Condition monitoring of ball bearings using Bayesian neural networks

Matheus de Moraes[1], João Paulo Dias[2], Hélio Fiori de Castro[1]

[1]*School of Mechanical Engineering, University of Campinas*
200 Mendeleyev Street, 13083-860, Campinas, São Paulo, Brazil
m263136@dac.unicamp.br, hfc@unicamp.br
[2]*Dept. of Civil and Mechanical Engineering, Shippensburg University of Pennsylvania*
1871 Old Main Drive, 17257, Shippensburg, Pennsylvania, United States
JPDias@ship.edu

Abstract. Rotating machines play a fundamental role in several engineering applications. In most of these applications, they are subjected to unexpected overloads that can cause the premature failure of their components. This work aims to present a condition monitoring framework that employs the Bayesian neural networks approach, which includes uncertainties quantification associated with the damage detection in ball bearings. Images of vibration signals of ball bearings were used to feed a Bayesian neural networks and the algorithm predictions were given in terms of the probability density function of the possibility of the vibration image belongs to a specific type and severity of damages. The results demonstrated that Bayesian neural networks (BNN) as a powerful technique for damage diagnosis, and it can quantify uncertainties in condition monitoring of ball bearings.

Keywords: Rotating Machines, Condition Monitoring, Bayesian Deep Learning, Variational inference

1 Introduction

Rotating machines play a fundamental role in several engineering applications. These machines have been widely used in different areas such as energy generation [1], transportation, and water waste treatment [2]. In most of these applications, rotating machines are often subjected to unexpected overloads. This can cause the premature failure of its components, which can lead to accidents, loss of revenue and waste of natural resources. Thus, condition monitoring is a fundamental tool to anticipate these premature failures and support predictive maintenance campaigns of rotating machines.

Furthermore, computational modeling and artificial intelligence are transforming industrial activities in many fields and are aiding decisions in manufacturing and maintenance tasks [3]. Nowadays, researchers are working on the development of data science to help decision-making tasks in condition monitoring [4]. Learning algorithms which include uncertainties quantification are cutting-edge in data-driven models, then there are related works in literature. Hoang and Kang [5] presented an automatic ball bearing damage diagnostic over four types of damage using vibration signals directly, the results achieved high performance even in noisy situations. Abdeljaber et al. [6] presented an online approach that identifies, quantifies the severity, and localizes bearing damages, in this work are shown an experimental validation with several types of defects, and the algorithm achieves high performance in these tasks. Variational inference (VI) was employed by San Martin et al. [7] for data dimensionality reduction, thus unsupervised learning algorithms were used in damage diagnostics of ball bearings elements and accomplished high accuracy. Lu et al. [8] employed Bayesian optimization to tune hyper-parameters of a deep convolutional network and this tool showed improvements in detection of bearings damages, even in the calculation of defect size.

Although important contributions had been made so far, the following research question arises: data science techniques as machine learning combined with uncertainty quantification can be a powerful tool to give more information and improve damage detection of ball bearings? The general goal of this work is to present a condition monitoring framework that employs a Bayesian neural networks (BNN) approach which includes the quantification of the uncertainties associated with the detection of ball bearings damages.

Thus, the specific aims of this work are: (i) to generate vibration images of ball bearings public dataset; (ii) to implement, test, and evaluate diverse convolutional Bayesian neural networks (BNN) in diagnostic tasks; and (iii) to classify the type of the damage, to quantify the damage severity, and to quantify uncertainties in these classifications.

2 Methodology

2.1 Vibration images

This work used the public dataset of different classes of ball bearing damage available on the Case Western Reserve University (CWRU) Bearing Data Center Website. The choice for this dataset aims the ensure reproducibility, and further, other researchers described the dataset in literature work [9].

