Sparse Bayesian Networks: Efficient Uncertainty Quantification in Medical Image Analysis

Zeinab Abboud[1] Herve Lombaert[1] and Samuel Kadoury[1,2]

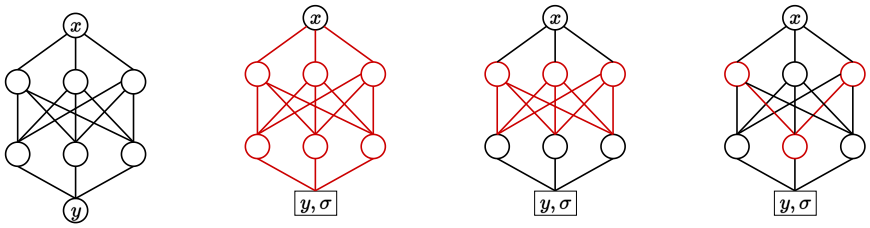
[1] Polytechnique Montreal, Montreal, QC, Canada
[2] CHUM Hospital Research Center, Montreal, QC, Canada
zeinab.abboud@polymtl.ca

Abstract. Efficiently quantifying predictive uncertainty in medical images remains a challenge. While Bayesian neural networks (BNN) offer reliable predictive uncertainty, they require substantial computational resources to train. Although Bayesian approximations such as ensembles have shown promise, they still suffer from high training costs. Existing approaches to reducing computational burden primarily focus on lowering the costs of BNN inference, with limited efforts to improve training efficiency and minimize parameter complexity. This study introduces a training procedure for a sparse (partial) Bayesian network. Our method selectively assigns a subset of parameters as Bayesian by assessing their deterministic saliency through gradient sensitivity analysis. The resulting network combines deterministic and Bayesian parameters, exploiting the advantages of both representations to achieve high task-specific performance and minimize predictive uncertainty. Demonstrated on multi-label ChestMNIST for classification and ISIC, LIDC-IDRI for segmentation, our approach achieves competitive performance and predictive uncertainty estimation by reducing Bayesian parameters by over 95%, significantly reducing computational expenses compared to fully Bayesian and ensemble methods.

Keywords: Uncertainty · Sparsity · Sensitivty Analysis · Bayesian Networks · Segmentation · Classification

1 Introduction

Unlocking the full potential of deep learning (DL) diagnostic systems in medical imaging crucially depends on precision in gauging predictive uncertainty. With a firm grasp on uncertainty, effectively quantifying and conveying the risks tied to model predictions becomes easier for clinicians [23]. As opposed to qualitative measures of uncertainty, such as saliency maps [27], predictive uncertainty is a quantitative measure of confidence or lack thereof in model prediction [23]. DL algorithms often fail to estimate model uncertainty, leading to unreliable predictions with overconfident false predictions [12,10]. This issue arises as DL models commonly assign poorly calibrated probabilities, posing risks in interpretation and decision-making [12,10]. BNNs represent parameters as random variables,



(a) Deterministic (b) Bayes (c) Layer-Partial (d) Sparse-Partial

Fig. 1: Various model implementations of deterministic (a), Bayesian (b), and partial Bayesian (c, d), where black connections are deterministic, and red are probabilistic (Bayesian). Partial Bayesian models can be implemented in two distinct approaches: (c) where a single or multiple layers can be set as Bayesian, or (d) sparse approach where a selected number of connections can be set as Bayesian.

characterized by distributions rather than single point estimates (Figure 1b). BNNs, therefore, provide average predictions and uncertainty estimates by sampling from parameter distributions. The major limitations of BNNs stem from their significant computational costs and noisy loss landscape associated with increased training complexity and high parameter count [18].

