XLIII CILAMCE
21–25, November, 2022
FOZ DO IGUAÇU – PR – BRAZIL

Comparison between Monte Carlo Dropout and Variational Inference Techniques for Bayesian Neural Network Models applied to Rotating Machinery Diagnostics

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

[1]*School of Mechanical Engineering, State University of Campinas*
200 Mendeleyev Street, 13083-860, Campinas, São Paulo, Brazil
o185244@dac.unicamp.br, 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 components play a crucial role in mechanical systems and are present in several industrial areas. These systems suffer from the adverse actions of loads and environmental condition. Thus, condition-based maintenance of these components is a fundamental technique to synchronize and support maintenance schedules, and machine learning algorithms are nowadays supporting these tasks. This paper aims to present a comparison between Monte Carlo Dropout and Variational Inference techniques applied to Bayesian neural network models in damage dignostics of ball bearings. The Bayesian Convolutional Neural Networks were tested, evaluated, validated against the physical data, and their prediction performances were compared. Results showed that both models had high performance in diagnosis. The comparison between the methods applied to Bayesian neural network models showed similar results but each method present their own characteristics that could provides advantages in certain specific situations.

Keywords: Rotating Machinery, Bayesian Neural Networks, Diagnosis.

1 Introduction

It is widely known that rotating components play a crucial role in mechanical systems and are present in a wide variety of projects including power generation, manufacturing, and transportation. Inherently to the nature of power transmission components, these systems suffer from the adverse effect of wear, overloading, assembly errors, lack of lubrication, etc. All these bad conditions can lead machine malfunction, which can result in accidents and loss of resources. In this context, condition-based maintenance of these components is a fundamental technique to synchronize and support maintenance schedules.

Machine fault detection using artificial intelligence techniques has been studied in the past decade and a thorough review on this subject was presented by Liu et al. [1]. The approach to Bayesian inference by applying the Monte Carlo Dropout technique was first described in the work of Gal and Ghahramani [2]. This technique has been employed in situations where a higher diagnostic reliability is needed, such as in the medical field as in the works of Lee and Kim [3] and Ju et al. [4]. Other applications can be found in the reliability of nuclear power plants in Bae et al. [5] and for monitoring of rotorcraft icing from aeroacoustics time-series data in Tong et al. [6].

Peng et al. [7] proposed two distinct architectures of Bayesian neural networks (BNN) that employed Variational Inference (VI) technique aiming to estimate the remaining useful life (RUL) of ball bearings of turbofan engines. In this work, whereas the first BNN received complex features of time-fequency signals as input, the second received time series; results for the RUL in terms of probability distributions were presented and a comparison between the real life of the ball bearing in experiments, for two studied architectures of BNN the RUL estimation shown. Wang et al. [8] proposed a comparison between BNN that uses VI and a Bayesian logistic regression (BLR) and the techniques were tested experimentally for the condition monitoring of a intern combustion engine, the results has a better performance in relation of common techniques. The work proposed by de Moraes et al. [9] presents a damage diagnostic of different failures in ball bearings starting from generation of vibration images of time signs; these data feds a convolutional BNN with VI and in the tests were done diagnostics of different kind of damages with different levels of severity, and it allowed the uncertainties to be quantified.

The main objective of this paper is to gather an understanding of different BNN techniques and bring them into the context of rotating machinery diagnostics. The research question of this research is: which parameters favor the choice of one technique over the other? Finally, the specific aims are: (i) to build the BNN models; (ii) to train, test and evaluate the BNN models; (iii) to compare the performance of the BNN models.

Beyond this introductory section, the paper is outlined in the methodology which is presented in Section 2. The main results and discussion about them are presented in Section 3. The final remarks and conclusions about this paper are presented in Section 4.

2 Methodology

2.1 Bayesian Inference

Bayesian models are produced through inference of their parameters, that is, by determining the probability distribution of the parameters by using prior information from available data. The formulation of Bayesian inference begins by considering a data set D, and the model parameters as θ. When Bayes' Theorem is applied we have the following equation:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{\int p(D|\theta')p(\theta')d\theta'}, \tag{1}$$

where, $p(D|\theta)$ are the likelihood function, $p(\theta)$ is the prior distribuitions of the parameters θ, and $\int p(D|\theta')p(\theta')d\theta'$ is the evidence, which is a likelihood function in which some parameters have been marginalized.

