A Review of Uncertainty Quantification in Deep Learning: Techniques, Applications and Challenges

Moloud Abdar*, Farhad Pourpanah, *Member, IEEE*, Sadiq Hussain, Dana Rezazadegan, Li Liu, *Senior Member, IEEE,* Mohammad Ghavamzadeh, Paul Fieguth, *Senior Member, IEEE,* Xiaochun Cao, *Senior Member, IEEE,* Abbas Khosravi, *Senior Member, IEEE,* U Rajendra Acharya, *Senior Member, IEEE,* Vladimir Makarenkov and Saeid Nahavandi, *Fellow, IEEE*

Abstract—Uncertainty quantification (UQ) plays a pivotal role in the reduction of uncertainties during both optimization and decision making, applied to solve a variety of real-world applications in science and engineering. Bayesian approximation and ensemble learning techniques are two of the most widely-used UQ methods in the literature. In this regard, researchers have proposed different UQ methods and examined their performance in a variety of applications such as computer vision (e.g., self-driving cars and object detection), image processing (e.g., image restoration), medical image analysis (e.g., medical image classification and segmentation), natural language processing (e.g., text classification, social media texts and recidivism risk-scoring), bioinformatics, etc. This study reviews recent advances in UQ methods used in deep learning, investigates the application of these methods in reinforcement learning, and highlight the fundamental research challenges and directions associated with the UQ field.

Index Terms—Artificial intelligence, Uncertainty quantification, Deep learning, Machine learning, Bayesian statistics, Ensemble learning, Reinforcement learning.

✦

1 INTRODUCTION

IN everyday scenarios, we deal with uncertainties in numerous fields, from invest opportunities and medical diagnosis to sporting games and weather forecast, with an objective to make decision based on collected observations and uncertain domain knowledge. Nowadays, we can rely on models developed using machine and deep learning

- *M. Abdar, A. Khosravi and S. Nahavandi are with the Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Australia (e-mails: m.abdar1987@gmail.com, mabdar@deakin.edu.au, abbas.khosravi@deakin.edu.au & saeid.nahavandi@deakin.edu.au).*
- *F. Pourpanah is with College of Mathematics and Statistics, Guangdong Key Lab. of Intelligent Information Processing, Shenzhen University, Shenzhen 518060, China (e-mail: farhad@szu.edu.cn).*
- *S. Hussain is with the System Administrator, Dibrugarh University, Dibrugarh, India (e-mail: sadiq@dibru.ac.in).*
- *D. Rezazadegan is with the Department of Computer Science and Software Engineering, Swinburne University of Technology, Melbourne, Australia (e-mail: drezazadegan@swin.edu.au).*
- *L. Liu is with the Center for Machine Vision and Signal Analysis, University of Oulu, Oulu, Finland (e-mail: li.liu@oulu.fi).*
- *M. Ghavamzadeh is with the Google research (e-mail: ghavamza@google.com).*
- *P. Fieguth is with the Department of Systems Design Engineering, University of Waterloo, Waterloo, Canada (e-mail: pfieguth@uwaterloo.ca).*
- *X. Cao is with the State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China (e-mail: caoxiaochun@iie.ac.cn).*
- *U. R. Acharya is with the Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Clementi, Singapore (e-mail: aru@np.edu.sg).*
- *V. Makarenkov is with the Department of Computer Science, University of Quebec in Montreal, Montreal (QC), Canada (e-mail: makarenkov.vladimir@uqam.ca).*
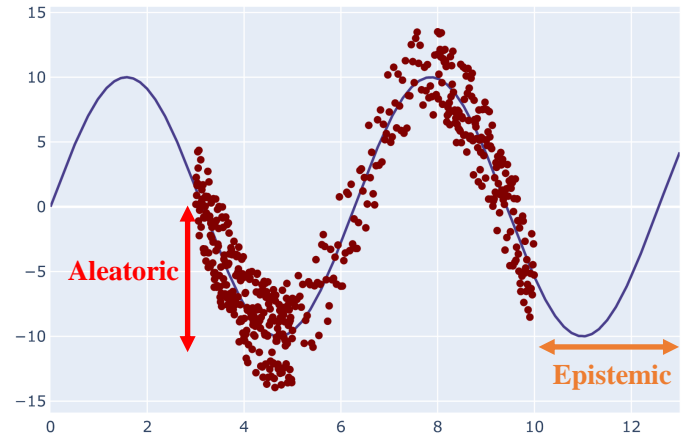- ** Corresponding author: Moloud Abdar, m.abdar1987@gmail.com*

Fig. 1: A schematic view of main differences between aleatoric and epistemic uncertainties.

techniques can quantify the uncertainties to accomplish statistical inference [1]. It is very important to evaluate the efficacy of artificial intelligence (AI) systems before its usage [2]. The predictions made by such models are uncertain as they are prone to noises and wrong model inference besides the inductive assumptions that are inherent in case of uncertainty. Thus, it is highly desirable to represent uncertainty in a trustworthy manner in any AI-based systems. Such automated systems should be able to perform accurately by handling uncertainty effectively. Principle of uncertainty plays an important role in AI settings such as concrete learning algorithms [3], and active learning (AL) [4], [5]. The sources of uncertainty occurs when the test and training data are mismatched and data uncertainty occurs because
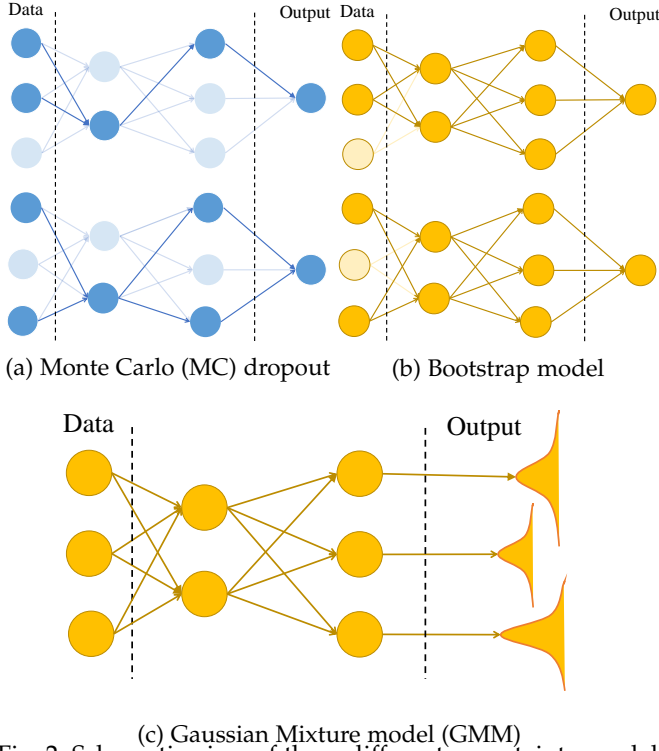
(a) Monte Carlo (MC) dropout   (b) Bootstrap model

(c) Gaussian Mixture model (GMM)

Fig. 2: Schematic view of three different uncertainty models with the related network architectures, reproduced based on [9].

(a) BNN   (b) OoD classifier

Fig. 3: A graphical representation of two different uncertainty-aware (UA) models, reproduced based on [14].

of class overlap or due to the presence of noise in the data [6]. Estimating knowledge uncertainty is more difficult compared to data uncertainty which naturally measures it as a result of maximum likelihood training. Sources of uncertainty in prediction are essential to tackle the uncertainty estimation problem [7]. There are two main sources of uncertainty, conceptually called aleatoric and epistemic uncertainties [8] (see Fig. 1).

Irreducible uncertainty in data giving rise to uncertainty in predictions is an aleatoric uncertainty (also known as data uncertainty). This type of uncertainty is not the property of the model, but rather is an inherent property of the data distribution; hence it is irreducible. Another type of uncertainty is epistemic uncertainty (also known as knowledge uncertainty) that occurs due to inadequate knowledge and data. One can define models to answer different human questions poised in model-based prediction. In the case of data-rich problem, there is a collection of massive data but it may be informatively poor [10]. In such cases, AI-based methods can be used to define the efficient models which characterize the emergent features from the data. Very often these data are incomplete, noisy, discordant and multimodal [1].

Uncertainty quantification (UQ) underpins many critical decisions today. Predictions made without UQ are usually not trustworthy and inaccurate. To understand the Deep Learning (DL) [11], [12] process life cycle, we need to comprehend the role of UQ in DL. The DL models start with the collection of most comprehensive and potentially relevant datasets available for decision making process. The DL scenarios are designed to meet some performance goals to select the most appropriate DL architecture after training the model using the labeled data. The iterative training
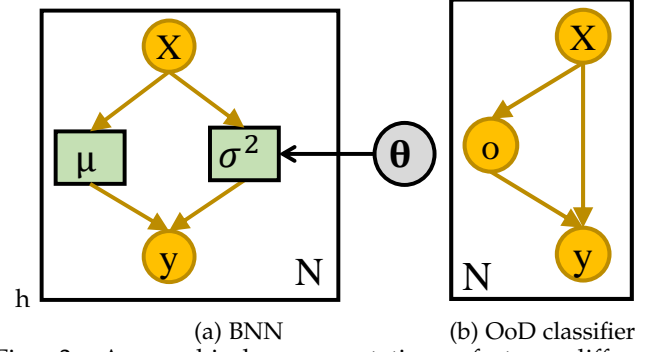
process optimizes different learning parameters, which will be 'tweaked' until the network provides a satisfactory level of performance.

There are several uncertainties that need to be quantified in the steps involved. The uncertainties that are obvious in these steps are the following: (i) selection and collection of training data, (ii) completeness and accuracy of training data, (iii) understanding the DL (or traditional machine learning) model with performance bounds and its limitations, and (iv) uncertainties corresponds to the performance of the model based on operational data [13]. Data driven approaches such as DL associated with UQ poses at least four overlapping groups of challenges: (i) absence of theory, (ii) absence of casual models, (iii) sensitivity to imperfect data, and (iv) computational expenses. To mitigate such challenges, ad hoc solutions like the study of model variability and sensitivity analysis are sometimes employed. Uncertainty estimation and quantification have been extensively studied in DL and traditional machine learning. In the following, we provide a brief summary of some recent studies that examined the effectiveness of various methods to deal with uncertainties.

A schematic comparison of the three different uncertainty models [9] (MC dropout, Boostrap model and GMM is provided in Fig. 2. In addition, two graphical representations of uncertainty-aware models (BNN) vs OoD classifier) is illustrated in Fig. 3.

1.1 Research Objectives and Outline

In the era of big data, ML and DL, intelligent use of different raw data has an enormous potential to benefit wide variety of areas. However, UQ in different ML and DL methods can significantly increase the reliability of their results. Ning et al. [15] summarized and classified the main contributions of the data-driven optimization paradigm under uncertainty. As can be observed, this paper reviewed the data-driven optimization only. In another study, Kabir et al. [16] reviewed Neural Network-based UQ. The authors focused on probabilistic forecasting and prediction intervals (PIs) as they are among most widely used techniques in the literature for UQ.

We have noticed that, from 2010 to 2020 (end of June), more than 2500 papers on UQ in AI have been published in various fields (e.g., computer vision, image processing, medical image analysis, signal processing, natural language processing, etc.). In one hand, we ignore large number of

papers due to lack of adequate connection with the subject of our review. On the other hand, although many papers that we have reviewed have been published in related conferences and journals, many papers have been found on open-access repository as electronic preprints (i.e. arXiv) that we reviewed them due to their high quality and full relevance to the subject. We have tried our level to best to cover most of the related articles in this review paper. It is worth mentioning that this review can, therefore, serve as a comprehensive guide to the readers in order to steer this fast-growing research field.

Unlike previous review papers in the field of UQ, this study reviewed most recent articles published in quantifying uncertainty in AI (ML and DL) using different approaches. In addition, we are keen to find how UQ can impact the real cases and solve uncertainty in AI can help to obtain reliable results. Meanwhile, finding important chats in existing methods is a great way to shed light on the path to the future research. In this regard, this review paper gives more inputs to future researchers who work on UQ in ML and DL. We investigated more recent studies in the domain of UQ applied in ML and DL methods. Therefore, we summarized few existing studies on UQ in ML and DL. It is worth mentioning that the main purpose of this study is not to compare the performance of different UQ methods proposed because these methods are introduced for different data and specific tasks. For this reason, we argue that comparing the performance of all methods is beyond the scope of this study. For this reason, this study mainly focuses on important areas including DL, ML and Reinforcement Learning (RL). Hence, the main contributions of this study are as follows:

- To the best of our knowledge, this is the first comprehensive review paper regarding UQ methods used in ML and DL methods which is worthwhile for researchers in this domain.
- A comprehensive review of newly proposed UQ methods is provided.
- Moreover, the main categories of important applications of UQ methods are also listed.
- The main research gaps of UQ methods are pointed out.
- Finally, few solid future directions are discussed.

2 PRELIMINARIES

In this section, we explained the structure of feed-forward neural network followed by Bayesian modeling to discuss the uncertainty in detail.

2.1 Feed-forward neural network

In this section, the structure of a single-hidden layer neural network [17] is explained, which can be extended to multiple layers. Suppose x is a D-dimensional input vector, we use a linear map W_1 and bias b to transform x into a row vector with Q elements, i.e., $W_1 x + b$. Next a non-linear transfer function $\sigma(.)$, such as rectified linear (ReLU), can be applied to obtain the output of the hidden layer. Then another linear function W_2 can be used to map hidden layer to the output:

$$\hat{y} = \sigma(xW_1 + b)W_2 \tag{1}$$

For classification, to compute the probability of X belonging to a label c in the set $\{1, ..., C\}$, the normalized score is obtained by passing the model output $\hat{y}$ through a softmax function $\hat{p}_d = \exp(\hat{y}_d)/(\sum_{d'} \exp(\hat{y}_{d'}))$. Then the softmax loss is used:

$$E^{W_1, W_2, b}(X, Y) = -\frac{1}{N} \sum_{i=1}^{N} \log(\hat{p}_{i, c_i}) \tag{2}$$

where $X = (x_1, ..., x_N)$ and $Y = (y_1, ..., y_N)$ are inputs and their corresponding outputs, respectively.

For regression, the Euclidean loss can be used:

$$E^{W_1, W_2, b}(X, Y) = \frac{1}{2N} \sum_{i=1}^{N} \|y_i - \hat{y}\|^2 \tag{3}$$

2.2 Uncertainty Modeling

As mentioned above, there are two main uncertainties: epistemic (model uncertainty) and aleatoric (data uncertainty) [18]. The aleatoric uncertainty has two types: *homoscedastic* and *heteroscedastic* [19]. The predictive uncertainty (PU) consists of two parts: *(i)* epistemic uncertainty (EU), and (ii) aleatoric uncertainty (AU), which can be written as sum of these two parts:

$$PU = EU + AU. \tag{4}$$

Epistemic uncertainties can be formulated as probability distribution over model parameters. Let $D_{tr} = \{X, Y\} = \{(x_i, y_i)\}_{i=1}^{N}$ denotes a training dataset with inputs $x_i \in \Re^D$ and their corresponding classes $y_i \in \{1, ..., C\}$, where C represents the number of classes. The aim is to optimize the parameters, i.e., ω, of a function $y = f^\omega(x)$ that can produce the desired output. To achieve this, the Bayesian approach defines a model likelihood, i.e., $p(y|x, \omega)$. For classification, the softmax likelihood can be used:

$$p(y = c|x, \omega) = \frac{\exp(f_c^\omega(x))}{\sum_{c'} \exp(f_{c'}^\omega(x))}. \tag{5}$$

and the Gaussian likelihood can be assumed for regression:

$$p(y|x, \omega) = \mathcal{N}(y; f^\omega(x), \tau^{-1}I), \tag{6}$$

where τ represents the model precision.

The posterior distribution, i.e., $p(\omega|x, y)$, for a given dataset D_{tr} over ω by applying Bayes' theorem can be written as follows:

$$p(\omega|X, Y) = \frac{p(Y|X, \omega)p(\omega))}{p(Y|X)}. \tag{7}$$

For a given test sample x^*, a class label with regard to the $p(\omega|X, Y)$ can be predicted:

$$p(y^*|x^*, X, Y) = \int p(y^*|x^*, \omega)p(\omega|X, Y)d\omega. \tag{8}$$

This process is called inference or marginalization. However, $p(\omega|X, Y)$ cannot be computed analytically, but it can be approximated by variational parameters, i.e., $q_\theta(\omega)$. The aim is to approximate a distribution that is close to

the posterior distribution obtained by the model. As such, the Kullback-Leibler (KL) [20] divergence is needed to be minimised with regard to θ. The level of similarity among two distributions can be measured as follows:

$$KL(q_\theta(\omega)\|p(\omega|X,Y)) = \int q_\theta(\omega) \log \frac{q_\theta(\omega)}{p(\omega|X,Y)} d\omega. \quad (9)$$

The predictive distribution can be approximated by minimizing KL divergence, as follows:

$$p(y^*|x^*,X,Y) \approx \int p(y^*|x^*,\omega)q_\theta^*(\omega)d\omega =: q_\theta^*(y^*,x^*), \quad (10)$$

where $q_\theta^*(\omega)$ indicates the optimized objective.

KL divergence minimization can also be rearranged into the **evidence lower bound** (ELBO) maximization [21]:

$$\mathcal{L}_{VI}(\theta) := \int q_\theta(\omega) \log p(Y|X,\omega)d\omega - KL(q_\theta(\omega)\|p(\omega)), \quad (11)$$

where $q_\theta(\omega)$ is able to describe the data well by maximizing the first term, and be as close as possible to the prior by minimizing the second term. This process is called variational inference (VI). Dropout VI is one of the most common approaches that has been widely used to approximate inference in complex models [22]. The minimization objective is as follows [23]:

$$\mathcal{L}(\theta,p) = -\frac{1}{N}\sum_{i=1}^{N} \log p(y_i|x_i,\omega) + \frac{1-p}{2N}\|\theta\|^2 \quad (12)$$

where N and P represent the number of samples and dropout probability, respectively.

To obtain data-dependent uncertainty, the precision τ in (6) can be formulated as a function of data. One approach to obtain epistemic uncertainty is to mix two functions: predictive mean, i.e., $f^\theta(x)$, and model precision, i.e., $g^\theta(x)$, and the likelihood function can be written as $y_i = \mathcal{N}(f^\theta(x), g^\theta(x)^{-1})$. A prior distribution is placed over the weights of the model, and then the amount of change in the weights for given data samples is computed. The Euclidian distance loss function (3) can be adapted as follows:

$$E^{W_1,W_2,b} :=$$
$$\frac{1}{2}(y - f^{W_1,W_2,b}(x))g^{W_1,W_2,b}(x)(y - f^{W_1,W_2,b}(x))^T -$$
$$\frac{1}{2}\log det\, g^{W_1,W_2,b} + \frac{D}{2}\log 2\pi$$
$$= -\log \mathcal{N}(f^\theta(x), g^\theta(x)^{-1}) \quad (13)$$

The predictive variance can be obtained as follows:

$$\widehat{Var}[x^*] := \frac{1}{T}\sum_{t=1}^{T} g^{\widetilde{\omega_t}}(x)\mathbf{I} + f^{\widetilde{\omega_t}}(x^*)^T f^{\widetilde{\omega_t}}(x^*)$$
$$- \widetilde{\mathbb{E}}[y^*]^T \widetilde{\mathbb{E}}[y^*] \xrightarrow[T\to\infty]{} Var_{q_\theta^*(y^*|x^*)}[y^*] \quad (14)$$

3 UNCERTAINTY QUANTIFICATION USING BAYESIAN TECHNIQUES

3.1 Bayesian Deep Learning/Bayesian Neural Networks

Despite the success of standard DL methods in solving various real-word problems, they cannot provide information about the reliability of their predictions. To alleviate this issue, Bayesian Deep Learning (BDL)/Bayesian Neural Networks (BNNs) [24], [25], [26], [27], [28], [29], [30], [31] can be used to interpret the model parameters. BNNs/BDL are robust to over-fitting problem and can be trained on both small and big datasets [32].

3.2 Monte Carlo (MC) dropout

As stated earlier, it is difficult to compute the exact posterior inference, but it can be approximated. In this regard, Monte Carlo (MC) [33] is an effective method. Nonetheless, it is a slow and computationally expensive method when integrated into a deep architecture. To combat this, MC (MC) dropout has been introduced, which uses dropout [34] as a regularization term to compute the prediction uncertainty [35]. Dropout is an effective technique that has been widely used to solve over-fitting problem in DNNs. During the training process, dropout randomly drops some units of NN to avoid them from co-tuning too much. Assume a NN with L layers, which W_l, b_l and K_l denote the weight matrices, bias vectors and dimensions of the lth layer, respectively. The output of NN and target class of the ith input x_i ($i = 1,...,N$) are indicated by $\hat{y}_i$ and y_i, respectively. The objective function using L_2 regularization can be written as:

$$\mathcal{L}_{dropout} := \frac{1}{N}\sum_{i=1}^{N} E(y_i,\hat{y}_i) + \lambda \sum_{l=1}^{L}(\|W_i\|_2^2 + \|b_i\|_2^2) \quad (15)$$

Dropout samples binary variables for each input data and every network unit in each layer (except the output layer), with the probability p_i for ith layer, if its value is 0, the unit i is dropped for a given input data. Same values are used in the backward pass to update parameters. Fig. 4 shows several visualization of variational distributions on a simple NN [36].

Several studies used MC dropout [37] to estimate UQ. Wang et al. [38] analyzed epistemic and aleatoric uncertainties for deep CNN-based medical image segmentation problems at both pixel and structure levels. They augmented the input image during test phase to estimate the transformation uncertainty. Specifically, the MC sampling was used to estimate the distribution of the output segmentation. Liu et al. [39] proposed a unified model using SGD to approximate both epistemic and aleatoric uncertainties of CNNs in presence of universal adversarial perturbations. The epistemic uncertainty was estimated by applying MC dropout with Bernoulli distribution at the output of neurons. In addition, they introduced the texture bias to better approximate the aleatoric uncertainty. Nasir et al. [40] conducted MC dropout to estimate four types of uncertainties, including variance of MC samples, predictive entropy, and Mutual Information (MI), in a 3D CNN to segment lesion from MRI sequences.

(a) Baseline neu- (b) Bernoulli (c) Gaussian
ral network DropConnect DropConnect

(d) Bernoulli (e) Gaussian (f) Spike-and-Slab
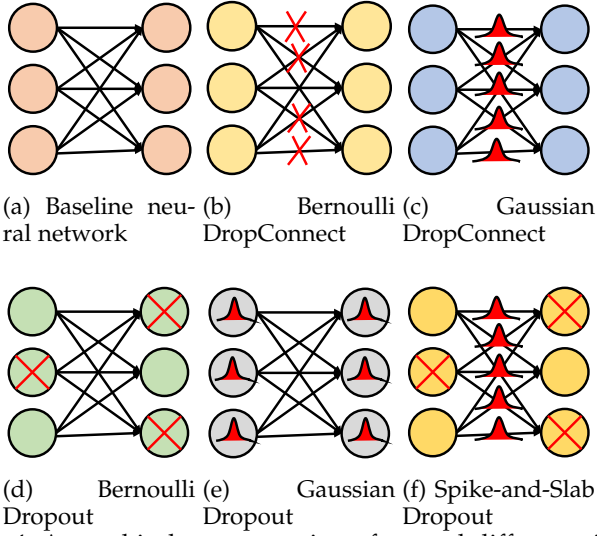Dropout Dropout Dropout

Fig. 4: A graphical representation of several different visualization of variational distributions on a simple NN which is reproduced based on [36].

In [41], two dropout methods, i.e. element-wise Bernoulli dropout [34] and spatial Bernoulli dropout [42] are implemented to compute the model uncertainty in BNNs for the end-to-end autonomous vehicle control. McClure and Kriegeskorte [36] expressed that sampling of weights using Bernoulli or Gaussian can lead to have a more accurate depiction of uncertainty in comparison to sampling of units. However, according to the outcomes obtained in [36], it can be argued that using either Bernoulli or Gaussian dropout can improve the classification accuracy of CNN. Based on these findings, they proposed a novel model (called spike-and-slab sampling) by combining Bernoulli or Gaussian dropout.

Do et al. [43] modified U-Net [44], which is a CNN-based deep model, to segment myocardial arterial spin labeling and estimate uncertainty. Specifically, batch normalization and dropout are added after each convolutional layer and resolution scale, respectively. Later, Teye et al. [45] proposed MC batch normalization (MCBN) that can be used to estimate uncertainty of networks with batch normalization. They showed that batch normalization can be considered as an approximate Bayesian model. Yu et al. [46] proposed a semi-supervised model to segment left atrium from 3D MR images. It consists of two modules including teacher and student, and used them in UA framework called UA self-ensembling mean teacher (UA-MT) model (see Fig. 5). As such, the student model learns from teacher model via minimizing the segmentation and consistency losses of the labeled samples and targets of the teacher model, respectively. In addition, UA framework based on MC dropout was designed to help student model to learn a better model by using uncertainty information obtained from teacher model. Table 1 lists studies that directly applied MC dropout to approximate uncertainty along with their applications.

3.2.1 Comparison of MC dropout with other UQ methods

Recently, several studies have been conducted to compare different UQ methods. For example, Foong et al. [65] empirically and theoretically studied the MC dropout and
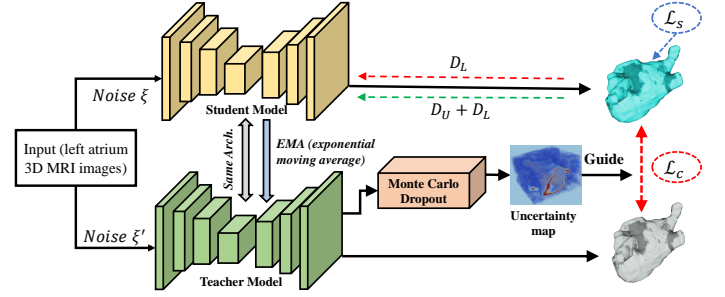

Fig. 5: A general view demonstrating the semi-supervised UA-MT framework applied to LA segmentation which is reproduced based on [46].
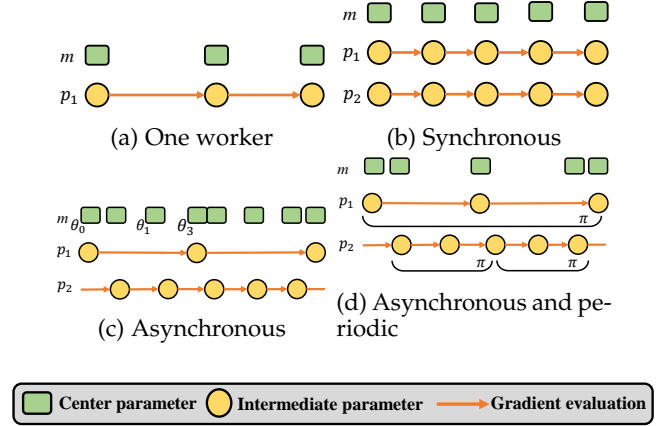

Fig. 6: A graphical implementations of different SG-MCMC models which is reproduced based on [47].

mean-field Gaussian VI. They found that both models can express uncertainty well in shallow BNNs. However, mean-field Gaussian VI could not approximate posterior well to estimate uncertainty for deep BNNs. Ng et al. [66] compared MC dropout with BBB using U-Net [44] as a base classifier. Siddhant et al. [67] empirically studied various DAL models for NLP. During prediction, they applied dropout to CNNs and RNNs to estimate the uncertainty. Hubschneider et al. [9] compared MC dropout with bootstrap ensembling-based method and a Gaussian mixture for the task of vehicle control. In addition, Mukhoti [68] applied MC dropout with several models to estimate uncertainty in regression problems. Kennamer et al. [69] empirically studied MC dropout in Astronomical Observing Conditions.

3.3 Markov chain Monte Carlo (MCMC)

Markov chain Monte Carlo (MCMC) [70] is another effective method that has been used to approximate inference. It starts by taking random draw z_0 from distribution $q(z_0)$ or $q(z_0|x)$. Then, it applies a stochastic transition to z_0, as follows:

$$Z_t \sim q(z_t|z_{t-1}, x). \tag{16}$$

This transition operator is chosen and repeated for T times, and the outcome, which is a random variable, converges in distribution to the exact posterior. Salakhutdinov et al. [71] used MCMC to approximate the predictive distribution rating values of the movies. Despite the success of the conventional MCMC, the sufficant number of iteration is unknown. In addition, MCMC requires long time to converge

TABLE 1: A summary of studies that applied the original MC dropout to approximate uncertainty along with their applications (Sorted by year).

Study	Year	Method	Application	Code
Kendal et al. [48]	2015	SegNet [49]	semantic segmentation	√
Leibig et al. [50]	2017	CNN	diabetic retinopathy	√
Choi et al. [51]	2017	mixture density network (MDN) [52]	regression	×
Jung et al. [53]	2018	full-resolution ResNet [54]	brain tumor segmentation	×
Wickstrom et al. [55]	2018	FCN [56] and SehNet [49]	polyps segmentation	×
Jungo et al. [57]	2018	FCN	brain tumor segmentation	×
Vandal et al. [58]	2018	Variational LSTM	predict flight delays	×
Devries and Taylor [59]	2018	CNN	medical image segmentation	×
Tousignant et al. [60]	2019	CNN	MRI images	×
Norouzi et al. [61]	2019	FCN	MRI images segmentation	×
Roy et al. [62]	2019	Bayesian FCNN	brain images (MRI) segmentation	√
Filos et al. [63]	2019	CNN	diabetic retinopathy	√
Harper and Southern [64]	2020	RNN and CNN	emotion prediction	×

to a desired distribution [33]. Several studies have been conducted to overcome these shortcomings. For example, Salimans et al. [72] expanded space into a set of *auxiliary random variables* and interpreted the stochastic Markov chain as a variational approximation.

The stochastic gradient MCMC (SG-MCMC) [73], [74] was proposed to train DNNs. It only needs to estimate the gradient on small sets of mini-batches. In addition, SG-MCMC can be converged to the true posterior by decreasing the step sizes [75], [76]. Gong et al. [77] combined amortized inference with SG-MCMC to increase the generalization ability of the model. Li et al. [47] proposed an accelerating SG-MCMC to improve the speed of the conventional SG-MCMC (see Fig. 6 for implementation of different SG-MCMC models). However, in short time, SG-MCMC suffers from a bounded estimation error [78] and it loses surface when applied to the multi-layer networks [79]. In this regard, Zhang et al. [80] developed a cyclical SG-MCMC (cSG-MCMC) to compute the posterior over the weights of neural networks. Specifically, a cyclical stepsize was used instead of the decreasing one. Large stepsize allows the sampler to take large moves, while small stepsize attempts the sampler to explore local mode.

Although SG-MCMC reduces the computational complexity by using a smaller subset, i.e. mini-batch, of dataset at each iteration to update the model parameters, those small subsets of data add noise into the model, and consequently increase the uncertainty of the system. To alleviate this, Luo et al. [81] introduced a sampling method called *the thermostat-assisted continuously tempered Hamiltonian Monte Carlo*, which is an extended version of the conventional Hamiltonian MC (HMC) [82]. Note that HMC is a MCMC method [83]. Specifically, they used Nosé-Hoover thermostats [84], [85] to handle the noise generated by mini-batch datasets. Later, dropout HMC (D-HMC) [83] was proposed for uncertainty estimation, and compared with SG-MCMC [73] and SGLD [86].

Besides, MCMC was integrated into the generative based methods to approximate posterior. For example, in [87], MCMC was applied to the stochastic object models, which

is learned by generative adversarial networks (GANs), to approximate the ideal observer. In [88], a visual tracking system based on a variational autoencoder (VAE) MCMC (VAE-MCMC) was proposed .
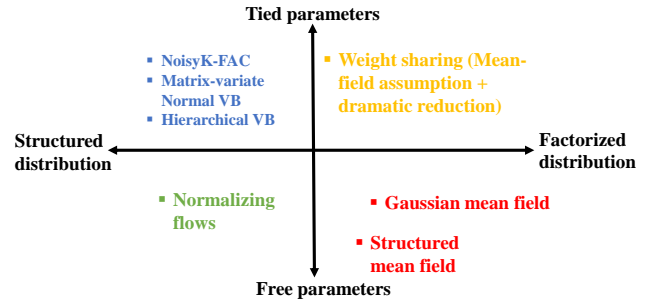
3.4 Variational Inference (VI)

Fig. 7: A summary of various VI methods for BDL which is reproduced based on [89]. Note that $Weight\ sharing\ (Mean-field\ assumption\ +\ dramatic\ reduction)$ is added based on the proposed method in [89].

The variational inference (VI) is an approximation method that learns the posterior distribution over BNN weights. VI-based methods consider the Bayesian inference problem as an optimization problem which is used by the SGD to train DNNs. Fig. 7 summaries various VI methods for BNN [89].

For BNNs, VI-based methods aim to approximate posterior distributions over the weights of NN. To achieve this, the loss can be defined as follows:

$$\mathcal{L}(\Phi) \approx \frac{1}{2|\mathcal{D}|}\sum_{i=1}^{|\mathcal{D}|}\mathcal{L}_R(y^{(i)}, x^{(i)}) + \frac{1}{|\mathcal{D}|}KL(q_\phi(w)\|p(w)) \quad (17)$$

where $|\mathcal{D}|$ indicates the number of samples, and

$$\mathcal{L}_R(y, x) = -\log(\hat{\tau}_x)^T 1 + \|\sqrt{\hat{\tau}_x} \odot (y - \hat{\mu}_x)\|^2 \quad (18)$$

$$\hat{\mu}_x = \hat{\mu}(x, w_\mu); \qquad w \sim q_\phi(w) \quad (19)$$

$$\hat{\tau}_x = \hat{\tau}(x, w_r). \tag{20}$$

where $\odot$ and 1 represent the element-wise product and vector filled with ones, respectively. Eq. (17) can be used to compute (10).

Posch et al. [90] defined the variational distribution using a product of Gaussian distributions along with diagonal covariance matrices. For each network layer, a posterior uncertainty of the network parameter was represented. Later, in [91], they replaced the diagonal covariance matrices with the traditional ones to allow the network parameters to correlate with each other. Inspired from transfer learning and empirical Bayes (EB) [92], MOPED [93] used a deterministic weights, which was derived from a pretrained DNNs with same architecture, to select meaningful prior distributions over the weight space. Later, in [94], they integrated an approach based on parametric EB into MOPED for mean field VI in Bayesian DNNs, and used fully factorized Gaussian distribution to model the weights. In addition, they used a real-world case study, i.e., diabetic retinopathy diagnosis, to evaluate their method. Subedar et al. [95] proposed an uncertainty aware framework based on multi-modal Bayesian fusion for activity recognition. They scaled BDNN into deeper structure by combining deterministic and variational layers. Marino et al. [96] proposed a stochastic modeling based approach to model uncertainty. Specifically, the DBNN was used to learn the stochastic learning of the system. Variational BNN [97], which is a generative-based model, was proposed to predict the superconducting transition temperature. Specifically, the VI was adapted to compute the distribution in the latent space for the model.

Louizos and Welling [98] adopted a stochastic gradient VI [99] to compute the posterior distributions over the weights of NNs. Hubin and Storvik [100] proposed a stochastic VI method that jointly considers both model and parameter uncertainties in BNNs, and introduced a latent binary variables to include/exclude certain weights of the model. Liu et al. [101] integrated the VI into a spatial–temporal NN to approximate the posterior parameter distribution of the network and estimate the probability of the prediction. Ryu et al. [102] integrated the graph convolutional network (GCN) into the Bayesian framework to learn representations and predict the molecular properties. Swiatkowski et al. [89] empirically studied the Gaussian mean-field VI. They decomposed the variational parameters into a low-rank factorization to make a more compact approximation, and improve the SNR ratio of the SG in estimating the lower bound of the variational. Franquhar et al. [103] used the mean-field VI to better train deep models. They argued that a deeper linear mean-field network can provide an analogous distribution of function space like shallowly full-co-variance networks. A schematic view of the proposed approach is demonstrated in Fig. 8.

3.5 Bayesian Active Learning (BAL)

Active learning (AL) methods aim to learn from unlabeled samples by querying an oracle [104]. Defining the right acquisition function, i.e., the condition on which sample is most informative for the model, is the main challenge of
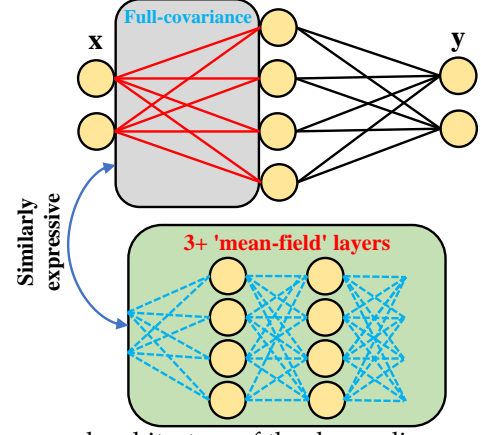

Fig. 8: A general architecture of the deeper linear mean-field network with three mean-field weight layers or more which is reproduced based on [103].

AL-based methods. Although existing AL frameworks have shown promising results in variety of tasks, they lack of scalability to high-dimensional data [105]. In this regard, the Baysian approaches can be integrated into DL structure to represent the uncertainty, and then combine with deep AL acquisition function to probe for the uncertain samples in the oracle.

DBAL [106], i.e., deep Bayesian AL, combine an AL framework with Bayesian DL to deal with high-dimensional data problems, i.e., image data. DBAL used batch acquisition to select the top n samples with the highest Bayesian AL by disagreement (BALD) [107] score. Model priors from empirical bayes (MOPED) [108] used BALD to evaluate the uncertainty. In addition, MC dropout was applied to estimate the model uncertainty. Later, Krisch et al. [109] proposed BatchBALD, which uses greedy algorithm to select a batch in linear time and reduce the run time. They modeled the uncertainty by leveraging the Bayesian AL (BAL) using Dropout-sampling. In [110], two types of uncertainty measures namely entropy and BALD [107], were compared.

ActiveHARNet [111], which is an AL-based framework for human action recognition, modeled the uncertainty by linking BNNs with GP using dropout. To achieve this, dropout was applied before each fully connected layer to estimate the mean and variance of BNN. DeepBASS [112], i.e., a deep AL semi-supervised learning, is an expectation-maximization [113] -based technique paired with an AL component. It applied MC dropout to estimate the uncertainty.

Scandalea et al. [114] proposed a framework based on U-Net structure for deep AL to segment biomedical images, and used uncertainty measure obtained by MC dropout, to suggest the sample to be annotated. Specifically, the uncertainty was defined based on the posterior probabilities' SD of the MC-samples. Zheng et al. [115] varied the number of Bayesian layers and their positions to estimate uncertainty through AL on MNIST dataset. The outcome indicated that few Bayesian layers near the output layer are enough to fully estimate the uncertainty of the model.

Inspired from [116], the Bayesian batch AL [117], which selects a batch of samples at each AL iteration to perform posterior inference over the model parameters, was proposed for large-scale problems. Active user training [118],
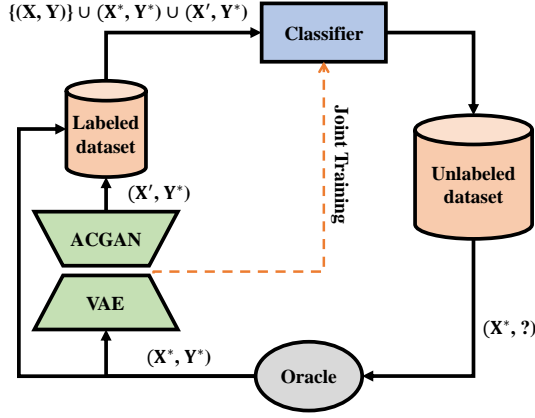
Fig. 9: Bayesian generative active deep learning (ADL). Note, ACGAN stands for the Auxiliary-classifier GAN which is reproduced based on [121].

which is a BAL-based crowdsourcing model, was proposed to tackle high-dimensional and complex classification problems. In addition, the Bayesian inference proposed in [119] was used to consider the uncertainty of the confusion matrix of the annotators.

Several generative-based AL frameworks have been introduced. In [120], a semi-supervised Bayesian AL model, which is a deep generative-based model that uses BNNs to give discriminative component, was developed. Tran et al. [121] proposed a Bayesian-based generative deep AL (BGADL) (Fig. 9) for image classification problems. They, firstly used the concept of DBAL to select the must informative samples and then VAE-ACGAN was applied to generate new samples based on the selected ones. Akbari et al. [122] proposed a unified BDL framework to quantify both aleatoric and epistemic uncertainties for activity recognition. They used an unsupervised DL model to extract features from the time series, and then their posterior distributions was learned through a VAE model. Finally, the Dropout [35] was applied after each dense layer and test phase for randomness of the model weights and sample from the approximate posterior, respectively.

3.6 Bayes by Backprop (BBB)

The learning process of a probability distribution using the weights of neural networks plays significant role for having better predictions results. Blundell et al. [123] proposed a novel yet efficient algorithm named Bayes by Backprop (BBB) to quantify uncertainty of these weights. The proposed BBB minimizes the compression cost which is known as the variational free energy (VFE) or the lower bound (expected) of the marginal likelihood. To do so, they defined a cost function as follows:

$$F(\mathcal{D}, \theta) = \mathrm{KL}[q(\mathbf{w}|\theta)\|P(\mathbf{w})] - \mathbb{E}_{q(\mathbf{w},\theta)}[\log P(\mathcal{D}|\mathbf{w})]. \quad (21)$$

The BBB algorithm uses unbiased gradient estimates of the cost function in 21 for learning distribution over the weights of neural networks. In another research, Fortunato et al. [124] proposed a new Bayesian recurrent neural network (BRNNs) using BBB algorithm. In order to improve the BBB algorithm, they used a simple adaptation of truncated back-propagation throughout the time. The proposed Bayesian RNN (BRNN) model is shown in Fig. 10.
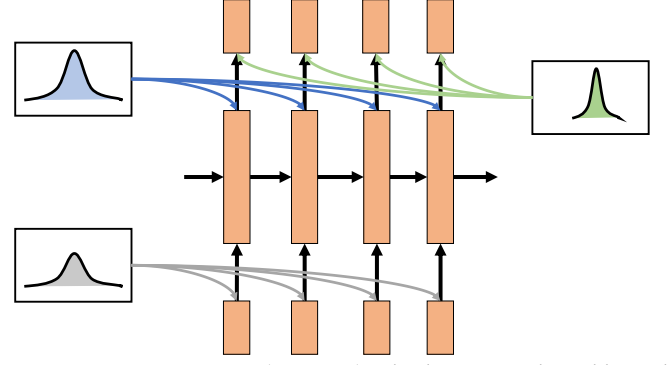

Fig. 10: Bayesian RNNs (BRNNs) which is reproduced based on the proposed model by Fortunato et al. [124].

Ebrahimi et al. [125] proposed an uncertainty-guided continual approach with BNNs (named UCB which stands for Uncertainty-guided continual learning (CL) technique with BNNs). The CL leads to learn a variety of new tasks while impound the aforetime knowledge obtained learned ones. The proposed UCB exploits the predicted uncertainty of the posterior distribution in order to formulate the modification in "important" parameters both in setting a hard-threshold as well as in a soft way. Recognition of different actions in videos needs not only big data but also is a time consuming process. To deal with this issue, de la Riva and Mettes [126] proposed a Bayesian-based deep learning method (named Bayesian 3D ConvNet) to analyze a small number of videos. In this regard, BBB was extended to be used by 3D CNNs and then employed to deal with uncertainty over the convolution weights in the proposed model. To do so, Gaussian distribution was applied to approximate the correct posterior in the proposed 3D Convolution layers using mean and STD (standard deviation) as follows:

$$\begin{cases} \theta = (\mu, \alpha), \\ \sigma^2 = \alpha.\mu^2, \\ q_\theta(\mathbf{w}_{ijhwt}|\mathcal{D}) = \mathcal{N}(\mu_{ijhwt}, \alpha_{ijhwt}\mu^2_{ijhwt}), \end{cases} \quad (22)$$

where i represents the input, j is the output, h is the filter height, w is the filter width and t is the time dimension. In another research, Ng et al. [66] compared the performance of two well-known uncertainty methods (MC dropout and BBB) in medical image segmentation (cardiac MRI) on a U-Net model. The obtained results showed that MC dropout and BBB demonstrated almost similar performances in medical image segmentation task.

3.7 Variational Autoencoders

An autoencoder is a variant of DL that consists of two components: (i) encoder, and (ii) decoder. Encoder aims to map high-dimensional input sample x to a low-dimensional latent variable z. While decoder reproduces the original sample x using latent variable z. The latent variables are compelled to conform a given prior distribution $P(z)$. Variational Autoencoders (VAEs) [99] are effective methods to model the posterior. They cast learning representations for high-dimensional distributions as a VI problem [127]. A

probabilistic model $P_\theta(x)$ of sample x in data space with a latent variable z in latent space can be written as follows:

$$p_\theta(x) = \int_z p_\theta(x|z)p(z), \tag{23}$$

The VI can be used to model the evidence lower bound $\log p_\theta(x)$ as follows:

$$\log p_\theta(x) = \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] - D_{KL}(q_\phi(z|x)\|p(x)), \tag{24}$$

where $q_\phi(z|x)$ and $p_\theta(x|z)$ are the encoder and decoder models, respectively, and ϕ and θ indicate their parameters.

