

# Calibrated Bayesian neural networks to estimate gestational age and its uncertainty on fetal brain ultrasound images

Lok Hin Lee<sup>1</sup>, Elizabeth Bradburn<sup>2</sup>, Aris T. Papageorgiou<sup>2</sup>, J. Alison Noble<sup>1</sup>

<sup>1</sup> Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, UK

lokhin.lee@eng.ox.ac.uk

<sup>2</sup> Nuffield Department of Women's & Reproductive Health, University of Oxford, UK

**Abstract.** We present an original automated framework for estimating gestational age (GA) from fetal ultrasound head biometry plane images. A novelty of our approach is the use of a Bayesian Neural Network (BNN), which quantifies uncertainty of the estimated GA. Knowledge of estimated uncertainty is useful in clinical decision-making, and is especially important in ultrasound image analysis where image appearance and quality can naturally vary a lot. A further novelty of our approach is that the neural network is not provided with images pixel size, thus making it rely on anatomical appearance characteristics and not size. We train the network using 9,299 scans from the INTERGROWTH-21st [22] dataset ranging from 14+0 weeks to 42+6 weeks GA. We achieve average RMSE and MAE of 9.6 and 12.5 days respectively over the GA range. We explore the robustness of the BNN architecture to invalid input images by testing with (i) a different dataset derived from routine anomaly scanning and (ii) scans of a different fetal anatomy.

**Keywords:** Fetal Ultrasound · Bayesian Neural Networks · Uncertainty in Regression · Gestational Age Estimation

## 1 Introduction

Knowledge of gestational age (GA) is important in order to schedule antenatal care, and to define outcomes, for example on whether a newborn is term or preterm [24]. Charts or formulas are used to estimate GA from ultrasound-based fetal measurements [21]. Fetal measurements are taken in standard imaging planes [24], which in the case of the head biometry plane, rely on the visibility of specific fetal brain structures. Classic GA estimation in a clinical setting relies on manual ellipse fitting by a trained sonographer, which has both significant intra-observer variability and skilled labour requirements [28].

In this paper we set out to assess the feasibility of estimating GA directly from image appearance rather than geometric fitting of ellipses or lines which is known to be difficult to automate across gestation. GA estimation based directly on image appearance would reduce the need for human interaction of automated

biometry after image capture. Further, significant progress has been made in automatic plane finding [2, 12], meaning that our algorithm might, in the future, be incorporated in a fully automated ultrasound-based GA estimation solution for minimally trained healthcare professional in a global health setting [23].

There have been previous attempts to automate GA estimation. These include GA estimation methods based on automating ellipse-fitting on fetal head images. Ciurte et al. [3] formalize the segmentation as a continuous min cut problem on ultrasound images represented as a graph. However, the method is semi-supervised and requires sonographer intervention. [10] directly estimate HC, but rely on predetermined ultrasound sweep protocols. [26] use a multi-task convolutional neural network to fit an ellipse on a fetal head image. However, all the methods above rely on clinical fetal growth charts for conversion of fetal measurements to GA, and therefore are subject to the same uncertainties.

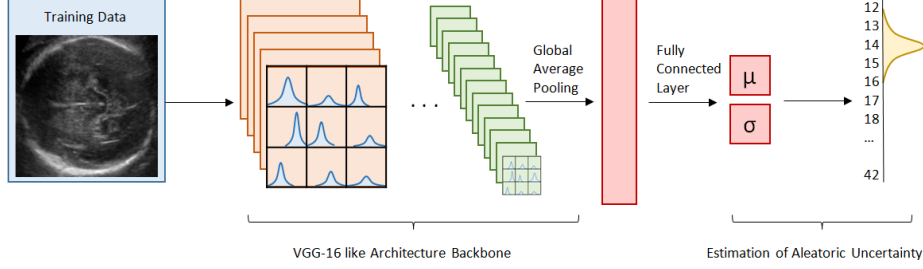
Namburete et al. [18] and [19] use regression forests with bespoke features and convolutional neural networks to directly regress GA from 3D ultrasound data without regressing from fetal measurements. However, such models only provide point estimates of GA without taking into account uncertainties caused by anatomical or image variations, and rely on costly 3D ultrasound volumes and probes.

