Bayesian Convolutional Neural Networks for Remaining Useful Life Prognostics of Solenoid Valves With Uncertainty Estimations

Tamir Mazaev ⓘ, Guillaume Crevecoeur ⓘ, and Sofie Van Hoecke ⓘ

Abstract—**Solenoid valves (SV) are essential components of industrial systems and therefore widely used. As they suffer from high failure rates in the field, fault prognosis of these assets plays a major role for improving their maintenance and reliability. In this work, Bayesian convolutional neural networks are used to predict the remaining useful life (RUL) of SV, by training them on the valve's current signatures. Predictive performance is further improved upon by using salient physical features obtained from an electromechanical model as the network's training input. Results show that our designed network architecture produces well-calibrated uncertainty estimations of the RUL predictive distributions, which is an important concern in prognostic decision-making.**

Index Terms—**Artificial neural networks, Bayes methods, machine learning, predictive maintenance, prognostics and health management, remaining life assessment, solenoid valve, occlusion, uncertainty.**

I. INTRODUCTION

THE purpose of prognostics and health management (PHM) is to enable optimal maintenance strategies, as to prevent machine failure, extend the lifetime of machines and reduce operational costs. This is achieved by detecting incipient faults, fault isolation, identification of different fault types (fault diagnostics), and fault prognosis. These techniques typically imply the analysis of healthy and/or faulty conditions indicated by process measurements.

Two main approaches exist for estimating the remaining useful life (RUL). Physics-based approaches rely on physical domain knowledge, which describe normal operation and physical degradation laws. Data-driven approaches are based on condition monitoring data, which are used for constructing statistical or machine learning models [1]. The term "hybrid approaches" is commonly used for approaches that combine physics-based and data-driven techniques.

Deep learning (DL) techniques are an important subcategory of data-driven approaches. Features are learned automatically at multiple levels of feature representations, which allows DL to learn complex relations mapping the input to the output directly. This contrasts DL with feature-based approaches such as decision tree ensembles or Gaussian process regression, which rely on the construction of features [2]. DL is particularly well positioned to solve the following issues in prognostics [3].

1) Automatic processing of massive amounts of data.
2) Automatic extraction of useful features from high-dimensional, heterogeneous data sources.
3) Learning temporal and functional relationships between and within time series.
4) Transfer knowledge between different units and operating conditions.

Despite advancements in DL-based prognostics, most works only provide deterministic RUL predictions, without capturing uncertainties [4]. When predicting the future behavior of systems in prognostics, various sources of uncertainty are involved, such as noisy sensor data and predictive uncertainty associated with randomness in future operating conditions. Predictive uncertainty quantification (UQ) thus has an important role in estimating the RUL of engineering systems. Pointwise predictions without UQ do not indicate how much confidence a DL method has in its predictions, which is a critical issue in prognostic decision-making.

Bayesian modeling is an effective framework for dealing with uncertainty in machine learning, including complex neural network models. Bayesian hypothesis reasoning is used to improve generalization performance, and to construct credible intervals for statistical inference. Modern Bayesian neural networks (BNNs) can be trained effectively through techniques such as Monte Carlo dropout or mean field variational inference (VI) [5]. Bayesian convolutional neural networks (BCNNs) have been successfully used for large-scale image classification tasks with UQ [6].

In this work, BCNNs are applied to predict the RUL of solenoid valves (SV), trained on image representations of its

Tamir Mazaev and Sofie Van Hoecke are with the Internet Technology and Data Science Lab (IDLab), imec, Ghent University, 9000 Ghent, Belgium (e-mail: tamir.mazaev@ugent.be; sofie.vanhoecke@ugent.be).

Guillaume Crevecoeur is with the Department of Electromechanical, Systems and Metal Engineering, Ghent University, 9000 Ghent, Belgium, and also with the EEDT-DC core lab, Flanders Make, 3920 Ghent, Belgium (e-mail: guillaume.crevecoeur@ugent.be).

current signals. SVs are essential components of many industrial systems in process industry. For example, approximately 2–4% of SVs in a chemical plant are used in critical safety instrumented functions, often forming the most important components in the safety loop [7]. Hence, reliable monitoring of these components is needed. In this work, an accelerated lifetime setup is used to obtain SV life time data under realistic cycling conditions. Physical domain knowledge is used to improve the predictive performance of a BCNN trained on the SV's current signals. Two physical forces obtained as features from the current signals are included to the BCNN input, since they contain salient information on the degradation state of the valve [8]. Feature attribution methods [9] are applied to the trained BCNN, highlighting which aspects of its input are important for its predictions. For the BCNN model, we design an architecture that is able to infer uncertainty in a flexible way. The standard deviation of the predictive RUL distribution is predicted as a separate output parameter of the network architecture. We study the reliability of these UQs by evaluating calibration curves of the RUL test predictions.

The remainder of this article is organized as follows. Section II discusses related work. In Section III, the setup used to create the data and the dataset itself are discussed. In Section IV, a brief overview of BNNs is provided, and the proposed RUL prediction methodology. In Section V, the results of the presented methodology on our dataset is discussed. Section VI concludes this article.

II. RELATED WORK

A. Deep Learning in PHM

A recent review on DL PHM applications for sensor condition monitoring data [3] by Fink *et al.*, presents an overview of current developments, challenges, potential solutions, and future research. Considering the large amount of work in this field, we focus on recent applications of 1) convolutional neural networks (CNNs) and 2) BNNs for RUL prognostics. Convolutional architectures are most relevant for our work, since SV sensor signals are represented as training data in the form of images.

1) CNNs: Aydemir *et al.* [10] designed a neural network architecture consisting of convolutional layers for extracting spatial features, an long short-term memory (LSTM) layer for tracking temporal information, and another architecture, where the dimension of images was reduced by a deep autoencoder. These models were applied to simulation data and a dataset of infrared image streams from rotating machinery. Yuan *et al.* propose a CNN-based, data-driven end-to-end framework for monitoring manufacturing systems, which is evaluated on ten representative datasets [11]. Yang *et al.* [12] integrated two CNN architectures into one framework: A first CNN for incipient failure threshold identification, and a second CNN for RUL prediction. This framework was applied to vibration data of bearings. Liu *et al.* [13] propose a joint-loss CNN architecture, which can implement bearing fault recognition and RUL prediction in parallel. Li *et al.* [14] use a multiscale CNN, applied to modular aeropropulsion system simulation data (C-MAPSS dataset) from NASA. Kim *et al.* [15] also study the C-MAPSS
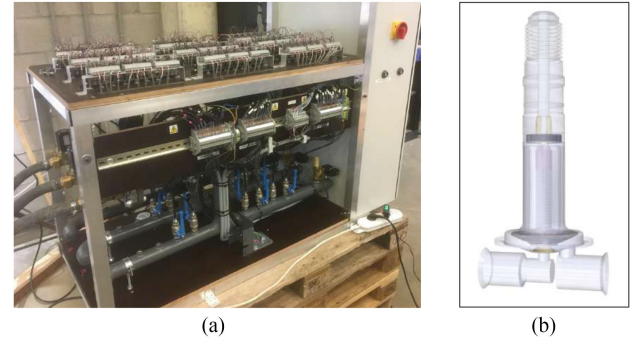

Fig. 1. (a) Experimental accelerated lifetime setup. 48 SVs (direct acting, 3/2 way, normally closed) are switched at a rate of 1 Hz for a total duration of 6 weeks, until EOL. Electrical current signals, the pressure of the air supplied to the valves, leakage, and temperature signals are measured. (b) CAD model of the internals of the SV.

dataset, using a CNN-based multitask learning method to reflect the relatedness of RUL estimation with a health status detection process. Ma *et al.* [16] propose a convolution-based LSTM that conducts convolutional operations on both the input-to-state and state-to-state transitions of an LSTM, applied to vibration data of rotating machinery.