Zamani et al. [128] developed a discrete VAE framework with Bernoulli latent variables as binary hashing code (Fig.11). The stochastic gradient was exploited to learn the model. They proposed a pairwise supervised hashing (PSH) framework to derive better hashing codes. PSH maximizes the ELBO with weighted KL regularization to learn more informative binary codes, and adapts a pairwise loss function to reward within-class similarity and between-class dissimilarity to minimize the distance among the hashing codes of samples from same class and vice versa.
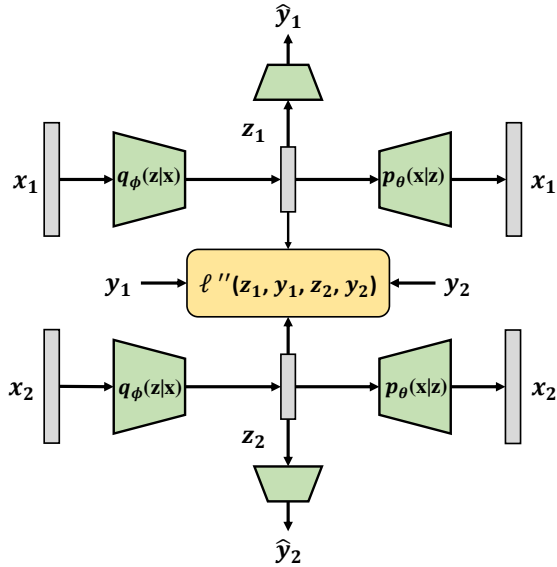

Fig. 11: Pairwise Supervised Hashing-Bernoulli VAE (PSH-BVAE) which is reproduced based on [128].

Bohm et al. [129] studied UQ for linear inverse problems using VAEs. Specifically, the vanilla VAE with mean-field Gaussian posterior was trained on uncorrupted samples under the ELBO. In addition, the EL_2O method [130] was adopted to approximate the posterior. Edupuganti et al. [131] studied the UQ tasks in magnetic resonance image recovery (see Fig. 12). As such, a VAE-GAN, which is a probabilistic recovery scheme, was developed to map the low quality images to high-quality ones. The VAE-GAN consists of VAE and multi-layer CNN as generator and discriminator, respectively. In addition, the Stein's unbiased risk estimator (SURE) was leveraged as a proxy to predict error and estimate the uncertainty of the model.

In [132], a framework based on variational U-Net [133] architecture was proposed for UQ tasks in reservoir simulations. Both simple U-Net and variational U-Net (VUNet)

are illustrated in Fig. 13. Cosmo VAE [134], which is a DL, i.e., U-Net, based VAE, was proposed to restore the missing observations of the cosmic microwave background (CMB) map. As such, the variational Bayes approximation was used to determine the ELBO of likelihood of the reconstructed image. Mehrasa et al. [135] proposed action point process VAE (APP VAE) for action sequences. APP VAE consists of two LSTM to estimate the prior and posterior distributions. Sato et al. [136] proposed a VAE-based UA for anomaly detection. They used MC sampling to estimate posterior.

Since VAEs are not stochastic processes, they are limited to encode finite-dimensional priors. To alleviate this limitation, Mishra et al. [137] developed the prior encoding VAE, i.e., πVAE. Inspired by the Gaussian process [138], πVAE is a stochastic process that learns the distribution over functions. To achieve this, πVAE encoder, firstly, transforms the locations to a high-dimensional space, and then, uses a linear mapping to link the feature space to outputs. While πVAE encoder aims to recreate linear mapping from the lower dimensional probabilistic embedding. Finally, the recreated mapping is used to get the reconstruction of the outputs. Guo et al. [139] used VAE to deal with data uncertainty under a just-in-time learning framework. The Gaussian distribution was employed to describe latent space features as variable-wise, and then the KL-divergence was used to ensure that the selected samples are the most relevant to a new sample. Daxberger et al. [140] tried to detect OoD samples during test phase. As such, the developed an unsupervised, probabilistic framework based on a Bayesian VAE. Besides, they estimated the posterior over the decoder parameters by applying SG-MCMC.

4 OTHER METHODS

In this section, we discuss few other proposed UQ methods used in machine and deep learning algorithms.

4.1 Deep Gaussian Processes

Deep Gaussian processes (DGPs) [141], [142], [143], [144], [145], [146], [147] are effective multi-layer decision making models that can accurately model the uncertainty. They represent a multi-layer hierarchy to Gaussian processes (GPs) [148], [149]. GPs is a non-parametric type of Bayesian model that encodes the similarity between samples using kernel function. It represents distributions over the latent variables with respect to the input samples as a Gaussian distribution $f_x \sim \mathcal{GP}(m(x), k(x, x'))$. Then, the output y is distributed based on a likelihood function $y|f_x \sim h(f_x)$. However, the conventional GPs can not effectively scale the large datasets. To alleviate this issue, inducing samples can be used. As such, the following variational lower bound can be optimized.

$$\log p(Y) \geq \sum_{y,x \in Y,X} \mathbb{E}_{q(f_x)}[\log p(y|f_x)] - \mathrm{KL}(q(f_Z)\|p(f_Z)), \tag{25}$$

where Z and $q(f_x)$ are the location of the inducing samples and the approximated variational to the distribution of f_x, respectively.
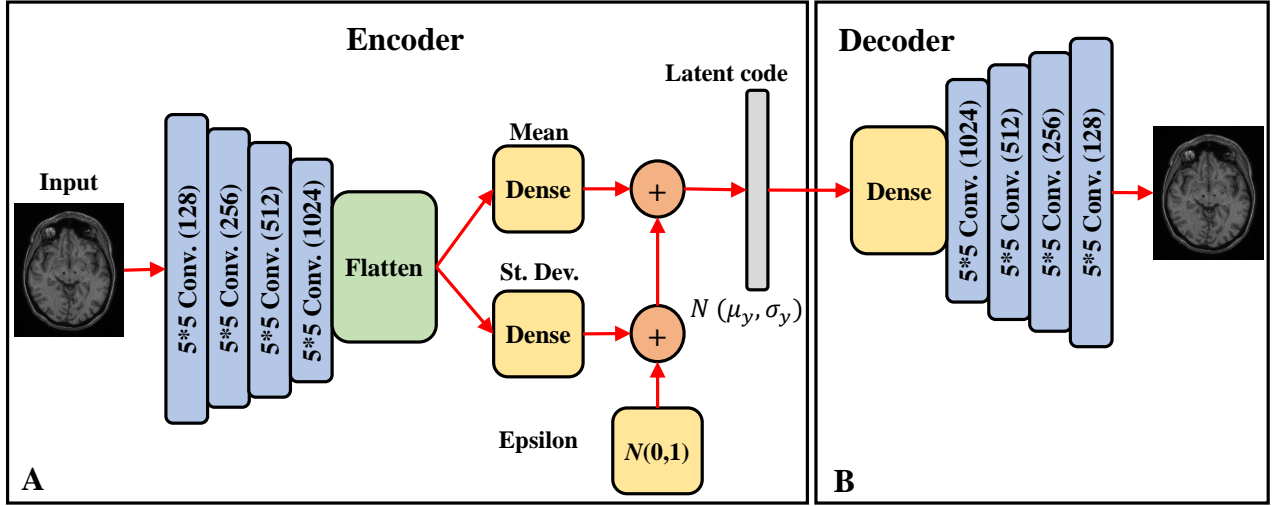
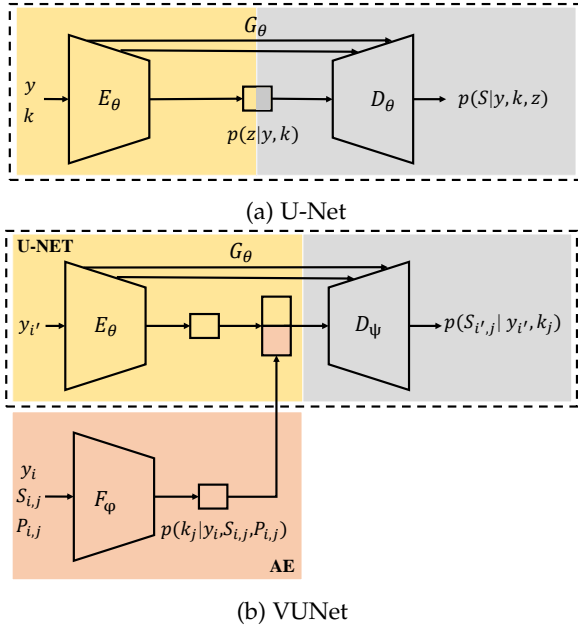Fig. 12: A schematic view of the proposed VAE model by Edupuganti et al. which is reproduced based on [131].

(a) U-Net

(b) VUNet

Fig. 13: A general view of U-Net and VUNet which are reproduced based on [132].

Fig. 14: A general Gaussian-based DNN model proposed by Bradshaw et al. [157] which is reproduced based on the same reference.

Oh et al. [150] proposed the hedged instance embedding (HIB), which hedges the position of each sample in the embedding space, to model the uncertainty when the input sample is ambiguous. As such, the probability of two samples matching was extended to stochastic embedding, and the MC sampling was used to approximate it. Specifically, the mixture of C Gaussians was used to represent the uncertainty. Havasi et al. [151] applied SGHMC into DGPs to approximate the posterior distribution. They introduced a moving window MC expectation maximization to obtain the maximum likelihood to deal with the problem of optimizing large number of parameters in DGPs. Maddox et al. [152] used stochastic weight averaging (SWA) [153] to build a Gaussian-baed model to approximate the true posterior. Later, they proposed SWA-G [154], which is SWA-Gaussian, to model Bayesian averaging and estimate uncertainty.

Most of the weight perturbation-based algorithms suffer from high variance of gradient estimation due to sharing same perturbation by all samples in a mini-batch. To al-

leviate this problem, flipout [155] was proposed. Filipout samples the pseudo-independent weight perturbations for each input to decorrelate the gradient within the mini-batch. It is able to reduce variance and computational time in training NNs with multiplicative Gaussian perturbations.

Despite the success of DNNs in dealing with complex and high-dimensional image data, they are not robust to adversarial examples [156]. Bradshaw et al. [157] proposed a hybrid model of GP and DNNs (GPDNNs) to deal with uncertainty caused by adversarial examples (see Fig. 14).

Choi et al. [158] proposed a Gaussian-based model to predict the localization uncertainty in YOLOv3 [159]. As such, they applied a single Gaussian model to the bbox coordinates of the detection layer. Specifically, the coordinates of each bbox is modeled as the mean (μ) and variance ($\sum$) to predict the uncertainty of bbox.

Khan et al. [160] proposed a natural gradient-based algorithm for Gaussian mean-field VI. The Gaussian distribution with diagonal covariances was used to estimate the probability. The proposed algorithm was implemented within the Adam optimizer. To achieve this, the network weights were perturbed during the gradient evaluation. In addition, they used a vector to adapt the learning rate to estimate uncertainty. Sun et al. [161] considered structural information of the model weights. They used the matrix variate Gaussian (MVG) [162] distribution to model structured correlations within the weights of DNNs, and introduced a reparametrization for the MVG posterior to make the posterior inference feasible. The resulting MVG model
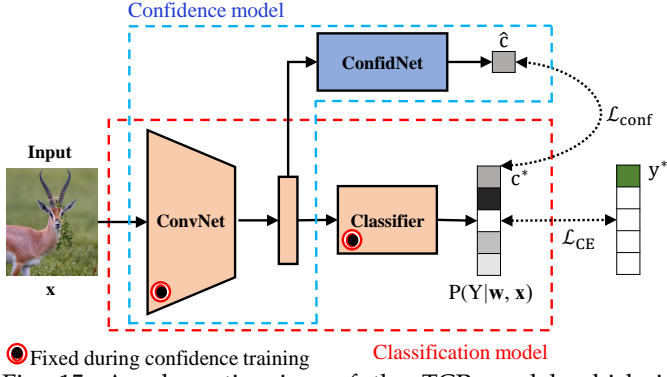
Fig. 15: A schematic view of the TCP model which is reproduced based on the same reference. [169].

was applied to a probabilistic BP framework to estimate posterior inference. Louizos and Welling [163] used MVG distribution to estimate the weight posterior uncertainty. They treated the weight matrix as a whole rather than treating each component of weight matrix independently. As mentioned earlier, GPs were widely used for UQ in deep learning methods. Van der Wilk et al. [164], Blomqvist et al. [165], Tran et al. [166], Dutordoir et al. [167] and Shi et al. [168] introduced convolutional structure into GP. In another study, Corbière et al. [169] expressed that the confidence of DDNs and predicting their failures is of key importance for the practical application of these methods. In this regard, they showed that the TCP ($True\,Class\,Probability$) is more suitable than the MCP ($Maximum\,Class\,Probability$) for failure prediction of such deep learning methods as follows:

$$TCP : \mathbb{R}^d \times \mathcal{Y} \to \mathbb{R}$$
$$(x, y^*) \to P(Y = y^*|w, x), \quad (26)$$

where $x_i \in \mathbb{R}^d$ represents a d-dimensional feature and $y_i^* \in \mathcal{Y} = \{1, ..., K\}$ is its correct class. Then, they introduced a new normalized type of the TCP confidence criterion:

$$TCP^r(x, y^*) = \frac{P(Y = y^*|w, x)}{P(Y = \hat{y}|w, x)}. \quad (27)$$

A general view of the proposed model in [169] is illustrated by Fig. 15:

In another research, Atanov et al. [170] introduced a probabilistic model and showed that Batch Normalization (BN) approach can maximize the lower bound of its related marginalized log-likelihood. Since inference computationally was not efficient, they proposed the Stochastic BN (SBN) approach for approximation of appropriate inference procedure, as an uncertainty estimation method. Moreover, the induced noise is generally employed to capture the uncertainty, check overfitting and slightly improve the performance via test-time averaging whereas ordinary stochastic neural networks typically depend on the expected values of their weights to formulate predictions. Neklyudov et al. [171] proposed a different kind of stochastic layer called variance layers. It is parameterized by its variance and each weight of a variance layer obeyed a zero-mean distribution. It implies that each object was denoted by a zero-mean distribution in the space of the activations. They demonstrated that these layers presented an upright defense

against adversarial attacks and could serve as a crucial exploration tool in reinforcement learning tasks.

4.2 Laplace approximations

Laplace approximations (LAs) are other popular UQ methods which are used to estimate the Bayesian inference [172]. They build a Gaussian distribution around true posterior using a Taylor expansion around the MAP, θ^*, as follows:

$$p(\theta|D) \approx p(\theta^*) \exp\{-\frac{1}{2}(\theta - \theta^*)'\mathbb{H}_{|\theta^*}(\theta - \theta^*)\} \quad (28)$$

where $\mathbb{H}_{|\theta} = \nabla_\theta p(y|\theta)\nabla_\theta p(y|\theta)'$ indicates the Hessian of the likelihood estimated at the MAP estimate. Ritter et al. [173] introduced a scalable LA (SLA) approach for different NNs. The proposed the model, then compared with the other well-known methods such as Dropout and a diagonal LA for the uncertainty estimation of networks. There are some more studies on LA such as [174], [175], [176], [177]. For example, Shinde et al. [174] with the help of Conditional Random Fields (CRF) on top of BNNs, they could figure out contextual information and semi-supervised learning. Thereafter, the authors compared the performance of LA with some variant of MC-dropout. Shinde et al. in another study [175] evaluated the performance of LA on autonomous driving application (KITTI dataset used). Their findings showed that even though uncertainty can be meaningful; however, they encouraged conducting more experimental evaluation on this point. In another study, Lee et al. [176] used LA-based inference engine for natural parameters/information form of Gaussian distribution. The authors managed to scale LA on the ImageNet dataset by spending considerable time for tuning the hyperparameters such that they could make a meaningful comparison. Finally, Humt et al. [177] applied the existing BO techniques to tune the hyperparameters of LA. The obtained outcomes indicated that the proposed BO approach required fewer iterations compared to when random search was performed.

5 UNCERTAINTY QUANTIFICATION IN REINFORCEMENT LEARNING

In decision making process, uncertainty plays a key role in decision performance in various fields such as Reinforcement Learning (RL) [178]. Different UQ methods in RL have been widely investigated in the literature [179]. Lee et al. [180] formulated the model uncertainty problem as Bayes-Adaptive Markov Decision Process (BAMDP). The general BAMDP defined by a tuple $\langle$ S, Φ, A, T, R, P_0, γ $\rangle$, where where S shows the underlying MDP's observable state space, Φ indicates the latent space, A represents the action space, T is the parameterized transition and finally R is the reward functions, respectively. Lets b_0 be an initial belief, a Bayes filter updates the posterior as follows:

$$b'(\phi'|s, b, a', s') = \eta \sum_{\phi \in \Phi} b(\phi)T(s, \phi, a', s', \phi') \quad (29)$$

Then, Bayesian Policy Optimization (BPO) method (see Fig. 16) is applied to POMDPs as a Bayes filter to compute the belief b of the hidden state as follows:

$$b'(s') = \psi(b, a', o') = \eta \sum_{s \in S} b(s)T(s, a', s')Z(s, a', o') \quad (30)$$
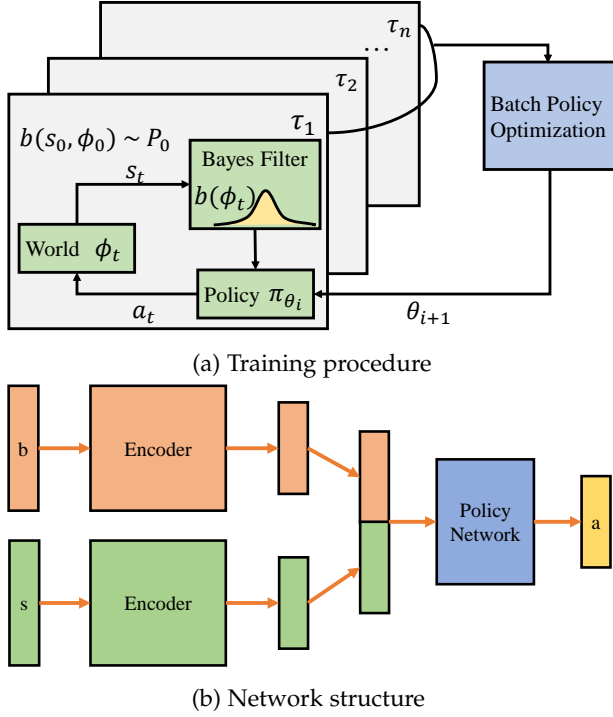
(a) Training procedure

(b) Network structure

Fig. 16: A general view of BPO which is reproduced based on [180].

In another research, O'Donoghue et al. [181] proposed the uncertainty Bellman equation (UBE) to quantify uncertainty. The authors used a Bellman-based which propagated the uncertainty (here variance) relationship of the posterior distribution of Bayesian. Kahn et al. a [182] presented a new UA model for learning algorithm to control a mobile robot. A review of past studies in RL shows that different Bayesian approaches have been used for handling parameters uncertainty [183]. Bayesian RL was significantly reviewed by Ghavamzadeh et al. [183] in 2015. Due to page limitation, we do not discuss the application of UQ in RL; but we summarise some of the recent studies here.

Kahn et al. a [182] used both Bootstrapping and Dropout methods to estimate uncertainty in NNs and then used in UA collision prediction model. Besides Bayesian statistical methods, ensemble methods have been used to quantify uncertainty in RL [184]. In this regard, Tschantz et al. [184] applied an ensemble of different point-estimate parameters $\theta = \{\theta_0, ..., \theta_B\}$ when trained on various batches of a dataset D and then maintained and treated by the posterior distribution $p(\theta|D)$. The ensemble method helped to capture both aleatoric and epistemic uncertainty. There are more UQ techniques used in RL, however, we are not able to discuss all of them in details in this work due to various reasons, such as page restrictions and the breadth of articles. Table 2 summarizes different UQ methods used in a variety of RL subjects.

6 ENSEMBLE TECHNIQUES

Deep neural networks (DNNs) have been effectively employed in a wide variety of machine learning tasks and have achieved state-of-the-art performance in different domains such as bioinformatics, natural language processing (NLP), speech recognition and computer vision [201], [202]. In

supervised learning benchmarks, NNs yielded competitive accuracies, yet poor predictive uncertainty quantification. Hence, it is inclined to generate overconfident predictions. Incorrect overconfident predictions can be harmful; hence it is important to handle UQ in a proper manner in real-world applications [203]. As empirical evidence of uncertainty estimates are not available in general, quality of predictive uncertainty evaluation is a challenging task. Two evaluation measures called calibration and domain shift are applied which usually are inspired by the practical applications of NNs. Calibration measures the discrepancy between long-run frequencies and subjective forecasts. The second notion concerns generalization of the predictive uncertainty to domain shift that is estimating if the network knows what it knows. An ensemble of models enhances predictive performance. However, it is not evident why and when an ensemble of NNs can generate good uncertainty estimates. Bayesian model averaging (BMA) believes that the true model reclines within the hypothesis class of the prior and executes soft model selection to locate the single best model within the hypothesis class. On the contrary, ensembles combine models to discover more powerful model; ensembles can be anticipated to be better when the true model does not lie down within the hypothesis class.

The authors in [204] devised Maximize Overall Diversity (MOD) model to estimate ensemble-based uncertainty by taking into account diversity in ensemble predictions across future possible inputs. Gustafsson et al. [205] presented an evaluation approach for measuring uncertainty estimation to investigate the robustness in computer vision domain. Researchers in [206] proposed a deep ensemble echo state network model for spatio-temporal forecasting in uncertainty quantification. Chua et al. [207] devised a novel method called probabilistic ensembles with trajectory sampling that integrated sampling-based uncertainty propagation with UA deep network dynamics approach. The authors in [201] demonstrated that prevailing calibration error estimators were unreliable in small data regime and hence proposed kernel density-based estimator for calibration performance evaluation and proved its consistency and unbiasedness. Liu et al. [208] presented a Bayesian nonparametric ensemble method which enhanced an ensemble model that augmented model's distribution functions using Bayesian nonparametric machinery and prediction mechanism. Hu et al. [209] proposed a model called margin-based Pareto deep ensemble pruning utilizing deep ensemble network that yielded competitive uncertainty estimation with elevated confidence of prediction interval coverage probability and a small value of the prediction interval width. In another study, the researchers [210] exploited the challenges associated with attaining uncertainty estimations for structured predictions job and presented baselines for sequence-level out-of-domain input detection, sequence-level prediction rejection and token-level error detection utilizing ensembles. Ensembles involve memory and computational cost which is not acceptable in many application [211]. There has been noteworthy work done on the distillation of an ensemble into a single model. Such approaches achieved comparable accuracy using ensembles and mitigated the computational costs. In posterior distribution $p(\theta|D)$, the uncertainty of model is captured. Let us consider from the posterior sam-

Study	Year	Application	Goal/Objective	UQ method	Code
Kalweit and Boedecker [185]	2017	Continuous Deep RL (CDRL)	Minimizing real-world interaction	Model-assisted Bootstrapped Deep Deterministic Policy Gradient (MA-BDDPG)	×
Tegho et al. [186]	2018	Dialogue management context	Dialogue policy optimisation	BBB propagation deep Q-networks (BBQN)	×
Riquelme et al. [187]	2018	Approximating the posterior sampling	Balancing both exploration and exploitation in different complex domains	Deep Bayesian Bandits Showdown using Thompson sampling	√
Pearce et al. [188]	2018	Exploration in RL	Confidence in action analysis	Bayesian inference using Anchored ensembles of NNs	√
Janz et al. [189]	2019	Temporal difference learning	Posterior sampling for RL (PSRL)	Successor Uncertainties (SU)	√
Shen and How [190]	2019	Discriminating potential threats	Stochastic belief space policy	Soft-Q learning	×
Benatan and Pyzer-Knapp [191]	2019	Safe RL (SRL)	The weights in RNN using mean and variance weights	Probabilistic Backpropagation (PBP)	×
Huang et al. [192]	2019	Model-based RL (MRL)	Better decision and improve performance	Bootstrapped model-based RL (BMRL)	×
Eriksson and Dimitrakakis [193]	2019	Risk measures and leveraging preferences	Risk-Sensitive RL (RSRL)	Epistemic Risk Sensitive Policy Gradient (EPPG)	×
Lötjens et al. [194]	2019	SRL	UA navigation	Ensemble of MC dropout (EMCD) and Bootstrapping	×
Clements et al. [195]	2019	Designing risk-sensitive algorithm	Disentangling aleatoric and epistemic uncertainties	Combination of distributional RL (DRL) and Approximate Bayesian computation (ABC) methods with NNs	√
Metelli et al. [196]	2019	Temporal difference learning	Balancing exploration and exploitation	WQL: Wasserstein Q-Learning	√
D'Eramo et al. [197]	2019	Drive exploration	Multi-Armed Bandit (MAB)	Bootstrapped deep Q-network with TS (BDQNTS)	×
Tschantz et al. [198]	2020	Model-based RL	Exploration and exploitation	Scaling active inference	×
Lee et al. [199]	2020	Multi-agent RL	Lack of entirely accurate knowledge of each agent in the model	robust Markov game	×
Ramakrishnan et al. [200]	2020	Blind spot detection	Mismatch between training and execution environments	Safe sim-to-real transfer	×

pled ensemble of models $\{P(y|x^\star, \theta^{(m)})\}_{m=1}^M$ as follows [211]:

$$\{P(y|x^\star, \theta^{(m)})\}_{m=1}^M \rightarrow \{P(y|\pi^{(m)})\}_{m=1}^M,$$
$$\pi^m = f(x^\star; \theta^{(m)}), \theta^{(m)} \sim p(\theta|D) \quad (31)$$

where x^* a test is input and π represents the parameters of a categorical distribution $[P(y = w_1), ..., P(y = w_k)]^T$. By taken into account the expectation with respect to the model posterior, predictive posterior or the expected predictive distribution, for a test input x^* is acquired. And then we have:

$$P(y|x^\star, D) = \mathbb{E}_{p(\theta|D)}[P(y|x^\star, \theta)] \quad (32)$$

Different estimate of data uncertainty are demonstrated by each of the models $P(y|x^\star, \theta^{(m)})$. The 'disagreement' or the level of spread of an ensemble sampled from the posterior is occurred due to the uncertainty in predictions as a result of model uncertainty. Let us consider an ensemble $\{P(y|x^\star, \theta^{(m)})\}_{m=1}^M$ that yields the expected set of behaviors, the entropy of expected distribution $P(y|x^\star, D)$ can be utilized as an estimate of total uncertainty in the prediction.

Measures of spread or 'disagreement' of the ensemble such as MI can be used to assess uncertainty in predictions due to knowledge uncertainty as follows:

$$\underbrace{\mathcal{MI}[y, \theta|x^\star, D]}_{Knowledge\ Uncertainty} = \underbrace{H[\mathbb{E}_{p(\theta|\mathcal{D})}[P(y|x^\star, \theta)]]}_{Total\ Uncertainty} - \underbrace{\mathbb{E}_{p(\theta|\mathcal{D})}[H[P(y|x^\star, \theta)]]}_{Expected\ Data\ Uncertainty} \quad (33)$$

The total uncertainty can be decomposed into expected data uncertainty and knowledge uncertainty via MI formulation. If the model is uncertain – both in out-of-domain and regions of severe class overlap, entropy of the total uncertainty or predictive posterior is high. If the models disagree, the difference of the expected entropy and entropy of predictive posterior of the individual models will be non-zero. For example, MI will be low and expected and predictive posterior entropy will be similar, and each member of the ensemble will demonstrate high entropy distribution in case of in regions of class overlap. In such scenario, data uncertainty dominates total uncertainty. The predictive posterior is near uniform while the expected entropy of each model may
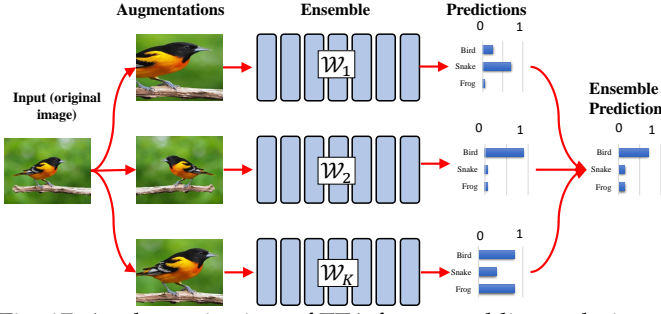
Fig. 17: A schematic view of TTA for ensembling techniques which is reproduced based on [212].

be low that yielded from diverse distributions over classes as a result of out-of-domain inputs on the other hand.In this region of input space, knowledge uncertainty is high because of the model's understanding of data is low. In ensemble distribution distillation, the aim is not only to capture its diversity but also the mean of the ensemble. An ensemble can be observed as a set of samples from an implicit distribution of output distributions:

$$\{P(y|x^\star, \theta^{(m)})\}_{m=1}^M \rightarrow \{P(y|\pi^{(m)})\}_{m=1}^M, \pi^{(m)} \sim p(\pi|x^\star, \mathcal{D}).$$
$$(34)$$

Prior Networks, a new class model was proposed that explicitly parameterize a conditional distribution over output distributions $p(\pi|x^\star, \hat{\emptyset})$ utilizing a single neural network parameterized by a point estimate of the model parameters $\hat{\emptyset}$. An ensemble can be emulated effectively by a Prior Networks and hence illustrated the same measure of uncertainty. By parameterizing the Dirichlet distribution, the Prior Network $p(\pi|x^\star, \hat{\emptyset})$ represents a distribution over categorical output distributions. Ensembling performance is measured by uncertainty estimation. Deep learning ensembles produces benchmark results in uncertainty estimation. The authors in [212] exploited in-domain uncertainty and examined its standards for its quantification and revealed pitfalls of prevailing matrices. They presented the deep ensemble equivalent score (DEE) and demonstrated how an ensemble of trained networks which is only few in number can be equivalent to many urbane ensembling methods with respect to test performance. For one ensemble, they proposed the test-time augmentation (TTA) in order to improve the performance of different ensemble learning techniques (see Fig. 17). There are some more studies on TTA for quantifying uncertainties such as [213], [38], [214], [215].

In another research, Wilson and Izmailov [216] showed how deep ensemble models can provide a better approximation to the Bayesian model average in DL compared to standard Bayesian methods (such as BBB). In other words, they believe that deep ensemble models can be considered as a Bayesian approach. Moreover, the authors discuss the importance of multi-basin marginalization (by applying the new procedure called MultiSWAG) for epistemic uncertainty representation, robustness to data corruption and finally performance. They also showed that it can entirely alleviate double descent behaviour. They analyzed priors over functions $P(f)$ when induced by Gaussian priors over

different parameters of NNs. Based on this procedure, they could show that the proposed process has several significant properties such as:

1) A valuable induced correlation function for a wide variety of images;
2) Impressive ability to generate beneficial features;
3) Excellent support for structured datasets versus corrupted datasets through the marginal likelihood;
4) Significant performance improvement over standard training.

The obtained outcomes showed how probability UQ in DL can help providing a prescription of model construction which eliminates ambiguous generalization properties of DNNs, such as the Having the capacity for fitting images with random classes (labels), double descent, and over parametrization.

However, deep ensembles [217] are a simple approach that presents independent samples from various modes of the loss setting. Under a fixed test-time computed budget, deep ensembles can be regarded as powerful baseline for the performance of other ensembling methods. It is a challenging task to compare the performance of ensembling methods. Different values of matrices are achieved by different models on different datasets. Interpretability is lacking in values of matrices as performance gain is compared with dataset and model specific baseline. Hence, Ashukha et al. [212] proposed DDE with an aim to introduce interpretability and perspective that applies deep ensembles to compute the performance of other ensembling methods. DDE score tries to answer the question: what size of deep ensemble demonstrates the same performance as a specific ensembling technique? The DDE score is based on calibrated log-likelihood (CLL). DDE is defined for an ensembling technique (m) and lower and upper bounds are depicted as below [212]:

$$\mathcal{DEE}_m(k) = min\{l \in R, l \geq 1 | CLL_{DE}^{mean}(l) \geq CLL_m^{mean}(k)\},$$
$$(35)$$

$$\mathcal{DEE}_m^{upper/lower}(k) = min\{l \in R, l \geq 1 | CLL_{DE}^{mean}(l)$$
$$\mp CLL_{DE}^{std}(l) \geq CLL_m^{mean}(k)\}, \quad (36)$$

where the mean and standard deviation of the calibrated log-likelihood yielded by an ensembling technique m with l samples is dubbed as $CLL_m^{mean/std}(l)$. They measured $CLL_{DE}^{std}(l)$ and $CLL_{DE}^{mean}(l)$ for natural numbers $l \in \mathbb{N}_{>0}$ and linear interpolation is applied to define them for real values $l \geq 1$. They depict $DDE_m(k)$ for different number of samples k for different methods m with upper and lower bounds $DEE_m^{upper}(k)$ and $DEE_m^{lower}(k)$.

Different sources of model uncertainty can be taken care by incorporating a presented ensemble technique to propose a Bayesian nonparametric ensemble (BNE) model devised by Liu et al. [208]. Bayesian nonparametric machinery was utilized to augment distribution functions and prediction of a model by BNE. The BNE measure the uncertainty patterns in data distribution and decompose uncertainty into discrete components that are due to error and noise. The model yielded precise uncertainty estimates from observational noise and demonstrated its utility with respect to model's

bias detection and uncertainty decomposition for an ensemble method used in prediction. The predictive mean of BNE can be expressed as below [208]:

$$E(y|X, \omega, \delta, G) = \sum_{k=1}^{K} f_k(X)\omega_k + \underbrace{\delta(X)}_{Due\ to\ \delta}$$
$$+ \underbrace{\int_{y \in \dagger} \left[\Phi((y|X, \mu) - G[\Phi((y|X, \mu)] \right] dy}_{Due\ to\ G}. \quad (37)$$

The predictive mean for the full BNE is comprised of three sections:

1) The predictive mean of original ensemble $\sum_{k=1}^{K} f_k(X)\omega_k$;
2) BNE's direct correction to the prediction function is represented by the term δ; and
3) BNE's indirect correction on prediction derived from the relaxation of the Gaussian assumption in the model cumulative distribution function is represented by the term $\int \left[\Phi((y|X, \mu) - G[\Phi((y|X, \mu)] \right] dy$. In addition, two error correction terms $\mathcal{D}_\delta(y|X)$ and $\mathcal{D}_G(y|X)$ are also presented.

To denote BNE's predictive uncertainty estimation, the term $\Phi_{\varepsilon, \omega}$ is used which is the predictive cumulative distribution function of the original ensemble (i.e. with variance σ_ε^2 and mean $\sum_k f_k \omega_k$). The BNE's predictive interval is presented as [208]:

$$U_q(y|X, \omega, \delta, G) = \left[\Phi_{\varepsilon, \omega}^{-1} \left(G^{-1}(1 - \frac{q}{2}|X) \right) + \delta(x), \right.$$
$$\left. \Phi_{\varepsilon, \omega}^{-1} \left(G^{-1}(1 + \frac{q}{2}|X) \right) + \delta(x) \right]. \quad (38)$$

Comparing the above equation to the predictive interval of original ensemble $\left[\Phi_{\varepsilon, \omega}^{-1} \left(G^{-1}(1 - \frac{q}{2}|X) \right), \Phi_{\varepsilon, \omega}^{-1} \left(G^{-1}(1 + \frac{q}{2}|X) \right) \right]$, it can be observed that the residual process δ adjusts the locations of the BNE predictive interval endpoints while G calibrates the spread of the predictive interval.

As an important part of ensemble techniques, loss functions play a significant role of having a good performance by different ensemble techniques. In other words, choosing the appropriate loss function can dramatically improve results. Due to page limitation, we summarise the most important loss functions applied for UQ in Table 3.

6.1 Deep Ensemble

Deep ensemble, is another powerful method used to measure uncertainty and has been extensively applied in many real-world applications [209]. To achieve good learning results, the data distributions in testing datasets should be as close as the training datasets. In many situations, the distributions of test datasets are unknown especially in case of uncertainty prediction problem. Hence, it is tricky for the traditional learning models to yield competitive performance. Some researchers applied MCMC and BNNs that relied on the prior distribution of datasets to work out the uncertainty prediction problems [204]. When these approaches are employed into large size networks, it becomes computationally expensive. Model ensembling is an effective technique which can be used to enhance the predictive performance of supervised learners. Deep ensembles are applied to get better predictions on test data and also produce model uncertainty estimates when learns are provided with OoD data. The success of ensembles depends on the variance-reduction generated by combining predictions that are prone to several types of errors individually. Hence, the improvement in predictions is comprehended by utilizing a large ensemble with numerous base models and such ensembles also generate distributional estimates of model uncertainty. A deep ensemble echo state network (D-EESN) model with two versions of the model for spatio-temporal forecasting and associated uncertainty measurement presented in [206]. The first framework applies a bootstrap ensemble approach and second one devised within a hierarchical Bayesian framework. Multiple levels of uncertainties and non-Gaussian data types were accommodated by general hierarchical Bayesian approach. The authors in [206] broadened some of the deep ESN technique constituents presented by Antonelo et al. [223] and Ma et al. [224] to fit in a spatio-temporal ensemble approach in the D-EESN model to contain such structure. As shown in previous section, in the following , we summarise few loss functions of deep ensembles in Table 4.

6.2 Deep Ensemble Bayesian

The expressive power of various ensemble techniques extensively shown in the literature. However, traditional learning techniques suffered from several drawbacks and limitations as listed in [228]. To overcome these limitations, Fersini et al. [228] utilized the ensemble learning approach to mitigate the noise sensitivity related to language ambiguity and more accurate prediction of polarity can be estimated. The proposed ensemble method employed Bayesian model averaging, where both reliability and uncertainty of each single model were considered. Study [229] presented one alteration to the prevailing approximate Bayesian inference by regularizing parameters about values derived from a distribution that could be set equal to the prior. The analysis of the process suggested that the recovered posterior was centered correctly but leaned to have an overestimated correlation and underestimated marginal variance. To obtain uncertainty estimates, one of the most promising frameworks is Deep BAL (DBAL) with MC dropout. Pop et al. [217] argued that in variational inference methods, the mode collapse phenomenon was responsible for overconfident predictions of DBAL methods. They devised Deep Ensemble BAL that addressed the mode collapse issue and improved the MC dropout method. In another study, Pop et al. [230] proposed a novel AL technique especially for DNNs. The statistical properties and expressive power of model ensembles were employed to enhance the state-of-the-art deep BAL technique that suffered from the mode collapse problem. In another research, Pearce et al. [231] a new ensemble of NNs, approximately Bayesian ensem-

TABLE 3: Main loss functions used by ensemble techniques for UQ.

Study	Dataset type	Base classifier(s)	Method's name	Loss equation	Code
TV et al. [218]	Sensor data	Neural Networks (LSTM)	Ordinal Regression (OR)	$\mathcal{L}_{OR}(y,\hat{y}) = -\frac{1}{N}\sum_{j=1}^{K} y_j \cdot \log(\hat{y}_j) + (1 - y_j) \cdot \log(1 - \hat{y}_j)$	$\times$
Sinha et al. [219]	Image	Neural Networks	Diverse Information Bottleneck in Ensembles (DIBS)	$L_G = \mathbb{E}_{\hat{z}_1 \sim q(\tilde{z}_i|x), \hat{z}_2 \sim q(\tilde{z}_j|x)}[\log \mathcal{D}(\hat{z}_1, \hat{z}_2)] + \mathbb{E}_{\hat{z}_1 \sim r(\tilde{z}), \hat{z}_2 \sim q(\tilde{z}_i|x)}[\log(1 - \mathcal{D}(\hat{z}_1, \hat{z}_2))] + \mathbb{E}_{\hat{z}_1 \sim q(\tilde{z}_i|x), \hat{z}_2 \sim q(\tilde{z}_i|x)}[\log 1 - \mathcal{D}(\hat{z}_1, \hat{z}_2))]$	$\checkmark$
Zhang et al. [201]	Image	Neural Networks	Mix-n-Match Calibration	$\mathbb{E}\|z - y\|_2^2$ (the standard square loss)	$\times$
Lakshminarayanan et al. [203]	Image	Neural Networks	Deep Ensembles	$\mathcal{L}(\theta) = -S(p_\theta, q)$	$\times$
Jain et al. [204]	Image and Protein DNA binding	Deep Ensembles	Maximize Overall Diversity (MOD)	$L(\theta_m; x_n, y_n) = -\log p_{\theta_m}(y_n|x_m)$	$\times$
Gustafsson et al. [205]	Video	Neural Networks	Scalable BDL	Regression: $L(\theta) = \frac{1}{N}\sum_{i=1}^{N} \frac{(y_i - \hat{\mu}(x_i))^2}{\sigma^2(x_i)} + \log \sigma^2(x_i) + \frac{1}{N}\theta^\top \theta$, Classification: $L(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{C} y_{i,k} \log \hat{s}(x_i)_k + \frac{1}{2N}\theta^\top \theta$	$\checkmark$
Chua et al. [207]	Robotics (video)	Neural Networks	Probabilistic ensembles with trajectory sampling (PETS)	$lossp(\theta) = -\sum_{n=1}^{N} \log \widetilde{f}_\theta(s_{n+1}|s_n, a_n)$	$\checkmark$
Hu et al. [209]	Image and tabular data	Neural Networks	margin-based Pareto deep ensemble pruning (MBPEP)	$Loss_{multi} = W_{CVAE} * Loss_{CVAE} + W_{CRNN} * Loss_{CRNN}$	$\times$
Malinin et al. [211]	Image	Neural Networks	Ensemble Distribution Distillation (EnD^2)	$\mathcal{L}(\phi, D_{ens}) = -\frac{1}{N}\sum_{i=1}^{N}\left[\ln \Gamma(\hat{\alpha}_0^{(i)}) - \sum_{c=1}^{K} \ln \Gamma(\hat{\alpha}_c^{(i)}) + \frac{1}{M}\sum_{m=1}^{M}\sum_{c=1}^{K}(\hat{\alpha}_0^{(i)} - 1)\ln \pi_c^{(}im)\right]$	$\checkmark$
Ashukha et al. [212]	Image	Neural Networks	Deep ensemble equivalent score (DEE)	$L(w) = -\frac{1}{N}\sum_{i=1}^{N} \log \hat{p}(y_i^\star|x_i, w) + \frac{\lambda}{2}\|w\|^2 \to \min_w$	$\checkmark$
Pearce et al. [220]	Tabular data	Neural Networks	Quality-Driven Ensembles (QD-Ens)	$Loss_{QD} = MPIW_{capt.} + \lambda\frac{n}{\alpha(1-\alpha)} \max(0, (1 - \alpha) - PICP)^2$	$\checkmark$
Ambrogioni et al. [221]	Tabular data	Bayesian logistic regression	Wasserstein variational gradient descent (WVG)	$\mathcal{L}(z^1) = -\mathbb{E}_{z \sim p(z|x)}[c(z^j, z)]$	$\times$
Hu et al. [222]	Image	Neural Networks	Bias-variance decomposition	$\mathcal{L} = \frac{1}{2}\exp(-s(x))\frac{\sum_r \|y_r(x) - \hat{y}(x)\|^2}{R} + \frac{1}{2}s(x)$	$\times$

TABLE 4: Main loss functions used by deep ensemble techniques for UQ.

Study	Dataset type	Base classifier(s)	Method's name	Loss equation	Code
Fan et al. [225]	GPS-log	Neural Networks	Online Deep Ensemble Learning (ODEL)	$\mathcal{L} = H(F_{ensemble}(X_{t-T:t-1}), one\_hot(X_t))$	$\times$
Yang et al. [226]	Smart grid	K-means	Least absolute shrinkage and selection operator (LASSO)	$\mathcal{L}(y_i^m, \hat{y}_i^{m,q}) = \frac{1}{Q}\sum_{q \in Q} \max\left((q - 1)H_\varepsilon\left(y_i^m, \hat{y}_i^{m,q}\right), qH_\varepsilon\left(y_i^m, \hat{y}_i^{m,q}\right)\right)$	$\times$
van Amersfoort et al. [227]	Image	Neural Networks	Deterministic UQ (DUQ)	$L(x, y) = -\sum_c y_c \log(K_c) + (1 - y_c)\log(1 - K_c)$	$\times$

bling approach, called "*anchoredensembling*". The proposed approach regularises the parameters regarding values attracted from a distribution.

6.3 Uncertainty Quantification in Traditional Machine Learning domain using Ensemble Techniques

It is worthwhile noting that UQ in traditional machine learning algorithms have extensively been studied using different ensemble techniques and few more UQ methods (e.g. please see [232]) or some other UQ methods in classification problems [233] in the literature. However, due to page limitation, we just summarized some of the ensemble techniques (as UQ methods) used in traditional machine learning domain. For example, Tzelepis et al. [232] proposed a maximum margin classifier to deal with uncertainty in input data. The proposed model is applied for classification task using SVM (Support Vector Machine) algorithm with multi-dimensional Gaussian distributions. The proposed model named SVM-GSU (SVM with Gaussian Sample Uncertainty) and it is illustrated by Fig. 18:

Fig. 18: A schematic view of SVM-GSU which is reproduced based on . [232].

