Deep Learning for Predicting in Hospital Mortality

Presenting Author: Christine Lee, MS, UC Irvine Department of Biomedical Engineering

Co-Authors: Ira Hofer, MD, UCLA Department of Anesthesiology **;** Maxime Cannesson, MD PhD, UCLA Department of Anesthesiology**;** Pierre Baldi, PhD, UC Irvine Department of Computer Science

Introduction: Patients undergoing surgery are often at higher risk of instability during surgery as well as poor postoperative outcomes. Being able to identify patients at higher risk for poor outcomes would allow for more effective care and allocation of hospital resources, and ideally avoid complication altogether. Current risk scores such as the ASA, POSPOM, and RQI have shown success in identifying patients at risk of mortality, however, they are limited to preoperative information. The Surgical Apgar score utilizes intraoperative data, however, has been shown to have limited accuracy. We hypothesize that deep neural network models (DNNs) can leverage the complexity of intraoperative data to improve the classification of in hospital mortality in surgical patients.

Methods: Data used in these experiments came from UCLA Medical Center with IRB approval. The data consists of 59,985 patients with 87 features calculated at the end of surgery. These variables include intraoperative vital signs, interventions, and anesthesia events. The 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. Missing values were filled with the means for that feature. Values that were greater than a clinically normal maximum were set to that maximum possible. Finally, all variables were rescaled to have mean 0 and standard deviation 1. The % occurrence of in hospital mortality was 0.81%. Thus, training data was augmented 100x for a mortality occurrence of ~45% by adding Gaussian noise with a standard deviation of 0.0001 to mortality patients only. All DNNs were trained on 80% of the data (n=47,988) with five-fold cross validation. 20% of the data was held out as a future test set. All DNNs were feedforward networks with a sigmoid output, and were trained using stochastic gradient descent with momentum. Dropout, L2 weight decay and early stopping were used to prevent overfitting. We also assessed improvement of the DNN with adding ASA as a feature, and robustness of the DNN to a reduced feature set of 46. A logistic regression model with the 87 features was also trained for comparison. Performance was assessed using mean and standard deviation of AUROC from cross validation. For comparison, the AUROC of ASA, Surgical Apgar, RQI, and POSPOM were also calculated on the training data. It should be noted that RQI could not be calculated for 25,621 training patients due to lack of RQI score weights for their CPT codes.

Results: The final DNN architecture consisted of 4 hidden layers with 300 neurons in each layer. RQI outperformed Surgical Apgar, POSPOM, ASA, logistic regression, and the DNN with all 87 features (Table 1). DNN with ASA added as a feature and DNN with the reduced feature set performed comparably to RQI. However, RQI could not be calculated on approximately half of the data, while the DNNs do not have this limitation. In addition, we see that there is improved performance with the addition of the preoperative feature ASA and DNNs are robust to a reduced feature set.

Table 1. Risk score AUC results and model 5 fold cross validated AUC results (mean ± std) on training data (n=47,988). *RQI could not be calculated for 25,621 patients.

Risk Score	**AUC**
Surgical Apgar	0.58
POSPOM	0.74
ASA	0.85
RQI Score*	0.92
Model	**5 Fold Cross Validated AUC**
Logistic Regression	0.87 ± 0.02
DNN w/ 87 Intraop Features	0.90 ± 0.01
DNN w/ 87 Intraop + ASA Features	0.92 ± 0.01
DNN w/ 46 Intraop + ASA Features	0.92 ± 0.01

Conclusion: In conclusion, DNNs exhibit potential for being able to not only classify patients at risk for inhospital mortality, but also for improving upon and leveraging preoperative risk.