2) BNNs: In contrast to the frequentist CNN (FCNN) models above, earlier work on BNNs for prognostics is now described. Peng *et al.* [4] use a Bayesian multiscale CNN and Bayesian bidirectional LSTM trained by VI, applied to a ball bearing dataset and the C-MAPPS dataset. Li *et al.* [5] propose recurrent NNs with gated recurrent units and a sequential Bayesian boosting algorithm applied to high-voltage circuit breakers, where epistemic and aleatoric uncertainty are considered separately. Kraus *et al.* [17] propose a Bayesian structured-effect neural network, evaluated on the C-MAPPS dataset. Gao *et al.* [6] also study this dataset, proposing a joint prognostic model, where Bayesian linear models are used in conjunction with an NN. Wang *et al.* [18] use VI to quantify the uncertainty of recurrent convolutional NNs in RUL prediction. The method is evaluated on vibration data from accelerated degradation tests of rolling element bearings and sensor data from life testing of milling cutters.

The above methodologies using BNNs all focus on purely data-driven prediction performance. In this work, we adopt a hybrid approach by adding salient physical features from a physical model [8] to the input of the BCNN, with the aim to lower RUL errors. Moreover, an important aspect not included in these works is that uncertainty estimations should show good calibration performance. Credible predictive intervals can only be relied on when they faithfully represent uncertainty.

B. Solenoid Valves

To the best of our knowledge, there exists no earlier work on UQ for SV diagnostics or prognostics. Previous work without UQ is now described. Jo *et al.* [19] propose a model-based fault detection method for SVs taken from real braking systems of urban railway vehicles. Guo *et al.* [20] use a data-driven fault diagnosis method for SVs using multikernel support vector
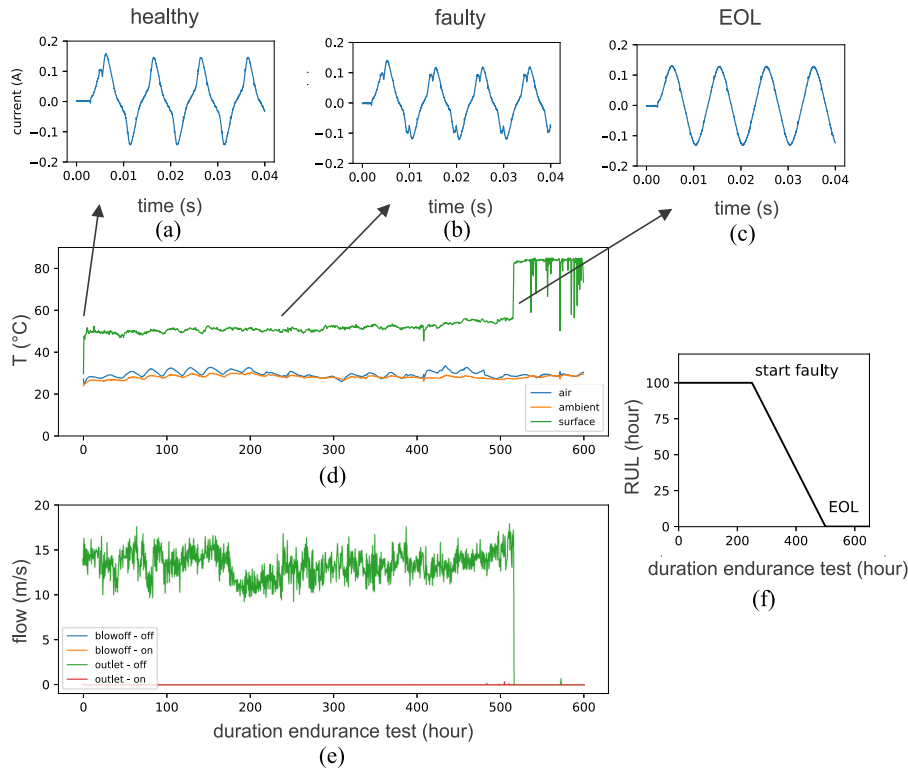
Fig. 2. Evolution of a SV current signal over the course of its lifetime. (a)–(c) Current of a valve deteriorating from healthy to faulty and EOL. In (d) and (e), temperature and leakage sensor measurements are shown respectively. In (f), a piece-wise linear function for modeling the RUL of the SV is shown.

machines. Liniger *et al.* [21] propose a model-based scheme for detecting the early signs of coil failure in SVs. The method employs a thermal model of the solenoid and an extended Kalman filter for generating coil current residuals. Sarwar *et al.* [22] developed an algorithm for fault isolation and diagnosis of high-pressure fuel pump SVs using current feedback. Mazaev *et al.* [23] proposed data-driven RUL prediction approaches using shallow feature-based approaches and an ensemble of CNNs to construct a health index of SVs, which are used to extrapolate RUL. This work directly compares the performance of the presented methodology with these results.

III. Data

A. Setup

An experimental endurance test setup, as shown in Fig. 1(a), was used for monitoring the degradation of 48 SVs. These are direct acting 3/2 way normally closed SVs (Burkert Type 6014). By switching the valves at a rate of 1 Hz for a total duration of 6 weeks, accelerated life testing was achieved. One acquisition of the valves switching ON and OFFwas captured every hour. Each valve was switched by an input voltage of 110-V ac at 50 Hz. Compressed air at 8 bar (g) was supplied to the valves at an ambient temperature of 25 °C. The current signal through the SV was monitored, as well as the temperature and pressure of the air being supplied to the valves. Supplementary sensors measured the surface temperatures of and the ambient temperatures around the valves. Thermal mass flow was measured for

TABLE I
SENSOR USED FOR EACH DATA MEASUREMENT
IN THE EXPERIMENTAL SETUP

Measurement	Sensor
Current	LEM CTSR 0.3-P
Temperature	Pt100
Pressure	GEMS 3100
Thermal mass flow	IST AG FS7

detecting leakage at the outlet ports and ventholes (blow-off holes) of the valves. The sensor type used for each measurement is summarized in Table I. The internals of the SV, such as the plunger and return spring, are shown in Fig. 1(b).

B. Dataset

Physical models derived from first principles couple the current signal of the SV to dynamical states such as the magnetic flux and the plunger position [8]. When energizing or deenergizing the valve, its mechanical deterioration is visible through this current signal.

Three classes of current signals can be distinguished over the lifetime of a solenoid when it is being energized, as shown in Fig. 2. These classes are very distinct from another, as Tod *et al.* [8] have shown that a CNN is able to classify them with 99% accuracy. In its 1) healthy state, closing the valve results in one isolated hit of the plunger. Following this healthy period, a

2) faulty regime begins that is distinguished by several hits of the plunger, which is visible in the current signal. The transition from 1) to 2) cannot be detected with flow or temperature measurements, but can be detected using the CNN in [8]. Lastly, 3) EOL is reached when the plunger remains stuck in the armature tube, and no hits appear anymore. It can be observed that the EOL of the valve is characterized with a steep increase of its 4) surface temperature. The EOL is also characterized with a steep drop of the flow at the outlet port right before the valve is 5) switched OFF. When the valve is still operating as intended, a flow is expected through the outlet port since the valve is normally closed. This flow being blocked validates that the plunger remains stuck. The flow and temperature measurements can thus be used jointly to clearly mark the EOL of the system.

These observations are characteristics for all valves reaching EOL. 40 out of the 48 valves monitored reached EOL at the termination of the accelerated lifetime test.

IV. METHODOLOGY

A. RUL Model

Since the current signals of the SV of the healthy class [Fig. 2(a)] show little change over the valve's lifetime, degradation is modeled starting from the first faulty signal [Fig. 2(f)]. From that point onward, RUL predictions are made until EOL.

B. Current Signals as Images

As described in Section III, temperature and flow measurements are needed to validate the exact moment of EOL of the SVs. However, we aim to limit the amount of expensive sensors needed. Our RUL model thus relies on current signals, which only require access to the circuit that carries the input power of the valve. Hence, the valve's degradation is detected in a noninvasive manner, and no extra sensors are required once the solenoid is deployed. For PHM in general, this is recognized as an important benefit of current signature analysis [24].