We therefore use a Bayesian Neural Network (BNN) [16] to perform regression directly on two dimensional fetal trans-thalamic (TT) plane images in both the second and third trimesters. A Bayesian Neural Network learns the distribution across each individual trainable parameter. This allows the BNN to predict probability distributions instead of point estimates, which allows the calculation of uncertainty estimates that can further inform clinical decision making. BNNs have been used in medical disease classification [6, 15] and brain segmentation [25, 17].

**Contributions** We use a Bayesian Neural Network (BNN) trained on fetal trans-thalamic (TT) plane images for GA estimation. Firstly, we directly estimate GA from images of the fetal TT plane. Secondly, we train the BNN and an auxiliary isotonic regression model [13] to predict calibrated aleatoric and epistemic uncertainties (definitions of these terms are given in Sect. 1.1). Thirdly, we show that the Bayesian treatment of inference allows for automatic detection against unexpected and out-of-distribution data. Specifically, we test the trained BNNs on datasets outside of the training distribution; (i) a test dataset of TT planes taken with a different ultrasound machine and (ii) a test dataset of fetal US images other than the TT plane on the original ultrasound machine.

### 1.1 Neural Network Architecture

Due to computer memory limitations, we investigate the use of VGG-16 [9] as the backbone of a Bayesian Neural Network in a regression context as a proof of concept (See Fig 1.). Other backbones e.g. ResNets [] could also be used.



**Fig. 1.** The overview of the Bayesian Neural Network. The VGG-like backbone consists of five convolutional blocks and a global average pooling layer at the top to act as a regression layer. Each convolutional block  $CB$  consists of two convolutional layers, a batch normalization layer and a max pooling layer. Each  $CB^{f,sz,st}$  was parameterized by the shared filter size, kernel size and stride respectively. The VGG-16 like backbone, in this notation, is  $CB^{64,3,2} - CB^{128,3,2} - CB^{256,3,2} - CB^{512,3,2} - CB^{512,3,2}$

**Epistemic and Aleatoric Uncertainties** We can break down uncertainty estimates into uncertainty over the network weights (epistemic uncertainty) and irreducible uncertainties over the noise inherent in the data (aleatoric uncertainty) [11]. As training data size increases, in theory epistemic uncertainty converges to zero. This allows us to discern between uncertainties where more image data would be useful and uncertainties that are inherent in the model population. Research has also found that predicted uncertainty estimates from BNNs may be mis-calibrated to actual empirical uncertainty [8, 14]. A well calibrated regressive model implies that a prediction  $y$  with a predicted  $p$  confidence interval will lie within the confidence interval  $p$  of the time for all  $p \in 0, 1$ . We therefore additionally perform calibration [13] of uncertainty estimates.

**Probabilistic Regression** We use a probabilistic network in order to estimate the aleatoric and epistemic uncertainties in image-based prediction.

To estimate epistemic uncertainty, we use mean-field variational inference [7] to approximate learnable network parameters with a fully factorized Gaussian posterior  $q(\mathbf{W}) \sim N(\mu, \sigma)$ , where  $\mathbf{W}$  represents learnable weights,  $\mu$  represents the parameter mean and  $\sigma$  represents the standard deviation of the parameter.

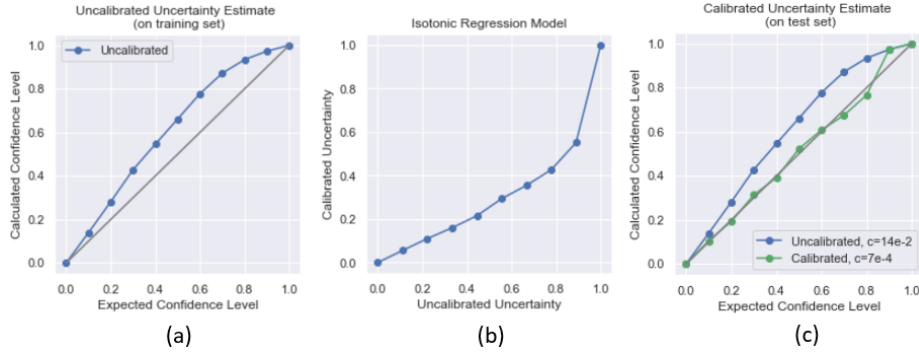
Gaussian posteriors and priors both reduce the computational cost of estimating the evidence lower bound (ELBO) [27], which is required as the prior  $p(\mathbf{W})$  is computationally intractable. The use of mean-field variational inference doubles the network size for each backbone network, as each weight is parameterized by the learned means and standard deviation weights.

