension of the proposed U-Net 4 network

Layer	Name	Filter	Total parameter	Output dimension
1	Input	-	-	(14400, 100, 1)
	Conv 1	16	96	(14400, 100, 16)
	Conv 2	16	816	(14400, 100, 16)
2	Conv 1	32	1632	(7200, 100, 32)
	Conv 2	32	3168	(7200, 100, 32)
3	Conv 1	64	6336	(3600, 100, 64)
	Conv 2	64	12480	(3600, 100, 64)
4	Conv 1	128	24960	(1800, 100, 128)
	Conv 2	128	49536	(1800, 100, 128)
	T-Conv 4	64	24640	(3600, 100, 64)
3	Conv 1	64	24768	(3600, 100, 64)
	Conv 2	64	12480	(3600, 100, 64)
	T-Conv 3	32	6176	(7200, 100, 32)
2	Conv 1	32	6240	(7200, 100, 32)
	Conv 2	32	3168	(7200, 100, 32)
	T-Conv 2	16	1552	(14400, 100, 16)
1	Conv 1	16	1584	(14400, 100, 16)
	Conv 2	16	816	(14400, 100, 16)
	Fully Connected	-	68	(14400, 100, 1)

4. Results and discussions

4.1 Neural network performance

The overall performance of different types of neural networks training and classification for both low and high rpm signal data are presented in Table 3. In terms of accuracy, LSTM marks the lowest performance of classification in either 880 rpm (88.4%) or 1160 rpm (90.8%) although the validation accuracy is adequate among all neural networks compared. Meanwhile, CNN shared the worst performance with LSTM at low motor speed 880 rpm, achieving 88.4% average classification rate among 4 classes. At 1160 rpm, the CNN model improved to 94.2%. Comparatively, hybrid CNN-LSTM networks showed better consistency in both motor speeds, with parallel CNN-LSTM network performing slightly better than sequential CNN-LSTM, achieving 91.1% at 880 rpm and 93.6% at 1160 rpm. This shows the effectiveness of hybrid architecture to reliably commit on various scenarios that occur in real applications.

Of all the neural networks tested in this work, U-Net shows the best accuracy performance for both motor speeds. Employing convolution technique, this architecture performs well even with the simplest 2 layers architecture (U-Net 2) to achieve a similar performance with hybrid CNN-LSTM for 880 rpm condition at an average classification accuracy of 89.1%. Increasing the U-Net layer looks promising in this context where U-Net 4 achieved the highest accuracy of 92.9% for 880 rpm motor speed. Nonetheless, although U-Net 4 achieved slightly better overall accuracy compared to U-Net 2, it sacrifices a lot of computational processing time with more than eighteen times longer training period and twelve times for classification period than U-Net 2. A higher classification period might be detrimental when it comes to applications that require fast real-time processing or monitoring.

Table 3: Performance comparison of different neural networks for (a) 880rpm; (b) 1160rpm

(a)

Accuracy (%)	Class	Neural Networks					
		CNN	LSTM	Sequential CNN-LSTM	Parallel CNN-LSTM	U-Net 2	U-Net 4
	H	90.0	84.6	86.8	83.6	93.7	91.1
	F	80.2	79.0	88.2	88.0	87.8	93.6
	FM	99.7	99.8	100	99.6	100	100
	FD	83.8	90.0	88.7	93.2	75	86.8
Average accuracy (%)		88.4	88.4	90.9	91.1	89.1	92.9
Validation accuracy (%)		92.0	96.4	90.4	92.5	97.3	95.9
Training period (s)		69	1278	2028	1140	115	2096
Classification period (s)		0.64	8.07	1.61	3.78	2.59	32.71

(b)

Accuracy (%)	Class	Neural Networks					
		CNN	LSTM	Sequential CNN-LSTM	Parallel CNN-LSTM	U-Net 2	U-Net 4
	H	91.2	82.5	90.6	87.8	92.8	91.1
	F	94.3	82.2	95.1	92.3	94.8	96.7
	FM	93.0	99.5	85.8	98.9	99.7	97.8
	FD	99.1	98.8	99.8	95.4	100	99.7
Average accuracy (%)		94.40	90.75	92.83	93.60	96.83	96.33
Validation accuracy (%)		98.06	94.69	93.83	97.15	95.67	99.69
Training period (s)		76	659	114	285	25	280
Classification period (s)		0.97	9.12	1.23	2.15	4.17	26.69

