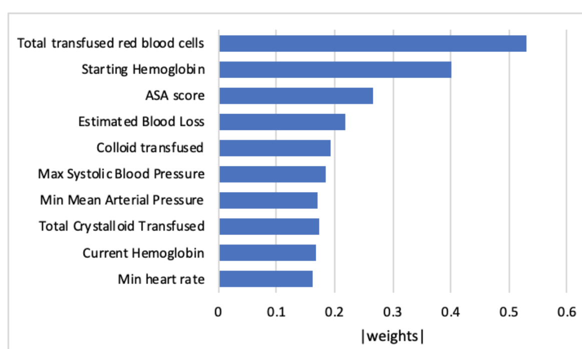
**Postoperative In-Hospital Mortality Prediction Using Bayesian Neural Networks for Interpretability**

**Presenting Author:** Muntaha Samad, MS
**Co-Authors:** Christine Lee, MS, Ira Hofer, MD, Pierre Baldi, PhD, Maxime Cannesson, MD, PhD

**Introduction:** Lee et al. recently showed that deep neural networks (DNNs) can predict postoperative in-hospital mortality with an AUC of 0.91.[1] The non-linearity of DNNs makes them powerful machine learning models, however they lack interpretability. DNNs effectively function as a black box, making it very difficult to determine how much each feature contributes to predictions and how confident the network is in each prediction. In healthcare applications, it is vital that models be interpretable in order to help clinicians make and clearly justify real-time decisions. To address this need, Overweg et al. showed that bayesian neural networks (BNNs) can be used to make predictions on clinical data (collected from the intensive care unit) with a high accuracy, while also being interpretable for clinicians.[2] In this study, we apply similar concepts and create a BNN to predict the  risk of postoperative in-hospital mortality. Using this BNN allows us to visualize the contribution of each feature to predictions and to evaluate the uncertainty of each prediction.

**Method:** The dataset used in this experiment came from the UCLA Medical Center with IRB approval and includes all surgical procedures performed since March 17, 2013, excluding cases without general anesthesia, and cases where patients were  >89 or < 18 years of age. The dataset includes 59,985 patients with 53 features such as: demographics, intraoperative vital signs, medication administration, anesthesia events, and type of surgery. Missing data was imputed using the mode for categorical features and using the mean for all other features. The data was then normalized and split using 80% for training and 20% for testing. This dataset is highly imbalanced, with the percent occurrence of mortality being less than 1% in the training set. To address this imbalance, we used the Synthetic Minority Over-Sampling Technique (SMOTE) to augment the mortality occurences to be approximately 50% in the training set.[3] The test set was not augmented.

**Results/Conclusion:** The 5 fold cross validation AUC for our BNN is 0.809 +/- 0.015. The BNN used in this experiment has 4 hidden layers and 100 neurons per layer. The first layer uses a sparsity inducing horseshoe prior and the subsequent layers use a gaussian prior. The network is designed so that weights connected to a single input feature all share the same prior which allows the network to suppress or promote all weights connected to a single feature and therefore suppress or promote the contribution of each feature to the prediction. By looking at the weights in the first layer connected to each input feature we are able to determine each feature's contribution to the prediction. The figure shows the top 10 contributing features with the highest impact on predictions. We can see that total transfused red blood cells has the highest contribution to outcome prediction.  An inherent property of BNNs is that all parameters and predictions are distributions rather than point estimates. By looking at the distribution of each prediction, we are also able to evaluate its uncertainty and provide clinicians with an even clearer understanding of predictions. The next steps for this experiment would be to further improve the AUC by fine tuning the model architecture and hyperparameters.



**References:**


1. Lee, C. K., Hofer, I., Gabel, E., Baldi, P., & Cannesson, M. (2018). Development and validation of a deep neural network model for prediction of postoperative in-hospital mortality. Anesthesiology: The Journal of the American Society of Anesthesiologists, 129(4), 649-662.
2. Anna-Lena Popkes, Hiske Overweg, Ari Ercole, Yingzhen Li, Jos´e Miguel Hern´andez-Lobato, Yordan Zaykov, and Cheng Zhang. Interpretable outcome prediction with sparse Bayesian neural networks in intensive care. arXiv preprint arXiv:1905.02599, 2019.
3. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” J. Artif. Intell. Res., vol. 16, pp. 321–357, 2002.