In another research, Pereira et al. [234] examined various techniques for transforming classifiers into uncertainty methods where predictions are harmonized with probability estimates with their uncertainty. They applied various uncertainty methods: Venn-ABERS predictors, Conformal Predictors, Platt Scaling and Isotonic Regression. Partalas et al. [235] presented a novel measure called Uncertainty Weighted Accuracy (UWA), for ensemble pruning through directed hill climbing that took care of uncertainty of present ensemble decision. The experimental results demonstrated that the new measure to prune a heterogeneous ensemble significantly enhanced the accuracy compared to baseline methods and other state-of-the-art measures. Peterson et al. [236] exploited different types of errors that might creep in atomistic machine learning, and addressed how uncertainty analysis validated machine-learning predictions. They applied a bootstrap ensemble of neural network based calculators, and exhibited that the width of the ensemble can present an approximation of the uncertainty.

7 FURTHER STUDIES OF UQ METHODS

In this section, we cover other methods used to estimate the uncertainty. In this regard, presented a summary of the proposed methods, but not the theoretical parts. Due to the page limitation and large number of references, we are

Fig. 19: A single block diagram for searching space in the architecture which is reproduced based on [237].

not able to review all the details of the methods. For this reason, we recommend that readers check more details of each method in the reference if needed.

The OoD is a common error appears in machine and deep learning systems when training data have different distribution. To address this issue, Ardywibowo et al. [237] introduced a new UA architecture called *Neural Architecture Distribution Search* ($NADS$). The proposed NADS finds an appropriate distribution of different architectures which accomplish significantly good on a specified task. A single block diagram for searching space in the architecture is presented by Fig. 19.

Unlike previous designing architecture methods, NADS allows to recognize common blocks amongst the entire UA architectures. On the other hand, the cost functions for the uncertainty oriented neural network (NN) are not always converging. Moreover, an optimized prediction interval (PI) is not always generated by the converged NNs. The convergence of training is uncertain and they are not customizable in the case of such cost functions. To construct optimal PIs, Kabir et al. [238] presented a smooth customizable cost function to develop optimal PIs to construct NNs. The PI coverage probability (PICP), PI-failure distances and optimized average width of PIs were computed to lessen the variation in the quality of PIs, enhance convergence probability and speed up the training. They tested their method on electricity demand and wind power generation data. In the case of non-Bayesian deep neural classification, uncertainty estimation methods introduced biased estimates for instances whose predictions are highly accurate. They argued that this limitation occurred because of the dynamics of training with SGD-like optimizers and possessed similar characteristics such as overfitting. Geifman et al. [239] proposed an uncertainty estimation method that computed the uncertainty of highly confident points by utilizing snapshots of the trained model before their approximations were jittered. The proposed algorithm outperformed all well-known techniques. In another research, Tagasovska et al. [240] proposed single-model estimates for DNNs of
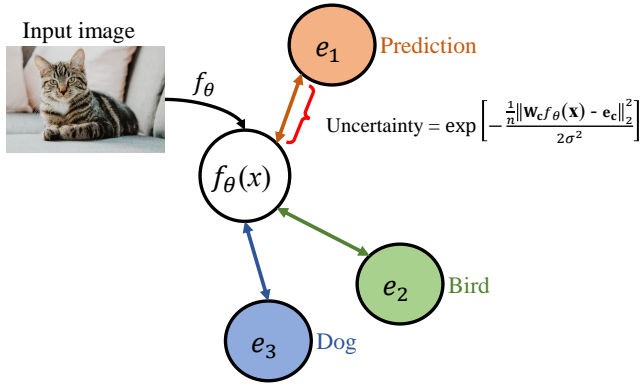
Fig. 20: A general view of the DUQ architecture which is reproduced based on [227], [241].

epistemic and aleatoric uncertainty. They suggested a loss function called Simultaneous Quantile Regression (SQR) to discover the conditional quantiles of a target variable to assess aleatoric uncertainty. Well-calibrated prediction intervals could be derived by using these quantiles. They devised Orthonormal Certificates (OCs), a collection of non-constant functions that mapped training samples to zero to estimate epistemic uncertainty. The OoD examples were mapped by these certificates to non-zero values.

van Amersfoort et al. [227], [241] presented a method to find and reject distribution data points for training a deterministic deep model with a single forward pass at test time. They exploited the ideas of RBF networks to devise deterministic UQ (DUQ) which is presented in Fig. 20. They scaled training in this with a centroid updating scheme and new loss function. Their method could detect out-of-distribution (OoD) data consistently by utilizing a gradient penalty to track changes in the input. Their method is able to enhance deep ensembles and scaled well to huge databases. Tagasovska et al. [242] demonstrated frequentist estimates of epistemic and aleatoric uncertainty for DNNs. They proposed a loss function, simultaneous quantile regression to estimate all the conditional quantiles of a given target variable in case of aleatoric uncertainty. Well-calibrated prediction intervals could be measured by using these quantiles. They proposed a collection of non-trivial diverse functions that map all training samples to zero and dubbed as training certificates for the estimation of epistemic uncertainty. The certificates signalled high epistemic uncertainty by mapping OoD examples to non-zero values. By using Bayesian deep networks, it is possible to know what the DNNs do not know in the domains where safety is a major concern. Flawed decision may lead to severe penalty in many domains such as autonomous driving, security and medical diagnosis. Traditional approaches are incapable of scaling complex large neural networks. Mobiny et al. [243] proposed an approach by imposing a Bernoulli distribution on the model weights to approximate Bayesian inference for DNNs. Their framework dubbed as MC-DropConnect demonstrated model uncertainty by small alternation in the model structure or computed cost. They validated their technique on various datasets and architectures for semantic segmentation and classification tasks. They also introduced a novel uncertainty quantification metrics. Their experimental results showed considerable enhancements in

uncertainty estimation and prediction accuracy compared to the prior approaches.

Uncertainty measures are crucial estimating tools in machine learning domain, that can lead to evaluate the similarity and dependence between two feature subsets and can be utilized to verify the importance of features in clustering and classification algorithms. There are few uncertainty measure tools to estimate a feature subset including rough entropy, information entropy, roughness, and accuracy etc. in the classical rough sets. These measures are not proper for real-valued datasets and relevant to discrete-valued information systems. Chen et al. [244] proposed the neighborhood rough set model. In their approach, each object is related to a neighborhood subset, dubbed as a neighborhood granule. Different uncertainty measures of neighborhood granules were introduced, that were information granularity, neighborhood entropy, information quantity, and neighborhood accuracy. Further, they confirmed that these measures of uncertainty assured monotonicity, invariance and non-negativity. In the neighborhood systems, their experimental results and theoretical analysis demonstrated that information granularity, neighborhood entropy and information quantity performed superior to the neighborhood accuracy measure. On the other hand, reliable and accurate machine learning systems depends on techniques for reasoning under uncertainty. The UQ is provided by a framework using Bayesian methods. But Bayesian uncertainty estimations are often imprecise because of the use of approximate inference and model misspecification. Kuleshov et al. [245] devised a simple method for calibrating any regression algorithm; it was guaranteed to provide calibrated uncertainty estimates having enough data when used to probabilistic and Bayesian models. They assessed their technique on recurrent, feedforward neural networks, and Bayesian linear regression and located outputs well-calibrated credible intervals while enhancing performance on model-based RL and time series forecasting tasks.

Gradient-based optimization techniques have showed its efficacy in learning overparameterized and complex neural networks from non-convex objectives. Nevertheless, generalization in DNNs, the induced training dynamics, and specific theoretical relationship between gradient-based optimization methods are still unclear. Rudner et al. [246] examined training dynamics of overparameterized neural networks under natural gradient descent. They demonstrated that the discrepancy between the functions obtained from non-linearized and linearized natural gradient descent is smaller in comparison to standard gradient descent. They showed empirically that there was no need to formulate a limit argument about the width of the neural network layers as the discrepancy is small for overparameterized neural networks. Finally, they demonstrated that the discrepancy was small on a set of regression benchmark problems and their theoretical results were steady with empirical discrepancy between the functions obtained from non-linearized and linearized natural gradient descent. Patro et al. [247] devised gradient-based certainty estimates with visual attention maps. They resolved visual question answering job. The gradients for the estimates were enhanced by incorporating probabilistic deep learning techniques. There are two key advantages: 1. enhancement in getting the certainty

estimates correlated better with misclassified samples and 2. state-of-the-art results obtained by improving attention maps correlated with human attention regions. The enhanced attention maps consistently improved different techniques for visual question answering. Improved certainty estimates and explanation of deep learning techniques could be achieved through the presented method. They provided empirical results on all benchmarks for the visual question answering job and compared it with standard techniques.

BNNs have been used as a solution for neural network predictions, but it is still an open challenge to specify their prior. Independent normal prior in weight space leads to weak constraints on the function posterior, permit it to generalize in unanticipated ways on inputs outside of the training distribution. Hafner et al. [14] presented noise contrastive priors (NCPs) to estimate consistent uncertainty. The prime initiative was to train the model for data points outside of the training distribution to output elevated uncertainty. The NCPs relied on input prior, that included noise to the inputs of the current mini batch, and an output prior, that was an extensive distribution set by these inputs. The NCPs restricted overfitting outside of the training distribution and produced handy uncertainty estimates for AL. BNNs with latent variables are flexible and scalable probabilistic models. They can record complex noise patterns in the data by using latent variables and uncertainty is accounted by network weights. Depeweg et al. [248] exhibited the decomposition and derived uncertainty into aleatoric and epistemic for decision support systems. That empowered them to detect informative points for AL of functions with bimodal and heteroscedastic noises. They further described a new risk-sensitive condition to recognize policies for RL that balanced noise aversion, model-bias and expected cost by applying decomposition.

Uncertainty modelling in DNNs is an open problem despite advancements in the area. BNNs, where the prior over network weights is a design choice, is a powerful solution. Frequently normal or other distribution supports sparsity. The prior is agnostic to the generative process of the input data. This may direct to unwarranted generalization for out-of- distribution tested data. Rohekar et al. [249] suggested a confounder for the relation between the discriminative function and the input data given the target label. They proposed for modelling the confounder by sharing neural connectivity patterns between the discriminative and generative networks. Hence, a novel deep architecture was framed where networks were coupled into a compact hierarchy and sampled from the posterior of local causal structures (see Fig. 21).

They showed that sampling networks from the hierarchy, an efficient technique, was proportional to their posterior and different types of uncertainties could be estimated. It is a challenging job to learn unbiased models on imbalanced datasets. The generalization of learned boundaries to novel test examples are hindered by concentrated representation in the classification space in rare classes. Khan et al. [250] yielded that the difficulty level of individual samples and rarity of classes had direct correlation with Bayesian uncertainty estimates. They presented a new approach for uncertainty based class
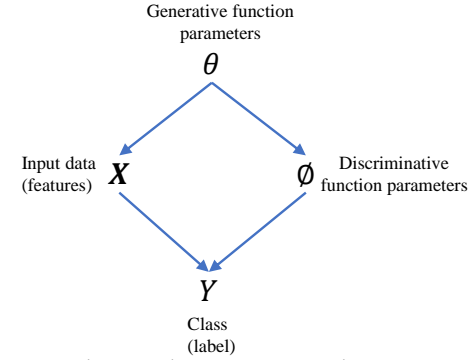

Fig. 21: A causal view demonstrating the main assumptions taken by Rohekar et al. [249] (this figure is reproduced based on the reference).

imbalance learning that exploited two-folded insights: 1. In rare (uncertain) classes, the classification boundaries should be broadened to evade overfitting and improved its generalization; 2. sample's uncertainty was defined by multivariate Gaussian distribution with a covariance matrix and a mean vector that modelled each sample. Individual samples and its distribution in the feature space should be taken care by the learned boundaries. Class and sample uncertainty information was used to obtain generalizable classification techniques and robust features. They formulated a loss function for max-margin learning based on Bayesian uncertainty measure. Their technique exhibited key performance enhancements on six benchmark databases for skin lesion detection, digit/object classification, attribute prediction and face verification.

Neural networks do not measure uncertainty meaningfully as it leans to be overconfident on incorrectly labelled, noisy or unseen data. Variational approximations such as Multiplicative Normalising Flows or BBB are utilized by BDL to overcome this limitation. However, current methods have shortcomings regarding scalability and flexibility. Pawlowski et al. [251] proposed a novel technique of variational approximation, termed as Bayes by Hypernet (BbH) that deduced hypernetworks as implicit distributions. It naturally scaled to deep learning architectures and utilized neural networks to model arbitrarily complex distributions. Their method was robust against adversarial attacks and yielded competitive accuracies. On the other hand, significant increase in prediction accuracy records in deep learning models, but it comes along with the enhancement in the cost of rendering predictions. Wang et al. [252] speculated that for many of the real world inputs, deep learning models created recently, it tended to "overthink" on simple inputs. They proposed I Don't Know" (IDK) prediction cascades approach to create a set of pretrained models systematically without a loss in prediction accuracy to speed up inference. They introduced two search based techniques for producing a new cost-aware objective as well as cascades. Their IDK cascade approach can be adopted in a model without further model retraining. They tested its efficacy on a variety of benchmarks.

Yang et al. [253] proposed a deep learning approach for propagating and quantifying uncertainty in models inspired by non-linear differential equations utilized by physics-informed neural networks. Probabilistic representations

for the system states were produced by latent variable models while physical laws described by partial differential equations were satisfied by constraining their predictions. It also forwards an adversarial inference method for training them on data. A regularization approach for efficiently training deep generative models was provided by such physics-informed constraints. Surrogates of physical models in which the training of datasets was usually small, and the cost of data acquisition was high. The outputs of physical systems were characterized by the framework due to noise in their observations or randomness in their inputs that bypassed the need of sampling costly experiments or numerical simulators. They proved efficacy of their method via a series of examples that demonstrated uncertainty propagation in non-linear conservation laws and detection of constitutive laws. For autonomous driving, 3D scene flow estimation techniques generate 3D motion of a scene and 3D geometry. Brickwedde et al. [254] devised a new monocular 3D scene flow estimation technique dubbed as Mono-SF that assessed both motion of the scene and 3D structure by integrating single-view depth information and multi-view geometry. A CNN algorithm termed as ProbDepthNet was devised for combining single-view depth in a statistical manner. The new recalibration technique, ProbDepth-Net, was presented for regression problems to guarantee well-calibrated distributions. ProbDepthNet design and Mono-SF method proved its efficacy in comparison to the state-of-the-art approaches.

Mixup is a DNN training technique where extra samples are produced during training by convexly integrating random pairs of images and their labels. The method had demonstrated its effectiveness in improving the image classification performance. Thulasidasan et al. [255] investigated the predictive uncertainty and calibration of models trained with mixup. They revealed that DNNs trained with mixup were notably better calibrated than trained in regular technique. They tested their technique in large datasets and observed that this technique was less likely to over-confident predictions using random-noise and OoD data. Label smoothing in mixup trained DNNs played a crucial role in enhancing calibration. They concluded that training with hard labels caused overconfidence observed in neural networks. The transparency, fairness and reliability of the methods can be improved by explaining black-box machine learning models. Model's robustness and users' trust raised concern as the explanation of these models exhibited considerable uncertainty. Zhang et al. [256] illustrated the incidence of three sources of uncertainty, viz. variation in explained model credibility, variation with sampling proximity and randomness in the sampling procedure across different data points by concentrating on a specific local explanation technique called Local Interpretable Model-Agnostic Explanations (LIME). Even the black-box models with high accuracy yielded uncertainty. They tested the uncertainty in the LIME technique on two publicly available datasets and synthetic data.

In the incidence of even small adversarial perturbations, employment of DNNs in safety-critical environments is rigorously restricted. Sheikholeslami et al. [257] devised a randomized approach to identify these perturbations

that dealt with minimum uncertainty metrics by sampling at the hidden layers during the DNN inference period. Adversarial corrupted inputs were identified by the sampling probabilities. Any pre-trained DNN at no additional training could be exploited by new detector of adversaries. The output uncertainty of DNN from the BNNs perspectives could be quantified by choosing units to sample per hidden layer where layer-wise components denoted the overall uncertainty. Low-complexity approximate solvers were obtained by simplifying the objective function. These approximations associated state-of-the-art randomized adversarial detectors with the new approach in addition to delivering meaningful insights. Moreover, consistency loss between various predictions under random perturbations is the basis of one of the effective strategies in semi-supervised learning. In a successful student model, teachers' pseudo labels must possess good quality, otherwise learning process will suffer. But the prevailing models do not evaluate the quality of teachers' pseudo labels. Li et al. [258] presented a new certainty-driven consistency loss (CCL) that employed predictive uncertainty information in the consistency loss to learn students from reliable targets dynamically. They devised two strategies i.e. Temperature CCL and Filtering CCL to either pay less attention on the uncertain ones or filter out uncertain predictions in the consistency regularization. They termed it FT-CCL by integrating the two strategies to enhance consistency learning approach. The FT-CCL demonstrated robustness to noisy labels and enhancement on a semi-supervised learning job. They presented a new mutual learning technique where one student was detached with its teacher and gained additional knowledge with another student's teacher.

Englesson et al. [259] introduced a modified knowledge distillation method to achieve computationally competent uncertainty estimates with deep networks. They tried to yield competitive uncertainty estimates both for out and in-of-distribution samples. Their major contributions were as follows: 1. adapting and demonstrating to distillation's regularization effect, 2. presenting a new target teacher distribution, 3. OoD uncertainty estimates were enhanced by a simple augmentation method, and 4. widespread set of experiments were executed to shed light on the distillation method. On the other hand, well calibrated uncertainty and accurate full predictive distributions are provided by Bayesian inference. High dimensionality of the parameter space limits the scaling of Bayesian inference methods to DNNs. Izmailov et al. [28] designed low-dimensional subspaces of parameter space that comprised of diverse sets of high performing approaches. They applied variational inference and elliptical slice sampling in the subspaces. Their method yielded well-calibrated predictive uncertainty and accurate predictions for both image classification and regression by exploiting Bayesian model averaging over the induced posterior in the subspaces.

Csáji et al. [260] introduced a data-driven strategy for uncertainty quantification of models based on kernel techniques. The method needed few mild regularities in the computation of noise instead of distributional assumptions such as dealing with exponential families or GPs. The uncertainty about the model could be

estimated by perturbing the residuals in the gradient of the objective function. They devised an algorithm to make it distribution-free, non-asymptotically guaranteed and exact confidence regions for noise-free and ideal depiction of function that they estimated. For the symmetric noises and usual convex quadratic problems, the regions were star convex centred on a specified small estimate, and ellipsoidal outer approximations were also efficiently executed. On the other hand, the uncertainty estimates can be measured while pre-training process. Hendrycks et al. [261] demonstrated that pre-training enhanced the uncertainty estimates and model robustness although it might not improve the classification metrics. They showed the key gains from pre-training by performing empirical experiments on confidence calibration, OoD detection, class imbalance, label corruption and adversarial examples. Their adversarial pre-training method demonstrated approximately10% enhancement over existing methods in adversarial robustness. Pre-training without task-specific techniques highlighted the need for pre-training, surpassed the state-of-the-art when examining the future techniques on uncertainty and robustness.

Trustworthy confidence estimates are required by high-risk domains from predictive models. Rigid variational distributions utilized for tractable inference that erred on the side of overconfidence suffered from deep latent variable models. Veeling et al. [262] devised Stochastic Quantized Activation Distributions (SQUAD) that executed a tractable yet flexible distribution over discretized latent variables. The presented technique is sample efficient, self-normalizing and scalable. Their method yielded predictive uncertainty of high quality, learnt interesting non-linearities, fully used the flexible distribution. Multi-task learning (MTL) is another domain that the impact of the importance of uncertainty methods on it can be considered. For example, MTL demonstrated its efficacy for MR-only radiotherapy planning as it can jointly automate contour of organs-at-risk - a segmentation task – and simulate a synthetic CT (synCT) scan - a regression task from MRI scans. Bragman et al. [263] suggested utilizing a probabilistic deep-learning technique to estimate the parameter and intrinsic uncertainty. Parameter uncertainty was estimated through a approximate Bayesian inference whilst intrinsic uncertainty was modelled using a heteroscedastic noise technique. This developed an approach for uncertainty measuring over prediction of the tasks and data-driven adaptation of task losses on a voxel-wise basis. They demonstrated competitive performance in the segmentation and regression of prostate cancer scans. More information can be found in Tables 5 and 6.

As discussed earlier, GP is a powerful technique used for quantifying uncertainty. However, it is complex to form a Gaussian approximation to the posterior distribution even in the context of uncertainty estimation in huge deep-learning models. In such scenario, prevailing techniques generally route to a diagonal approximation of the co-variance matrix in spite of executing low uncertainty estimates by these matrices. Mishkin et al. [444] designed a novel stochastic, low-rank, approximate natural-gradient (SLANG) technique for VI in huge deep models to tackle

this issue. Their technique computed a "diagonal plus low-rank" structure based on back-propagated gradients of the network log-likelihood. Their findings indicate that the proposed technique in forming Gaussian approximation to the posterior distribution. As a fact, the safety of the AI systems can be enhanced by estimating uncertainty in predictions. Such uncertainties arise due to distributional mismatch between the training and test data distributions, irreducible data uncertainty and uncertainty in model parameters. Malinin et al. [382] devised a novel framework for predictive uncertainty dubbed as Prior Networks (PNs) that modelled distributional uncertainty explicitly. They achieved it by parameterizing a prior distribution over predictive distributions. Their work aimed at uncertainty for classification and scrutinized PNs on the tasks of recognizing OoD samples and identifying misclassification on the CIFAR-10 and MNIST datasets. Empirical results indicate that PNs, unlike non-Bayesian methods, could successfully discriminate between distributional and data uncertainty.

7.1 Other UQ Techniques

In this sub-section, we aim to summary few other UQ techniques applied in the literature. We have selected the most relevant ones for this part of study. BNNs have been reinvigorating research interest by enhancing accurate predictions presented by neural networks with well-calibrated predictive uncertainties. But selection of model and number of nodes remain a challenge. Ghosh et al. [445] exploited the horseshoe, continuous shrinkage priors and the regularized horseshoe distributions for selection of model in BNNs (see Fig. 22). The strong shrinkage provided by the horseshoe was effective in turning of nodes that did not help explaining data when placed over node pre-activations and coupled with appropriate variational approximations. Their model selection technique over the number of nodes did not come at the expense of computational or predictive performance. In another research, Hernandez-Lobato et al. [446] introduced a novel approximate inference method based on the minimization of α-divergences termed as black-box alpha (BB-α). It could be implemented using stochastic gradient descent as BB-α scaled to large datasets. Only the likelihood function and its gradients were required as input to implement BB-α in complex probabilistic models. Automatic differentiation could be applied to obtain these gradients. Their method was able to interpolate between variational Bayes and an algorithm similar to expectation propagation by changing the divergence parameter α.

Patro et al. [447] presented a probabilistic approach for solving the task of 'Visual Dialog'. Common sense knowledge to answer and understanding and reasoning of language modality, and visual modality are required to solve this task. They believed that the sources of uncertainty were critical in solving this task. Their framework helped a varied generation of answers and in estimating uncertainty. Their framework comprised of probabilistic representation module that provided representations for conversation history, question and image and an uncertainty representation module that selected the appropriate answer that minimized uncertainty. They achieved an enhanced visual dialog system that is also more explainable utilizing the presented probabilistic

TABLE 5: More UQ methods in the main three categories proposed in the literature. Note that we provide in the row related to other methods the names of the proposed UQ methods for each reference. But, because of the importance of mentioning the proposed method, we also did the same in some other parts (General information).

| UQ category	Studies
Bayesian	Dzunic and Fisher III [264] (SSIM (state-space switching interaction model) with DBN (Dynamic Bayesian Network), Balan et al. [265] (BPE: Bayesian parameter estimation), Houthooft et al. [266] (VIME: VI Maximizing Exploration), Springenberg et al. [267], Lakshminarayanan et al. [268] (Extend Mondrian forests), Ilg et al. [269], Heo et al. [270], Henderson et al. [271], Ahn et al. [272], Zhang et al. [273], Sensoy et al. [274], Khan et al. [160], Acerbi [275] (VBMC: Variational Bayesian MC), Tóthová et al. [276], Haußmann et al. [277], Gong et al. [77], De Ath et al. [278], Foong et al. [279], Hasanzadeh et al. [280], Chang et al. [281], Stoean et al. [282], Xiao et al. [283], Repetti et al. [284], Cardelli et al. [285] (BNN robustness estimation), Moss et al. [286], Dutordoir et al. [287], Luo et al. [288], Gafni et al. [289], Jin et al. [290], Han et al. [291], Stoean et al. [292], Oh et al. [293], Dusenberry et al. [294], Havasi et al. [295], Krishnan et al. [296] (MOPED: MOdel Priors with Empirical Bayes using DNN), Filos et al. [297], Huang et al. [298], Amit and Meir [299], Bhattacharyya et al. [300], Yao et al. [301], Laves et al. [302] (UCE: uncertainty calibration error), Yang et al. [303] (OC-BNN: Output-Constrained BNN), Thakur et al. [304] (LUNA: Learned UA), Yacoby et al. [305] (NCAI: Noise Constrained Approximate Inference), Masood and Doshi-Velez [306] (PVI Particle-based VI), Abdolshah et al. [307] (MOBO: Multi-objective Bayesian optimisation), White et al. [308] (BO), Balandat et al. [309] (BOTORCH), Galy-Fajou et al. [310] (CMGGPC: Conjugate multi-class GP classification), Lee et al. [311] (BTAML: Bayesian Task Adaptive Meta-Learning), Vadera and Marlin [312] (BDK: Bayesian Dark Knowledge), Siahkoohi et al. [313] (SGLD: Stochastic gradient Langevin dynamics), Sun et al. [314], Patacchiola et al. [315], Cheng et al. [316], Oliveira et al. [317] (UCB: Upper confidence bound), Caldeira and Nord [318], Wandzik et al. [319] (MCSD: MC Stochastic Depth), Deng et al. [320] (DBSN: DNN Structures), González-López et al. [321], Foong et al. [322] (ConvNP: Convolutional Neural Process), Yao et al. [323] (SI: Stacked inference), Prijatelj et al. [324], Herzog et al. [325], Prokudin et al. [326] (CVAE: conditional VAE), Tuo and Wang [327], Acerbi [328] (VBMC+EIG (expected information gain)/VIQR (variational interquantile range)), Zhao et al. [329] (GEP: generalized expectation propagation), Li et al. [330] (DBGP: deep Bayesian GP), Jacot et al. [331] (NTK: Neural Tangent Kernel), Wang and Ročková [332] (Gaussian approximability), Jesson et al. [333] (BCEVAE: Bayesian Causal Effect VAE), De Sousa Ribeiro et al. [334] (LIRVI: local iterative routing with VI), Zhao et al. [335] (GKDE: Graph-based Kernel Dirichlet distribution Estimation), Mukherjee and Ahmed Hassan [336] (UST: Uncertainty-aware self-training), Hu et al. [337] (Bayesian uncertainty estimation), Yang et al. [338] (OC-BNN: Output-Constrained BNN), Farquhar et al. [339] (Deep MFV weight posteriors), Charpentier et al. [340] (PostNet: Posterior Network), Guénais et al. [341] (BaCOUn: Bayesian Classifier with OOD Uncertainty), Li et al. [342] (DNN-MFBO: DNN Multi-Fidelity BO), Lyle et al. [343], Lee and Seok. [344] (cGAN: Conditional GAN), Fan et al. [345] (BAM: Bayesian Attention Module), Chauhan et al. [346] (Neural Heteroscedastic Regression and MC dropout), Zhou et al. [347] (Collaborating Network), Chan et al. [348] (Transductive dropout), Wang and Zhou [349] (Thompson sampling via local uncertainty), Joo et al. [350] (Belief matching framework), Wang and Van Hoof [351] (DSVNP: Doubly Stochastic Variational Neural Process), Hortúa et al. [352] (NF: Normalizing Flows), Lyu et al. [353] (MC-BLEU-VAR: MC dropout BLEU Score Variance), Notin et al. [354] (MI), Jarvenpaa et al. [355] (BABC: Bayesian Approximate Bayesian Computation), Huggins et al. [356] (Validated VI via PPEB (Practical Posterior Error Bounds)), Boluki et al. [357] (Learnable Bernoulli dropout), Barbano et al. [358] (BDGD: Bayesian Deep Gradient Descent), Wenzel et al. [359], Suzuki et al. [360] (Multi-objective BO using Pareto-frontier entropy)
Ensemble	Zhang et al. [273], Buckman et al. [361] (STEVE: Stochastic ensemble value expansion), Chang et al. [281], He et al. [362] (BDE: Bayesian Deep Ensembles), Schwab et al. [363], Smith et al. [364], Malinin and Gales [365], Jain et al. [366], Valdenegro-Toro [367], Juraska et al. [368], Oh et al. [369], Brown et al. [370], Salem et al. [371], Wen et al. [372], Wenzel et al. [373] (HDE: hyper-deep ensembles), Wang et al. [374] (DynSnap: Dynamic snapshot ensemble), Grönquist et al. [375], Lu et al. [376] (Ensemble GP with spectral features), Duan et al. [377] (NGBoost: Natural Gradient Boosting)
Others	Jiang et al. [2] (Trust score), Qin et al. [378] (infoCAM: informative class activation map), Wu et al. [379] (Deep Dirichlet mixture networks), Qian et al. [380] (Margin preserving metric learning), Gomez et al. [381] (Targeted dropout), Malinin and Gales [382] (Prior networks), Dunlop et al.et al. [383] (DGP: deep GP), Hendrycks et al. [384] (Self-supervision), Kumar et al. [385] (Scaling-binning calibrator), [386] (AugMix as a data processing approach), Możejko et al. [387] (Softmax output), Xie et al. [388] (DiffChaser), Boiarov et al. [389] (SPSA: Simultaneous Perturbation Stochastic Approximation), Ye et al. [390] (Lasso bootstrap), Monteiro et al. [391] (SSN: Stochastic Segmentation Network), Maggi et al. [392] (Superposition semantics), Amiri et al [393] (LCORPP: learning-commonsense reasoning and probabilistic planning), Sensoy et al. [394] (GEN: Generative Evidential Neural Network), Belakaria, et al. [395] (USeMO: UA Search framework for optimizing Multiple Objectives), Liu et al. [396] (UaGGP: UA Graph GP), Northcutt et al. [397] (Confident learning), Manders et al. [398] (Class Prediction Uncertainty Alignment), Chun et al. [399] (Regularization Method), Mehta et al. [400] (Uncertainty metric), Liu et al. [401] (SNGP: Spectral-normalized Neural GP), Scillitoe et al. [402] (MF's: Mondrian forests), Ovadia et al. [403] (Dataset shift), Biloš et al. [404] (FD-Dir (Function Decomposition-Dirichlet) and WGP-LN (Weighted GP-Logistic-Normal)), Zheng and Yang [405] (MR: memory regularization), Zelikman et al. [406] (CRUDE: Calibrating Regression Uncertainty Distributions Empirically), Da Silva et al. [407] (RCMP: Requesting Confidence-Moderated Policy advice), Thiagarajan et al. [408] (Uncertainty matching), Zhou et al. [409] (POMBU: Policy Optimization method with Model-Based Uncertainty), Standvoss et al. [410] (RGNN: recurrent generative NN), Wang et al. [411] (TransCal: Transferable Calibration), Grover and Ermon [412] (UAE: uncertainty autoencoders), Cakir et al. [413], [414] (MI), Yıldız et al. [415] (ODE^2VAE: Ordinary Differential Equation VAE), Titsias, Michalis et al. [416] and Lee et al. [417] (GP), Ravi and Beatson [418] (AVI: Amortized VI), Lu et al. [419] (DGPM: DGP with Moments), Wang et al. [420] (NLE loss: negative log-likelihood error), Tai et al. [421] (UIA: UA imitation learning), Selvan et al. [422] (cFlow: conditional Normalizing Flow), Poggi et al. [423] (Self-Teaching), Cui et al. [424] (MMD: Maximum Mean Discrepancy), Lindinger et al. [425] (SDGP: Structured DGP), Wang et al. [426] (A non-asymptotic confidence set for the MLE (maximum likelihood estimate)), Meronen et al. [427] (Matérn activation in GP), Rudner et al. [428] (DREG-SNR: Doubly-reparameterized Gradient Estimates-Signal-to-Noise Ratio), Zhao and Udell [429] (LRGC: Low Rank Gaussian Copula), Shi et al. [430] (Multi-source uncertainty aware ADL), Kopetzki et al. [431] (Certifiable robustnes and uncertainty attacks for DBU Dirichlet-based uncertainty), Chung et al. [432] (MAQR: Model Agnostic Quantile Regression), Finzi et al. [433] (PNCNN: Probabilistic Numeric CNN), Alaa and Van Der Schaar [434] (BJ: Blockwise infinitesimal Jackknife), Liu et al. [435] (Energy-based OoD), Aushev et al. [436] (Likelihood-Free inference with DGP), Antorán et al. [437] (MLL: Marginal Log Likelihood), Huo et al. [438] (MEL: maximum entropy learning), Bondesan and Welling [439] (Quantum deformed layer), Ardywibowo et al. [440] (Switching GP), Sadeghi et al. [441] (SPGD (Stochastic proximal gradient descent) and SPGDA (SPGD-ascent)), Vadera et al. [442] (URSABench: Uncertainty, Robustness, Scalability and Accuracy Benchmark), Cai et al. [443] (Entrywise confidence intervals)
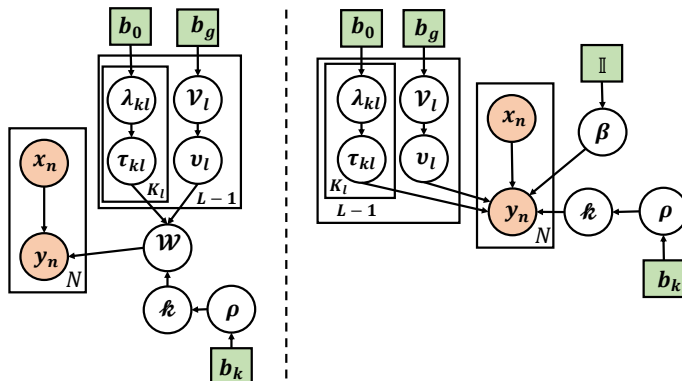
Fig. 22: A summary of graphical models for the conditional dependencies of BNNs with Horseshoe priors which is reproduced based on [445]. Note, the left part of the image is the centered parameterization and the right part is the on-centered parameterization.

approach. Farquhar et al. [448] introduced a variational approximate posterior for BNNs termed as Radial BNNs which scales well to large models. Radial BNNs maintained full support: avoiding the a priori implausibility of discrete support and letting them acted as a prior for CL. Their technique evaded a sampling issue in mean-field variational inference (MFVI) occurred due to 'soap-bubble' pathology of multivariate Gaussians. They demonstrated that Radial BNNs are robust to hyperparameters unlike MFVI and proved its efficacy in real world tasks without needing intensive tuning and ad-hoc tweaks.

Novak et al. [449] proposed neural tangents; a library intended to facilitate research into infinite-width neural networks. It permitted a high-level API for denoting hierarchical and complex neural network architectures. These networks could be trained and estimated either in their infinite-width or at finite-width limit as usual. Infinite-width networks can be trained analytically using gradient descent or using exact Bayesian inference via the Neural Tangent Kernel. All computations were distributed automatically over several accelerators with near-linear scaling in the different devices. Yıldız et al. [415] proposed Ordinary Differential Equation Variational Auto-Encoder (ODE^2VAE), a latent second order ODE model for high-dimensional sequential data. Their model could concurrently learn the embedding of high dimensional trajectories and deduce capriciously complex continuous-time latent dynamics. Their approach explicitly decomposed the latent space into position and momentum components and cracked a second order ODE system that was contrary to RNN based time series approaches and recently presented black-box ODE methods. They further presented probabilistic latent ODE dynamics parameterized by deep BNN for tackling uncertainty. They tested their method on bouncing balls, image rotation and motion capture datasets.

Liu et al. [450] proposed a novel technique to train a robust neural network against adversarial attacks. They observed that although fusing randomness could enhance the robustness of neural networks, incorporating noise blindly to all the layers was not the optimal way to add randomness. They formally learnt the posterior distribution of models in a scalable way by modelling randomness under the framework of BNN. They devised the mini-max problem in BNN to learn the best model distribution under adversarial attacks. Their method yielded state-of-the-art performance under strong attacks. As mentioned earlier, DNNs have yielded outstanding performances in several noteworthy domains including autonomous driving, security and medical diagnosis. In these domains, safety is very crucial, hence knowing what DNNs do not know is highly desirable. Most BNNs are trained by minimizing a suitable ELBO on a variational approximation or sampled through MC methods because of the intractability of the resulting optimization problem. Pomponi et al. [451] devised a variant of former and replaced KL divergence in the ELBO term with a Maximum Mean Discrepancy (MMD) estimator. Their method based on the properties of the MMD exhibited numerous advantages including robustness to the choice of a prior over the weights, better calibrated and higher accuracy. They estimated the uncertainty as well as it performed in a robust manner against the injection of noise and adversarial attacks over their inputs. BNNs show promising results in enhancing UQ and robustness of modern DL methods. However, they have problem with parameter efficiency and underfitting at scale. On the contrary, deep ensembles outperformed BNNs on certain problems, but they also have efficiency issues. The strengths of these two approaches need to be combined to remedy their common problems. Dusenberry et al. [294] devised a rank-1 parameterization of BNNs and also utilized mixture approximate posteriors to capture multiple modes. Rank-1 BNNs demonstrated state-of-the-art performance across OoD variants, calibration on the test sets, accuracy and log-likelihood.

Indeed, there are different types of uncertainties in machine and deep learning domains which are needed to be handled in different ways. Harang et al. [452] investigated three types of uncertainties- open set uncertainty, intrinsic data uncertainty, and model capacity uncertainty and review methods to address each one. They proposed a unified hierarchical technique that integrated techniques from invertible latent density inference, Bayesian inference and discriminative classification in a single end-to-end DNN topology to demonstrate competent per-sample uncertainty estimation in a detection context. Their method could accommodate base/prior rates for binary detection and addressed all three uncertainties. In addition, it is critical for the safety of using an AI application to know the reliability of different classification accuracies. The standard procedure to access it is to use the confidence score of the classifier. Jiang et al. [2] presented a novel score dubbed as trust score that estimated the agreement between a modified nearest-neighbor classifier and classifier on the testing example. They empirically demonstrated that high trust score exhibited high precision at recognizing correctly classified examples, outperforming the classifier's confidence score and other baselines as well. In another work, Rohekar et al. [453] introduced a technique that covers both model selection and model averaging in the same framework. Their technique combined bootstrap with constraint-based learning to tackle prime limitation of constraint-based learning—sensitivity to errors in the independence tests. They formulated an algorithm for learning a tree, in which each node denoted a scored CPDAG for a subset of variables and the level of the node correspond

to the maximal order of conditional independencies that were encoded in the graph. Greater computational efficiency was guaranteed by reusing stable low order independencies. Their algorithm learnt better MAP models, scaled well to hundreds of variables and more reliable causal relationships between variables.

The Bayesian probabilistic framework presents an ethical way to perform model comparison and derive meaningful metrics for guiding decisions. However, many models are intractable with standard Bayesian methods, as their likelihood is computationally too expensive to evaluate or lack a closed-form likelihood function. Radev et al. [454] presented a new approach for performing Bayesian model comparison using specialized deep learning architectures. They also introduced a new way to measure epistemic uncertainty in model comparison problems. They argued that their measure of epistemic uncertainty offers a distinctive proxy to quantify absolute evidence even in a framework which believed that the true data-generating model was within a finite set of candidate models. In another study, Belakaria et al. [395] tackled the issue of multi-objective (MO) blackbox utilizing expensive function evaluations, where the goal was to approximate the true Pareto set of solutions while reducing the number of function evaluations. They introduced a new UA search framework called as USeMO to select efficiently the sequence of inputs for assessment to crack this issue. The selection process of USeMO comprised of cracking a cheap MO optimization problem via surrogate models of the true functions to recognize the most potential candidates and choosing the best candidate based on a measure of uncertainty. They also presented theoretical analysis to characterize the efficiency of their method.

Extensive tuning of hyperparameters are required in many machine learning models to perform well. BO and variety of other methods are utilized to expedite and automate this process. As tuning usually requires repeatedly fully training models, hence it remains tremendously costly. Ariafar et al. [455] introduced to hasten the Bayesian optimization (BO) method by applying the relative amount of information supplied by each training example. They leveraged importance sampling (IS) to do so. That enhanced the quality of the black-box function evaluations and their run-time and hence must be executed carefully. Indeed, NNs with binary weights are hardware-friendly and computation-efficient, but as it involves a discrete optimization problem, their training is challenging. Applying gradient-based methods, such as Straight-Through Estimator and ignoring the discrete nature of the problem surprisingly works well in practice. Meng et al. [456] presented a principled approach that justified such methods applying Bayesian learning rule. The rule resulted in an algorithm when applied to compute a Bernoulli distribution over the binary weights. The algorithm enabled uncertainty estimation for CL to avoid catastrophic forgetting and also achieved state-of-the-art performance.

UQ methods have also been used in semi-supervised learning, zero-shot learning as well as meta-learning. Semi-supervised learning models, such as co-training, could present a powerful method to influence unlabeled data. Xia et al. [457] introduced a new approach, UMCT (UA multi-view co-training), to address semi-supervised learning on

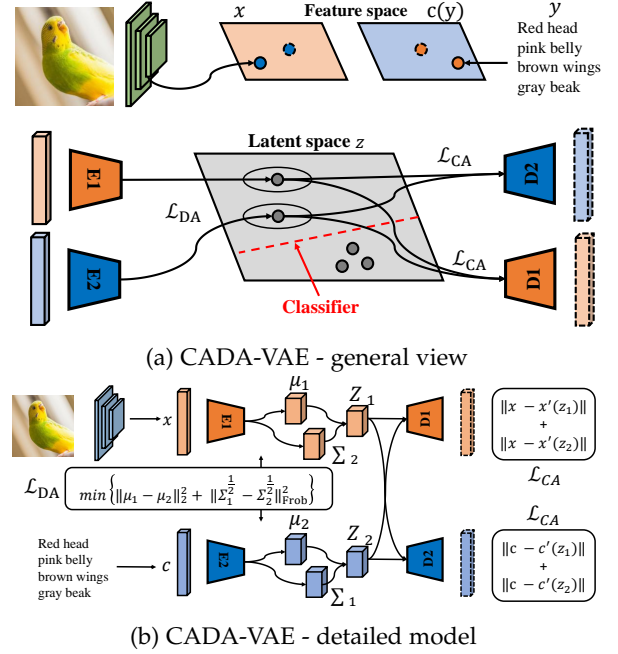(a) CADA-VAE - general view

(b) CADA-VAE - detailed model

Fig. 23: A schematic view of CADA-VAE (Cross- and Distribution Aligned VAE) which is reproduced based on [458].

3D data, such as volumetric data from medical imaging. Co-training was attained by exploring multi-viewpoint consistency of 3D data. They produced different views by permuting or rotating the 3D data and used asymmetrical 3D kernels to support diversified features in different sub-networks. Additionally, they presented an uncertainty-weighted label fusion technique to measure the reliability of each view's prediction with BDL. On the other hand, many models in generalized zero-shot learning depend on cross-modal mapping between the class embedding space and the image feature space. Schönfeld et al. [458] devised an approach where class embeddings and a shared latent space of image features were learned by modality-specific aligned VAE (named CADA-VAE). The key to their model is that they aligned the distributions learned from images and from side-information to produce latent features that contained the essential multi-modal information associated with unseen classes. They examined their learned latent features on several benchmark datasets and confirmed a novel state-of-the-art on generalized on few-shot or zero-shot learning. The general view and detailed CADA-VAE model is illustrated by Fig. 23.

Meta-learning models [459] are subject to overfitting when there are no enough training tasks for the metalearners to generalize. Tseng et al. [460] proposed an effective and simple approach to mitigate the risk of overfitting for gradient-based meta-learning. They randomly dropped the gradient in the inner-loop optimization during the gradient-based adaptation stage such that the augmented gradients enhanced generalization to novel tasks. They proposed a general form of the presented gradient dropout regularization and demonstrated that this term could be sampled from either the Gaussian or Bernoulli distribution. They empirically yielded that the gradient dropout regularization alleviated the overfitting issue and enhanced the performance on different gradient-based meta-learning frameworks.

There are some more UQ methods which have been proposed in the literature. For example, Variational Bayes (VB) is computationally efficient, theoretically grounded and generally applicable among methods to realize probabilistic inference in DNNs. Wu et al. [461] devised two methods to turn VB into a robust inference tool for BNNs. First method presented a new deterministic method to approximate moments in neural networks, got rid of gradient variance. Second method proposed a hierarchical prior for parameters and a new Empirical Bayes technique for automated selection of prior variances. The resulting method combining these two methods is very robust and efficient. Another research direction is related to GP models for using in BNNs. A flexible and simple technique to generating expressive priors in GP models produces new kernels from a combination of basic kernels. Despite the link between BNNs and GPs, the BNN analogue of this has not yet been investigated. Pearce et al. [462] explored BNN architectures mirroring such kernel combinations. They showed further how BNNs could generate periodic kernels that were often helpful in this context. Their empirical experiments demonstrated the practical value of these ideas in reinforcement and supervised settings. On the other hand, two prime obstacles in adoption of variational BNN are the high parameter overhead and difficulty of implementation that occurred due to "programming overhead". MC dropout tackles these obstacles well, but has limitation in model performance when applied in networks with batch normalization layers. Chang et al. [281] designed a general variational family for ensemble-based BNN that included dropout as a special case. They further proposed two members of the family that worked well with batch normalization layers while preserving the advantages of low parameter and programming overhead.