At higher motor speed of 1160 rpm, both U-Net 2 and U-Net 4 outperformed other models with 96.8% and 96.3% average accuracy respectively. This finding indicates that the higher excitation frequencies could be more effectively analyzed by simpler architecture of U-Net. This could be due to training data containing more information as there are more vibration

cycles within the same segmentation length compared to lower excitation frequencies. Therefore U-Net 2 could be trained as effectively as U-Net 4. It only took 25 seconds to train the input data, almost 10 times faster than U-Net 4. In terms of classification speed, U-Net 2 required an average of 4.17 s compared to 26.29 s for U-Net 4. In this context, it is worth noting that the conventional CNN marked the fastest training and classification times among all neural networks compared in the low motor speed scenario. As for the high-speed scenario, CNN achieved the fastest classification time and ranked second in training time after U-Net 2. This shows that CNN model could be a good candidate for applications which require fast processing speed, albeit slightly underperformed in terms of accuracy compared to U-Net models.

Overall, it can be concluded that U-Net 4 achieved the best performance at 880 rpm while U-Net 2 is the best at 1160 rpm. In applications where real-time monitoring is needed, however, the U-Net 2 and CNN might be considered as well due to the fast performance with adequate accuracy offered.

4.2 Analysis of signal segmentation for U-Net models

As U-Net models demonstrated robust performance, this section further analyzes the effects of segmenting input signals on the classification accuracy. Due to deeper data feature learning along the multiple U-Net layers, U-Net 4 model was chosen to be evaluated by using raw input signal with various segmentation lengths from 200 ms to 2000 ms with an increment of 200 ms. Figure 5 presents the overall performance result of low and high motor speed vibration signals in terms of accuracy for each class along with training and classification periods.