As described in the works on CNNs summarized in Section II, a highly effective way to perform fault diagnostics and prognostics on CBM sensor data, is to train CNN models on time-series sensor data represented as images. In Fig. 3(a), it is shown how images are constructed from the valve's current signals. Every column of the image corresponds to a signal from one data acquisition. These are combined into an image by concatenating them horizontally. Each acquisition is spaced one hour of accelerated life time testing apart, and every image consists of 30 signals in total. As a result, every image encompasses 30 h of accelerated lifetime testing, still representing the current signals in the time domain. Every current signal is a time series of 200 values, so the image has a dimension of [200 × 30]. By taking this sliding window of 30 h over a valve's lifetime, these images were constructed for every valve. The step size used between every subsequent image is 17 acquisitions (=17 h). The two parameters with values 30 and 17 were chosen as heuristic values, by manually looking at the evolution of the current signals. These have been found to have a negligible effect on RUL prediction results (including uncertainties). Every image is
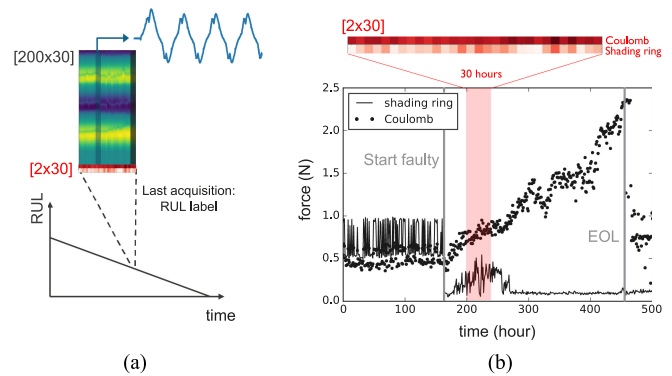

Fig. 3. (a) Image representation of the valve current signals. 30 subsequent current acquisitions of one valve, spaced 1 h of accelerated lifetime testing apart, are combined in an image. Every column (=1 current signal) is extended with two physical features computed from the corresponding current signal (a different color scale in red for the features is used for clarity). Every image is labeled with the RUL value corresponding to the last current acquisition in the image. (b) Evolution of shading ring force and Coulomb friction over the lifetime of a valve. It is shown how the force values can be represented as pixel values in an image. Larger force values correspond to darker pixel values.

labeled with the RUL value corresponding to the last acquisition in the image. The total dataset consists of 1487 labeled images.

The image is extended from a [200 × 30] matrix to a [202 × 30] matrix, by extending every column (= 1 acquisition) with two physical features computed from the current signal corresponding to that acquisition. All current signals are multiplied with a factor 5, such that they have similar magnitudes compared to the physical features. These physical features are described in the subsection below.

C. Addition of Physical Features

In the work of Tod et al. [8], two physical forces were identified based on the current signals of the SVs used in our study. The first force is a shading ring force. Next to the main coil of the SV, this is a secondary source of electromagnetic force originating from a shading ring, which is modeled as periodic function delayed from the main coil force. The purpose of the shading ring is to avoid a zero electromagnetic force when the control current is zero. The second force is a kinetic Coulomb friction force, including a static and dynamic term, for which the need was revealed by visual and manual actuation of damaged valves. Both the shading ring and Coulomb force were used by Tod et al. to extend a classical SV physics model. We refer to [8] for a full discussion of the complete model. An important consideration is that the variations of these two forces over the lifetime of the valve are found to be sufficient to explain the occurrence of the different current signal classes, enabling model-based fault diagnostics.

The evolution of the two forces in a single example valve is shown in Fig. 3(b). The correspondence between the force values and their representation as pixels in an image is also shown. In Section IV-E below, it is explained how these images are used as CNN training data for the hybrid models presented. The performance of all CNN models is compared for the case when

only the current signals are used (images of dimension $[200 \times 30]$) versus the case when the physical features are included as well ($[202 \times 30]$), or when using physical features only ($[2 \times 30]$).

D. Bayesian Neural Networks

In the Bayesian view of statistics, unknown latent parameters (or model weights $\boldsymbol{\omega}$) are treated as random variables. A distribution of the latent parameters is learned conditionally on the training data [2]. The posterior distribution over the model weights can be obtained through Bayes theorem

$$\pi(\boldsymbol{\omega}|\mathcal{D}) = \frac{p(\boldsymbol{\omega})p(\mathcal{D}|\boldsymbol{\omega})}{\int p(\boldsymbol{\omega})p(\mathcal{D}|\boldsymbol{\omega})\mathrm{d}\boldsymbol{\omega}} \tag{1}$$

where $\mathcal{D}$ denotes the training data $(\mathbf{x}, \mathbf{y})$.

Quantities f, such as the predictive mean and variance, are obtained as an expectation value over the posterior distribution

$$\mathbb{E}_\pi[f] = \int f(\boldsymbol{\omega})\pi(\boldsymbol{\omega}|\mathcal{D})\mathrm{d}\boldsymbol{\omega}. \tag{2}$$

Current research in BNNs is mostly focused on VI approaches since training through backpropagation, which is used for frequentist NNs, can also be implemented. VI is an approximate inference method that casts Bayesian inference as an optimization problem. An approximating family of distributions $q_{\boldsymbol{\theta}}(\boldsymbol{\omega})$, parametrized through variables $\boldsymbol{\theta}$, is used to represent the posterior distribution. An optimization objective is defined to capture the similarity between the approximate distribution and the true posterior distribution. The Kullback–Leibler (KL) divergence is used as the optimization objective, which can be expanded as follows:

$$\begin{aligned} \mathrm{KL}\left(q_{\boldsymbol{\theta}}(\boldsymbol{\omega})||\pi(\boldsymbol{\omega}|\mathcal{D})\right) &= \int q_{\boldsymbol{\theta}}(\boldsymbol{\omega})\log\frac{q_{\boldsymbol{\theta}}(\boldsymbol{\omega})}{p(\boldsymbol{\omega}|\mathcal{D})}\mathrm{d}\boldsymbol{\omega} \\ &= \mathbb{E}_q\left[\log\frac{q_{\boldsymbol{\theta}}(\boldsymbol{\omega})}{p(\boldsymbol{\omega})} - \log p(\mathcal{D}|\boldsymbol{\omega})\right] + \log p(\mathcal{D}) \\ &= \mathrm{KL}\left(q_{\boldsymbol{\theta}}(\boldsymbol{\omega})||p(\boldsymbol{\omega})\right) - \mathbb{E}_q\left[\log p(\mathcal{D}|\boldsymbol{\omega})\right] + \log p(\mathcal{D}) \\ &= -\mathcal{F}(q_{\boldsymbol{\theta}}(\boldsymbol{\omega})) + \log p(\mathcal{D}) \end{aligned} \tag{3}$$

where $\mathcal{F}(q_{\boldsymbol{\theta}}) = -\mathrm{KL}(q_{\boldsymbol{\theta}}(\boldsymbol{\omega})||p(\boldsymbol{\omega})) + \mathbb{E}_q[\log p(\mathcal{D}|\boldsymbol{\omega})]$ is the evidence lower bound (ELBO). The ELBO is the only term of the KL-divergence that needs to be optimized through backpropagation, since the log marginal likelihood of the expression does not depend on the model variational parameters $\boldsymbol{\theta}$.

To deal with the large number of model weights encountered in deep NNs, the mean-field approximation is commonly used. In mean-field variational inference (MFVI), the posterior distribution is factorized through independent Gaussians

$$q_{\boldsymbol{\theta}}(\boldsymbol{\omega}) = \prod_{i=1}^{P} \mathcal{N}\left(\omega_i|\mu_i, \sigma_i^2\right) \tag{4}$$

which allows for a tractable optimization of the network.