To estimate aleatoric uncertainty, a network is trained to minimize the negative log likelihood between the ground truth GA and a predicted factorized Gaussian, which is used as an approximation of the aleatoric uncertainty. During the forward pass, samples are drawn from the parameterized network weight distributions. The variation in the mean predicted GA from model weight per-

turbations can then be used as an estimate of epistemic uncertainty. The total predictive uncertainty is therefore the sum of epistemic and aleatoric uncertainties:

$$Var(y) \approx \underbrace{E(\hat{y}^2) - E(\hat{y})^2}_{\text{Epistemic Uncertainty}} + \underbrace{E(\hat{\sigma}^2)}_{\text{Aleatoric Uncertainty}} \quad (1)$$

where  $\hat{y}$  represents the predicted mean,  $\hat{\sigma}^2$  the predicted variance and expectations are sampled using Monte Carlo inference at test time.



**Fig. 2.** An overview of the uncertainty calibration results, here on a trained VGG-16 network with cyclical KL annealing. (a) is the uncalibrated uncertainty interval plot on the training set, (b) is the predicted auxiliary regression model, and (c) shows the final calibrated uncertainty on the test set with reduced calibration error.

**Uncertainty Calibration** Intuitively, for every predicted confidence interval, the predicted GA should be within the interval with the predicted probability  $p_j$ . However, the use of variational inference and approximate methods means that predicted uncertainties may not accurately reflect actual empirical uncertainties [13]. We therefore calibrate uncertainty predictions using an auxiliary isotonic regression model trained on the training set (Fig 2). We quantify the quality of the uncertainty calibration using the calibration error:

$$Err = \frac{1}{m} \sum_m (p_j - \hat{p}_j)^2 \quad (2)$$

where  $p_j$  and  $\hat{p}_j$  are the predicted confidence level and actual confidence levels respectively, and  $m$  is the number of confidence levels picked. In this paper, we pick  $m=10$  intervals with  $p_j$  confidence intervals equally distributed between  $[0, 1]$ . Confidence intervals were calculated based on a two-sided confidence interval test of the predicted Normal distribution.

**Implementation Details** Both networks were trained from scratch with learnable parameters initialized from a  $N(0, 1)$  distribution. The network architecture is outlined in Figure 1. The loss function therefore takes the term:

$$\mathcal{L}_{total} = \frac{1}{N} \sum_n -\frac{1}{2} \left( \frac{y - \mu_i}{\sigma_i} \right)^2 + \lambda(epoch) \mathcal{L}_{KL} \quad (3)$$

where  $y$  is the ground truth GA,  $\mu_i$  and  $\sigma_i$  indicates the predicted mean GA and scale of the aleatoric uncertainty,  $\mathcal{L}_{KL}$  is the Kullback-Leibler loss of the weight parameters.

Networks empirically do not successfully converge with a  $\lambda(epoch) = k$  if  $k$  is constant. We therefore explore two annealing schedules such that  $\lambda(epoch)$  is (i) either a monotonic function that starts at zero at the first epoch and scales linearly to one by the hundredth epoch [1] or a (ii) cyclical annealing schedule with  $\lambda(epoch)$  cyclically annealing between 0 and 1 every 100 epochs [5].

Networks were trained with the Adam optimizer under default parameters with a batch size of 50, and trained for a minimum of 300 epochs until validation loss stopped decreasing. All networks were implemented using Tensorflow Probability [4], and trained on a Intel Xeon E5-2630 CPU with a NVIDIA GTX 1080 Ti GPU.

## 2 Dataset

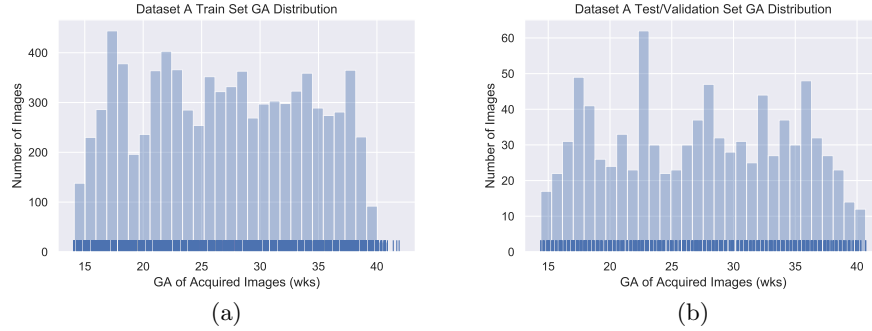
We use three clinical ultrasound datasets, designated Dataset A, B and C. Dataset A was used to train and validate the network, whilst Dataset B and C were only used during external testing.