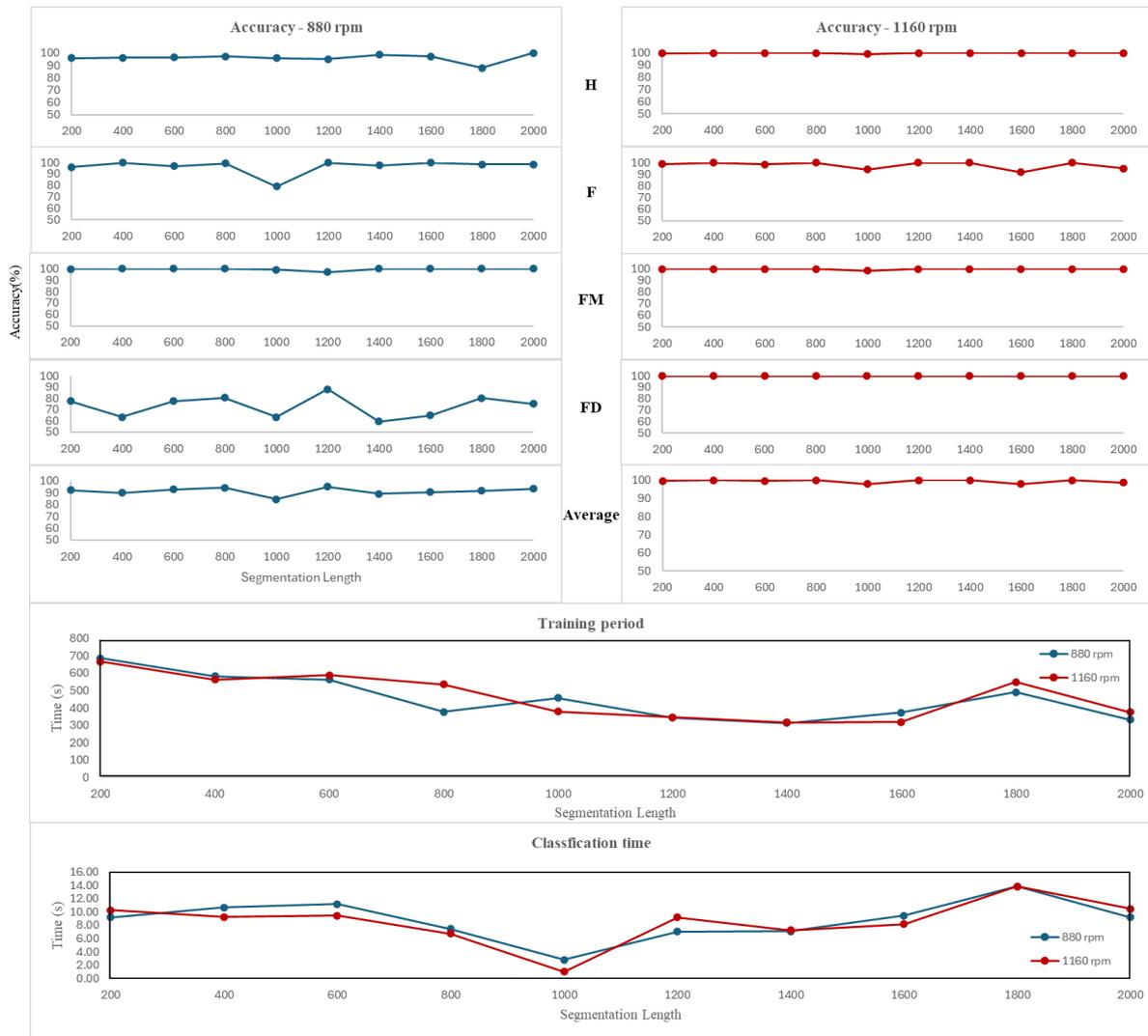


Figure 5: U-Net 4 accuracy, training period and classification period by varying signal segmentation length

For the low motor speed scenario (880 rpm), it is apparent that classification accuracy is more sensitive to the variation of input signal segmentation length compared to high motor speed scenario. This indicates finding optimal segmentation length is crucial in this regard. Among all classes, faulty mildly destroyed (FM) class achieved consistently high accuracy across all segmentation lengths. The performance of healthy (H) class was similar to that of the FM class except for a slight drop of accuracy at the segmentation length of 1800 ms. In the faulty (F) class, a larger drop of accuracy at 1000 ms segmentation length was experienced. Meanwhile for faulty destroyed (FD) class, the trend shows more fluctuations with the highest

accuracy of 88% observed at 1200 ms segmentation length. Overall, this outcome makes 1200 ms the best segmentation length across all classes for the low 880 rpm motor speed scenario.

Segmentation length refers to the period in which a signal is divided for processing. In the low 880 rpm scenario, the motor's vibrations may have periodic patterns or cyclic behavior, especially for the FD class where results showed that the classification accuracy is sensitive to the segmentation length. To provide further insights into this phenomenon, the vibration signal is analyzed using empirical mode decomposition (EMD). The EMD, also known as a component of the Hilbert-Huang transform, was developed based on the simple assumption that a signal is composed of a series of intrinsic modes of oscillations or functions (IMF). EMD is popular in analyzing non-linear and non-stationary time series signals and eliminates unnecessary noises by decomposing them into a pre-determined number of IMFs and a residual signal.

Figure 6 shows the EMD for FD class with 1200 ms segmentation window. The mechanical vibration frequency is usually observed in low frequency range, in this case, the main vibration frequency was around 17.52 Hz as shown in the frequency spectrum of IMF4.