For the large-scale datasets typically used in DL, it is infeasible to evaluate the full log likelihood for training. The dominant method used to train DL is minibatch gradient descent, a variation of the stochastic gradient descent algorithm. The training dataset is split into smaller batches that are used to calculate model error and update model weights. For VI, the variational objective becomes

$$\mathcal{L}(\boldsymbol{\omega}, \boldsymbol{\theta}) = -\frac{N}{M}\sum_{i=1}^{N}\mathbb{E}_q\left[\log p(\mathcal{D}_i|\boldsymbol{\omega})\right] + KL\left(q_{\boldsymbol{\theta}}(\boldsymbol{\omega})||p(\boldsymbol{\omega})\right) \tag{5}$$

where $\mathcal{D}_i \subset \mathcal{D}$ and each minibatch is of size $|M|$.

E. Proposed Models

Fig. 4 shows our proposed FCNN and BCNN architectures for the RUL prognosis task. The training data consists of SV current signals and two physical features represented as images with their corresponding RUL labels, as discussed in Section IV-B.

Since the CNN shown in Fig. 4(a) is a frequentist network, its predictions consists of pointwise RUL values. In order to avoid negative RUL predictions, which do not appear in the dataset, a rectified linear unit (ReLU) is used in the last network layer. For the case, where only the physical features are used as training data, Conv2D kernels of size $[2 \times 2]$ are used to match the input dimension of size $[30 \times 2]$. The FCNN is trained with the Adam optimizer on an mse loss, with a learning rate of 0.001, and a batch size of 64. Early stopping (patience set to 8 epochs) is used to prevent overfitting. The model is reset to the best model evaluated on a validation set. Leave-one-out cross validation is used for testing. As the CNN uses images from 35 valves for training, and 4 valves for validation, the remaining valve is used for testing. This is repeated for all 40 valves to cross validate. The total dataset consists of 1487 images. The valve with the longest lifetime contributes to 75 labeled images in the dataset, while the valve with the shortest lifetime contributes to 20 images. As such, the minimum amount of images used for training and validation is 1412, while the maximum amount is 1467.

A first approach to convert the FCNN into a BCNN is shown in Fig. 4(b), as proposed in earlier work by Peng *et al.* [4] for RUL prognosis tasks. Dropout layers (dropout rate = 0.2) are added just before the convolutional layers, in addition to the dropout layers already used for the dense layers. Posterior RUL predictions are obtained by applying stochastic forward passes through the trained CNN with dropouts enabled at inference time. Gal *et al.* have proven that this procedure approximates an FCNN to its Bayesian counterpart trained through VI [6]. This method is referred to as "BCNN, MC dropout." To obtain a predictive RUL distribution for every test sample, a truncated Gaussian distribution is fitted to 40 RUL values obtained through stochastic forward passes. This distribution is a Gaussian distribution of which the random variables are bounded to nonnegative values (as illustrated in Fig. 4). As a consequence, negative RUL predictions are avoided, similar to the FCNN. For training, the same settings are used as for the FCNN.

We also propose a second approach to convert the FCNN into a Bayesian architecture, as shown in Fig. 4(c). A key difference of this architecture concerns the dense part of the network after flattening the reduced features of the convolutional and maxpool layers, which is split into two parts. One dense subnetwork predicts the mean $\hat{\mu}_{\mathrm{RUL}}$ of a truncated Gaussian distribution
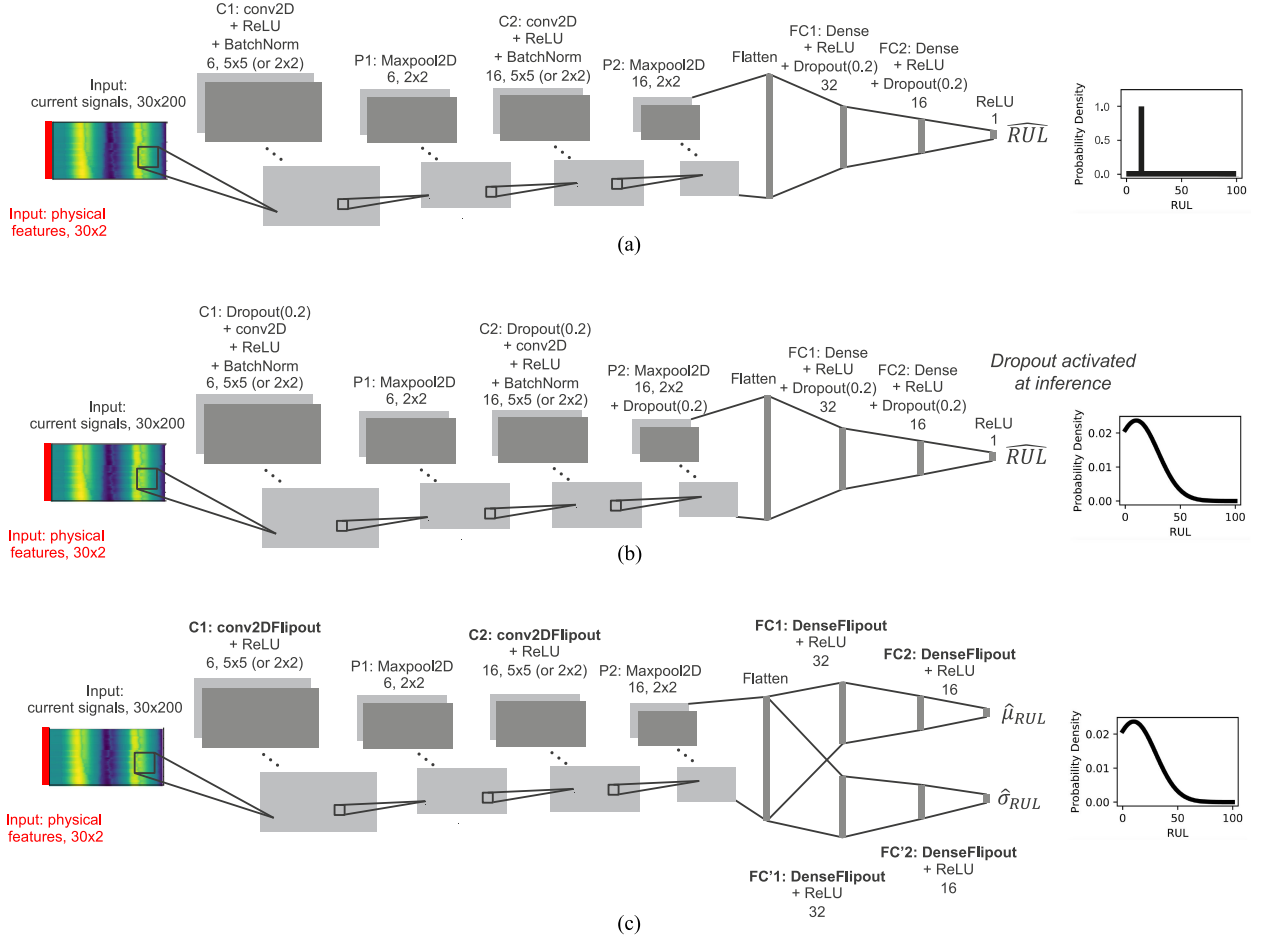
Fig. 4. Comparison between the architectures of the (a) frequentist and (b), (c) Bayesian CNNs designed for the RUL prognosis task. When training on physical features only, conv2D kernels of dimension [2 × 2] (instead of [5 × 5]) are used. The Bayesian networks are able to fit the mean and standard deviation of a truncated normal distribution, while the frequentist network is limited to pointwise predictions. Bayesian layers trained through MFVI in (c) are indicated in bold.

