Gearbox Fault Diagnostics Using Deep Learning with Simulated Data

Ozhan Gecgel
Department of Mechanical Engineering
Texas Tech University
Lubbock, USA
ozhan.gecgel@ttu.edu

Stephen Ekwaro-Osire
Department of Mechanical Engineering
Texas Tech University
Lubbock, USA
stephen.ekwaro-osire@ttu.edu

João Paulo Dias
Department of Mechanical Engineering
Texas Tech University
Lubbock, USA
joao-paulo.dias@ttu.edu

Abdul Serwadda
Department of Computer Science
Texas Tech University
Lubbock, USA
abdul.serwadda@ttu.edu

Fisseha M. Alemayehu
School of Engineering, Computer Science and Mathematics
West Texas A&M University
Canyon, USA
falemayehu@wtamu.edu

Abraham Nispel
Department of Mechanical Engineering
Texas Tech University
Lubbock, USA
abraham.nispel@ttu.edu

***Abstract*—Transmission components are prone to fatigue damage due to high and intermittent loading cycles, that cause premature failure of gearboxes. Recently, several vibration-based diagnostics approaches using Machine Learning (ML) and Deep Learning (DL) algorithms have been proposed to identify gearboxes faults. However, most of them rely on a large amount of training data collection from physical experiments, which is often associated with high costs. This paper offers an ML and DL classification performance comparison of several algorithms to diagnose faults in a gearbox based on realistic simulated vibration data. A dynamic model of a single-stage gearbox was developed to generate data for different health conditions. Generated datasets were fed to ML and DL algorithms and accuracy results were compared. Results revealed the superiority of Convolutional Neural Network compared to other classifiers. This research contributes to the prevention of catastrophic failures in gearboxes by early crack detection and maintenance schedule optimization.**

Keywords—dynamic model, vibration-based diagnostics, deep learning, image encoding, tooth profile error

I. Introduction

Gearboxes are the main transmission components in power and transportation systems. A wide range of applications of transmission gearboxes can be found in vehicles, wind turbines, helicopters and bucket wheel excavators examples [1]. However, when transmitting power, gear teeth are prone to fatigue damage due to high and intermittent loading cycles, which are the main cause of premature failure of gearboxes [2]. The use of robust fault diagnostics techniques can provide indications of early faults in gearboxes, which is crucial to avoid unexpected catastrophic failures and reduce costs of maintenance [3]. Among the existing fault diagnostics techniques applied to gearbox diagnostics, vibration signal analysis is the most popular in the industry. Such techniques consist of the application of signal processing and feature extraction tools to identify fault signatures hidden in the vibration signals. Recently, several vibration-based diagnostics approaches have been proposed using Machine Learning (ML) and Deep Learning (DL) classification algorithms to automatically identify and classify faults in gearboxes [4]–[6]. However, the majority of these approaches rely on the collection of large amounts of training data from physical experiments or from the field, which is often associated with high costs in test-rig building and instrumentation [7]. The use of dynamic models to simulate the vibration response of gear systems under many faulty conditions has been shown to be a viable alternative to the expensive collection of data from experiments or field measurements [8], [9]. Additionally, realistic dynamic models that consider imperfections in gear tooth profile and noise artificially added to the vibration signal are preferred to predict the effects of faults in gearboxes, which can enhance simulation-based diagnostic algorithms. Several dynamic models considering manufacturing and assembly error in gears have been proposed. Bartelmus [10] presented a 6-DOF non-linear dynamic model of single-stage gearbox considering a flexible coupling and an error mode random function to describe the gear TPE. A similar dynamic model was proposed by Divandari et al. [11] who also incorporated the effect of the gear eccentricity error caused by the shaft misalignment. Xun et al. [12] presented a formulation based on a stochastic perturbation method to study the dynamic performance of planetary gear transmissions with random TPE. Dadon et al. [13] proposed a generic dynamic model for the prediction of gear vibration which included the modeling of the contact between gear teeth, the effect of faults on the mesh stiffness, and the integration with geometric errors. All these works have pointed out a considerable increase vibration response of the gear model as a result of the increase in the inter-tooth forces.

Real vibration signals present non-stationary and noisy characteristics which represent a considerable challenge for most of traditional ML diagnostics algorithms. Some of these traditional algorithms, such as Decision Tree Classifier (DTC), and Support Vector Machine (SVC), use manual feature extraction which is highly prone to uncertainty and require the expertise of the analyst to create a valid feature vector [14].

978-1-5386-8357-6/19/$31.00 ©2019 IEEE

978-1-5386-8357-6/19/$31.00 ©2019 IEEE

Unlike traditional feature-based methods, DL algorithms do not require hand-crafted features. DL algorithms such as Convolutional Neural Networks (CNN) are among the most used for classification problems involving complex signals [15]. In these methods, an image representation of time-series introduces different feature types that are not available for 1-D signals, and therefore time series classification can be treated as a texture image recognition task [16]. These algorithms have been successfully applied to fault diagnostics of gears and bearing using encoded images of vibration signals [14], [17], [18]. Although these image encoding methods have been reported to achieve good results in fault diagnosis of rotary machinery, CNN has also been successfully applied to problems involving fault identification in vibrations time series data without any intermediary image conversion process [19], [20]. Additionally, LSTM network has also shown satisfactory results in diagnostics and prognostics of machines using a complex time series of measured vibration and other signals [21], [22]. Different from CNN algorithm, LSTM network is a variant of the Recurrent Neural Network (RNN) that retains both recent and long-term memories of input patterns. However, none of the above-mentioned works have been used to classify faults using artificially generated realistic vibration signals.