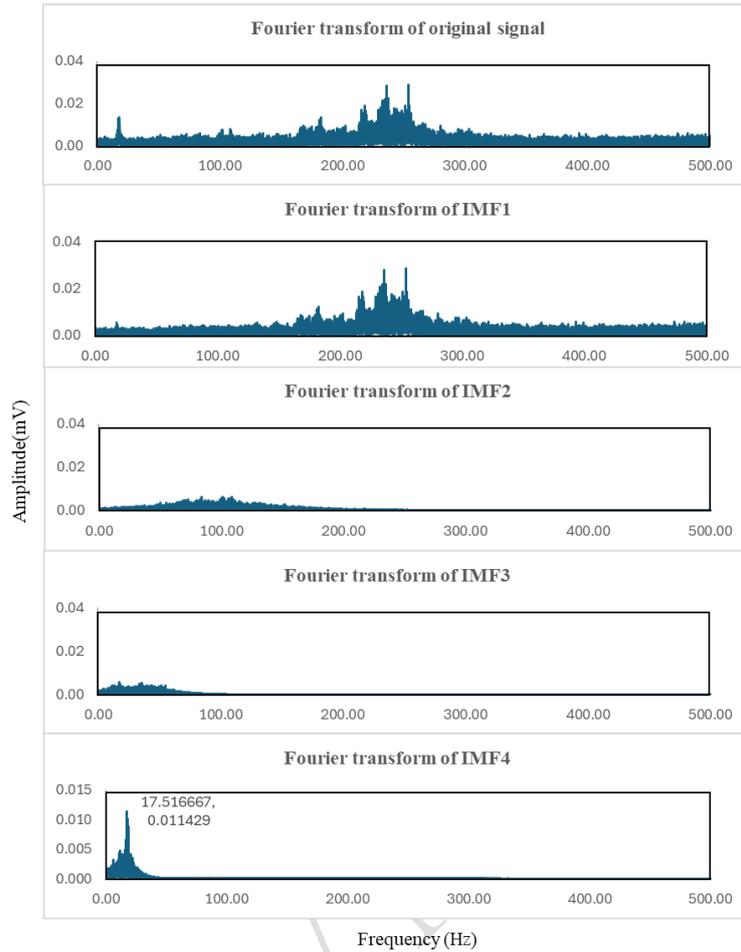


Figure 6: Empirical mode decomposition (EMD) of 880 rpm motor vibration signal for FD class

To provide further insights into the fluctuations in classification accuracy for FD class, lower accuracies were recorded beyond the 1200 ms segmentation length, with the decrement of 32.4% to the lowest point at 1400 ms followed by a slight increment thereafter. This may be attributed to the drawback of U-Net that requires larger dataset for effective model training [17]. Also, a larger signal segmentation length causes a lower amount of dataset to be used during the training process. In this context, it is essential to note that 880 rpm motor speed is equivalent to an excitation frequency of 14.6 Hz and this translates to roughly 68 data points acquired for each cycle at a data sampling rate of 1 kHz. Consider a signal segmentation length of 10 cycles, equivalent to 680 data points (680 ms), the classification accuracy indeed was not

the highest as observed in Figure 5, between the range from 600 ms to 800 ms. This suggests that choosing a segmentation length that tallies with the excitation frequency may not yield the optimal classification performance.

In the case of high motor speed scenario (1160 rpm), the average accuracy was rather consistent and not sensitive to the segmentation lengths, it also recorded higher overall classification accuracy compared to the low motor speed scenario. The reason for this phenomenon may be attributed to the noise appearance of the raw signal output from both motor speeds as visualised in Figure 7. It is hard to recognize the numbers of complete cycle in the 880 rpm vibration signal due to noise contamination. Comparatively, the signal acquired from the 1160 rpm showed clearer cycles of vibrations. As a result, noises contained in the signal may hinder the neural network learning progress and lead to lower average classification accuracy for the case of 880 rpm as shown in Figure 7. This suggests that test conditions could affect the neural network performance. Furthermore, it is essential to note that the deeper layers of U-Net model with multiple convolutions across the encoder and decoder could capture the signal features more efficiently in the high motor speed scenario regardless of the segmentation length.

