Alyssa N James. Machine Learning Risk Assessment Model for Hospital Acquired Pressure Injuries. A Master's Paper for the M.S. in I.S. degree. November 2021. 38 pages
Advisor: Yue Wang

Hospital Acquired Pressure Injuries (HAPI) adversely affect patient outcomes, increase health care costs, and despite considerable efforts to avoid, are increasing among critically ill patients. Many factors contribute to the development and advancement of pressure injuries in acute and critical care settings. Manual assessment tools, such as the Braden Scale, are currently utilized to predict HAPIs. This project aims to generate a model to predict HAPIs with better accuracy than the Braden Scale using a machine learning approach. Data sets were developed so predictions could be made after 48-hours from admission and 72-hours from admission. Through quantitative evaluation of the Receiver Operating Characteristic (ROC) curve, the Explainable Boosting Machine (EBM) algorithm produced a model with a greater area under the ROC curve (AUC) of 0.79 while the Braden Scale produced an AUC of 0.70 at both 48-hours and 72-hours.

Headings:

Hospital Acquired Pressure Injuries

Clinical Risk Prediction

Braden Scale Score

Machine Learning

MACHINE LEARNING RISK ASSESSMENT MODEL FOR HOSPITAL ACQUIRED PRESSURE INJURIES

by
Alyssa N James

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

November 2021

Approved by

Yue Wang

Yue Wang

Table of Contents

LITERATURE REVIEW

Introduction

Hospital Acquired Pressure Injuries (HAPIs) adversely affect patient outcomes, increase health care costs, and despite considerable efforts to avoid, are increasing among critically ill patients (Darvall et al., 2018).

Pressure injuries (PI) occur when pressure over skin and underlying tissue results in localized areas of damage. Bony prominences and areas under medical devices are particularly susceptible. Pressure injuries are categorized by stage: Stage 1 (non-blanchable erythema of intact skin) through stage 4 (full-thickness skin and tissue loss), plus categories of unstageable, deep tissue, and mucus membrane PIs (National Pressure Ulcer Advisory Panel, n.d.).

A pressure injury is considered hospital acquired when it is discovered and documented 24 hours after admission to an inpatient unit. If it is discovered and documented within the 24-hour time frame, it is considered community acquired and not due to quality of care received from the hospital. Hospital acquired pressure injuries increase length of stay for patients, increase suffering, and increase morbidity and mortality (Padula et al., 2015). Furthermore, stage 3 HAPIs add an average additional cost of $43,180 per injury (Edger, 2017; Zaratkiewicz et al., 2010), and are not reimbursed by the Centers for Medicare & Medicaid Services (CMS) since HAPIs are considered avoidable and were

added to the list of non-reimbursable hospital acquired conditions (Padula et al., 2015). Further, the Patient Protection and Affordable Care Act (PPACA) penalizes hospitals with high HAPI rates. In Fiscal Years 2018 and 2019, 137 pressure injuries assessed as stage 3 or higher occurred at a large academic medical center in the southeast, resulting in an estimated $5,915,660 lost revenue. Despite this hospital's considerable investment and efforts to reduce HAPIs, their HAPI rates continue to be among the highest in the nation. There are considerable incentives, both to the patient and the institution, to accurately predict HAPIs and provide tailored prevention interventions based on the patient's risk.

Pressure Injuries

Many factors contribute to the development and advancement of pressure injuries in acute and critical care settings. Intrinsic patient characteristics affect the architecture and integrity of the skin and include factors such as disease or injury status, sensory perception, and age. Extrinsic factors reduce tissue tolerance (the ability of the skin and underlying tissue to avoid adverse effects of pressure) to pressure through exposure to moisture, friction, and/or shearing forces (Braden & Bergstrom, 1987; Edsberg et al., 2014). Hospital staff can apply interventions to prevent pressure injuries; however, the additional cost of prevention is passed to the patient during their stay at the hospital. For this reason, PI risk assessment tools are critical to identify patients who need interventions. The widely adopted risk assessment tool is the Braden Scale, which was developed in 1987.

The Braden Scale score is developed by assessing the patient on 6 sub-scales and then summing for a total score. The sub-scales include sensory perception, activity, mobility,

moisture, nutrition, with scores ranging from 1 to 4, and friction/shear with scores ranging from 1 to 3 (Cox, 2012). Clinicians must physically assess the patient and assign a subscale score that most closely aligns with the descriptors.

Since the development of the Braden Scale, advances in critical care technologies now save sicker and more compromised patients who would not have survived in the past (Cox & Roche, 2015). However, these same technologies, unavoidable due to their life-saving attributes, may contribute to the development of HAPIs. Assessment tools, such as the Braden Scale, intended to identify patients at risk for PIs, were developed prior to technological advancement of critical care settings, and do not reliably and specifically predict which patients will develop HAPIs (Cox, 2017). Cox and Tescher et al. have found the Braden Scale score to be useful in predicting pressure injuries in non-critical care areas but cannot account for the complexities in critical care settings, and it is not useful in helping the clinicians target prevention treatments (2012). The academic medical center that is partnering with this project struggles to apply targeted interventions in their large critical care population. New knowledge is needed to identify which combinations of intrinsic and extrinsic, modifiable and non-modifiable risk factors lead to HAPIs.

Machine learning is a useful approach for complex questions because it's able to quickly handle statistical computations across a vast number of independent variables. The application of machine learning in health care is growing due to rapid analysis of routinely collected data and has increasingly demonstrated its value in hospital quality improvement (Tran et al., 2014). Machine learning techniques have been utilized in

healthcare settings for the purpose of developing predictive models, which we will explore in the next section.

Machine Learning

There is a growing burden on nurses to document repetitive assessments in the electronic health record (EHR). Due to the manual nature of the Braden Scale tool, nurses spend a significant amount of time assessing and documenting findings on patients. Using a statistical approach, there is an opportunity to automate pressure injury risk assessment as well as improve accuracy by assessing variables that are already documented within the patient's electronic health record. An automated risk assessment tool would alleviate the burden on nurses and create more time to care for patients.

Machine learning is grounded in the field of statistical learning, which is the application of quantitative methods to explain a phenomenon through data. Machine learning is a method of positivist research, in which the assumptions are that the phenomenon exists independently of observation and that it can be simplified and predicted (Wildemuth, 2017). Machine learning is made possible by accessible and powerful software that is designed to predict simple linear and complex non-linear relationships (James et al., 2017). Machine learning works by rapidly measuring the relationship between input variables (or a combination of input variables) and the output variable, and then calculating a formula to predict the output variable.

There are a variety of predictive models that can be employed by machine learning, either linear or non-linear. The most common form of predictive model is the linear regression.