**Dataset A** In Dataset A, derived from the INTERGROWTH-21st study [22], ultrasound scans of enrolled women were performed every  $5 \pm 1$  weeks of gestation, leading to ultrasound image scans from GAs ranging from 14+0 weeks to 41+6 weeks. We used the TT plane image to regress to find the GA, as the plane shows features that vary in anatomical appearance as the fetus grows [20].

Ground truth GAs for Dataset A are defined as the date from the last menstrual period (LMP) and further validated with a crown-rump measurement in the first trimester that agreed with the LMP to within 7 days. Images were captured using the Philips HD-9 (Philips Ultrasound, Bothell, WA, USA) with curvilinear abdominal transducers (C52, C63, V73).

Using a 90/1/9 train/validation/test split on the subjects, we generated a training set (8369/3016 images/subjects), validation set (91/35) and test set (849/300) (Fig 3).

**Dataset B** This dataset consists of routine second and third trimester scans. Ultrasound scans had been acquired using a Voluson E8 version BT18 (General Electric Healthcare, Zipf, Austria) which had different imaging characteristics



**Fig. 3.** An overview of the GA distribution in Dataset A for (a) train and (b) validation and test sets.

and processing algorithms to the machine used for Dataset A. We extracted  $N=20$  TT planes from clinical video recordings from different subjects. GAs inferred from growth charts ranged from 18+6 weeks to 22+1 weeks.

**Dataset C** This dataset consists of 6,739 images taken with the same image acquisition parameters and GA ranges as dataset A. However, the images were of the fetal femur plane instead of the TT plane.

**Dataset Pre-Processing and Augmentation** All images had sonographer measurement markings removed using automated template matching and a median filter, and were resized to 224x224 pixels. Images were then further visually checked to be satisfactory TT plane images by an independent sonographer. Pixel intensities were normalized to  $[-1, 1]$ . During training, we augment the dataset by randomly flipping the image horizontally, random resizing of the image by  $\pm 10\%$  and by shifting the image by up to 20 pixels vertically and horizontally. We find this reasonable, as variation in TT plane image appearance in each dataset for the same GA exceed this range. We also normalize GA to  $[-1, 1]$ . However, we do not give pixel size information to the network to reduce the impact of biometric measurement information.

### 3 Results

We present the results of both BNNs on the datasets separately.

**Dataset A** The results for networks with both annealing functions are summarized in Table 1. The network that trained with cyclical Kullback-Leibler annealing outperformed monotonic annealing. This may be due to the fact that

**Table 1.** Regression metrics for the investigated BNNs backbone architectures. Best results for each metric for 2D images are in bold. RMSE and MAE Metrics are calculated with the maximum likelihood for both aleatoric and epistemic uncertainties. For completeness, we include the results of [19] as a comparison. However, our results are based on regression on a single 2D ultrasound image, compared to a 3D ultrasound volume in their work.

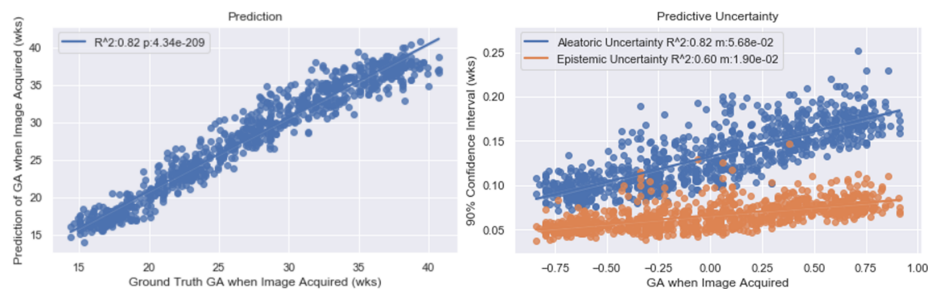
	$\mu$ RMSE (days)	$\mu$ MAE (days)	Calibration Error	No. of Parameters
VGG-16 with KL Cycling	<b>9.6</b>	<b>12.5</b>	0.24/ <b>7.4e-4</b>	18.8M
VGG-16 with KL Annealing	12.2	15.4	<b>0.21</b> /1.2e-3	18.8M
VGG-16 (Deterministic)	11.9	14.5	N/A	9.4M
3D Convolutional Regression Network [19]		7.72	N/A	6.1M