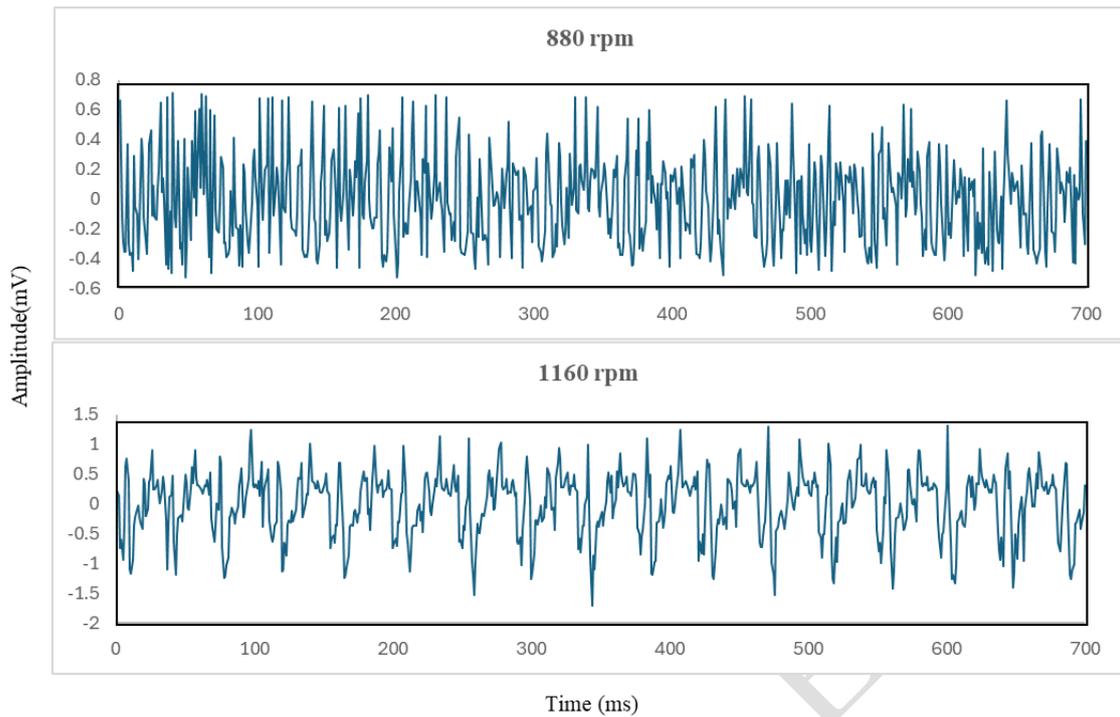


Figure 7: Raw vibration signals response of FD class under both 880 and 1160 rpm motor speeds.

The importance of data segmentation is also reported in other related studies where developing segmentation strategies on raw signal could help in achieving better neural network performance. Kim and Choi presented a new method for gear fault classification based on the idea of segmenting the original signal corresponding to the number of teeth of the gear via autocorrelation, after which the CNN is implemented for faults classification (Kim and Choi, 2018). It was reported that the segmented signal helps the neural networks distinguish different types of faults better, hence improving the overall accuracy. In another work, utilizing ECG recordings that share similar shapes among different leads to isolating individual heartbeats using R-peaks enabled easy extraction of heartbeats for analysis and training by splitting ECG signals into segments around each R-peak of ~ 1 second interval, which aids in accurate classification of waveform (Chen et al., 2023). Meanwhile in this study, as discussed

previously, choosing segmentation length that tallies with the excitation frequency may not yield the best performance. Nevertheless, analyzing the neural network performance by varying the signal segmentation length revealed a better performance of 95% average classification accuracy at 1200 ms segmentation length compared to 92.9% at default segmentation length of 100 ms (Table 3) for low motor speed scenario.

In terms of processing time, results in Figure 5 showed that both training and classification consumed less time within the segmentation length of 800 to 1400 ms. Specifically, at 1400 ms segmentation length, training the U-Net 4 model required least time of only 313 seconds for both low and high motor speed scenarios. Meanwhile, the classification time is at the best performance when the data was segmented at 1000 ms. The classification time increased thereafter until 1800 ms before it dropped slightly at 2000 ms, which is not favorable for an efficient computational processing system.

Nonetheless, to distinguish the best segmentation in terms of both accuracy and processing time, there is a trade-off between these two criteria at segmentation lengths of 1000 ms and 1200 ms. At 1000 ms, although the average accuracy is considered low among other neural networks with 84.4% and 97.9% for low and high motor speed respectively, the time taken for classification process is outstanding, which indicates the most reliable system for real-time result. In contrast, the best average accuracy in both low and high rpm was achieved at 1200 ms with 95% and 100% respectively, despite requiring longer training and classification time compared to 1000 ms segmentation length. This finding might be a useful reference for which the application of the system is crucial as the requirement on accuracy level and processing time may vary. So, tuning the right signal segmentation length is essential in order to achieve desirable performance with the proposed U-Net model. The main criteria of SHM systems in practical fields are accuracy and detection speed. For instance, if the application requires high accuracy, the relative processing time may take longer and vice versa

for a system that prioritizes a faster response. Therefore, tuning the right signal segmentation length could achieve better accuracy or faster response, depending on the application's requirements.