The acceleration signals obtained with a sampling rate of 42 kHz were transformed in several vibration images in the same way proposed by Hoang and Kang [5]. This procedure consists of transforming the time-series as follows: first, the signal is segmented in an appropriate length of 400 samples; next, the segment of the signal is normalized with zero mean in values that range from -1 to 1; finally, the segment of the signal is transformed by the following equation:

$$Pixel[i,j] = A[(i-1)M + j], \tag{1}$$

— where $i = 1 : N$, $j = 1 : M$, $Pixel[i, j]$ is the value of the corresponding pixel. This procedure generates a grayscale vibration image. In this work, the vibration image was set with the size of $M = 20$ and $N = 20$ pixels.

2.2 Bayesian neural networks

In order to make data-driven diagnosis of the ball bearings defects, this work employed Bayesian neural networks (BNN). The initial layers were modeled with a two-dimensional convolutional (Conv2D) layer and a sub-sampling (pooling) layer for the feature extraction task. The advantage of using convolution layers in feature extraction is that it does not require experience in the choice of the features. In condition monitoring, experience means previous knowledge about physical systems (physical models and equipment history data).Some limitations of using convolutional layers are associated with the requirement of large volumes of data in the training of learning algorithms and the absence of physical meaning. For more information and a rigorous mathematical explanation of convolution in the images feature extraction, the following literature is recommend [10–12]. The final layers were built with fully connected dense layers, and one hot categorical layer as activation function.

BNN has probabilistic layers with parameters θ, which differs from the traditional convolutional neural networks (CNN) that have deterministic parameters. This work is limited to the application of Bayesian learning in the first convolutional reparametrization layer and the last dense variational layer. The convolutional reparametrization layer takes the aleatoric uncertainty into account. On the other hand, the dense variational layer captures the epistemic uncertainty that emerges from the lack of information about the modeling process. The central idea to estimate the posterior density function of the parameters θ given the data D comes from the Bayes' rule:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{\int p(D|\theta')p(\theta')d\theta'}, \tag{2}$$

where $p(\theta|D)$ is the posterior density function of the parameters θ, $p(D|\theta)$ is the likelihood function, $p(\theta)$ is the prior distribution of the parameters θ, and $\int p(D|\theta')P(\theta')d\theta'$ is a normalization constant.

Variational inference was employed to find the approximation of posterior distributions minimizing the Kullback-Leibler divergence $\mathbf{D_{KL}}$ between some distribution $q(\theta|\omega)$ that belongs to a Q family of distributions and the posterior $p(\theta|D)$. This operation, is equivalent to maximizing the Evidence Lower Bound function (ELBO)[4, 7, 13]:

$$ELBO(q(\theta|\omega)) = \mathbb{E}_{q(\theta|\omega)}(\log p(\theta|D) - \mathbf{D_{KL}}(q(\theta|\omega)||p(\theta)), \tag{3}$$

where $\mathbb{E}_{q(\theta|\omega)}(\log p(\theta|D))$ is the expectation of the log-evidence, and $p(\theta)$ is the prior distribution of the set of parameters θ.

In this work, Tensorflow and Tensorflow probability (TFP) modules were used to construct the BNN. Variational layers learn posterior distributions over the weights of the model incorporating distributions into the model architecture. The inclusion of uncertainty in neural network models is done by changing each weight from a point estimation (deterministic) to a probability distribution which is the building blocks of TFP. Then, the model learns the parameters of these distributions.

The procedure of the Bayes' algorithm by back propagation [13, 14] starts with the definition of the prior distributions over these parameters, which means the belief about model parameters θ in the lack of information about any data D. The implementation of the convolutional reparametrization layer passed by the definition of priors and posterior distributions over the kernel and bias. It was inserted in the layer by the kernel posterior function that requires callable objects with mean and standard deviation parameters. The kernel matrix receives the posterior. In turn, the kernel prior function argument was defined by the spherical Gaussian with zero mean and standard deviation equal to 1. In the same way, the dense variational layer was defined. For both, convolution reparametrization and dense variational layers, the posterior distributions were learned using variational inference, in other words, by optimization of the $\mathbf{D_{KL}}$ loss function. Figure 1 shows an schematic representation of the BNN architecture used in this work, which has posterior distributions in the kernel matrix of convolution and dense variational reparametrization layers.