(bounded to nonnegative values), the other predicts its standard deviation $\hat{\sigma}_{RUL}$. In contrast to the MC dropout architecture, the mean and variance of the truncated Gaussian distribution per RUL test sample are learned explicitly as predictive outputs. The BCNN is trained through MFVI of the distributions (4), which allows for training through backpropagation of objective (5). Thereby, we refer to this architecture as "BCNN, MFVI." Note that μ_{RUL} and σ_{RUL} do not need to be explicitly labeled in the dataset, but are optimized through the log likelihood of the data [in the first term of (5)]. Just as for the FCNN, we use an Adam optimizer on an mse loss, with a learning rate of 0.001, and a batch size of 64. The prior distribution [represented as $p(\boldsymbol{\omega})$ in (1)] set for every model weight is a standard normal distribution. Leave-one-out cross validation is again used for testing on the same dataset. We use Flipout for training the Bayesian layers, which is an effective method to decorrelate the gradients within a minibatch, as to obtain unbiased stochastic gradient estimates [25]. Flipout layers implicitly sample pseudoindependent weight perturbations for every update in a variational BNN. These layers are indicated in bold in Fig. 4(b).

Given the relatively small size of the neural networks, offline training on a GPU can occur in a few minutes. Inference can occur in the order of seconds. Since the transition of a healthy SV toward EOL needs weeks of accelerated lifetime testing and training, and inference can thus be performed on timescales that are in the order of magnitudes smaller.

In the next section, the performance of the FCNN is compared to the BCNNs. We study the effect of including the physical features to the input of the networks. Considering the truncation of the Gaussian predictive output, we also evaluate its effect on RUL prediction performance and uncertainty calibration. Well-calibrated predictive credible intervals (e.g. 90% confidence) should match the true observed RUL outcomes (e.g. 90% of the samples fall within the interval).

V. RESULTS

A. RUL Prediction

An example of RUL values predicted by the CNN models for test valve 10 is shown in Fig. 5. On the top of the figure, results are shown when only using the SV current signals as training input. In the middle, this is shown when only physical features are used. The bottom figure shows the combination. In contrast to the FCNNs, the BCNNs construct credible intervals,
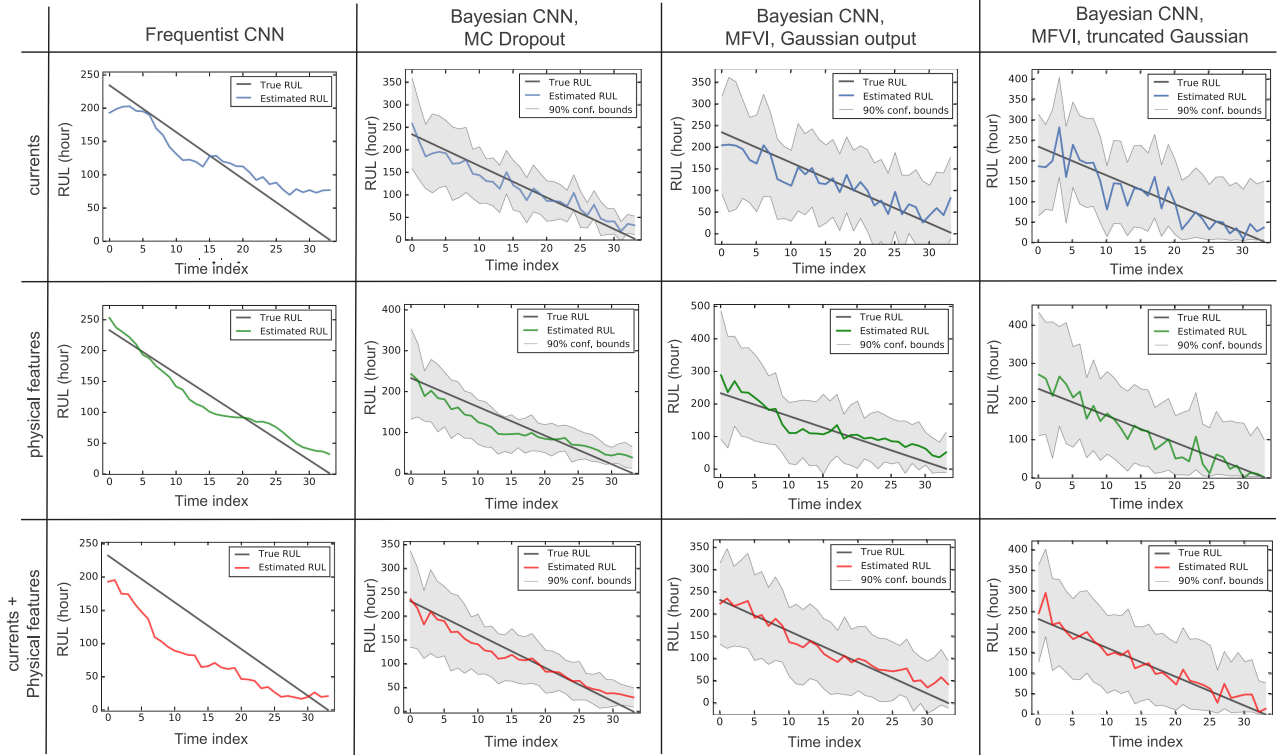
Fig. 5.   Comparison of RUL test results for valve 10. Bayesian CNNs allow for the construction of a 90% credible interval on the RUL estimations. Note that a Gaussian output without truncation allows for negative RUL predictions.
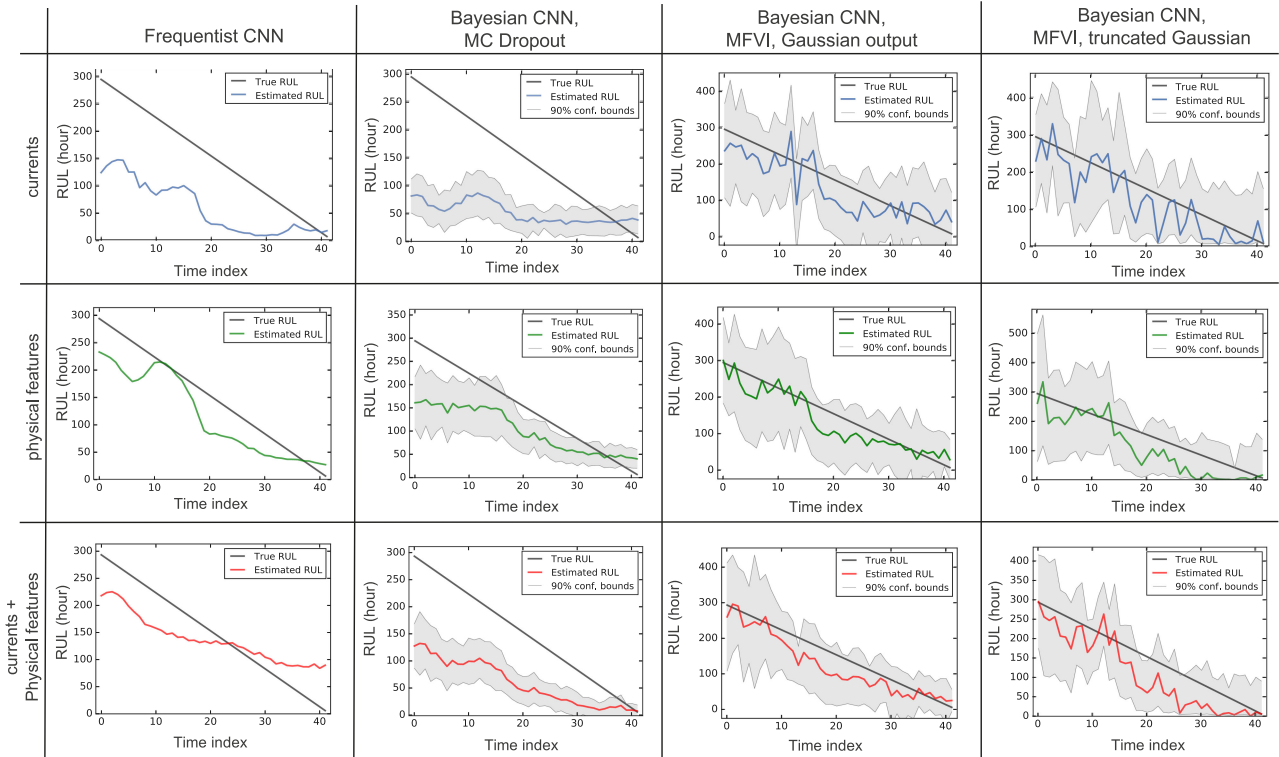