2.2 Monte Carlo Dropout

The dropout procedure was formalized by Srivastava et al. [10] as a regularization method for neural networks, and consists of modeling the probability of a neuron participating or not in the model by a Bernoulli distribution. By applying the Monte Carlo test to neural network multiple predictions are obtained on the neural network with dropout. According to Gal and Ghahramani [2] optimization based on stochastic gradient descent combined with the application of dropout during training produces the occurrence of the deep Gaussian process described by Damianou and Lawrence [11] responsible for approximating Bayesian inference in the model. To determine the uncertainty related to prediction, multiple predictions are sampled from a neural network with dropout, and the uncertainty can be determined by calculating the standard deviation, the entropy, and the negative log likelihood.

Finally, the response of the Monte Carlo dropout neural network is given by the average of its predictions combined with the uncertainty calculated by one of the methods described. In this paper, the standard deviation method was chosen for the attribution of uncertainty because it is related to direct values of the probability scale of the softmax function applied in the last layer of the network.

2.3 Variational inference

Once the central purpose of parametric Bayesian methods is obtaining the posterior distribution of the non-observable parameters θ that is density function, which represents the uncertainties over θ, with the observable data set D. In general, the integration of the evidence of eq. (1) is untreatable. Therefore, variational inference is employed to approximate the posterior analytically. A family of treatable distributions $\mathcal{Q}$ is chosen with the objective of to search, between the probability density functions $q(\theta)$ that belongs to the family, a probability density function $q^*(\theta)$ in the way of:

$$q^*(\theta) = \text{argmin} D_{KL}(q(\theta)||p(D|\theta)), \tag{2}$$

where, $D_{KL}(q(\theta)||p(D|\theta)$ is the Kullback-Leibler divergence between $q(\theta)$ and $p(D|\theta)$ that is given by:

$$D_{KL}(q(\theta)||p(D|\theta)) = \int \log \frac{q(\theta)}{p(D|\theta)} q(\theta) d\theta. \tag{3}$$

after some algebraic manipulations we can demonstrate that when the D_{KL} is minimized, the evidence lower bound objective (ELBO) function is maximized with respect of $q(\theta)$. Finally:

$$\log p(D) = D_{KL}(q(\theta)||p(\theta|D) + \mathcal{L}(q(\theta))), \tag{4}$$

thus, the ELBO function is the lower boundary of the evidence and it is treatable.

2.4 Data augmentation, Vibration Images and Neural Network Architecture

Data augmentation is a technique to creat a new samples from an already existing dataset, in which usually only a few data points are known. In this technique, noise is added to copies of samples of the original dataset in order to artificially generate more samples points. In this work, noise was added to the signal using the signal-to-noise ratios (SNR) with the MatLab function White Gaussian Noise. This procedure was already used in in some previous works [12], [13].

As was done in the paper de Moraes et al. [9] to feed the Bayesian Convolution Neural Networks were used vibration images. The raw data presented in [14] were transformed into images as the proposal of Hoang and Kang [12].

To perform the comparison a common structure was found for both types of Bayesian models, in the Table 1 all the constructive characteristics of the neural networks are shown. For the layers that contain in their name the term *Reparametrization*, normal distributions were assumed for the parameters.

Table 1. Architecture of the Neural Networks Used

Layers		**Parameters**	**Value**	**Parameters**	**Value**	**Parameters**	**Value**
Variational Inference	**Monte Carlo Dropout**						
Convolution 2D Reparametrization	Convolution 2D	Filters	60	Kernel Size	8	Activaation	Relu
Maxpolling 2D	Maxpolling 2D	Poll Size	(2,2)	-	-	-	-
Convolutional 2D	Convolutional 2D	Filters	60	Kernel Size	(3,3)	Activaation	Relu
Maxpolling 2D	Maxpolling 2D	Poll Size	(2,2)	-	-	-	-
Flatten	Flatten	-	-	-	-	-	-
Dense	Dense (with Dropout)	Activation	Relu	Number of Neurons	90	-	-
Dense	Dense (with Dropout)	Activation	Relu	Number of Neurons	90	-	-
Dense Reparametrization	Dense (with Dropout)	Activation	Relu	Number of Neurons	90	-	-
Dense	Dense	Activation	Softmax	Number of Neurons	10	-	-

Figure 1 shows the flowchart of the methodology of this work.