In a linear model the predictors and response variables move together in a straight line. This type of model is very stable with low variation, but also has high bias, because it assumes a linear relationship (Hastie et al., 2009). Non-linear models can calculate highly accurate predictions of non-linear relationships with low bias because they easily adapt to the input data. However, the adaptability of a non-linear model risks becoming unstable with high variation (Hastie et al., 2009). In machine learning, if a model accommodates too much variation in the training data, it loses it predictive power when applied to the test data. Determining best fit for a model is balancing bias and variance in order to produce consistent and reliable results that represent the true form of the data (James et al., 2017).

Prior to Lou et al.'s work in expanding learning Generalized Additive Models (GAMs), interpretability was lost at the expense of higher accuracy complex models (2012). Interpretability in this context means that users can understand how each variable contributes to and impacts the predictor in the model. Lou et al.'s original work did not model interactions between variables, which limited the accuracy compared to full complexity models. They later adapted GAM to include pairwise interactions, calling it Generalized Additive Models with Pairwise Interactions (GA^2M), in order to achieve higher accuracy, while maintaining intelligibility (2013). Then, Caruana et al. demonstrated the application of the GA^2M in constructing useful models in healthcare (2015).

Automated risk assessments are the most common application of machine learning in medicine (Deo, 2015). Machine learning is currently applied in predicting a patient's risk of acquiring pneumonia, sepsis, central line-associated bloodstream infections, heart disease, and recently pressure injuries (Caruana et al., 2015; Yee et al., 2019; Parreco et al., 2018; Nahar et al., 2013; Alderdan et al., 2018).

Alderden et al. demonstrated that a machine learning approach predicted pressure injuries with greater accuracy than the Braden Scale despite being limited in aspects of study design and scope (2018). A predictive modeling that accurately identifies critically or acutely ill patients at risk for developing HAPIs remains greatly underdeveloped. The intention of this study is to extend the work they initiated and develop a more robust prediction tool, using GA^2M, for a wider patient population.

GUIDING QUESTION

Can an algorithm identify patients at risk of developing a pressure injury better than the manual risk assessment tool by using available clinical data?

METHODS

Data Collection

The institutional review board at the University of North Carolina at Chapel Hill approved the study. Data was extracted from an enterprise data warehouse for Electronic Health Record (EHR) data, the Carolina Data Warehouse for Health (CDW-H). Two data sets were developed using the same query, but each data set had a different time period. The first data set gathered patient information within the first 48 hours of admission with 56,726 records. And the second data set gathered patient information within 72 hours of admission with 43,629 records. The first data set accounted for the largest sample size as its time criteria fit a larger number of patients. The number of records in the second data set was reduced, because patients either were discharged or deceased between the 48-hour or 72-hour interval. The query used to build the data set was refined in an iterative fashion, where each variable was validated by clinical experts before developing the next variable, which then informed necessary updates to the query. Provided in the Appendix, Table 1 lists the variable names and definitions. The final two data sets were validated by

randomly sampling 10 patients, 5 with pressure injuries and 5 without from each data set, and manually comparing the results to the clinicians view in the EHR. The final data sets were compiled and stored in a SQL database.

Sample Selection Procedure

The sample consisted of all patients admitted to critical care and acute care units at UNC Hospitals, an academic medical center with a level one trauma center, from January 2017 - December 2018. All patients under the age of 18 were excluded, due to this patient population needing a different risk assessment scale, the Braden QD Scale (Curley et al., 2018). HAPIs that were indicated as present on admission, documented within 24 hours of admission, were excluded from the target variable to avoid misattribution of community acquired pressure injuries as HAPIs. Only the first encounter was recorded when a patient had repeated encounters during the sample time period regardless of HAPI status.

Variable Selection Procedure

Variables were selected and engineered based on relevant publications, and interviews with clinical experts, staff nurses, and Wound Ostomy and Continence Nurses (WOCNS). Interviews with clinicians were critical in development of the data set as it guided the selection of known and suspected contributing factors based upon their experience caring for patients with pressure injuries, and to better understand their workflow and patient population. There was a total of 139 variables describing each patient in the data set. Variables ranged from patient history such as age and gender to

continuously monitored variables like heart rate and blood pressure, to lab tests like Hemoglobin and Creatinine levels, to duration of hospital stay and procedures or therapies. Necessary steps to protect privacy were taken including, de-identification of all medical records and storage on a secured database in the health system's secured network. The target was a binary variable of either a HAPI occurring or not occurring during the sample time period. A HAPI occurring included HAPIs stage 1 to 4, deep tissue injury, or unstageable. Only stage 2 and above HAPIs have financial repercussions, but stage 1 HAPIs were included in this study due to positive benefits of early identification and intervention. In the 48-hour data set, 222 patients developed a HAPI (0.39%), and in the 72-hour data set, 384 developed a HAPI (0.88%).

Analysis

The analysis described below was applied to each of the two data sets separately, in order to develop two independent models that produced predictions at 48 hours and 72 hours of admission to the hospital.

The purpose of developing two models with varying time intervals was to identify key predictive variables as a patient's condition changes over time. The time intervals were chosen as a first step to establish a need for time series predictions. The current manual risk assessment system has a protocol to conduct a full skin assessment to identify pressure injuries and determine risk within 24 hours of admission, and then once per day for the entire length of stay at the hospital. The sample data sets were designed to closely mimic the current manual system for comparison purposes. If a pressure injury is

discovered within the first 24 hours, it is considered a community acquired pressure; this is considered a predictor variable instead of the target variable, as it is impossible to predict community acquired pressure injuries. For this reason, we did not include a separate 0-24-hour model in this study.

This study will train two separate models.

- The first model will predict the first hospital acquired pressure injury from data collected between 0-48 hours of admission.

  Pr(first HAPI in 0-48 hours)

Along with HAPI predictions, this model will display the interactions between predictor variables and the target variable, so clinicians can determine which interventions to apply within the first 48 hours of admission in order to prevent a pressure injury from occurring.

- The second model will predict the first pressure injury between 0-72 hours of admission.

  Pr(first HAPI in 0-72 hours)

Compared to the first model, this model will offer insights into new variables that alter the patient's risk for acquiring a HAPI in a longer time interval after admission.

Explainable Boosting Machine

In health care, the most useful models are the ones where clinicians find the predictions accurate but also explainable. For this reason, the Explainable Boosting Machine (EBM) model was selected. EBM is a generalized additive model with pairwise Interaction that balances the intelligibility of a linear model, like logistic regression, with the

performance of a non-linear model, like random forest (Caruana et al, 2015). EBM is considered an intelligible model because the direction and magnitude of the predictor variables' interactions with the predicted outcome can be graphed as line graphs, and the pairwise interactions can be graphed as heat maps. These learned interactions can be edited based on clinical knowledge, which again contributes to the intelligibility of EBM.