Bayesian inference facilitates a general framework for incorporating specific properties or prior knowledge into machine learning techniques through selecting a prior distribution carefully. Atanov et al. [463] presented a novel type of prior distributions for CNN, deep weight prior (DWP), that examined generative models to persuade a certain structure of trained convolutional filters. They devised a technique for VI with implicit priors and denoted DWP in a form of an implicit distribution. The experimental results empirically showed that DWP enhanced the performance of BNN when training data is small and initialization of weights with samples from DWP hastened training of CNN. The catastrophic forgetting is an unavoidable issue in CL models for dynamic environments. Li et al. [464] introduced a technique termed as Continual Bayesian Learning Networks (CBLN) to address this problem that facilitates the networks to distribute supplementary resources to acclimatize to new tasks without forgetting the formerly learned tasks. CBLN preserved a mixture of Gaussian posterior distributions that are combined with diverse tasks utilizing a BNN. The presented technique did not require accessing the past training data and could select proper weights to classify the data points during the test time automatically based on an uncertainty criterion. Along with all listed UQ methods, there are some other effective UQ methods which we would include them here as well. Alaa and van der Schaar [465] developed the discriminative jackknife (DJ)

procedure. The proposed DJ approach is flexible procedure which is usable to a wide range of DL methods.

Shekhovtsov et al. [466] exploited the cause of enhanced the generalization performance of deep networks due to Batch Normalization (BN). They argued that randomness of batch statistics was one of the prime reasons. The randomness emerged in the parameters rather than in activations and declared an explanation as a handy Bayesian learning. They utilized the idea to other deterministic normalization methods that were ignorant of the batch size. One of the prime drawbacks of NN-based UQ is the high memory requirement during training; which hinders their application to processing shallower architectures and/or smaller field-of-views (FOVs). Gantenbein et al. [467] examined the efficacy of applying reversible blocks for constructing memory-efficient NN for quantification of segmentation uncertainty. The reversible architecture yielded memory saving by precisely computing the activations from the outputs of the successive layers during backpropagation instead of accumulating the activations for each layer. They incorporated the reversible blocks into an architecture termed as PHiSeg that was devised for UQ in medical image segmentation. The reversible architecture, RevPHiSeg, permitted training NNs for quantifying segmentation uncertainty on GPUs with restricted memory and processing larger FOVs.

The authors in [10] reengineered DeepLab-v3+, to produce its Bayesian counterpart applying Concrete dropout (CD) and MC dropout inference techniques. The UQ and surrogate modeling job for PDE systems are considered as supervised learning problems in most of the circumstances where output and input pairs are used for training [468]. Yang et al. [469] designed a framework for UQ for image registration using a low-rank Hessian approximation, a pathological image registration using an image synthesis deep network and predicted registration parameter with a patch-based deep learning approach by applying image appearances only. Their network predicted the initial momentum for the Deformation Diffeomorphic Metric Mapping (LDDMM) model for both multi-modal and uni-modal registration problems. The researchers in [470] proposed a BDL model that integrated epistemic uncertainty with input-dependent aleatoric uncertainty. Their explicit uncertainty formulation produced a novel loss functions that could be interpreted as learned attenuation.

Zhu et al. [468] devised a physics-constrained deep learning model for high-dimensional UQ and surrogate modeling without labeled data. The ensembles of DNNs that belong to the mixture models class can be employed to quantify the prediction uncertainty [471]. Vishnu et al. [218] presented deep learning framework for prognostics with UQ that were helpful in conditions where (i) inherent noise was there in the sensor readings, (ii) operational conditions for future were not estimated, and (iii) labeled failure data was rare because of scarcity of failures. Constructing a Gaussian distribution over the weights, and sample it to produce a distribution over the categorical output are employed for approximation of distributions over the output of classification neural networks in BDL, however, the process is costly [472]. Begoli et al. [13] presented the challenges for the adoption of UQ due to the absence of sound underlying theory and new research opportunities for the advancement

of theoretical methods and practical approaches of UQ in the area of AI assisted medical decision making.

Kendall et al. [473] proposed a real-time robust six-degree of freedom monocular visual relocalization system by applying a Bayesian convolutional neural network to single RGB images to regress the 6 degrees of freedom (6-DOF) camera pose. They obtained an estimation of relocalization uncertainty of their system and enhanced the state-of-the-art localization accuracy on an outdoor database of large scale in nature. Semantic segmentation with BDL has been revolutionary to attain uncertainty maps from deep models in semantic class prediction [10]. To estimate whether a BDL model records an improved uncertainty estimates than another model, we need new metrics. The authors in [471] utilized both variational Bayesian inference and maximum likelihood to train compound density networks, and demonstrated that they can obtain competitive uncertainty estimates on OoD data which are robust in terms of adversarial examples. To guarantee high operational availability of equipment and condition-based maintenance, multi-sensor time series data extracted from Remaining Useful Life (RUL) or prognostics estimation are crucial. Hobbhahn et al. [472] utilized the Dirichlet approximation to devise a lightweight uncertainty-aware output ranking for the setup of ImageNet. In this regard, they used the $LaplaceBridge$ to map a Gaussian distribution onto a Dirichlet one. Bayesian neural networks (BNNs) do not scale well to computer vision jobs as it is tricky to train and show meager generalization under dataset-shift [219]. This generates the need of effective ensembles that can generalize and produce trustworthy uncertainty estimates. The authors [219] obtained diversity in the output predictions used for multi-modal data modeling by optimizing the diversity inducing adversarial loss for learning latent variables. The novel BDL tools make it possible to model epistemic uncertainty in computer vision [470].

8 APPLICATIONS

In this section, we discuss few most important applications of different UQ techniques used in machine and deep learning methods. In this regard, we first summarise the application of UQ techniques in image processing and computer vision followed by medical image analysis. Afterwards, we show how UQ has been applied to Natural Language Processing (NLP) and some more applications of UQ techniques in the following.

8.1 Image Processing and Computer Vision

Nowadays, deep learning algorithms are being vastly used to map high dimensional data to output arrays, while these mappings can be inaccurate in many cases, e.g. the wrong identification of two African Americans as gorillas in an image classification system [474], has lead to racial discrimination [470]. Therefore, it's important to take uncertainty into account, where the predictions made by deep learning based computer vision algorithms. To date, there have been number of studies addressed uncertainty in deep learning algorithms for various applications including but not limited to image/video retrieval [475], [476], depth estimation [477], [478], object detection [479], [480], [481], semantic segmentation and scene understanding [482], [483], [484], [485], [10], optical flow estimation and motion prediction [269], [486], [487], human pose estimation and pedestrian localization [488], [489], [326], person re-identification and face recognition [490], [491], [492], camera re-localization [473], avoiding adversarial attacks [493], [494], during the years 2016 to 2020.

As a fact, most of research studies in deep learning applications are concentrating on prediction accuracy. Unlike those studies, untangling the complexity of various DNNs and addressing uncertainty for a variety of computer vision tasks has attracted significant interest [495]. There still has been a good record of using BNNs and MC dropout for uncertainty estimation on use of deep learning architectures. Nine studies have reported MC dropout as the most effective uncertainty quantity technique [473], [48], [269], [496], [470], [489], [10], [487], [494], applicable on various deep learning architectures. Kendall et al showed that uncertainty of their Bayesian convolutional neural networks model came from appearance and pose dissimilarity of images to the training set and could estimate the model's re-localization uncertainty, which improved localization accuracy on a large outdoor dataset [473]. Same authors have developed a measure of model uncertainty by MC sampling with dropout and enhanced the semantic segmentation performance compared to the state-of-the-art methods in 2016 [48]. Eldesokey et al. [497] proposed an UA model for CNNs and tested on the KITTI dataset. The proposed model identified disturbed measurements of the input data after learning an input confidence estimator in a self-supervised procedure using the normalized CNNs (NCNNs).

Indeed, epistemic uncertainty estimation is a challenging problem, and while several scalable techniques lately have appeared, no widespread assessment has been carried out in a real-world setting. Gustafsson et al. [498] devised a comprehensive assessment framework for scalable epistemic uncertainty estimation techniques in deep learning. Their framework tested for the robustness needed in real-world computer vision applications. They also utilized their framework to compare conclusively and extensively two scalable techniques: MC-dropout and ensembling. Postels et al. [485] proposed a sampling-free method for estimating the epistemic uncertainty of a neural network. Epistemic uncertainty is crucial in safety-critical applications, since it denotes the reliability of predictions using new data. Their prime contribution was the approximation of epistemic uncertainty estimated by these techniques which did not necessitate sampling, thus remarkably mitigating the computational overhead. They used their method to volumetric visual jobs to showcase the advantages of their techniques in terms of computational overhead as well as uncertainty estimates. Cai et al. [499] worked on the hand segmentation generalization issue without using segmentation labels in the target domain. They designed a Bayesian CNN-based model adaptation approach for hand segmentation, which devised and considered two vital factors: 1) general information about hand shapes shared across domains and 2) prediction uncertainty when the model was used in a new domain. Accordingly, they introduced iterative self-training strategy hand segmentation in the novel domain,

which was directed by the model uncertainty approximated by a Bayesian CNN. But Bayesian techniques have not been exploited extensively for 3D modalities such as point clouds often utilized for autonomous systems and robots. Bhandary et al. [500] examined three uncertainty quantification techniques viz. MC-DropConnect, MC dropout and deep ensemble on the DarkNet21Seg 3D semantic segmentation model and analyzed the impact of different parameters such as drop probability values on task performance, number of models in ensembles or forward passes and uncertainty estimate quality. They demonstrated that deep ensembles generated better results than other methods in terms of uncertainty metrics and performance.

Weakly-supervised semantic segmentation using image-level labels is accomplished by acquiring object response maps. However, prevailing techniques depend on the classifier that can result in a response map only attending on discriminative object regions as the network does not require seeing the complete object for optimizing the classification loss. Chang et al. [501] introduced a principled and end-to-end trainable approach to let the network paying attention to other parts of the object, while generating a more uniform and complete response map. They proposed specifically Mixup data augmentation strategy into the classification network and devised two uncertainty regularization terms to better act together with the Mixup scheme. More information regarding different UQ methods applied in computer vision and image processing tasks is illustrated in Table 7.

8.2 Medical Applications

An automated analysis of medical image has come into existence as soon as it was possible load and scan medical images into the computer [502]. At the outset, from the 1970s to the 1990s, medical image analysis was done with sequential application of low-level pixel processing (region growing, edge and line detector filters) and mathematical modeling (fitting lines, ellipses and circles) to build compound rule-based systems that solved particular tasks. It is analogous to expert systems with many if-then-else statements that were popular in artificial intelligence in the same period. At the end of the 1990s, supervised methods, where training samples are used to develop a system, is becoming gradually more popular in medical image analysis. Examples include active shape models, atlas, concept of feature extraction and use of statistical classifiers. This machine learning approach is still very popular and forms the foundation for various booming commercially available medical image analysis systems. Hence, there is a shift from systems that are entirely devised by humans to systems that are trained by computers utilizing example data from which feature vectors are derived. Computer algorithms establish the optimal decision boundary in the high-dimensional feature space. Both monetary and ethical costs of erroneous predictions can be noteworthy in medicine, and the complexity of the issue imposes progressively more complex models. Although DL methods have achieved outstanding performances in medical image analysis, most of them have not been employed in extremely automated disease monitoring systems due to lack of reliability of the model [503]. For example, Dusenberry et al. [504] studied the role of model uncertainty strategies in the medical domain. They demonstrated

that population-level metrics, such as calibration error, log-likelihood, AUC-ROC and AUC-PR did not capture model uncertainty and was shown by applying RNN ensembles and different BRNNs. They showcased that RNNs with only Bayesian embeddings could be a competent way to tackle model uncertainty compared to ensembles.

As we all know, medical well-annotated data is extremely expensive for conducting medical image segmentation. However, unlabeled data are very appropriate solution which can be used both in semi-supervised and unsupervised learning domains. As discussed earlier, Xia et al. [505] introduced the UMCT model as a semi-supervised framework and tested it on various medical image datasets. They extended the Dice loss for ULF (uncertainty-weighted label fusion) as follows:

$$\mathcal{L}_{Dice} = \frac{1}{D} \Sigma_{d=0}^{D} \frac{2\Sigma_{i=1}^{N} y_i^d \hat{y}_i^d}{\Sigma_{i=1}^{N}(y_i^d)^2 + \Sigma_{i=1}^{N}(\hat{y}_i^d)^2}, \qquad (39)$$

According to the obtained results, the proposed UMCT method outperformed the other applied methods to the same datasets. As a result, they concluded that having a proper uncertainty method can assist having a better medical image analysis performance.

Blood oxygen saturation (sO_2) measurement by optical imaging oximetry offers insight into local tissue metabolism and functions. Traditional methods for quantifying sO_2 suffer from uncertainties due to variations in the experimental conditions, systemic spectral bias, light spectral bias, tissue geometry and biological variability. Liu et al. [506] devised deep spectral learning (DSL), a novel data-driven approach to yield oximetry that was robust to experimental variations and also facilitated uncertainty quantification for each sO_2 prediction. Predictions calculated by DSL were highly adaptive to the depth-dependent backscattering spectra as well as to experimental variabilities. The DSL-predicted sO_2 demonstrated notably lower mean-square errors than those of the traditional least-squares fitting method. Inherent ambiguities cause many real-world vision problems. It is difficult to access for example which region contains cancerous tissue from a CT scan in clinical applications. Kohl et al. [507] devised a generative segmentation model based on the combination of a U-Net with a conditional VAE which is capable of generating large number of plausible hypotheses. They exhibited that on a Cityscapes segmentation task and a lung abnormalities segmentation task approach regenerated all the possible segmentation variants as well as the frequencies with which they outperformed the existing methods.

In another research, Araújo et al. [508] proposed an uncertainty-aware deep learning model (named DR—GRADUATE) for grading diabetic retinopathy (DR) using eye fundus images. In this regard, they introduced a new Gaussian-sampling technique on a Multiple Instance Learning (MIL) framework and used the proposed system as a second-opinion DR diagnostic system.

UQ methods have also been used in prostate cancer domain. Karimi et al. [509] studied the prostate cancer using ultrasound images. In this regard, they proposed a robust and accurate deep learning (CNN) segmentation model. Moreover, due to the importance of uncertainty in medical

image analysis, they computed the uncertainty as follows:

$$Q = 1 - \bar{p}^2 - (1 - \bar{p})^2, \tag{40}$$

where $\bar{p}$ is the average of the applied probability maps. The obtained results confirmed that adding uncertainty resulted to having better prostate cancer segmentation outcomes. As discussed above, the MC dropout has demonstrated impressive performance for quantifying uncertainty in deep learning methods. Combalia et al. [510] applied the MC dropout in DNNs for UQ of dermoscopic (skin lesion) image classification. Their results indicated that using different uncertainty metrics are appropriate solution to explore difficult and OoD samples.

The cardiovascular disease detection by machine and deep learning is another research topic for application of UQ methods. 2D echocardiography is a widespread imaging modality for cardiovascular diseases. Deep learning techniques have widely been used in 2D echocardiography for structural and functional assessment and automated view classification. Most of the models do not estimate uncertainty in this regard which is very crucial. Dahal et al. [511] compared three ensemble based uncertainty techniques utilizing four different metrics to achieve an insight of uncertainty modeling for left ventricular segmentation from Ultrasound (US) images. They further showed how uncertainty estimation could be utilized to reject inferior quality images and hence enhanced the segmentation results.

Registration is a basic task in medical image analysis which can be used in numerous tasks including motion analysis, multi-modal image alignment, intra-operative tracking and image segmentation. Zhu et al. [512] proposed a neural registration framework (NeurReg) with a hybrid loss of displacement fields and data similarity, which considerably enhanced the existing state-of-the-art of registrations. They simulated different transformations by a registration simulator which created fixed image and displacement field ground truth for training. They devised three segmentation frameworks based on the proposed registration framework: 1) MTL with atlas-based segmentation as an intermediate feature, 2) joint learning of both registration and segmentation tasks, and 3) atlas-based segmentation. Different probable ailments can be detected by accurate and automatic segmentation of anatomical structures on medical images. Bian et al. [513] introduced an uncertainty-aware domain alignment approach to tackle the domain shift issue in the cross-domain UDA (Unsupervised Domain Adaptation) task. Domain shift is an issue related to performance of the segmentation of various deep neural networks and segmentation task may deteriorate several devices or modalities due to the noteworthy dissimilarity across the domains. In this regard, they devised specifically an UESM (Uncertainty Estimation and Segmentation Module) to attain the uncertainty map estimation. Then, they proposed a new UCE (Uncertainty-aware Cross Entropy) loss to leverage the uncertainty information to enhance the performance of segmentation on extremely uncertain regions. The optimal target samples by uncertainty guidance were selected by an UST (Uncertainty-aware Self- Training) method to further boost the performance in the UDA task.

Kohl et al. [514] devised a segmentation network with a conditional variational auto-encoder (cVAE) termed it as Hierarchical Probabilistic U-Net that applied a hierarchical latent space decomposition. They demonstrated that their model formulation permitted reconstruction and sampling of segmentations with high fidelity while providing the flexibility to learn complex structured distributions across scales. Their model split automatically an inductive bias that they estimated useful in structured output prediction tasks beyond segmentation. In another research, Yin et al. [515] stated that uncertainty related to FFR (fractional flow reserve) of coronary artery disease (CAD) in few properties such as anatomic and physiologic is common. For this reason, they proposed a predictive probabilistic model for FFR using the BO approach. The obtained outcomes clearly acknowledge the importance of dealing with uncertainty in the diagnosis of CAD.

Li et al. [516] exploited uncertainty calibration within an AL framework for medical image segmentation. Uncertainty estimation is specifically crucial in the data-driven AL setting where the goal is to attain definite accuracy with least labeling effort. The model learns to choose the most enlightening unlabeled samples for annotation derived from its estimated uncertainty. Different acquisition strategies and uncertainty estimation techniques were explored. They argued that choosing regions to annotate instead of full images led to more well-calibrated models. We provide further information about UQ methods applied in different medical application tasks in Table 8.

8.3 Natural Language Processing and Text Mining

Natural language processing (NLP) focuses on understanding, analysing and generating languages that humans utilize naturally [517]. In recent years, significant and practical real-world problems have been addressed and large-scale systems are also deployed in this research domain. Novel machine and deep learning approaches such as continuous space methods and DNNs have inferred language patterns from the huge data of real world and make accurate predictions about the new data. One noteworthy challenge is to describe a language in a form that can be effectively processed by a learning system. NLP is an interdisciplinary field between linguistics and artificial intelligence [518]. One of the most broadly studied areas of NLP is text mining (TM) that collects vital information from free (unstructured) texts. In this way, new knowledge can be extracted from a huge amount of texts. But the acquisition of reliable information from texts is not straightforward because of human linguistic ability of speaking about non-existing and non-realistic things or events. There are some propositions whose truth value cannot be unambiguously determined as these propositions are uncertain and they may be false true in some possible worlds but may be true in other ones. Uncertainty is a significant linguistic incident that is pertinent in many fields of language processing. In most general case, it can be termed as lack of information as the reader or listener is uncertain for a piece of information. Hence, uncertain propositions are those whose reliability or truth value cannot be determined due to lack of information. Distinguishing between uncertain and factual (i.e. true or false) propositions is of prime importance both in natural

language processing and linguistics applications. It is essential to recognize linguistic cues of uncertainty since the target and source language may differ in their framework to express uncertainty in machine translation. In clinical document classification, medical reports can be grouped depending on whether the patient probability suffers, does not suffer or suffers from an ailment. There are several different NLP applications that try to investigate uncertainty in natural language texts in a couple of domains (e.g. news or biomedical texts). Most of these approaches use annotated databases for assessment. Various uncertainty corpora like the CoNLL-2010 Shared Task, FactBank, Genia and BioScope corpora has been produced in the recent years. Comparison of these corpora is not possible for the lack of unified annotation principles. The prevailing uncertainty detectors can hardly be applied across domains, and novel resource creation for each domain is costly and time consuming. Instead, a unified widespread approach is needed that can be adapted to a particular need of each domain without much effort and language independence of the model would also be preferred. But we reported a table which includes a summary of the most important UQ methods applied in NLP domain in Table 9.

8.4   Summary of some applied UQ methods in NLP

In this part of review, we briefly summarize some studies have been conducted on UQ in the domain of NLP. It should be noted that we do not disclose the details of the methods due to page limitations. For this reason, we strongly recommend that if the readers need more information of the proposed UQ methods, they can refer to the main sources.
A high-dimensional hidden layer and a large dictionary size make training of the RNN- language model (RNN-LM) as an ill-posed challenge. Chien et al. [517] proposed a Bayesian approach to regularize the RNN-LM and utilized it for continuous speech recognition. The uncertainty of the estimated model parameters that was presented by a Gaussian prior was compensated by penalizing the too complicated RNN-LM. The Gaussian hyperparameter was estimated by maximizing the marginal likelihood and regularized parameters were computed with reference to maximum a posterior criterion was utilized to construct regularized model. A small set of salient outer-products were selected to devise the Bayesian RNN-LM (BRNN-LM) by developing a rapid approximation to a Hessian matrix. As we know, clinical named entity recognition (NER) is one of the basic tasks for devising clinical NLP systems. Domain experts are required for annotating large amount of samples to achieve good performance by a machine learning (ML) system. This is an expensive exercise. A sample selection technique called active learning (AL) tries to mitigate the annotation cost. Chen et al. [519] introduced and examined both novel and existing AL techniques for a clinical NER job to recognize medical treatments, problems and laboratory tests from the clinical notes. They simulated the AL experiments by applying different novel and prevailing algorithms in three categories including baseline sampling, diversity-based, and uncertainty-based techniques. Based on number of sentences vs. the learning curves of F-measure, uncertainty sampling performed superior to all its

counterparts in terms of the area under the learning curve (ALC) score. Most diversity-based techniques yielded better performance than random sampling in ALC. In another research, Kong et al. [520] introduced a novel theoretical perspective of data noising in RNN language models. They demonstrated that variants of data noising were instances of Bayesian RNN with a particular variational distribution. They presented natural extensions to data noising under the variational framework and a more principled method to apply at prediction time by utilizing this insight. They devised an element-wise variational smoothing technique and variational smoothing with tied input and output embedding matrices. Their model was empirically tested on two language modelling datasets and exhibited superior performance than the prevailing data noising techniques. As we know, factuality is a major concern in many domains especially in social media where informal texts are in abundance. The dependence of the existing methods in social media is on lexical cues where phrases are either omitted from the sentences or is expressed in substandard form. Han et al. [521] introduced ANFU, an Attention-based Neural Framework for Uncertainty identification on social media texts. ANFU incorporated CNN to capture the most vital semantics and attention-based Long Short-Term Memory (LSTM) networks to denote the semantics of words. The experiments were performed on four benchmark datasets (2 English + 2 Chinese). Their proposed ANFU method performed better than any state-of-the-art techniques in terms of F1 score using four social media datasets. Zhang et al. [522] demonstrated that a huge deep learning model could utilize dropout variational inference to predict price movements from limit order books (LOBs), the data source with pricing and trading movements. To enhance profits by avoiding needless trades and position sizing, uncertainty information extracted from posterior predictive distributions could be applied. Their experimental results showcased that Bayesian techniques enhanced predictive performance as stochastic regularisers and uncertainty information could be utilised in trading. In another research, the authors in [523] designed a measure of uncertainty for long sequences of discrete random variables related to the words in the output sentence. This measure took care of epistemic uncertainty similar to the MI that is applied for single discrete random variables such as in classification. Their uncertainty measures cracked a prime intractability in the raw application of prevailing methods on long sentences. They utilized Europarl and WMT 13 for German- English translation task to train a Transformer model with dropout.
As we know, machine translation is a hot topic in neural sequence-to-sequence models. A lack of diversity is observed in the final translations and performance degradation is reported with large beams. The study [524] tried to uncover extrinsic uncertainty caused by noisy training data and related to some of the concerns associated to the inherent uncertainty of the task, due to the existence of numerous valid translations for a single source sentence. They proposed metrics and tools to examine how uncertainty in the data was recorded by the model distribution and the effects of searching techniques in translation. They also presented tools for examining model calibration and some limitations of the current models could be fixed by it.

Accordingly, model calibration in classification is evaluated by Vaicenavicius et al. [525].

The authors in [121] designed a module for rapid experimentation with neural network uncertainty and dubbed it as Bayesian Layers. Neural network libraries with drop-in replacements for common layers were extended by it. These layers recorded activations ("stochastic output layers"), pre-activation units (dropout), uncertainty overweight (Bayesian neural nets) or the function itself (GP). They fused a 5-billion parameter "Bayesian Transformer" on 512 TPUv2 cores for uncertainty in a Bayesian dynamics and machine translation model for model-oriented planning. Bayesian Layers could be utilized for probabilistic programming with stochastic processes used within the Edward2 language. On the other hand, the complexity of machine learning models pose uncontrolled risks, the lack of control and knowledge of the internals of each component applied generate unavoidable effects, such as difficulty in auditability and lack of transparency. Mena et al. [526] presented a wrapper that given a black-box model augmented its output prediction with an assessment of uncertainty. Decision rejection mechanism was employed to lessen the uncertainty or risk. They advocated for a rejection system based on the resulting uncertainty measure that discarded more uncertain predictions but selected more confident predictions; hence improved trustability of the system. They empirically showcased their method in simulated sentiment analysis framework for different domains.

As we know, reliable UQ is a prime step towards devising accountable, transparent, and explainable artificial intelligent systems and BDL plays a crucial role in such quantification. Xiao et al. [527] presented new strategies to examine the data uncertainties and the benefits of characterizing model for NLP tasks. They utilized recurrent and CNN models to experiment empirically on language modelling, named entity recognition, and sentiment analysis to demonstrate that explicitly modelling uncertainties was not only improved model performances but also essential to compute output confidence levels in different NLP tasks. More studies have been conducted on the impact of Bayesian methods in improving the results of deep learning methods in NLP. The authors in [124] investigated a variational Bayes scheme for RNN. At first, they demonstrated that good quality uncertainty estimates and superior regularisation could be adapted by using truncated backpropagation with an extra computational cost during training and also mitigating the number of parameters by 80%. Secondly, they illustrated that the performance of Bayesian RNNs could be enhanced further by employing a new kind of posterior approximation. The current batch statistics could be sharpened by incorporating local gradient information into the approximate posterior. This technique could be utilized broadly in training of BNNs. They empirically yielded that Bayesian RNNs performed better on an image captioning task and a language modelling benchmark than traditional RNNs. The authors in [528] proposed an intelligent framework to enhance en-route flight safety by trajectory prediction where a Bayesian approach was utilized for model prediction uncertainty. Four steps were employed. In the first step, huge raw messages were processed with a distributed computing engine Apache Spark to derive trajectory information efficiently. Two deep learning models were then trained to predict the flight trajectory from different perspectives. The deep learning models were blended together to create multi-fidelity prediction in the third step. Then, the multi-fidelity technique was expanded to multiple flights to examine safety based on vertical and horizontal separation distance between two flights. The blended models showed promising results in en-route safety and flight trajectory prediction.

8.5 Further Applications

In this section, we summarize more applications of various UQ methods. This section tries to cover few most important recent studies.

As mentioned in the previous sections, BDL has been dealing with both epistemic and aleatoric uncertainties in predictions and successful in different domains such as climate change. Vandal et al. [529] devised a discrete-continuous BDL technique with lognormal and Gaussian likelihoods for uncertainty quantification. They presented a super-resolution based DL model dubbed as "DeepSD" for Statistical Downscaling (SD) in climate utilized in precipitation that followed highly skewed distribution. Their discrete-continuous models performed superior to Gaussian distribution with respect to uncertainty calibration and predictive accuracy. As a fact, traditional ANNs lack the capability to model uncertainty and hence not suitable for long-term planning tasks. ANN long-term estimations are deviated from the real behaviour of the system due to approximation errors and process noise. In another research, Nalisnick et al. [530] presented two structured priors—automatic depth determination (ADD) and joint automatic relevance determination (ARD)-ADD—to permit Bayesian reasoning about a neural network's depth. The implementation led to runtime costs or little extra memory to BBB. Future work includes the use of structured variational approximations, comparison against other variational inference strategies and experiments on larger datasets. Dusenberry et al. [504] examined the role of model uncertainty techniques in the medical domain. They demonstrated that population-level metrics, such as calibration error, log-likelihood, AUC-ROC and AUC-PR did not capture model uncertainty by applying different BRNNs and RNN ensembles. The need for estimating model uncertainty was motivated by considerable variability in patient-specific optimal decisions and predictions. They further demonstrated that RNNs with only Bayesian embeddings yielded better results in model uncertainty compared to ensembles.

As we know, new challenges come up in prevailing pixel-based prediction techniques with the advancement of remote sensing imagery. Although deep learning methods achieved a breakthrough in semantic segmentation of high-resolution images, most of the methods yielded predictions with poor boundaries. Bischke et al. [531] proposed a novel cascaded multi-task loss for preserving semantic segmentation boundaries in satellite imagery. Their method outperformed the state-of-the-art techniques by 8.3% without an extra post-processing step. However, in autonomous

driving, object detection plays a crucial role. Localizing the objects and recognize objects perfectly is infeasible due to incomplete data and sensor noise. Hence, the uncertainty associated with the predictions should be computed by the detector. Meyer et al. [532] devised a method that enhanced the learning of probability distribution by taking into account potential noise in the ground-truth labeled data. Their method enhanced not only the object detection performance but also the accuracy of the learned distribution. RNNs have been applied to forecast increasingly complicated systems. Although the RNN literature is highly developed and expansive, UQ is often not taken into account. If considered, then also the uncertainty is usually quantified without the utilization of a rigorous approach. McDermott et al. [533] proposed a Bayesian RNN model for nonlinear spatio-temporal forecasting while quantifying uncertainty in a more formal framework. Unique nature of nonlinear spatio-temporal data was accommodated by modifying the basic RNN. They tested their model with two nonlinear spatio-temporal forecasting frameworks and a Lorenz simulation. On the other hand, RNN language models (RNNLMs) have proved its superiority in several different tasks including speech recognition. Learning appropriate representation of contexts for word prediction can be achieved through the hidden layers of RNNLMs. Fixed hidden vectors and deterministic model parameters in conventional RNNLMs have limitation in modelling the uncertainty over hidden representations. Yu et al. [534] presented a comparative study of hidden and parametric representation uncertainty modelling techniques based variational RNNLMs and Bayesian gates respectively was examined on gated recurrent units (GRU) and LSTM language models. Performance improvements were observed over conventional RNNLMs by their model in terms of word error rate and perplexity.

Predictive accuracy in black-box turbulence models is enhanced by tuning Reynolds-Averaged Stokes (RANS) simulations and applying machine learning algorithms. Geneva et al. [535] presented a new data-driven approach to provide probabilistic bounds for fluid quantities and enhanced RANS predictions. The anisotropic tensor component of Reynolds stress was predicted by using an invariant BDNN. Stein variational gradient decent algorithm was applied to train the model. Based on the proposed method, the associated probabilistic bounds and prediction enhancement of the data-driven model were addressed. Following the research for dealing with uncertainty, we came across the study of Feng et al. [536] which proposed a novel extreme learning machine (ELM) termed as rough ELM (RELM). RELM utilized rough set to divide data into lower approximation set and upper approximation set, and they were used to train lower approximation neurons and upper approximation neurons. RELM showed a comparable accuracy and repeatability in most classification tasks. In another study, Walmsley et al. [537] applied Bayesian CNN and a new generative model of Galaxy Zoo volunteer responses to infer posteriors for the visual morphology of galaxies. The probability of each possible label can be predicted by using Bayesian CNN to learn from galaxy images with uncertain labels. Their posteriors were reliable for practical use as they were well-calibrated. They utilized BALD AL strategy applying their posteriors to request volunteer responses for

the subset of galaxies. They demonstrated that training their Bayesian CNNs utilizing AL needed up to 35-60% fewer labelled galaxies relying on the morphological features.

The distribution of states at execution time may differ from the distribution observed during training makes learning a policy utilizing only observational data a challenging task. Henaff et al. [538] introduced to train a policy by unrolling a learned model of environment dynamics over multiple time steps while explicitly penalizing two costs. The original cost the policy sought to optimize, and an uncertainty cost that represented its divergence from the states it was trained on. They examined their strategy utilizing huge observational dataset of driving behaviour recorded from traffic cameras. In drug discovery, as another application of UQ methods, it is a challenge to predict physical properties and bioactivity of small molecules. Zhang et al. [539] used Bayesian semi-supervised graph convolutional neural networks to achieve UQ and AL. Sampling from the posterior distribution was applied in the Bayesian approach that estimates uncertainty in a statistically principled way. Semi-supervised learning untangled regression and representation learning allowing the model to start AL from a small training data and keeping uncertainty estimates accurate in the low data limit. Their method highlighted the promise of BDL for chemistry.

According to the literature, it is obvious that machine learning has the prospective to present valuable assistance in clinical decision making especially in the Intensive Care Unit (ICU) of a hospital. Traditional machine learning models do not take into account uncertainty in predictions. Ruhe et al. [540] exhibited how the predictive uncertainty and Bayesian modelling could be utilized to recognize out-of-domain examples and reduce the risk of faulty predictions. They utilized BNN to predict risk of mortality of ICU patients. Their empirical results show that uncertainty could detect out-of-domain patients and avert probable errors. Many machine learning techniques need human supervision to yield optimal performance. The quality of manual annotations essentially limited in tasks such as DensePose. Neverova et al. [541] addressed the issue by augmenting neural network predictors with the ability to output a distribution over labels, thus introspectively and explicitly capturing the aleatoric uncertainty in the annotations. A new state-of-the-art accuracy in the benchmark could be achieved by understanding uncertainty better and hence solving the original DensePose task more accurately. The uncertainty estimates produced by multiple models could be used in fusing predictions in a better way to model ensembling that could enhance the accuracy further.

As mentioned earlier, uncertainty estimates in RL tasks and large vision models can be obtained via dropout. A grid-search over the dropout probabilities is essential — an impossible one with RL and a prohibitive operation with large models. Gal et al. [542] devised a novel dropout variant that provided better performance and improved calibrated uncertainties. They used a continuous relaxation of dropout's discrete masks depending on the recent advancements in BDL. They analysed their variant on several tasks and provided insights into usual practice in the area where larger dropout probabilities are often utilized in deeper model layers.

Mobile robots for indoor use depend on 2D laser scanners

for navigation, localization and mapping. These sensors are unable to measure the full occupancy of complex objects and cannot detect transparent surfaces. These estimates are prone to uncertainty and thus make the evaluation of confidence a significant issue for autonomous navigation and mapping. Verdoja et al. [543] proposed a solution to the problem using fully CNN, as another application of UQ methods. They demonstrated that uncertainty over obstacle distances was however better modelled with a Laplace distribution. They created maps based on DNN uncertainty models. Their algorithm was used to create a map that included information over obstacle distance estimates while taking care of the level of uncertainty in each estimate. Traditional high-dimensional data reduction methods such as projection pursuit regression (PPR), reduced rank regression (RRR), partial least squares (PLS), and principal component analysis (PCA) are all shallow learners. Polson et al. [544] examined DL counterparts that exploited multiple deep layers of data reduction and provided predictive performance gains. Dropout regularization and SGD training optimisation provided variable selection and estimation. They illustrated their technique by providing an analysis of international bookings on Airbnb.

Energy (e.g. electricity markets) is another common domain for application of different UQ methods with machine and deep learning. A successful participation to liberalized electricity markets can be achieved by forecasting accurate day-ahead energy prices. Brusaferri et al. [545] presented a new approach based on BDL strategies for probabilistic energy price forecast. Scalability to complex network was guaranteed by executing a specific training method. They examined their system on two day-ahead markets characterized by different behaviours. Robust performance of the system was achieved by providing forecast uncertainty indications in out-of-sample conditions. Moreover, class imbalance in remote sensing poses a challenge for land cover mapping where small objects get less attention to yield better accuracy. Uncertainty quantification on pixel level using CNN is another issue in remote sensing. Kampffmeyer et al. [546] devised a deep CNN in remote sensing images for land cover mapping with prime aim on urban areas. Their method tried to achieve good accuracy for small objects. They applied recent technologies for UQ in the domain of remote sensing. They yielded overall classification accuracy of 87%. On the other hand, accurately predicting net load arising from distributed photovoltaic (PV) generation is a great challenge. Sun et al. [547] presented a new probabilistic day-ahead net load forecasting technique to capture both aleatoric uncertainty and epistemic uncertainty utilizing BDL. The performance of aggregated net load forecasting was improved by considering residential rooftop PV outputs as input features and exploited clustering in subprofiles. The proposed scheme proved its efficacy with high PV visibility and subprofile clustering.

9 LITERATURE GAPS AND OPEN ISSUES

We reviewed most of UQ works in machine and deep learning methods and briefly discussed them in the previous sections. Nevertheless, there are several important literature gaps and open issues. In the following, we list the most important gaps and open issues that should be addressed in the near future. Furthermore, we have listed few future research directions for further studies. The most important gaps, open issues and future directions are discussed in the following.

- This review shows that most of the proposed UQ methods are presented in the supervised, followed by unsupervised learning methods. However, we realized that there are fewer studies on the semi-supervised learning methods. We believe that this is an important gap in the domain of UQ which could be filled in the future.

- Our findings reveal that most of UQ methods have been proposed overwhelmingly for different types of NNs, especially DL methods. However, there are many other methods in the field of ML that uncertainty has either not been investigated or superficially discussed. The probable reason for this possibility is that DL methods have been able to be almost the best (SOTA: the State Of The Art) methods in various fields (e.g. computer vision, medical image analysis, NLP, time series analysis, signal processing, etc.). But as a matter of fact, it can be claimed that different types of traditional ML methods have a significant performance on the analysis of small data whereas DL is almost incapable.

- Fusion-based methods (e.g., multi-modal Bayesian fusion [95], multi-level fusion [548], image fusion [549], data fusion [550], etc.) have shown great ability to optimize different machine and deep learning methods. This led us to investigate the effects of a variety of fusion-based methods to deal with uncertainty in machine and deep learning methods. We realized that fusion-based methods have significant potential to address the uncertainty of models. Therefore, we would suggest that more fusion-based methods can be used in future work for quantifying uncertainty.

- The results of our research show that although a variety of ensemble methods [551], [552], [553], [554], [555], [556], [557] also have a high ability to deal with uncertainty along with good performance and optimizing the performance of other methods, these high capabilities of these methods have not been used more significantly. In other words, we noticed that these methods have performed remarkably well in few studies. But we realized that the ensemble methods are less commonly used in recent studies. Therefore, we strongly recommend further studies on ensemble methods and their substantial impact on quantifying uncertainty in machine and deep learning domain. For example, Caldeira and Nord [318] presented a comparative study to compare the performance of the three UQ important methods: BNN, CD and DE. Based on the obtained outcomes, they recommend DE for further investigations and applications as it achieved the best or comparable results in the study.

- Decision-making is a cognitive process which results in choosing the best possible action/belief among
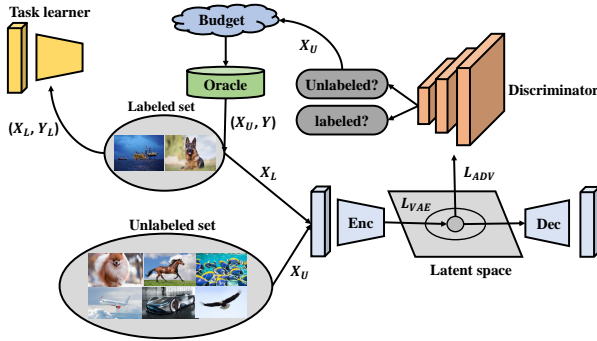
Fig. 24: A schematic view of the VAAL model which is reproduced based on [565].

all other alternative options. There are few well-known theories such as three-way decisions [558] and Info-Gap (IG) decision [559], [560] which can be used as UQ methods. Our findings reveal that these theories have been able to significantly help in dealing with uncertainty. For this reason, we think using various decision-making theories can be used during the decision-making process of machine and deep learning methods.

- Active learning (AL, also sometimes called optimal experimental design) plays a key role in dealing with lack of labeled data to label new data with the desirable outputs. As most researchers in the domain of machine and deep learning are aware that having labeled data is very difficult, costly and time consuming. In few very extremely sensitive areas such as healthcare and self-driving cars, this importance becomes more apparent. We have reviewed several good studies in this area, including: Gordon et al. [561], Lee et al. [311], Nguyen et al. [562], Hu et al. [563], Qu et al. [564] and some few more. Our reviews; however, reveal that although uncertainty in this area is quite important, but very few studies have been done in this subject. For example, Sinha et al. [565] proposed a new Al in an adversarial manner which is called VAAL (Variational Adversarial AL). The proposed model trained a latent space by using a VAE and an adversarial network trained to discriminate between labeled and unlabeled data (see Fig. 24). They showed the dramatic impact of this type of method on UQ. For this reason, we believe that researchers can fill this gap with further studies in order to improve the data labeling quality by having far more certainty than previous studies.
- Transfer learning is an idea technique to deal with well-training of different machine and deep learning methods whereas there is no enough data to train the models properly. In other words, transfer learning is a technique for repurposing and reusing already trained machine and deep learning methods in a set of new situations. Based on our review study, this technique also has uncertainties. Hence, we would recommend conducting more research on this area and proposing new UQ methods for transfer learning.
- Neural architecture search (NAS) methods are a set of techniques for automatically designing of ANNs.

Uncertainty awareness of such methods is an important and sensitive area for the use of UQ methods. However, the results of our research reveal that very few studies have been done so far. Hence, we list this case as another open research gap for further investigations. Moreover, neural ensemble search methods [566] are well-developed techniques for uncertainty calibration.

- Self supervised learning (SSL) [261], [567] is an important subset of unsupervised learning for generating output labels from different data objects by finding a relationship between various parts of the object, or a variety of views of the given object. We think that the SSL methods have several uncertainties and therefore further investigation of these methods has a high potential as an important research gap of UQ.
- The attention mechanism [568], [569] is a powerful strategy in NNs and DL methods to concentrate on adequate parts of the input data more than the unnecessary (irrelevant) parts while conducting a prediction task. However, we found that selecting relevant and irrelevant parts of data is accompanied by uncertainty. Our reviews show that fewer studies of UQ have been conducted in this area. For this reason, we also list this area as a research gap for further investigations.
- The OoD [570], [571], [572], [573], [574], [575] inputs (samples) can improve different models robustness as well as uncertainty. For example, Lee et al. [311] a new model called BTAML for dealing with imbalanced data and also detection OoD samples. According to previous studies, we can see that detection of OoD can help for outstanding performance of different NN and DL methods. However, quantifying uncertainties for detecting OoD samples needs further investigations [576].
- Hypernetworks [577], [578], [579] are very powerful methods to generate the weights for one network using another network. We found out that quantifying uncertainties in hypernetworks can be very useful approach to have a better defense to deal with adversarial samples. However, we found a very few studies dealing with uncertainties in hypernetworks [580]. Hence, we can also suggest this open issues to conduct more research for dealing with uncertainties in hypernetworks.
- Continual learning (CL) [581], [582], [583], [584], [585], [586] (or continual deep learning [587]) is a subset of machine learning which provides the ability of different models to continually learn from a stream of data. In other words, an algorithm deals with sequence of samples (on-stationary data) while reusing former knowledge and then exploiting it in a better adapting manner to a changeable environment. Our review shows that there are only a few studies introducing UQ methods in CL. Therefore, quantifying uncertainties in CL domain is another open research field for applying new UQ methods of dealing with uncertainties.
- Graph Neural Networks (GNNs) [588], [589], [590], [591], [592] are powerful graph representation learn-

ing methods to extract high-level features for related nodes from their topological neighborhoods. Uncertainty quantification of GNNs in graph analytic tasks is still an open challenge. Our findings reveal that even though Bayesian-based GNNs/GCNs (Graph Convolutional Networks) [280], [593], [594], [595] and Deep graph BO [596] have shown promising outcomes for quantifying uncertainties in GNNs/GCNs, but there are a limited number of studies in this research domain. Hence, we would recommend proposing more new yet efficient UQ methods for dealing with uncertainties in GNNs/GCNs.

- Bayesian optimization (BO) [597], [598] is a sample-efficient global optimization method for optimizing time-consuming black-box objective functions which take a quite long time to evaluate. We found out that BO can be very efficient approach not only for optimizing a wide range of applications, Hyperparameter tuning [599] but also for quantifying uncertainties in machine and deep learning methods or quantifying uncertainties in BO remains as an open issue for future investigation [600], [601].

- Uncertainty calibration is another approach for measuring the model's calibration errors which has been used in different case studies [201], [385], [602], [603]. However, we noticed that there are less studies on calibration of modern NNs, DLs and many more ML methods [411], [602], [604], [605], [606], [607]. Calibration measures can be then used for quantifying uncertainties in ML and DL predictions. Hence, we would recommend developing different uncertainty calibration approaches.