In order to address the above-mentioned gaps, this research offers an ML and DL classification performance comparison of several algorithms to diagnose faults in a simple gearbox configuration based on realistic simulated vibration data. First, a 6-degrees-of-freedom (DOF) non-linear dynamic model of a single-stage gearbox was developed, in which the effect of the tooth profile error (TPE) in the gear teeth was considered and the noise was added artificially to the simulated signal. Secondly, the model was used to generate a dataset of samples for different scenarios of operation for the healthy and damaged gearbox. Thirdly, the generated data was preprocessed for traditional ML and DL algorithms. Statistical condition indicators were extracted from vibration signals to be used for training and testing of traditional ML algorithms namely Support Vector Classifier (SVC), Random Forest Classifier (RFC) and Decision Tree Classifier (DTC). For Convolutional Neural Network (CNN), vibration signals were converted to images using three different image encoding methods. Moreover, 1-D vibration signals were reshaped to 2-D to be able to feed them to 2-D CNN algorithm. Although preliminary performance results using 1-D CNN and 2-D CNN were found to be very similar, in this work, 2-D CNN was preferred to have a consistent model structure that is used for image inputs. No preprocessing was done for Long Short Term Memory (LSTM) input and the raw data was fed directly. All the algorithms were tuned for their best-performing architectures. Finally, the obtained fault classification accuracy rates were compared. The test results revealed the superiority of the CNN algorithm with 2-D time series of data compared to image input methods. Moreover, CNN showed a promising potential of application of gearbox diagnostics techniques using realistic simulated data.

II. Methodology

A. One-Stage Gearbox Dynamic Model

A 6-DOF one-stage gearbox dynamic model that consists of a pinion with 19 teeth and a gear with 48 teeth was used in the proposed diagnostics framework. Fig. 1 illustrates the free-body diagram of the dynamic model. This model was used to compute the dynamic response of a single-stage gear system [9], [10], [23]. The parameters shown in Fig. 1 are described in Table A.I in the appendix.

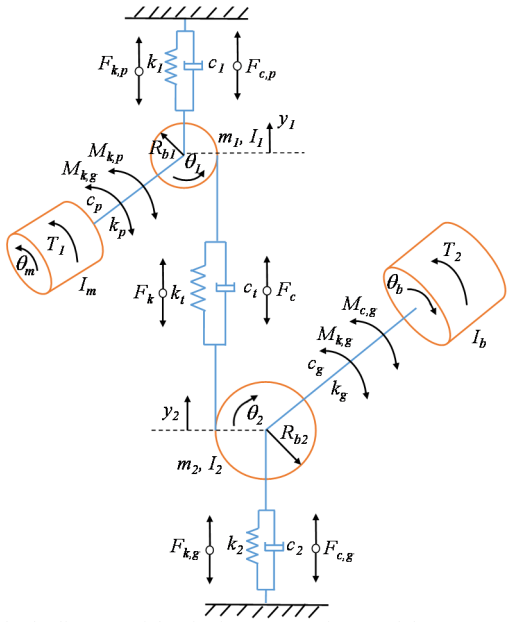

Fig. 1. Free-body diagram of the single-stage gearbox model.

The equations of motion for the system considering the linear displacement, velocity, and acceleration in the *y*-direction and the rotary displacements, velocities, and accelerations of the pinion, gear, motor, and load are given by [9], [10], [23]:

$$m_1\ddot{y}_1 = F_k + F_c - k_1 y_1 - c_1\dot{y}_1, \tag{1}$$

$$m_2\ddot{y}_2 = F_k + F_c - k_2 y_2 - c_2\dot{y}_2, \tag{2}$$

$$I_1\ddot{\theta}_1 = k_p(\theta_m - \theta_1) + c_p(\dot{\theta}_m - \dot{\theta}_1) - R_{b1}(F_k + F_c), \tag{3}$$

$$I_2\ddot{\theta}_2 = R_{b2}(F_k + F_c) - k_g(\theta_2 - \theta_b) - c_g(\dot{\theta}_2 - \dot{\theta}_b), \tag{4}$$

$$I_m\ddot{\theta}_m = M_1 - k_p(\theta_m - \theta_1) - c_p(\dot{\theta}_m - \dot{\theta}_1), \tag{5}$$

$$I_b\ddot{\theta}_b = -M_2 + k_g(\theta_2 - \theta_b) - c_g(\dot{\theta}_2 - \dot{\theta}_b), \tag{6}$$

$$F_k = k_t(R_{b1}\theta_1 - R_{b2}\theta_2 - y_1 + y_2), \tag{7}$$

$$F_c = c_t(R_{b1}\dot{\theta}_1 - R_{b2}\dot{\theta}_2 - \dot{y}_1 + \dot{y}_2), \tag{8}$$

Equations (1)-(8) were solved simultaneously to obtain the vibration signals, i.e. displacement, velocity, and acceleration of the pinion and gear. The solution of this set of differential equations depends on the calculation of the total effective mesh stiffness for different fault conditions of the pinion and gear. To determine the total effective mesh stiffness, a simplified model based on the potential energy method was adopted [24], as illustrated in Fig. 2. This model assumes that the crack is initiated at the root of the 15th tooth of the pinion and propagates through a straight line with size *a* and constant inclination angle $\beta = 45^o$ with respect to the centerline of the tooth. Fig. 2 also shows the angle between centerline and radius intersecting the tangent line drawn from the addendum of the root, α_g, the half

height of the top land of the tooth, h_r and the half tooth angle, α_2.

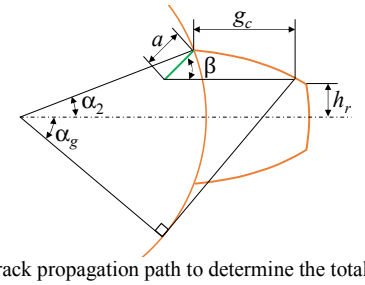

Fig. 2. Assumed crack propagation path to determine the total mesh stiffness.

B. Tooth Profile Error Model

In order to have a more realistic representation of the signal, the effects of TPE were included in the dynamic response of the system. The schematic representation of the TPE model is presented in Fig. 3. In this figure, the TPE, E_r, is interpreted as the relative difference between the design profile, P_D, and the actual profile, P_A, caused by the lack of precision of machinery employed in the manufacturing process, L_A, is an infinitesimal section of the tooth profile that is amplified to show the effect of the tooth profile error.

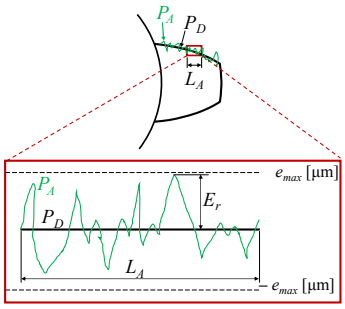

Fig. 3. Tooth profile error model.