setting  $\lambda(epoch)$  to zero dramatically changes the hyper surface of the loss function, whilst the monotonic annealing creates a smoother change of the hyper surface which the network can get comfortable with especially in local minima.

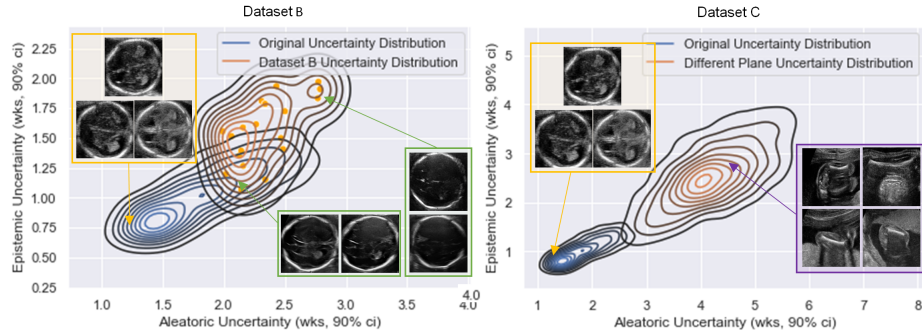
During inference, we estimate epistemic uncertainty with Monte Carlo iterations of  $n=100$ . We find that the network tends to overestimate predictive uncertainties, leading to high uncertainty calibration error (Fig. 2). However, the errors are consistent and therefore lend themselves to easy calibration, significantly reducing uncertainty error once calibrated.

Plotting predictive uncertainty against GA shows that uncertainty increases with increasing gestational age (Fig. 4). This may be because biological variation increases with increasing fetal growth. This is observed in clinical fetal growth charts. This is also supported by the higher slope that aleatoric uncertainty has with increasing GA compared to epistemic uncertainty.

**Evaluation on Dataset B** We evaluate the performance of the model on Dataset B, where the image is taken on a different ultrasound machine and in a different context, but the image is of the TT plane. We find that the distribution of uncertainties is higher than the original test dataset (Fig. 5).



**Fig. 4.** An overview of the results of the model trained with Kullback-Leibler cyclical annealing. On the left, we show a scatterplot of the overall predicted GA against GT GA with  $r^2$  of 0.8. On the right, we plot uncertainties against GT GA, and find that aleatoric uncertainty greater than epistemic uncertainty as a function of GA, where  $m$  refers to the gradient of the best fit line.



**Fig. 5.** An overview of kernel density estimates of the aleatoric and epistemic confidence intervals (ci) for out of distribution datasets. Images enclosed in yellow are from the dataset A test set, green from the external Dataset B imaging the same anatomy, and purple from the dataset A(FL). Figure is best viewed digitally.

We empirically find that the images from the dataset B with higher image contrast and a more visible CSP or midline have reduced predictive uncertainties compared with images that do not. This is in line with the observation that the images with i) clearly visible brain anatomical structures and ii) low GAs tend to have lower predictive uncertainties in the test dataset A. The MSE and RMSE of GA of images in Dataset B are 19.6 and 18.1 days respectively. However, this is expected, as the “ground truth” gestational ages for Dataset B were taken from fetal growth charts using HC measurements, compared to the gold standard LMP ground truth available in the Dataset A. A fetus with a HC-measurement based GA of 24 weeks has a clinical 90% confidence interval of 1.4 weeks [21].

**Evaluation on Dataset C** Dataset C is a dataset where image acquisition parameters are the same, but the images are of a different anatomy. We find that the uncertainty estimates provide a useful metric which can potentially inform the health professional that an invalid fetal plane is being used for regression. This may not be as obvious or possible with traditional GA estimation methods, which will predict a gestational age no matter the validity of the input image.

