Bayesian Learning and Uncertainty Quantification on Diabetes Data

Dwyer Deighan, Gurvinder Singh
University at Buffalo,NY

Abstract.
In recent years, many works in the Machine Learning literature have shown the benefit of Uncertainty Quantification for the Model, As it gave confidence over the prediction which is required for the life-critical and patient safety related decision making. In this paper we proposed the Bayesian Learning for the aleatoric and epistemic uncertainty quantification. We used simple Bayesian Neural Network, Bayesian Tab Transformer and Bayesian Network for Uncertainty Quantification and done a comparison together. We also compared with their frequentist counterpart and showed the bayesian models get the identical accuracy with additional benefit of producing confidence interval over the predictions.

Keywords. Bayesian Neural Network; Bayesian Structure Learning; Uncertainty Quantification; Tab Transformer; Bayesian Tab Transformer

Introduction. The diagnostic process in the medical field is particularly difficult to model. This is due to the uncertainty that characterizes the medical knowledge regarding the possible diseases and the data regarding the patient's condition; not to mention the different approaches to the diagnosis and the doctor's reasoning in the diagnostic process. Uncertainty Quantification is to Know when a classifier's prediction can be trusted or not and it is useful in many applications and critical for safely using AI. While the bulk of the effort in machine learning research has been towards improving classifier performance, understanding when a classifier's predictions should and should not be trusted has received far less attention Jiang et al. (2018). Sources of uncertainty arise when the test and training data are mismatched, while data uncertainty occurs because of class overlap or due to the presence of noise in the data. however, estimating knowledge uncertainty is significantly more difficult than estimating data uncertainty. the two main types of uncertainty, are aleatoric and epistemic uncertainties Hüllermeier & Waegeman (2021).The uncertainty in data that gives rise to uncertainty in predictions is called aleatoric uncertainty while epistemic uncertainty or Modelling uncertainty occurs due to inadequate knowledge.

Bayesian approximation and ensemble learning techniques are two widely-used types of uncertainty quantification methods. here, We propose to model reasoning under uncertainty, using Bayesian Neural Network and Bayesian Network for a diabetes dataset. So that results were more interpretable and it should be able to cooperate with human doctors in the diagnostic procedure.

Previous Work. There are some work in the past to do Bayesian learning on medical dataset. where Lixandru-Petre (2020) proposed a Bayesian Network for a Diabetes, where they used the Graphical Lasso Method to find the correlation between different features and then using the Bayesian Hill climbing method to find the causal direction between these features. whereas one of the first work of using the bayesian neural network on Diabetes dataset is done by Kahramanli & Allahverdi (2008) in which they designed a hybrid system for diabetes disease. The work of using larger models like transformer to do bayesian modelling is proposed in paper by Xue et al. (2021) where they used the bayesian transformer for language models in speech recognition. The use of Transformer on Tabular dataset is proposed in TabTransformer Huang et al. (2020) where

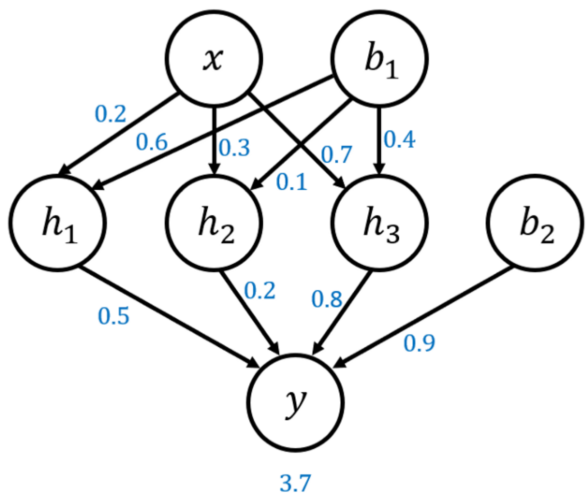

Figure 1. The Left network shows the Frequentist Neural Network with single weight values whereas the Right network shows the Bayesian Neural network with the weights as a Gaussian distribution.

they used both categorical and continuous variables separately and later combined together in the model.

Experiment. We done experiment in two different fronts and then performed a comparison between these two broad categories.