In the absence of experimental data, the randomness of TPE can be quantified by [11], [10]:

$$E_r = [1 - r_{ces}(1 - l_i)]e_{max}, \tag{9}$$

in which r_{ces} is the coefficient of error scope ranging from 0 to 1, l_i is a random number between 0 to 1, and e_{max} is the maximum tolerance deviation in μm. Thus, the effect of the profile errors on the dynamic response of the system is calculated by,

$$F_o = F_k + F_c + k_t E_r, \tag{10}$$

where F_o is the total dynamic force of the system, F_k and F_c are the stiffness and damping contact forces given by equations (7) and (8). The additional dynamic force produced by deviations in the profile was computed based on the total mesh stiffness, k_t, and the values of E_r [11].

C. Additive Gaussian White Noise

After all data sets were generated with different TPE levels, white Gaussian noise with different signal-to-noise ratios (SNR) were added to all signals to simulate close to real-life conditions. The SNR values were calculated using an expression given by [25]:

$$\text{SNR}=10\times \log_{10}\left(\frac{P_{\text{signal}}}{P_{\text{noise}}}\right) \tag{11}$$

where P_{signal} is the power of the signal and P_{noise} is the power of the added noise. Commonly used values of 5, 10, and 15 have been considered as the noise levels in this study [26], [27].

D. Feature Extraction and Traditional Machine Learning Classifiers

The first fault classification framework proposed in this study involves the application of traditional ML algorithms in vibration-based diagnostics of a faulty gearbox. For this purpose, features were extracted manually from the simulated vibration signal using 24 statistical condition indicators [28]. The condition indicators used were (1) maximum value, (2) minimum value, (3) mean, (4) peak-to-peak, (5) harmonic mean, (6) trimmed mean, (7) variance, (8) standard deviation, (9) mean absolute deviation, (10) median absolute deviation, (11) crest factor, (12) peak to RMS, (13) skewness, (14) kurtosis, (15) shape factor, (16) RMS, (17) mean frequency, (18) frequency center, (19) RMS frequency, (20) figure-of-merit zero (FM0), (21) standard deviation frequency, (22) largest sideband amplitude, (23) sideband index, and (24) sideband level factor. Where the first 16 features are in time domain and the remaining 8 features are in frequency domain respectively. The 24 condition indicators were calculated for each data set and used for training and testing of classifiers.

In faults diagnostics applications, ML classification models (also termed classifiers), are able to train with the fault signature presented in the collected data and to classify the fault condition. There are several classifiers, each one suitable for a specific problem, and for this study, the three classifiers used were RFC, SVM, and DTC [29] due to their performance metrics in the previous study [9]. Each classifier has several algorithm parameters that have a significant effect on estimation accuracy. Thus, all the classifiers were tuned for their optimum algorithm parameters. The classifiers were then trained with 70% of the data selected randomly out of the dataset for each fault class and tested with the remaining 30%.

E. Convolutional Neural Network

CNN is one of the most common DL methods that are very popular for visual recognition. In this research, to take advantage of visual recognition of CNN, the time series of vibration data were encoded to images and fed to CNN. Three different imaging methods were used to transform the time series of vibration data into 2D images. These methods are called spectrogram, Gramian Angular Displacement Field (GADF) and Markov Transition Field (MTF). Alternatively, to make a head-to-head comparison with LSTM, time series of data input was also fed to CNN. Each 1-D time series of vibration data with 10,000 sample points, were reshaped to 100×100 and fed to the same CNN structure.

A spectrogram is a tool to visualize the spectrum of frequencies of a signal where x and y-axes are time and frequency and the color scale indicates the amplitude of the frequency. It is a commonly used signal imaging technique to take advantage of the CNN algorithm [30], [31]. A MATLAB function spectrogram function was used for this purpose.

GADF is an encoding method to encode the time series of data to images. It transforms the 1-D data into a symmetric matrix called Gram Matrix [32]**.** The advantage of GADF is that it preserves the temporal dependency. The time increases along the main diagonal from top left to bottom right where the original values of the scaled time series are contained. Therefore, it will be possible to reconstruct the times series from the high-level features learned by the deep neural network [32], [33]. To generate the GADF images, a Python package called PYTS was used [34].

MTF is another encoding method that encodes time series of data to images that was offered by Wang et. al [32]. The MTF encodes the multi-span transition probabilities of the times series by assigning the probability from the quantile at t_i to the quantile at t_j. Like for GADF images, MTF images were also generated with PYTS.

As shown in Fig. 4, the resized images were fed to CNN architecture that consists of consecutive convolutional layers, an activation layer, pooling layer, and fully connected layers. In between these layers, drop out layers were used to avoid overfitting. Overfitted networks memorize the training data; when it is tested with new data the network may not recognize the new data even if the difference is very subtle. Dropout layers delete connections between the neurons randomly. Drop out ratio is selected as 40%. During the training, 100×100 RGB images are fed to the first convolutional layer with a receptive field of 4×4. Convolutional layer scans the input image with 4×4 pixel windows and maps that to the max-pooling layer. Convolutional layer keeps sliding to the next window until all the image is scanned completely. In the max-pooling layer, 2×2 pixel window is used. Max-pooling layer scans the 4×4 pixel image from the convolutional layer, and picks the 4 biggest values and passes it to the activation layer. The convolutional and pooling layers are repeated 4 times with neuron numbers of 32, 64, 32, 32 respectively and connected to the fully connected 3 layers. The 3 fully connected dense layers have the neuron numbers 64, 64 and 4 respectively. While the "relu" and "tanh" activation functions are used throughout the convolutional and fully connected layers, the last layer was activated with "softmax" function.

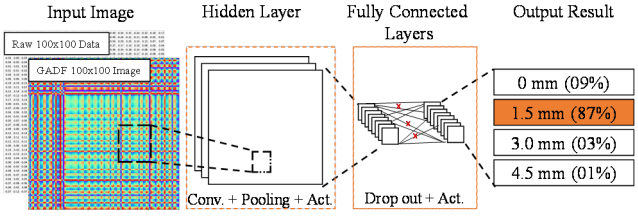

Fig. 4. CNN layers.