Overfitting is troublesome to a model, because it will follow the training data set too closely and will lower the accuracy of the model when used on test data sets. Bagging refers to the collection of many random subsamples of data with replacement, and each subsample is referred to as a bag. Bagging is used widely in machine learning because it reduces variance in the data while retaining bias, which improves the accuracy of the model (James et al., 2017). For these reasons, Caruana et al's EBM model integrated into an open source InterpretML python package was utilized for this analysis (Nori et al., 2019). The EBM model utilizes techniques like bagging, gradient boosting, and automatic interaction detection, so no additional bagging or boosting was done at this time.

Data Processing and Missing Value Imputation

Data processing and model creation were executed in python. The same data processing steps were utilized for both the 48-hour and 72-hour data sets. The Pandas python package has a profiling report function, which I executed first to explore data types, descriptive statistics, correlations, and missing values and potentially erroneous values (Brugman, 2021). From this report, variables that needed preprocessing were identified. The clinical team was consulted when defining the logic to impute values for each

variable. Below is the plan that was followed to impute missing values and handle erroneous values.

For lab value variables, an out-of-range value of “9999999” was recorded if a patient’s lab value was incorrect. This recording was done in the lab, and no other values could be retrieved from the data warehouse. The out-of-range value is significantly higher than actual lab values, which skewed the data. For this reason, out-of-range values were replaced with the max value within the range. For example, excluding the out-range-value, if hemoglobin values in the data sets range between 1.7 - 20.3, then the out-of-range value of “9999999” would be replaced with 20.3.

Null values for continuous or categorical features like, Recovery Time, First BMI, First Weight, Severity of Illness, Total Urine Output, and lab values were replaced with the mean value of that feature. Null values in indicator fields, like Has Diabetes, Has Anemia, Has GI Bleed, etc. were replaced with zero, which indicate that they were not present during the patient’s encounter. Patient sex was encoded as zero for male and one for female. Null values in Lowest Glasgow Score were replaced with 15, which indicates that the patient has no brain injury. Null values in Average RASS were replaced with zero, which indicates that the patient is not sedated. Finally, the Average Braden Score was encoded as either zero, not at risk, or one, at risk, using 16 as the cutoff (Hyun et al., 2013; Alderden et al., 2018). The encoded Average Braden Score was not used in the training of the model but used in the evaluation of the model.

To ensure that there was independence between predictor variables, the Variance Inflation Factor (VIF) was reviewed for each variable. The VIF is widely used to assess multicollinearity between variables and was calculated for continuous variables in python using the variance_inflaction_factor from the statsmodels package (Hastie et al., 2009; Perktold, 2019). I then removed the variables with the largest VIF one at a time and reran the analysis. I repeated that process for all variables with a VIF over ten. Ten is the recommended cutoff point (Hastie et al., 2009).

Finally, 105 independent variables remained for model creation.

Model Creation

The data set was divided into training and test data sets using a 70:30 train:test split (Caruana et al, 2015). To divide the data set the train_test_split function from the scikit-learn Python package was utilized (Pedgrosa et al., 2011). A sampling seed was set, so that results could be replicated, and the shuffle parameter was set to true. The target in the 48-hour data set was the "HAS_HAPI_48" variable and the target in the 72-hour data set was the "HAS_HAPI_72" variable. The 48-hour train data set contained 39,708 records while the 48-hour test data set contained 17,018 records. The 72-hour train data set contained 30,540 records while the 72-hour test data set contained 13,089 records.

The test data sets were set aside and were not used to train the model. The EBM model was developed using the InterpretML Python package on the training data sets. The test data sets were then used to evaluate the model's performance.

Evaluation

The EBM model used on the test data set produced true-positive and false-positive results that were fitted to the receiver operating curve (ROC curve) using the scikit-learn Python package. The area under the curve (AUC) was calculated as a score of accuracy. Models that have a larger AUC do a better job of predicting true-positives while minimizing false-positives (Hasite et al., 2009). That same information can be produced from Braden Scale scores on the same sets of data using the encoded "AVERAGE_BRADEN" feature and the target feature. AUC scores of both models for each time frame (48-hour and 72-hour) were compared to evaluate stronger performance (Alderden et al., 2018).

The benchmark AUC score for the 48-hour Braden Scale was 0.70 and the 72-hour Braden Scale was also 0.70, as displayed in Figures 1 and 2.

The first iteration of the EBM model was developed using 103 independent predictor variables. The results of the first model returned an Area Under the Curve (AUC) of 0.77. To identify useful variables, reduce variance and improve AUC of the EBM model, variables were systematically removed one by one, a model was created, and AUC measured. If the resulting AUC was higher or remained consistent with the first 0.77, then the variable was kept out, if the AUC was lowered then the variable was added back in. This process was repeated until 34 variables were identified to improve AUC to 0.79, these variables were used in the final model creation and are identified in Table 1 in the Appendix.

To further evaluate the performance of the EBM model, I built two other models in the InterpretML Python package, a Decision Tree model with a max depth of 10, and a Logistic Regression model, using the same 34 features. The performance of each model is depicted in the ROC Curve graph in Figures 1 and 2.

Figure 1:

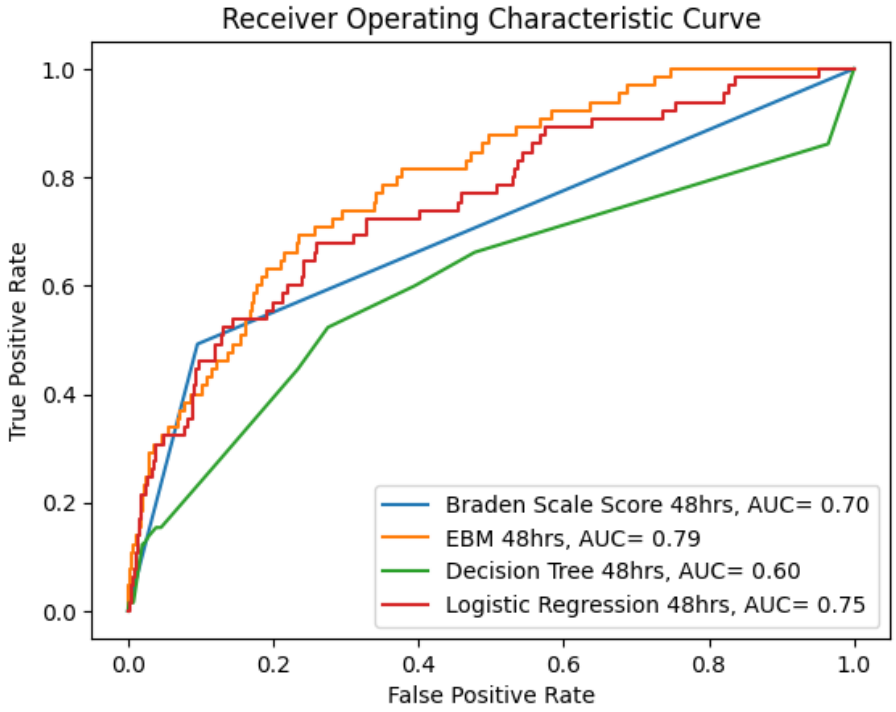
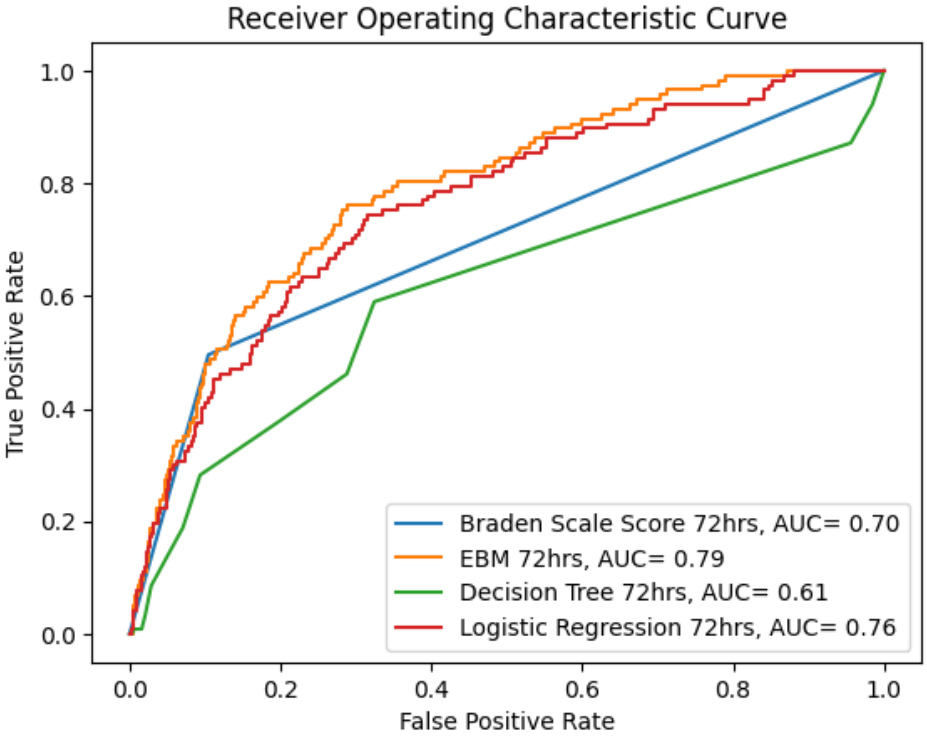


Figure 2:



In review of Figures 1 and 2, the final EBM model had the best performance, with an AUC of 0.79. At the 48-hour and 72-hour time periods, the Braden Scale Score had an AUC of 0.70, which required time dedicated to manual patient assessments at regular intervals during each period. In this study, the EBM model performed better than the Braden Scale Score and has the additional benefit of being automated, which could free up care time from the nursing team.

DISCUSSION

The EBM model's relatively strong performance suggests that the model would be a useful replacement to the Braden Scale Score. A reliable and automated risk prediction system for pressure injuries offers many advantages over the current manual assessments. The first major advantage would be time savings for the nursing care teams by removing the need for manual assessments at regular intervals. At the hospital in this study, patients must be assessed every twelve hours if they are not at risk and then every four hours if they are to be found at risk using the Braden Scale Score. If the patient is at risk, then the nursing care team must implement additional prevention strategies, like turning every 2-4 hours, nutrition changes, pressure redistribution, etc., which increases time and cost passed on to the patient. The EBM model has a higher true positivity rate and lower false positivity rate, which means that additional time and cost will only be dedicated to patients who need the care. Similarly, if prevention strategies are targeted to patients who are truly at risk, then hopefully rates of HAPI will reduce and lower the financial burden for the hospital. Another advantage of the EBM model is the flexibility to adapt to different patients and settings. As discussed later in this section, the model can easily be trained for different units and levels of care to provide better predictions of risk. A final advantage of an implemented automated EBM model would be its real timeliness. As discussed earlier, nurses must assess for pressure injuries at regular intervals, and in the time between assessments a patient's risk factors can change and put them at risk of

developing a pressure injury. Since the EBM model uses some continuously monitored variables, it has the potential to provide an updated prediction of risk as soon as the patient's condition changes.

In Figures 3 and 4, the most predictive variables are ranked by the EBM model. In the 48-hour model, the six most important variables were age (at time of admission), the interaction between age and the number of community acquired pressure injuries, BUN level, BUN creatinine ratio, diabetic, and receiving blood thinners subcutaneously. In the 72-hour model, the six most important variables shifted to BUN creatinine ratio, diabetic, BUN level, receiving skeletal muscle relaxant orally, age (at time of admission), and alkaline phosphatase level.