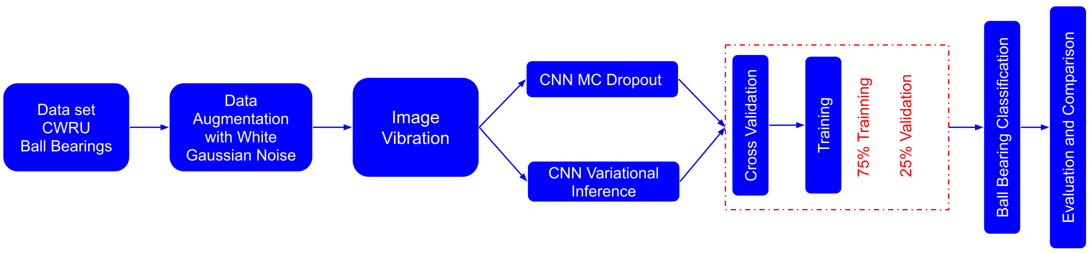


Figure 1. Paper's methodology flowchart.

3 Results and Discussion

As one of the main points of any Deep Learning model, training and validation play a central role in the applicability of diagnostic techniques, an unsatisfactory performance in this aspect makes the project unfeasible. In this sense, Figure 2a below shows a graph comparing the curves of training and validation accuracy over the training epochs for the variational inference and Monte Carlo Dropout methods. Training was stopped at 200 epochs as the accuracy stabilized around 100 training epochs. As shown in the graph of Figure 2a, the Bayesian model with Variational Inference stands out by demonstrating a higher speed in training and a higher accuracy rate against training data than the model with Monte Carlo Dropout. However, the model with Monte Carlo Dropout demonstrated a better balance between test accuracy and validation accuracy, the point curves, from Monte Carlo Dropout, demonstrate a much higher correlation over epochs than the solid lines from Variational Inference. Thus, for this problem, the Monte Carlo Dropout technique tended to be able to produce a model that best generalized the problem to the validation data, achieving superior accuracy with this data set.

In order to highlight this capability of Monte Carlo Dropout relative to Variational Inference, a cross-validation procedure was performed on the two models, the graph of Figure 2b, where the training and validation accuracy data are shown for each of the ten different combination of samples in the procedure, called K-fold. The trend observed in Figure 2a is again verified, the now supported by the cross validation data shown in Figure 2b. The red dashed line representing the test accuracy with Monte Carlo Dropout is higher in all test K-folds finding its best value near 95% accuracy, while the model with Variational Inference finds the best accuracy value near 90%, maintaining an average test accuracy close to 87% accuracy.