F. Long-Short-Term Memory

LSTM neural networks tend to perform well with time series data, especially if the prediction depends on the historical context. Examples of applications that meet this property include speech recognition, music, vibration signals, text. As shown in Fig. 5, the cell state starts from the first input to the last input. At each time step, LSTM stores only the necessary information of the inputs to cell state by forgetting gate. This way, important information is conserved and moved to the next state [21], [22]. Finally, the output state is calculated from the state cell with the according to weights.

Each signal with 10,000 sample points was directly fed to LSTM. Three LSTM layers with 64 nodes and with the activation functions of "sigmoid", "relu" and "tanh" were employed respectively in the LSTM structure. All layers were followed by dropout layers with 0.3 ratio and the structure was finalized with the "softmax" layer. The number of layers and activation function were determined by grid search to select the best performing structure.

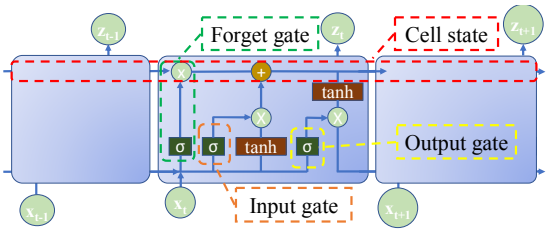

Fig. 5. LSTM framework

G. Proposed Single-Stage Gearbox Diagnostics Framework

Fig. 6 summarizes the proposed vibration-based diagnostics framework. The input parameters for the healthy and faulty gear conditions are provided to dynamic model with the TPE model to generate the signals for 4 different classes (healthy gear and gear with three different crack sizes). In total, 12 datasets are generated for the training/testing procedure by adding 3 different levels of noise to the initial datasets. As shown in TABLE I, the first three rows present datasets of individual 3 TPE levels (e_{max} parameters of 3, 5, 10 µm) and the fourth row presents a combination of each individual TPE level datasets (first 3 rows). After these 4 datasets were generated, 3 different noise levels (5, 10, 15 SNR) were considered and 3 different variations of the initial 4 datasets were obtained. While individual TPE level datasets have a sample size of 3,200 signals, datasets with the combination of TPE levels have a sample size of 9,600 signals. The 3,200 sample signals result from the combination of 20 values of input motor torque (uniformly spaced between 15 and 75 N.m), 20 values of input shaft rotational speed (uniformly spaced between 1,440 and 2,160 rpm), and 2 values of TPE r_{ces} parameter (0.15 and 0.30) for each one of the four classes. The last 3 datasets are obtained by a combination of all TPE levels with three different noise levels with a total of 9,600 sample signals for each dataset. After the signals are generated, white Gaussian noise is added to the simulated signal.

TABLE I OBTAINED DATASETS

e_{max}	SNR	Sample size	# of data sets
3µm	5 - 10 - 15	3200 signals	3
5µm	5 - 10 - 15	3200 signals	3
10µm	5 - 10 - 15	3200 signals	3
3µm, 5µm, 10µm	5 - 10 - 15	9600 signals	3

The resulting raw data is then pre-processed depending on the ML algorithm adopted. The features of the signal are extracted manually for traditional ML algorithms, in which a training/testing sample proportion of 70%/30% is used. For the CNN algorithm, spectrogram, GADF and MTF images are

created and a training/validation/testing proportion of 50%/25%/25% is adopted. Alternatively, the raw data is used directly in the CNN algorithm, as well as in the LSTM with the same training/validation/testing proportion of 50%/25%/25%.

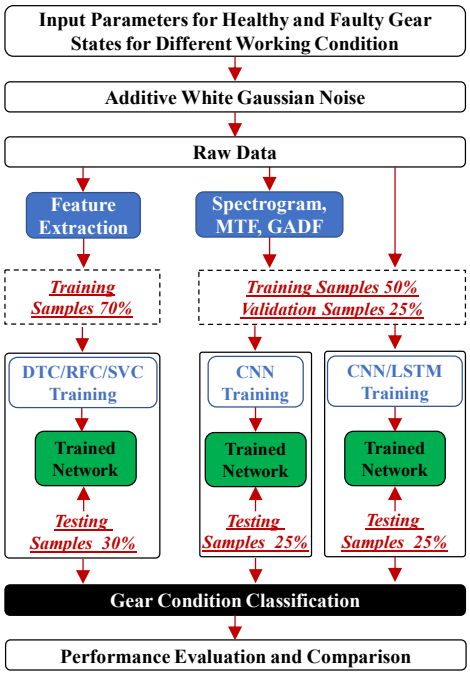

Fig. 6. Flowchart of the proposed single-stage gearbox vibration-based diagnostics framework.

III. Results and Discussion

A. Dynamic Response of the Single-Stage Gearbox

Fig. 7 presents the results showing the comparison between the simulated vibration signal of the pinion, with and without the TPE for different crack sizes in the pinion tooth. The vibration data from the dynamic model with TPE was obtained using TPE parameters of e_{max} = 5 μm and r_{ces} = 0.15. The impact on the dynamic response of the gear system is clearly evidenced by the sudden variations on the acceleration magnitude of the pinion. One can observe that the visual regularity of the simulated acceleration signals without the TPE is lost when this effect is considered. Fig. 7(a)-(d), show the change in the vibration pattern during the meshing of the healthy and 3 different levels of the faulty tooth. As it is highlighted with red dashed squares, the irregularity in the vibration when the cracked tooth is meshed becomes visually less distinguishable in the signals obtained with the TPE model as the crack size is reduced. The results indicate that, despite the micro-geometrical nature of the TPE, this effect can represent some random characteristics observed in real vibration data, which can make the detection of early crack faults difficult for conventional feature extraction techniques.