BAYESIAN NEURAL NETWORK. We first started with a Two layer neural network with cross entropy loss and Adam optimizer and trained the 330 total parameters to predict one out of three categories. Our next step is take the same two layer neural network and converted it into Bayesian Neural network where we took Normal Distribution as prior and posterior for the weights and bias of the model and use the Variation inference to learn the parameters for the distribution of weight and biases. Now the 330 parameters of the Frequentist network get doubled as these 660 parameters are now representing the distribution parameters not the single weight values.

Our next set of experiment is to use a complex model for the same task here we first used Tabtransformer model proposed by Huang et al. (2020) where Categorical Features first column embed and then passed to the Multihead attention modules where it concat with the Continuous features which were layer normalized into a vector. A multi-layer perceptron is then output the category by taking the input of combined features. In the Bayesian Tab transformer the multi-layer perceptron and Feed forward layers of attention modules has started learning the distribution parameters instead of single weights. Here again the number of parameters get doubled in the model which correspond to the longer training and inference time.

For the Uncertainty Quantification to figure out the effect of the out-of-distribution samples during inference time. we created a category holdout dataset where one category is put on hold and the model is trained on the rest of the categories and during the inference time will see the difference in the entropy while using in-distribution vs the out-of-distribution samples. This experiment is tried to simulate the situation in the real world live environment where the introduction of new category can break the model without giving any signals.

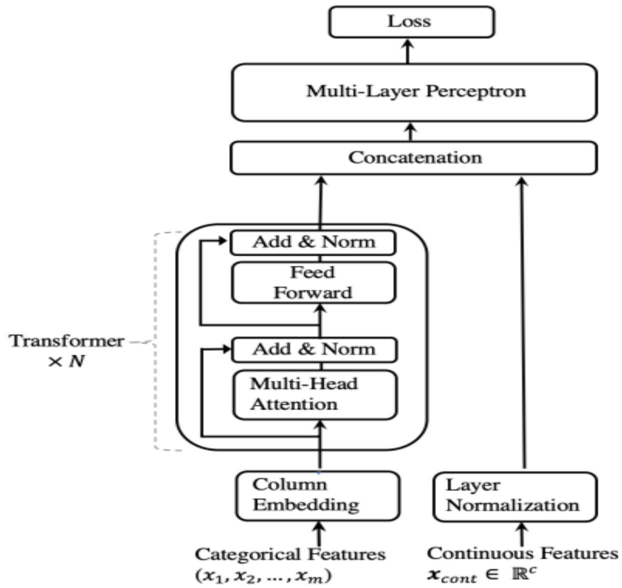

Figure 2. Tab transformer model architecture

BAYESIAN NETWORK. For the Bayesian Network experiment we first learned the Structure, then parameters then moved onto the inference over this model. For the Structure learning Hill-Climbing Search With Bayesian Dirichlet equivalent uniform (Bdeu) Score to find the max scored network out of different candidate networks which can represent the distribution P. The learned DAG shows the relation in a form of cause and effect between the variables. The Parameters for the variables are learned using the Maximumlikelihood Estimator, these represent as the conditional probability distribution tables. For the Inference part, the Monte-carlo sampling named metropolis-hasting is been used.

Dataset. For our experiment we used the Diabetes (Dataset) by Behavioral Risk Factor Surveillance System (BRFSS) which were prepared by telephone survey on health-related issues by the CDC. This original dataset contains responses from 253,680 individuals and has 21 feature variables with target variables has 3 classes. 0 is for no diabetes or only during pregnancy, 1 is for prediabetes, and 2 is for diabetes. Some of the Features are High Blood Pressure, High Cholestrol, Smoker, Stroke.