Recent attempts have been made to reduce the computational costs associated with Bayesian *inference* by post-hoc processing methods [17,31,29]; however, they do not address the costs associated with training. Few studies have utilized a strategy of conducting Bayesian inference on the output layer of a given neural network [3,18,30] as a means of reducing the computational overhead (see Figure 1c). While previous research has explored strategies for optimizing Bayesian layer placement and justifying its selection [36,26,30], there remains a lack of investigation into further decreasing the parameter complexity of Bayesian training and inference. A Bayesian subnetwork selection using second-order Hessian analysis approximating a fully Hamiltonian-Monte-Carlo (HMC) [25] Bayesian network was proposed, achieving competitive uncertainty quantification [6]. However, the method requires computing and storing the full covariance matrix, making it impractical for more complex tasks such as semantic segmentation [6]. While there exist other Bayesian approximation methods, such as Monte-Carlo (MC) dropout [11] and deep ensembles [21], they have their limitations. MC-dropout exhibits overconfidence, posing risks in safety-critical applications, while deep ensembles incur high computational costs due to the requirement of training and storage of multiple models for inference.

In contrast to Bayesian models, there have been successful attempts to utilize sparsity for reducing model parameter complexities in deterministic models [13,22,7,8]. Parameter sparsity involves pruning connections within a neural network to accelerate model convergence and reduce computational overhead. Early studies involve reducing network size by pruning the weights and re-training

through saliency analysis [22]. LeCun et al. [22] show that reducing the number of parameters by more than 30% resulted in no impact on the model performance. Other methods have used second-order Hessian-based analysis [24,14] to prune weights; however, such approaches are impractical with larger and more complex modern networks. More recent approaches introduce parameter sparsity through pre-defined randomized sparsity [7], structured sparsity [34], and dynamic in-training drop-and-grow algorithm [8,9]. Introducing sparsity in neural networks has improved model performance, accelerated convergence, and regularized the network, improving model generalization [16].

This study aims to reduce the computational cost of training Bayesian neural networks by decreasing the number of Bayesian parameters. We achieve this by:

1. Promoting sparsity among Bayesian parameters based on strong deterministic predictive performance (Figure 1d).
2. Introducing a training method to integrate mixed deterministic and Bayesian parameters into the network architecture.
3. Demonstrating our training approach for medical image classification and segmentation, achieving competitive performance with over 95% reduction in Bayesian parameters.

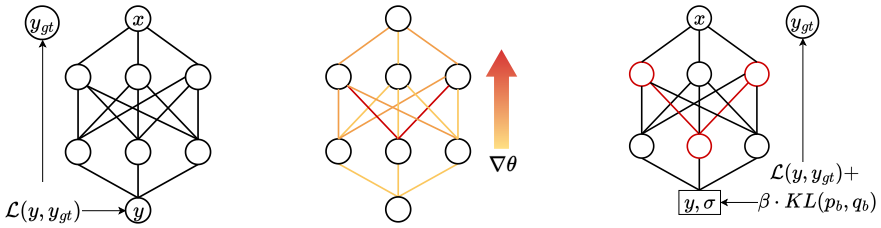
The proposed method, depicted in Figure 2, initializes a partial Bayesian NN[3] using estimated points from a trained deterministic NN. Specific parameters are designated as Bayesian based on their saliency, with the degree of "Bayesian-ness" controlled by a hyperparameter. Experimental results in classification and segmentation demonstrate that assigning 1% the network parameters as Bayesian can yield high-quality predictive uncertainty while preserving model performance. This approach accelerates BNN convergence with significant savings in computational cost.

2 Methods

2.1 Problem Setup

Consider a dataset $\mathcal{D} = \{x_n, y_n\}_{n=1}^{N}$ where $x_n \in \mathbb{R}^{W \times H \times C}$, where W, H, C, represent the width, height, and number of channels for an input image, respectively. For segmentation $\{y_n^{(r)} \in \mathcal{Y}^{W \times H}\}$ represents the mask for rater r, for classification $\{y_n \in 1, ..., c\}$ represents the categorical class c. The goal is to train a neural network to model the probabilistic distribution $p_\theta(y|x)$ where θ represents the network parameters. Let θ_d represent the deterministic point estimates, and θ_b represent Bayesian parameters, such that $\theta_b^{(i)}$ for the i^{th} parameter is parameterized by $\{\mu_i, \sigma_i\}$, the mean and standard deviation of the Gaussian distribution $\mathcal{N}_i(\mu_i, \sigma_i)$. The objective is to learn the probabilistic distribution of $p_\theta(y|x)$ where $\theta = \{\theta_d, \theta_b\}$ comprises a blend of deterministic and Bayesian parameters and reduce the number of Bayesian parameters to optimize model complexity and performance.