5. Conclusions

In this paper, fault diagnosis has been performed on a vibration platform based on the damage severity of the rubber vibration isolators by quantifying the number of induced cracks.

Based on the results:

- i. U-Net model achieved the best accuracy with the lowest computational time for training. LSTM and hybrid CNN-LSTM setup, however, achieved a lower classification accuracy for every scenario along with longer computational time.
- ii. Due to the most favorable performance of U-Net 4 architecture, further analysis has been conducted by varying the segmentation length of the input signals. The result shows that data segmentation potentially improves the model average accuracy as well as processing time with peak performance observed at segmentation length of 1200 ms and 1000 ms respectively, despite the trend fluctuating throughout the testing range.
- iii. U-Net performance shows different behaviors on both low and high motor speed scenarios considered in this work. As the optimal 1200 ms segmentation length was achieved at low-speed scenario with the highest average accuracy for all classes, the high-speed scenario recorded a consistently high average accuracy, and it is not sensitive towards various segmentations. Contradictory to past research using CNN for segmentation test on vibration signals, U-Net 4 in this study shows a better performance in capturing significant features of high-speed vibration signal, partly due to less embedded noise.

- iv. However, for both motor speed scenarios, the processing time is more favoured at 1000 ms despite slightly low classification accuracy. This concludes that there is a trade-off between model accuracy and computational speeds that might be at a different tier of crucialness in different applications that can act as a guide into creating a reliable SHM system.
- v. Future works revolve around investigating the performance of neural network models in classifying other failure mechanisms on rubber vibration isolator such as thermal degradation effect where rubber materials are exposed to high temperature environment in some industrial practical applications. In addition, degradation of rubber isolators subjected to other excitation scenarios such as traffic loadings and earthquakes could be investigated as well in order to explore the robustness of neural network models in identifying those defects.

Acknowledgments

This work is supported by FRGS/1/2022/TK0/UNIM/02/3 grant from the Ministry of Higher Education, Malaysia.

References

Adam, J. M., Parisi, F., and Sagaseta, J., (2018). "Research and practice on progressive collapse and robustness of building structures in the 21st century" *Engineering Structures*, 173, 122-149, <https://doi.org/10.1016/j.engstruct.2018.06.082>.

Aksan, F., Li, Y., Suresh, V., and Janik, P., (2023). “CNN-LSTM vs. LSTM-CNN to Predict Power Flow Direction: A Case Study of the High-Voltage Subnet of Northeast Germany” *Sensors*, 23, 901, <https://doi.org/10.3390/s23020901>.

Alshingiti, Z., Alaql, R., Al-Muhtadi, J., Haq, Q.E.U., Saleem, K., and Faheem, M.H., (2023). “A Deep Learning-Based Phishing Detection System Using CNN, LSTM, and LSTM-CNN” *Electronics*, 12, 232, <https://doi.org/10.3390/electronics12010232>.

Belkhiria, S., Hamdi, A., and Fathallah, R. (2020). “Strain-based criterion for uniaxial fatigue life prediction for an SBR rubber: Comparative study and development” *Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications*, 234(7), 897-909, <https://doi.org/10.1177/1464420720913595>.

Chen, C. C., Liu, Z., and Yang, G., (2021). “An improved fault diagnosis using 1d-convolutional neural network model”, *Electronics*, 10(1), 1–19, <https://doi.org/10.3390/electronics10010059>.

Chen, Z., Wang, M., Zhang, M., Huang, W., Gu, H., and Xu, J., (2023). “Post-processing refined ECG delineation based on 1D-UNet”, *Biomedical Signal Processing and Control*, 79, 104106, <https://doi.org/10.1016/j.bspc.2022.104106>.

Gil-Negrete, N., Viñolas, J., and Kari, L. (2006). “A simplified methodology to predict the dynamic stiffness of carbon-black filled rubber isolators using a finite element code”, *Journal of Sound and Vibration*, 296(4-5): 757-776, <https://doi.org/10.1016/j.jsv.2006.03.038>.