Fig. 6.   Comparison of RUL test results for valve 16. The BCNN with MC dropout severely underestimates the true RUL and empirical uncertainty for this valve.
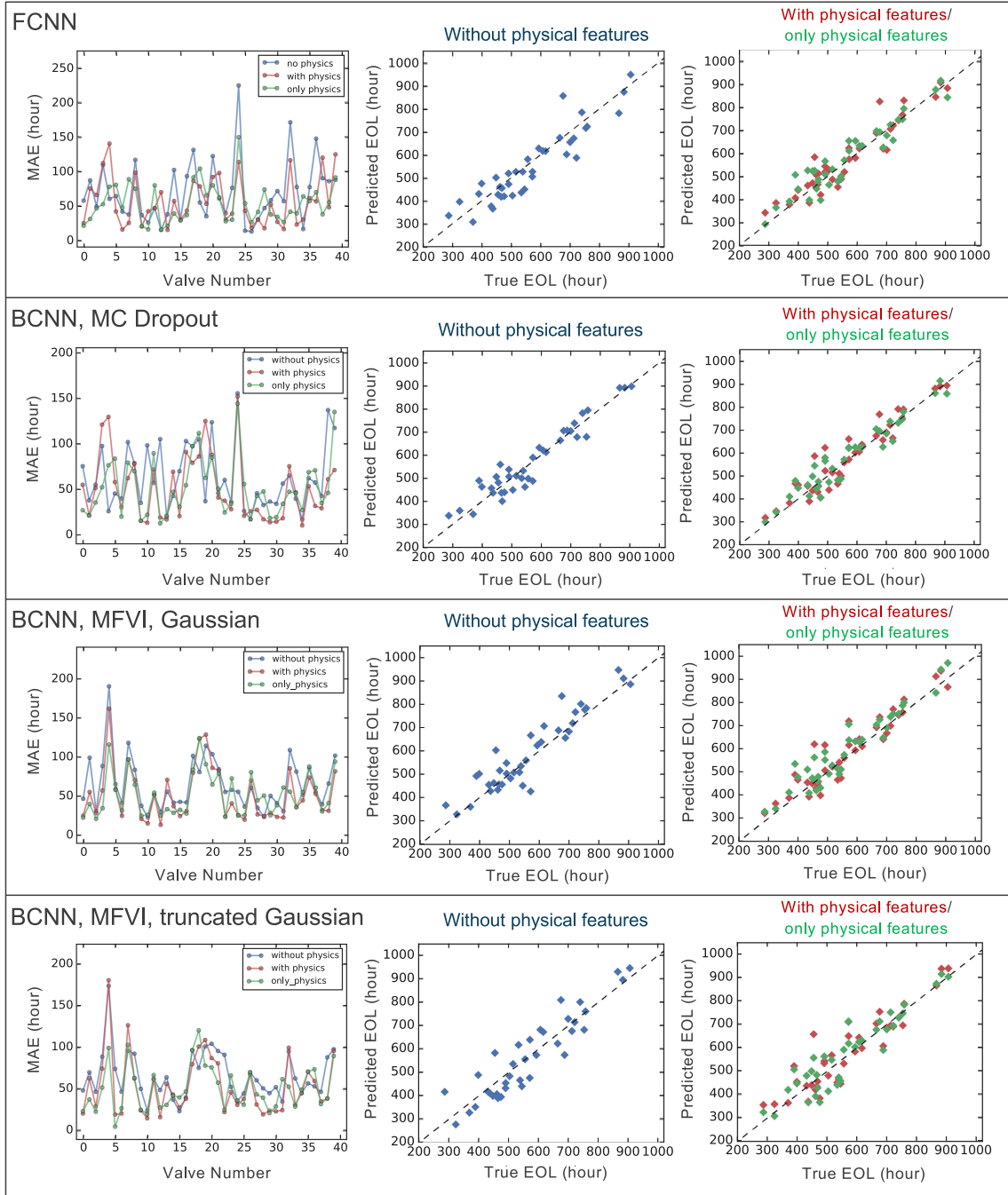
Fig. 7.   Comparison of predicted EOL test results for all 40 test valves. On the leftmost figures: Comparison of MAE averaged over all EOL values predicted per valve. On the middle and rightmost figures: True versus predicted EOL at 70% degradation of each valve.

hence 90% credible intervals are shown for these CNNs. For illustration purposes, results for the BCNN trained by MFVI are shown when the Gaussian distribution is not truncated. When the output of the BCNN is not restricted toward nonnegative values, negative values can occur in the RUL predictions as can be seen in the figure. Another observation is that the BCNN trained and evaluated by MC dropout results in uncertainty intervals that are more narrow. However, this is not always justified for other test valves. For example, RUL prediction results for valve 16 in Fig. 6 show that the BCNN trained with

MC dropout underestimates the RUL by a large margin. Our proposed architecture, trained by MFVI, does not exhibit this problem. In Section V-B, the trustworthiness of the uncertainty intervals constructed by the BCNN models is compared over all test valves through calibration plots.

By computing the mean absolute error (MAE) over the predicted and true RUL value over the degradation region (as shown in Fig. 5), an averaged prediction error is obtained per valve. This MAE is shown per valve in the left part of Fig. 7. The errors for every CNN model are compared. On the middle and right part

TABLE II
MAE AVERAGED PER VALVE, AND MAE AT 70% DEGRADATION PER VALVE

Method	Average MAE	MAE at 70%
FCNN, no physics	69.5 ± 44.5	53.8 ± 39.5
FCNN, with physics	57.0 ± 35.2	41.3 ± 35.0
FCNN, only physics	56.7 ± 28.2	42.8 ± 30.2
BCNN, MC Dropout, no physics, trunc. Gaussian	63.2 ± 34.8	39.0 ± 28.8
BCNN, MC Dropout, with physics, trunc. Gaussian	51.8 ± 36.1	36.9 ± 33.8
BCNN, MC Dropout, only physics, trunc. Gaussian	51.7 ± 32.0	40.16 ± 28.8
BCNN, no physics, Gaussian	65.5 ± 33.4	48.0 ± 43.4
BCNN, with physics, Gaussian	52.5 ± 33.9	44.6 ± 39.6
BCNN, only physics, Gaussian	52.5 ± 27.4	46.7 ± 34.2
BCNN, no physics, trunc. Gaussian	66.6 ± 28.6	60.7 ± 38.9
BCNN, with physics, trunc. Gaussian	54.0 ± 35.1	53.0 ± 43.6
BCNN, only physics, trunc. Gaussian	51.1 ± 26.4	51.9 ± 36.1
AdaNet (CNN, feature learning)	-	82.7 ± 79.3
Linear regression (feature-based)	-	76.6 ± 81.5
GBRT (feature-based)	-	97.6 ± 116.4

The mean and standard deviation of these values over all 40 test valves are given.

Fig. 8. Calibration curves for the RUL test results of the BCNNs.

of the figure, the predicted versus true EOL at 70% degradation per valve is shown for every CNN model.

The mean and standard deviation of the average MAE per valve, and MAE at 70% degradation per valve can be computed over all 40 test valves. These values are summarized in Table II. These metrics serve as an indication of the overall performance of the proposed CNN models for the RUL prediction task. The CNN models are compared with the deep and shallow learning methods used in [23]. In this work, health indices are computed up to 70% degradation, after which these health indices are extrapolated in order to compute the RUL. Therefore, only the MAE at 70% degradation can be used as a comparison.