Figure 3: 48-hour EBM

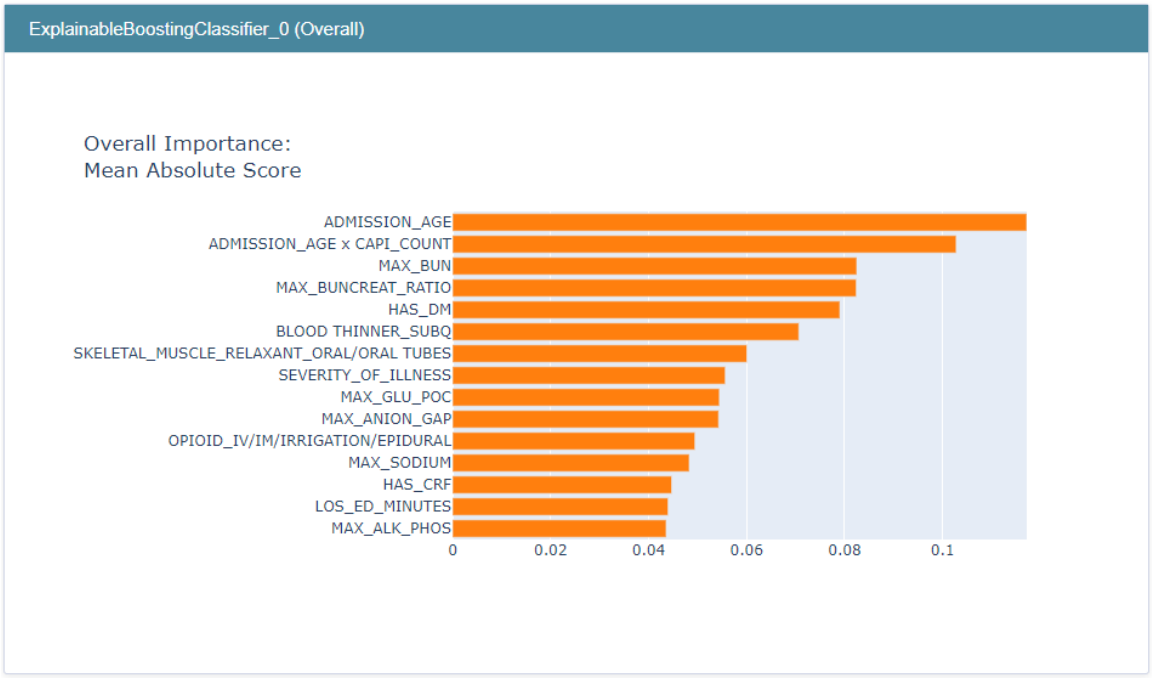
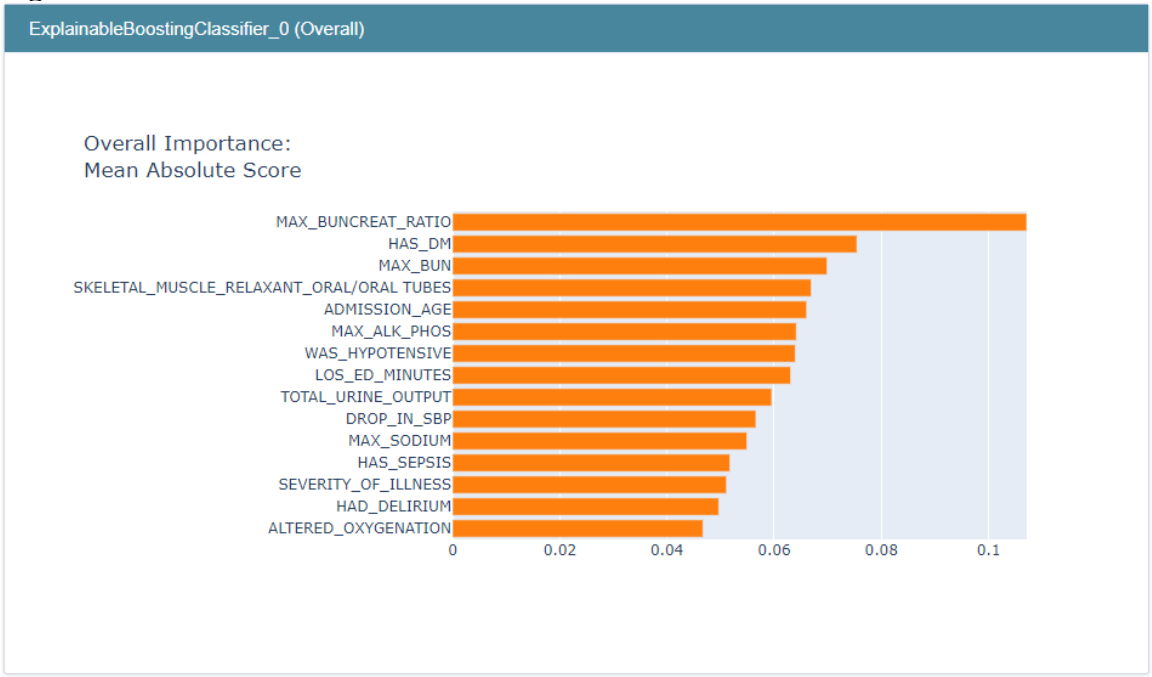


Figure 4: 72-hour EBM



Surprisingly, variables that were found to be predictive of pressure injuries in other studies, like BMI or weight, were excluded from these models due to their high multicollinearity. Similarly, I initially included medications, like vasopressors, since they have been predictive of pressure injuries in other research, but ultimately removed them because they did not improve the accuracy of these models.

To improve the model performance and prepare them for implementation in a hospital setting, I would like to train an EBM model for each major adult care level, like Intensive Care Unit (ICU), Stepdown, and Acute. I believe this would be beneficial as patients in Stepdown and ICU have more complex cases and likely different predictive variables influencing pressure injury risk.

This study supports the need for a time series predictive model as patients' risk factors change during the length of their admission. Time series predictions would also improve the accuracy of the model during implementation in a hospital setting.

CONCLUSION

I developed two models to predict risk for pressure injuries among all adult patients within 48 hours and 72 hours of admission by using a machine learning Explainable Boosting Machine approach. These models rely on information that is readily available in an Electronic Health Record and does not require clinicians to manually assess patient against a tool, like the Braden Scale. The next step will be to review the interactions learned by the models with the clinical team and retrain ICU, Stepdown, and Acute models to optimize their predictive power. Finally, a time series approach will be explored as part of implementing the model in a hospital setting.

APPENDIX

Table 1: Variable List

	Variable List	Definition	Used in 48-hour Model	Used in 72-hour Model
1	PATIENT_ENCOUNTER_NUMBER	Unique identifier for patient admission		
2	ADMISSION_INSTANT	Date/time of hospital admission		
3	INPATIENT_ADMISSION_INSTANT	Date/time for Inpatient Admission (when the patient is classified as inpatient)		
4	LOS_ED_MINUTES	Length of stay in the Emergency Department (ED)	✓	✓
5	DISCHARGE_INSTANT	Date/time of hospital discharge		
6	ADMISSION_AGE	Patient's age at time of admission	✓	✓
7	PATIENT_SEX	Patient's sex		
8	HAS_HAPI_48	Indicates if the patient had a HAPI within 48 hours of admission		
9	HAS_HAPI_72	Indicates if the patient had a HAPI within 72 hours of admission		
10	HAS_HAPI	Indicates if the patient had a HAPI within their entire admission		
11	HAS_CAPI	Indicates if the patient arrived with a community acquired pressure injury	✓	✓
12	CAPI_COUNT	The number of community acquired pressure injuries the patient arrived with	✓	✓
13	TOTAL_SURGERY_TIME	Length of time in minutes spent in the Operating Room	✓	✓
14	TOTAL_RECOVERY_TIME	Length of time in minutes spent in Recovery	✓	✓
15	HAS_DM	Indicates if the patient had diabetes during the time period. ICD-10 codes were used. Codes that start with E08, codes that start	✓	✓