	Model	F1 score
1	Linear Baseline	83.2
2	Naive Bayes	81.9
3	XgBoost	84.4
4	2 Layer NN	84.2
5	Bayesian NN	84
6	Tab Transformer	85.1
7	Bayesian Tab Transformer	84.6

Table 1. Table showing F1 score on the Test data for Different Models

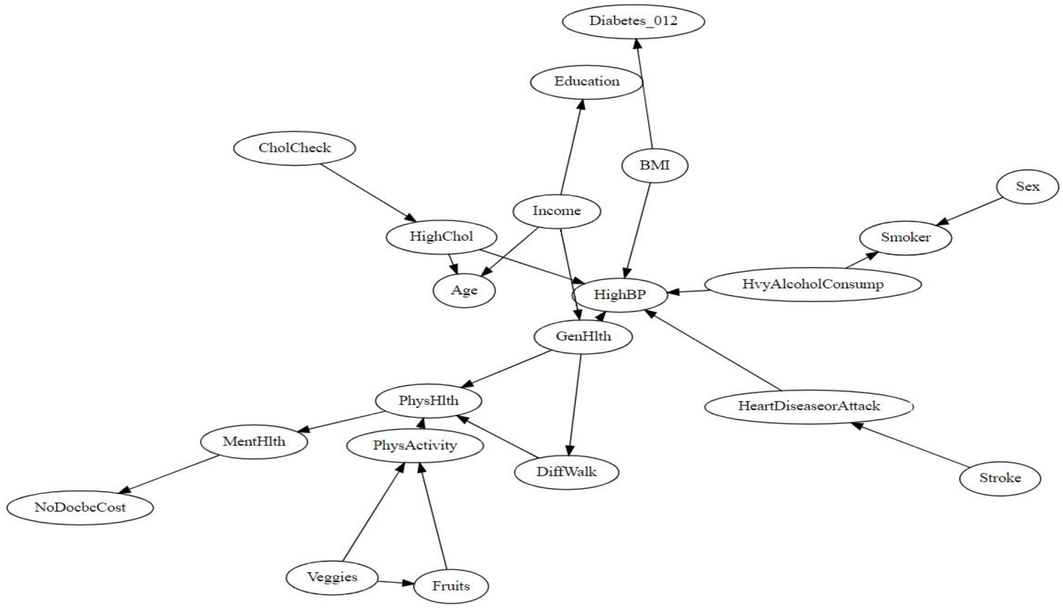

Figure 3. Learned Bayesian network structure from data

Results. In Figure 4. of the Results we showed the Uncertainty over the prediction of two samples taken randomly from the test dataset. The Non-Bayesian and Transformer model are Frequentist model so we just get the same prediction over multiple inferences But on the other side the Bayesian models gave a confidence interval over the predictions while doing inferences 50 times.

In the Figure 5. We used the experiment of hold-out category for calculating the difference between Entropy and Negative likelihood values of out-of-distibution and in-distribution data.

In figure 7 we use -log(likelihood) of models' predictions of diabetes class conditioned on either known (aka training) or unknown (aka holdout) data. In figure 6 we use surprisal (aka information content), as an alternative measure for uncertainty. To test if it is 'more uncertain' of its predictions for the holdout class dataset (where the class is: pre-diabetes). In both cases larger values indicate more uncertainty so we expected to see higher uncertainty for the holdout class data.

Strangely as can be seen in figures 6 and 7 it appears that the probabilistic graphical models weren't as effective at identifying unseen data via uncertainty quantification. This could be because for the neural network case the soft-max activation function makes the probabilities follow an exponential distribution while this isn't the case for the PGM (the Bayesian Markov Networks). The effects of this can seen in figures 4 5, since the uncertainty measures take a bimodal distribution of sorts.

Conclusion. We were Able to Train and compare Bayesian Neural network, Bayesian Network and Markov Network Uncertainty Using Hold out Class Approach. In Future work, we want to use the Causal learning from Bayesian Network models to be incorporated in Bayesian Neural network models. This effects of this can be seen in the **??**

We Also Want to thanks the Prof. Mingchen gao for there continuous Support during the project.