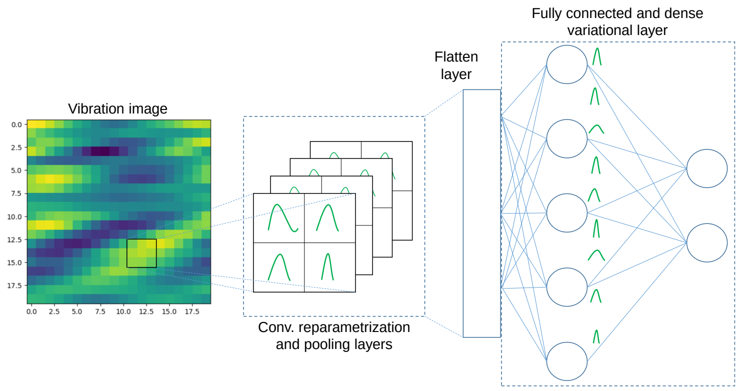


Figure 1. Scheme of the Bayesian neural network architecture

2.3 Data augmentation

Data augmentation is a technique to enlarge the data set, in order to consider the same number of signals in each class, which is better explained in section 3.1 and 3.3. A common way to enlarge the dataset is adding noise to the signal. The level of noise is calculate by the signal-to-noise ratios (SNR), given by [5, 6, 15]:

$$SNR = 10\log_{10}\left(\frac{P_{signal}}{P_{noise}}\right), \tag{4}$$

where P_{signal} and P_{noise} are the power of signal and noise in the signal, respectively.

3 Results and discussion

3.1 Vibration images organization and size of training set

In this work, only the operation with null motor load and an approximated motor speed of 1797 RPM were considered. Beyond the normal baseline operation, *i. e.*, without damage, there are three damage types: ball, inner race, outer race. Furthermore, for each damage type there are three different severity levels expressed in damage diameters: 0.007", 0.014", 0.021". Furthermore, we choose ten classes of damage in the diagnosis, Table 1 shows the quantity of vibration images for each class condition.

Table 1. Vibration images data set

Condition	Damage diameter [in]			Total of vibration images
	0.007	0.014	0.021	
Normal	–	–	–	606
Ball	606	606	606	1818
Inner race	606	158	606	1370
Outer race	606	606	606	1818

Note that, in Table 1 the class pertaining to the 0.014 in damage in the inner race contains only 158 vibration images, whereas all other classes contain 606 vibration images. Because of that, augmented data is considered for this specific class.

For the first numerical experimentation with the model was performed with the full dataset composed of 5612 images. The results for the training and testing accuracies with different sizes of the training set are presented in Table 2. Its possible to notice that, the training set with 4500 vibration images presented the higher performance in the test.

Table 2. Training and testing accuracies for different sizes of the training set

Training set size	Training accuracy	Testing accuracy
2806 vibration images (50%)	85.99%	80.22%
4209 vibration images (75%)	87.36%	84.25%
4500 vibration images (80%)	89.76%	85.79%
5050 vibration images (90%)	89.94%	85.23%

3.2 Tunning hyper-parameters and diagnostic

A comparison between models with different architectures of BNN were tested in this section:

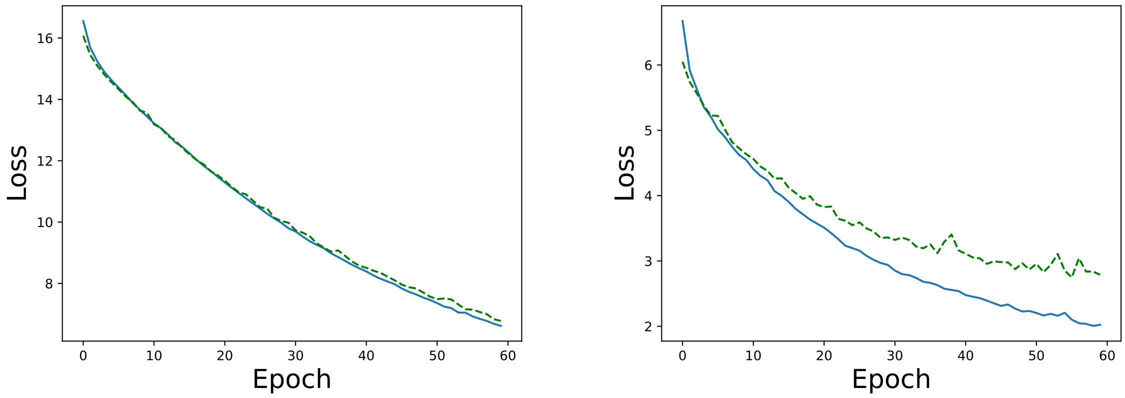
- Model #1: sequential model called convolutional Bayesian neural network with five layers: convolutional 2D reparametrization; maxpooling 2D; flatten layer; dense reparameterization layer;
- Model #2: sequential model called convolutional Bayesian neural network with five layers: convolutional 2D reparametrization; maxpooling 2D; convolutional 2D; maxpooling 2D; flatten layer; dense reparameterization layer;
- Model #3: sequential model called convolutional Bayesian neural network with five layers: convolutional 2D reparametrization; maxpooling 2D; convolutional 2D; maxpooling 2D; flatten layer; dense with 64 units; dense with 64 units; dense reparameterization layer.

Table 3. Training and testing accuracies for different architetures of BNN

Model	Train accuracy	Test accuracy
#1	89.76%	85.79%
#2	90.49%	85.52%
#3	93.91%	84.17%

Its is possible to notice that even with increases in training accuracy, the accuracy on the testing set did not improve. And it is due the overfitting that harms the predictive ability of the model. Overffiting occurs when the model fits excessively the data during the training, when the models is tested with new data in validation, it is not able to classify new data.

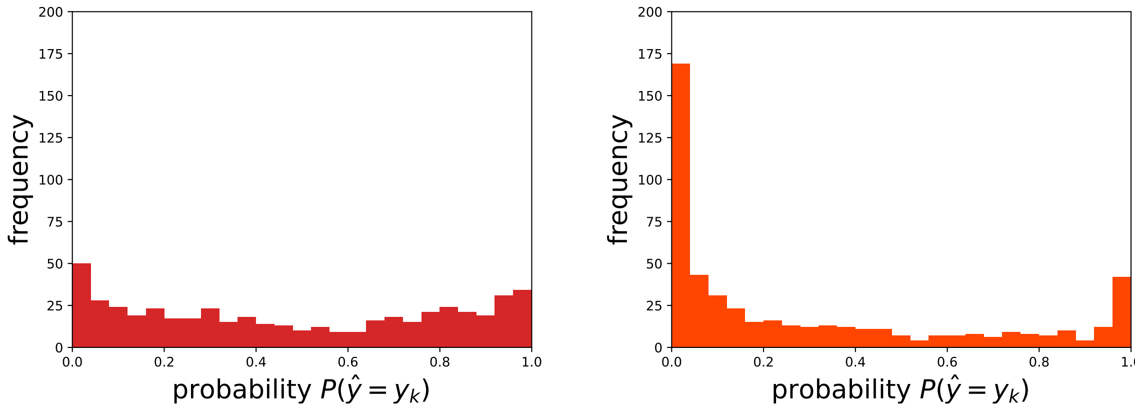
Figure 2 shows a comparison between the graphic of the loss function value *versus* epochs of training of the model #1 (Fig. 2-a) and the model #3 (Fig. 2-b). These results indicate that there is overfitting in the training of model #3 (Fig. 2-b), since the curve of validation set distanced from the curve of training set as far as the number of epochs increases. The same overfitting is not observed in the graphic of model #1.



