5-2022

Forecasting hypotension by learning from multivariate mixed responses..

Jodie Ritter
University of Louisville

FORECASTING HYPOTENSION BY LEARNING FROM MULTIVARIATE MIXED RESPONSES

By: Jodie Ritter

B.S, University of Central Florida, 1996

A Thesis

Submitted to the Faculty of the

J.B. Speed School of Engineering of the University of

Louisville in Partial Fulfillment of the Requirements

for the Degree of

Master of Science

in Industrial Engineering

Department of Industrial Engineering

University of Louisville

Louisville, Kentucky

Louisville, KY

May 2022

FORECASTING HYPOTENSION BY LEARNING FROM MULTIVARIATE MIXED RESPONSES

Submitted by: _____
Jodie Ritter

A Thesis Approved on

4/28/2022
_____
(Date)

by the Following Thesis Committee:

_____
Lihui Bai

_____
Xiaoyu Chen

_____
Jiapeng Huang

_____
Hui Zhang

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Bai. Not only my committee chair but also my academic advisor, she has given me opportunities to expand my knowledge and grow academically. Without her continued support, guidance and encouragement this thesis would not have been possible.

I would also like to thank Dr. Chen who has pushed me to grow in the area of machine learning, writing and presentations. His expertise in data science was vital to my success. His commitment to my growth and future success is sincerely appreciated.

Additionally, I would like to thank Dr. Huang and Dr. Zhang. Dr. Huang entrusted me with this research and was available to answer my endless questions regarding anesthesiology and patient care. Dr. Zhang's expertise in data visualization was helpful in this thesis.

Finally, I would like to thank my husband and my children for their understanding and support throughout this process.

ABSTRACT

FORECASTING HYPOTENSION BY LEARNING FROM MULTIVARIATE MIXED RESPONSES

Jodie Ritter

April 28, 2022

Blood Pressure is the main determinant of blood flow to organs. Hypotension is defined as a systolic blood pressure less than 90 mmHg or a diastolic blood pressure less than 50 mmHg. The severity and duration of hypotension is associated with low blood flow to organs often result in organ damage and a high mortality rate. Predicting hypotension prior to surgery and during the surgery can reduce the incidence and duration resulting in better patient outcomes. This thesis uses preoperative bloodwork and vital signs as well as perioperative vital signs in 5-minute increments as inputs to forecast hypotension. Hypotension can be represented by multivariate mixed responses which follows both continuous and binary distributions. The main focus of this thesis is to apply a new method known as an "Interpretable Neural Network" (INN) to this clinical predictive application by simultaneously modeling mixed hypotension responses considering human domain knowledge. The customized INN method was developed and tested with a dataset containing 588 hysterectomy surgeries. It was benchmarked against other models including an Artificial Neural Network (ANN), logistic regression, k-nearest neighbors, support vector classifier, stochastic gradient descent, decision tree, random forest and

extra trees. The results suggest while the ANN classification model had the best test accuracy overall, the customized INN model produced better test accuracy with the mixed response.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

Perioperative hypotension is associated with adverse outcomes in patients undergoing surgery (Lin et al, 2011). Hypotension during noncardiac surgery can cause postoperative complications such as renal insufficiency, myocardial injury, and increased mortality. Predicting hypotension prior to the episode and taking preventative measures early can be crucial in improving patient outcomes.

Currently, management of perioperative hypotension is reactive (Wijnberge et al., 2020). Many factors contribute to perioperative hypotension such as patient comorbidities, medications taken, and medications used for induction of anesthesia (Kendale et al., 2018). Additionally, hypotension during surgery is preceded by subtle hemodynamic changes. These changes are challenging to detect because the cardiovascular system is interdependent, has complicated networks, and is influenced by compensatory mechanisms (Saugel et al., 2019). In a study by Lin et al., when anesthesiologists used the prevalent methods in practice to predict perioperative hypotensive episodes, they scored an average accuracy of 51.6%. Due to the complexity of prediction and importance in improving patient outcomes, artificial intelligence has been an area of interest in recent years because it can incorporate large amounts of data and develop robust predictive analytics (Wijnberge et al., 2020).