[3] Note we will use sparse and partial Bayesian interchangeably, as promoting sparsity in the Bayesian parameters renders the network *partially* Bayesian.



Step 1: Deterministic **Step 2**: Sensitivity Analysis **Step 3**: Sparse Bayes

Fig. 2: Our proposed training of sparse (partial) Bayesian network. **Step 1**: Train a deterministic model by minimizing the negative log likelihood $\mathcal{L}(y, y_{gt})$ where the parameters are represented as point estimates. **Step 2**: Perform a gradient-based sensitivity analysis, denoted as $\nabla\theta$, and identify the Topk connections corresponding to the highest gradients (in red). **Step 3**: Train a sparse (partial) Bayesian model with the Topk connections as Bayesian parameters and the remaining network as deterministic by minimizing the evidence lower bound (ELBO) loss $\mathcal{L}(y, y_{gt}) + \beta \cdot D_{KL}\left(q_b(\theta), p_b(\theta)\right)$, where $p_b(\theta)$ and $q_b(\theta)$ are the prior and posterior distributions for the θ_b Bayesian parameters.

2.2 Sparse Bayesian Networks

We introduce a simple procedure to train a mixed-parameter model, regardless of the chosen architecture or application: (1) train a deterministic model, (2) compute parameter gradients to select the Topk predictive parameters, and (3) train a partial Bayesian model. The following is a detailed description of each step:

(1) Train a deterministic model: given a dataset $\mathcal{D}$, train a neural network $f_d(\cdot)$ by minimizing the negative log likelihood $\mathcal{L} = -\log p_\theta(y|x)$ (Figure 2-Step 1).

(2) Sensitivity analysis: Performing a gradient-based sensitivity analysis is seamlessly integrated into the backward pass through the network, thus incurring no additional computational overhead. Subsequently, select the Topk parameters, i.e. Topk($|\nabla\theta_i|$), where k is the number of parameters designated as Bayesian. (see Figure 2-Step 2).

(3) Train a partial Bayesian model: A sparse masked-gradient approach is employed to have a mixture of deterministic and Bayesian parameters within the network layers.

Initialization: Firstly, point estimates from the deterministic model are set to the mean μ_i parameter values to initialize the partial Bayesian model. This

initialization ensures a robust starting point for optimizing uncertainty learning within the Bayesian network. Then, for each layer within the model, a mask is generated based on the Topk gradient magnitudes of its weights. These masks serve two primary purposes: initializing the σ_i parameter associated with Bayesian weights and facilitating sparse gradient updates. The masks, therefore, enable modeling mixed-parameter types within a given layer, i.e., deterministic and Bayesian. The deterministic weights do not contribute to the Kullback-Leibler (KL) divergence term. Consequently, deterministic weights are represented as $\delta_i(\mu_i)$ Dirac-delta functions, whereas Bayesian weights are modeled as normal distributions $\mathcal{N}_i(\mu_i, \sigma_i)$. The extent of the Bayesian expression of the model is hyper-parameterized by r_{bayes}, which determines the rate (or percentage) of Bayesian parameters over the total number of parameters.

Training: The partial Bayesian model is trained with variational inference [19] using the reparameterization trick [20]. Five posterior samples were used at train and inference time to reduce the computational overhead associated with larger posterior samples. The network is trained by minimizing the evidence lower bound (ELBO) loss $\mathbb{E}\left[-\log p_\theta(y|x)\right] + \beta \cdot D_{KL}\left(q(\theta_b), p(\theta_b)\right)$ where β is the weight of the KL divergence term that minimizes the difference between the prior distribution $p(\theta_b)$ and the posterior $q(\theta_b)$. The prior distribution is the standard normal $\mathcal{N}(0, 1)$ (see Figure 2-Step 3).