		with E09, codes that start with E10, codes that start with E11, codes that start with 024, codes that start with E13		
16	HAS_ANEMIA	Indicates if the patient had anemia during the time period. ICD-10 codes were used. Consider all ICD-10 codes that start with D50, D51, D52, D53	✓	✓
17	HAS_GI_BLEED	Indicates if the patient had anemia during the time period. ICD-10 codes were used. Consider ICD-10 Codes = K92.0, K92.1, K92.2	✓	✓
18	HAS_PVD	Indicates if the patient had Peripheral Vascular Disease during the time period. ICD-10 codes were used. Consider all ICD-10 codes that start with I73	✓	✓
19	HAS_ARF	Indicates if the patient had Acute Renal Failure during the time period. ICD-10 codes were used. Consider all ICD-10 codes that start with N17		
20	HAS_CRF	Indicates if the patient had Chronic Renal Failure during the time period. ICD-10 codes were used. Consider all ICD-10 codes that start with N18		
21	HAS_TXP	Indicates if the patient received a transplant during the time period. ICD-10 codes were used. Consider all ICD-10 codes that start with Z94		
22	HAS_SEPSIS	Indicates if the patient had sepsis during the time period. ICD-10 codes were used. Consider all ICD-10 codes that start with A41	✓	✓
23	HX_NICOTINE	Indicates if the patient had a history of nicotine use	✓	✓
24	TOBACCO_USE	Indicates if the patient had a history of tobacco use		
25	FIRST_BMI	First BMI recorded of encounter		
26	FIRST_WEIGHT	First weight recorded of encounter		
27	LOWEST_GLASGOW	Lowest Glasgow Coma Scale score within time period		

28	HAD_DELIRIUM	Indicates if the patient was delirious during the time period. Used the Confusion Assessment Method (CAM).	✓	✓
29	WAS_HYPOTENSIVE	Indicates if the patient was hypotensive during the time period. The Mean Arterial Pressure (MAP) was reviewed. If MAP was less than 60 mm Hg then the patient was indicated to be hypotensive, otherwise not hypotensive	✓	✓
30	ALTERED_OXYGENATION	Indicates if the patient had altered oxygenation during time period. If oxygen saturation was less than 90% by pulse oximetry, then the patient was indicated to have altered oxygenation, otherwise not	✓	✓
31	AVERAGE_RASS	The average value recorded within time period for Richmond Agitation Assessment Scale (RASS)	✓	✓
32	SEVERITY_OF_ILLNESS	Maximum Mew's Score recorded within time period	✓	✓
33	FEVER	Indicates if the patient had a fever, if temperature was greater than thirty-eight degrees Celsius during time period		
34	HYPOTHERMIA	Indicates if the patient had hypothermia, if temperature was less than thirty-six degrees Celsius during time period		
35	RECEIVED_CRRT	Indicates if the patient received Continuous Renal Replacement Therapy (CRRT) during time period		
36	DROP_IN_SBP	Indicates if the patient experienced a drop in systolic blood pressure during time period. If systolic blood pressure was less than 90 mmHg, then the patient was indicated to have a drop in blood pressure.		
37	TOTAL_URINE_OUTPUT	Total measure urine output from patient during time period		

38	AVERAGE_BRADEN	Average Braden Scale Score recorded during time period. Used for evaluation of model, and not used in the model training.		
39	MIN_HGB	Minimum Lab Value: Hemoglobin within time period		
40	MIN_HCT	Minimum Lab Value: Hematocrit within time period		
41	MAX_A1C	Maximum Lab Value: Hemoglobin A1C within time period		
42	MIN_ALBUMIN	Minimum Lab Value: Albumin within time period		
43	MAX_LACTATE	Maximum Lab Value: Lactate within time period	✓	✓
44	MAX_CREATININE	Maximum Lab Value: Creatinine within time period	✓	✓
45	MAX_GLU_POC	Maximum Lab Value: Glucose within time period	✓	✓
46	MAX_SODIUM	Maximum Lab Value: Sodium within time period	✓	✓
47	MIN_SODIUM	Minimum Lab Value: Sodium within time period		
48	MAX_POTASSIUM	Maximum Lab Value: Potassium within time period		
49	MIN_POTASSIUM	Minimum Lab Value Potassium within time period		
50	MAX_CHLORIDE	Maximum Lab Value: Chloride within time period		
51	MIN_CHLORIDE	Minimum Lab Value: Chloride within time period		
52	MAX_CO2	Maximum Lab Value: CO2 within time period		
53	MAX_BUN	Maximum Lab Value: BUN within time period	✓	✓
54	MIN_BUN	Minimum Lab Value: BUN within time period		
55	MAX_BUNCREAT_RATIO	Maximum Lab Value: BUN/Creatinine Ratio within time period	✓	✓
56	MAX_ANION_GAP	Maximum Lab Value: Anion Gap within time period	✓	✓
57	MAX_CALCIUM	Maximum Lab Value: Calcium within time period		
58	MAX_MAGNESIUM	Maximum Lab Value: Magnesium within time period		

59	MIN_MAGNESIUM	Minimum Lab Value: Magnesium within time period		
60	MAX_PHOSPHOROUS	Maximum Lab Value: Phosphorous within time period		
61	MIN_TOTAL_PROTEIN	Minimum Lab Value: Total Protein within time period		
62	MAX_TOTAL_BILI	Maximum Lab Value: Total Bilirubin within time period	✓	✓
63	MAX_AST	Maximum Lab Value: AST within time period		
64	MAX_ALT	Maximum Lab Value: ALT within time period	✓	✓
65	MAX_ALK_PHOS	Maximum Lab Value: Alkaline Phosphatase within time period	✓	✓
66	ANTIDEPRESSANT_ORAL/ ORAL TUBES	Indicates if the patient received this medication and route combination within the time period		
67	ANTIDIABETIC_INHALED	Indicates if the patient received this medication and route combination within the time period		
68	ANTIDIABETIC_IV/IM/IRR IGATION/EPIDURAL	Indicates if the patient received this medication and route combination within the time period		
69	ANTIDIABETIC_ORAL/OR AL TUBES	Indicates if the patient received this medication and route combination within the time period		
70	ANTIDIABETIC_OTHER	Indicates if the patient received this medication and route combination within the time period		
71	ANTIDIABETIC_SUBQ	Indicates if the patient received this medication and route combination within the time period		
72	ANTIPSYCHOTIC_IV/IM/IR RIGATION/EPIDURAL	Indicates if the patient received this medication and route combination within the time period		
73	ANTIPSYCHOTIC_ORAL/O RAL TUBES	Indicates if the patient received this medication and route combination within the time period		