Artificial intelligence has been applied in many solutions in medicine to date. The early detection of atrial fibrillation was developed in 2014 and available commercially on Apple watches beginning with the Apple watch 4 in 2019. Particularly, neural networks have been used to predict outcomes in gastrointestinal bleeding, survival of esophageal cancer, inflammatory bowel disease, and metastasis in colorectal cancer and esophageal squamous cell carcinoma. There has also been success with artificial intelligence in medical image diagnosis where the accuracy matches that of radiologists (Briganti and Le Moine, 2020).

This thesis aims to use preoperative bloodwork and vital signs as well as perioperative vital signs as inputs to forecast hypotension. In particular, the perioperative medical data is collected at 5-minute interval. This research contributes to the literature of applying machine learning to hypotension prediction in two ways. First, we not only consider binary response (hypotension or non-hypotension), but also mixed response consisting of three output variables. Second, we explore a new machine learning method known as the "interpretable neural network" (INN) in predicting perioperative hypotension, thus integrating anesthesiologists' expert opinion into the prediction.

The research utilized a dataset consisting of medical records for 1,463 hysterectomy patients at University of Louisville Health from June 2018 to June 2021. The customized INN method is developed and tested with a dataset containing 588 hysterectomy surgeries. It Additionally, the performance of the proposed INN model is benchmarked against other commonly used machine learning models including Artificial Neural Network (ANN), logistic regression, k-nearest neighbors, support vector

classifier, stochastic gradient descent, decision tree, random forest and extra trees. Overall, the ANN classification model had the best test accuracy, and the customized INN model yields better test accuracy with the mixed response.

The remainder of this thesis is organized as follows. Section II reviews literature on predictive models for perioperative hypotension. Section III presents exploratory data analysis and ANOVA results for the 1,463 medical records. Section IV discusses the architecture of the INN, as well as the proposed techniques to handle missing values, intraoperative data incorporation and risk group clustering. Section V describes parameter settings and computation results for classification and a mixed response Finally, Section VI concludes the thesis with future research directions.

II. LITERATURE REVIEW

The literature review will introduce works related to predicting intraoperative hypotension in the healthcare field. It will focus on two main streams: perioperative hypotension prediction using machine learning and novel approaches in machine learning.

A. Perioperative Hypotension Prediction Using Machine Learning

Predicting hypotension using pre- and intraoperative data is a young field and the literature on this is rather scant. One closely related work by Kang et al. (2020) conducted a binary classification of hypotension using preoperative and intraoperative data as we do. Intraoperative data was recorded from induction to incision and Naïve Bayes, logistic regression, random forest, and ANN models were used to predict hypotension. The random forest model performed the best with an area under the receiver operating characteristic(AUROC) curve of 0.84 (Kang et al., 2020). In a similar stream, Kendale et al. used preoperative data as well as intraoperative vital signs from anesthesia induction to 10 minutes post induction for the classification of hypotension using similar machine learning models as Kang et al. The random forest model again performed the best with an AUROC of 0.74 (Kendale et al., 2018). Unlike the previous two studies, Hatib et al. used arterial blood pressure waveform data in a logistic regression model for the classification of hypotension. This model was successful in

predicting intraoperative hypotension 15 minutes before it occurred with a sensitivity of 88% and specificity of 87% (Hatib et al., 2018).

B. Novel Approaches in Machine Learning

In the second stream of literature review, we focus on novel approaches in machine learning. Specifically, we restrict our review on those works that address missing data imputation, clustering, interpretability of machine learning models and mixed responses.

A common challenge in healthcare analytics is to acknowledge and deal with missing data that occurs due to either clinicians' simply not collecting them, a monitoring device connection problem or not working, or the random glitches in the electronic health record system. Many approaches to impute missing data include zero filling, means filling, k-nearest neighbor filling and Expectation-Maximization (EM) filling. A study by Hegde et al. (2019) simulated missing data in healthcare records and compared Principal Component Analysis (PCA) to Multiple Imputation for Chained Equations (MICE) for imputation of healthcare data. PCA implements feature reduction and the EM algorithm to fill missing data. According to Hegde et al., PCA outperformed MICE in overall missing value imputation accuracy and root mean square error. Regarding imputation of missing data not at random (MNAR) commonly seen in healthcare data, a study Le and Tan by compared imputation algorithms and concluded that the use of more information of the same medical context improves the imputation of missing values (Le & Tan, 2018). PCA has also been used with Gaussian Mixture Models (GMM) to improve machine learning performance. A study by Guo and Chen used PCA with GMM in a HVAC fault diagnosis model using a Bayesian network which reduced computation time and improved accuracy compared to using GMM alone (Guo & Chen,

2020). In this same stream, PCA with GMM was used in population stratification for ancestry estimation and demonstrated superior performance in less time compared to GMM, k-means, and k-means with PCA (Budiarto et al., 2021).