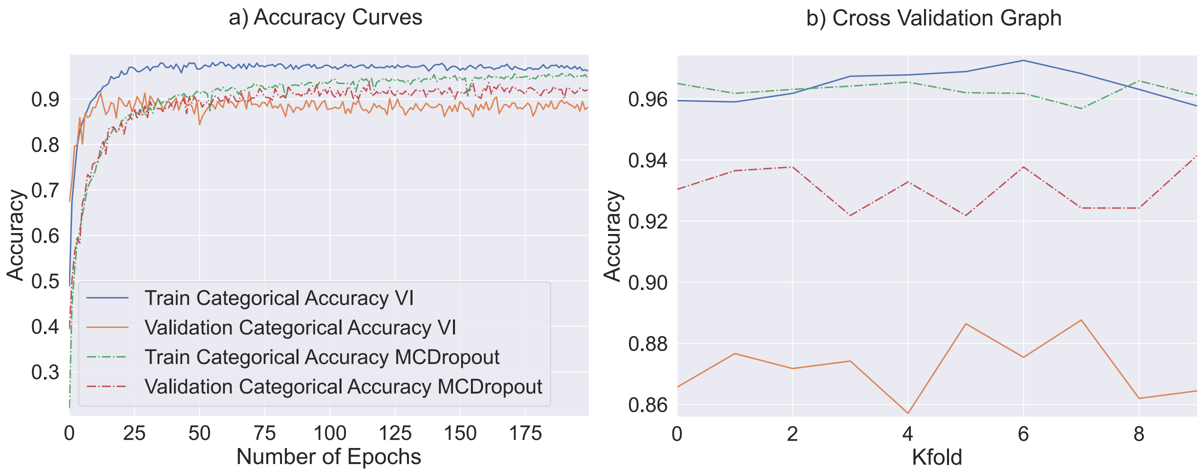


Figure 2. a) Graphic of accuracy *versus* number of epochs. b) Graphic of accuracy in each K-fold during cross-validation.

The Figures 3 and 4 shows the confusion matrices of the mean values calculated from a sample size of 500. Comparing the two confusion matrices it is possible to note an equivalence between the methods in most classes the Variational Inference (Fig. 4) performed slightly better than the Monte Carlo Dropout model (Fig. 3). The largest differences in accuracy were observed in the Outer Race 7 classes in favor of Variational Inference (Fig. 4) and in the Ball 14 mils (mil $= 10^{-3}$ in) class favoring Monte Carlo Dropout (Fig. 3).

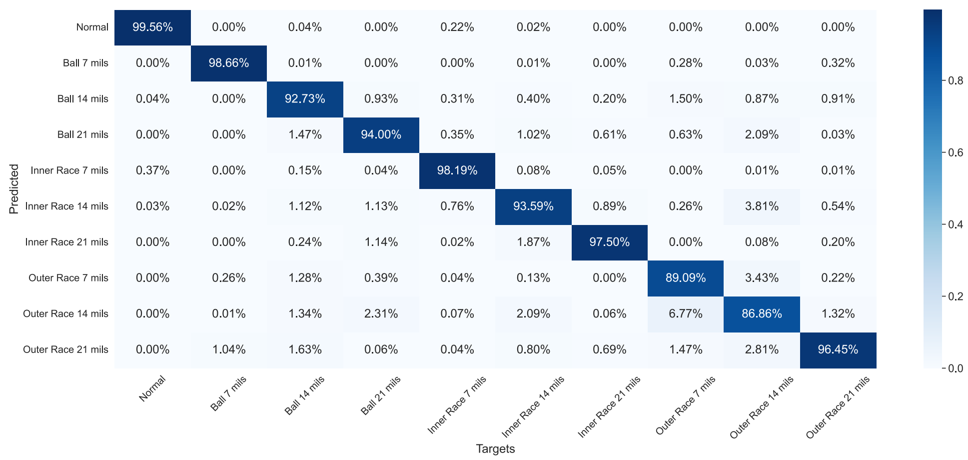


Figure 3. Mean Confusion Matrix for Monte Carlo Dropout Model.

With these differences in accuracy raised, it is of interest to relate them to the levels of uncertainty produced by the models. In this sense, Figures 5 and 6 shows in confusion matrices format the average uncertainty of the predictions for each class. In general, the levels of uncertainty in the Monte Carlo Dropout model (Fig. 5) are higher with respect to those of the model with Variational Inference (Fig. 6). It is relevant to emphasize that in the Monte Carlo Dropout model, the highest levels of uncertainty are correlated to the Outer Race 7 class, in which a lower classification accuracy was observed in relation to the Variational Inference. Nevertheless, in both models the levels of greater uncertainty were also concentrated in Outer Race classes, demonstrating that the separation zone of these classes is more complex and shows greater difficulty in classification than the others.