74	ANTIPSYCHOTIC_OTHER	Indicates if the patient received this medication and route combination within the time period		
75	ANTIPSYCHOTIC_RECTAL	Indicates if the patient received this medication and route combination within the time period		
76	BARBITURATE_IV/IM/IRRIGATION/EPIDURAL	Indicates if the patient received this medication and route combination within the time period		
77	BARBITURATE_ORAL/ORAL TUBES	Indicates if the patient received this medication and route combination within the time period		
78	BARBITURATE_OTHER	Indicates if the patient received this medication and route combination within the time period		
79	BLOOD THINNER_ET/INTRA-PLACES	Indicates if the patient received this medication and route combination within the time period		
80	BLOOD THINNER_IV/IM/IRRIGATION/EPIDURAL	Indicates if the patient received this medication and route combination within the time period		
81	BLOOD THINNER_ORAL/ORAL TUBES	Indicates if the patient received this medication and route combination within the time period		
82	BLOOD THINNER_OTHER	Indicates if the patient received this medication and route combination within the time period		
83	BLOOD THINNER_PERITONEAL	Indicates if the patient received this medication and route combination within the time period		
84	BLOOD THINNER_SUBQ	Indicates if the patient received this medication and route combination within the time period	✓	✓

85	BLOOD THINNER_TOPICAL/MISC/ OTHER	Indicates if the patient received this medication and route combination within the time period
86	BZD_ET/INTRA-PLACES	Indicates if the patient received this medication and route combination within the time period
87	BZD_IV/IM/IRRIGATION/E PIDURAL	Indicates if the patient received this medication and route combination within the time period
88	BZD_NOSE	Indicates if the patient received this medication and route combination within the time period
89	BZD_ORAL/ORAL TUBES	Indicates if the patient received this medication and route combination within the time period
90	BZD_OTHER	Indicates if the patient received this medication and route combination within the time period
91	GENERAL_ANESTHETIC_I NHALED	Indicates if the patient received this medication and route combination within the time period
92	GENERAL_ANESTHETIC_I V/IM/IRRIGATION/EPIDUR AL	Indicates if the patient received this medication and route combination within the time period
93	GENERAL_ANESTHETIC_ NOSE	Indicates if the patient received this medication and route combination within the time period
94	GENERAL_ANESTHETIC_ ORAL/ORAL TUBES	Indicates if the patient received this medication and route combination within the time period
95	GENERAL_ANESTHETIC_ OTHER	Indicates if the patient received this medication and route combination within the time period
96	LAXATIVE_ET/INTRA-PLACES	Indicates if the patient received this medication and route

		combination within the time period		
97	LAXATIVE_IV/IM/IRRIGATION/EPIDURAL	Indicates if the patient received this medication and route combination within the time period		
98	LAXATIVE_ORAL/ORAL TUBES	Indicates if the patient received this medication and route combination within the time period		
99	LAXATIVE_OTHER	Indicates if the patient received this medication and route combination within the time period		
100	LAXATIVE_RECTAL	Indicates if the patient received this medication and route combination within the time period		
101	LAXATIVE_SUBQ	Indicates if the patient received this medication and route combination within the time period		
102	LAXATIVE_TOPICAL/MISC/OTHER	Indicates if the patient received this medication and route combination within the time period		
103	NONBZD_HYPNOTIC_ORAL/ORAL TUBES	Indicates if the patient received this medication and route combination within the time period		
104	NONBZD_HYPNOTIC_OTHER	Indicates if the patient received this medication and route combination within the time period		
105	OPIOID_INHALED	Indicates if the patient received this medication and route combination within the time period		
106	OPIOID_INTRATHECAL	Indicates if the patient received this medication and route combination within the time period		
107	OPIOID_IV/IM/IRRIGATION/EPIDURAL	Indicates if the patient received this medication and route combination within the time period	✓	✓

108	OPIOID_NERVE BLOCK	Indicates if the patient received this medication and route combination within the time period		
109	OPIOID_NOSE	Indicates if the patient received this medication and route combination within the time period		
110	OPIOID_ORAL/ORAL TUBES	Indicates if the patient received this medication and route combination within the time period		
111	OPIOID_OTHER	Indicates if the patient received this medication and route combination within the time period		
112	OPIOID_RECTAL	Indicates if the patient received this medication and route combination within the time period		
113	OPIOID_SUBQ	Indicates if the patient received this medication and route combination within the time period		
114	OREXIN_ANTAGONIST_ORAL/ORAL TUBES	Indicates if the patient received this medication and route combination within the time period		
115	SKELETAL_MUSCLE_RELAXANT_INTRATHECAL	Indicates if the patient received this medication and route combination within the time period		
116	SKELETAL_MUSCLE_RELAXANT_IV/IM/IRRIGATION/EPIDURAL	Indicates if the patient received this medication and route combination within the time period		
117	SKELETAL_MUSCLE_RELAXANT_ORAL/ORAL TUBES	Indicates if the patient received this medication and route combination within the time period	✓	✓
118	SKELETAL_MUSCLE_RELAXANT_TOPICAL/MISC/OTHER	Indicates if the patient received this medication and route combination within the time period		
119	STEROID_EARS	Indicates if the patient received this medication and route		

		combination within the time period
120	STEROID_ET/INTRA-PLACES	Indicates if the patient received this medication and route combination within the time period
121	STEROID_EYES	Indicates if the patient received this medication and route combination within the time period
122	STEROID_INHALED	Indicates if the patient received this medication and route combination within the time period
123	STEROID_INTRATHECAL	Indicates if the patient received this medication and route combination within the time period
124	STEROID_IV/IM/IRRIGATION/EPIDURAL	Indicates if the patient received this medication and route combination within the time period
125	STEROID_NOSE	Indicates if the patient received this medication and route combination within the time period
126	STEROID_ORAL/ORAL TUBES	Indicates if the patient received this medication and route combination within the time period
127	STEROID_OTHER	Indicates if the patient received this medication and route combination within the time period
128	STEROID_RECTAL	Indicates if the patient received this medication and route combination within the time period
129	STEROID_TOPICAL/MISC/OTHER	Indicates if the patient received this medication and route combination within the time period
130	VASOPRESSOR_ET/INTRA-PLACES	Indicates if the patient received this medication and route combination within the time period

131	VASOPRESSOR_EYES	Indicates if the patient received this medication and route combination within the time period
132	VASOPRESSOR_INHALED	Indicates if the patient received this medication and route combination within the time period
133	VASOPRESSOR_INTRATHECAL	Indicates if the patient received this medication and route combination within the time period
134	VASOPRESSOR_IV/IM/IRRIGATION/EPIDURAL	Indicates if the patient received this medication and route combination within the time period
135	VASOPRESSOR_NOSE	Indicates if the patient received this medication and route combination within the time period
136	VASOPRESSOR_ORAL/ORAL TUBES	Indicates if the patient received this medication and route combination within the time period
137	VASOPRESSOR_OTHER	Indicates if the patient received this medication and route combination within the time period
138	VASOPRESSOR_SUBQ	Indicates if the patient received this medication and route combination within the time period
139	VASOPRESSOR_TOPICAL/MISC/OTHER	Indicates if the patient received this medication and route combination within the time period

BIBLIOGRAPHY

Alderden, J., Pepper, G. A., Wilson, A., Whitney, J. D., Richardson, S., Butcher, R., Jo, Y., & Cummins, M. R. (2018) Predicting pressure injury in critical care patients: A machine-learning model. American Journal of Critical Care, 27, 461-468. doi:https://doi.org/10.4037/ajcc2018525

Braden, B., & Bergstrom, N. (1987). A Conceptual Schema for the Study of the Etiology of Pressure Sores. Rehabilitation Nursing: The Official Journal of the Association of Rehabilitation Nurses, 12(1), 8–16.