In an attempt to explore the impact of the TPE model parameters on the dynamic response of the gear system, the acceleration of the pinion was simulated with different values of e_{max} and r_{ces}, and is presented in, Fig. 8 and Fig. 9, respectively. From the plots, it is clearly seen that the magnitude of the pinion acceleration increases when e_{max} or r_{ces} is increased. In addition to the change in magnitude, it can be observed that changes in r_{ces} have an impact on the pattern of the acceleration signal as well. This occurs due to the greater r_{ces} value increasing the effect of the random numbers (l_i) utilized to assess the values of $E_{r.}$. Since the TPE parameters e_{max} and r_{ces} may vary depending on the manufacturing process and on the wear state of the gear surface, experimental data is required to have a better characterization of such parameters.

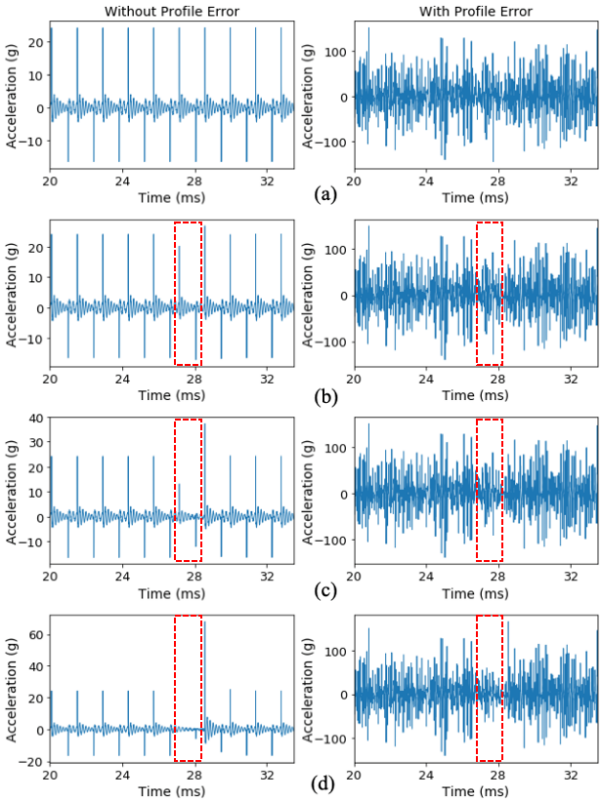

Fig. 7. Comparison of simulated vibration data obtained with and without the TPE for different crack sizes in the pinion: (a) 0.0 mm, (b) 1.5 mm, (c) 3.0 mm, (d) 4.5 mm (simulation conditions T = 15 N.m, f = 2160 rpm, e_{max}=5 μm).

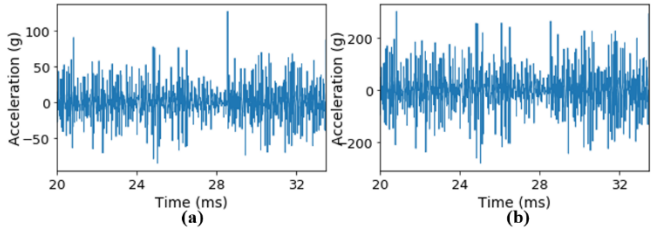

Fig. 8. Effect of the maximum error of the tooth profile on the simulated vibration data with a 4.5 mm crack in the pinion for (a) e_{max}=3 μm, and (b) e_{max}=10 μm (simulation conditions T=15 N.m, f=2160 rpm, and r_{ces}= 0.15).

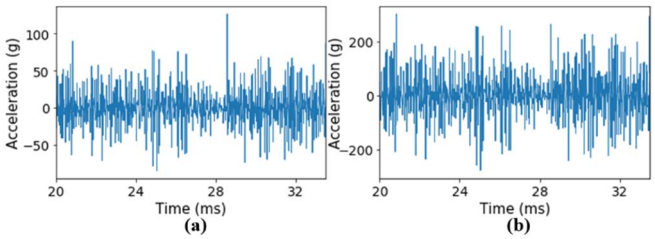

Fig. 9. Effect of the parameter r_{ces} of the tooth profile on the simulated vibration data with a 3.0 mm crack in the pinion for (a) r_{ces}=0.15, and (b) r_{ces}=0.30 (simulation conditions T=15 N.m, f=2160 rpm, and e_{max}=10 μm).

B. Classification

In this part, firstly, the classification results of traditional ML algorithms will be shown. Secondly, example images of encoding time series to images will be shown followed by classification results of the CNN with the encoded images. Lastly, CNN and LSTM classification results with raw vibration data will be analyzed and discussed.

1) Traditional Machine Learning Results with Manual Feature Extraction

All the features were extracted and used as inputs to the SVC, RFC and DTC algorithms. All three classifiers were tuned for their optimum parameters for each dataset and all the classifiers were run 20 times to see the variations in accuracy. Fig. 10 shows accuracies and error bars for all datasets in 4 groups. Each group represents the rows 1 to 4 of TABLE I respectively. As shown in this figure, all the classifiers resulted in very low accuracies. Even with the least challenging data set, lowest TPE level and lowest noise level (e_{max}=3 µm, 15 SNR), the highest accuracy achieved was 80.4% by RFC. As the noise level and TPE level were increased, a decrease in mean accuracy and an increase in standard deviation were observed as expected. As aforementioned, the results of traditional ML algorithms are highly depended on the features that are used. These results could be improved by exploring more features and trying to see which features might improve the accuracy by trial error. Yet, since this approach is very cumbersome and the initial results were not promising at all, no further analyses have been done with traditional ML classifiers. Thus, research was moved on to DL algorithms.

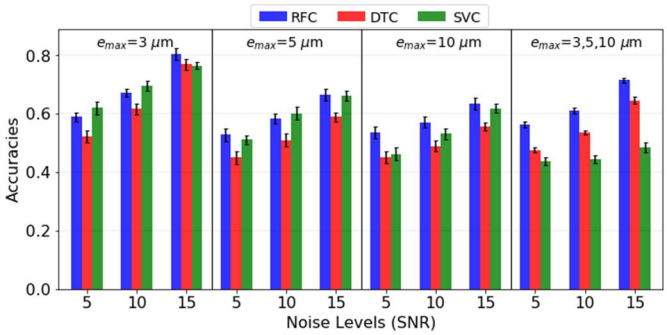

Fig. 10. Accuracies of traditional ML algorithms.

