Abstract No. MSR17

Modernizing Pressure Injury Risk Assessment in the ICU in the COVID Era: Ensemble Super-Learning and Explainable AI

Yunzhe Zhou*, PhD(c), MA[1]; Reis, Lua, PhD[2]; Jenny Alderden, PhD, RN[3]; Yuwei Zhang, MD, PhD, MPH, MBA[2]; Andy R. Wilson, PhD, MStat[2,4]

[1]*University of California, Berkeley*, California; [2]*Parexel*, Raleigh, North Carolina; [3]*Boise State*, Boise, Idaho; [4]*University of Utah*, Salt Lake City, Utah

*ztzyz615@berkeley.edu

Abstract

Objectives: Hospital-acquired pressure injuries (HAPrI) are areas of injury to the skin and/or underlying tissues. Risk stratification is essential for guiding prevention in the ICU, but current risk assessment tools require labor-intensive input. This motivates a tactical, parsimonious, and automatic risk profiling algorithm based on readily available clinical measures (e.g., COVID status, race, Medicare/Medicaid status). Additionally, International Pressure Injury Prevention guidelines call for developing machine learning-based risk assessment algorithms that are clinician-interpretable and context-informed.

Methods: Adult patients admitted to one of two ICUs between April 2020 and April 2021 were eligible for inclusion. Discrete and ensemble super-learning models, adjusting for class imbalance, were created from a rich library of candidate base learners. For explainability, SHAP (SHapley Additive exPlanations) global and local values were derived to help explain variable average marginal contributions (across all permutations) to the model. An iteration of clinical expert review was performed with the SHAP values, and simulations of patient profiles and results were used to reformat and re-weight predictor variables. All analysis was run in open Python (version 3.7), and code/results will be available via a GitHub page.

Results: The final sample consisted of 1,911 patients (removing 9 with missing pressure injury status). Hospital-acquired pressure injuries (defined as stage 2, or worse) occurred in 18.5% of the sample (n=354). We achieved the best overall performance on the testing data with a stacked ensemble using three base models: random forest (rf), gradient boosted machine (gbm), and neural network (NN) (Performance on 20% holdout: Accuracy: 81%; AUC: 0.78; AUCPR: 0.53).

Conclusion: Prediction engineering should be done in collaboration with clinical experts to optimize tactical implementation to both optimize performance, with minimal interruption to workflow. XAi enhanced adoption of the experts' advice based on the selected model features.

Objectives

The goal is twofold: (1) using modern machine learning methods to enhance the performance of risk assessment for hospital-acquired pressure injuries. (2) developing explainable-AI techniques to better understand potential HAPrI etiology and extract features for more accurate risk assessment.

Methods

Ensemble machine learning: In machine learning, ensemble methods use multiple learning algorithms to obtain better predictive performance than each base learner alone [1]. Figure 1 depicts the structure of the stacked ensemble model with three base models: random forest (rf), gradient-boosted machine (gbm), and neural network (NN). The training dataset goes through each base learner, and a meta-learner combines each learner's generated predictions to enhance the prediction performance. We compare different models with the stacked ensemble model and evaluate the performance on the testing data.

Explainable-AI (XAi): SHAP (SHapley Additive exPlanations) is a method to explain individual predictions and interpret the feature importance based on the game-theoretically optimal Shapley values [2]. We use SHAP values to help explain the variable average marginal contributions to the black-box stacked ensemble model and perform an iteration of clinical expert review to reformat and re-weight predictor variables.

Results

After preprocessing and data cleaning (removing 9 with missing pressure injury status), the final sampling contains 1,911 patients, which is divided into training (80%) and testing (20%) data. Table 1 shows the performance of different models for predicting the occurrence of HAPrI. We observe that the stacked ensemble model achieves the best overall performance on the testing data (Accuracy: 81%; AUC: 0.78; AUCPR: 0.53).

Figure 2 depicts the SHAP summary plot for this stacked ensemble model. Each point on the summary plot is a SHAP value for a feature and an instance where the overlapping points are jittered in the y-axis direction to represent the distribution of SHAP values. The position on the y-axis is determined by the feature and on the x-axis by the SHAP value, with the color representing the value of the feature from low to high. We order the features according to their importance. From the plot, we can see that the top three features are: total ICU hours, maximum Modified Early Warning Score (MEWS), and the length of stay for the hospital admission. These features are further reviewed by clinical experts to optimize tactical implementation.

Conclusions

Most critical care patients are designated high-risk when using industry-standard assessment tools, such as the Braden Scale. Machine learning approaches may better discriminate high-risk critical care patients than these standard tools.

Using existing EHR data to develop risk algorithms allows for near real-time risk assessment and reduction in redundant charting. Furthermore, using XAi techniques, we were able to understand how specific features in the model influenced the predicted outcome.