We compare the error distributions by using paired *t*-tests. An assumption of this statistical test, is that the underlying differences between two sets of values are normally distributed. By using the Shapiro–Wilk test, the null hypothesis of the distribution being normally distributed cannot be rejected for all error distribution comparisons ($p > 0.05$ for all) in Table II. Thus, paired *t*-tests can be used for comparing all methods. A first observation is that every CNN model proposed in our work performs better than AdaNet or the feature-based approaches ($p \sim 0$ for every combination, RUL MAE is reduced by $\sim 40\%$). This signifies that the proposed methodology is better suited for the RUL prediction task than current state-of-the-art solutions. A second observation is that by adding the physical features
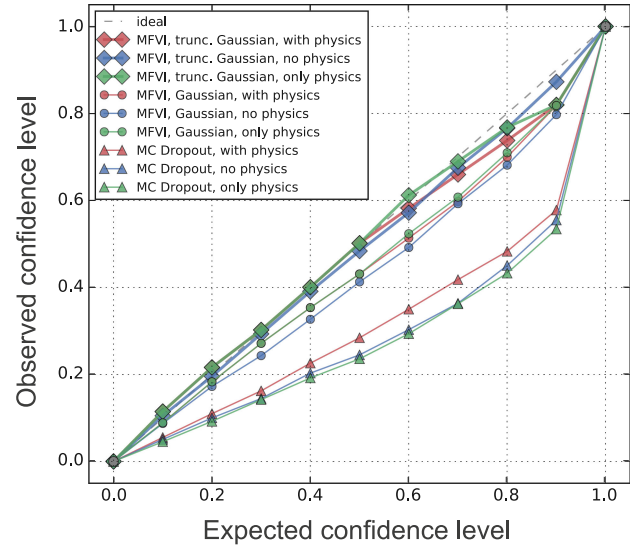
to the input of the CNN models proposed in this work, every model shows improved performance at the RUL prediction task ($p < 0.01$ for the FCNN and BCNNs, RUL MAE is reduced by $\sim 20\%$.). A third observation is that adding the current signals to the training input (next to the physical features) brings no performance gain ($p > 0.05$ for every network architecture). A fourth observation, is that the FCNNs and BCNNs amongst themselves perform similarly when they are trained by using the same input (with or without physical features, $p > 0.05$ for every comparison). Note that analogous observations can be found in the field of image classification, where earlier work [26] has shown that Bayesian variants of FCNNs achieve similar performances on benchmark datasets.

B. Calibration

Calibration curves for the RUL predictions of the BCNNs are shown in Fig. 8. The curves were constructed using all RUL predictions per valve over all 40 test valves. The ideal calibration curve is shown in the figure. For instance, if a credible interval of 60% is considered for every RUL output prediction, ideally 60% of the true RUL values should fall within this interval. Since it is not possible to construct these intervals for an FCNN, only BCNNs are evaluated in this figure. The first observation is that our proposed architecture using MFVI shows a much better calibration performance than the MC dropout architecture. The latter model underestimates predictive uncertainty by a large margin. In contrast, the calibration curves for the MFVI architecture lie close to the ideal line, with only a slight tendency to underestimate the uncertainty toward higher confidence levels.

For illustration purposes, the BCNN MFVI architecture with nontruncated Gaussians is also included in the figure. It can be observed that the use of nontruncated Gaussians as predictive outputs, which is a less appropriate modeling choice, results in underestimated uncertainty intervals. This is the case when physical features are either included or excluded.
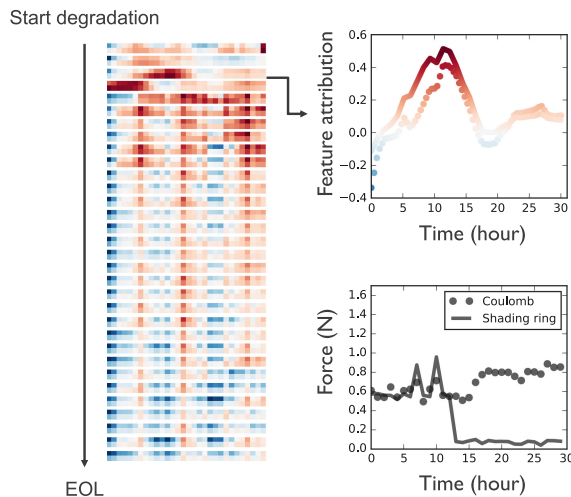
Fig. 9. Feature attributions for the physical features of valve 10. On the left: Attributions for every data acquisition are shown, from the start of degradation up to EOL. Red indicates a positive attribution, the opposite holds for blue. On the right: The feature attributions of the third data acquisition (now as graphs instead of a "top down" view), together with the corresponding physical features.

A regular finding with VI approaches is that the models are overconfident. Predictive means can be accurate, while variances are considerably underestimated. Formulating fully reliable and calibrated uncertainty estimates for BNNs using VI remains, however, unsolved [2]. With the presented BCNN MFVI architecture and modeling choices in this work, reliable credible intervals can nevertheless be obtained. Therefore, the resulting uncertainty estimates can be trusted, which is an important concern in a prognostic setting.

C. Interpretability

Occlusion is a perturbation-based approach to compute feature importance in the input images of a CNN [27]. It involves replacing contiguous rectangular regions of the input image with grey pixels, and computing the difference in predictive output. This technique has been used to highlight areas in solenoid valve current signals that characterize degradation in [8]. Feature attributions obtained by the occlusion method are shown in Fig. 9 for test valve 10 (the same valve as in Fig. 5) for the FCNN trained on the valve current signals together with the physical features. On the left of the figure, feature attributions are shown for the physical features from start of degradation until EOL. Every 2×30 rectangle corresponds to the physical features of one data acquisition, as indicated in red for the input data of the CNNs shown in Fig. 4. Red indicates features that contribute toward a larger RUL value, while blue indicates a lower value. On the right of the figure, feature attributions of the third data acquisition (now as graphs, instead of a "top down" view), together with the corresponding physical features are shown.

The first observation, is the shift in color on the left part of the figure. An evolution from predominantly positive contributions, to predominantly negative contributions toward EOL can be identified. This can be interpreted as if the physical features are

helping the network to distinguish between early and late RUL values. This evolution can also be seen for the other test valves in general. A second distinctive observation that applies to the test set in general, is the peak in positive feature attributions as shown on the right of the figure. This peak occurs at the moment when the shading ring force drops to zero, which only happens early in the degradation process. Thus, this event contributes toward a larger RUL prediction when observed by the CNN.

By considering these two observations, we gain insight into why the use of the physical features as training data, as described in Section V-A, results in improved RUL estimations.

VI. CONCLUSION

This article proposes a BCNN-based methodology for direct RUL prediction of solenoid valves using current signals as training data. Results from earlier work using DL and feature-based approaches were improved upon, reducing RUL MAE by ∼40%. Predictive performance was further improved by using salient physical features as the BCNN input (∼20% lower RUL MAE), making it a kind of hybrid model. Two explainable insights into the raised predictive performance were gained by using occlusion, a feature attribution method. It was shown that hybrid models outperform data-driven approaches when doing early RUL prediction. In contrast to earlier work on BCNN-based RUL predictions using MC dropout for training and inference, our proposed BCNN MFVI architecture showed well-calibrated predictive UQ. Its credible intervals for the RUL closely match the empirical predictive uncertainty. As a result, the predictive uncertainty allowed for reliable prognostic decision-making.

Future work will focus on an extensive comparative study of the calibration performance of Bayesian neural networks for prognostics. Further research into the fusion of physics-based domain knowledge and Bayesian DL is of interest as well.

REFERENCES

[1] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mech. Syst. Signal Process.*, vol. 104, pp. 799–834, 2018.

[2] E. Goan and C. Fookes, "Bayesian neural networks: An introduction and survey," in *Case Studies in Applied Bayesian Data Science: CIRM Jean-Morlet Chair*, Springer International Publishing, 2018, pp. 45–87.