Brugman, S. (2021). *Pandas Profiling.* pandas profiling. https://pandas-profiling.github.io/pandas-profiling/docs/master/rtd/.

Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 15*. doi: 10.1145/2783258.2788613

Cheng, C.-W., & Wang, M. D. (2015). Improving Personalized Clinical Risk Prediction Based on Causality-Based Association Rules. ACM-BCB ... ... : The ACM Conference on Bioinformatics, Computational Biology and Biomedicine, 2015, 386–392.

Cox, J. (2017). Pressure Injury Risk Factors in Adult Critical Care Patients: A Review of the Literature. Ostomy/wound Management, 63(11), 30–43.

Cox, J., & Roche, S. (2015). Vasopressors and development of pressure ulcers in adult critical care patients. American Journal of Critical Care: An Official Publication, American Association of Critical-Care Nurses, 24(6), 501–510.

Cox, J. (2012). Predictive Power of the Braden Scale for Pressure Sore Risk in Adult Critical Care Patients. Journal of Wound, Ostomy and Continence Nursing, 39(6), 613–621. doi: 10.1097/WON.0b013e31826a4d83.

Curley, M. A. Q., Hasbani, N. R., Quigley, S. M., Stellar, J. J., Pasek, T. A., Shelley, S. S., Kulik, L. A., Chamblee, T. B., Dilloway, M. A., Caillouette, C. N., McCabe, M. A., & Wypij, D. (2018). Predicting pressure injury risk in pediatric patients: The Braden QD scale. *The Journal of Pediatrics*, 192. https://doi.org/10.1016/j.jpeds.2017.09.045

Deo R. C. (2015). Machine Learning in Medicine. *Circulation*, *132*(20), 1920–1930. https://doi.org/10.1161/CIRCULATIONAHA.115.001593

Darvall, J. N., Mesfin, L., & Gorelik, A. (2018). Increasing frequency of critically ill patient turns is associated with a reduction in pressure injuries. Critical Care and Resuscitation: Journal of the Australasian Academy of Critical Care Medicine, 20(3), 217–222.

Edger, M. (2017). Effect of a Patient-Repositioning Device in an Intensive Care Unit On Hospital-Acquired Pressure Injury Occurrences and Cost: A Before-After Study. Journal of Wound, Ostomy, and Continence Nursing: Official Publication of the Wound, Ostomy and Continence Nurses Society / WOCN, 44(3), 236–240.

Hastie, T., Friedman, J. H., Jerome H., & Tibshirani, R. (2009). *The elements of statistical learning : data mining, inference, and prediction.* Springer.

Hyun, S., Vermillion, B., Newton, C., & Fall, M. (2013). *PREDICTIVE VALIDITY OF THE BRADEN SCALE FOR PATIENTS IN INTENSIVE CARE UNITS* American Association of Critical-Care Nurses. doi:10.4037/ajcc2013991

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An introduction to statistical learning: with applications in R*. New York: Springer.

Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine Learning and Data Mining Methods in Diabetes Research. Computational and Structural Biotechnology Journal, 15, 104–116.

Lou, Y., Caruana, R., & Gehrke, J. (2012). Intelligible models for classification and regression. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 12*, 150–158. doi: 10.1145/2339530.2339556

Lou, Y., Caruana, R., Gehrke, J., & Hooker, G. (2013). Accurate intelligible models with pairwise interactions. *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 13*, 623–631. doi: 10.1145/2487575.2487579

Nahar, J., Imam, T., Tickle, K. S., & Chen, Y.-P. P. (2013). Association rule mining to detect factors which contribute to heart disease in males and females. Expert Systems with Applications, 40(4), 1086–1093.

Nori, H., Jenkins, S., Koch, P., & Caruana, R. (2019). InterpretML: A Unified Framework for Machine Learning Interpretability. *ArXiv, abs/1909.09223*.

Padula, W. V., Makic, M. B. F., Wald, H. L., Campbell, J. D., Nair, K. V., Mishra, M. K., & Valuck, R. J. (2015). Hospital-Acquired Pressure Ulcers at Academic Medical Centers in the United States, 2008-2012: Tracking Changes Since the CMS Nonpayment Policy. Joint Commission Journal on Quality and Patient Safety / Joint Commission Resources, 41(6), 257–263.

Parreco, J. P., Hidalgo, A. E., Badilla, A. D., Ilyas, O., & Rattan, R. (2018). Predicting central line-associated bloodstream infections and mortality using supervised machine learning. *Journal of critical care*, *45*, 156–162. https://doi.org/10.1016/j.jcrc.2018.02.010

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(12), 2825–2830. Retrieved 2021, from http://jmlr.org/papers/v12/pedregosa11a.html.

Perktold, J., Seabold, S., & Taylor, J. (2019). *Statsmodels.stats.outliers_influence.variance_inflation_factor. statsmodels*. https://www.statsmodels.org/devel/generated/statsmodels.stats.outliers_influence.variance_inflation_factor.html.

Zaratkiewicz, S., Whitney, J. D., Lowe, J. R., Taylor, S., O'Donnell, F., & Minton-Foltz, P. (2010). Development and implementation of a hospital-acquired pressure ulcer incidence tracking system and algorithm. Journal for Healthcare Quality: Official Publication of the National Association for Healthcare Quality, 32(6), 44–51.

Tescher, A. N. , Branda, M. E. , Byrne, T. O. & Naessens, J. M. (2012). All At-Risk Patients Are Not Created Equal. Journal of Wound, Ostomy and Continence Nursing, 39(3), 282–291. doi: 10.1097/WON.0b013e3182435715.

Futoma, J., Morris, J., & Lucas, J. (2015). A comparison of models for predicting early hospital readmissions. *Journal of biomedical informatics*, *56*, 229–238. https://doi.org/10.1016/j.jbi.2015.05.016

Tran, T., Luo, W., Phung, D., Gupta, S., Rana, S., Kennedy, R. L., . . . Venkatesh, S. (2014). A framework for feature extraction from hospital medical data with applications in risk prediction. *BMC Bioinformatics, 15* doi:http://dx.doi.org.libproxy.lib.unc.edu/10.1186/s12859-014-0425-8

Wildemuth, B. M. (Ed.). (2017). Applications of social research methods to questions in information and library science. Libraries Unlimited, an imprint of ABC-CLIO, LLC.