It is important to emphasize that the training procedure described above applies to supervised training, regardless of the task. The experimental setup presented next aims to illustrate the effectiveness of our proposed training scheme for both medical image classification and segmentation tasks, enabling the estimation of uncertainty in both objectives.

3 Experimental Results

The objective of the following experimental setup is to showcase the applicability of the training paradigm in both medical image classification and segmentation tasks rather than achieving state-of-the-art performance. A standard deterministic model, a fully Bayesian model, and an ensemble approach are comparative baselines for task performance and uncertainty quantification. The proposed method is demonstrated on the classification of ChestMNIST dataset [35,33] and binary segmentation of LIDC-IDRI [1,2,4] and ISIC [5,32] datasets. ChestMNIST contains 112,120 X-Ray images from 30,805 patients with 78,468/11,219/22,433 as a train/validation/test split with an input size of 224×224. LIDC-IDRI contains 1,018 clinical thoracic CT scans, cropped into 15,096 patches of size 128×128, each with four manual segmentation masks; the data is divided into four bags for 4-fold train/test splits (0.75/0.25). ISIC dataset of skin lesions contains 2,594/100/1,000 train/validation/test images, all resized to $224 \times 224 \times 3$.

3.1 Evaluation Metrics & Training Details

The performance metrics are determined by the primary evaluation criteria specific to each dataset. For ChestMNIST classification, accuracy and the area under the receiver operating characteristic curve (AUC). The reported multi-label AUC score is the average of the binary AUC scores. Dice and intersection-over-union (IoU) scores evaluate segmentation performance for LIDC-IDRI and ISIC datasets, respectively. The Brier score is used as a proper scoring rule for evaluating the accuracy of the output probabilities Brier Score $= \frac{1}{c}\sum_{i=1}^{c}(\delta_c^* - p(y = c \mid x^*))^2$ where $\delta_c^* = 1$ if $c = y^*$, and $\delta_c^* = 0$ otherwise. Entropy of Expectation is used to compute the total uncertainty EoE $= -p\left(y = c \mid x\right)\log\left(p\left(y = c \mid x\right)\right)$. Lastly, the Expected Calibration Error [12] is used to evaluate model calibration. Floating point operations (FLOPs) are used to measure training computational cost.

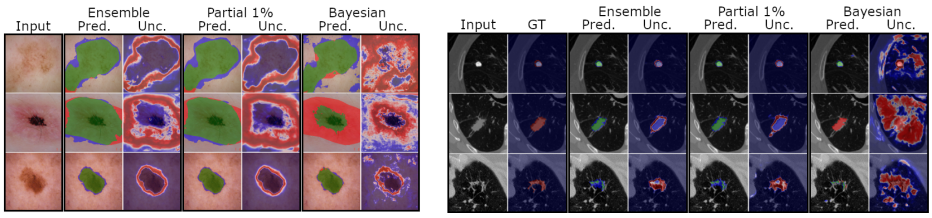
Deterministic and partial Bayesian models were trained for 50 epochs, with a fixed learning rate of 0.01, and a weight decay of 1×10^{-5} for L_2 regularization with a stochastic gradient descent optimizer. Batch normalization was used in all models, with a batch size of 50 for classification and 10 for segmentation. The models were trained using a weighted cross-entropy loss to address class imbalances effectively. All ensemble models for comparison were executed with five ensemble members. For all datasets, we demonstrate the impact of increasing the Bayesian expression of the network on the performance and uncertainty estimation, given a fixed computational budget by varying $r_{\text{bayes}} = (1\%, 5\%, 10\%, 20\%, 40\%, 80\%)$. The partial and fully Bayesian models were trained with ELBO loss and a static weight for the KL term of $\beta = 0.01$ for classification and annealed for segmentation from 0.2 to 0.01 as a function of epochs $\beta_i = \beta_{target} - (\beta_{target} - \beta_{init}) \times (epoch_i / epoch_{total})$. Models were trained with NVIDIA RTX A6000 GPUs. Code available at repository.

