# Effect of Liver Segmentation Confidence Score on Multi-Modal Diabetes Prediction *

Anonymous*, Anonymous*,*, Anonymous*, and Anonymous*

[1] Anonymous
[2] Anonymous ***@**.**
Anonymous
[3] Anonymous
***@**.**

**Abstract.** In the era of deep learning, the prognosis of diabetes using deep learning is not a new thing. But due to insufficient annotated ground truths, training a model is not feasible many a time. In general, fine-tuning and uncertainty/confidence measure is a widely accepted approach to alleviate the low labeled sample issue with the power to quantify the model confidence on a new dataset. In this study, we inferences liver segmentation from uncertainty based models with their respective uncertainty/confidence score. We then use a statistical model to inference diabetes from segmentation volume and confidence score. We test it on another data-set with less labeled samples to see, if the addition of confidence score is of any significant change for predicting diabetes.

**Keywords:** Segmentation · Confidence Score · Diabetes.

## 1 Introduction

- Segmentation required for the extraction of imaging biomarkers from scans - While we are working on improving the performance, we will never have a method that will perfectly segment the organ of interest in all imaging situations - Hence, there is an interest in not only predicting the segmentation but also the uncertainty of the segmentation

- Bayesian approaches enable to estimate the posterior distribution - Several methods for Bayesian neural networks have been proposed - And many applications of these techniques on medical images - However, the integration of the segmentation uncertainty into follow-up analysis of extracted biomarkers has not yet been well studied

In this work, we study the integration of segmentation uncertainty in the statistical analysis of image-derived biomarkers. The two main questions that we investigate are: How do we integrate segmentation uncertainty in statistical models for group analysis? Which method for uncertainty quantification is best suited for this purpose? For the integration, we evaluate .. For the uncertainty quantification, we compare .. We perform experiments for the segmentation of liver in abdominal MRI scans in subjects with diabetes mellitus.

*

## 2 Material and Methods

### 2.1 Data

We selected two distinct dataset with diabetes analysis data for our experiments. The first dataset is Cooperative Health Research in the Augsburg Region(KORA) with 305 subjects, each with whole body magnetic resonance images and label map for liver. The volumetric scans are 260x320x160 anisotropic ($1.5mm^3$) voxels. We split the dataset into 155 training and 150 testing subjects by equally distributing diabetic and non-diabetic volumes. The second dataset is UK-BioBank, we had 20,000 subjects with 224x174x44 anisotropic ($2.2x2.2x4.5)mm^3$ voxels with manually segmentated liver organ label map for 10 volumes. Both the datasets are different in their orientation and resolution. For a better performance, we reorient and change resolution of both the dataset to a standard format with normalising pixel intensities to the range [0, 1]. We use UK-Biobank solely for cross-dataset testing purpose.

### 2.2 Bayesian Neural Networks

We use four different varieties of bayesian neural network to predict segmentation and model confidence score. For the convenience, we use quicknat[6] as base architecture and develop below variations on top of it.

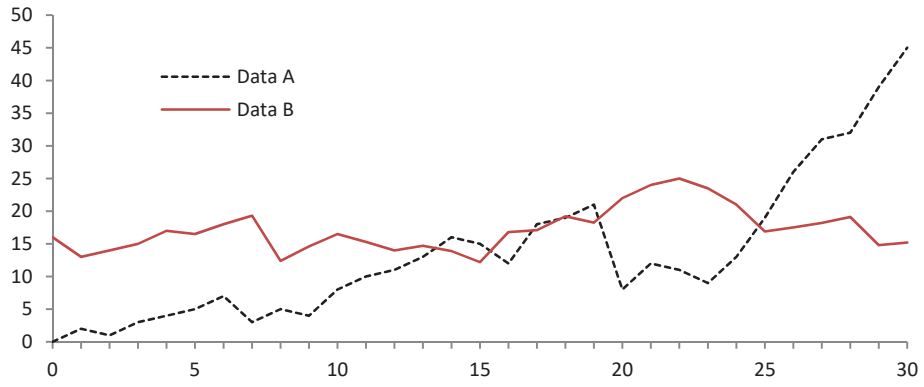


**Fig. 1.** A figure with all four architectures [TODO].

*I. Monte Carlo Dropout:* In MC Dropout quicknat architecture [5], Roy et. al. has introduced test time dropout layer into quicknat[6] to add stochasticity during inference. He added dropout layer after each encoder and decoder block. Generally dropout has been used in training time to avoid over-fitting, but test time dropout enables the network to rewire different neurons to inference different outputs.

*II. probabilistic U-Net:* Semantic segmentation is all about assigning a particular class to the pixels. In medical domain, even if over-ally we get a better segmentation output, but there exist ambiguity in pixel level in many ambiguous images, which may lead to wrong diagnosis. Kohl et. al. with Probabilistic U-Net [4] architecture, suggested to learn the low-dimensional latent space of segmentation maps to understand possible segmentation variants. A random sample than added to the base architecture to inference multiple plausible segmentation output.