Artificial neural network (ANN) is arguably one of most implemented machine learning methods in research as well as in practice. One commonly known challenge of neural networks is the lack of insight into how decisions are made or the "black box" effect. In an effort to address this, Chen et al. developed an "Interpretable Neural Network" (INN) to provide transparency in the neural network. The INN takes rules established using human domain knowledge and optimizes their thresholds for better prediction performance. It is customizable via its architecture and can accommodate any number of input variables, number of output variables and number of rules. The advantages of the INN are that it preserves current domain knowledge and rules, optimizes current rule thresholds and takes advantage of existing rules and data driven methods providing high accuracy with interpretability. (Chen et al., 2022).

Using a mixed response in machine learning is currently being studied. However, modeling a mixed response is prevalent in the field of statistics. A study by Kang et al. jointly modeled a binary and continuous response. Prediction accuracy was measured by the root mean square error of the continuous response and the misclassification error of the binary response. The results of simulated data were that the joint predictions were more accurate than if modeled separately (Kang et al. 2018). Another study by Hwang and Pennell demonstrated better fit modeling a joint response where the binary response is correlated with the continuous response. In other words, the continuous variable underlies the binary variable (Hwang and Pennell, 2013).

III. EXPLORATORY DATA ANALYSIS

Data was collected for 1,463 hysterectomy patients at University of Louisville Health between September 2018 and June 2021. The variables collected are shown in Appendix A. Of these patients, 642 patients or 44% of the total patient population experienced intraoperative hypotension. This data was manually recorded using patient records and any systolic blood pressure reading less than 90 or diastolic blood pressure reading less than 50 defined hypotension during surgery. Descriptive statistics of this patient population are shown in Table 1 below.

Table 1: Statistics of 1,463 Patients

Variable	N	Mean	SE Mean	St Dev	Min	Q1	Median	Q3	Max
Age	1463	50.05	0.34	12.81	19	41	48	58	97
ASA Class	1463	2.56	0.02	0.58	1	2	3	3	4
Pre Temp	1423	98.03	0.01	0.54	95.4	97.7	98.1	98.4	100.7
Pre SBP	1458	131.14	0.47	17.96	11	118	130	144	195
Pre DBP	1458	74.73	0.32	12.08	39	67	74	83	117
Pre HR	1442	78.24	0.36	13.52	18	69	77	86	138
Pre SpO2	1457	98.51	0.08	2.96	10	98	99	100	100
Pre RR	1438	17.23	0.11	4.21	4	15	17	19	82
WBC	1393	7.79	0.08	2.86	2.6	5.8	7.3	9.2	28.9
Hgb	1393	12.76	0.05	1.8	4.9	11.8	13	14	17.6
Platelet	1392	277.53	2.28	85.16	47	223	268	319	981
Sodium	1116	137.48	0.08	2.68	99	136	138	139	151
Potassium	1118	3.82	0.02	0.73	2.5	3.6	3.8	4	25
Chloride	1116	103.96	0.1	3.24	87	102	104	106	115
Pre CO2	1116	23.7	0.07	2.35	16	22	24	25	32
Anion Gap	1115	9.85	0.07	2.21	1	8	10	11	18
Glucose	1118	110.03	0.99	33.12	68	92	101	115	347
BUN	1116	10.91	0.21	6.94	1	8	10	13	147
Creatinine	1116	0.82	0.01	0.45	0.31	0.68	0.77	0.87	8.78
eGFR Af	1114	19.17	0.71	23.7	1.9	10.08	13.2	17.33	248