Clinicians are reluctant to trust and incorporate Ai, even when it outperforms clinician judgment. By enabling clinicians to understand algorithm decision-making on the patient level, XAi will enhance trust and support collaborations between clinicians and data scientists to detect anomalies and augment algorithms to enable expert-augmented machine learning. Additionally, there are pressing issues related to clinical algorithm development, including ethics, algorithm-encoded bias, trust, regulation, and encouraging expert interaction, which have implications for the future of XAi design and motivates additional study.

Key Points

- Ensemble models are known to outperform component (base) models and have great performance
- XAi provided intuitive and clinically plausible explanations for why a patient might develop a pressure injury in the ICU

Figure 1. Flowchart for the Stacked Ensemble Model

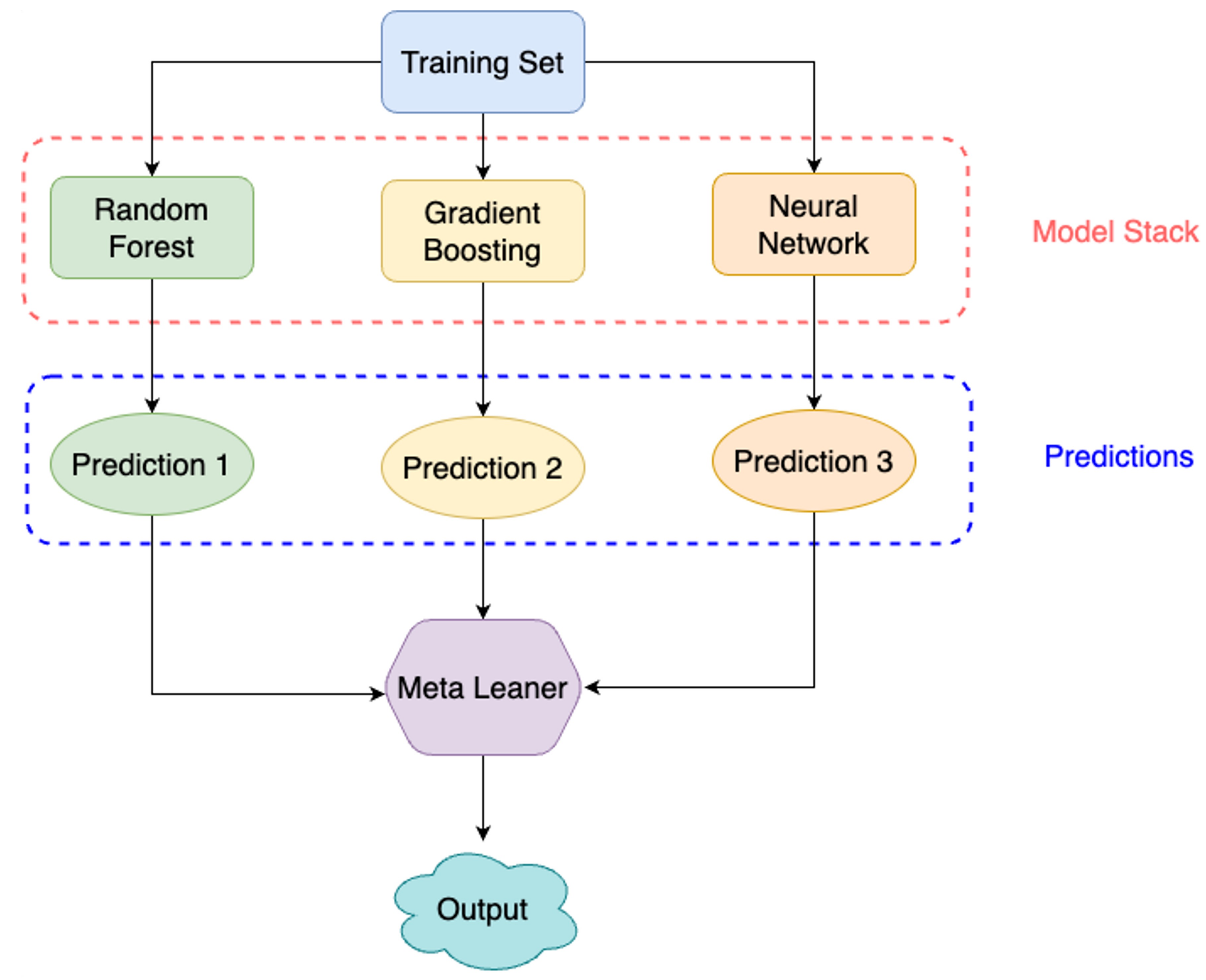

Figure 2. SHAP Summary Plot for the Stacked Ensemble Model.