In order to better understand the behavior of the models in the Outer Race 7 mils and 14 mils classification interface, Figures 7 that shows histograms for a dignostic. Enabling the study of the stochastic behavior of the

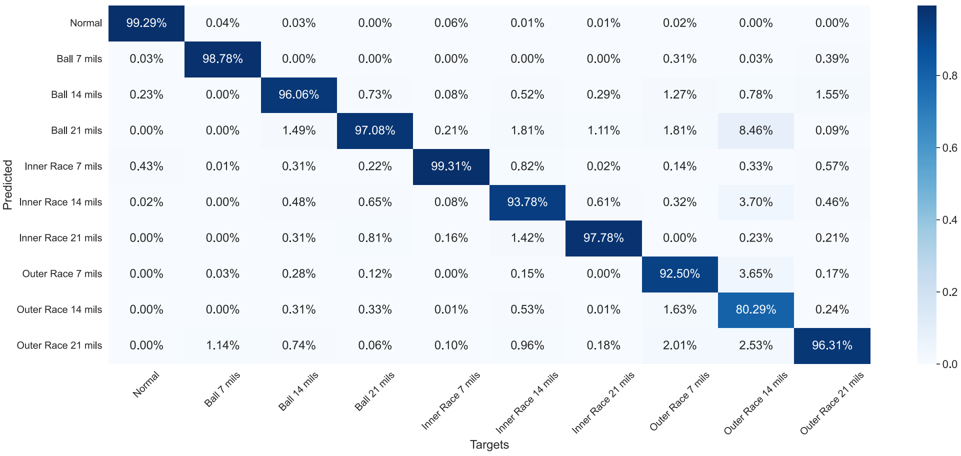


Figure 4. Mean Confusion Matrix for Variational Inference Model.

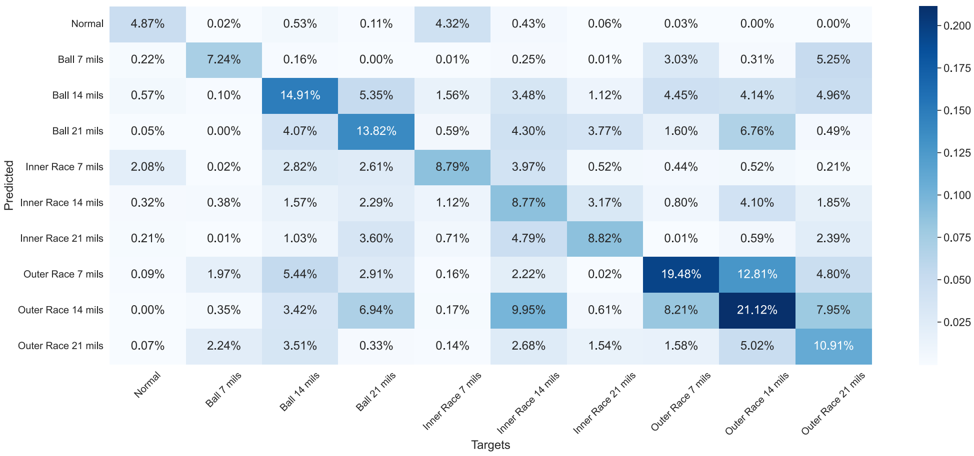


Figure 5. Confusion matrices of the standard deviation values related to the Uncertainty in each class for Monte Carlo Dropout Model.