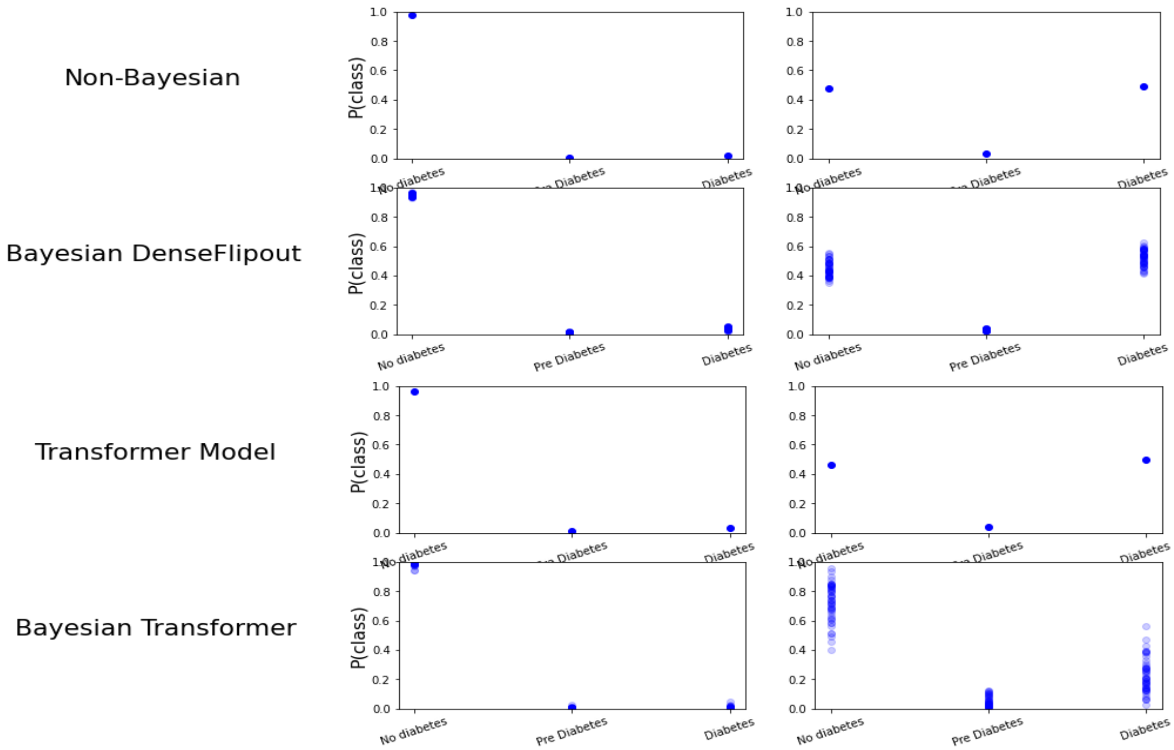

Figure 4. Compare the predictions for a known and unknown sample

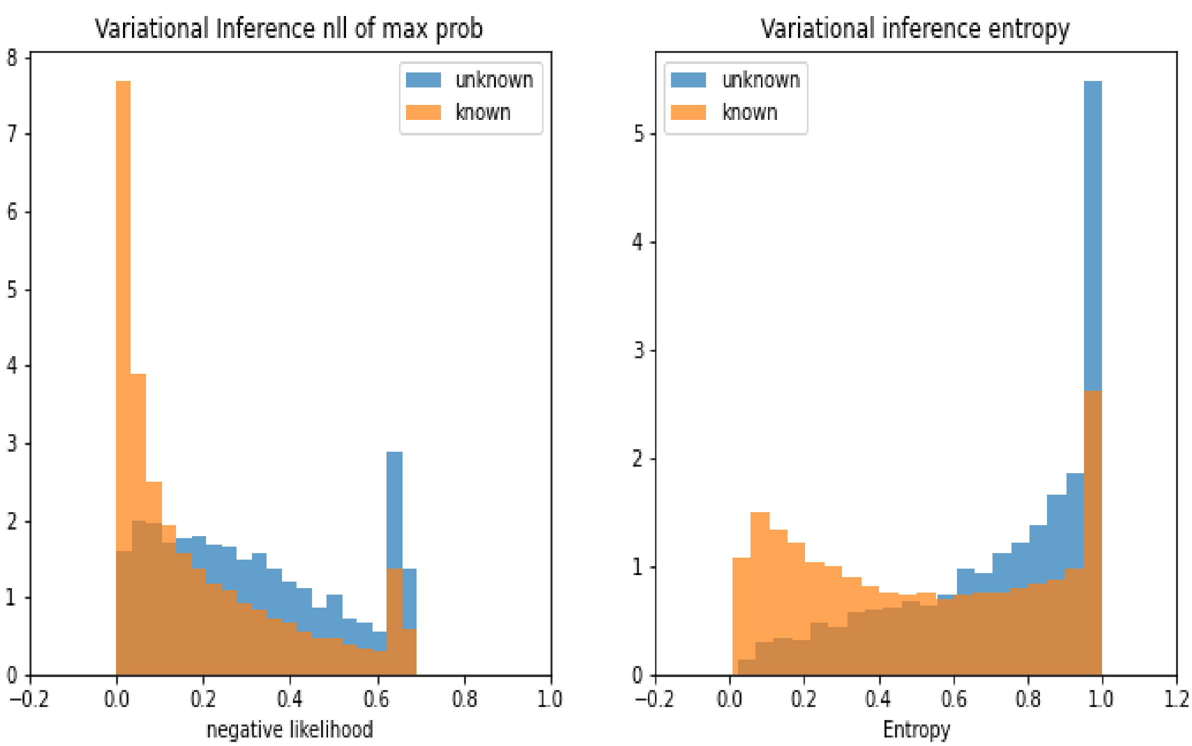

Figure 5. Compare the uncertainty measures for all known and unknown classes

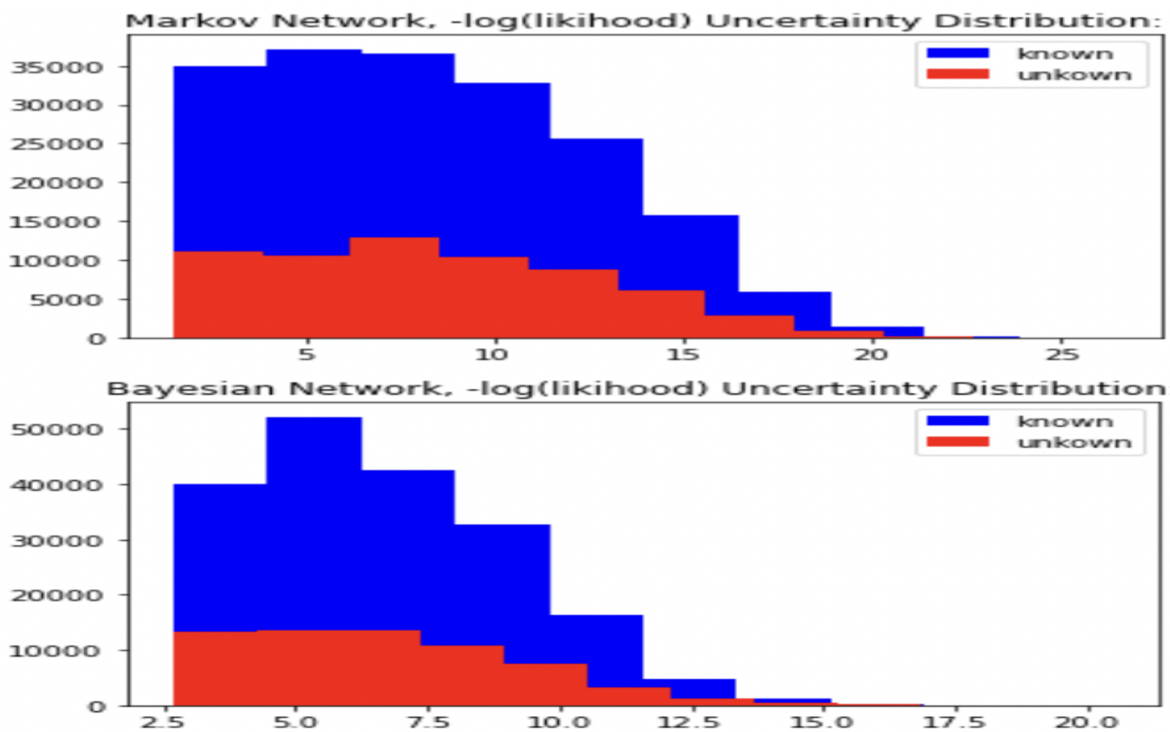

Figure 6. PGM surprisal model Uncertainty

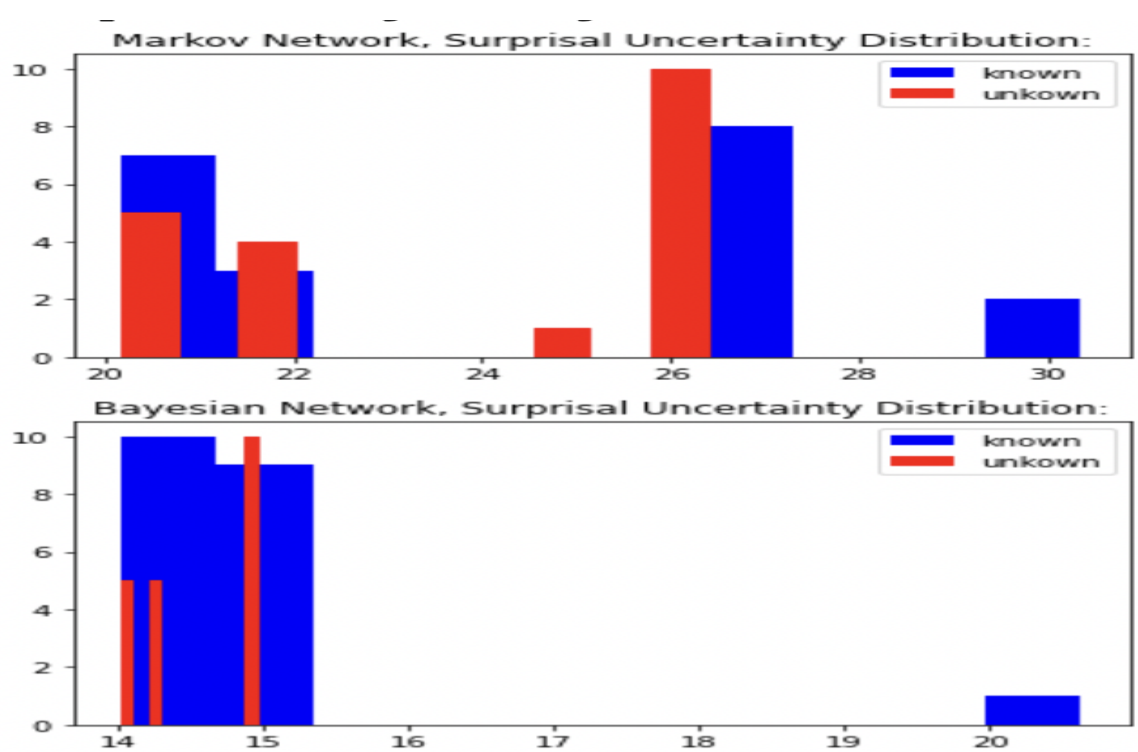

Figure 7. PGM -log(likelihood) Uncertainty

References

Huang, Xin, Ashish Khetan, Milan Cvitkovic & Zohar Karnin. 2020. Tabtransformer: Tabular data modeling using contextual embeddings. *arXiv preprint arXiv:2012.06678* .

Hüllermeier, Eyke & Willem Waegeman. 2021. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning* 110(3). 457–506.

Jiang, Heinrich, Been Kim, Melody Guan & Maya Gupta. 2018. To trust or not to trust a classifier. *Advances in neural information processing systems* 31.

Kahramanli, Humar & Novruz Allahverdi. 2008. Design of a hybrid system for the diabetes and heart diseases. *Expert systems with applications* 35(1-2). 82–89.

Lixandru-Petre, Irina-Oana. 2020. Modeling a bayesian network for a diabetes case study. In *2020 international conference on e-health and bioengineering (ehb)*, 1–4. IEEE.

Xue, Boyang, Jianwei Yu, Junhao Xu, Shansong Liu, Shoukang Hu, Zi Ye, Mengzhe Geng, Xunying Liu & Helen Meng. 2021. Bayesian transformer language models for speech recognition. In *Icassp 2021-2021 ieee international conference on acoustics, speech and signal processing (icassp)*, 7378–7382. IEEE.