3.2 ChestMNIST Classification

We first evaluate a ResNet18 [15] on the multi-label classification of ChestMNIST dataset [35,33]. Figure 4 summarizes the comparative performance of sparse Bayes with different r_{bayes} values, with ensemble and fully Bayesian approaches. With increasing r_{bayes} and a fixed computational budget, there's a decline in test performance metrics, both on classification error and uncertainty metrics (Figure 4). Comparing the performance of a 1% model to a 5-member ensemble, it's evident that the sparse model performs competitively, particularly in uncertainty metrics with $\sim 80\%$ fewer parameters compared to ensembles. The fully Bayesian model performs poorly, with a 20% decline in accuracy and a $10\times$ entropy and $30\times$ FLOPs of the 1% partial model. For further details and metrics, please refer to the supporting information (SI) for numerical data.

3.3 LIDC-IDRI & ISIC Segmentation

Next, we evaluate a UNet [28] on the segmentation of LIDC-IDRI and ISIC datasets. For LIDC-IDRI, we randomly picked a mask from the four ground



(a) ISIC (b) LIDC-IDRI

Fig. 3: Segmentation samples for a 5-member ensemble, 1% partial Bayesian, and fully Bayesian models with input image on the far left. Predictions mask overlays show true positive (green), false positive (blue), and false negative (red). The uncertainty map is the entropy of the output probability, showing regions of high (red) and low (blue) uncertainty. (b) LIDC-IDRI includes inter-rater variability (2nd column). Our partial 1% is at par with ensembles at a lower cost. (Zoom in for a better view of the details.)

truth masks at each training step. Our experiments demonstrate the effectiveness of our training recipe with sparse Bayesian parameter representation, achieving competitive performance against benchmark Bayesian and ensemble methods (Figure 4, Table 1 in supplementary information (SI)). Setting only 1% of the network's parameters to Bayesian yields comparable performance to ensembles while requiring fewer FLOPs by 30% and significantly fewer parameters 80% less than ensembles (FLOPs details in SI). Comparing the 1% partial Bayesian model to the fully Bayesian model (Figure 4 and Figure 3), our training approach demonstrates significantly better performance with a fraction of the required FLOPs. For instance, in the segmentation of LIDC lung nodules, the Partial 1% model achieves a Dice score of 0.80 ± 0.01 and an entropy of 0.010 ± 0.001, while the fully Bayesian model achieves a Dice score of 0.67 ± 0.09 and an entropy of 0.3 ± 0.1. Moreover, our approach incurs less than 20% of the computational cost (FLOPs) of training the fully Bayesian model and approximately 50% fewer parameters.

Qualitatively, Figure 3 demonstrates the consistency of our model across both segmentation tasks. In particular, Figure 3b illustrates how our approach aligns with the ground truth majority vote and the uncertainty of the ground truth segmentation for LIDC-IDRI. Despite low Bayesian rates (1%), the partial Bayesian model effectively expresses uncertainty, particularly evident in regions with high uncertainty in LIDC-IDRI and ISIC datasets, consistent with areas of higher ambiguity (second example), such as near borders of ROIs or unclear boundaries of skin lesions (Figure 3a). Conversely, training a fully Bayesian network results in inferior performance, higher uncertainty, and higher false negative rate (demonstrated in the red overlay regions), with significantly increased computational costs due to slow model convergence (Figure 4). The distribution of the Bayesian parameters per our selection criterion is consistent with previous findings [30,36], where the Topk selected parameters are distributed near the last