## 4 Conclusion

In this paper, we have described a Bayesian Neural Network framework with calibrated uncertainties to directly predict gestational age from a TT plane image across a wide range of GA. This was done without knowledge of pixel size. The best performing network achieved a RMSE of 9.6 days across the entire gestational age range. We demonstrated that the biased predictive uncertainties from variational inference can be calibrated, and are useful to detect images that are not within the network’s training. This is potentially a useful feature to prevent the model from being used out of the intended in a real world setting.



## References

1. Bowman, S.R., Vilnis, L., Vinyals, O., Dai, A.M., Jozefowicz, R., Bengio, S.: Generating sentences from a continuous space. Tech. rep. (2016)
2. Cai, Y., Sharma, H., Chatelain, P., Noble, J.A.: Multi-task SonoEyeNet: Detection of Fetal Standardized Planes Assisted by Generated Sonographer Attention Maps. *MICCAI* **11070** (Sep 2018)
3. Ciurte, A., Bresson, X., Cuisenaire, O., Houhou, N., Nedevschi, S., Thiran, J.P., Cuadra, M.B.: Semi-Supervised Segmentation of Ultrasound Images Based on Patch Representation and Continuous Min Cut. *PLoS ONE* **9**(7), e100972 (Jul 2014)
4. Dillon, J.V., Langmore, I., Tran, D., Brevdo, E., Vasudevan, S., Moore, D., Patton, B., Alemi, A., Hoffman, M., Saurous, R.A.: TensorFlow Distributions (nov 2017)
5. Fu, H., Li, C., Liu, X., Gao, J., Celikyilmaz, A., Carin, L.: Cyclical Annealing Schedule: A Simple Approach to Mitigating KL Vanishing. In: *NAACL*. pp. 240–250. Association for Computational Linguistics, Minneapolis, Minnesota (Jun 2019)
6. Gill, R.S., Caldairou, B., Bernasconi, N., Bernasconi, A.: Uncertainty-informed detection of epileptogenic brain malformations using bayesian neural networks. In: *LNCS*. vol. 11767 LNCS, pp. 225–233. Springer (oct 2019)
7. Graves, A.: Practical variational inference for neural networks. In: *NIPS* (2011)
8. Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q.: On calibration of modern neural networks. In: *ICML*. vol. 3, pp. 2130–2143 (2017)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *CVPR*. vol. 2016-Decem, pp. 770–778. IEEE Computer Society (dec 2016)
10. Heuvel, T.L.A.v.d., Petros, H., Santini, S., Korte, C.L.d., Ginneken, B.v.: Automated Fetal Head Detection and Circumference Estimation from Free-Hand Ultrasound Sweeps Using Deep Learning in Resource-Limited Countries. *Ultrasound in Medicine and Biology* **45**(3), 773–785 (Mar 2019)
11. Kendall, A., Gal, Y.: What uncertainties do we need in Bayesian deep learning for computer vision? In: *NIPS*. pp. 5580–5590. NIPS’17, Curran Associates Inc., Long Beach, California, USA (Dec 2017)
12. Kong, P., Ni, D., Chen, S., Li, S., Wang, T., Lei, B.: Automatic and Efficient Standard Plane Recognition in Fetal Ultrasound Images via Multi-scale Dense Networks. In: Melbourne, A., Licandro, R., DiFranco, M., Rota, P., Gau, M., Kampel, M., Aughwane, R., Moeskops, P., Schwartz, E., Robinson, E., Makropoulos, A. (eds.) *Data Driven Treatment Response Assessment and Preterm, Perinatal, and Paediatric Image Analysis*. pp. 160–168. Lecture Notes in Computer Science, Springer International Publishing, Cham (2018)
13. Kuleshov, V., Fenner, N., Ermon, S.: Accurate uncertainties for deep learning using calibrated regression. In: *ICML*. vol. 6, pp. 4369–4377 (2018)
14. Lakshminarayanan, B., Pritzel, A., Blundell, C.: Simple and scalable predictive uncertainty estimation using deep ensembles. In: *NIPS*. vol. 2017-Decem, pp. 6403–6414 (2017)
15. Leibig, C., Allken, V., Ayhan, M.S., Berens, P., Wahl, S.: Leveraging uncertainty information from deep neural networks for disease detection. *Sci. Rep.* **7**(1), 1–14 (dec 2017)
16. MacKay, D.J.C.: A Practical Bayesian Framework for Backpropagation Networks. *Neural Comput.* **4**(3), 448–472 (may 1992)