2) CNN with Image Input

Using all the data sets, spectrogram, MTF and GADF images were generated for each signal in all the 12 data sets. All the images were resized to 100×100 pixel to make the data easier to handle and to save on the computational power. Feeding images with the initial pixel size of 430×430 require higher memory and computational time. 100×100 was determined as the ideal pixel size due to computational efficiency and no loss in critical information. Fig. 11 shows the example images with their compressed versions for spectrogram, MTF, and GADF respectively. As a result, 36 different datasets were generated were used as inputs to CNN. Fig. 12 shows classification accuracies with the error bars for 20 different runs. It can be easily seen that for all levels of noise and TPE, encoding time series into images with GADF resulted in better accuracy compared to spectrogram and MTF. As expected, as the noise level and TPE level increases, accuracy decreases. Moreover, it resulted in a greater span of an error bar. It can also be observed that as all TPE levels were combined, overall accuracy increases. To elaborate, for GADF encoding at 5 SNR level accuracies are 88%, 81% and 77% for 3µm, 5µm, 10µm TPE levels respectively, while they are fed one by one, with an average of 82% of the 20 runs. On the other hand, when they are fed all together the average accuracy goes to 86%. In this paper, although it is not a commonly used TPE level, 3µm TPE level was also used as a data augmentation tool to see if it will have any effect on the accuracy as the number of the sample size increases. Moreover, it was shown that it has a positive effect when combined with the other signals.

Confusion matrix of CNN with GADF encoding method for the data set with 5 SNR and all TPE signals is shown in Fig. 13. In this figure, the condition classification and misclassification results can be observed in detail. The vertical axis represents the actual label of the pinion condition in terms of crack size, and the horizontal axis shows the predicted label. The diagonal of a confusion matrix (from top left to bottom right) represents the classification accuracy of each condition. A diagonal value lower than unity indicates that the classifier has confused certain crack size with a different one. For instance, while 80% of the healthy signals predicted correctly, 18% of healthy signals are classified as a 1.5mm crack fault and 2% of the healthy signals are classified as a 3mm crack fault. Most of the misclassification was observed in the distinction of the healthy gear and 1.5mm crack size. This has already been pointed out as one of the main challenges of diagnostics of small size cracks in gears [8].

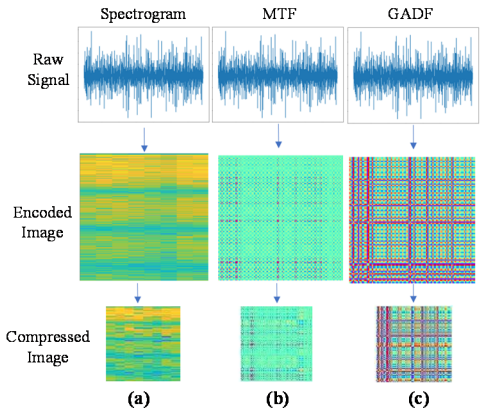

Fig. 11. Encoding time series of data to images with a) Spectrogram b) MTF c) GADF.

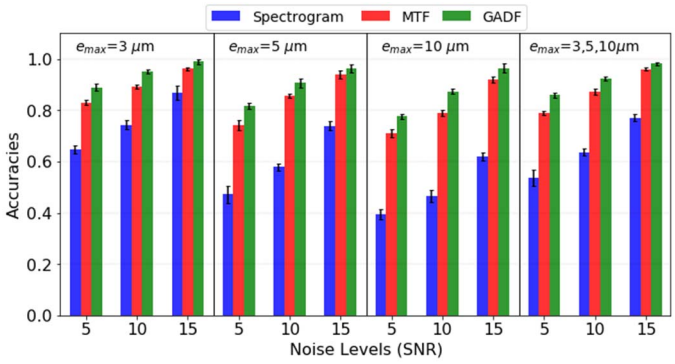

Fig. 12. Accuracies of CNN with different encoding methods.

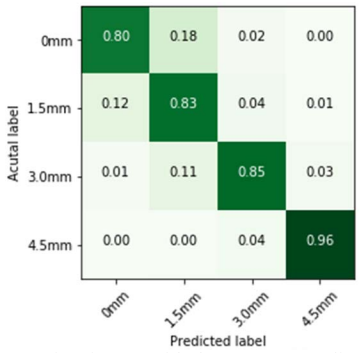

Fig. 13. Confusion matrix of CNN with the GADF encoding method.

3) LSTM and CNN with Raw Data Inputs

In this part, classification results of LSTM and CNN with raw data inputs are presented. While all the signals were fed to LSTM directly, for CNN, each signal was reshaped to 2-D to be able to use the same layer structure that is used for image inputs. Finally, as stated before, data was divided into three, with 50% for training, 25% for validation and another 25% for testing. Fig. 14 shows classification accuracies with the error bars for each data set for LSTM and CNN. It can be seen that for all levels of noise and TPE level, CNN outperforms LSTM and all other methods discussed above. Moreover, from the error bars, we can see that CNN with raw data input has the lowest variation among all other methods. We can also observe that LSTM benefits from combining all the data together greatly. For instance, while the mean accuracy of each run for 5 SNR with single TPE level is 64% when all the TPE levels were combined the mean accuracy increases to 84%. This accuracy result is even greater than the accuracy of the least challenging dataset (e_{max}=3µm, 15 SNR) with an accuracy of 68%. CNN results with raw data input shows similar accuracy trends as image inputs. It benefits from the combination of all TPE levels at a similar rate to the image inputs. We can say that CNN performs better than LSTM with less amount of data and data augmentation can improve the accuracy of LSTM dramatically. Fig. 15 shows the confusion matrix of CNN with raw data input. Similar to Fig. 13 most of the misclassification is done at the smaller crack sizes. Although classification accuracies of healthy and 1.5mm crack sizes are close to each other for both methods, the classification accuracy of 3mm crack size with CNN raw data performs much better.