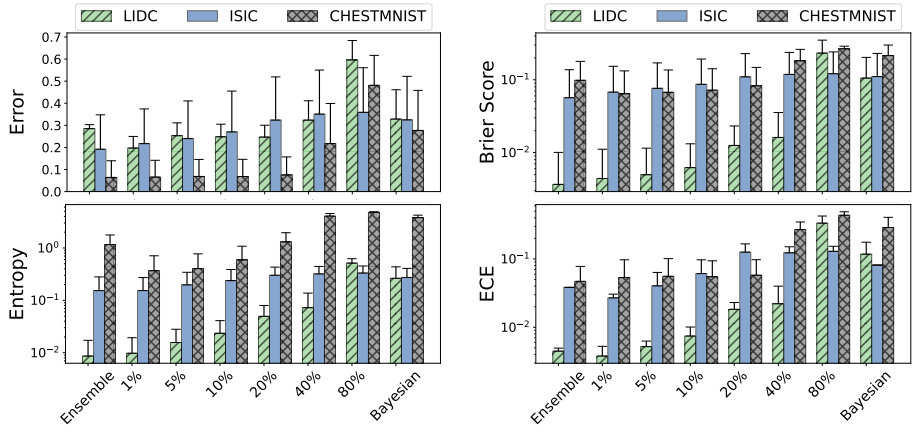


Fig. 4: Performance comparison of partial Bayesian models with $r_{\text{bayes}} = (1\%, 5\%, 10\%, 20\%, 40\%, 80\%)$, 5-member ensembles, and fully Bayesian models for classification and segmentation tasks. Bars represent the mean ± standard deviation. Error is computed as $(1 - \text{accuracy/Dice/IoU})$. Numerical data in SI.

block, input, and output layers offering a computationally low-cost approach for Bayesian subnetwork selection as opposed to second-order Hessian-based criterion [6].

4 Conclusion

This paper presents a training procedure for efficiently training sparse Bayesian networks with variational inference, achieving competitive performance and predictive uncertainty estimation compared to fully Bayesian and ensemble methods. Our approach demonstrates significantly fewer parameters and lower computational requirements without compromising task performance. Specifically, a network with only 1% Bayesian parameters matches or surpasses ensemble performance and consistently outperforms fully Bayesian networks with orders of magnitude fewer FLOPs. This enables more cost-effective predictive uncertainty quantification without compromising performance, thus facilitating uncertainty-guided decision-making in medical image analysis. Future research will establish the minimum Bayesian parameters necessary for a given architecture to express uncertainty effectively.

Acknowledgments. We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) for PGS-D award, NSERC Discovery grant, and Canada Research Chair in Shape Analysis in Medical Imaging.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Armato III, S.G., McLennan, G., Bidaut, et al.: The lung image database consortium and image database resource initiative: a completed reference database of lung nodules on CT scans. Medical Physics **38**(2) (2011)
2. Armato III, S.G., McLennan, G., Bidaut, L., et al.: Data From LIDC-IDRI. The Cancer Imaging Archive (2015), Data set
3. Azizzadenesheli, K., Brunskill, E., Anandkumar, A.: Efficient exploration through Bayesian deep Q-networks. In: 2018 Information Theory and Applications Workshop. IEEE (2018)
4. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., Prior, F.: The Cancer Imaging Archive: Maintaining and Operating a Public Information Repository. Journal of Digital Imaging **26**(6) (2013)
5. Codella, N., Rotemberg, V., Tschandl, P., et al.: Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration. arXiv preprint arXiv:1902.03368 (2019)
6. Daxberger, E., Nalisnick, E., Allingham, J.U., Antorán, J., Hernández-Lobato, J.M.: Bayesian deep learning via subnetwork inference. In: ICML. PMLR (2021)
7. Dey, S., Huang, K.W., Beerel, P.A., Chugg, K.M.: Pre-defined sparse neural networks with hardware acceleration. IEEE Journal on Emerging and Selected Topics in Circuits and Systems **9** (2019)
8. Evci, U., Gale, T., Menick, J., Castro, P.S., Elsen, E.: Rigging the lottery: Making all tickets winners. In: ICML. PMLR (2020)
9. Frankle, J., Carbin, M.: The lottery ticket hypothesis: Finding sparse, trainable neural networks. In: ICLR (2019)
10. Gal, Y.: Uncertainty in deep learning. Ph.D. thesis, University of Cambridge (2016)
11. Gal, Y., Ghahramani, Z.: Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In: ICML. PMLR (2016)
12. Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q.: On calibration of modern neural networks. In: ICML. vol. 70. PMLR (2017)
13. Guo, Y., Yao, A., Chen, Y.: Dynamic network surgery for efficient DNNs. In: NeurIPS. vol. 29 (2016)
14. Hassibi, B., Stork, D.: Second order derivatives for network pruning: Optimal brain surgeon. In: NeurIPS. vol. 5 (1992)
15. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on CVPR (2016)
16. Hoefler, T., Alistarh, D., Ben-Nun, T., Dryden, N., Peste, A.: Sparsity in deep learning: Pruning and growth for efficient inference and training in neural networks. JMLR **22** (2021)
17. Jia, X., Yang, J., Liu, R., Wang, X., Cotofana, S.D., Zhao, W.: Efficient computation reduction in Bayesian neural networks through feature decomposition and memorization. IEEE Transactions on Neural Networks and Learning Systems **32** (2021)
18. Jospin, L.V., Laga, H., Boussaid, F., Buntine, W., Bennamoun, M.: Hands-on Bayesian neural networks—A tutorial for deep learning users. IEEE Computational Intelligence Magazine **17** (2022)
19. Kingma, D.P., Welling, M.: Auto-encoding variational Bayes. In: ICLR. vol. 28 (2014)