9.1 Future Directions

UQ methods have been achieving highly regarded achievement and obtaining distinguished performance when applied to different machine and deep learning methods. However, there are some open challenges that should be addressed. We provide several future directions of UQ methods for the three main research topics: computer vision and image processing, medical image processing and NLP.

Computer vision and image processing: As discussed earlier, computer vision and image processing are two main research domains for the application of different UQ methods. Although various studies have been conducted in these areas, but there are certainly still many open research directions which should be considered in the future researches. In the following, we aim to list few most important future directions in these domains.

Theoretical analysis and a more resilient inference methodology of various UQ methods should be investigated in future studies. For example, integration of semi-supervised and AL can be developed for acquiring new samples. In addition, data labelling is a time consuming and costly process in all domains not only for computer vision and image processing. Therefore, we recommend further studies on automated data labelling techniques and investigating the impact of UQ methods. Applications of the cascade structures have proven to be a powerful mechanism for improving various machine and deep learning methods.

However, we think that simplifying these methods and their integration with UQ methods for different computer vision and image processing tasks is valuable. Moreover, integration of dynamic/multi-modal image restoration issues with some advanced inversion approaches (e.g. different plug-and-play type schemes) for applying UQ methods for revealing relevant point estimates, is another interesting future direction. In addition, the review outcomes reveal that ensemble methods are still among the best approaches especially for detecting epistemic uncertainties (OoD issue). For this reason, application of new ensemble methods is another interesting research direction in computer vision and image processing.

Integration of UQ methods with different human pose architectures and then use the estimated uncertainty in future frame prediction and motion tracking issues, is another engaging open research direction. Also, although we mentioned above some UQ methods for BAL, but we noticed that better uncertainty estimates in BAL as well as more complex methods should be proposed in this domain. In addition, we found that detecting adversarial samples is an interesting challenge for BNNs [608]. Thus, we highly recommend further studies to develop more effective UQ methods for detecting adversarial samples. Sampling-free learning methods (e.g. Bayesian quantized networks (BQNs) [609], Sampling-free VI [610]) are powerful techniques for learning an appropriate posterior distribution over their discrete parameters with truly-calibrated uncertainty. Furthermore, embedding techniques have obtained outstanding performance in different computer vision and image processing tasks. However, we found that there are very few studies on probabilistic embedding strategies [611], [612] to quantify uncertainties. We also noticed that even though Bayesian methods (i.e., Variational Bayes [613] have been used for UQ in capsule routing methods, but calibrated uncertainty estimates of predictions by using different capsule networks is an open future work direction. Online applications of different BNNs is an open issue for future investigations due to various reasons such as the limitations of variational techniques and having risks for selecting appropriate approximations of batch posteriors [614]. Uncertainty of CL [272] is another open research direction in computer vision and image processing. For example, Nguyen et al. [615], [616] proposed variational CL (VCL) to deal with uncertainty and showed the effectiveness of such a approach. Finally, we found out that quantifying uncertainties in multitask transfer learning [617] is a very important research domain in which further investigations are highly warranted.

Medical image analysis: One possible research direction that could be considered in the future is a closer collaboration between medical and artificial intelligence researchers. Due to the high sensitivity in this field of science, we strongly recommend collecting larger medical data in the domain. This can be very helpful in resolving uncertainty, as a result, the proposed machine and deep learning methods can perform better in predicting various diseases and cancers. As we know, ground truth data for medical image segmentation plays a critical role in the correctness of the obtained results. For this reason, closer cooperation between the two groups can provide platforms for optimizing existing machine and deep learning models. Furthermore, refer-

ral of incorrect predicted data to specialists has a great role in dealing with uncertainty. Hence, there is a need for close collaboration between medical and computer researchers in the field of medical image segmentation.

We also noticed that various fusion methods have good potential for segmentation uncertainty predictions of medical images. Moreover, we found out that in most of previous studies standard losses for medical segmentation have been used whereas there are some new, yet effective, losses which could be used. On the other hand, combining both visualization of uncertainty and medical image segmentation can be used within AL computer assisted diagnostic (CAD) systems to improve the segmentation output. Another important future direction can be concentrated on big medical data collection. As we know, having a bigger data can dramatically improve the performance of various deep and machine learning methods. Our comprehensive review reveals that the problem of having enough medical data is still open. But if this is not possible, transfer learning techniques can be an ideal solution for improving the training process. Using this technique we can properly tune the applied DL methods; however, we know there are few uncertainties. As mentioned above, this can be considered as an open gap for future researchers.

The development of different semi-supervised-based methods for medical image segmentation is another promising approach for dealing with medical data shortages. We found out that UQ methods can have great impact in semi-supervised-based medical image segmentation which is another interesting research direction for future investigations. Along with all of these open directions in medical data analysis, MTL has also showed promising performance for medical data analysis. But as Nguyen et al. [618] showed adding UQ methods to MTL can significantly quantify uncertainty for prediction of clinical risks. However, as Gast and Roth [486] stated although probabilistic approaches have widely been used for several years; however, probabilistic approaches have not been comprehensively applied in practice since sampling techniques are frequently too slow. To solve this issue, development of proper probabilistic approaches [486] can be used in the real medical applications.

Natural language processing (NLP): There are few NLP-based subjects such as neural machine translation [619] and some other interdisciplinary subjects such as image captioning [620], [621], [622], [623], visual question answering (VQA) [624], [625], [626], [627], [628] which are closely associated with NLP. Finding the right caption for an image has always been a challenge in computer science. Especially since this issue can be accompanied by some uncertainties due to the merging of two important disciplines (i.e., image processing and NLP). On the other hand, medical VQA is a very important task for health organisations to find out the most appropriate answers to the given questions. Indeed, this topic handles both image processing and NLP tasks at the same time. We believe that because of the essence of the matter, adding methods to deal with uncertainties can greatly contribute to the productivity of this branch. In addition, classification of the data stream text environments is an interesting domain for application of UQ methods for finding the uncertain sentences.

Further directions: In this phase, we discuss few research directions in various subjects such as RL, numerical data analysis, signal processing, toy and synthetic data, etc. For instance, application of meta RL (MRL) is effective and efficient in optimizing decision performance. However, decision making process is always come with uncertainty. Hence, we suggest adding UQ methods with various MRL models can yield better decisions with more certainty. The proposed natural-gradient approach [160] can be generalized with some other approximation types (e.g., exponential-family distributions) and apply it on RL and stochastic optimization. Moreover, proper understanding of the interaction between choosing the inference algorithm and approximating family is another future research direction for synthetic data (regression task). Developing various stochastic regularization methods is another open direction for researchers. We also noticed that leveraging proper weights of the Frank-Wolfe optimization algorithm [117] and finding how this technique interacts with some alternative procedures approximate inference can be interesting avenues for further investigations.

Moreover, the digital healthcare is a very important research area which can help to make medicine more precise and personalized. Quantifying uncertainty of the digital healthcare and deploy them in some real-world clinical settings is another open research path. Approximate Bayesian inference in the CL and sequential decision making applications should be used as an inner procedure of a larger method. And this procedure needs robust version of BNNs. Hence, application of Deterministic VI [461] with different BNNs is an ideal approach. Accordingly, learning of different BNNs [629], [161] can be optimized using various Assumed Density Filtering (ADF) techniques [630] and apply them for different machine and deep learning tasks (e.g. NLP, computer vision and image processing, signal processing and many more). In addition, ensemble-based sampling [631], [632] methods have shown the capability to approximate sampling techniques (i.e., Thompson sampling) and properly deal with uncertainty in complex methods such as different NNs. Finally, quantifying uncertainties for multi-agent systems [633], [634] is also another important future direction since an individual agent cannot solve problems as are impossible or difficult.

9.2 Lack of Data and Code Availability

As we know, availability of data and codes play significant role for improving prior methods. In other words, having data and codes will assist researchers to go through proposed methods and find main gaps. Our comprehensive review shows that most of previous studies (especially medical case studies) do not share their data and codes with others. We understand that because of few cases the authors may not be able to make the data/code public, but we believe sharing both data and code will be very helpful to improve the quality and performance of different machine and deep learning methods. In other words, having a code and data of a research paper will accelerate the optimization of the code in the study.

10 CONCLUSION

Uncertainty quantification is one of the key parts of decision making process. The UQ methods are becoming popular to evaluate the uncertainty in various real-life applications. Nowadays, uncertainty has become an inseparable part of traditional machine and deep leering methods. This study has comprehensively reviewed the most important UQ methods which have been applied in traditional machine learning and deep learning. Hence, we provided a comprehensive description and comparative analysis of current state-of-the-art UQ approaches. Also, in this study we reviewed many important UQ perspectives, including common UQ applications, in-depth discussions, main research gaps, ending by presenting several solid future research directions in the area. We believe that this review paper focusing on the use of the UQ methods in artificial intelligence will benefit researchers in a variety of fields, and may be considered as a guideline for the use of UQ in deep learning-based applications.

REFERENCES

[1] A. Malinin, "Uncertainty estimation in deep learning with application to spoken language assessment," Ph.D. dissertation, University of Cambridge, 2019.

[2] H. Jiang, B. Kim, M. Guan, and M. Gupta, "To trust or not to trust a classifier," in *Advances in neural information processing systems*, 2018, pp. 5541–5552.

[3] T. M. Mitchell, *The need for biases in learning generalizations*. Department of Computer Science, Laboratory for Computer Science Research . . . , 1980.

[4] V.-L. Nguyen, S. Destercke, and E. Hüllermeier, "Epistemic uncertainty sampling," in *International Conference on Discovery Science*. Springer, 2019, pp. 72–86.

[5] C. C. Aggarwal, X. Kong, Q. Gu, J. Han, and S. Y. Philip, "Active learning: A survey," in *Data Classification*. Chapman and Hall/CRC, 2014, pp. 599–634.

[6] B. T. Phan, "Bayesian deep learning and uncertainty in computer vision," Master's thesis, University of Waterloo, 2019.

[7] Y. Gal, "Uncertainty in deep learning," Ph.D. dissertation, University of Cambridge, 2016.

[8] E. Hüllermeier and W. Waegeman, "Aleatoric and epistemic uncertainty in machine learning: A tutorial introduction," *arXiv preprint arXiv:1910.09457*, 2019.

[9] C. Hubschneider, R. Hutmacher, and J. M. Zöllner, "Calibrating uncertainty models for steering angle estimation," in *IEEE Intelligent Transportation Systems Conference*, 2019, pp. 1511–1518.

[10] J. Mukhoti and Y. Gal, "Evaluating bayesian deep learning methods for semantic segmentation," *arXiv preprint arXiv:1811.12709*, 2018.

[11] H. Kabir, M. Abdar, S. M. J. Jalali, A. Khosravi, A. F. Atiya, S. Nahavandi, and D. Srinivasan, "Spinalnet: Deep neural network with gradual input," *arXiv preprint arXiv:2007.03347*, 2020.

[12] X. Wang, Y. Zhao, and F. Pourpanah, "Recent advances in deep learning," *International Journal of Machine Learning and Cybernetics*, 2020.

[13] E. Begoli, T. Bhattacharya, and D. Kusnezov, "The need for uncertainty quantification in machine-assisted medical decision making," *Nature Machine Intelligence*, vol. 1, no. 1, pp. 20–23, 2019.

[14] D. Hafner, D. Tran, T. Lillicrap, A. Irpan, and J. Davidson, "Noise contrastive priors for functional uncertainty," *arXiv preprint arXiv:1807.09289*, 2018.

[15] C. Ning and F. You, "Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming," *Computers & Chemical Engineering*, vol. 125, pp. 434–448, 2019.

[16] H. D. Kabir, A. Khosravi, M. A. Hosen, and S. Nahavandi, "Neural network-based uncertainty quantification: A survey of methodologies and applications," *IEEE access*, vol. 6, pp. 36 218–36 234, 2018.

[17] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," California Univ San Diego La Jolla Inst for Cognitive Science, Tech. Rep., 1985.

[18] A. D. Kiureghian and O. Ditlevsen, "Aleatory or epistemic? does it matter?" *Structural Safety*, vol. 31, no. 2, pp. 105 – 112, 2009, risk Acceptance and Risk Communication.

[19] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" in *Advances in neural information processing systems*, 2017, pp. 5574–5584.

[20] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[21] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *Journal of the American statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.

[22] Y. Gal and Z. Ghahramani, "Bayesian convolutional neural networks with bernoulli approximate variational inference," *arXiv preprint arXiv:1506.02158*, 2015.

[23] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, "An introduction to variational methods for graphical models," *Machine learning*, vol. 37, no. 2, pp. 183–233, 1999.

[24] K.-C. Wang, P. Vicol, J. Lucas, L. Gu, R. Grosse, and R. Zemel, "Adversarial distillation of bayesian neural network posteriors," *arXiv preprint arXiv:1806.10317*, 2018.

[25] L. V. Jospin, W. Buntine, F. Boussaid, H. Laga, and M. Bennamoun, "Hands-on bayesian neural networks–a tutorial for deep learning users," *arXiv preprint arXiv:2007.06823*, 2020.

[26] H. Wang and D.-Y. Yeung, "Towards bayesian deep learning: A survey," *arXiv preprint arXiv:1604.01662*, 2016.

[27] J. Maroñas, R. Paredes, and D. Ramos, "Calibration of deep probabilistic models with decoupled bayesian neural networks," *Neurocomputing*, 2020.

[28] P. Izmailov, W. J. Maddox, P. Kirichenko, T. Garipov, D. Vetrov, and A. G. Wilson, "Subspace inference for bayesian deep learning," in *Uncertainty in Artificial Intelligence*. PMLR, 2020, pp. 1169–1179.

[29] A. D. Cobb and B. Jalaian, "Scaling hamiltonian monte carlo inference for bayesian neural networks with symmetric splitting," *arXiv preprint arXiv:2010.06772*, 2020.

[30] T. Karaletsos and T. D. Bui, "Hierarchical gaussian process priors for bayesian neural network weights," *arXiv preprint arXiv:2002.04033*, 2020.

[31] A. Foong, D. Burt, Y. Li, and R. Turner, "On the expressiveness of approximate inference in bayesian neural networks," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[32] A. Kucukelbir, D. Tran, R. Ranganath, A. Gelman, and D. M. Blei, "Automatic differentiation variational inference," *J. Mach. Learn. Res.*, vol. 18, no. 1, p. 430–474, 2017.

[33] R. M. Neal, *Bayesian learning for neural networks*. Springer Science & Business Media, 2012, vol. 118.

[34] 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, no. 56, pp. 1929–1958, 2014.

[35] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*, 2016, pp. 1050–1059.

[36] P. McClure and N. Kriegeskorte, "Representing inferential uncertainty in deep neural networks through sampling," in *International Conference on Learning Representations, ICLR 2017-Conference Track Proceedings*, 2016.

[37] K. Brach, B. Sick, and O. Dürr, "Single shot mc dropout approximation," *arXiv preprint arXiv:2007.03293*, 2020.

[38] G. Wang, W. Li, M. Aertsen, J. Deprest, S. Ourselin, and T. Vercauteren, "Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks," *Neurocomputing*, vol. 338, pp. 34 – 45, 2019.

[39] H. Liu, R. Ji, J. Li, B. Zhang, Y. Gao, Y. Wu, and F. Huang, "Universal adversarial perturbation via prior driven uncertainty approximation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 2941–2949.

[40] T. Nair, D. Precup, D. L. Arnold, and T. Arbel, "Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation," *Medical Image Analysis*, vol. 59, p. 101557, 2020.

[41] A. Amini, A. Soleimany, S. Karaman, and D. Rus, "Spatial uncertainty sampling for end-to-end control," *arXiv preprint arXiv:1805.04829*, 2018.

[42] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler, "Efficient object localization using convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 648–656.

[43] H. P. Do, Y. Guo, A. J. Yoon, and K. S. Nayak, "Accuracy, uncertainty, and adaptability of automatic myocardial asl segmentation using deep cnn," *Magnetic Resonance in Medicine*, vol. 83, no. 5, pp. 1863–1874, 2020.

[44] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015.

[45] M. Teye, H. Azizpour, and K. Smith, "Bayesian uncertainty estimation for batch normalized deep networks," in *Proceedings of the 35th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, J. Dy and A. Krause, Eds., vol. 80. Stockholmsmässan, Stockholm Sweden: PMLR, 10–15 Jul 2018, pp. 4907–4916.

[46] L. Yu, S. Wang, X. Li, C.-W. Fu, and P.-A. Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation," in *Medical Image Computing and Computer Assisted Intervention*, D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, Eds. Cham: Springer International Publishing, 2019, pp. 605–613.

[47] C. Li, C. Chen, Y. Pu, R. Henao, and L. Carin, "Communication-efficient stochastic gradient mcmc for neural networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 4173–4180.

[48] A. Kendall, V. Badrinarayanan, and R. Cipolla, "Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding," *arXiv preprint arXiv:1511.02680*, 2015.

[49] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[50] C. Leibig, V. Allken, M. S. Ayhan, P. Berens, and S. Wahl, "Leveraging uncertainty information from deep neural networks for disease detection," *Scientific reports*, vol. 7, no. 1, p. 17816, 2017.

[51] S. Choi, K. Lee, S. Lim, and S. Oh, "Uncertainty-aware learning from demonstration using mixture density networks with sampling-free variance modeling," 2017.

[52] C. M. Bishop, "Mixture density networks," Neural Network Research Group, Tech. Rep., 1994.

[53] A. Jungo, R. McKinley, R. Meier, U. Knecht, L. Vera, J. Pérez-Beteta, D. Molina-García, V. M. Pérez-García, R. Wiest, and M. Reyes, "Towards uncertainty-assisted brain tumor segmentation and survival prediction," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, A. Crimi, S. Bakas, H. Kuijf, B. Menze, and M. Reyes, Eds. Cham: Springer International Publishing, 2018, pp. 474–485.

[54] T. Pohlen, A. Hermans, M. Mathias, and B. Leibe, "Full-resolution residual networks for semantic segmentation in street scenes," 2016.

[55] K. Wickstrøm, M. Kampffmeyer, and R. Jenssen, "Uncertainty modeling and interpretability in convolutional neural networks for polyp segmentation," in *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2018, pp. 1–6.

[56] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2017.

[57] A. Jungo, R. Meier, E. Ermis, E. Herrmann, and M. Reyes, "Uncertainty-driven sanity check: Application to postoperative brain tumor cavity segmentation," *arXiv preprint arXiv:1806.03106*, 2018.

[58] T. Vandal, M. Livingston, C. Piho, and S. Zimmerman, "Prediction and uncertainty quantification of daily airport flight delays," in *International Conference on Predictive Applications and APIs*, 2018, pp. 45–51.

[59] T. DeVries and G. W. Taylor, "Leveraging uncertainty estimates for predicting segmentation quality," *ArXiv*, vol. abs/1807.00502, 2018.

[60] A. Tousignant, P. Lemaître, D. Precup, D. L. Arnold, and T. Arbel, "Prediction of disease progression in multiple sclerosis patients using deep learning analysis of mri data," in *International Conference on Medical Imaging with Deep Learning*, 2019, pp. 483–492.

[61] A. Norouzi, A. Emami, K. Najarian, N. Karimi, S. samavi, and S. M. R. Soroushmehr, "Exploiting uncertainty of deep neural networks for improving segmentation accuracy in mri images," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 2322–2326.

[62] A. G. Roy, S. Conjeti, N. Navab, C. Wachinger, A. D. N. Initiative *et al.*, "Bayesian quicknat: model uncertainty in deep whole-brain segmentation for structure-wise quality control," *NeuroImage*, vol. 195, pp. 11–22, 2019.

[63] A. Filos, S. Farquhar, A. N. Gomez, T. G. J. Rudner, Z. Kenton, L. Smith, M. Alizadeh, A. de Kroon, and Y. Gal, "A systematic comparison of bayesian deep learning robustness in diabetic retinopathy tasks," 2019.

[64] R. Harper and J. Southern, "A bayesian deep learning framework for end-to-end prediction of emotion from heartbeat," *IEEE Transactions on Affective Computing*, 2020.

[65] A. Y. Foong, D. R. Burt, Y. Li, and R. E. Turner, "On the expressiveness of approximate inference in bayesian neural networks," *arXiv*, pp. arXiv–1909, 2019.

[66] M. Ng, F. Guo, L. Biswas, and G. A. Wright, "Estimating uncertainty in neural networks for segmentation quality control," Technical Report. URL: https://www. doc. ic. ac. uk/~ bglocker/public . . . , Tech. Rep., 2018.

[67] A. Siddhant and Z. C. Lipton, "Deep Bayesian active learning for natural language processing: Results of a large-scale empirical study," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels, Belgium: Association for Computational Linguistics, Oct.-Nov. 2018, pp. 2904–2909.

[68] J. Mukhoti, P. Stenetorp, and Y. Gal, "On the importance of strong baselines in bayesian deep learning," *arXiv preprint arXiv:1811.09385*, 2018.

[69] N. Kennamer, A. T. Ihler, and D. Kirkby, "Empirical study of mc-dropout in various astronomical observing conditions," in *CVPR Workshops*, 2019.

[70] M. A. Kupinski, J. W. Hoppin, E. Clarkson, and H. H. Barrett, "Ideal-observer computation in medical imaging with use of markov-chain monte carlo techniques," *JOSA A*, vol. 20, no. 3, pp. 430–438, 2003.

[71] R. Salakhutdinov and A. Mnih, "Bayesian probabilistic matrix factorization using markov chain monte carlo," in *Proceedings of the International Conference on Machine Learning*, ser. ICML '08. Association for Computing Machinery, 2008, p. 880–887.

[72] T. Salimans, D. Kingma, and M. Welling, "Markov chain monte carlo and variational inference: Bridging the gap," in *International Conference on Machine Learning*, 2015, pp. 1218–1226.

[73] T. Chen, E. Fox, and C. Guestrin, "Stochastic gradient hamiltonian monte carlo," in *International conference on machine learning*, 2014, pp. 1683–1691.

[74] N. Ding, Y. Fang, R. Babbush, C. Chen, R. D. Skeel, and H. Neven, "Bayesian sampling using stochastic gradient thermostats," in *Advances in neural information processing systems*, 2014, pp. 3203–3211.

[75] C. Chen, N. Ding, and L. Carin, "On the convergence of stochastic gradient mcmc algorithms with high-order integrators," in *Advances in Neural Information Processing Systems*, 2015, pp. 2278–2286.

[76] C. Li, A. Stevens, C. Chen, Y. Pu, Z. Gan, and L. Carin, "Learning weight uncertainty with stochastic gradient mcmc for shape classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5666–5675.

[77] W. Gong, S. Tschiatschek, S. Nowozin, R. E. Turner, J. M. Hernández-Lobato, and C. Zhang, "Icebreaker: Element-wise efficient information acquisition with a bayesian deep latent gaussian model," in *Advances in Neural Information Processing Systems 32*, 2019, pp. 14 820–14 831.

[78] Y. W. Teh, A. H. Thiery, and S. J. Vollmer, "Consistency and fluctuations for stochastic gradient langevin dynamics," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 193–225, 2016.

[79] A. Choromanska, M. Henaff, M. Mathieu, G. B. Arous, and Y. LeCun, "The loss surfaces of multilayer networks," in *Artificial intelligence and statistics*, 2015, pp. 192–204.

[80] R. Zhang, C. Li, J. Zhang, C. Chen, and A. G. Wilson, "Cyclical stochastic gradient mcmc for bayesian deep learning," *arXiv preprint arXiv:1902.03932*, 2019.

[81] R. Luo, J. Wang, Y. Yang, J. WANG, and Z. Zhu, "Thermostat-assisted continuously-tempered hamiltonian monte carlo for bayesian learning," in *Advances in Neural Information Processing Systems 31*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman,

N. Cesa-Bianchi, and R. Garnett, Eds. Curran Associates, Inc., 2018, pp. 10 673–10 682.

[82] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid monte carlo," *Physics letters B*, vol. 195, no. 2, pp. 216–222, 1987.

[83] S. Hernández, D. Vergara, M. Valdenegro-Toro, and F. Jorquera, "Improving predictive uncertainty estimation using dropout–hamiltonian monte carlo," *Soft Computing*, vol. 24, no. 6, pp. 4307–4322, 2020.

[84] W. G. Hoover, "Canonical dynamics: Equilibrium phase-space distributions," *Physical review A*, vol. 31, no. 3, p. 1695, 1985.

[85] S. Nosé, "A unified formulation of the constant temperature molecular dynamics methods," *The Journal of chemical physics*, vol. 81, no. 1, pp. 511–519, 1984.

[86] M. Welling and Y. W. Teh, "Bayesian learning via stochastic gradient langevin dynamics," in *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011, pp. 681–688.

[87] W. Zhou and M. A. Anastasio, "Markov-chain monte carlo approximation of the ideal observer using generative adversarial networks," in *Medical Imaging 2020: Image Perception, Observer Performance, and Technology Assessment*, vol. 11316. International Society for Optics and Photonics, 2020, p. 113160D.

[88] J. Kwon, "Robust visual tracking based on variational auto-encoding markov chain monte carlo," *Information Sciences*, vol. 512, pp. 1308 – 1323, 2020.

[89] J. Swiatkowski, K. Roth, B. S. Veeling, L. Tran, J. V. Dillon, S. Mandt, J. Snoek, T. Salimans, R. Jenatton, and S. Nowozin, "The k-tied normal distribution: A compact parameterization of gaussian mean field posteriors in bayesian neural networks," *arXiv preprint arXiv:2002.02655*, 2020.

[90] K. Posch, J. Steinbrener, and J. Pilz, "Variational inference to measure model uncertainty in deep neural networks," *arXiv preprint arXiv:1902.10189*, 2019.

[91] K. Posch and J. Pilz, "Correlated parameters to accurately measure uncertainty in deep neural networks," 2019.

[92] H. Robbins, "An empirical bayes approach to statistics," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. University of California Press, 1956, pp. 157–163.

[93] R. Krishnan, M. Subedar, and O. Tickoo, "Efficient priors for scalable variational inference in bayesian deep neural networks," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 2019, pp. 773–777.

[94] R. Krishnan, M. Subedar, O. Tickoo, A. Filos, and Y. Gal, "Improving mfvi in bayesian neural networks with empirical bayes: a study with diabetic retinopathy diagnosis," in *Fourht workshop on Bayesian Deep Learning (NeurIPS 2019)*, 2019.

[95] M. Subedar, R. Krishnan, P. L. Meyer, O. Tickoo, and J. Huang, "Uncertainty-aware audiovisual activity recognition using deep bayesian variational inference," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 6300–6309.

[96] D. L. Marino and M. Manic, "Modeling and planning under uncertainty using deep neural networks," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 8, pp. 4442–4454, 2019.

[97] T. D. Le, R. Noumeir, H. L. Quach, J. H. Kim, J. H. Kim, and H. M. Kim, "Critical temperature prediction for a superconductor: A variational bayesian neural network approach," *IEEE Transactions on Applied Superconductivity*, vol. 30, no. 4, pp. 1–5, 2020.

[98] C. Louizos and M. Welling, "Multiplicative normalizing flows for variational bayesian neural networks," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2017, pp. 2218–2227.

[99] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2013.

[100] A. Hubin and G. Storvik, "Combining model and parameter uncertainty in bayesian neural networks," *arXiv preprint arXiv:1903.07594*, 2019.

[101] Y. Liu, H. Qin, Z. Zhang, S. Pei, Z. Jiang, Z. Feng, and J. Zhou, "Probabilistic spatiotemporal wind speed forecasting based on a variational bayesian deep learning model," *Applied Energy*, vol. 260, p. 114259, 2020.

[102] S. Ryu, Y. Kwon, and W. Y. Kim, "A bayesian graph convolutional network for reliable prediction of molecular properties with uncertainty quantification," *Chemical Science*, vol. 10, no. 36, pp. 8438–8446, 2019.

[103] S. Farquhar, L. Smith, and Y. Gal, "Try depth instead of weight correlations: Mean-field is a less restrictive assumption for variational inference in deep networks," in *Bayesian Deep Learning Workshop at NeurIPS*, 2020.

[104] H. S. Hossain, M. A. A. H. Khan, and N. Roy, "Active learning enabled activity recognition," *Pervasive and Mobile Computing*, vol. 38, pp. 312–330, 2017.

[105] S. Tong, *Active learning: theory and applications*. Stanford University USA, 2001, vol. 1.

[106] Y. Gal, R. Islam, and Z. Ghahramani, "Deep bayesian active learning with image data," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2017, pp. 1183–1192.

[107] N. Houlsby, F. Huszár, Z. Ghahramani, and M. Lengyel, "Bayesian active learning for classification and preference learning," 2011.

[108] R. Krishnan, M. Subedar, and O. Tickoo, "Specifying weight priors in bayesian deep neural networks with empirical bayes," *arXiv:1906.05323*, 2019.

[109] A. Kirsch, J. van Amersfoort, and Y. Gal, "Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning," 2019.

[110] S. Burkhardt, J. Siekiera, and S. Kramer, "Semisupervised bayesian active learning for text classification," in *Bayesian Deep Learning Workshop at NeurIPS*, 2018.

[111] G. K. Gudur, P. Sundaramoorthy, and V. Umaashankar, "Activeharnet: Towards on-device deep bayesian active learning for human activity recognition," in *The 3rd International Workshop on Deep Learning for Mobile Systems and Applications*, 2019, pp. 7–12.

[112] R. Matthias, K. Karsten, and G. Hanno, "Deep bayesian active semi-supervised learning," in *IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2018, pp. 158–164.

[113] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[114] M. L. di Scandalea, C. S. Perone, M. Boudreau, and J. Cohen-Adad, "Deep active learning for axon-myelin segmentation on histology data," 2019.

[115] J. Zeng, A. Lesnikowski, and J. M. Alvarez, "The relevance of bayesian layer positioning to model uncertainty in deep bayesian active learning," *arXiv preprint arXiv:1811.12535*, 2018.

[116] J. Huggins, T. Campbell, and T. Broderick, "Coresets for scalable bayesian logistic regression," in *Advances in Neural Information Processing Systems*, 2016, pp. 4080–4088.

[117] R. Pinsler, J. Gordon, E. Nalisnick, and J. M. Hernández-Lobato, "Bayesian batch active learning as sparse subset approximation," in *Advances in Neural Information Processing Systems*, 2019, pp. 6356–6367.

[118] M. Servajean, A. Joly, D. Shasha, J. Champ, and E. Pacitti, "Crowdsourcing thousands of specialized labels: A bayesian active training approach," *IEEE Transactions on Multimedia*, vol. 19, no. 6, pp. 1376–1391, 2017.

[119] E. Simpson and S. Roberts, "Bayesian methods for intelligent task assignment in crowdsourcing systems," in *Decision Making: Uncertainty, Imperfection, Deliberation and Scalability*. Springer, 2015, pp. 1–32.

[120] J. Gordon and J. M. Hernández-Lobato, "Bayesian semisupervised learning with deep generative models," *arXiv preprint arXiv:1706.09751*, 2017.

[121] T. Tran, T.-T. Do, I. Reid, and G. Carneiro, "Bayesian generative active deep learning," *arXiv preprint arXiv:1904.11643*, 2019.

[122] A. Akbari and R. Jafari, "Personalizing activity recognition models with quantifying different types of uncertainty using wearable sensors," *IEEE Transactions on Biomedical Engineering*, 2020.

[123] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, "Weight uncertainty in neural networks," *arXiv preprint arXiv:1505.05424*, 2015.

[124] M. Fortunato, C. Blundell, and O. Vinyals, "Bayesian recurrent neural networks," *arXiv preprint arXiv:1704.02798*, 2017.

[125] S. Ebrahimi, M. Elhoseiny, T. Darrell, and M. Rohrbach, "Uncertainty-guided continual learning with bayesian neural networks," *arXiv preprint arXiv:1906.02425*, 2019.

[126] M. de la Riva and P. Mettes, "Bayesian 3d convnets for action recognition from few examples," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2019, pp. 0–0.

[127] P. Ghosh, M. S. M. Sajjadi, A. Vergari, M. Black, and B. Schölkopf, "From variational to deterministic autoencoders," 2019.

[128] S. Z. Dadaneh, S. Boluki, M. Yin, M. Zhou, and X. Qian, "Pairwise supervised hashing with bernoulli variational auto-encoder and self-control gradient estimator," *arXiv preprint arXiv:2005.10477*, 2020.

[129] V. Böhm, F. Lanusse, and U. Seljak, "Uncertainty quantification with generative models," *arXiv preprint arXiv:1910.10046*, 2019.

[130] U. Seljak and B. Yu, "Posterior inference unchained with EL$_2$O," 2019.

[131] V. Edupuganti, M. Mardani, J. Cheng, S. Vasanawala, and J. Pauly, "Uncertainty analysis of vae-gans for compressive medical imaging," *arXiv preprint arXiv:1901.11228*, 2019.

[132] L. Jin, H. Lu, and G. Wen, "Fast uncertainty quantification of reservoir simulation with variational u-net," 2019.

[133] P. Esser, E. Sutter, and B. Ommer, "A variational u-net for conditional appearance and shape generation," 2018.

[134] K. Yi, Y. Guo, Y. Fan, J. Hamann, and Y. G. Wang, "Cosmovae: Variational autoencoder for cmb image inpainting," 2020.

[135] N. Mehrasa, A. A. Jyothi, T. Durand, J. He, L. Sigal, and G. Mori, "A variational auto-encoder model for stochastic point processes," 2019.

[136] K. Sato, K. Hama, T. Matsubara, and K. Uehara, "Predictable uncertainty-aware unsupervised deep anomaly segmentation," in *2019 International Joint Conference on Neural Networks (IJCNN)*, 2019, pp. 1–7.

[137] S. Mishra, S. Flaxman, and S. Bhatt, "πvae: Encoding stochastic process priors with variational autoencoders," 2020.

[138] M. Garnelo, D. Rosenbaum, C. J. Maddison, T. Ramalho, D. Saxton, M. Shanahan, Y. W. Teh, D. J. Rezende, and S. Eslami, "Conditional neural processes," *arXiv preprint arXiv:1807.01613*, 2018.

[139] F. Guo, R. Xie, and B. Huang, "A deep learning just-in-time modeling approach for soft sensor based on variational autoencoder," *Chemometrics and Intelligent Laboratory Systems*, vol. 197, p. 103922, 2020.

[140] E. Daxberger and J. M. Hernández-Lobato, "Bayesian variational autoencoders for unsupervised out-of-distribution detection," 2019.

[141] A. Damianou and N. Lawrence, "Deep gaussian processes," in *Artificial Intelligence and Statistics*, 2013, pp. 207–215.

[142] D. Duvenaud, O. Rippel, R. Adams, and Z. Ghahramani, "Avoiding pathologies in very deep networks," in *Artificial Intelligence and Statistics*, 2014, pp. 202–210.

[143] H. Salimbeni and M. Deisenroth, "Doubly stochastic variational inference for deep gaussian processes," in *Advances in Neural Information Processing Systems*, 2017, pp. 4588–4599.

[144] A. Borovykh, "A gaussian process perspective on convolutional neural networks," *arXiv preprint arXiv:1810.10798*, 2018.

[145] H. Yu, Y. Chen, Z. Dai, K. H. Low, and P. Jaillet, "Implicit posterior variational inference for deep gaussian processes," *arXiv preprint arXiv:1910.11998*, 2019.

[146] S. Sun, W. Dong, and Q. Liu, "Multi-view representation learning with deep gaussian processes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[147] S. W. Ober and L. Aitchison, "Global inducing point variational posteriors for bayesian neural networks and deep gaussian processes," *arXiv preprint arXiv:2005.08140*, 2020.

[148] C. E. Rasmussen, "Gaussian processes for machine learning," in *Advanced Lectures on Machine Learning*. MIT Press, 2006.

[149] T. Teng, J. Chen, Y. Zhang, and B. K. H. Low, "Scalable variational bayesian kernel selection for sparse gaussian process regression." in *AAAI*, 2020, pp. 5997–6004.

[150] S. J. Oh, K. Murphy, J. Pan, J. Roth, F. Schroff, and A. Gallagher, "Modeling uncertainty with hedged instance embedding," *arXiv preprint arXiv:1810.00319*, 2018.

[151] M. Havasi, J. M. Hernández-Lobato, and J. J. Murillo-Fuentes, "Inference in deep gaussian processes using stochastic gradient hamiltonian monte carlo," in *Advances in neural information processing systems*, 2018, pp. 7506–7516.

[152] W. Maddox, T. Garipov, P. Izmailov, D. Vetrov, and A. G. Wilson, "Fast uncertainty estimates and bayesian model averaging of dnns," in *Uncertainty in Deep Learning Workshop at UAI*, 2018.

[153] P. Izmailov, D. Podoprikhin, T. Garipov, D. Vetrov, and A. G. Wilson, "Averaging weights leads to wider optima and better generalization," *arXiv preprint arXiv:1803.05407*, 2018.

[154] W. J. Maddox, P. Izmailov, T. Garipov, D. P. Vetrov, and A. G. Wilson, "A simple baseline for bayesian uncertainty in deep learning," in *Advances in Neural Information Processing Systems*, 2019, pp. 13 153–13 164.

[155] Y. Wen, P. Vicol, J. Ba, D. Tran, and R. Grosse, "Flipout: Efficient pseudo-independent weight perturbations on mini-batches," *arXiv preprint arXiv:1803.04386*, 2018.

[156] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," 2013.

[157] J. Bradshaw, A. G. d. G. Matthews, and Z. Ghahramani, "Adversarial examples, uncertainty, and transfer testing robustness in gaussian process hybrid deep networks," *arXiv preprint arXiv:1707.02476*, 2017.

[158] J. Choi, D. Chun, H. Kim, and H.-J. Lee, "Gaussian yolov3: An accurate and fast object detector using localization uncertainty for autonomous driving," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 502–511.

[159] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," 2018.

[160] M. E. Khan, D. Nielsen, V. Tangkaratt, W. Lin, Y. Gal, and A. Srivastava, "Fast and scalable bayesian deep learning by weight-perturbation in adam," *arXiv preprint arXiv:1806.04854*, 2018.

[161] S. Sun, C. Chen, and L. Carin, "Learning structured weight uncertainty in bayesian neural networks," in *Artificial Intelligence and Statistics*, 2017, pp. 1283–1292.

[162] A. K. Gupta and D. K. Nagar, *Matrix variate distributions*. CRC Press, 2018, vol. 104.

[163] C. Louizos and M. Welling, "Structured and efficient variational deep learning with matrix gaussian posteriors," in *International Conference on Machine Learning*, 2016, pp. 1708–1716.

[164] M. Van der Wilk, C. E. Rasmussen, and J. Hensman, "Convolutional gaussian processes," in *Advances in Neural Information Processing Systems*, 2017, pp. 2849–2858.

[165] K. Blomqvist, S. Kaski, and M. Heinonen, "Deep convolutional gaussian processes," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2019, pp. 582–597.

[166] G.-L. Tran, E. V. Bonilla, J. Cunningham, P. Michiardi, and M. Filippone, "Calibrating deep convolutional gaussian processes," in *The 22nd International Conference on Artificial Intelligence and Statistics*, 2019, pp. 1554–1563.

[167] V. Dutordoir, M. van der Wilk, A. Artemev, M. Tomczak, and J. Hensman, "Translation insensitivity for deep convolutional gaussian processes," *arXiv preprint arXiv:1902.05888*, 2019.

[168] J. Shi, M. Titsias, and A. Mnih, "Sparse orthogonal variational inference for gaussian processes," in *International Conference on Artificial Intelligence and Statistics*, 2020, pp. 1932–1942.

[169] C. Corbière, N. Thome, A. Bar-Hen, M. Cord, and P. Pérez, "Addressing failure prediction by learning model confidence," in *Advances in Neural Information Processing Systems*, 2019, pp. 2902–2913.

[170] A. Atanov, A. Ashukha, D. Molchanov, K. Neklyudov, and D. Vetrov, "Uncertainty estimation via stochastic batch normalization," *arXiv preprint arXiv:1802.04893*, 2018.

[171] K. Neklyudov, D. Molchanov, A. Ashukha, and D. Vetrov, "Variance networks: When expectation does not meet your expectations," *arXiv preprint arXiv:1803.03764*, 2018.

[172] D. J. MacKay and D. J. Mac Kay, *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[173] H. Ritter, A. Botev, and D. Barber, "A scalable laplace approximation for neural networks," in *6th International Conference on Learning Representations, ICLR 2018-Conference Track Proceedings*, vol. 6. International Conference on Representation Learning, 2018.

[174] J. Feng, M. Durner, Z.-C. Marton, F. Balint-Benczedi, and R. Triebel, "Introspective robot perception using smoothed predictions from bayesian neural networks," 2019.

[175] K. Shinde, J. Lee, M. Humt, A. Sezgin, and R. Triebel, "Learning multiplicative interactions with bayesian neural networks for visual-inertial odometry," *arXiv preprint arXiv:2007.07630*, 2020.

[176] J. Lee, M. Humt, J. Feng, and R. Triebel, "Estimating model uncertainty of neural networks in sparse information form," *arXiv preprint arXiv:2006.11631*, 2020.

[177] M. Humt, J. Lee, and R. Triebel, "Bayesian optimization meets laplace approximation for robotic introspection," *arXiv preprint arXiv:2010.16141*, 2020.

[178] T. Doan, B. Mazoure, A. Durand, J. Pineau, and R. D. Hjelm, "Attraction-repulsion actor-critic for continuous control reinforcement learning," *arXiv preprint arXiv:1909.07543*, 2019.

[179] X. Zhao, S. Hu, J.-H. Cho, and F. Chen, "Uncertainty-based decision making using deep reinforcement learning," in *2019 22th International Conference on Information Fusion (FUSION)*. IEEE, 2019, pp. 1–8.

[180] G. Lee, B. Hou, A. Mandalika, J. Lee, S. Choudhury, and S. S. Srinivasa, "Bayesian policy optimization for model uncertainty," *arXiv preprint arXiv:1810.01014*, 2018.

[181] B. O'Donoghue, I. Osband, R. Munos, and V. Mnih, "The uncertainty bellman equation and exploration," *arXiv preprint arXiv:1709.05380*, 2017.

[182] G. Kahn, A. Villaflor, V. Pong, P. Abbeel, and S. Levine, "Uncertainty-aware reinforcement learning for collision avoidance," *arXiv preprint arXiv:1702.01182*, 2017.

[183] M. Ghavamzadeh, S. Mannor, J. Pineau, A. Tamar *et al.*, "Bayesian reinforcement learning: A survey," *Foundations and Trends® in Machine Learning*, vol. 8, no. 5-6, pp. 359–483, 2015.

[184] A. Tschantz, B. Millidge, A. K. Seth, and C. L. Buckley, "Reinforcement learning through active inference," *arXiv preprint arXiv:2002.12636*, 2020.

[185] G. Kalweit and J. Boedecker, "Uncertainty-driven imagination for continuous deep reinforcement learning," in *Conference on Robot Learning*, 2017, pp. 195–206.

[186] C. Tegho, P. Budzianowski, and M. Gašić, "Benchmarking uncertainty estimates with deep reinforcement learning for dialogue policy optimisation," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 6069–6073.

[187] C. Riquelme, G. Tucker, and J. Snoek, "Deep bayesian bandits showdown: An empirical comparison of bayesian deep networks for thompson sampling," *arXiv preprint arXiv:1802.09127*, 2018.

[188] T. Pearce, N. Anastassacos, M. Zaki, and A. Neely, "Bayesian inference with anchored ensembles of neural networks, and application to exploration in reinforcement learning," *arXiv preprint arXiv:1805.11324*, 2018.

[189] D. Janz, J. Hron, P. Mazur, K. Hofmann, J. M. Hernández-Lobato, and S. Tschiatschek, "Successor uncertainties: exploration and uncertainty in temporal difference learning," in *Advances in Neural Information Processing Systems*, 2019, pp. 4509–4518.

[190] M. Shen and J. P. How, "Active perception in adversarial scenarios using maximum entropy deep reinforcement learning," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 3384–3390.

[191] M. Benatan and E. O. Pyzer-Knapp, "Fully bayesian recurrent neural networks for safe reinforcement learning," *arXiv preprint arXiv:1911.03308*, 2019.

[192] W. Huang, J. Zhang, and K. Huang, "Bootstrap estimated uncertainty of the environment model for model-based reinforcement learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 3870–3877.

[193] H. Eriksson and C. Dimitrakakis, "Epistemic risk-sensitive reinforcement learning," *arXiv preprint arXiv:1906.06273*, 2019.

[194] B. Lötjens, M. Everett, and J. P. How, "Safe reinforcement learning with model uncertainty estimates," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 8662–8668.

[195] W. R. Clements, B.-M. Robaglia, B. Van Delft, R. B. Slaoui, and S. Toth, "Estimating risk and uncertainty in deep reinforcement learning," *arXiv preprint arXiv:1905.09638*, 2019.

[196] A. M. Metelli, A. Likmeta, and M. Restelli, "Propagating uncertainty in reinforcement learning via wasserstein barycenters," in *Advances in Neural Information Processing Systems*, 2019, pp. 4333–4345.

[197] C. D'Eramo, A. Cini, and M. Restelli, "Exploiting action-value uncertainty to drive exploration in reinforcement learning," in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–8.

[198] A. Tschantz, M. Baltieri, A. K. Seth, and C. L. Buckley, "Scaling active inference," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–8.