17. McClure, P., Rho, N., Lee, J.A., Kaczmarzyk, J.R., Zheng, C.Y., Ghosh, S.S., Nielson, D.M., Thomas, A.G., Bandettini, P., Pereira, F.: Knowing What You Know in Brain Segmentation Using Bayesian Deep Neural Networks. *Front. Neuroinform.* **13**, 67 (oct 2019)
18. Namburete, A.I., Stebbing, R.V., Kemp, B., Yaqub, M., Papageorgiou, A.T., Alison Noble, J.: Learning-based prediction of gestational age from ultrasound images of the fetal brain. *Med. Image Anal.* **21**(1), 72–86 (apr 2015)
19. Namburete, A.I., Xie, W., Noble, J.A.: Robust regression of brain maturation from 3D fetal neurosonography using CRNs. In: LNCS. vol. 10554 LNCS, pp. 73–80. Springer Verlag (2017)
20. Paladini, D., Malinger, G., Monteagudo, A., Pilu, G., Timor-Tritsch, I., Toi, A.: Sonographic examination of the fetal central nervous system: Guidelines for performing the 'basic examination' and the 'fetal neurosonogram' (jan 2007)
21. Papageorgiou, A.T., Kemp, B., et al.: Ultrasound-based gestational-age estimation in late pregnancy. *Ultrasound Obstet. Gynecol.* **48**(6), 719–726 (dec 2016). <https://doi.org/10.1002/uog.15894>
22. Papageorgiou, A.T., Kennedy, S.H., Salomon, L.J., Altman, D.G., Ohuma, E.O., Stones, W., Gravett, M.G., Barros, F.C., Victora, C., Purwar, M., Jaffer, Y., Noble, J.A., Bertino, E., Pang, R., Cheikh Ismail, L., Lambert, A., Bhutta, Z.A., Villar, J.: The INTERGROWTH-21 st fetal growth standards: toward the global integration of pregnancy and pediatric care. *American Journal of Obstetrics and Gynecology* **218**(2), S630–S640 (Feb 2018)
23. Papageorgiou, A.T., Ohuma, E.O., Altman, D.G., Todros, T., Ismail, L.C., Lambert, A., Jaffer, Y.A., Bertino, E., Gravett, M.G., Purwar, M., Noble, J.A., Pang, R., Victora, C.G., Barros, F.C., Carvalho, M., Salomon, L.J., Bhutta, Z.A., Kennedy, S.H., Villar, J.: International standards for fetal growth based on serial ultrasound measurements: The Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project. *Lancet* (9946), 869–879 (sep). [https://doi.org/10.1016/S0140-6736\(14\)61490-2](https://doi.org/10.1016/S0140-6736(14)61490-2)
24. Salomon, L.J., Alfrevic, Z., Da Silva Costa, F., Deter, R.L., Figueras, F., Ghi, T., Glanc, P., Khalil, A., Lee, W., Napolitano, R., Papageorgiou, A., Sotiradis, A., Stirnemann, J., Toi, A., Yeo, G.: ISUOG Practice Guidelines: ultrasound assessment of fetal biometry and growth. *Ultrasound Obstet. Gynecol.* **53**(6), 715–723 (2019)
25. Sedai, S., Antony, B., Rai, R., Jones, K., Ishikawa, H., Schuman, J., Gadi, W., Garnavi, R.: Uncertainty guided semi-supervised segmentation of retinal layers in OCT images. In: LNCS. vol. 11764 LNCS, pp. 282–290. Springer (oct 2019)
26. Sobhaninia, Z., Raffei, S., Emami, A., Karimi, N., Najarian, K., Samavi, S., Reza Soroushmehr, S.M.: Fetal Ultrasound Image Segmentation for Measuring Biometric Parameters Using Multi-Task Deep Learning. In: EMBC. pp. 6545–6548. IEEE, Berlin, Germany (Jul 2019)
27. Wen, Y., Vicol, P., Ba, J., Tran, D., Grosse, R.: Flipout: Efficient pseudo-independent weight perturbations on mini-batches. In: ICLR (2018)
28. Zador, I.E., Salari, V., Chik, L., Sokol, R.J.: Ultrasound measurement of the fetal head: computer versus operator. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology* **1**(3), 208–211 (May 1991)