Guo, H., Lin, X. and Zhu, K. (2022). “Pyramid LSTM network for tool condition monitoring”, *IEEE Transactions on Instrumentation and Measurement*, 71, 1-11, <https://doi.org/10.1109/TIM.2022.3173278>.

Guo, S., Yang, T., and Gao, W., (2018). “A novel fault diagnosis method for rotating machinery based on a convolutional neural network”, *Sensors*, 18(5), 1429, <https://doi.org/10.3390/s18051429>.

Hoang, N.D., and Tran, D.V., (2024). “Machine Learning-Based Estimation of Concrete Compressive Strength: A Multi-Model and Multi-Dataset Study”, *Civil Engineering Infrastructures Journal*, 57(2): 247-265, https://ceij.ut.ac.ir/article_93331.

Hochreiter, S. and Schmidhuber, J. (1997). “Long Short-Term Memory”, *Neural Computation*, 9, 1735–1780, <https://doi.org/10.1162/neco.1997.9.8.1735>.

Hu, Q., He, Z., and Zhang, Z., (2007). “Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble”, *Mechanical Systems and Signal Processing*, 21(2): 688–705, <https://doi.org/10.1016/j.ymsp.2006.01.007>.

Huang, N.E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.C., Tung, C.C. and Liu, H.H., (1998). “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis”, *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971): 903-995, <https://doi.org/10.1098/rspa.1998.0193>.

Janssens, O., Slavkovikj, V., and Vervisch, B., (2016). “Convolutional Neural Network Based Fault Detection for Rotating Machinery”, *Journal of Sound and Vibration*, 377, 331–345, <https://doi.org/10.1016/j.jsv.2016.05.027>.

Jansson, A., Humphrey, E., and Montecchio, N., (2017). “Singing voice separation with deep U-Net convolutional networks”, *Proceedings of the 18th International Society for Music Information Retrieval Conference, Suzhou, China*, <https://openaccess.city.ac.uk/id/eprint/19289/>.

- Karim, F., Majumdar, S., and Darabi, H. (2017). “LSTM fully convolutional networks for time series classification”, *IEEE Access*, 6, 1662-1669, <https://doi.org/10.1109/ACCESS.2017.2779939>.
- Karim, F., Majumdar, S., and Darabi, H. (2019). “Multivariate LSTM-FCNs for time series classification”, *Neural Networks*, 116, 237-245, <https://doi.org/10.1016/j.neunet.2019.04.014>.
- Karim, F., Majumdar, S., and Darabi, H. (2019). “Insights into LSTM fully convolutional networks for time series classification”, *IEEE Access*, 7, 67718-67725, <https://doi.org/10.1109/ACCESS.2019.2916828>.
- Kim, S. and Choi, J-H. (2018). “Convolutional neural network for gear fault diagnosis based on signal segmentation approach”, *Structural Health Monitoring*, 18(5-6), 1401–1415, <https://doi.org/10.1177/1475921718805683>.
- Koszewski, D., Görne, T., and Korvel, G. (2023). “Automatic music signal mixing system based on one-dimensional Wave-U-Net autoencoders”, *EURASIP Journal on Audio, Speech, and Music Processing*, 2023(1), 1, <https://doi.org/10.1186/s13636-022-00266-3>.
- Lapčík, Jr L., Augustin, and P., Pištěk, A. (2001). “Measurement of the dynamic stiffness of recycled rubber based railway track mats according to the DB-TL 918.071 standard”, *Applied Acoustics*, 62(9), 1123-1128, [https://doi.org/10.1016/S0003-682X\(00\)00098-0](https://doi.org/10.1016/S0003-682X(00)00098-0).
- Li, P., Abdel-Aty, M., and Yuan, J. (2020). “Real-time crash risk prediction on arterials based on LSTM-CNN”, *Accident Analysis & Prevention*, 135, 105371, <https://doi.org/10.1016/j.aap.2019.105371>.
- Long, J., Shelhamer, E. and Darrell, T. (2015). “Fully convolutional networks for semantic segmentation”, *Proceedings of the IEEE conference on computer vision and pattern recognition*, Boston, 3431- 3440, <https://doi.org/10.48550/arXiv.1411.4038>.

