

## An Interpretable Neural Network for Prediction of Postoperative In-hospital Mortality

**Presenting Author:** Christine Lee, PhD

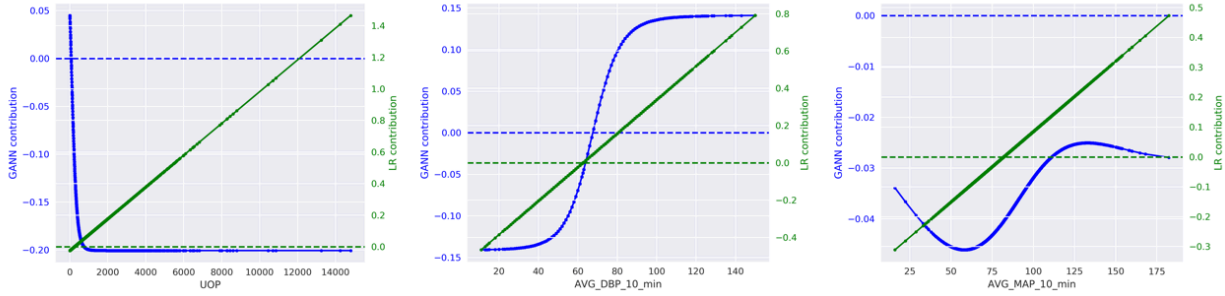
**Co-authors:** Muntaha Samad, Ira Hofer, Maxime Cannesson, Pierre Baldi

**Introduction:** We recently showed that deep neural networks (DNNs) can successfully predict postoperative in-hospital mortality with an AUC of 0.91.<sup>1</sup> While DNNs are great machine learning models and often have higher accuracy than more simple models like logistic regression, they are not intelligible. In healthcare, intelligible models not only help clinicians to understand the problem and create more targeted action plans, they also help to gain the clinicians' trust. Caruana et al. demonstrated generalized additive models with pairwise interactions can be applied to real healthcare problems such as pneumonia risk with high accuracy.<sup>2</sup> Through a graphical representation of each model feature's learned contribution to the predicted risk, the intelligible model helps us to visualize learned patterns and identify new patterns in the data or confirm what clinicians already know. In this study, we applied the same idea and created a "generalized additive neural network" (GANN) to help visualize feature patterns related to risk of in-hospital mortality.

**Methods:** Data used in these experiments came from UCLA Medical Center with IRB approval. The data consists of 59,985 patients with 46 EMR features extracted at the end of surgery as well as 33 HCUP codes (found in >1% of the data). These EMR features included ASA, intraoperative vital signs, interventions, and anesthesia events. Data included all surgical procedures performed since March 1, 2013. Cases not done with general anesthesia, and patients > 89 or < 18 years of age were excluded. This data was then randomly split into 80% for training (n= 47,988) and 20% for test (n=11,997). Missing values were filled with the means for that feature. Values that were greater than a clinically normal maximum were set to a maximum possible. Finally, all features were rescaled to have mean 0 and standard deviation 1. In the training data, the % occurrence of in-hospital mortality was 0.81% (n=389). All DNNs were trained on 80% of the data with 5-fold cross validation, and were all feedforward networks with sigmoid outputs trained using Adam optimization. L2 regularization and early stopping were used to prevent overfitting. A logistic regression model with the same features was also trained for comparison. Results reported are on the test set with 95% confidence intervals from bootstrapping.

**Results and Conclusions:** The best performing GANN model had an AUC 0.921 (0.895-0.95). The final model hyperparameters were 1 hidden layer and 50 neurons with tanh activations; and trained with a dropout probability of 0.5 and L2 regularization lambda of 0.0001. When we visualize the top 9 features from the GANN, we see that for each feature the GANN learned a different relationship between the feature and mortality risk (Figure 2). In contrast, a logistic regression model only learned a linear relationship.

**Figure 1.** Sample of 3 continuous features for mortality risk GANN contributions across all patients, in order of highest to lowest (left to right, top to bottom). Logistic regression (LR) contributions were also plotted for comparison. The more negative the risk contribution, the less contribution the feature's value has to the risk of mortality. The features in order are urine output (UOP), average diastolic blood pressure of the last 10 minutes of the case (AVG\_DBP\_10\_min), and average mean blood pressure of the last 10 minutes of the case (AVG\_MAP\_10\_min)



## 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. Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015, August). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1721-1730). ACM.