[199] H.-S. Lee, Y. Zhang, W. Zame, C. Shen, J.-W. Lee, and M. van der Schaar, "Robust recursive partitioning for heterogeneous treatment effects with uncertainty quantification," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[200] R. Ramakrishnan, E. Kamar, D. Dey, E. Horvitz, and J. Shah, "Blind spot detection for safe sim-to-real transfer," *Journal of Artificial Intelligence Research*, vol. 67, pp. 191–234, 2020.

[201] J. Zhang, B. Kailkhura, and T. Y.-J. Han, "Mix-n-match: Ensemble and compositional methods for uncertainty calibration in deep learning," in *International Conference on Machine Learning*. PMLR, 2020, pp. 11 117–11 128.

[202] J. Z. Liu, "Variable selection with rigorous uncertainty quantification using bayesian deep neural networks," in *Bayesian Deep Learning Workshop at NeurIPS*, 2019.

[203] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Advances in neural information processing systems*, 2017, pp. 6402–6413.

[204] S. Jain, G. Liu, J. Mueller, and D. Gifford, "Maximizing overall diversity for improved uncertainty estimates in deep ensembles," *arXiv preprint arXiv:1906.07380*, 2019.

[205] F. K. Gustafsson, M. Danelljan, and T. B. Schön, "Evaluating scalable bayesian deep learning methods for robust computer vision," *arXiv preprint arXiv:1906.01620*, 2019.

[206] P. L. McDermott and C. K. Wikle, "Deep echo state networks with uncertainty quantification for spatio-temporal forecasting," *Environmetrics*, vol. 30, no. 3, p. e2553, 2019.

[207] K. Chua, R. Calandra, R. McAllister, and S. Levine, "Deep reinforcement learning in a handful of trials using probabilistic dynamics models," in *Advances in Neural Information Processing Systems*, 2018, pp. 4754–4765.

[208] J. Liu, J. Paisley, M.-A. Kioumourtzoglou, and B. Coull, "Accurate uncertainty estimation and decomposition in ensemble learning," in *Advances in Neural Information Processing Systems*, 2019, pp. 8950–8961.

[209] R. Hu, Q. Huang, S. Chang, H. Wang, and J. He, "The mbpep: a deep ensemble pruning algorithm providing high quality uncertainty prediction," *Applied Intelligence*, vol. 49, no. 8, pp. 2942–2955, 2019.

[210] A. Malinin and M. Gales, "Uncertainty in structured prediction," *arXiv preprint arXiv:2002.07650*, 2020.

[211] A. Malinin, B. Mlodozeniec, and M. Gales, "Ensemble distribution distillation," *arXiv preprint arXiv:1905.00076*, 2019.

[212] A. Ashukha, A. Lyzhov, D. Molchanov, and D. Vetrov, "Pitfalls of in-domain uncertainty estimation and ensembling in deep learning," *arXiv preprint arXiv:2002.06470*, 2020.

[213] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using convolutional neural networks with test-time augmentation," in *International MICCAI Brainlesion Workshop*. Springer, 2018, pp. 61–72.

[214] A. Lyzhov, Y. Molchanova, A. Ashukha, D. Molchanov, and D. Vetrov, "Greedy policy search: A simple baseline for learnable test-time augmentation," in *Conference on Uncertainty in Artificial Intelligence*. PMLR, 2020, pp. 1308–1317.

[215] D. Shanmugam, D. Blalock, G. Balakrishnan, and J. Guttag, "When and why test-time augmentation works," *arXiv preprint arXiv:2011.11156*, 2020.

[216] A. G. Wilson and P. Izmailov, "Bayesian deep learning and a probabilistic perspective of generalization," in *Advances in Neural Information Processing Systems 31*, 2020, pp. 1–12.

[217] R. Pop and P. Fulop, "Deep ensemble bayesian active learning: Addressing the mode collapse issue in monte carlo dropout via ensembles," *arXiv preprint arXiv:1811.03897*, 2018.

[218] V. TV, P. Malhotra, L. Vig, G. Shroff *et al.*, "Data-driven prognostics with predictive uncertainty estimation using ensemble of deep ordinal regression models," *arXiv preprint arXiv:1903.09795*, 2019.

[219] S. Sinha, H. Bharadhwaj, A. Goyal, H. Larochelle, A. Garg, and F. Shkurti, "Dibs: Diversity inducing information bottleneck in model ensembles," *arXiv preprint arXiv:2003.04514*, 2020.

[220] T. Pearce, A. Brintrup, M. Zaki, and A. Neely, "High-quality prediction intervals for deep learning: A distribution-free, ensembled approach," in *International Conference on Machine Learning*. PMLR, 2018, pp. 4075–4084.

[221] L. Ambrogioni, U. Guclu, and M. van Gerven, "Wasserstein variational gradient descent: From semi-discrete optimal transport to ensemble variational inference," *arXiv preprint arXiv:1811.02827*, 2018.

[222] S. Hu, N. Pezzotti, D. Mavroeidis, and M. Welling, "Simple and accurate uncertainty quantification from bias-variance decomposition," *arXiv preprint arXiv:2002.05582*, 2020.

[223] E. A. Antonelo, E. Camponogara, and B. Foss, "Echo state networks for data-driven downhole pressure estimation in gas-lift oil wells," *Neural Networks*, vol. 85, pp. 106–117, 2017.

[224] Q. Ma, L. Shen, and G. W. Cottrell, "Deep-esn: A multiple projection-encoding hierarchical reservoir computing framework," *arXiv preprint arXiv:1711.05255*, 2017.

[225] Z. Fan, X. Song, T. Xia, R. Jiang, R. Shibasaki, and R. Sakuramachi, "Online deep ensemble learning for predicting citywide human mobility," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, pp. 1–21, 2018.

[226] Y. Yang, W. Hong, and S. Li, "Deep ensemble learning based probabilistic load forecasting in smart grids," *Energy*, vol. 189, p. 116324, 2019.

[227] J. van Amersfoort, L. Smith, Y. W. Teh, and Y. Gal, "Simple and scalable epistemic uncertainty estimation using a single deep deterministic neural network," *arXiv preprint arXiv:2003.02037*, 2020.

[228] E. Fersini, E. Messina, and F. A. Pozzi, "Sentiment analysis: Bayesian ensemble learning," *Decision support systems*, vol. 68, pp. 26–38, 2014.

[229] T. Pearce, M. Zaki, A. Brintrup, N. Anastassacos, and A. Neely, "Uncertainty in neural networks: Bayesian ensembling," *arXiv preprint arXiv:1810.05546*, 2018.

[230] R. Pop and P. Fulop, "Deep ensemble bayesian active learning," in *Bayesian Deep Learning Workshop at NeurIPS*, 2020.

[231] T. Pearce, F. Leibfried, A. Brintrup, M. Zaki, and A. Neely, "Uncertainty in neural networks: Approximately bayesian ensembling," in *The 23rd International Conference on Artificial Intelligence and Statistics, AISTATS 2020*, 2020.

[232] C. Tzelepis, V. Mezaris, and I. Patras, "Linear maximum margin classifier for learning from uncertain data," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 12, pp. 2948–2962, 2017.

[233] T. Kanamori, A. Takeda, and T. Suzuki, "Conjugate relation between loss functions and uncertainty sets in classification problems," *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 1461–1504, 2013.

[234] T. Pereira, S. Cardoso, M. Guerreiro, S. C. Madeira, A. D. N. Initiative *et al.*, "Targeting the uncertainty of predictions at patient-level using an ensemble of classifiers coupled with calibration methods, venn-abers, and conformal predictors: A case study in ad," *Journal of Biomedical Informatics*, vol. 101, p. 103350, 2020.

[235] I. Partalas, G. Tsoumakas, and I. Vlahavas, "An ensemble uncertainty aware measure for directed hill climbing ensemble pruning," *Machine Learning*, vol. 81, no. 3, pp. 257–282, 2010.

[236] A. A. Peterson, R. Christensen, and A. Khorshidi, "Addressing uncertainty in atomistic machine learning," *Physical Chemistry Chemical Physics*, vol. 19, no. 18, pp. 10 978–10 985, 2017.

[237] R. Ardywibowo, S. Boluki, X. Gong, Z. Wang, and X. Qian, "Nads: Neural architecture distribution search for uncertainty awareness," *arXiv preprint arXiv:2006.06646*, 2020.

[238] H. Kabir, A. Khosravi, A. Kavousi-Fard, S. Nahavandi, and D. Srinivasan, "Optimal uncertainty-guided neural network training," *arXiv preprint arXiv:1912.12761*, 2019.

[239] Y. Geifman, G. Uziel, and R. El-Yaniv, "Bias-reduced uncertainty estimation for deep neural classifiers," *arXiv preprint arXiv:1805.08206*, 2018.

[240] N. Tagasovska and D. Lopez-Paz, "Single-model uncertainties for deep learning," in *Advances in Neural Information Processing Systems*, 2019, pp. 6414–6425.

[241] J. van Amersfoort, L. Smith, Y. W. Teh, and Y. Gal, "Uncertainty estimation using a single deep deterministic neural network," in *Proceedings of the 37th International Conference on Machine Learning*, 2020.

[242] N. Tagasovska and D. Lopez-Paz, "Frequentist uncertainty estimates for deep learning," *arXiv preprint arXiv:1811.00908*, 2018.

[243] A. Mobiny, H. V. Nguyen, S. Moulik, N. Garg, and C. C. Wu, "Dropconnect is effective in modeling uncertainty of bayesian deep networks," *arXiv preprint arXiv:1906.04569*, 2019.

[244] Y. Chen, Y. Xue, Y. Ma, and F. Xu, "Measures of uncertainty for neighborhood rough sets," *Knowledge-Based Systems*, vol. 120, pp. 226–235, 2017.

[245] V. Kuleshov, N. Fenner, and S. Ermon, "Accurate uncertainties for deep learning using calibrated regression," *arXiv preprint arXiv:1807.00263*, 2018.

[246] T. G. Rudner, F. Wenzel, Y. W. Teh, and Y. Gal, "The natural neural tangent kernel: Neural network training dynamics under natural gradient descent," in *Fourht workshop on Bayesian Deep Learning (NeurIPS 2019)*, 2019.

[247] B. N. Patro, M. Lunayach, S. Patel, and V. P. Namboodiri, "U-cam: Visual explanation using uncertainty based class activation maps," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 7444–7453.

[248] S. Depeweg, J. M. Hernández-Lobato, F. Doshi-Velez, and S. Udluft, "Decomposition of uncertainty in bayesian deep learning for efficient and risk-sensitive learning," *arXiv preprint arXiv:1710.07283*, 2017.

[249] R. Y. Rohekar, Y. Gurwicz, S. Nisimov, and G. Novik, "Modeling uncertainty by learning a hierarchy of deep neural connections," in *Advances in Neural Information Processing Systems*, 2019, pp. 4246–4256.

[250] S. Khan, M. Hayat, S. W. Zamir, J. Shen, and L. Shao, "Striking the right balance with uncertainty," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 103–112.

[251] N. Pawlowski, A. Brock, M. C. Lee, M. Rajchl, and B. Glocker, "Implicit weight uncertainty in neural networks," *arXiv preprint arXiv:1711.01297*, 2017.

[252] X. Wang, Y. Luo, D. Crankshaw, A. Tumanov, F. Yu, and J. E. Gonzalez, "Idk cascades: Fast deep learning by learning not to overthink," *arXiv preprint arXiv:1706.00885*, 2017.

[253] Y. Yang and P. Perdikaris, "Adversarial uncertainty quantification in physics-informed neural networks," *Journal of Computational Physics*, vol. 394, pp. 136–152, 2019.

[254] F. Brickwedde, S. Abraham, and R. Mester, "Mono-sf: Multi-view geometry meets single-view depth for monocular scene flow estimation of dynamic traffic scenes," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 2780–2790.

[255] S. Thulasidasan, G. Chennupati, J. A. Bilmes, T. Bhattacharya, and S. Michalak, "On mixup training: Improved calibration and predictive uncertainty for deep neural networks," in *Advances in Neural Information Processing Systems*, 2019, pp. 13 888–13 899.

[256] Y. Zhang, K. Song, Y. Sun, S. Tan, and M. Udell, "why should you trust my explanation?" understanding uncertainty in lime explanations," *arXiv preprint arXiv:1904.12991*, 2019.

[257] F. Sheikholeslami, S. Jain, and G. B. Giannakis, "Minimum uncertainty based detection of adversaries in deep neural networks," *arXiv preprint arXiv:1904.02841*, 2019.

[258] Y. Li, L. Liu, and R. T. Tan, "Decoupled certainty-driven consistency loss for semi-supervised learning," *arXiv*, pp. arXiv–1901, 2019.

[259] E. Englesson and H. Azizpour, "Efficient evaluation-time uncertainty estimation by improved distillation," *arXiv preprint arXiv:1906.05419*, 2019.

[260] B. C. Csáji and K. B. Kis, "Distribution-free uncertainty quantification for kernel methods by gradient perturbations," *Machine Learning*, vol. 108, no. 8-9, pp. 1677–1699, 2019.

[261] D. Hendrycks, K. Lee, and M. Mazeika, "Using pre-training can improve model robustness and uncertainty," *arXiv preprint arXiv:1901.09960*, 2019.

[262] B. S. Veeling, R. v. d. Berg, and M. Welling, "Predictive uncertainty through quantization," *arXiv preprint arXiv:1810.05500*, 2018.

[263] F. J. Bragman, R. Tanno, Z. Eaton-Rosen, W. Li, D. J. Hawkes, S. Ourselin, D. C. Alexander, J. R. McClelland, and M. J. Cardoso, "Quality control in radiotherapy-treatment planning using multi-task learning and uncertainty estimation," in *International Conference on Medical Imaging with Deep Learning*, 2018.

[264] Z. Dzunic and J. Fisher III, "Bayesian switching interaction analysis under uncertainty," in *Artificial Intelligence and Statistics*, 2014, pp. 220–228.

[265] A. K. Balan, V. Rathod, K. P. Murphy, and M. Welling, "Bayesian dark knowledge," in *Advances in Neural Information Processing Systems*, 2015, pp. 3438–3446.

[266] R. Houthooft, X. Chen, Y. Duan, J. Schulman, F. De Turck, and P. Abbeel, "Vime: Variational information maximizing exploration," in *Advances in Neural Information Processing Systems*, 2016, pp. 1109–1117.

[267] J. T. Springenberg, A. Klein, S. Falkner, and F. Hutter, "Bayesian optimization with robust bayesian neural networks," in *Advances in neural information processing systems*, 2016, pp. 4134–4142.

[268] B. Lakshminarayanan, D. M. Roy, and Y. W. Teh, "Mondrian forests for large-scale regression when uncertainty matters," in *Artificial Intelligence and Statistics*, 2016, pp. 1478–1487.

[269] E. Ilg, Ö. Çiçek, S. Galesso, A. Klein, O. Makansi, F. Hutter, and T. Brox, "Uncertainty estimates for optical flow with multi-hypotheses networks," *arXiv preprint arXiv:1802.07095*, p. 81, 2018.

[270] J. Heo, H. B. Lee, S. Kim, J. Lee, K. J. Kim, E. Yang, and S. J. Hwang, "Uncertainty-aware attention for reliable interpretation and prediction," in *Advances in Neural Information Processing Systems*, 2018, pp. 909–918.

[271] P. Henderson, T. Doan, R. Islam, and D. Meger, "Bayesian policy gradients via alpha divergence dropout inference," *arXiv preprint arXiv:1712.02037*, 2017.

[272] H. Ahn, S. Cha, D. Lee, and T. Moon, "Uncertainty-based continual learning with adaptive regularization," in *Advances in Neural Information Processing Systems*, 2019, pp. 4394–4404.

[273] R. Zhang, C. Li, C. Chen, and L. Carin, "Learning structural weight uncertainty for sequential decision-making," *arXiv preprint arXiv:1801.00085*, 2017.

[274] M. Sensoy, L. Kaplan, and M. Kandemir, "Evidential deep learning to quantify classification uncertainty," in *Advances in Neural Information Processing Systems*, 2018, pp. 3179–3189.

[275] L. Acerbi, "Variational bayesian monte carlo," in *Advances in Neural Information Processing Systems*, 2018, pp. 8213–8223.

[276] K. Tóthová, S. Parisot, M. C. Lee, E. Puyol-Antón, L. M. Koch, A. P. King, E. Konukoglu, and M. Pollefeys, "Uncertainty quantification in cnn-based surface prediction using shape priors," in *International Workshop on Shape in Medical Imaging*. Springer, 2018, pp. 300–310.

[277] M. Haussmann, S. Gerwinn, and M. Kandemir, "Bayesian prior networks with pac training," *arXiv preprint arXiv:1906.00816*, 2019.

[278] G. De Ath, R. M. Everson, J. E. Fieldsend, and A. A. Rahat, "ϵ-shotgun: ϵ-greedy batch bayesian optimisation," *arXiv preprint arXiv:2002.01873*, 2020.

[279] A. Y. Foong, Y. Li, J. M. Hernández-Lobato, and R. E. Turner, "'in-between'uncertainty in bayesian neural networks," *arXiv preprint arXiv:1906.11537*, 2019.

[280] A. Hasanzadeh, E. Hajiramezanali, S. Boluki, M. Zhou, N. Duffield, K. Narayanan, and X. Qian, "Bayesian graph neural networks with adaptive connection sampling," *arXiv preprint arXiv:2006.04064*, 2020.

[281] O. Chang, Y. Yao, D. Williams-King, and H. Lipson, "Ensemble model patching: A parameter-efficient variational bayesian neural network," *arXiv preprint arXiv:1905.09453*, 2019.

[282] C. Stoean, R. Stoean, M. Atencia, M. Abdar, L. Velázquez-Pérez, A. Khosravi, S. Nahavandi, U. R. Acharya, and G. Joya, "Automated detection of presymptomatic conditions in spinocerebellar ataxia type 2 using monte carlo dropout and deep neural network techniques with electrooculogram signals," *Sensors*, vol. 20, no. 11, p. 3032, 2020.

[283] T. Z. Xiao, A. N. Gomez, and Y. Gal, "Wat zei je? detecting out-of-distribution translations with variational transformers," *arXiv preprint arXiv:2006.08344*, 2020.

[284] A. Repetti, M. Pereyra, and Y. Wiaux, "Scalable bayesian uncertainty quantification in imaging inverse problems via convex optimization," *SIAM Journal on Imaging Sciences*, vol. 12, no. 1, pp. 87–118, 2019.

[285] L. Cardelli, M. Kwiatkowska, L. Laurenti, N. Paoletti, A. Patane, and M. Wicker, "Statistical guarantees for the robustness of bayesian neural networks," *arXiv preprint arXiv:1903.01980*, 2019.

[286] H. B. Moss, D. S. Leslie, and P. Rayson, "Mumbo: Multi-task max-value bayesian optimization," *arXiv preprint arXiv:2006.12093*, 2020.

[287] V. Dutordoir, M. Wilk, A. Artemev, and J. Hensman, "Bayesian image classification with deep convolutional gaussian processes," in *International Conference on Artificial Intelligence and Statistics*, 2020, pp. 1529–1539.

[288] Y. Luo, Z. Huang, Z. Zhang, Z. Wang, M. Baktashmotlagh, and Y. Yang, "Learning from the past: Continual meta-learning via bayesian graph modeling," *arXiv preprint arXiv:1911.04695*, 2019.

[289] Y. Gafni, R. Lavi, and M. Tennenholtz, "Vcg under sybil (false-name) attacks-a bayesian analysis." in *AAAI*, 2020, pp. 1966–1973.

[290] X. Jin, C. Lan, W. Zeng, and Z. Chen, "Uncertainty-aware multi-shot knowledge distillation for image-based object re-identification," *arXiv preprint arXiv:2001.05197*, 2020.

[291] L. Han, R. Gao, M. Kim, X. Tao, B. Liu, and D. N. Metaxas, "Robust conditional gan from uncertainty-aware pairwise comparisons." in *AAAI*, 2020, pp. 10 909–10 916.

[292] R. Stoean, C. Stoean, M. Atencia, R. Rodríguez-Labrada, and G. Joya, "Ranking information extracted from uncertainty quantification of the prediction of a deep learning model on medical time series data," *Mathematics*, vol. 8, no. 7, p. 1078, 2020.

[293] C. Oh, K. Adamczewski, and M. Park, "Radial and directional posteriors for bayesian deep learning." in *AAAI*, 2020, pp. 5298–5305.

[294] M. W. Dusenberry, G. Jerfel, Y. Wen, Y.-a. Ma, J. Snoek, K. Heller, B. Lakshminarayanan, and D. Tran, "Efficient and scalable bayesian neural nets with rank-1 factors," *arXiv preprint arXiv:2005.07186*, 2020.

[295] M. Havasi, J. Snoek, D. Tran, J. Gordon, and J. M. Hernández-Lobato, "Refining the variational posterior through iterative optimization," in *International Conference on Learning Representations*, 2019.

[296] R. Krishnan, M. Subedar, and O. Tickoo, "Specifying weight priors in bayesian deep neural networks with empirical bayes." in *AAAI*, 2020, pp. 4477–4484.

[297] A. Filos, P. Tigas, R. McAllister, N. Rhinehart, S. Levine, and Y. Gal, "Can autonomous vehicles identify, recover from, and adapt to distribution shifts?" *arXiv preprint arXiv:2006.14911*, 2020.

[298] Y. Huang, W. Huang, L. Li, and Z. Li, "Meta-learning pac-bayes priors in model averaging." in *AAAI*, 2020, pp. 4198–4205.

[299] R. Amit and R. Meir, "Meta-learning by adjusting priors based on extended pac-bayes theory," in *International Conference on Machine Learning*, 2018, pp. 205–214.

[300] A. Bhattacharyya, M. Fritz, and B. Schiele, "Bayesian prediction of future street scenes through importance sampling based optimization," *arXiv preprint arXiv:1806.06939*, 2018.

[301] J. Yao, W. Pan, S. Ghosh, and F. Doshi-Velez, "Quality of uncertainty quantification for bayesian neural network inference," *arXiv preprint arXiv:1906.09686*, 2019.

[302] M.-H. Laves, S. Ihler, K.-P. Kortmann, and T. Ortmaier, "Calibration of model uncertainty for dropout variational inference," *arXiv preprint arXiv:2006.11584*, 2020.

[303] W. Yang, L. Lorch, M. A. Graule, S. Srinivasan, A. Suresh, J. Yao, M. F. Pradier, and F. Doshi-Velez, "Output-constrained bayesian neural networks," *arXiv preprint arXiv:1905.06287*, 2019.

[304] S. Thakur, C. Lorsung, Y. Yacoby, F. Doshi-Velez, and W. Pan, "Learned uncertainty-aware (luna) bases for bayesian regression using multi-headed auxiliary networks," *arXiv preprint arXiv:2006.11695*, 2020.

[305] Y. Yacoby, W. Pan, and F. Doshi-Velez, "Learning deep bayesian latent variable regression models that generalize: When non-identifiability is a problem," *arXiv preprint arXiv:1911.00569*, 2019.

[306] M. A. Masood and F. Doshi-Velez, "A particle-based variational approach to bayesian non-negative matrix factorization." *J. Mach. Learn. Res.*, vol. 20, pp. 90–1, 2019.

[307] M. Abdolshah, A. Shilton, S. Rana, S. Gupta, and S. Venkatesh, "Cost-aware multi-objective bayesian optimisation," *arXiv preprint arXiv:1909.03600*, 2019.

[308] C. White, W. Neiswanger, and Y. Savani, "Bananas: Bayesian optimization with neural architectures for neural architecture search," *arXiv preprint arXiv:1910.11858*, 2019.

[309] M. Balandat, B. Karrer, D. R. Jiang, S. Daulton, B. Letham, A. G. Wilson, and E. Bakshy, "Botorch: Programmable bayesian optimization in pytorch," *arXiv preprint arXiv:1910.06403*, 2019.

[310] T. Galy-Fajou, F. Wenzel, C. Donner, and M. Opper, "Multi-class gaussian process classification made conjugate: Efficient inference via data augmentation," *arXiv preprint arXiv:1905.09670*, 2019.

[311] H. B. Lee, H. Lee, D. Na, S. Kim, M. Park, E. Yang, and S. J. Hwang, "Learning to balance: Bayesian meta-learning for imbalanced and out-of-distribution tasks," *arXiv preprint arXiv:1905.12917*, 2019.

[312] M. P. Vadera and B. M. Marlin, "Assessing the robustness of bayesian dark knowledge to posterior uncertainty," *arXiv preprint arXiv:1906.01724*, 2019.

[313] A. Siahkoohi, G. Rizzuti, and F. J. Herrmann, "A deep-learning based bayesian approach to seismic imaging and uncertainty quantification," *arXiv preprint arXiv:2001.04567*, 2020.

[314] S. Sun, G. Zhang, J. Shi, and R. Grosse, "Functional variational bayesian neural networks," *arXiv preprint arXiv:1903.05779*, 2019.

[315] M. Patacchiola, J. Turner, E. J. Crowley, M. O'Boyle, and A. Storkey, "Deep kernel transfer in gaussian processes for few-shot learning," *arXiv preprint arXiv:1910.05199*, 2019.

[316] Z. Cheng, M. Gadelha, S. Maji, and D. Sheldon, "A bayesian perspective on the deep image prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5443–5451.

[317] R. Oliveira, L. Ott, and F. Ramos, "Bayesian optimisation under uncertain inputs," *arXiv preprint arXiv:1902.07908*, 2019.

[318] J. Caldeira and B. Nord, "Deeply uncertain: Comparing methods of uncertainty quantification in deep learning algorithms," *Machine Learning: Science and Technology*, vol. 2, no. 1, pp. 1–9, 2020.

[319] L. Wandzik, R. V. Garcia, and J. Krüger, "Uncertainty quantification in deep residual neural networks," *arXiv preprint arXiv:2007.04905*, 2020.

[320] Z. Deng, Y. Luo, J. Zhu, and B. Zhang, "Dbsn: Measuring uncertainty through bayesian learning of deep neural network structures," *arXiv preprint arXiv:1911.09804*, 2019.

[321] J. González-López, S. Ventura, and A. Cano, "Distributed selection of continuous features in multilabel classification using mutual information," *IEEE transactions on neural networks and learning systems*, 2019.

[322] A. Y. Foong, W. P. Bruinsma, J. Gordon, Y. Dubois, J. Requeima, and R. E. Turner, "Meta-learning stationary stochastic process prediction with convolutional neural processes," *arXiv preprint arXiv:2007.01332*, 2020.

[323] Y. Yao, A. Vehtari, and A. Gelman, "Stacking for non-mixing bayesian computations: The curse and blessing of multimodal posteriors," *arXiv preprint arXiv:2006.12335*, 2020.

[324] D. S. Prijatelj, M. McCurrie, and W. J. Scheirer, "A bayesian evaluation framework for ground truth-free visual recognition tasks," *arXiv preprint arXiv:2007.06711*, 2020.

[325] L. Herzog, E. Murina, O. Dürr, S. Wegener, and B. Sick, "Integrating uncertainty in deep neural networks for mri based stroke analysis," *Medical Image Analysis*, p. 101790, 2020.

[326] S. Prokudin, P. Gehler, and S. Nowozin, "Deep directional statistics: Pose estimation with uncertainty quantification," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 534–551.

[327] R. Tuo and W. Wang, "Uncertainty quantification for bayesian optimization," *arXiv preprint arXiv:2002.01569*, 2020.

[328] L. Acerbi, "Variational bayesian monte carlo with noisy likelihoods," *arXiv preprint arXiv:2006.08655*, 2020.

[329] J. Zhao, S. He, and S. Sun, "Probabilistic inference of bayesian neural networks with generalized expectation propagation," *Neurocomputing*, 2020.

[330] Y. Li, S. Rao, A. Hassaine, R. Ramakrishnan, Y. Zhu, D. Canoy, G. Salimi-Khorshidi, T. Lukasiewicz, and K. Rahimi, "Deep bayesian gaussian processes for uncertainty estimation in electronic health records," *arXiv preprint arXiv:2003.10170*, 2020.

[331] A. Jacot, F. Gabriel, and C. Hongler, "Neural tangent kernel: Convergence and generalization in neural networks," in *Advances in neural information processing systems*, 2018, pp. 8571–8580.

[332] Y. Wang and V. Ročková, "Uncertainty quantification for sparse deep learning," *arXiv preprint arXiv:2002.11815*, 2020.

[333] A. Jesson, S. Mindermann, U. Shalit, and Y. Gal, "Identifying causal-effect inference failure with uncertainty-aware models," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[334] F. De Sousa Ribeiro, G. Leontidis, and S. Kollias, "Introducing routing uncertainty in capsule networks," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[335] X. Zhao, F. Chen, S. Hu, and J.-H. Cho, "Uncertainty aware semi-supervised learning on graph data," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[336] S. Mukherjee and A. Awadallah, "Uncertainty-aware self-training for few-shot text classification," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[337] P. Hu, S. Sclaroff, and K. Saenko, "Uncertainty-aware learning for zero-shot semantic segmentation," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[338] W. Yang, L. Lorch, M. Graule, H. Lakkaraju, and F. Doshi-Velez, "Incorporating interpretable output constraints in bayesian neural networks," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[339] S. Farquhar, L. Smith, and Y. Gal, "Liberty or depth: Deep bayesian neural nets do not need complex weight posterior approximations," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[340] B. Charpentier, D. Zügner, and S. Günnemann, "Posterior network: Uncertainty estimation without ood samples via density-based pseudo-counts," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[341] T. Guénais, D. Vamvourellis, Y. Yacoby, F. Doshi-Velez, and W. Pan, "Bacoun: Bayesian classifiers with out-of-distribution uncertainty," *arXiv preprint arXiv:2007.06096*, 2020.

[342] S. Li, W. Xing, R. Kirby, and S. Zhe, "Multi-fidelity bayesian optimization via deep neural networks," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[343] C. Lyle, L. Schut, R. Ru, Y. Gal, and M. van der Wilk, "A bayesian perspective on training speed and model selection," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[344] M. Lee and J. Seok, "Estimation with uncertainty via conditional generative adversarial networks," *arXiv preprint arXiv:2007.00334*, 2020.

[345] X. Fan, S. Zhang, B. Chen, and M. Zhou, "Bayesian attention modules," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[346] L. Chauhan, J. Alberg, and Z. Lipton, "Uncertainty-aware lookahead factor models for quantitative investing," in *International Conference on Machine Learning*. PMLR, 2020, pp. 1489–1499.

[347] T. Zhou, Y. Li, Y. Wu, and D. Carlson, "Estimating uncertainty intervals from collaborating networks," *arXiv preprint arXiv:2002.05212*, 2020.

[348] A. Chan, A. Alaa, Z. Qian, and M. Van Der Schaar, "Unlabelled data improves bayesian uncertainty calibration under covariate shift," in *International Conference on Machine Learning*. PMLR, 2020, pp. 1392–1402.

[349] Z. Wang and M. Zhou, "Thompson sampling via local uncertainty," in *International Conference on Machine Learning*. PMLR, 2020, pp. 10 115–10 125.

[350] T. Joo, U. Chung, and M.-G. Seo, "Being bayesian about categorical probability," *arXiv preprint arXiv:2002.07965*, 2020.

[351] Q. Wang and H. Van Hoof, "Doubly stochastic variational inference for neural processes with hierarchical latent variables," in *International Conference on Machine Learning*. PMLR, 2020, pp. 10 018–10 028.

[352] H. J. Hortúa, L. Malago, and R. Volpi, "Constraining the reionization history using bayesian normalizing flows," *arXiv preprint arXiv:2005.07694*, 2020.

[353] Z. Lyu, D. Duolikun, B. Dai, Y. Yao, P. Minervini, T. Z. Xiao, and Y. Gal, "You need only uncertain answers: Data efficient multilingual question answering," in *TWorkshop on Uncertainty and Ro-bustness in Deep Learning*, 2020.

[354] P. Notin, J. M. Hernández-Lobato, and Y. Gal, "Principled uncertainty estimation for high dimensional data," in *TWorkshop on Uncertainty and Ro-bustness in Deep Learning*, 2020.

[355] M. Jarvenpaa, A. Vehtari, and P. Marttinen, "Batch simulations and uncertainty quantification in gaussian process surrogate approximate bayesian computation," in *Conference on Uncertainty in Artificial Intelligence*. PMLR, 2020, pp. 779–788.

[356] J. Huggins, M. Kasprzak, T. Campbell, and T. Broderick, "Validated variational inference via practical posterior error bounds," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2020, pp. 1792–1802.

[357] S. Boluki, R. Ardywibowo, S. Z. Dadaneh, M. Zhou, and X. Qian, "Learnable bernoulli dropout for bayesian deep learning," *arXiv preprint arXiv:2002.05155*, 2020.

[358] R. Barbano, C. Zhang, S. Arridge, and B. Jin, "Quantifying model uncertainty in inverse problems via bayesian deep gradient descent," *arXiv preprint arXiv:2007.09971*, 2020.

[359] F. Wenzel, K. Roth, B. S. Veeling, J. Swikatkowski, L. Tran, S. Mandt, J. Snoek, T. Salimans, R. Jenatton, and S. Nowozin, "How good is the bayes posterior in deep neural networks really?" *arXiv preprint arXiv:2002.02405*, 2020.

[360] S. Suzuki, S. Takeno, T. Tamura, K. Shitara, and M. Karasuyama, "Multi-objective bayesian optimization using pareto-frontier entropy," in *International Conference on Machine Learning*. PMLR, 2020, pp. 9279–9288.

[361] J. Buckman, D. Hafner, G. Tucker, E. Brevdo, and H. Lee, "Sample-efficient reinforcement learning with stochastic ensemble value expansion," in *Advances in Neural Information Processing Systems*, 2018, pp. 8224–8234.

[362] B. He, B. Lakshminarayanan, and Y. W. Teh, "Bayesian deep ensembles via the neural tangent kernel," *arXiv preprint arXiv:2007.05864*, 2020.

[363] P. Schwab and W. Karlen, "Cxplain: Causal explanations for model interpretation under uncertainty," in *Advances in Neural Information Processing Systems*, 2019, pp. 10 220–10 230.

[364] L. Smith and Y. Gal, "Understanding measures of uncertainty for adversarial example detection," *arXiv preprint arXiv:1803.08533*, 2018.

[365] A. Malinin and M. Gales, "Reverse kl-divergence training of prior networks: Improved uncertainty and adversarial robustness," in *Advances in Neural Information Processing Systems*, 2019, pp. 14 520–14 531.

[366] S. Jain, G. Liu, J. Mueller, and D. Gifford, "Maximizing overall diversity for improved uncertainty estimates in deep ensembles." in *AAAI*, 2020, pp. 4264–4271.

[367] M. Valdenegro-Toro, "Deep sub-ensembles for fast uncertainty estimation in image classification," *arXiv preprint arXiv:1910.08168*, 2019.

[368] J. Juraska, P. Karagiannis, K. K. Bowden, and M. A. Walker, "A deep ensemble model with slot alignment for sequence-to-sequence natural language generation," *arXiv preprint arXiv:1805.06553*, 2018.

[369] M.-h. Oh, P. A. Olsen, and K. N. Ramamurthy, "Crowd counting with decomposed uncertainty." in *AAAI*, 2020, pp. 11 799–11 806.

[370] K. E. Brown, F. A. Bhuiyan, and D. A. Talbert, "Uncertainty quantification in multimodal ensembles of deep learners," in *The Thirty-Third International Flairs Conference*, 2020.

[371] T. S. Salem, H. Langseth, and H. Ramampiaro, "Prediction intervals: Split normal mixture from quality-driven deep ensembles," *arXiv preprint arXiv:2007.09670*, 2020.

[372] Y. Wen, G. Jerfel, R. Muller, M. W. Dusenberry, J. Snoek, B. Lakshminarayanan, and D. Tran, "Improving calibration of batchensemble with data augmentation," in *TWorkshop on Uncertainty and Ro-bustness in Deep Learning*, 2020.

[373] F. Wenzel, J. Snoek, D. Tran, and R. Jenatton, "Hyperparameter ensembles for robustness and uncertainty quantification," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[374] L. Wang, D. Ghosh, M. Gonzalez Diaz, A. Farahat, M. Alam, C. Gupta, J. Chen, and M. Marathe, "Wisdom of the ensemble: Improving consistency of deep learning models," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[375] P. Grönquist, C. Yao, T. Ben-Nun, N. Dryden, P. Dueben, S. Li, and T. Hoefler, "Deep learning for post-processing ensemble weather forecasts," *arXiv preprint arXiv:2005.08748*, 2020.

[376] Q. Lu, G. Karanikolas, Y. Shen, and G. B. Giannakis, "Ensemble gaussian processes with spectral features for online interactive learning with scalability," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2020, pp. 1910–1920.

[377] T. Duan, A. Anand, D. Y. Ding, K. K. Thai, S. Basu, A. Ng, and A. Schuler, "Ngboost: Natural gradient boosting for probabilistic prediction," in *International Conference on Machine Learning*. PMLR, 2020, pp. 2690–2700.

[378] Z. Qin and D. Kim, "Rethinking softmax with cross-entropy: Neural network classifier as mutual information estimator," *arXiv preprint arXiv:1911.10688*, 2019.

[379] Q. Wu, H. Li, L. Li, and Z. Yu, "Quantifying intrinsic uncertainty in classification via deep dirichlet mixture networks," *arXiv preprint arXiv:1906.04450*, 2019.

[380] Q. Qian, J. Tang, H. Li, S. Zhu, and R. Jin, "Large-scale distance metric learning with uncertainty," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8542–8550.

[381] A. N. Gomez, I. Zhang, K. Swersky, Y. Gal, and G. E. Hinton, "Learning sparse networks using targeted dropout," *arXiv preprint arXiv:1905.13678*, 2019.

[382] A. Malinin and M. Gales, "Predictive uncertainty estimation via prior networks," in *Advances in Neural Information Processing Systems*, 2018, pp. 7047–7058.

[383] M. M. Dunlop, M. A. Girolami, A. M. Stuart, and A. L. Teckentrup, "How deep are deep gaussian processes?" *The Journal of Machine Learning Research*, vol. 19, no. 1, pp. 2100–2145, 2018.

[384] D. Hendrycks, M. Mazeika, S. Kadavath, and D. Song, "Using self-supervised learning can improve model robustness and uncertainty," in *Advances in Neural Information Processing Systems*, 2019, pp. 15 663–15 674.

[385] A. Kumar, P. S. Liang, and T. Ma, "Verified uncertainty calibration," *Advances in Neural Information Processing Systems*, vol. 32, pp. 3792–3803, 2019.

[386] D. Hendrycks, N. Mu, E. D. Cubuk, B. Zoph, J. Gilmer, and B. Lakshminarayanan, "Augmix: A simple data processing method to improve robustness and uncertainty," *arXiv preprint arXiv:1912.02781*, 2019.

[387] M. Możejko, M. Susik, and R. Karczewski, "Inhibited softmax for uncertainty estimation in neural networks," *arXiv preprint arXiv:1810.01861*, 2018.

[388] X. Xie, L. Ma, H. Wang, Y. Li, Y. Liu, and X. Li, "Diffchaser: Detecting disagreements for deep neural networks." in *IJCAI*, 2019, pp. 5772–5778.

[389] A. Boiarov, O. Granichin, and O. Granichina, "Simultaneous perturbation stochastic approximation for few-shot learning," *arXiv preprint arXiv:2006.05152*, 2020.

[390] C. Ye, Y. Li, and X. Zeng, "An improved deep network for tissue microstructure estimation with uncertainty quantification," *Medical Image Analysis*, vol. 61, p. 101650, 2020.

[391] M. Monteiro, L. L. Folgoc, D. C. de Castro, N. Pawlowski, B. Marques, K. Kamnitsas, M. van der Wilk, and B. Glocker, "Stochastic segmentation networks: Modelling spatially correlated aleatoric uncertainty," *arXiv preprint arXiv:2006.06015*, 2020.

[392] F. M. Maggi, M. Montali, and R. Peñaloza, "Temporal logics over finite traces with uncertainty." in *AAAI*, 2020, pp. 10 218–10 225.

[393] S. Amiri, M. S. Shirazi, and S. Zhang, "Learning and reasoning for robot sequential decision making under uncertainty." in *AAAI*, 2020, pp. 2726–2733.

[394] M. Sensoy, L. Kaplan, F. Cerutti, and M. Saleki, "Uncertainty-aware deep classifiers using generative models," *arXiv preprint arXiv:2006.04183*, 2020.

[395] S. Belakaria, A. Deshwal, and J. R. Doppa, "Uncertainty aware search framework for multi-objective bayesian optimization with constraints," *arXiv preprint arXiv:2008.07029*, 2020.

[396] Z.-Y. Liu, S.-Y. Li, S. Chen, Y. Hu, and S.-J. Huang, "Uncertainty aware graph gaussian process for semi-supervised learning." in *AAAI*, 2020, pp. 4957–4964.

[397] C. G. Northcutt, L. Jiang, and I. L. Chuang, "Confident learning: Estimating uncertainty in dataset labels," *arXiv preprint arXiv:1911.00068*, 2019.

[398] J. Manders, E. Marchiori, and T. van Laarhoven, "Simple domain adaptation with class prediction uncertainty alignment," *arXiv preprint arXiv:1804.04448*, vol. 1, no. 2, p. 3, 2018.

[399] S. Chun, S. J. Oh, S. Yun, D. Han, J. Choe, and Y. Yoo, "An empirical evaluation on robustness and uncertainty of regularization methods," *arXiv preprint arXiv:2003.03879*, 2020.

[400] R. Mehta, A. Filos, Y. Gal, and T. Arbel, "Uncertainty evaluation metric for brain tumour segmentation," *arXiv preprint arXiv:2005.14262*, 2020.

[401] J. Z. Liu, Z. Lin, S. Padhy, D. Tran, T. Bedrax-Weiss, and B. Lakshminarayanan, "Simple and principled uncertainty estimation with deterministic deep learning via distance awareness," *arXiv preprint arXiv:2006.10108*, 2020.

[402] A. Scillitoe, P. Seshadri, and M. Girolami, "Uncertainty quantification for data-driven turbulence modelling with mondrian forests," *arXiv preprint arXiv:2003.01968*, 2020.

[403] Y. Ovadia, E. Fertig, J. Ren, Z. Nado, D. Sculley, S. Nowozin, J. Dillon, B. Lakshminarayanan, and J. Snoek, "Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift," in *Advances in Neural Information Processing Systems*, 2019, pp. 13 991–14 002.

[404] M. Biloš, B. Charpentier, and S. Günnemann, "Uncertainty on asynchronous time event prediction," in *Advances in Neural Information Processing Systems*, 2019, pp. 12 851–12 860.

[405] Z. Zheng and Y. Yang, "Unsupervised scene adaptation with memory regularization in vivo," *arXiv preprint arXiv:1912.11164*, 2019.

[406] E. Zelikman and C. Healy, "Improving regression uncertainty estimates with an empirical prior," *arXiv preprint arXiv:2005.12496*, 2020.

[407] F. L. Da Silva, P. Hernandez-Leal, B. Kartal, and M. E. Taylor, "Uncertainty-aware action advising for deep reinforcement learning agents." in *AAAI*, 2020, pp. 5792–5799.

[408] J. J. Thiagarajan, B. Venkatesh, P. Sattigeri, and P.-T. Bremer, "Building calibrated deep models via uncertainty matching with auxiliary interval predictors." in *AAAI*, 2020, pp. 6005–6012.

[409] Q. Zhou, H. Li, and J. Wang, "Deep model-based reinforcement learning via estimated uncertainty and conservative policy optimization." in *AAAI*, 2020, pp. 6941–6948.

[410] K. Standvoss, S. C. Quax, and M. A. Van Gerven, "Visual attention through uncertainty minimization in recurrent generative models," *BioRxiv*, 2020.

[411] X. Wang, M. Long, J. Wang, and M. Jordan, "Transferable calibration with lower bias and variance in domain adaptation," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[412] A. Grover and S. Ermon, "Uncertainty autoencoders: Learning compressed representations via variational information maximization," in *The 22nd International Conference on Artificial Intelligence and Statistics*, 2019, pp. 2514–2524.

[413] F. Cakir, K. He, S. A. Bargal, and S. Sclaroff, "Hashing with mutual information," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 10, pp. 2424–2437, 2019.

[414] F. Cakir, K. He, S. Adel Bargal, and S. Sclaroff, "Mihash: Online hashing with mutual information," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 437–445.

[415] C. Yildiz, M. Heinonen, and H. Lahdesmaki, "Ode2vae: Deep generative second order odes with bayesian neural networks," in *Advances in Neural Information Processing Systems*, 2019, pp. 13 412–13 421.

[416] M. K. Titsias, J. Schwarz, A. G. d. G. Matthews, R. Pascanu, and Y. W. Teh, "Functional regularisation for continual learning with gaussian processes," *arXiv preprint arXiv:1901.11356*, 2019.

[417] J. Lee, Y. Bahri, R. Novak, S. S. Schoenholz, J. Pennington, and J. Sohl-Dickstein, "Deep neural networks as gaussian processes," *arXiv preprint arXiv:1711.00165*, 2017.

[418] S. Ravi and A. Beatson, "Amortized bayesian meta-learning," in *International Conference on Learning Representations*, 2018.