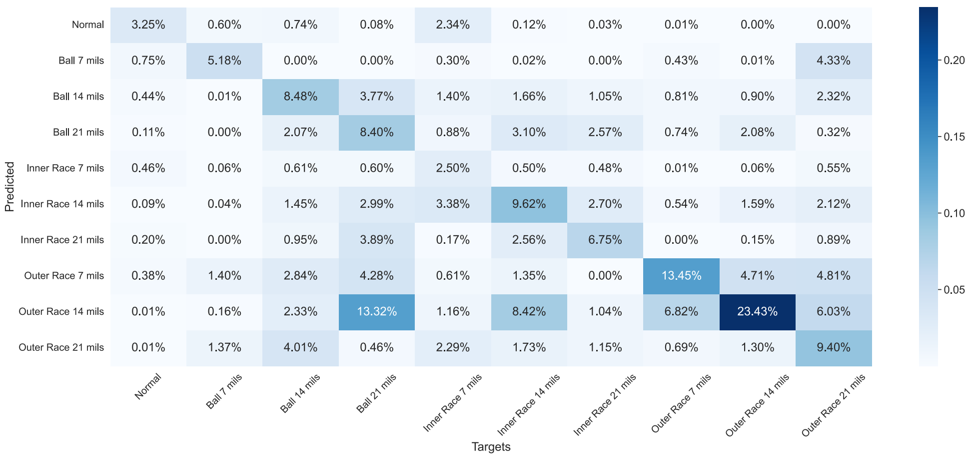
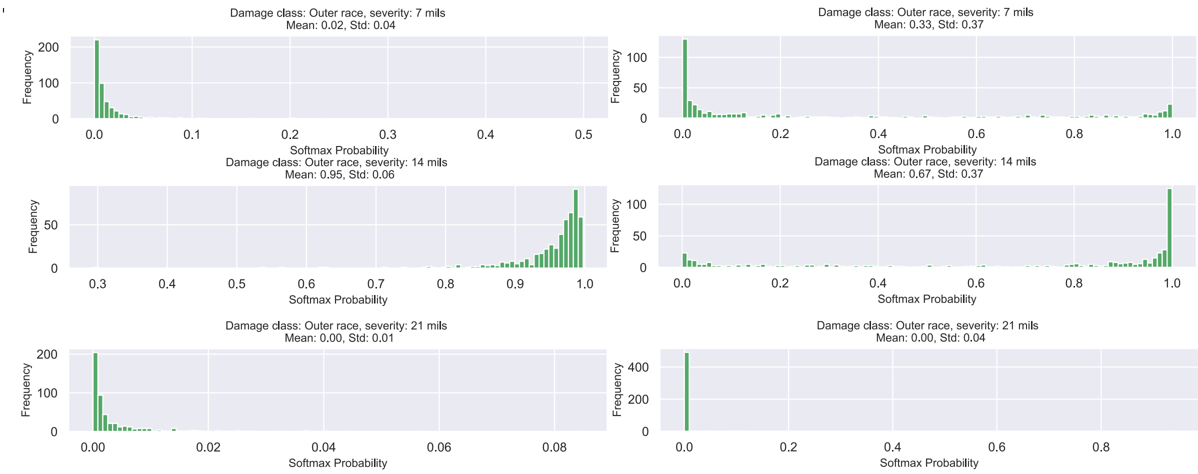


Figure 6. Confusion matrices of the standard deviation values related to the Uncertainty in each class for Variational Inference Model.

models and how they are producing uncertainties. From the appreciation of the results for this Outer Race 14 mils record, it is notable that the Monte Carlo Dropout model (Fig. 7) has produced more sparse predictions distributions, however, in all predictions the classification was correct. However, in the application of Variational Inference (Fig. 7b) there is in some samples a misclassification, causing confusion between the Outer Race 7 and 14 mils classes.



a) MCDropout b) Variational Inference

Figure 7. Histograms of a diagnosis of a outer race damage with a diameter of 0.014" for Monte Carlo Dropout and Variational Inference Model.

Finally, it is of due purpose that the behavior of the model is analyzed by taking into account the uncertainties in the classification. This aspect is central when the subject is Bayesian models. The graph of Figure 8 shows the behavior of model accuracy *versus* increasing uncertainty in the predictions. It is easy to conclude that both models decreased their accuracy as the average uncertainty in the sample predictions increased. This aspect indicates that the greater the uncertainty in the prediction, the lower the observed accuracy. The uncertainty of the two models is well expressed when this phenomenon is observed.