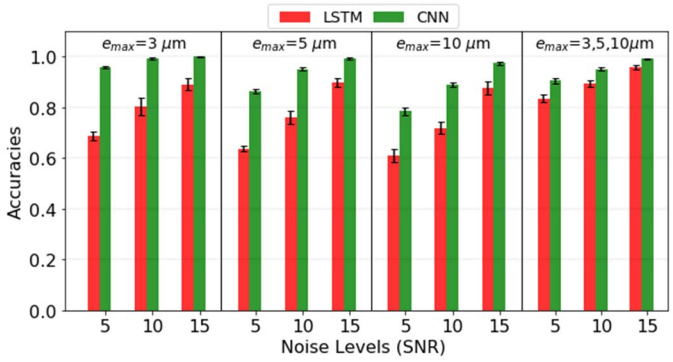

Fig. 14. Accuracies of CNN and LSTM with raw data inputs.

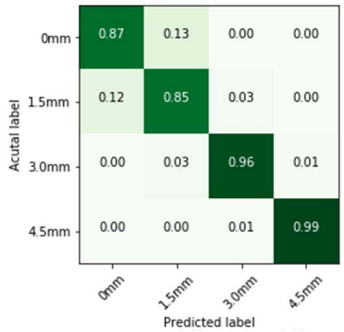

Fig. 15. Confusion matrix of CNN with raw data input.

IV. Conclusion

In this paper, the performance of an ML and DL fault classification algorithms in vibration-based diagnostics of a single-stage gearbox. Six DOF dynamic model of a single-stage gearbox was developed to simulate vibration data. 6 DOF dynamic model of a single-stage gearbox was developed. To make the simulated vibration signals more realistic, TPE was considered, and Gaussian white noise was added to the signal. The model was used to generate data for different operating conditions for both healthy and damaged gearbox scenarios. Statistical condition indicators were extracted from vibration signals to be used for training and testing of traditional ML algorithms namely SVC, RFC and DTC. For CNN, vibration signals were converted to images using several image encoding methods. Moreover, 1-D vibration signals were reshaped to 2-D to be able to feed them to CNN algorithm. No preprocessing has been done for LSTM and the raw data was fed directly. All the algorithms were tuned for their best-performing architectures. Finally, the obtained fault classification accuracy rates were compared.

The results have shown that the micro-geometrical nature of the TPE added considerable complexity to the simulated signal, which brought more difficulty to some classification algorithms to detect and identify the presence of the crack in the gear tooth. CNN showed superior performance compared to LSTM and traditional ML classifiers. Although CNN is a very popular tool for image recognition and image classification, it is proven that it performs the best among all other methods with raw vibration signal. The proposed framework can contribute to the development of a new generation DL diagnostics algorithms trained only with simulated data. However, further refinements on the dynamic model to incorporate different types of faults in the gear (e.g. wear, crack, misalignment) and the interaction with other faulty components (e.g. bearings, shafts) need to be done. Furthermore, the proposed approach still lacks validation with experimental data of a real gear test rig, which could bring considerable improvements to the diagnostics framework. For future work, simulated data will be used to train the proposed algorithm and experimental data will be collected to test its performance.

References

[1] X. Liang, M. J. Zuo, and Z. Feng, "Dynamic modeling of gearbox faults: A review," *Mech. Syst. Signal Process.*, vol. 98, pp. 852–876, 2018.

[2] F. Zhao, Z. Tian, and Y. Zeng, "Overview on Gear Health Prognostics," in *Probabilistic Prognostics and Health Management of Energy Systems*, S. Ekwaro-Osire, A. C. Goncalves, and F. M. Alemayehu, Eds. Springer