- Luo, R. K. (2021). “Multiaxial fatigue prediction on crack initiation for rubber antivibration design–location and orientation with stress ranges”, *Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications*, 235(3), 469-480, <https://doi.org/10.1177/1464420720969857>.
- Ma, M. and Mao, Z. (2021). “Deep-convolution-based LSTM network for remaining useful life prediction”, *IEEE Transactions on Industrial Informatics*, 17(3), 1658-1667, <https://doi.org/10.1109/TII.2020.2991796>.
- Moosavian, A., Ahmadi, H., and Tabatabaeefar, A. (2013). “Comparison of two classifiers; K-nearest neighbor and artificial neural network, for fault diagnosis on a main engine journal-bearing”, *Shock and Vibration*, 20(2), 263-272, <https://doi.org/10.3233/SAV-2012-00742>.
- Perslev, M., Jensen, M. H., and Darkner, S. (2019). “U-Time: A fully convolutional network for time series segmentation applied to sleep staging”, *Advances in Neural Information Processing Systems*, 2019, 32, <https://doi.org/10.48550/arXiv.1910.11162>.
- Ronneberger, O., Fischer, P. and Brox, T. (2015). “U-net: Convolutional networks for biomedical image segmentation”, *Proceedings of Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich*, 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.
- Shen, J., Zhang, L., Kusunoki, K. (2023). “Structural floor acceleration denoising method using generative adversarial network”, *Soil Dynamics and Earthquake Engineering*, 173, 108061, <https://doi.org/10.1016/j.soildyn.2023.108061>.
- Tse, P. W., Yang, W. X., and Tam, H. Y. (2004). “Machine fault diagnosis through an effective exact wavelet analysis”, *Journal of Sound and Vibration*, 277(4–5), 1005–1024, <https://doi.org/10.1016/j.jsv.2003.09.031>.

Wahid, A., Breslin, J. G., and Intizar, M. A. (2022). "Prediction of machine failure in industry 4.0: a hybrid CNN-LSTM framework", *Applied Sciences*, 12(9), 4221, <https://doi.org/10.3390/app12094221>.

Wang, J., Cheng, S., and Tian, J., (2023). "A 2D CNN-LSTM hybrid algorithm using time series segments of EEG data for motor imagery classification", *Biomedical Signal Processing and Control*, 83, 104627, <https://doi.org/10.1016/j.bspc.2023.104627>.

Wu, Y. T. and Stewart, R. R. (2023). "Attenuating coherent environmental noise in seismic data via the U-net method", *Frontiers in Earth Science*, 11, 1082435, <https://doi.org/10.3389/feart.2023.1082435>.

Xie, P., Zhang, L., and Li, M. (2024). "Elevator vibration signal denoising by deep residual U-Net", *Measurement.*, 15, 225, 113976, <https://doi.org/10.1016/j.measurement.2023.113976>.

Zeng, Y., He, Z., and Pan, P. (2023). "A deep learning method to monitor axial pressure and shear deformation of rubber bearings under coupled compression and shear loading" *Earthquake Engineering & Structural Dynamics*, 52(11), 3304-3321, <https://doi.org/10.1002/eqe.3895>.

Zhang, D., Zhang, L., and Zhang, N. (2023). "Early Fault Diagnosis of Rolling Bearing Based on Threshold Acquisition U-Net", *Machines*, 11(1), 119, <https://doi.org/10.3390/machines11010119>.

Zhang, P. and Chen, C. (2023). "Time–frequency analysis for planetary gearbox fault diagnosis based on improved U-Net++", *Journal of Failure Analysis and Prevention*, 23(3), 1068-80, <https://doi.org/10.1007/s11668-023-01651-6>.

Zhao, H., Zhou, Y., and Bai, T. (2023). “A U-Net Based Multi-Scale Deformable Convolution Network for Seismic Random Noise Suppression”, *Remote Sensing*, 17, 15(18), 4569, <https://doi.org/10.3390/rs15184569>.

Zhou, S., Xiao, M., and Bartos, P. (2020). “Remaining useful life prediction and fault diagnosis of rolling bearings based on Short-time Fourier Transform and convolutional neural network”, *Shock and Vibration*, 2020, 1-14, <https://doi.org/10.1155/2020/8857307>.

Accepted / Not Edited