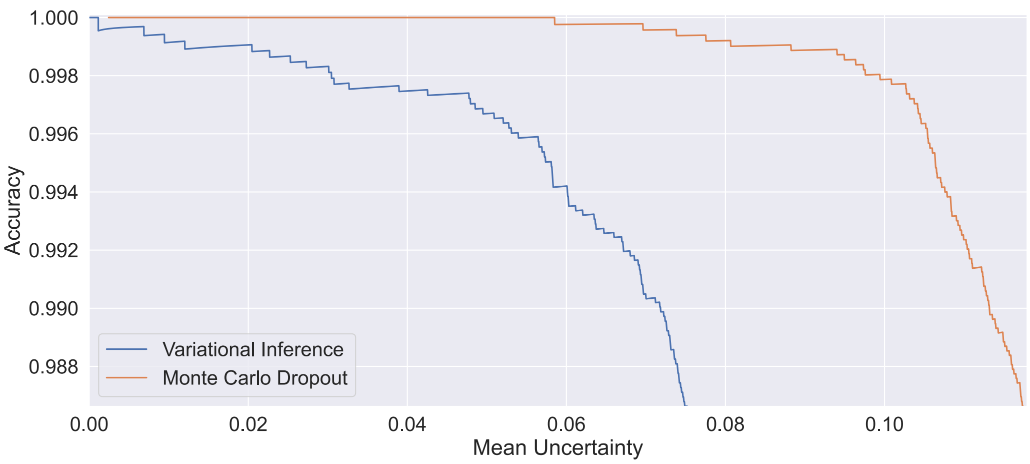


Figure 8. Graphic of Accuracy *versus* Mean Uncertainty.

4 Conclusion

To compare the techniques, the article presented the theoretical background and an application with a benchmark data set using validation accuracy, test and uncertainty metrics. Trends were observed that may be relevant in solving new problems, such as the better generalization of the model using the Monte Carlo Dropout technique, and the higher training speed when using Variational Inference.

In a certain way the Monte Carlo Dropout technique is easier to use considering the aspects of theoretical understanding and code construction, however, understanding and using uncertainty requires the comprehension of Bayesian inference concepts that are common to the application of both techniques.

References

[1] 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.

[2] Y. Gal and Z. Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning, 2015.

[3] H. H. Lee and H. Kim. Bayesian deep learning–based 1h-mrs of the brain: Metabolite quantification with uncertainty estimation using monte carlo dropout. *Magnetic Resonance in Medicine*, vol. 88, n. 1, pp. 38–52, 2022.

[4] L. Ju, X. Wang, L. Wang, D. Mahapatra, X. Zhao, Q. Zhou, T. Liu, and Z. Ge. Improving medical images classification with label noise using dual-uncertainty estimation. *IEEE Transactions on Medical Imaging*, vol. 41, n. 6, pp. 1533–1546, 2022.

[5] J. Bae, J. W. Park, and S. J. Lee. Limit surface/states searching algorithm with a deep neural network and monte carlo dropout for nuclear power plant safety assessment. *Applied Soft Computing*, vol. 124, pp. 109007, 2022.

[6] H. Tong, J. M. Hauth, X. Huan, B. Y. Zhou, N. R. Gauger, M. C. Morelli, and A. Guardone. Bayesian recurrent neural networks for monitoring rotorcraft icing from aeroacoustics time-series data. In *AIAA Scitech 2022 Forum*, pp. 2358, 2022.

[7] W. Peng, Z.-S. Ye, and N. Chen. Bayesian deep-learning-based health prognostics toward prognostics uncertainty. *IEEE Transactions on Industrial Electronics*, vol. 67, n. 3, pp. 2283–2293, 2020.

[8] R. Wang, H. Chen, and C. Guan. A bayesian inference-based approach for performance prognostics towards uncertainty quantification and its applications on the marine diesel engine. *ISA transactions*, vol. 118, pp. 159–173, 2021.

[9] de M. Moraes, J. P. Dias, and de H. F. Castro. Condition monitoring of ball bearings using bayesian neural networks. In *Proceedings of the XLII Ibero-Latin-American Congress on Computational Methods in Engineering and III Pan-American Congress on Computational Mechanics, ABMEC-IACM*, pp. 1–7, Rio de Janeiro, 2021.

[10] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.

[11] A. Damianou and N. D. Lawrence. Deep Gaussian processes. In C. M. Carvalho and P. Ravikumar, eds, *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, volume 31 of *Proceedings of Machine Learning Research*, pp. 207–215, Scottsdale, Arizona, USA. PMLR, 2013.

[12] 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.

[13] 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.

[14] 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.

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.