International Publishing, 2017, pp. 49–65.
[3] Y. Qu, D. He, J. Yoon, B. Van Hecke, E. Bechhoefer, and J. Zhu, "Gearbox Tooth Cut Fault Diagnostics Using Acoustic Emission and Vibration Sensors — A Comparative Study," pp. 1372–1393, 2014.
[4] N. Saravanan and K. I. Ramachandran, "Expert Systems with Applications Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN)," *Expert Syst. Appl.*, vol. 37, no. 6, pp. 4168–4181, 2010.
[5] H. Chen, Y. Sun, Z. Shi, and J. Lin, "Intelligent Analysis Method of Gear Faults Based on FRWT and SVM," vol. 2016, 2016.
[6] L. Jedli and J. Jonak, "A disassembly-free method for evaluation of spiral bevel gear assembly," vol. 88, no. March 2015, pp. 399–412, 2017.
[7] Z. Tian, M. J. Zuo, and S. Wu, "Crack propagation assessment for spur gears using model-based analysis and simulation," *J. Intell. Manuf.*, vol. 23, no. 2, pp. 239–253, 2012.
[8] L. Wang and Y. Shao, "Fault mode analysis and detection for gear tooth crack during its propagating process based on dynamic simulation method," vol. 71, pp. 166–178, 2017.
[9] O. Gecgel, S. Ekwaro-Osire, J. P. Dias, A. Nispel, F. M. Alemayehu, and A. Serwadda, "Machine Learning in Crack Size Estimation of a Spur Gear Pair Using Simulated Vibration Data," in *Proceedings of the 10th International Conference on Rotor Dynamics – IFToMM Vol. 2*, vol. 61, 2019, pp. 175–190.
[10] W. Bartelmus, "Mathematical modelling and computer simulations as an aid to gearbox diagnostics," *Mech. Syst. Signal Process.*, vol. 15, no. 5, pp. 855–871, 2001.
[11] M. Divandari, B. H. Aghdam, and R. Barzamini, "Tooth profile modification and its effect on spur gear pair vibration in presence of localized tooth defect," *J. Mech.*, vol. 28, no. 2, pp. 373–381, 2012.
[12] C. Xun, X. Long, and H. Hua, "Effects of random tooth profile errors on the dynamic behaviors of planetary gears," *J. Sound Vib.*, vol. 415, pp. 91–110, 2018.
[13] I. Dadon, N. Koren, R. Klein, and J. Bortman, "A realistic dynamic model for gear fault diagnosis," *Eng. Fail. Anal.*, vol. 84, no. November 2017, pp. 77–100, 2018.
[14] D. Verstraete, A. Ferrada, E. L. Droguett, V. Meruane, and M. Modarres, "Deep Learning Enabled Fault Diagnosis Using Time-Frequency Image Analysis of Rolling Element Bearings," *Shock Vib.*, pp. 1–17, 2017.
[15] G. Zhao, G. Zhang, Q. Ge, and X. Liu, "Research Advances in Fault Diagnosis and Prognostic based on Deep Learning," in *Proceedings of 2016 Prognostics and System Health Management Conference (PHM-Chengdu)*, 2017, pp. 1–6.
[16] N. Hatami, Y. Gavet, and J. Debayle, "Classification of Time-Series Images Using Deep Convolutional Neural Networks," 2017.
[17] J. Wang, J. Zhuang, L. Duan, and W. Cheng, "A Multi-Scale Convolution Neural Network for Featureless Fault Diagnosis," in *Proceedings of International Symposium on Flexible Automation*, 2016, pp. 1–6.
[18] P. Cao, S. Zhang, and J. Tang, "Preprocessing-Free Gear Fault Diagnosis Using Small Datasets with Deep Convolutional Neural Network-Based Transfer Learning," *IEEE Access*, vol. 6, pp. 26241–26253, 2018.
[19] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors (Switzerland)*, vol. 17, no. 2, 2017.
[20] L. Wen, L. Gao, X. Li, L. Wang, and J. Zhu, "A Jointed Signal Analysis and Convolutional Neural Network Method for Fault Diagnosis," *Procedia CIRP*, vol. 72, pp. 1084–1087, 2018.
[21] J. Zhang, P. Wang, R. Yan, and R. X. Gao, "Long short-term memory for machine remaining life prediction," *J. Manuf. Syst.*, vol. 48, pp. 78–86, 2018.
[22] H. Luo, M. Huang, and Z. Zhou, "Integration of Multi-Gaussian fitting and LSTM neural networks for health monitoring of an automotive suspension component," *J. Sound Vib.*, vol. 428, pp. 87–103, 2018.
[23] H. B. Endeshaw, S. Ekwaro-Osire, F. M. Alemayehu, and J. P. Dias, "Evaluation of fatigue crack propagation of gears considering uncertainties in loading and material properties," *Sustain.*, vol. 9, no. 12, 2017.
[24] S. Wu, M. J. Zuo, and A. Parey, "Simulation of spur gear dynamics and estimation of fault growth," *J. Sound Vib.*, vol. 317, no. 3–5, pp. 608–624, 2008.
[25] C. Li, W. E. I. Zhang, G. Peng, and S. Liu, "Bearing Fault Diagnosis Using Fully-Connected Winner-Take-All Autoencoder," *Access*, pp. 6103–6115, 2018.
[26] M. Zhao, J. Lin, Y. Miao, and X. Xu, "Detection and recovery of fault impulses via improved harmonic product spectrum and its application in defect size estimation of train bearings," *Meas. J. Int. Meas. Confed.*, vol. 91, pp. 421–439, 2016.
[27] Y. Zhang, B. Tang, Z. Liu, and R. Chen, "An adaptive demodulation approach for bearing fault detection based on adaptive wavelet filtering and spectral subtraction," *Meas. Sci. Technol.*, vol. 27, no. 2, p. 025001, 2016.
[28] Z. Liu, J. Qu, M. J. Zuo, and H. B. Xu, "Fault level diagnosis for planetary gearboxes using hybrid kernel feature selection and kernel Fisher discriminant analysis," *Int. J. Adv. Manuf. Technol.*, vol. 67, no. 5–8, pp. 1217–1230, 2013.
[29] "Scikit-learn: machine learning in Python — Scikit-learn 0.19.1 documentation," *Scikit-Learn*, 2017. [Online]. Available: http://scikit-learn.org/stable/. [Accessed: 04-Feb-2018].
[30] D. Verstraete, A. Ferrada, E. L. Droguett, V. Meruane, and M. Modarres, "Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings," *Shock Vib.*, vol. 2017, 2017.
[31] H. Liu, L. Li, and J. Ma, "Rolling Bearing Fault Diagnosis Based on STFT-Deep Learning and Sound Signals," *Shock Vib.*, 2016.
[32] Z. Wang and T. Oates, "Spatially Encoding Temporal Correlations to Classify Temporal Data Using Convolutional Neural Networks," 2015.
[33] J.-F. Chen, W.-L. Chen, C.-P. Huang, S.-H. Huang, and A.-P. Chen, "Financial Time-Series Data Analysis Using Deep Convolutional Neural Networks," *2016 7th Int. Conf. Cloud Comput. Big Data*, pp. 87–92, 2016.
[34] J. Faouzi, "pyts: a Python package for time series transformation and classification," 2018.

APPENDIX A

TABLE A. I MODEL PARAMETERS OF THE SINGLE-STAGE GEARBOX MODEL

Parameter	Description
F_k, F_c	Stiffness and damping contact (inter-tooth) forces
$F_{k,p}$, $F_{c,p}$	Stiffness and damping force of input shaft bearing
$F_{k,g}$, $F_{c,g}$	Stiffness and damping force of output shaft bearing
$M_{k,p}$, $M_{c,p}$	Stiffness and damping moment of the input shaft
$M_{k,g}$, $M_{c,g}$	Stiffness and damping moment of the output shaft
T_1, T_2	Input motor torque and output load torque
m_1, m_2	Masses of the pinion and gear
I_1, I_2	Moments of inertia of the pinion and gear assembly
I_m, I_b	Moments of inertia of the motor and load
R_{b1}, R_{b2}	Base circle radius of the pinion and gear
θ_1, θ_2	Angular displacement of the pinion and gear
θ_m, θ_b	Angular displacement of the motor and load
y_1, y_2	Linear displacement of the pinion and gear in the y-direction
k_1, k_2	Vertical radial stiffness of input and output bearings
k_p, k_g	Torsional stiffness of input and output flexible couplings
c_1,c_2	Vertical radial damping coefficient of input and output bearings
c_p, c_g	Damping coefficient of input and output flexible couplings
k_t, c_t	Total effective mesh stiffness and damping coefficient