20. Kingma, D.P., Salimans, T., Welling, M.: Variational dropout and the local reparameterization trick. In: NeurIPS. vol. 28 (2015)
21. Lakshminarayanan, B., Pritzel, A., Blundell, C.: Simple and scalable predictive uncertainty estimation using deep ensembles. In: NeurIPS. vol. 30 (2017)
22. LeCun, Y., Denker, J., Solla, S.: Optimal brain damage. In: NeurIPS. vol. 2 (1989)
23. McCrindle, B., Zukotynski, K., Doyle, T.E., Noseworthy, M.D.: A radiology-focused review of predictive uncertainty for AI interpretability in computer-assisted segmentation. Radiology: AI **3** (2021)
24. Mozer, M.C., Smolensky, P.: Skeletonization: A technique for trimming the fat from a network via relevance assessment. In: NeurIPS. vol. 1 (1988)
25. Neal, R.M.: Bayesian Learning for Neural Networks. Ph.D. thesis, University of Toronto, Toronto, Canada (1995)
26. Prabhudesai, S., Hauth, J., Guo, D., et al.: Lowering the computational barrier: Partially Bayesian neural networks for transparency in medical imaging AI. Frontiers in Computer Science **5** (2023)
27. Reyes, M., Meier, R., Pereira, S., et al.: On the interpretability of artificial intelligence in radiology: challenges and opportunities. Radiology: AI **2** (2020)
28. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI (2015)
29. Sharma, H., Jennings, E.: Bayesian neural networks at scale: a performance analysis and pruning study. The Journal of Supercomputing **77** (2021)
30. Sharma, M., Farquhar, S., Nalisnick, E., Rainforth, T.: Do Bayesian neural networks need to be fully stochastic? In: AISTATS. PMLR (2023)
31. Subedar, M., Krishnan, R., Kashyap, S.N., Tickoo, O.: Quantization of Bayesian neural networks and its effect on quality of uncertainty. In: Workshop on Uncertainty and Robustness in Deep Learning, ICML (2021)
32. Tschandl, P., Rosendahl, C., Kittler, H.: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific data **5**(1) (2018)
33. Wang, X., Peng, Y., Lu, L., et al.: Chestx-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE conference on CVPR (2017)
34. Wen, W., Wu, C., Wang, Y., Chen, Y., Li, H.: Learning structured sparsity in deep neural networks. In: NeurIPS. vol. 29 (2016)
35. Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., Ni, B.: MedMNIST V2-a large-scale lightweight benchmark for 2D and 3D biomedical image classification. Scientific Data **10**(1) (2023)
36. Zeng, J., Lesnikowski, A., Alvarez, J.M.: The relevance of Bayesian layer positioning to model uncertainty in deep Bayesian active learning. In: Third workshop on Bayesian Deep Learning in NeurIPS (2018)