[419] C.-K. Lu, S. C.-H. Yang, X. Hao, and P. Shafto, "Interpretable deep gaussian processes with moments," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2020, pp. 613–623.

[420] B. Wang, J. Lu, Z. Yan, H. Luo, T. Li, Y. Zheng, and G. Zhang, "Deep uncertainty quantification: A machine learning approach for weather forecasting," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 2087–2095.

[421] L. Tai, P. Yun, Y. Chen, C. Liu, H. Ye, and M. Liu, "Visual-based autonomous driving deployment from a stochastic and uncertainty-aware perspective," *arXiv preprint arXiv:1903.00821*, 2019.

[422] R. Selvan, F. Faye, J. Middleton, and A. Pai, "Uncertainty quantification in medical image segmentation with normalizing flows," *arXiv preprint arXiv:2006.02683*, 2020.

[423] M. Poggi, F. Aleotti, F. Tosi, and S. Mattoccia, "On the uncertainty of self-supervised monocular depth estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3227–3237.

[424] P. Cui, W. Hu, and J. Zhu, "Calibrated reliable regression using maximum mean discrepancy," *arXiv preprint arXiv:2006.10255*, 2020.

[425] J. Lindinger, D. Reeb, C. Lippert, and B. Rakitsch, "Beyond the mean-field: Structured deep gaussian processes improve the predictive uncertainties," *arXiv preprint arXiv:2005.11110*, 2020.

[426] H. Wang, L. Xie, A. Cuozzo, S. Mak, and Y. Xie, "Uncertainty quantification for inferring hawkes networks," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[427] L. Meronen, C. Irwanto, and A. Solin, "Stationary activations for uncertainty calibration in deep learning," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[428] T. G. Rudner, O. Key, Y. Gal, and T. Rainforth, "On signal-to-noise ratio issues in variational inference for deep gaussian processes," *arXiv preprint arXiv:2011.00515*, 2020.

[429] Y. Zhao and M. Udell, "Matrix completion with quantified uncertainty through low rank gaussian copula," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[430] W. Shi, X. Zhao, F. Chen, and Q. Yu, "Multifaceted uncertainty estimation for label-efficient deep learning," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[431] A.-K. Kopetzki, B. Charpentier, D. Zügner, S. Giri, and S. Günnemann, "Evaluating robustness of predictive uncertainty estimation: Are dirichlet-based models reliable?" *arXiv preprint arXiv:2010.14986*, 2020.

[432] Y. Chung, W. Neiswanger, I. Char, and J. Schneider, "Beyond pinball loss: Quantile methods for calibrated uncertainty quantification," *arXiv preprint arXiv:2011.09588*, 2020.

[433] M. Finzi, R. Bondesan, and M. Welling, "Probabilistic numeric convolutional neural networks," *arXiv preprint arXiv:2010.10876*, 2020.

[434] A. Alaa and M. Van Der Schaar, "Frequentist uncertainty in recurrent neural networks via blockwise influence functions," in *International Conference on Machine Learning*. PMLR, 2020, pp. 175–190.

[435] W. Liu, X. Wang, J. Owens, and S. Y. Li, "Energy-based out-of-distribution detection," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[436] A. Aushev, H. Pesonen, M. Heinonen, J. Corander, and S. Kaski, "Likelihood-free inference with deep gaussian processes," *arXiv preprint arXiv:2006.10571*, 2020.

[437] J. Antorán, J. Allingham, and J. M. Hernández-Lobato, "Depth uncertainty in neural networks," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[438] Z. Huo, A. PakBin, X. Chen, N. Hurley, Y. Yuan, X. Qian, Z. Wang, S. Huang, and B. Mortazavi, "Uncertainty quantification for deep context-aware mobile activity recognition and unknown context discovery," *arXiv preprint arXiv:2003.01753*, 2020.

[439] R. Bondesan and M. Welling, "Quantum deformed neural networks," *arXiv preprint arXiv:2010.11189*, 2020.

[440] R. Ardywibowo, G. Zhao, Z. Wang, B. Mortazavi, S. Huang, and X. Qian, "Adaptive activity monitoring with uncertainty quantification in switching gaussian process models," *arXiv preprint arXiv:1901.02427*, 2019.

[441] A. Sadeghi, G. Wang, M. Ma, and G. B. Giannakis, "Learning while respecting privacy and robustness to distributional uncertainties and adversarial data," *arXiv preprint arXiv:2007.03724*, 2020.

[442] M. P. Vadera, A. D. Cobb, B. Jalaian, and B. M. Marlin, "Ursabench: Comprehensive benchmarking of approximate bayesian inference methods for deep neural networks," *arXiv preprint arXiv:2007.04466*, 2020.

[443] C. Cai, H. V. Poor, and Y. Chen, "Uncertainty quantification for nonconvex tensor completion: Confidence intervals, heteroscedasticity and optimality," in *International Conference on Machine Learning*. PMLR, 2020, pp. 1271–1282.

[444] A. Mishkin, F. Kunstner, D. Nielsen, M. Schmidt, and M. E. Khan, "Slang: Fast structured covariance approximations for bayesian deep learning with natural gradient," in *Advances in Neural Information Processing Systems*, 2018, pp. 6245–6255.

[445] S. Ghosh, J. Yao, and F. Doshi-Velez, "Model selection in bayesian neural networks via horseshoe priors." *Journal of Machine Learning Research*, vol. 20, no. 182, pp. 1–46, 2019.

[446] J. Hernandez-Lobato, Y. Li, M. Rowland, T. Bui, D. Hernández-Lobato, and R. Turner, "Black-box alpha divergence minimization," in *International Conference on Machine Learning*, 2016, pp. 1511–1520.

[447] B. N. Patro, V. P. Namboodiri *et al.*, "Probabilistic framework for solving visual dialog," *arXiv preprint arXiv:1909.04800*, 2019.

[448] S. Farquhar, M. A. Osborne, and Y. Gal, "Radial bayesian neural networks: Beyond discrete support in large-scale bayesian deep learning," *stat*, vol. 1050, p. 7, 2020.

[449] R. Novak, L. Xiao, J. Hron, J. Lee, A. A. Alemi, J. Sohl-Dickstein, and S. S. Schoenholz, "Neural tangents: Fast and easy infinite neural networks in python," *arXiv preprint arXiv:1912.02803*, 2019.

[450] X. Liu, Y. Li, C. Wu, and C.-J. Hsieh, "Adv-bnn: Improved adversarial defense through robust bayesian neural network," *arXiv preprint arXiv:1810.01279*, 2018.

[451] J. Pomponi, S. Scardapane, and A. Uncini, "Bayesian neural networks with maximum mean discrepancy regularization," *arXiv preprint arXiv:2003.00952*, 2020.

[452] R. Harang and E. M. Rudd, "Towards principled uncertainty estimation for deep neural networks," *arXiv preprint arXiv:1810.12278*, 2018.

[453] R. Y. Rohekar, Y. Gurwicz, S. Nisimov, G. Koren, and G. Novik, "Bayesian structure learning by recursive bootstrap," in *Advances in Neural Information Processing Systems*, 2018, pp. 10 525–10 535.

[454] S. T. Radev, M. D'Alessandro, P.-C. Bürkner, U. K. Mertens, A. Voss, and U. Kothe, "Amortized bayesian model comparison with evidential deep learning," *arXiv preprint arXiv:2004.10629*, 2020.

[455] S. Ariafar, Z. Mariet, E. Elhamifar, D. Brooks, J. Dy, and J. Snoek, "Weighting is worth the wait: Bayesian optimization with importance sampling," *arXiv preprint arXiv:2002.09927*, 2020.

[456] X. Meng, R. Bachmann, and M. E. Khan, "Training binary neural networks using the bayesian learning rule," *arXiv preprint arXiv:2002.10778*, 2020.

[457] Y. Xia, F. Liu, D. Yang, J. Cai, L. Yu, Z. Zhu, D. Xu, A. Yuille, and H. Roth, "3d semi-supervised learning with uncertainty-aware multi-view co-training," in *The IEEE Winter Conference on Applications of Computer Vision*, 2020, pp. 3646–3655.

[458] E. Schonfeld, S. Ebrahimi, S. Sinha, T. Darrell, and Z. Akata, "Generalized zero-and few-shot learning via aligned variational autoencoders," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8247–8255.

[459] J. Yoon, T. Kim, O. Dia, S. Kim, Y. Bengio, and S. Ahn, "Bayesian model-agnostic meta-learning," in *Advances in Neural Information Processing Systems*, 2018, pp. 7332–7342.

[460] H.-Y. Tseng, Y.-W. Chen, Y.-H. Tsai, S. Liu, Y.-Y. Lin, and M.-H. Yang, "Regularizing meta-learning via gradient dropout," *arXiv preprint arXiv:2004.05859*, 2020.

[461] A. Wu, S. Nowozin, E. Meeds, R. E. Turner, J. M. Hernández-Lobato, and A. L. Gaunt, "Deterministic variational inference for robust bayesian neural networks," *arXiv preprint arXiv:1810.03958*, 2018.

[462] T. Pearce, R. Tsuchida, M. Zaki, A. Brintrup, and A. Neely, "Expressive priors in bayesian neural networks: Kernel combinations and periodic functions," *arXiv preprint arXiv:1905.06076*, 2019.

[463] A. Atanov, A. Ashukha, K. Struminsky, D. Vetrov, and M. Welling, "The deep weight prior," *arXiv preprint arXiv:1810.06943*, 2018.

[464] H. Li, P. Barnaghi, S. Enshaeifar, and F. Ganz, "Continual learning using bayesian neural networks," *arXiv preprint arXiv:1910.04112*, 2019.

[465] A. Alaa and M. Van Der Schaar, "Discriminative jackknife: Quantifying uncertainty in deep learning via higher-order influence functions," in *International Conference on Machine Learning*. PMLR, 2020, pp. 165–174.

[466] A. Shekhovtsov and B. Flach, "Stochastic normalizations as bayesian learning," in *Asian Conference on Computer Vision*. Springer, 2018, pp. 463–479.

[467] M. Gantenbein, E. Erdil, and E. Konukoglu, "Revphiseg: A memory-efficient neural network for uncertainty quantification in medical image segmentation," *arXiv preprint arXiv:2008.06999*, 2020.

[468] Y. Zhu, N. Zabaras, P.-S. Koutsourelakis, and P. Perdikaris, "Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data," *Journal of Computational Physics*, vol. 394, pp. 56–81, 2019.

[469] X. Yang, "Uncertainty quantification, image synthesis and deformation prediction for image registration," Ph.D. dissertation, University of North Carolina at Chapel Hill, 2017.

[470] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" in *Advances in neural information processing systems*, 2017, pp. 5574–5584.

[471] A. Kristiadi, S. Däubener, and A. Fischer, "Predictive uncertainty quantification with compound density networks," *arXiv preprint arXiv:1902.01080*, 2019.

[472] M. Hobbhahn, A. Kristiadi, and P. Hennig, "Fast predictive uncertainty for classification with bayesian deep networks," *arXiv preprint arXiv:2003.01227*, 2020.

[473] A. Kendall and R. Cipolla, "Modelling uncertainty in deep learning for camera relocalization," in *2016 IEEE international conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 4762–4769.

[474] J. Guynn, "Google photos labeled black people 'gorillas'," *USA Today*, vol. 1, 2015.

[475] A. Taha, Y.-T. Chen, T. Misu, A. Shrivastava, and L. Davis, "Unsupervised data uncertainty learning in visual retrieval systems," *arXiv preprint arXiv:1902.02586*, 2019.

[476] G. Dorta, S. Vicente, L. Agapito, N. D. Campbell, and I. Simpson, "Structured uncertainty prediction networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5477–5485.

[477] A. Asai, D. Ikami, and K. Aizawa, "Multi-task learning based on separable formulation of depth estimation and its uncertainty." in *CVPR Workshops*, 2019, pp. 21–24.

[478] C. Liu, J. Gu, K. Kim, S. G. Narasimhan, and J. Kautz, "Neural rgb (r) d sensing: Depth and uncertainty from a video camera," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 986–10 995.

[479] A. Harakeh, M. Smart, and S. L. Waslander, "Bayesod: A bayesian approach for uncertainty estimation in deep object detectors," *arXiv preprint arXiv:1903.03838*, 2019.

[480] M. T. Le, F. Diehl, T. Brunner, and A. Knol, "Uncertainty estimation for deep neural object detectors in safety-critical applications," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 3873–3878.

[481] Y. He, C. Zhu, J. Wang, M. Savvides, and X. Zhang, "Bounding box regression with uncertainty for accurate object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2888–2897.

[482] P.-Y. Huang, W.-T. Hsu, C.-Y. Chiu, T.-F. Wu, and M. Sun, "Efficient uncertainty estimation for semantic segmentation in videos," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 520–535.

[483] G. Pascual, S. Seguí, and J. Vitrià, "Uncertainty gated network for land cover segmentation." in *CVPR Workshops*, 2018, pp. 276–279.

[484] C. Martinez, K. M. Potter, M. D. Smith, E. A. Donahue, L. Collins, J. P. Korbin, and S. A. Roberts, "Segmentation certainty through uncertainty: Uncertainty-refined binary volumetric segmentation under multifactor domain shift," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.

[485] J. Postels, F. Ferroni, H. Coskun, N. Navab, and F. Tombari, "Sampling-free epistemic uncertainty estimation using approximated variance propagation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 2931–2940.

[486] J. Gast and S. Roth, "Lightweight probabilistic deep networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3369–3378.

[487] A. Loquercio, M. Segù, and D. Scaramuzza, "A general framework for uncertainty estimation in deep learning," *arXiv preprint arXiv:1907.06890*, 2019.

[488] N. B. Gundavarapu, D. Srivastava, R. Mitra, A. Sharma, and A. Jain, "Structured aleatoric uncertainty in human pose estimation." in *CVPR Workshops*, vol. 2, 2019.

[489] L. Bertoni, S. Kreiss, and A. Alahi, "Monoloco: Monocular 3d pedestrian localization and uncertainty estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 6861–6871.

[490] J. Zheng, "Augmented deep representations for unconstrained still/video-based face recognition," Ph.D. dissertation, University of Maryland, College Park, 2019.

[491] T. Yu, D. Li, Y. Yang, T. M. Hospedales, and T. Xiang, "Robust person re-identification by modelling feature uncertainty," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 552–561.

[492] J. Zheng, R. Yu, J.-C. Chen, B. Lu, C. D. Castillo, and R. Chellappa, "Uncertainty modeling of contextual-connections between tracklets for unconstrained video-based face recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 703–712.

[493] J. C. Peterson, R. M. Battleday, T. L. Griffiths, and O. Russakovsky, "Human uncertainty makes classification more robust," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 9617–9626.

[494] G. Carbone, M. Wicker, L. Laurenti, A. Patane, L. Bortolussi, and G. Sanguinetti, "Robustness of bayesian neural networks to gradient-based attacks," *arXiv preprint arXiv:2002.04359*, 2020.

[495] O. Makansi, E. Ilg, O. Cicek, and T. Brox, "Overcoming limitations of mixture density networks: A sampling and fitting framework for multimodal future prediction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 7144–7153.

[496] A. Bhattacharyya, M. Fritz, and B. Schiele, "Long-term on-board prediction of people in traffic scenes under uncertainty," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4194–4202.

[497] A. Eldesokey, M. Felsberg, K. Holmquist, and M. Persson, "Uncertainty-aware cnns for depth completion: Uncertainty from beginning to end," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 014–12 023.

[498] F. K. Gustafsson, M. Danelljan, and T. B. Schon, "Evaluating scalable bayesian deep learning methods for robust computer vision," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 318–319.

[499] M. Cai, F. Lu, and Y. Sato, "Generalizing hand segmentation in egocentric videos with uncertainty-guided model adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 14 392–14 401.

[500] N. Hochgeschwender, P. Plöger, F. Kirchner, M. Valdenegro-Toro *et al.*, "Evaluating uncertainty estimation methods on 3d semantic segmentation of point clouds," *arXiv preprint arXiv:2007.01787*, 2020.

[501] Y.-T. Chang, Q. Wang, W.-C. Hung, R. Piramuthu, Y.-H. Tsai, and M.-H. Yang, "Mixup-cam: Weakly-supervised semantic segmentation via uncertainty regularization," *arXiv preprint arXiv:2008.01201*, 2020.

[502] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[503] Z. W. Lim, M. L. Lee, W. Hsu, and T. Y. Wong, "Building trust in deep learning system towards automated disease detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 9516–9521.

[504] M. W. Dusenberry, D. Tran, E. Choi, J. Kemp, J. Nixon, G. Jerfel, K. Heller, and A. M. Dai, "Analyzing the role of model uncertainty for electronic health records," in *Proceedings of the ACM Conference on Health, Inference, and Learning*, 2020, pp. 204–213.

[505] Y. Xia, D. Yang, Z. Yu, F. Liu, J. Cai, L. Yu, Z. Zhu, D. Xu, A. Yuille, and H. Roth, "Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation," *Medical Image Analysis*, p. 101766, 2020.

[506] R. Liu, S. Cheng, L. Tian, and J. Yi, "Deep spectral learning for label-free optical imaging oximetry with uncertainty quantification," *Light: Science & Applications*, vol. 8, no. 1, pp. 1–13, 2019.

[507] S. Kohl, B. Romera-Paredes, C. Meyer, J. De Fauw, J. R. Ledsam, K. Maier-Hein, S. A. Eslami, D. J. Rezende, and O. Ronneberger, "A probabilistic u-net for segmentation of ambiguous images," in *Advances in Neural Information Processing Systems*, 2018, pp. 6965–6975.

[508] T. Araújo, G. Aresta, L. Mendonça, S. Penas, C. Maia, Â. Carneiro, A. M. Mendonça, and A. Campilho, "Dr— graduate: uncertainty-aware deep learning-based diabetic retinopathy grading in eye fundus images," *Medical Image Analysis*, p. 101715, 2020.

[509] D. Karimi, Q. Zeng, P. Mathur, A. Avinash, S. Mahdavi, I. Spadinger, P. Abolmaesumi, and S. E. Salcudean, "Accurate and robust deep learning-based segmentation of the prostate clinical target volume in ultrasound images," *Medical image analysis*, vol. 57, pp. 186–196, 2019.

[510] M. Combalia, F. Hueto, S. Puig, J. Malvehy, and V. Vilaplana, "Uncertainty estimation in deep neural networks for dermoscopic image classification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 744–745.

[511] L. Dahal, A. Kafle, and B. Khanal, "Uncertainty estimation in deep 2d echocardiography segmentation," *arXiv preprint arXiv:2005.09349*, 2020.

[512] W. Zhu, A. Myronenko, Z. Xu, W. Li, H. Roth, Y. Huang, F. Milletari, and D. Xu, "Neurreg: Neural registration and its application to image segmentation," in *The IEEE Winter Conference on Applications of Computer Vision*, 2020, pp. 3617–3626.

[513] C. Bian, C. Yuan, J. Wang, M. Li, X. Yang, S. Yu, K. Ma, J. Yuan, and Y. Zheng, "Uncertainty-aware domain alignment for anatomical structure segmentation," *Medical Image Analysis*, p. 101732, 2020.

[514] S. A. Kohl, B. Romera-Paredes, K. H. Maier-Hein, D. J. Rezende, S. Eslami, P. Kohli, A. Zisserman, and O. Ronneberger, "A hierarchical probabilistic u-net for modeling multi-scale ambiguities," *arXiv preprint arXiv:1905.13077*, 2019.

[515] M. Yin, A. Yazdani, and G. E. Karniadakis, "One-dimensional modeling of fractional flow reserve in coronary artery disease: Uncertainty quantification and bayesian optimization," *Computer Methods in Applied Mechanics and Engineering*, vol. 353, pp. 66–85, 2019.

[516] B. Li and T. S. Alstrøm, "On uncertainty estimation in active learning for image segmentation," *arXiv preprint arXiv:2007.06364*, 2020.

[517] J.-T. Chien and Y.-C. Ku, "Bayesian recurrent neural network for language modeling," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 2, pp. 361–374, 2015.

[518] V. Vincze, "Uncertainty detection in natural language texts," Ph.D. dissertation, szte, 2015.

[519] Y. Chen, T. A. Lasko, Q. Mei, J. C. Denny, and H. Xu, "A study of active learning methods for named entity recognition in clinical text," *Journal of biomedical informatics*, vol. 58, pp. 11–18, 2015.

[520] L. Kong, G. Melis, W. Ling, L. Yu, and D. Yogatama, "Variational smoothing in recurrent neural network language models," *arXiv preprint arXiv:1901.09296*, 2019.

[521] X. Han, B. Li, and Z. Wang, "An attention-based neural framework for uncertainty identification on social media texts," *Tsinghua Science and Technology*, vol. 25, no. 1, pp. 117–126, 2019.

[522] Z. Zhang, S. Zohren, and S. Roberts, "Bdlob: Bayesian deep convolutional neural networks for limit order books," *arXiv preprint arXiv:1811.10041*, 2018.

[523] T. Z. Xiao, A. N. Gomez, and Y. Gal, "Wat heb je gezegd? detecting out-of-distribution translations with variational transformers," in *Third workshop on Bayesian Deep Learning (NeurIPS 2018)*, 2019.

[524] M. Ott, M. Auli, D. Grangier, and M. Ranzato, "Analyzing uncertainty in neural machine translation," *arXiv preprint arXiv:1803.00047*, 2018.

[525] J. Vaicenavicius, D. Widmann, C. Andersson, F. Lindsten, J. Roll, and T. B. Schön, "Evaluating model calibration in classification," *arXiv preprint arXiv:1902.06977*, 2019.

[526] J. Mena, O. Pujol, and J. Vitrià, "Dirichlet uncertainty wrappers for actionable algorithm accuracy accountability and auditability," *arXiv preprint arXiv:1912.12628*, 2019.

[527] Y. Xiao and W. Y. Wang, "Quantifying uncertainties in natural language processing tasks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 7322–7329.

[528] X. Zhang and S. Mahadevan, "Bayesian neural networks for flight trajectory prediction and safety assessment," *Decision Support Systems*, vol. 131, p. 113246, 2020.

[529] T. Vandal, E. Kodra, J. Dy, S. Ganguly, R. Nemani, and A. R. Ganguly, "Quantifying uncertainty in discrete-continuous and skewed data with bayesian deep learning," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 2377–2386.

[530] E. Nalisnick and J. M. Hernández-Lobato, "Automatic depth determination for bayesian resnets," in *Third workshop on Bayesian Deep Learning (NeurIPS 2018)*, 2018.

[531] B. Bischke, P. Helber, J. Folz, D. Borth, and A. Dengel, "Multitask learning for segmentation of building footprints with deep neural networks," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 1480–1484.

[532] G. P. Meyer and N. Thakurdesai, "Learning an uncertainty-aware object detector for autonomous driving," *arXiv preprint arXiv:1910.11375*, 2019.

[533] P. L. McDermott and C. K. Wikle, "Bayesian recurrent neural network models for forecasting and quantifying uncertainty in spatial-temporal data," *Entropy*, vol. 21, no. 2, p. 184, 2019.

[534] J. Yu, M. W. Lam, S. Hu, X. Wu, X. Li, Y. Cao, X. Liu, and H. Meng, "Comparative study of parametric and representation uncertainty modeling for recurrent neural network language models." in *INTERSPEECH*, 2019, pp. 3510–3514.

[535] N. Geneva and N. Zabaras, "Quantifying model form uncertainty in reynolds-averaged turbulence models with bayesian deep neural networks," *Journal of Computational Physics*, vol. 383, pp. 125–147, 2019.

[536] L. Feng, S. Xu, F. Wang, S. Liu, and H. Qiao, "Rough extreme learning machine: A new classification method based on uncertainty measure," *Neurocomputing*, vol. 325, pp. 269–282, 2019.

[537] M. Walmsley, L. Smith, C. Lintott, Y. Gal, S. Bamford, H. Dickinson, L. Fortson, S. Kruk, K. Masters, C. Scarlata *et al.*, "Galaxy zoo: probabilistic morphology through bayesian cnns and active learning," *Monthly Notices of the Royal Astronomical Society*, vol. 491, no. 2, pp. 1554–1574, 2020.

[538] M. Henaff, A. Canziani, and Y. LeCun, "Model-predictive policy learning with uncertainty regularization for driving in dense traffic," *arXiv preprint arXiv:1901.02705*, 2019.

[539] Y. Zhang *et al.*, "Bayesian semi-supervised learning for uncertainty-calibrated prediction of molecular properties and active learning," *Chemical Science*, vol. 10, no. 35, pp. 8154–8163, 2019.

[540] D. Ruhe, G. Cina, M. Tonutti, D. de Bruin, and P. Elbers, "Bayesian modelling in practice: Using uncertainty to im-

prove trustworthiness in medical applications," *arXiv preprint arXiv:1906.08619*, 2019.

[541] N. Neverova, D. Novotny, and A. Vedaldi, "Correlated uncertainty for learning dense correspondences from noisy labels," in *Advances in Neural Information Processing Systems*, 2019, pp. 920–928.

[542] Y. Gal, J. Hron, and A. Kendall, "Concrete dropout," in *Advances in neural information processing systems*, 2017, pp. 3581–3590.

[543] F. Verdoja, J. Lundell, and V. Kyrki, "Deep network uncertainty maps for indoor navigation," in *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2019, pp. 112–119.

[544] N. G. Polson, V. Sokolov *et al.*, "Deep learning: A bayesian perspective," *Bayesian Analysis*, vol. 12, no. 4, pp. 1275–1304, 2017.

[545] A. Brusaferri, M. Matteucci, P. Portolani, and A. Vitali, "Bayesian deep learning based method for probabilistic forecast of day-ahead electricity prices," *Applied Energy*, vol. 250, pp. 1158–1175, 2019.

[546] M. Kampffmeyer, A.-B. Salberg, and R. Jenssen, "Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2016, pp. 1–9.

[547] M. Sun, T. Zhang, Y. Wang, G. Strbac, and C. Kang, "Using bayesian deep learning to capture uncertainty for residential net load forecasting," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 188–201, 2019.

[548] B. Xu and Z. Chen, "Multi-level fusion based 3d object detection from monocular images," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2345–2353.

[549] H. Li, X.-J. Wu, and J. Kittler, "Infrared and visible image fusion using a deep learning framework," in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 2705–2710.

[550] S. Nemati, R. Rohani, M. E. Basiri, M. Abdar, N. Y. Yen, and V. Makarenkov, "A hybrid latent space data fusion method for multimodal emotion recognition," *IEEE Access*, vol. 7, pp. 172 948–172 964, 2019.

[551] K. Tian, Y. Xu, S. Zhou, and J. Guan, "Versatile multiple choice learning and its application to vision computing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6349–6357.

[552] A. Wasay, B. Hentschel, Y. Liao, S. Chen, and S. Idreos, "Mothernets: Rapid deep ensemble learning," in *Proceedings of the 3rd MLSys Conference (MLSys)*, 2020.

[553] K. Lee, C. Hwang, K. Park, and J. Shin, "Confident multiple choice learning," *arXiv preprint arXiv:1706.03475*, 2017.

[554] W. H. Beluch, T. Genewein, A. Nürnberger, and J. M. Köhler, "The power of ensembles for active learning in image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9368–9377.

[555] Y. Tan, B. Jin, X. Yue, Y. Chen, and A. S. Vincentelli, "Exploiting uncertainties from ensemble learners to improve decision-making in healthcare ai," *arXiv preprint arXiv:2007.06063*, 2020.

[556] H. J. Lee, S. T. Kim, N. Navab, and Y. M. Ro, "Efficient ensemble model generation for uncertainty estimation with bayesian approximation in segmentation," *arXiv preprint arXiv:2005.10754*, 2020.

[557] L. Tran, B. S. Veeling, K. Roth, J. Swiatkowski, J. V. Dillon, J. Snoek, S. Mandt, T. Salimans, S. Nowozin, and R. Jenatton, "Hydra: Preserving ensemble diversity for model distillation," *arXiv preprint arXiv:2001.04694*, 2020.

[558] Y. Yao, "An outline of a theory of three-way decisions," in *International Conference on Rough Sets and Current Trends in Computing*. Springer, 2012, pp. 1–17.

[559] Y. Ben-Haim, *Info-gap decision theory: decisions under severe uncertainty*. Elsevier, 2006.

[560] V. A. Marchau, W. E. Walker, P. J. Bloemen, and S. W. Popper, *Decision making under deep uncertainty: From theory to practice*. Springer Nature, 2019.

[561] J. Gordon, J. Bronskill, M. Bauer, S. Nowozin, and R. E. Turner, "Meta-learning probabilistic inference for prediction," *arXiv preprint arXiv:1805.09921*, 2018.

[562] C. Nguyen, T.-T. Do, and G. Carneiro, "Uncertainty in model-agnostic meta-learning using variational inference," in *The IEEE Winter Conference on Applications of Computer Vision*, 2020, pp. 3090–3100.

[563] S. X. Hu, P. G. Moreno, Y. Xiao, X. Shen, G. Obozinski, N. D. Lawrence, and A. Damianou, "Empirical bayes transductive meta-learning with synthetic gradients," *arXiv preprint arXiv:2004.12696*, 2020.

[564] M. Qu, T. Gao, L.-P. Xhonneux, and J. Tang, "Few-shot relation extraction via bayesian meta-learning on relation graphs," in *International Conference on Machine Learning*. PMLR, 2020, pp. 7867–7876.

[565] S. Sinha, S. Ebrahimi, and T. Darrell, "Variational adversarial active learning," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 5972–5981.

[566] S. Zaidi, A. Zela, T. Elsken, C. Holmes, F. Hutter, and Y. W. Teh, "Neural ensemble search for performant and calibrated predictions," *arXiv preprint arXiv:2006.08573*, 2020.

[567] W. Van Gansbeke, S. Vandenhende, S. Georgoulis, M. Proesmans, and L. Van Gool, "Learning to classify images without labels," *arXiv preprint arXiv:2005.12320*, 2020.

[568] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[569] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "Eca-net: Efficient channel attention for deep convolutional neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 534–11 542.

[570] D. Hendrycks, M. Mazeika, and T. Dietterich, "Deep anomaly detection with outlier exposure," *arXiv preprint arXiv:1812.04606*, 2018.

[571] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 6023–6032.

[572] J. Ren, P. J. Liu, E. Fertig, J. Snoek, R. Poplin, M. Depristo, J. Dillon, and B. Lakshminarayanan, "Likelihood ratios for out-of-distribution detection," in *Advances in Neural Information Processing Systems*, 2019, pp. 14 707–14 718.

[573] S. Padhy, Z. Nado, J. Ren, J. Liu, J. Snoek, and B. Lakshminarayanan, "Revisiting one-vs-all classifiers for predictive uncertainty and out-of-distribution detection in neural networks," *arXiv preprint arXiv:2007.05134*, 2020.

[574] L. Kong, J. Sun, and C. Zhang, "Sde-net: Equipping deep neural networks with uncertainty estimates," *arXiv preprint arXiv:2008.10546*, 2020.

[575] J. Postels, H. Blum, C. Cadena, R. Siegwart, L. Van Gool, and F. Tombari, "Quantifying aleatoric and epistemic uncertainty using density estimation in latent space," *arXiv preprint arXiv:2012.03082*, 2020.

[576] D. Ulmer and G. Cinà, "Know your limits: Monotonicity & softmax make neural classifiers overconfident on ood data," *arXiv preprint arXiv:2012.05329*, 2020.

[577] D. Ha, A. Dai, and Q. V. Le, "Hypernetworks," *arXiv preprint arXiv:1609.09106*, 2016.

[578] A. Navon, A. Shamsian, G. Chechik, and E. Fetaya, "Learning the pareto front with hypernetworks," *arXiv preprint arXiv:2010.04104*, 2020.

[579] T. Galanti and L. Wolf, "On the modularity of hypernetworks," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[580] D. Krueger, C.-W. Huang, R. Islam, R. Turner, A. Lacoste, and A. Courville, "Bayesian hypernetworks," *arXiv preprint arXiv:1710.04759*, 2017.

[581] C. Zeno, I. Golan, E. Hoffer, and D. Soudry, "Task agnostic continual learning using online variational bayes," *arXiv preprint arXiv:1803.10123*, 2018.

[582] S. Farquhar and Y. Gal, "A unifying bayesian view of continual learning," *arXiv preprint arXiv:1902.06494*, 2019.

[583] S. Kessler, V. Nguyen, S. Zohren, and S. Roberts, "Indian buffet neural networks for continual learning," *arXiv preprint arXiv:1912.02290*, 2019.

[584] Y. Chen, T. Diethe, and N. Lawrence, "Facilitating bayesian continual learning by natural gradients and stein gradients," *arXiv preprint arXiv:1904.10644*, 2019.

[585] A. Kumar, S. Chatterjee, and P. Rai, "Bayesian structure adaptation for continual learning," *arXiv*, pp. arXiv–1912, 2019.

[586] H. Li, P. Barnaghi, S. Enshaeifar, and F. Ganz, "Continual learning using bayesian neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[587] P. Pan, S. Swaroop, A. Immer, R. Eschenhagen, R. E. Turner, and M. E. Khan, "Continual deep learning by functional regularisation of memorable past," *arXiv preprint arXiv:2004.14070*, 2020.

[588] J. Zhou, G. Cui, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *arXiv preprint arXiv:1812.08434*, 2018.

[589] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superglue: Learning feature matching with graph neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 4938–4947.

[590] F. d. A. Belbute-Peres, T. Economon, and Z. Kolter, "Combining differentiable pde solvers and graph neural networks for fluid flow prediction," in *International Conference on Machine Learning*. PMLR, 2020, pp. 2402–2411.

[591] D. Luo, W. Cheng, D. Xu, W. Yu, B. Zong, H. Chen, and X. Zhang, "Parameterized explainer for graph neural network," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[592] W. Shi and R. Rajkumar, "Point-gnn: Graph neural network for 3d object detection in a point cloud," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1711–1719.

[593] F. Teimury, S. Pal, A. Amini, and M. Coates, "Estimation of time-series on graphs using bayesian graph convolutional neural networks," in *Wavelets and Sparsity XVIII*, vol. 11138. International Society for Optics and Photonics, 2019, p. 111380Y.

[594] P. Elinas, E. V. Bonilla, and L. Tiao, "Variational inference for graph convolutional networks in the absence of graph data and adversarial settings," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[595] J. Sun, W. Guo, D. Zhang, Y. Zhang, F. Regol, Y. Hu, H. Guo, R. Tang, H. Yuan, X. He *et al.*, "A framework for recommending accurate and diverse items using bayesian graph convolutional neural networks," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 2030–2039.

[596] L. Ma, J. Cui, and B. Yang, "Deep neural architecture search with deep graph bayesian optimization," in *2019 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. IEEE, 2019, pp. 500–507.

[597] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," *Advances in neural information processing systems*, vol. 25, pp. 2951–2959, 2012.

[598] P. I. Frazier, "A tutorial on bayesian optimization," *arXiv preprint arXiv:1807.02811*, 2018.

[599] J. Wu, S. Toscano-Palmerin, P. I. Frazier, and A. G. Wilson, "Practical multi-fidelity bayesian optimization for hyperparameter tuning," in *Uncertainty in Artificial Intelligence*. PMLR, 2020, pp. 788–798.

[600] M. Balandat, B. Karrer, D. Jiang, S. Daulton, B. Letham, A. G. Wilson, and E. Bakshy, "Botorch: A framework for efficient monte-carlo bayesian optimization," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[601] F. Sorourifar, G. Makrygirgos, A. Mesbah, and J. A. Paulson, "A data-driven automatic tuning method for mpc under uncertainty using constrained bayesian optimization," *arXiv preprint arXiv:2011.11841*, 2020.

[602] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," *arXiv preprint arXiv:1706.04599*, 2017.

[603] K. Patel, W. Beluch, B. Yang, M. Pfeiffer, and D. Zhang, "Multi-class uncertainty calibration via mutual information maximization-based binning," *arXiv preprint arXiv:2006.13092*, 2020.

[604] D. Widmann, F. Lindsten, and D. Zachariah, "Calibration tests in multi-class classification: A unifying framework," in *Advances in Neural Information Processing Systems*, 2019, pp. 12 257–12 267.

[605] A. Rahimi, A. Shaban, C.-A. Cheng, R. Hartley, and B. Boots, "Intra order-preserving functions for calibration of multi-class neural networks," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[606] S. Utpala and P. Rai, "Quantile regularization: Towards implicit calibration of regression models," *arXiv preprint arXiv:2002.12860*, 2020.

[607] S. Zhao, T. Ma, and S. Ermon, "Individual calibration with randomized forecasting," in *International Conference on Machine Learning*. PMLR, 2020, pp. 11 387–11 397.

[608] M. Rawat, M. Wistuba, and M.-I. Nicolae, "Harnessing model uncertainty for detecting adversarial examples," in *NIPS Workshop on Bayesian Deep Learning*, 2017.

[609] J. Su, M. Cvitkovic, and F. Huang, "Sampling-free learning of bayesian quantized neural networks," *arXiv preprint arXiv:1912.02992*, 2019.

[610] M. Haußmann, F. A. Hamprecht, and M. Kandemir, "Sampling-free variational inference of bayesian neural networks by variance backpropagation," in *Uncertainty in Artificial Intelligence*. PMLR, 2020, pp. 563–573.

[611] Y. Shi and A. K. Jain, "Probabilistic face embeddings," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 6902–6911.

[612] T. R. Scott, K. Ridgeway, and M. C. Mozer, "Stochastic prototype embeddings," *arXiv preprint arXiv:1909.11702*, 2019.

[613] F. D. S. Ribeiro, G. Leontidis, and S. D. Kollias, "Capsule routing via variational bayes." in *AAAI*, 2020, pp. 3749–3756.

[614] R. Loftin, M. E. Taylor, M. L. Littman, J. MacGlashan, B. Peng, and D. L. Roberts, "Open problems for online bayesian inference in neural networks," in *Bayesian Deep Learning Workshop at NeurIPS*, 2016.

[615] C. V. Nguyen, Y. Li, T. D. Bui, and R. E. Turner, "Variational continual learning," *arXiv preprint arXiv:1710.10628*, 2017.

[616] H. Tseran, M. E. Khan, T. Harada, and T. D. Bui, "Natural variational continual learning," in *Continual Learning Workshop@ NeurIPS*, vol. 2, 2018.

[617] A. Lacoste, B. Oreshkin, W. Chung, T. Boquet, N. Rostamzadeh, and D. Krueger, "Uncertainty in multitask transfer learning," *arXiv preprint arXiv:1806.07528*, 2018.

[618] T. A. Nguyen, H. Jeong, E. Yang, and S. J. Hwang, "Clinical risk prediction with temporal probabilistic asymmetric multi-task learning," *arXiv preprint arXiv:2006.12777*, 2020.

[619] S. Wang, Y. Liu, C. Wang, H. Luan, and M. Sun, "Improving back-translation with uncertainty-based confidence estimation," *arXiv preprint arXiv:1909.00157*, 2019.

[620] K. Hama, T. Matsubara, K. Uehara, and J. Cai, "Exploring uncertainty measures for image-caption embedding-and-retrieval task," *arXiv preprint arXiv:1904.08504*, 2019.

[621] L. Zhou, H. Palangi, L. Zhang, H. Hu, J. J. Corso, and J. Gao, "Unified vision-language pre-training for image captioning and vqa." in *AAAI*, 2020, pp. 13 041–13 049.

[622] M. Cornia, M. Stefanini, L. Baraldi, and R. Cucchiara, "Meshed-memory transformer for image captioning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10 578–10 587.

[623] Y. Pan, T. Yao, Y. Li, and T. Mei, "X-linear attention networks for image captioning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10 971–10 980.

[624] S. Antol, A. Agrawal, J. Lu, M. Mitchell, D. Batra, C. Lawrence Zitnick, and D. Parikh, "Vqa: Visual question answering," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 2425–2433.

[625] J. Lu, J. Yang, D. Batra, and D. Parikh, "Hierarchical question-image co-attention for visual question answering," in *Advances in neural information processing systems*, 2016, pp. 289–297.

[626] P. Anderson, X. He, C. Buehler, D. Teney, M. Johnson, S. Gould, and L. Zhang, "Bottom-up and top-down attention for image captioning and visual question answering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 6077–6086.

[627] Y. Goyal, T. Khot, D. Summers-Stay, D. Batra, and D. Parikh, "Making the v in vqa matter: Elevating the role of image understanding in visual question answering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6904–6913.

[628] Z. Yu, J. Yu, Y. Cui, D. Tao, and Q. Tian, "Deep modular co-attention networks for visual question answering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2019, pp. 6281–6290.

[629] T. E. Boult, S. Cruz, A. R. Dhamija, M. Gunther, J. Henrydoss, and W. J. Scheirer, "Learning and the unknown: Surveying steps toward open world recognition," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 9801–9807.

[630] S. Ghosh, F. M. Delle Fave, and J. S. Yedidia, "Assumed density filtering methods for learning bayesian neural networks." in *AAAI*, 2016, pp. 1589–1595.

[631] X. Lu and B. Van Roy, "Ensemble sampling," in *Advances in neural information processing systems*, 2017, pp. 3258–3266.

[632] M. Karamanis and F. Beutler, "Ensemble slice sampling," *arXiv preprint arXiv:2002.06212*, 2020.

[633] F. Pourpanah, C. J. Tan, C. P. Lim, and J. Mohamad-Saleh, "A q-learning-based multi-agent system for data classification," *Applied Soft Computing*, vol. 52, pp. 519 – 531, 2017.

[634] F. Pourpanah, R. Wang, C. P. Lim, X. Wang, M. Seera, and C. J. Tan, "An improved fuzzy artmap and q-learning agent model for pattern classification," *Neurocomputing*, vol. 359, pp. 139 – 152, 2019.

[635] R. Zhang, C. Li, C. Chen, and L. Carin, "Learning structural weight uncertainty for sequential decision-making," in *International Conference on Artificial Intelligence and Statistics*, 2018, pp. 1137–1146.

[636] A. Y. Foong, D. R. Burt, Y. Li, and R. E. Turner, "Pathologies of factorised gaussian and mc dropout posteriors in bayesian neural networks," *stat*, vol. 1050, p. 2, 2019.

[637] J. Lambert, O. Sener, and S. Savarese, "Deep learning under privileged information using heteroscedastic dropout," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8886–8895.

[638] A. Kendall, Y. Gal, and R. Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7482–7491.

[639] R. Yasarla and V. M. Patel, "Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image deraining," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8405–8414.

[640] C. Xue, Q. Dou, X. Shi, H. Chen, and P.-A. Heng, "Robust learning at noisy labeled medical images: Applied to skin lesion classification," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 1280–1283.

[641] E. Abbasnejad, Q. Wu, Q. Shi, and A. v. d. Hengel, "What's to know? uncertainty as a guide to asking goal-oriented questions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4155–4164.

[642] V. Peretroukhin, B. Wagstaff, and J. Kelly, "Deep probabilistic regression of elements of so (3) using quaternion averaging and uncertainty injection." in *CVPR Workshops*, 2019, pp. 83–86.

[643] C. Zhang and B. Jin, "Probabilistic residual learning for aleatoric uncertainty in image restoration," *arXiv preprint arXiv:1908.01010*, 2019.

[644] E. Harris, A. Marcu, M. Painter, M. Niranjan, A. Prügel-Bennett, and J. Hare, "Understanding and enhancing mixed sample data augmentation," *arXiv preprint arXiv:2002.12047*, 2020.

[645] N. Miolane and S. Holmes, "Learning weighted submanifolds with variational autoencoders and riemannian variational autoencoders," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 14 503–14 511.

[646] Q. Zhou, Z. Feng, G. Cheng, X. Tan, J. Shi, and L. Ma, "Uncertainty-aware consistency regularization for cross-domain semantic segmentation," *arXiv preprint arXiv:2004.08878*, 2020.

[647] J. Zhang, D.-P. Fan, Y. Dai, S. Anwar, F. S. Saleh, T. Zhang, and N. Barnes, "Uc-net: uncertainty inspired rgb-d saliency detection via conditional variational autoencoders," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8582–8591.

[648] G.-H. Lee and S.-W. Lee, "Uncertainty-aware mesh decoder for high fidelity 3d face reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6100–6109.

[649] K. Wang, X. Peng, J. Yang, S. Lu, and Y. Qiao, "Suppressing uncertainties for large-scale facial expression recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6897–6906.

[650] N. Yang, L. v. Stumberg, R. Wang, and D. Cremers, "D3vo: Deep depth, deep pose and deep uncertainty for monocular visual odometry," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1281–1292.

[651] J. Chang, Z. Lan, C. Cheng, and Y. Wei, "Data uncertainty learning in face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5710–5719.

[652] M. Polic, S. Steidl, C. Albl, Z. Kukelova, and T. Pajdla, "Uncertainty based camera model selection," in *Proceedings of the*

[653] Y. Nan and H. Ji, "Deep learning for handling kernel/model uncertainty in image deconvolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2388–2397.

[654] A. Kumar, T. K. Marks, W. Mou, Y. Wang, M. Jones, A. Cherian, T. Koike-Akino, X. Liu, and C. Feng, "Luvli face alignment: Estimating landmarks' location, uncertainty, and visibility likelihood," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8236–8246.