[3] O. Fink, Q. Wang, M. Svensén, P. Dersin, W. Lee, and M. Ducoffe, "Potential, challenges and future directions for deep learning in prognostics and health management applications," *Eng. Appl. Artif. Intell.*, vol. 92, 2020, Art. no. 103678.

[4] W. Peng, Z. Ye, and N. Chen, "Bayesian deep-learning-based health prognostics toward prognostics uncertainty," *IEEE Trans. Ind. Electron.*, vol. 67, no. 3, pp. 2283–2293, Mar. 2020.

[5] G. Li, L. Yang, C. Lee, X. Wang, and M. Rong, "A Bayesian deep learning RUL framework integrating epistemic and aleatoric uncertainties," *IEEE Trans. Ind. Electron.*, to be published, doi: 10.1109/TIE.2020.3009593.

[6] Y. Gal and Z. Ghahramani, "Bayesian convolutional neural networks with bernoulli approximate variational inference," in *Proc. 4th Int. Conf. Learn. Representations Workshop Track*, 2016, pp. 1050–1059.

[7] N. J. Jameson, M. H. Azarian, and M. Pecht, "Fault diagnostic opportunities for solenoid operated valves using physics-of-failure analysis," in *Proc. Int. Conf. Prognostics Health Manage.*, 2014, pp. 1–6.

[8] G. Tod, T. Mazaev, K. Eryilmaz, A. P. Ompusunggu, E. Hostens, and S. V. Hoecke, "A convolutional neural network aided physical model improvement for AC solenoid valves diagnosis," in *Proc. Prognostics Syst. Health Manage. Conf.*, 2019, pp. 223–227.

[9] M. Ancona, E. Ceolini, C. Oztireli, and M. Gross, "Towards better understanding of gradient-based attribution methods for deep neural networks," in *Proc. Int. Conf. Learn. Representations*, 2017, pp. 1–16.

[10] G. Aydemir and K. Paynabar, "Image-based prognostics using deep learning approach," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 5956–5964, Sep. 2020.

[11] Y. Yuan *et al.*, "A general end-to-end diagnosis framework for manufacturing systems," *Nat. Sci. Rev.*, vol. 7, no. 2, pp. 418–429, Nov. 2020.

[12] B. Yang, R. Liu, and E. Zio, "Remaining useful life prediction based on a double-convolutional neural network architecture," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9521–9530, Dec. 2019.

[13] R. Liu, B. Yang, and A. G. Hauptmann, "Simultaneous bearing fault recognition and remaining useful life prediction using joint-loss convolutional neural network," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 87–96, Jan. 2020.

[14] H. Li, W. Zhao, Y. Zhang, and E. Zio, "Remaining useful life prediction using multi-scale deep convolutional neural network," *Appl. Soft Comput.*, vol. 89, 2020, Art. no. 106113.

[15] T. Kim and S. Sohn, "Multitask learning for health condition identification and remaining useful life prediction: Deep convolutional neural network approach," *J. Intell. Manuf.*, pp. 1–11, Jul. 2020.

[16] M. Ma and Z. Mao, "Deep convolution-based LSTM network for remaining useful life prediction," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 1658-1667, Mar. 2021.

[17] M. Kraus and S. Feuerriegel, "Forecasting remaining useful life: Interpretable deep learning approach via variational Bayesian inferences," *Decis. Support Syst.*, vol. 125, 2019, Art. no. 113100.

[18] B. Wang, Y. Lei, T. Yan, N. Li, and L. Guo, "Recurrent convolutional neural network: A new framework for remaining useful life prediction of machinery," *Neurocomputing*, vol. 379, pp. 117–129, 2020.

[19] S. Jo, B. Seo, H. Oh, B. D. Youn, and D. Lee, "Model-based fault detection method for coil burnout in solenoid valves subjected to dynamic thermal loading," *IEEE Access*, vol. 8, pp. 70387–70400, 2020.

[20] W. Guo, J. Cheng, Y. Tan, and Q. Liu, "Solenoid valve fault diagnosis based on genetic optimization MKSVM," in *Proc. IOP Conf. Series: Earth Environ. Sci.*, 2018, vol. 170, no. 4, Art. no. 042134.

[21] J. Liniger, S. Stubkier, M. Soltani, and H. C. Pedersen, "Early detection of coil failure in solenoid valves," *IEEE/ASME Trans. Mechatronics*, vol. 25, no. 2, pp. 683–693, Apr. 2020.

[22] A. Sarwar and X. Lu, "Fault isolation and diagnosis of high pressure fuel pump solenoid valves using current feedback," in *Proc. IEEE Int. Conf. Prognostics Health Manage.*, 2018, pp. 1–8.

[23] T. Mazaev, A. P. Ompusunggu, G. Tod, G. Crevecoeur, and S. V. Hoecke, "Data-driven prognostics of alternating current solenoid valves," in *Proc. Prognostics Health Manage. Conf.*, 2020, pp. 109–115.

[24] S. Singh, A. Kumar, and N. Kumar, "Motor current signature analysis for bearing fault detection in mechanical systems," *Procedia Materials Sci.*, vol. 6, pp. 171–177, 2014.

[25] Y. Wen, P. Vicol, J. Ba, D. Tran, and R. Grosse, "Flipout: Efficient pseudo-independent weight perturbations on mini-batches," in *Proc. Int. Conf. Learn. Representations*, 2018, pp. 1–16.

[26] K. S. F. Laumann and M. Liwicki, "Uncertainty estimations by softplus normalization in Bayesian convolutional neural networks with variational inference," vol. abs/1806.05978, 2020.

[27] M. Zeiler and R. Fergus, "Visualizing and understanding convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.*, 2013, pp. 818–833.

Tamir Mazaev received the M.Sc. degree in physics and astronomy from Ghent University, Ghent, Belgium, in 2016, and the M.Sc degree in artificial intelligence from KU Leuven, Leuven, Belgium, in 2017. He is currently working toward the Ph.D. degree in computer science engineering with IDLab, Ghent University-imec.

His research interests mainly concern the use of machine learning for applications in predictive maintenance.

Mr. Mazaev was the recipient of a personal grant strategic basic research (SB) from Research Foundation - Flanders (FWO), in 2020.

Guillaume Crevecoeur was born in 1981. He received the master's and Ph.D. degree in engineering physics from Ghent University, Ghent, Belgium, in 2004 and 2009, respectively.

He was appointed Associate Professor with the Department of Electromechanical, System and Metal Engineering, the Ghent University, in 2014. He currently leads the Ghent University activities on sensing, monitoring, control and decision-making within the Flanders Make. His research interests include intersection of system identification, control and machine learning for mechatronic, and industrial robotic systems. His goal is to endow physical dynamic systems with improved functionalities and capabilities when interacting with uncertain environments, other systems and humans.

Dr. Crevecoeur was the recipient of a Research Foundation Flanders postdoctoral fellowship from the Ghent University, in 2009.

Sofie Van Hoecke received the graduate degree and the Ph.D. degree in computer science engineering, from the Engineering Department and the Department of Information Technology, Ghent University, Ghent, Belgium, in 2003 and 2009, respectively.

After being a Postdoctoral Research Engineer, In 2009, she started as Lecturer ICT and ICT Research Coordinator with the University College West-Flanders, Kortrijk, Belgium. In 2013, she returned to the Ghent University, where, after finalizing her tenure track, she is currently an Associate Professor and leads the multidisciplinary research team PreDiCT within the IDLab research group of Ghent University-imec. Her research interests are funded through Flemish, national, and European industry-driven and fundamental projects. Her research interest includes multisensor architectures, condition monitoring, machine learning, semantic dashboards, and the incorporation of semantic technologies, user feedback and/or physical models into machine learning to design hybrid AI solutions, with applications in predictive maintenance, smart buildings and predictive healthcare.