

Exploring ASA Score Predictability Using Machine Learning

Presenting Author: Priya Ramaswamy, MD, MEng, Beth Israel Deaconess Medical Center

Co-authors: Matthias Eikermann MD, PhD, Beth Israel Deaconess Medical Center; Peter Santer, MD PhD Beth Israel Deaconess Medical Center; Stephanie Grabitz, MD Beth Israel Deaconess Medical Center

Background/Introduction: The goal of this study is to investigate whether machine learning (ML) can accurately estimate ASA scores for pre-operative patients based on objective data contained in hospital Electronic Medical Records, and to incorporate a rule-based ground-truth misclassification correction method into algorithmic training of ASA score classification algorithms. ASA scores (range 1-6) is a subjective assessment of a patient's overall health or physical status. ASA scores are used for resource allocation, improving workflow, and reimbursement. ASA misclassification can affect observed versus expected mortality ratios of surgical patients (Helkin et al. 2017). ASA misclassification can also affect billing, since ASA 4's and 5's generally receive more reimbursement compared to ASA 1-3. Thus it is important to document accurate ASA scores. Since ASA scores are a function of patient features such as age and medical history, ML algorithms are ideally suited for automating the ASA classification system using patient characteristics, thus removing the subjectivity of human error. Prior work for ASA score prediction using ML is either based on small sample sets of few hundred patients (Karpagavalli et al. 2009, Lazouni et al. 2013) or converts the full ASA score prediction problem to an easier, binary prediction problem, which may be unsuitable for clinical use (Zhang 2016). In this study, we utilize retrospective, single-center, de-identified patient data from 202,353 patients to train an ML model that performs 4-way classification between ASA scores 1-4. Our dataset excluded scores of 5 and above.

Methods and Results: We split the dataset into training, validation, and test subsets in roughly a 60%/20%/20% split. We use 78 z-normalized features per patient for training the model; these include demographic features, medical history, and medication history. Since ASA scores are unevenly distributed across the patient population, the dataset contains significant class imbalance---the most common class contains roughly four times the samples than the least common class. Since ML algorithms do not perform well for such data, we utilize balanced sampling to convert the data to a uniform class distribution. We train four different algorithms---decision tree, bagging, random forest, and multi-class support vector machines (SVM)---and perform hyperparameter tuning using the validation set to find the algorithm with the best accuracy. We find that random forests obtain the best balanced accuracy of 71% with class-wise (1-4) recall of 90%, 50%, 54% and 83%, respectively. Note that a random classifier would obtain a balanced accuracy of 25%, showing that our method performs quite well. A previously unexplored aspect of research on ASA score prediction is that the "ground truth" ASA scores obtained from EHRs, used for training and validating ML models, are likely to contain noise due to physician errors in coding. Using the ASA guidelines (e.g., "the minimum ASA score for a patient who has had a stroke within the past three months should be 4"), we build a rule-based algorithm to correct ASA score labels in our patient records. We find that a full 8.5% of our dataset contains coding errors (17,277 cases.) We re-train our best ML algorithm using the corrected ASA score labels, which improves our prediction performance to 77% (class-wise recall of 93%, 64%, 58%, 84%), which is an improvement of 6%. This should be an essential step in future research.

Conclusion: In summary, we show that machine learning can predict ASA scores from objective data contained in EHRs and prediction performance can be improved by incorporating a rule-based ASA score correction method. Future work will focus on improving prediction performance using natural language processing-based features and automated machine learning methods.

References:

[1] Helkin A, Jain SV, Gruessner A, et al. Impact of ASA score misclassification on NSQIP predicted mortality: a retrospective analysis. *Perioperative Medicine*. 2017 6:23.

- [2] Karpagavalli, S., Jamuna, K.S., & Vijaya, M.S. Machine learning approach for preoperative anaesthetic risk prediction. *International Journal of Recent Trends in Engineering*, 1(2), 19-22. 2009
- [3] Lazouni, M. Daho, M., Settouti, N., Chikh, M., & Mahmoudi, S. Machine learning tool for automatic ASA detection. *Modeling Approaches and Algorithms for Advanced Computer Applications*. Volume 488, 9-16 2013
- [4] Zhang, L. Data-driven system for perioperative acuity prediction. Thesis, Graduate School of Vanderbilt University. Dec, 2016.