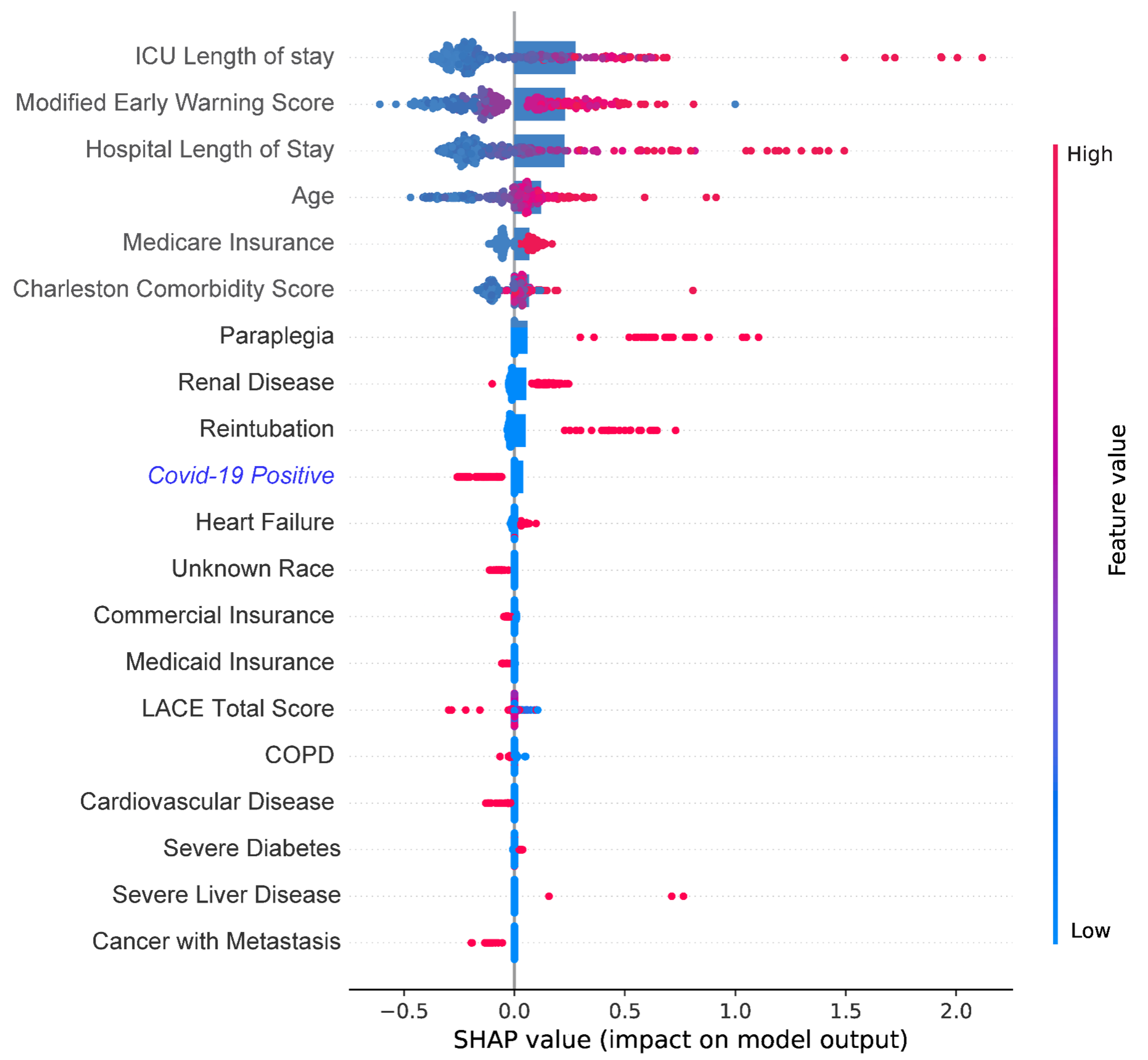

Table 1. Performance of different models for predicting the occurrence of HAPrI.

Type	Random Forest	K nearest Neighbor	Logistic Regression	Neural Network	Gradient Boosting	Stacked Ensemble
AUC	0.7761	0.7157	0.7565	0.73	0.7698	**0.7759**
AUCPR	0.5231	0.4415	0.5016	0.4775	0.5119	**0.5289**
ACCURACY	0.8061	0.7976	0.8015	0.8	0.8064	**0.8086**

References

1. Opitz, D. and Maclin, R. (1999). "Popular ensemble methods: An empirical study." Journal of Artificial Intelligence Research.
2. Lundberg, Scott M., and Su-In Lee (2017). "A unified approach to interpreting model predictions." Advances in Neural Information Processing Systems.