*III. Hierarchical U-Net:* Hierarchical U-Net[3] is an extension of Probabilistic U-Net[4], where Kohl. et al. suggested, instead of only looking at low dimensional ambiguity, the model learns multi-scale latent space of the segmentation maps while training and generates plausible segmentation output now from a broader understanding of segmentation space.

*IV. Bayesian F-CNN with Re-parameterization:* In this method, We replace traditional conv layer with bayesian layer [fig 1] which has been developed using re-parameterization trick [1] 1. 2 conv layer has been used whose outputs are further processed to add non-linearity by adding a tanh activation layer each. We consider the outputs from tanh layer as "mu" and "sigma". A gaussian white noise of 0 mean and 0.1 std. has been added to "sigma" to add stochasticity and has been finally getting added to "mu".

$$g_\theta(\epsilon) = \mu_\theta + \epsilon\sigma_\theta \tag{1}$$

$$\epsilon \approx N(0, 0.1) \tag{2}$$

Instead of 0 mean, 1 std. advised by Kingma et. al. in [1], we use a narrower std. to restrict activation's weight variations under certain boundary for better performance.

### 2.3 Confidence Measure

For network prediction confidence measure, we use intersection over union(IoU) [5] metric. IoU works over N MC samples of a specific structure s.

$$IoU_s = \frac{|(S_1 == s) \cap (S_2 == s) \cap ...(S_N == s)|}{|(S_1 == s) \cup (S_2 == s) \cup ...(S_N == s)|} \tag{3}$$

We use N=10 and a specific structure i.e Liver.

### 2.4 Statistical Methods

For statistical analysis of imaging bio-markers that we inferenced from 4 different bayesian neural networks with their confidence score, we figure out below different ways to incorporate imaging and non-imaging bio-markers. We focused on the task of classifying diabetes state and introduce confidence score to eradicate imperfection in inferenced segmentation volumes.

*I. Base Model:* We will use basic non-imaging bio-markers for diabetes analysis i.e. Age, Sex and Body Mass Index (BMI). We have a binary diabetes class to classify i.e. Diabetic/Non-Diabetic. Age and BMI has been z-score normalized with 0 mean and 1 std.

$$DiabetesState = \alpha_0 + \alpha_1 * Age + \alpha_2 * Sex + \alpha_3 * BMI + \epsilon \tag{4}$$

*II. Base Model with Segmentation Volumes:* Segmentation volumes has been z-score normalised with 0 mean and 1 std.

$$DiabetesState = RHS(eqn4) + \alpha_4 * Volumes + \epsilon \tag{5}$$

*RHS(eqn 4) represents right hand side of equation 4

*III. Base Model with Segmentation Volumes and Confidence Score:* We use both confidence and volumes as con-founder in this stage.

$$DiabetesState = RHS(eqn4) + \alpha_4 * Volumes + \alpha_5 * ConfidenceScore + \epsilon \tag{6}$$

*IV. Base Model with Segmentation Volumes regularized by Confidence Score:* In this stage, we are using confidence score as a regularizing factor only for volume.

$$DiabetesState = RHS(eqn4) + \alpha_4 * (ConfidenceScore * Volumes) + \epsilon \tag{7}$$

*V. Base Model with Segmentation Volumes and Confidence Score as Instance Weight:* In contrast to equation 7, here we are using confidence score as an instance weight which is regularising the whole equation consisting non-imaging bio-markers.

$$ConfidenceScore * (DiabetesState = RHS(eqn4) + \alpha_4 * Volumes + \epsilon) \tag{8}$$

## 3 Experimental Results

## 4 Conclusion

## References

1. Kingma, D.P., Salimans, T. and Welling, M., 2015. Variational dropout and the local reparameterization trick. In Advances in neural information processing systems (pp. 2575-2583).
2. Gal, Y. and Ghahramani, Z., 2016, June. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In international conference on machine learning (pp. 1050-1059).
3. Kohl, S.A., Romera-Paredes, B., Maier-Hein, K.H., Rezende, D.J., Eslami, S.M., Kohli, P., Zisserman, A. and Ronneberger, O., 2019. A Hierarchical Probabilistic U-Net for Modeling Multi-Scale Ambiguities. arXiv preprint arXiv:1905.13077.

4. Kohl, S., Romera-Paredes, B., Meyer, C., De Fauw, J., Ledsam, J.R., Maier-Hein, K., Eslami, S.A., Rezende, D.J. and Ronneberger, O., 2018. A probabilistic u-net for segmentation of ambiguous images. In Advances in Neural Information Processing Systems (pp. 6965-6975).
5. Roy, A.G., Conjeti, S., Navab, N., Wachinger, C. and Alzheimer's Disease Neuroimaging Initiative, 2019. Bayesian QuickNAT: model uncertainty in deep whole-brain segmentation for structure-wise quality control. NeuroImage, 195, pp.11-22.
6. Roy, A.G., Conjeti, S., Navab, N., Wachinger, C. and Alzheimer's Disease Neuroimaging Initiative, 2019. QuickNAT: A fully convolutional network for quick and accurate segmentation of neuroanatomy. NeuroImage, 186, pp.713-727.