[655] S. Cheng, Z. Xu, S. Zhu, Z. Li, L. E. Li, R. Ramamoorthi, and H. Su, "Deep stereo using adaptive thin volume representation with uncertainty awareness," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2524–2534.

[656] Y. Tang, Z. Ni, J. Zhou, D. Zhang, J. Lu, Y. Wu, and J. Zhou, "Uncertainty-aware score distribution learning for action quality assessment," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 9839–9848.

[657] E. D. Carvalho, R. Clark, A. Nicastro, and P. H. Kelly, "Scalable uncertainty for computer vision with functional variational inference," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 003–12 013.

[658] A. Angelopoulos, S. Bates, J. Malik, and M. I. Jordan, "Uncertainty sets for image classifiers using conformal prediction," *arXiv preprint arXiv:2009.14193*, 2020.

[659] A. Jungo, R. McKinley, R. Meier, U. Knecht, L. Vera, J. Pérez-Beta, D. Molina-García, V. M. Pérez-García, R. Wiest, and M. Reyes, "Towards uncertainty-assisted brain tumor segmentation and survival prediction," in *International MICCAI Brainlesion Workshop*. Springer, 2017, pp. 474–485.

[660] O. Ozdemir, B. Woodward, and A. A. Berlin, "Propagating uncertainty in multi-stage bayesian convolutional neural networks with application to pulmonary nodule detection," *arXiv preprint arXiv:1712.00497*, 2017.

[661] R. Tanno, D. E. Worrall, A. Ghosh, E. Kaden, S. N. Sotiropoulos, A. Criminisi, and D. C. Alexander, "Bayesian image quality transfer with cnns: exploring uncertainty in dmri super-resolution," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 611–619.

[662] Y. Kwon, J.-H. Won, B. J. Kim, and M. C. Paik, "Uncertainty quantification using bayesian neural networks in classification: Application to biomedical image segmentation," *Computational Statistics & Data Analysis*, vol. 142, p. 106816, 2020.

[663] M. S. Ayhan and P. Berens, "Test-time data augmentation for estimation of heteroscedastic aleatoric uncertainty in deep neural networks," in *1st Conference on Medical Imaging with Deep Learning,*, 2018.

[664] A. Jungo, R. Meier, E. Ermis, M. Blatti-Moreno, E. Herrmann, R. Wiest, and M. Reyes, "On the effect of inter-observer variability for a reliable estimation of uncertainty of medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 682–690.

[665] G. Wang, W. Li, M. A. Zuluaga, R. Pratt, P. A. Patel, M. Aertsen, T. Doel, A. L. David, J. Deprest, S. Ourselin *et al.*, "Interactive medical image segmentation using deep learning with image-specific fine tuning," *IEEE transactions on medical imaging*, vol. 37, no. 7, pp. 1562–1573, 2018.

[666] S. Moccia, S. J. Wirkert, H. Kenngott, A. S. Vemuri, M. Apitz, B. Mayer, E. De Momi, L. S. Mattos, and L. Maier-Hein, "Uncertainty-aware organ classification for surgical data science applications in laparoscopy," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 11, pp. 2649–2659, 2018.

[667] P. McClure, C. Y. Zheng, J. Kaczmarzyk, J. Rogers-Lee, S. Ghosh, D. Nielson, P. A. Bandettini, and F. Pereira, "Distributed weight consolidation: a brain segmentation case study," in *Advances in Neural Information Processing Systems*, 2018, pp. 4093–4103.

[668] A. Jungo and M. Reyes, "Assessing reliability and challenges of uncertainty estimations for medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 48–56.

[669] J. I. Orlando, P. Seeböck, H. Bogunović, S. Klimscha, C. Grechenig, S. Waldstein, B. S. Gerendas, and U. Schmidt-Erfurth, "U2-net: A bayesian u-net model with epistemic uncertainty feedback for photoreceptor layer segmentation in patho-

logical oct scans," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 1441–1445.

[670] F. C. Ghesu, B. Georgescu, E. Gibson, S. Guendel, M. K. Kalra, R. Singh, S. R. Digumarthy, S. Grbic, and D. Comaniciu, "Quantifying and leveraging classification uncertainty for chest radiograph assessment," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 676–684.

[671] C. F. Baumgartner, K. C. Tezcan, K. Chaitanya, A. M. Hötker, U. J. Muehlematter, K. Schawkat, A. S. Becker, O. Donati, and E. Konukoglu, "Phiseg: Capturing uncertainty in medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 119–127.

[672] Ł. Raczkowski, M. Możejko, J. Zambonelli, and E. Szczurek, "Ara: accurate, reliable and active histopathological image classification framework with bayesian deep learning," *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019.

[673] Z. Eaton-Rosen, T. Varsavsky, S. Ourselin, and M. J. Cardoso, "As easy as 1, 2... 4? uncertainty in counting tasks for medical imaging," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 356–364.

[674] M. L. di Scandalea, C. S. Perone, M. Boudreau, and J. Cohen-Adad, "Deep active learning for axon-myelin segmentation on histology data," *arXiv preprint arXiv:1907.05143*, 2019.

[675] A. Filos, S. Farquhar, A. N. Gomez, T. G. Rudner, Z. Kenton, L. Smith, M. Alizadeh, A. de Kroon, and Y. Gal, "Benchmarking bayesian deep learning with diabetic retinopathy diagnosis," *Preprint*, 2019.

[676] M. Ravanbakhsh, T. Klein, K. Batmanghelich, and M. Nabi, "Uncertainty-driven semantic segmentation through human-machine collaborative learning," *arXiv preprint arXiv:1909.00626*, 2019.

[677] R. Jena and S. P. Awate, "A bayesian neural net to segment images with uncertainty estimates and good calibration," in *International Conference on Information Processing in Medical Imaging*. Springer, 2019, pp. 3–15.

[678] R. Tanno, D. Worrall, E. Kaden, A. Ghosh, F. Grussu, A. Bizzi, S. N. Sotiropoulos, A. Criminisi, and D. C. Alexander, "Uncertainty quantification in deep learning for safer neuroimage enhancement," *arXiv preprint arXiv:1907.13418*, 2019.

[679] R. D. Soberanis-Mukul, N. Navab, and S. Albarqouni, "Uncertainty-based graph convolutional networks for organ segmentation refinement," *arXiv preprint arXiv:1906.02191*, 2019.

[680] S. Hu, D. Worrall, S. Knegt, B. Veeling, H. Huisman, and M. Welling, "Supervised uncertainty quantification for segmentation with multiple annotations," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 137–145.

[681] G. Luo, S. Dong, W. Wang, K. Wang, S. Cao, C. Tam, H. Zhang, J. Howey, P. Ohorodnyk, and S. Li, "Commensal correlation network between segmentation and direct area estimation for biventricle quantification," *Medical image analysis*, vol. 59, p. 101591, 2020.

[682] K. Hoebel, V. Andrearczyk, A. Beers, J. Patel, K. Chang, A. Depeursinge, H. Müller, and J. Kalpathy-Cramer, "An exploration of uncertainty information for segmentation quality assessment," in *Medical Imaging 2020: Image Processing*, vol. 11313. International Society for Optics and Photonics, 2020, p. 113131K.

[683] P. Seeböck, J. I. Orlando, T. Schlegl, S. M. Waldstein, H. Bogunović, S. Klimscha, G. Langs, and U. Schmidt-Erfurth, "Exploiting epistemic uncertainty of anatomy segmentation for anomaly detection in retinal oct," *IEEE transactions on medical imaging*, vol. 39, no. 1, pp. 87–98, 2019.

[684] Y. Hiasa, Y. Otake, M. Takao, T. Ogawa, N. Sugano, and Y. Sato, "Automated muscle segmentation from clinical ct using bayesian u-net for personalized musculoskeletal modeling," *IEEE Transactions on Medical Imaging*, vol. 39, no. 4, pp. 1030–1040, 2019.

[685] Y. Xue, S. Cheng, Y. Li, and L. Tian, "Reliable deep-learning-based phase imaging with uncertainty quantification," *Optica*, vol. 6, no. 5, pp. 618–629, 2019.

[686] T. LaBonte, C. Martinez, and S. A. Roberts, "We know where we don't know: 3d bayesian cnns for uncertainty quantification of binary segmentations for material simulations," *arXiv preprint arXiv:1910.10793*, 2019.

[687] Z. Liao, H. Girgis, A. Abdi, H. Vaseli, J. Hetherington, R. Rohling, K. Gin, T. Tsang, and P. Abolmaesumi, "On modelling label uncertainty in deep neural networks: Automatic estimation of intra-observer variability in 2d echocardiography quality assessment," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1868–1883, 2019.

[688] M. Raghu, K. Blumer, R. Sayres, Z. Obermeyer, B. Kleinberg, S. Mullainathan, and J. Kleinberg, "Direct uncertainty prediction for medical second opinions," in *International Conference on Machine Learning*, 2019, pp. 5281–5290.

[689] Z. Zhang, A. Romero, M. J. Muckley, P. Vincent, L. Yang, and M. Drozdzal, "Reducing uncertainty in undersampled mri reconstruction with active acquisition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2049–2058.

[690] C. Donnat, N. Miolane, F. d. S. P. Bunbury, and J. Kreindler, "A bayesian hierarchical network for combining heterogeneous data sources in medical diagnoses," *arXiv preprint arXiv:2007.13847*, 2020.

[691] A. Mehrtash, W. M. Wells, C. M. Tempany, P. Abolmaesumi, and T. Kapur, "Confidence calibration and predictive uncertainty estimation for deep medical image segmentation," *IEEE Transactions on Medical Imaging*, 2020.

[692] K. Wickstrøm, M. Kampffmeyer, and R. Jenssen, "Uncertainty and interpretability in convolutional neural networks for semantic segmentation of colorectal polyps," *Medical Image Analysis*, vol. 60, p. 101619, 2020.

[693] G. Carneiro, L. Z. C. T. Pu, R. Singh, and A. Burt, "Deep learning uncertainty and confidence calibration for the five-class polyp classification from colonoscopy," *Medical Image Analysis*, p. 101653, 2020.

[694] P. Natekar, A. Kori, and G. Krishnamurthi, "Demystifying brain tumor segmentation networks: Interpretability and uncertainty analysis," *Frontiers in Computational Neuroscience*, vol. 14, p. 6, 2020.

[695] X. Li, Y. Zhou, N. C. Dvornek, Y. Gu, P. Ventola, and J. S. Duncan, "Efficient shapley explanation for features importance estimation under uncertainty," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 792–801.

[696] H. Zheng, S. M. M. Perrine, M. K. Pitirri, K. Kawasaki, C. Wang, J. T. Richtsmeier, and D. Z. Chen, "Cartilage segmentation in high-resolution 3d micro-ct images via uncertainty-guided self-training with very sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 802–812.

[697] J. Wang, Y. Yan, Y. Zhang, G. Cao, M. Yang, and M. K. Ng, "Deep reinforcement active learning for medical image classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 33–42.

[698] L. Quan, Y. Li, X. Chen, and N. Zhang, "An effective data refinement approach for upper gastrointestinal anatomy recognition," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 43–52.

[699] P. Yuan, A. Mobiny, J. Jahanipour, X. Li, P. A. Cicalese, B. Roysam, V. M. Patel, M. Dragan, and H. Van Nguyen, "Few is enough: Task-augmented active meta-learning for brain cell classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 367–377.

[700] E. Chiou, F. Giganti, S. Punwani, I. Kokkinos, and E. Panagiotaki, "Harnessing uncertainty in domain adaptation for mri prostate lesion segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 510–520.

[701] Y. Wang, Y. Zhang, J. Tian, C. Zhong, Z. Shi, Y. Zhang, and Z. He, "Double-uncertainty weighted method for semi-supervised learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 542–551.

[702] Y. Li, J. Chen, X. Xie, K. Ma, and Y. Zheng, "Self-loop uncertainty: A novel pseudo-label for semi-supervised medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 614–623.

[703] H. Yang, C. Shan, A. F. Kolen *et al.*, "Deep q-network-driven catheter segmentation in 3d us by hybrid constrained semi-supervised learning and dual-unet," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 646–655.

[704] L. Venturini, A. T. Papageorghiou, J. A. Noble, and A. I. Namburete, "Uncertainty estimates as data selection criteria to boost omni-supervised learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 689–698.

[705] S. Yu, H.-Y. Zhou, K. Ma, C. Bian, C. Chu, H. Liu, and Y. Zheng, "Difficulty-aware glaucoma classification with multi-rater consensus modeling," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 741–750.

[706] Z. Huang, Y. Gan, T. Lye, H. Zhang, A. Laine, E. D. Angelini, and C. Hendon, "Heterogeneity measurement of cardiac tissues leveraging uncertainty information from image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 782–791.

[707] P. Khairnar, P. Thiagarajan, and S. Ghosh, "A modified bayesian convolutional neural network for breast histopathology image classification and uncertainty quantification," *arXiv preprint arXiv:2010.12575*, 2020.

[708] F. C. Ghesu, B. Georgescu, A. Mansoor, Y. Yoo, E. Gibson, R. Vishwanath, A. Balachandran, J. M. Balter, Y. Cao, R. Singh *et al.*, "Quantifying and leveraging predictive uncertainty for medical image assessment," *Medical Image Analysis*, p. 101855, 2020.

[709] S. Eggenreich, C. Payer, M. Urschler, and D. Štern, "Variational inference and bayesian cnns for uncertainty estimation in multifactorial bone age prediction," *arXiv preprint arXiv:2002.10819*, 2020.

[710] R. D. Soberanis-Mukul, N. Navab, and S. Albarqouni, "Uncertainty-based graph convolutional networks for organ segmentation refinement," in *Medical Imaging with Deep Learning*. PMLR, 2020, pp. 755–769.

[711] ——, "An uncertainty-driven gcn refinement strategy for organ segmentation," *arXiv preprint arXiv:2012.03352*, 2020.

[712] J.-S. Prassni, T. Ropinski, and K. Hinrichs, "Uncertainty-aware guided volume segmentation," *IEEE transactions on visualization and computer graphics*, vol. 16, no. 6, pp. 1358–1365, 2010.

[713] D. Ulmer, L. Meijerink, and G. Cinà, "Trust issues: Uncertainty estimation does not enable reliable ood detection on medical tabular data," in *Machine Learning for Health*. PMLR, 2020, pp. 341–354.

[714] X. Cao, H. Chen, Y. Li, Y. Peng, S. Wang, and L. Cheng, "Uncertainty aware temporal-ensembling model for semi-supervised abus mass segmentation," *IEEE Transactions on Medical Imaging*, 2020.

TABLE 6: More information regarding additional UQ techniques proposed in the literature (Detailed information, sorted by year).

Study	Year	Subject	# datasets	Uncertainty method's name	Code
Balan et al. [265]	2015	Image processing, toy and numerical data and	4	PBE	×
Houthooft et al. [266]	2016	Toy data (regression)	1	VIME	√
Springenberg et al. [267]	2016	Numerical data (regression)	4	BNN	√
Zhang et al. [273]	2017	Image processing	7	UCF (uncertain convolutional features)	√
Khan et al. [160]	2018	Numerical data	8	Vadam	√
Malinin et al. [382]	2018	Image processing	2	PN	√
Ilg et al. [269]	2018	Computer vision	1	FlowNetH-Pred-Merged	×
Heo et al. [270]	2018	Medical signal	3	UA	√
Sensoy et al. [274]	2018	Image processing	2	EDL	√
Prokudin et al. [326]	2018	Image processing	3	CVAE	√
Smith et al. [364]	2018	Image processing	2	MI	√
Qian et al. [380]	2018	Image processing	4	MaPML	×
Dunlop et al. [383]	2018	Synthetic data	2	DGP	×
Manders et al. [398]	2018	Image processing	2	CPUA	×
Lee et al. [417]	2018	Image processing	2	GP	√
Acerbi [275]	2018	N/A	2	VBMC	√
Zhang et al. [635]	2018	Numerical data	10	S^2VGD: Structural Stein Variational Gradient Descent	×
Gong et al. [77]	2019	Numerical data	5	Icebreaker	√
Sun et al. [314]	2019	Numerical data	10	functional BNNs(fBNNs)	√
Vadera and Marlin [312]	2019	Image processing	1	BDk	×
Patacchiola et al. [315]	2019	Image processing	2	GP	√
Cheng et al. [316]	2019	Image processing	2	DIP: deep image prior	√
Ravi and Beatson [418]	2019	Image processing	2	AVI	√
Hendrycks et al. [261]	2019	Image processing	2	Self-supervision	√
Ilg et al. [269]	2019	Computer vision	1	SGDR (Stochastic Gradient Descent with warm Restarts) and Bootstrapped ensembles	×
Ahn et al. [272]	2019	Image processing	2	UCL	√
Haußmann et al. [277]	2019	Image processing	2	BEDL (Bayesian Evidential DL) + Reg (Regularization)	√
Foong et al. [279]	2019	Numerical data	9	Laplace approximation	×
Abdolshah et al. [307]	2019	Image processing	1	MOBO	×
White et al. [308]	2019	Image processing	1	BO	√
Balandat et al. [309]	2019	Geographical data	1	BOTORCH	√
Galy-Fajou et al. [310]	2019	Toy datasets	7	CMGGPC	√
Lee et al. [311]	2019	Image processing	4	BTAML	√
Schwab et al. [363]	2019	Image processing	2	CXPlain	√

Continued on next page

TABLE 6 – *Continued from previous page*

Study	Year	Subject	# datasets	Uncertainty method's name	Code
Malinin and Gales [365]	2019	Image processing	5	N/A	√
Wu et al. [379]	2019	Medical image processing	1	DDMN: deep Dirichlet mixture networks	×
Gomez et al. [381]	2019	Image processing	3	Targeted dropout	√
Northcutt et al. [397]	2019	Image processing	2	CL	√
Ovadia et al. [403]	2019	Image, text and categorical data	3	Data shift	√
Biloš et al. [404]	2019	Toy data	2	FD-Dir and WGP-LN	√
Zheng and Yang [405]	2019	Image processing	2	MR	√
Yildiz et al. [415]	2019	Image processing	3	ODE^{2}VAE	√
Wang et al. [420]	2019	Time series	1	NLE loss	√
Tai et al. [421]	2019	Computer vision	1	UIA	×
Foong et al. [636]	2019	Regression (synthetic data)	1	Factorised Gaussian assumption and MC dropout	×
De Ath et al. [278]	2020	Synthetic data	10	$\in$-shotgun	×
Foong et al. [322]	2020	Image processing	5	ConvNP	×
Yao et al. [323]	2020	Image processing and Numerical data	3	SI	√
Prijatelj et al. [324]	2020	Image processing	4	Bayesian evaluation	√
Herzog et al. [325]	2020	Medical image analysis	1	Bayesian aggregation	×
Tuo and Wang [327]	2020	N/A	N/A	BO	×
Acerbi [328]	2020	N/A	5	VBMC+EIG/VIQR	√
Zhao et al. [329]	2020	Numerical data	5	GEP	×
Li et al. [330]	2020	Care (network) data	1	DBGP	×
He et al. [362]	2020	Image processing and toy data	3	NTK	√
Salem et al. [371]	2020	Numerical data	10	SNM-QD+: split normal mixture-quality-driven loss	√
Hendrycks et al. [386]	2020	Image processing	3	AugMix	√
Boiarov et al. [389]	2020	NLP	1	SPSA	×
Chun et al. [399]	2020	Image processing	7	Regularization techniques	×
Wang et al. [411]	2020	Image processing	5	TransCal	×
Lu et al. [419]	2020	Numerical data	2	DGPM	×
Selvan et al. [422]	2020	Medical image analysis	2	cFlow	√
Poggi et al. [423]	2020	Computer vision	1	Self-Teaching	√
Cui et al. [424]	2020	Time series	5	MMD	×

Note: UA: Uncertainty-aware attention, EDL: Evidential Deep Learning, Vadam: Variational Adam, MaPML: Margin Preserving Metric Learning

TABLE 7: A summary of various UQ methods applied in computer vision and image processing tasks (sorted by year)

Study	Year	Data source	Application	# Images/videos	# Classes	Classifier	UQ method	Code
Kendall et al. [473]	2016	aCambridge Landmarks dataset (outdoor), 7 Scenes dataset (indoor)	Camera Re-localization	covering a ground area of up to 50,000 m2 including 1920×1080 images, 640×480 images	4 classes, 7 classes	Bayesian CNN	Averaging MC dropout samples obtained from the posterior Bernoulli distribution of the Bayesian CNN's weights	√
Kendall et al. [48]	2016	SUN RGB-D (Indoor), CamVid (Outdoor)	Scene understanding	SUN RGB-D: 5285 training and 5050 testing images, while images were resized to 224x224, CamVid: 367 training images and 233 testing images of day and dusk scenes, while images were resized to 360x480	CamVid: 11 classes, SUN RGB-D: 37 classes	Bayesian SegNet	MC sampling with dropout	√
Kendall et al. [470]	2017	CamVid and NYU v2, Make3D	Semantic segmentation	600 and 1449, 534	11 and 40	BDL	MC dropout	√
Gal et al. [542]	2017	Synthetic dataset, UCI datasets, MNIST	RL	N/A	N/A	NNs	Continuous relaxation of dropout's discrete masks, using BDL	√
Ilg et al. [269]	2018	Sintel train clean, Sintel train final, KITTI 2012+2015, FlyingThings3D	Optical flow estimation	N/A	N/A	Multi-headed network architecture that yields multiple hypotheses in a single network without the need of sampling	Uncertainty estimates efficiently a single forward pass and without the need for sampling or ensembles	√

Continued on next page

TABLE 7 – *Continued from previous page*

Study	Year	Data source	Application	# Images/videos	# Classes	Classifier	UQ method	Code
Bhattacharyya et al. [496]	2018	Cityscapes dataset	Long-term on-board prediction Of people	2975 training, 500 validation and 1525 test video sequences of length 1.8 seconds (30 frames) having resolution of 2048×1024 pixels	20	RNN encoder-decoder + CNN + LSTM-Bayesian	MC and minimizing the KL divergence of its approximate weight distribution	×
Gast and Roth [486]	2018	FlyingChairs, Sintel, CIFAR10 and MNIST	Uncertainty prediction on CNN	N/A	N/A	Lightweight probabilistic CNNs (FlowNet-ADF and FlowNet-ProbOut and ProbOut)	Apply uncertainty propagating layers using Gaussians (Building upon standard maximum conditional likelihood learning while concentrating on probabilistic outputs)	×
Lambert et al. [637]	2018	ImageNet, Multi-30K	Image classification and machine translation	N/A	30 thousand Flickr images	CNN+LSTM	Proposed LUPI (Learning Under Privileged Information) which makes variance of a function of the privileged information. Then using the privileged information in heteroscedastic dropout to estimate uncertainty	√
Kendall et al. [638]	2018	CityScapes	Scene geometry and semantics	2,975 training and 500 validation images at 2048×1024 resolution. 1,525 images are withheld for testing on an online evaluation server	20	Deep convolutional encoder followed by convolutional decoders	A principled loss function which can learn a relative weighting automatically from the data and is robust to the weight initialization	√

TABLE 7 – *Continued from previous page*

Study	Year	Data source	Application	# Images/videos	# Classes	Classifier	UQ method	Code
Dorta et al. [476]	2018	Splines and Ellipses (synthetic), CelebA, CIFAR10	Image reconstruction	CelebA (202,599 images of faces), CIFAR10(182,637, 19,962)	N/A	Deep probabilistic generative models (AE and VAE)	Structured Gaussian (Gaussian random field)	√
Prokudin et al. [326]	2018	PASCAL 3D+, IDIAP, CAVIAR-o, Town-Centre coarse gaze estimation	Pose estimation	36000, (42304 images for training, 11995 images for validation, 11996 images for testing), (10802 images for training, 5444 images for validation, 5445 images for testing), (6916 images for training, 874 images for validation, 904)	12	Probabilistic deep learning model on top of Biternion network approach and the finite VM mixture model	Probabilistic deep learning model by extending von Mises distributions	√
Le et al. [480]	2018	KITTI	Object detection in safety-critical applications	N/A	N/A	Single Shot Detector (SSD)	A MC integration by sampling the network outputs through the softmax function	×
Pascual et al. [483]	2018	DeepGlobe Land Cover classification challenge	Land cover semantic segmentation	1.146 satellite RGB images of size 2448x2448 pixels, split into training/validation/test, each with 803/171/172 images	7	Gated Convolutional Network (GCN)	Uncertainty GCNN	×
Huang et al. [482]	2018	CamVid	Semantic segmentation in videos	Total 701 labeled frames split into 367 training frames, 101 validation frames and 233 test frames	11	Bayesian SegNet model, Tiramisu model, FlowNet for optical flow estimation	TA-MC (Temporal aggregation method) and RTA-MC (Region-Based Temporal Aggregation)	×
Taha et al. [475]	2019	Clothing1M, Honda driving	Image and video retrieval	1M images, 104 hours of driving	14	CNN+LSTM	Extension to triplet loss that models data uncertainty for each input Also, it models local noise in the embedding space.	×

Continued on next page

TABLE 7 – *Continued from previous page*

Study	Year	Data source	Application	# Images/videos	# Classes	Classifier	UQ method	Code
Zheng [490]	2019	Cast Search in Movies (CSM), IARPA Janus Surveillance Video Benchmark (IJB-S)	Video-based face recognition	N/A	2	CNN	Uncertainty-Gated Graph (UGG) which conducts graph-based identity propagation between tracklets, which are represented by nodes in a graph.	×
Yasarla et al. [639]	2019		Single image de-raining	12700 images	2	UMRL (Uncertainty guided Multi-scale Residual Learning) method based on cycle Spinning+RNN	UMRL	√
Xue et al. [640]	2019	ISIC	Skin lesion classification	3,582 images for training	2	CNN	OUSM (Online uncertainty sample mining method)	×
Hama et al. [620]	2019	MS COCO, Flickr30k	Image-caption embedding-and-retrieval task	COCO (113,287 images for training, 5,000 images for validation, and 5,000 images for testing), Flickr30k (30,000 images for training, 1,000 images for validation, and 1,000 images for testing)	5	DNN+CNN	Model averaging obtained over feature uncertainty+model averaging achieved over posterior uncertainty	×
Mukhoti et al. [10]	2019	Cityscapes dataset	Semantic segmentation	2975 imagesfor training, 500 images for validation, 1525 images for testing	50	Bayesian DeepLab	MC dropout	×

TABLE 7 – *Continued from previous page*

Study	Year	Data source	Application	# Images/videos	# Classes	Classifier	UQ method	Code
Harakeh et al. [479]	2019	Berkley Deep Drive (BDD) 100K Dataset (BDD), KITTI, MS COCO, Pascal VOC	Object detection	BDD: 80K frames (70K/10K training/validation), KITTI: 7; 481 frames, MS COCO: 223K frames (118K/5K training/testing), Pascal VOC: 5823 frames (testing)	BDD and KITTI: 7 common road scene, MS COCO: 81, Pascal VOC: 20	DNN	BayesOD (Bayesian-based object detectors)	√
He et al. [481]	2019	MS-COCO, PASCAL VOC 2007	Object detection	For PASCAL VOC: 20075k voc_2007_trainval images and 5k voc_2007_test images	N/A	Faster R-CNN+FPN (Feature Pyramid Network)+ Mask-R-CNN	A new bounding box regression loss (modeling bounding box predictions as well as ground-truth bounding boxes as Gaussian distribution and Dirac delta function, respectively)	√
Liu et al. [478]	2019	Multiple indoor and outdoor datasets	Depth estimation for 3D scene reconstruction	N/A	N/A	D-Net (CNNbased) + K-net (Kalman filter) + R-Net (based on U-Net with skip connections)	BDL	×
Abbasnejad et al. [641]	2019	GuessWhat dataset	Asking goal-oriented questions	155,281 dialogues related to 821, 955 question/answer pairs with vocabulary size of 11,465 on 66,537 images as well as 134,074 objects	N/A	RNN+LSTM	Bayesian-based model	×
Peterson et al. [493]	2019	CIFAR10H (for training), CIFAR10, IFAR10.1v6,v4, CINIC10, ImageNet-Far for testing	Improving robustness of adversarial attacks	10,000 images, 50,000, 2,000, 210,000, N/A images	N/A	CNN	Increasing distributional shift	√

TABLE 7 – *Continued from previous page*

Study	Year	Data source	Application	# Images/videos	# Classes	Classifier	UQ method	Code
Bertoni et al. [489]	2019	KITTI, nuScenes	3d pedestrian localization	7,481 training images, N/A	N/A	DNN (2D joints) + a light-weight feed-forward network (the 3D location)	MC dropout	√
Asai et al. [477]	2019	NYU depth dataset V2	Depth estimation	N/A	N/A	CNN Pixel-wise regression	Formulating regression with estimation of uncertainty as MTL	×
Loquercio et al. [487]	2019	Udacity dataset	End-to-end steering angle prediction, obstacle motion prediction and closed-loop control of a Quadrotor	N/A	N/A	DNN	Bayesian inference MC	×
Martinez, et al. [484]	2019	CT scans of woven composite materials	Automatically segmenting a diverse set of Volumetric CT scans of woven composite materials	They divided entire 1001x1150x1150 volume into a set of 48 sub-volume for training and sets of 8 for both validation and testing steps	2	CNN (VNet 3D)	MC dropout both during training and inference	×
Postels et al. [485]	2019	CamVid	Semantic segmentation and depth regression	N/A	11 out of 32	CNN	Sampling free noise injection	√
He et al. [488]	2019	Human 3.6M, MPII validation	Human pose estimation	N/A	N/A	CNN	Multivariate Gaussian	×
Peretroukhin et al. [642]	2019	Synthetic dataset, 7-Scenes, KITTI	Probabilistic regression of elements of SO(3)	DNN called HydraNet (extened version of multi-headed networks)	N/A	N/A	Regression units	×
Yu et al. [491]	2019	Market-1501, DukeMTMC-ReID, CUHK01, CUHK03	Person re-identification	N/A	2	Distribution Net (CNN, with random feature vectors)	Gaussian distribution	√

Continued on next page

TABLE 7 – *Continued from previous page*

Study	Year	Data source	Application	# Images/videos	# Classes	Classifier	UQ method	Code
Zhang et al. [643]	2019	Train400, Test12, Berkeley segmentation dataset (BSD68)	Image restoration	N/A	N/A	PRL (Probabilistic Residual Learning)	UQ (aleatoric) with CVAE (Conditional Variational Auto-encoders) loss	√
Carbone et al. [494]	2020	MNIST and Fashion MNIST	Gradient-based adversarial attacks	1000 test images from both datasets	N/A	BNNs	HMC and and VI support	×
Harris et al. [644]	2020	CIFAR-10, Fashion MNIST, ImageNet	Image, audio, text and point cloud classification (mixed sample data augmentation)	N/A	N/A	MixUp, FMix, FastText-300d, CNN, FastText, bi-directional LSTM, BERT	Relative entropy objective	√
Miolane and Holmes [645]	2020	Synthetic datasets	Representation learning	N/A	N/A	VAE	Riemannian VAE	×
Zhou et al. [646]	2020	Synthetic and GTAV, SYNTHA and Cityscapes datasets	Cross-domain semantic segmentation	N/A	N/A	Teacher-student network	UA consistency regularization and uncertainty-guided consistency loss	×
Zhang et al. [647]	2020	Six challenging benchmark datasets	RGB-D saliency detection	N/A	N/A	Conditional VAE	UC-Net	×
Lee and Lee [648]	2020	CASIA-Webface, COMA, 300W-LP, CelebA, AFLW2000-3D	3D face reconstruction	CASIA-Webface: 494k images, COMA: 20,000 different meshes for 12 various subjects, 300W-LP: 60k images, CelebA: 200k images, AFLW2000-3D: N/A	N/A	Graph CNN and GAN	Uncertainty-Aware Mesh Decoder	×

Continued on next page

TABLE 7 – *Continued from previous page*

Study	Year	Data source	Application	# Images/videos	# Classes	Classifier	UQ method	Code
Wang et al. [649]	2020	RAF-DB, FERPlus, AffectNet and WebEmotion	Facial expression	RAF-DB: 30,000 different facial images, FERPlus: 28,709, 3,589 and 3,589 images for training , validation and test, respectively, AffectNet: 450,000 images, WebEmotion: 41,000 videos	N/A	CNN	SCN: Self-Cure Network	√
Yang et al. [650]	2020	KITTI and EuRoC MAV	Monocular visual odometry	N/A	N/A	Self-supervised network	D3VO	×
Chang et al. [651]	2020	MS-Celeb-1M, LFW, MegaFace, CFP, YTF and IJB-C	Face recognition	MS-Celeb-1M: 3,648,176 images, the rest of datasets used: N/A	MS-Celeb-1M: 79,891 subjects, the rest of datasets used: N/A	CNN (ResNet+SE-block)	DUL (Data uncertainty learning)	×
Polic et al. [652]	2020	13 different ETH synthetic datasets and KITTI datasets	Camera model selection	ETH: 14-76 images, 2k-85k 3D points as well as 50k-795k observations, KITTI: N/A	N/A	N/A	AC (Accuracy-based Criterion), ACS (AC-based camera model Selection) and LACS (learning ACS)	√
Nan and Ji [653]	2020	SDS500 and three other standard benchmark datasets	Image de-convolution	BSDS500: 500 latent images, the rest of datasets used: N/A	N/A	CNN (U-Net)	TLS (Total least squares)	×
Kumar et al. [654]	2020	10 different face alignment datasets	Face alignmen	See Table 1 in [654]	See Table 1 in [654]	U-Net	LUVLi loss (Location, Uncertainty, and Visibility Likelihood) with CEN (Cholesky Estimator Network) and VEN (Visibility Estimator Network)	×

TABLE 7 – *Continued from previous page*

Study	Year	Data source	Application	# Images/videos	# Classes	Classifier	UQ method	Code
Cheng et al. [655]	2020	DTU, Tanks and Temple	Multi-stage depth prediction	N/A	N/A	CNN	UCS-Net: Uncertainty-aware Cascaded Stereo Network	√
Tang et al. [656]	2020	AQA-7, MTL-AQA and JIGSAWS	Action quality assessment	N/A	N/A	CNN	MUSDL: Multi-path uncertainty-aware score distributions learning	√
Carvalho et al. [657]	2020	CamVid	Semantic segmentation and pixel-wise depth regression	N/A	N/A	CNN	Functional VI	×
Angelopoulos et al. [658]	2020	ImageNet and ImageNet-V2	Image classifiers	N/A	N/A	CNN	Conformal prediction	√

TABLE 8: A summary of various UQ methods applied in medical application tasks (sorted by year).

Study	Year	Classifier	Application	Disease/cancer	# datasets	UQ method	Code
Leibig et al. [50]	2017	CNN	Classification	Diabetic Retinopathy	3	MC dropout	×
Jungo et al. [659]	2017	CNN	Segmentation	Brain tumor	1	MC dropout	×
Ozdemir et al. [660]	2017	Bayesian CNN	Segmentation	Nodule detection	1	VI	×
Tanno et al. [661]	2017	CNN	Classification	Brain tumor	2	VI	×
Jungo et al. [57]	2018	CNN	Segmentation	Brain tumor	1	MC dropout	×
Kwon et al. [662]	2018	BNNs	Classification	Cardiovascular	2	VI	×
Ayhan and Berens [663]	2018	DNNs (ResNet50)	Data augmentation	Fundus images	1	MC dropout	×
Jungo et al. [664]	2018	U-net	Segmentation	Brain tumor	2	MC dropout	×
Wang et al. [665]	2018	CNN	Segmentation	Brain tumor	1	Weighted loss function, network-based uncertainty (Softmax output) and scribble-based uncertainty (the geodesic distance to different scribbles)	×
Moccia et al. [666]	2018	CNN	Classification and tagging	Surgical data of various diseases	2	Superpixel (Spx)-based classification of anatomical structure	×

TABLE 8 – *Continued from previous page*

Study	Year	Classifier	Application	Disease/cancer	# datasets	UQ method	Code
McClure et al. [667]	2018	CL	Segmentation	Brain segmentation	5	DWC: Distributed weight consolidation	×
Wang et al. [38]	2019	CNN	Segmentation	Brain tumor	1	Ensemble	×
Tousignant et al. [60]	2019	CNN	Classification	Disability progression (brain)	2	MC dropout	×
Roy et al. [62]	2019	Bayesian QuickNAT	Segmentation	Brain segmentation	4	MC samples for voxel-wise model	√
Jungo and Reyes [668]	2019	U-Net	Segmentation	Brain tumor	2	Softmax entropy, MC dropout, Ensembles	√
Orlando et al. [669]	2019	U-Net	Segmentation	OCT scans	2	MC dropout	×
Ghesu et al. [670]	2019	DenseNet-121	Classification	Thoracic disease	2	MC dropout	×
Baumgartner et al. [671]	2019	U-Net	Segmentation	Thoracic and prostate	2	Ensemble	√
Raczkowski et al. [672]	2019	Bayesian CNN	Classification	Colorectal cancer	1	VI	×
Xue et al. [640]	2019	ResNet-101	Classification	Skin cancer	1	Ensemble	×
Eaton-Rosen et al. [673]	2019	U-Net	Segmentation and regression	Histopathological cell and white matter hyperintensity counting	2	MC dropout	×
di Scandalea et al. [674]	2019	U-Net	Segmentation	Axon myelin	2	MC dropout	√
Filos et al. [675]	2019	BDL	Classification	DR	1	MC dropout, ensembles and VI	×
Ravanbakhsh et al. [676]	2019	conditional GAN (cGAN)	Semantic segmentation	Cardiovascular disease	1	The scores generated by using the adversarial discriminator	×
Jena and Awate. [677]	2019	Bayesian DNN	Segmentation	Brain tumor, cell membrane and chest Radiograph organ	3	MC dropout	×
Tanno et al. [678]	2019	CNN	Classification	Brain tumour (Glioma) and MS (multiple sclerosis)	4	VI	×
Soberanis-Mukul et al. [679]	2019	CNN and GCN	Segmentation	Organ segmentation (pancreas)	1	MC dropout	×
Hu et al. [680]	2019	Probabilistic U-Net	Segmentation	Lung nodule CT dataset and MICCAI2012 prostate MRI	2	VI	×
Hu et al. [222]	2020	U-net and the Adaptive-CS-Net	MRI reconstruction and Curve fitting	Knee and brain MRI	2	MC dropout and deep ensembles	×

Continued on next page

TABLE 8 – *Continued from previous page*

Study	Year	Classifier	Application	Disease/cancer	# datasets	UQ method	Code
Luo et al. [681]	2020	DCN (Deep commensal network)	Segmentation	Cardiovascular disease	4	MC dropout	×
Hoebel et al. [682]	2020	U-Net	Segmentation	Lung disease	1	MC dropout	×
Liu et al. [506]	2019	CNN	Classification	sO_2	2	DSL (deep spectral learning)	×
Seeböck et al. [683]	2019	Bayesian U-Net	Segmentation	Retinal OCT scans	3	MC dropout	×
Hiasa et al. [684]	2019	Bayesian U-Net	Segmentation	Cancer	2	MC dropout	×
Xue et al. [685]	2019	BNN and U-Net	Classification	Gigapixel phase images	1	MC dropout and ensembles	×
LaBonte et al. [686]	2019	3D Bayesian CNN	Segmentation	Material data (CT scans)	2	VI	√
Liao et al. [687]	2019	DenseNet, LSTM	Regression	Cardiovascular diseases	1	MC dropout	×
Raghu et al. [688]	2019	N/A	Classification	DR	1	DUP (Direct Uncertainty Prediction)	√
Zhang et al. [689]	2019	ResNet	Reconstruction and classification	Knee MRI	2	Active acquisition	×
Ye et al. [390]	2020	Separable dictionary-LSTM	Tissue micro-structure estimation	Brain dMRI (diffusion magnetic resonance imaging) scans	1	Residual bootstrap strategy	×
Xia et al. [457]	2020		Segmentation	Pancreas and liver tumor	2	UMCT	×
Gantenbein et al. [467]	2020	Revesible PHiSeg	Segmentation	Lung and prostate	2	Variational function	×
Bian et al. [513]	2020	CNN	Segmentation	Cardiovascular and retinal OCT	2	UESM (Uncertainty Estimation and Segmentation Module) + UCE (Uncertainty-aware Cross Entropy) loss	×
Donnat et al. [690]	2020	Hierarchial bayesian network	Classification	COVID-19	1	Stochastic Expectation-Maximization	×
Mehrtash et al. [691]	2020	U-Net	Segmentation	Brain, heart and prostate	5	Ensembles	×
Wickstrøm et al. [692]	2020	CNN	Segmentation	Colorectal cancer	1	MC dropout	×
Carneiro et al. [693]	2020	ResNet, DenseNet	Classification	Polyp	2	MC integration	×
Natekar et al. [694]	2020	DenseUnet, ResUnet, SimUnet	Segmentation	Brain tumor	1	TTD (test time dropout) for VI	×

TABLE 8 – *Continued from previous page*

Study	Year	Classifier	Application	Disease/cancer	# datasets	UQ method	Code
Li and Alstrøm [516]	2020	ResNet	Segmentation	Colon and skin cancers	2	MC dropout	×
Dahal et al. [511]	2020	ResNet	Segmentation	Cardiovascular disease	2	TTA, HSE (Horizontal stacked ensemble), MC dropout	×
Li et al. [695]	2020	MLP (Multilayer Perceptron)	Segmentation	Autism	2	DistDeepSHAP (Distribution-based Deep Shapley value explanation)	√
Zheng et al. [696]	2020	ag-FCN (attention gated fully convolutional network)	Segmentation	Glands and infant brain tissues	2	dd-AL (Distribution discrepancy-based AL)	×
Wang et al. [697]	2020	ResNet	Classification	Lung disease and DR	2	DRLA (Deep Reinforcement AL)	×
Quan et al. [698]	2020	ResNet	Classification	EGD (Esophagogastroduodenoscopy)	2	Bayesian uncertainty estimates and ensemble	×
Yuan et al. [699]	2020	DNNs	Classification	Brain cell	2	Bayesian uncertainty	×
Chiou et al. [700]	2020	CycleGAN	Segmentation	Prostate lesion	2	Gaussian sampling	×
Wang et al. [701]	2020	Teacher-student model	Segmentation	Left Atrium and kidney segmentation	2	Double-uncertainty weighted	×
Li et al. [702]	2020	ResUNet	Segmentation	Organ and skin lesion segmentation	2	Self-loop uncertainty	×
Yang et al. [703]	2020	Deep Q learning and Dual-UNet	Segmentation	Catheter segmentation	1	Hybrid constraints	×
Venturini et al. [704]	2020	U-Net	Segmentation	Brain volumes and healthy pregnant females	2	test-time augmentation and test-time dropout	×
Yu et al. [705]	2020	ResNet	Classification	Glaucoma	2	FusionBranch	×
Huang et al. [706]	2020	ReLayNet	Segmentation	OCT (Optical coherence tomography)	1	MC dropout	×
Khairnar et al. [707]	2020	Bayesian CNN (BCNN)	Classification	Breast cancer	1	Modified BCNN	×
Ghesu et al. [708]	2020	Encoder-decoder network, U-Net, CNN)	Classification and segmentation	MRI scans, Prostate, Lung, Colorectal and Ovarian cancer	3	Uncertainty-driven Bootstrapping and Dempster-Shafer evidence theory	×
Eggenreich et al. [709]	2020	BCNN	Regression	Bone age prediction	1	VI and BCNNs	×

TABLE 8 – *Continued from previous page*

Study	Year	Classifier	Application	Disease/cancer	# datasets	UQ method	Code
Soberanis-Mukul et al. [710], [711]	2020	U-Net	Segmentation	Organ segmentation	2	GCNs	√
Prassni et al. [712]	2020	Random walker	Segmentation	Volume segmentation	N/A	Guided probabilistic volume segmentation	×
Ulmer et al. [713]	2020		Classification	Clinical data	2	OoD detection	×
Cao et al. [714]	2020	Dense U-Net	Segmentation	Automated breast ultrasound	2	Temporal ensembling	×

TABLE 9: A summary of various UQ methods used in NLP (sorted by year).

Study	Year	Published	# dataset	Classifier	UQ method	Code
Chien et al. [517]	2015	IEEE TNNLS	3	Bayesian learning	R-BRNN-LM	×
Chen et al. [519]	2015	JBI	1	Active learning	Uncertainty sampling	×
Ott et al. [524]	2018	ICML	3	Deep learning	Beam search and sampling	√
Zhang et al. [522]	2018	NeurIPS	5	Deep learning	BDLOB	×
Yijun et al. [527]	2019	AAAI	4	Deep learning	N/A	×
Kong et al. [520]	2019	ICLR	2	Deep learning	Element-wise variational smoothing	×
Tran et al. [121]	2019	NeurIPS	1	Deep learning	Bayesian layers	×
Fortunato et al. [124]	2019	arXiv	1	Deep learning	BBB (Bayesian RNN)	√
Zhang et al. [256]	2019	ICML	2	Deep learning	LIME	×
Roldán et al. [526]	2019	ACM FAccT	3	Black-box models (Deep learning)	Uncertainty wrapper using Dirichlet distribution	×
Xiao et al. [523]	2019	NeurIPS	2	Deep learning	Variational transformers	×
Han et al. [521]	2019	TST	4	Deep learning	ANFU	×
Zhang and Mahadevan [528]	2020	DSS	1	Deep learning	BNN	×