(a) Model #1 (b) Model #3

Figure 2. Graphics of loss *versus* epochs, the blue curve (-) represents the training set and green dashed curve (- -) represents the validation set, for (a) model #1 and (b) model #3

The histograms were obtained running 500 times the model prediction with the same vibration image (sampling method), in which the median and the standard deviation were obtained. Figure 3 shows two histograms of a diagnosis in which the true label is inner race with a diameter of 0.014" (IR 14), showing that the distribution of probabilities of the vibration image belongs to inner race with diameter of 0.014" (Fig. 3-a) and to outer race with diameter of 0.021" (OR 21)(Fig. 3-b). The calculated values of the median and the standard deviation were $= 0.45$ and std $= 0.33$, respectively, for IR 14 (Fig. 3-a) and $= 0.11$ and $= 0.35$, respectively, for OR 21(Fig. 3-b) . Thus, its is possible to infer that, in this case, the model did not learn to classify well, possibly because of the large uncertainties for both classes. This occurs due to the small number of images for the class IR 14. This is the main advantages of applying Bayesian Neural network. In the case of a conventional neural network, a false diagnosis probably should be indicated. In that case, a doubt is presented, and the correct diagnosis is possible, even its probability is not the highest. Therefore, possible indications should be considered and analyzed.



(a) Inner race, diameter 0.014" (b) Outer race, diameter 0.021"

Figure 3. Histograms of a diagnosis of inner race with a diameter of 0.014", showing the distribution of probabilities of the vibration image belongs (a) to inner race with diameter of 0.014" and (b) to outer race with diameter of 0.021"

3.3 Diagnostic with augmented data

In order to enlarge the dataset, we choose and add three levels of Gaussian white noise (- 4 dB, -3 dB and -2 dB), and add them into the vibration signals of the measure of inner race damage with a diameter of 0.014”, which has the smallest number of vibration images (see Tab. 1). After the procedure of data augmentation over the case in which had few vibration images, BNN reached training accuracy of 95.40% and the accuracy in the test set achieves 87.26%.

Figure 4 shows the confusion matrix of this trial. It is possible to notice that there are not false negative cases, for all classifications of the normal condition none of them was a failure state Fig. 4. This is important in condition monitoring task. From the observation of Fig. 4 it is also possible to notice that we obtain high performance in IR 14 detection. Although, it is possible to notice that the performance did not increase and it is explained by the model’s need for more vibration images.

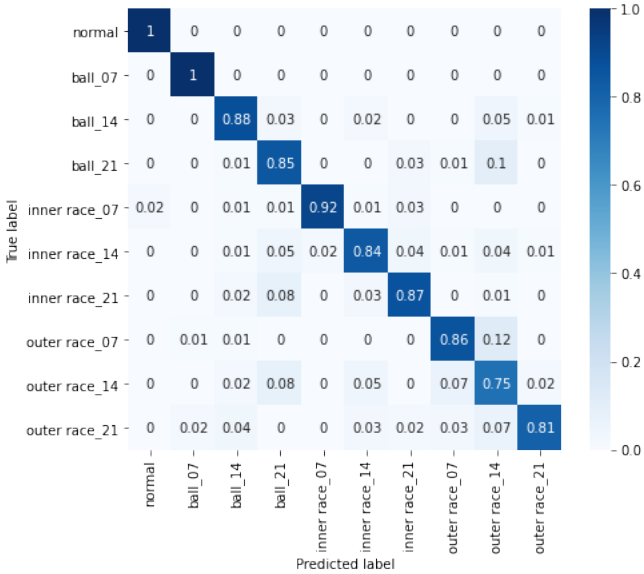


Figure 4. Confusion matrix for the test with the augmented data

4 Conclusions

This paper’s main conclusions are:

- the proposed data-driven model performs well in diagnoses task, which means achieves high scores during the test with several classes, and including the detection of the type and severity of the damage;
- there are not false positives of baseline classification during the tests with the proposed model, which is relevant in condition monitoring context;
- beyond the high scores in the diagnostic task, Bayesian neural networks are able to take into account of uncertainties of diagnostic and this is relevant in decision-making under risks, as condition monitoring;
- limitation of the model is related to the need for a large number of images, thus the technique of data augmentation helped to improve the ability of the model to classify with few samples of vibration signal.

Acknowledgements. The authors acknowledge the financial support given by Brazilian Coordination for the Improvement of Higher Education Personnel (CAPES) and the São Paulo Research Foundation (FAPESP) [grant number 2019/00974-1].

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] Y. Du, S. Zhou, X. Jing, Y. Peng, H. Wu, and N. Kwok. Damage detection techniques for wind turbine blades: A review. *Mechanical Systems and Signal Processing*, vol. 141, pp. 106445, 2020.

[2] S.-B. Ma, S. Kim, and J.-H. Kim. Optimization design of a two-vane pump for wastewater treatment using machine-learning-based surrogate modeling. *Processes*, vol. 8, n. 9, pp. 1170, 2020.

[3] R. Liu, B. Yang, E. Zio, and X. Chen. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, 2018.

[4] M. A. Vega and M. D. Todd. A variational bayesian neural network for structural health monitoring and cost-informed decision-making in miter gates. *Structural Health Monitoring*, vol. 00, pp. 1–15, 2020.

[5] D.-T. Hoang and H.-J. Kang. Rolling element bearing fault diagnosis using convolutional neural network and vibration image. *Cognitive Systems Research*, vol. 53, pp. 42–50, 2019.

[6] O. Abdeljaber, S. Sassi, O. Avci, S. Kiranyaz, A. A. Ibrahim, and M. Gabbouj. Fault detection and severity identification of ball bearings by online condition monitoring. *IEEE Transactions on Industrial Electronics*, vol. 66, n. 10, pp. 8136–8147, 2018.

[7] G. San Martin, E. López Droguett, V. Meruane, and das M. Chagas Moura. Deep variational auto-encoders: A promising tool for dimensionality reduction and ball bearing elements fault diagnosis. *Structural Health Monitoring*, vol. 18, n. 4, pp. 1092–1128, 2019.

[8] Y. Lu, Z. Wang, R. Xie, J. Zhang, Z. Pan, and S. Y. Liang. Bayesian optimized deep convolutional network for bearing diagnosis. *International Journal of Advanced Manufacturing Technology*, vol. 108, 2020.

[9] W. A. Smith and R. B. Randall. Rolling element bearing diagnostics using the case western reserve university data: A benchmark study. *Mechanical systems and signal processing*, vol. 64, pp. 100–131, 2015.

[10] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, vol. 1, n. 4, pp. 541–551, 1989.

[11] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, vol. 86, n. 11, pp. 2278–2324, 1998.

[12] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer series in statistics. Springer, 2009.

[13] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra. Weight uncertainty in neural network. In *International Conference on Machine Learning*, pp. 1613–1622. PMLR, 2015.

[14] K. Shridhar, F. Laumann, and M. Liwicki. Uncertainty estimations by softplus normalization in bayesian convolutional neural networks with variational inference. *arXiv preprint arXiv:1806.05978*, vol. 1, n. 1, 2018.

[15] O. Gecgel, S. Ekwaro-Osire, J. P. Dias, A. Serwadda, F. M. Alemayehu, and A. Nispel. Gearbox fault diagnostics using deep learning with simulated data. In *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*, pp. 1–8. IEEE, 2019.