eGFR nonAf	1113	98.62	0.85	28.21	6	83	97	113	241
BUN/Creatinine	1113	77.73	0.84	28.16	3.3	66	79	92	199
Calcium	1116	9.03	0.01	0.47	6.6	8.7	9	9.3	11.4
Protein	880	6.91	0.02	0.6	4.3	6.6	7	7.3	8.7
Albumin	881	3.72	0.02	0.48	1.4	3.5	3.8	4	5.2
Globulin	880	3.2	0.02	0.62	2	3	3	4	6
A/G	881	1.21	0.01	0.3	0.4	1	1.2	1.4	2.3
Bilirubin	880	0.72	0.01	0.33	0.1	0.5	0.7	0.88	2.9
Alkaline Phosphatase	882	71.6	1	29.61	23	54	67	83	469
AST	880	25.63	0.72	21.33	9	18	22	27	478
ALT	881	23.33	0.54	16.08	4	15	20	26	180

Of the binary variables, the number of occurrences and percentage of total patient population is shown in Table 2 below.

Table 2: Number of Patients with Hypertension and Types of Heart Pathologies

Variable	**Number of Patients**	**% Total Population**
Hypertension	633	43%
Taking Anti-Hypertensive	445	30%
Dyslipidemia	250	17%
ACEI	196	13%
BB	195	13%
ARBs	109	7%
Dysrhythmia	73	5%
CAD	45	3%
Emergency	30	2%
CHF	27	2%
Abnormal EKG	22	2%
Valvular Disease	12	1%
Hypotension	7	< 1%
Peripheral Vascular Disease	3	< 1%
Syncope	1	< 1%

As shown, 43% of the patient population had hypertension or high blood pressure, 30% were taking medication to control hypertension and 17% had dyslipidemia. Patients with heart pathologies took three different medications: 13% angiotensin-converting enzyme

inhibitors (ACEI), 13% beta blockers (BB) and 7% took angiotensin II receptor blockers (ARB). In order to identify significant factors for predicting hypotension, logistic regression was used for the binary response and the Analysis of Variance (ANOVA) from it is shown in Table 3 and the odds ratio for continuous predictors in Table 4 below.

Table 3: ANOVA for 1,463 Patients

		Wald Test	
Source	**DF**	**Chi-Square**	**P-Value**
Regression	48	73.01	0.011
Age	1	5.18	**0.023**
Pre Temp	1	0.12	0.734
Pre SBP	1	0.24	0.626
Pre DBP	1	11.97	**0.001**
Pre HR	1	7.06	**0.008**
Pre SpO2	1	1.49	0.223
Pre RR	1	2.79	0.095
WBC	1	0.75	0.387
Hgb	1	3.49	0.062
Platelet	1	0.22	0.640
Sodium	1	0.00	0.980
Potassium	1	0.76	0.384
Chloride	1	0.00	0.980
Pre CO2	1	0.00	0.982
Anion Gap	1	0.00	0.970
Glucose	1	0.48	0.490
BUN	1	0.23	0.630
Creatinine	1	0.13	0.714
eGFR Af	1	0.02	0.895
eGFR nonAf	1	0.10	0.756
BUN/Creatinine	1	0.00	0.986
Calcium	1	0.44	0.506
Protein	1	1.30	0.254
Albumin	1	2.72	0.099
Globulin	1	1.03	0.309
A/G	1	0.07	0.795
Billirubin	1	0.01	0.935
Alk Phos	1	1.02	0.313
AST	1	0.33	0.568
ALT	1	0.05	0.817
ASA Class	3	4.11	0.250
Emergency	1	0.00	0.959
Taking Anti-hypertensive	1	2.18	0.140
ACEI	1	0.04	0.843
ARBs	1	0.87	0.350
BB	1	0.35	0.552
Hypertension	1	4.56	**0.033**

CHF	1	3.87	**0.049**
Valvular Disease	1	0.33	0.563
Dysrythmia	1	0.29	0.587
Hypotension	1	0.43	0.512
CAD	1	1.48	0.224
Dyslipidemia	1	0.30	0.581
Peripheral Vascular Disease	1	0.00	0.954
Syncope	1	0.00	0.969
Abnormal EKG	1	0.58	0.446

Table 4: Odds Ratio for Continuous Predictors

	Odds Ratio	**95% CI**
Age	1.0188	(1.0026, 1.0352)
Pre Temp	0.9526	(0.7199, 1.2604)
Pre SBP	0.9974	(0.9870, 1.0079)
Pre DBP	0.9744	(0.9603, 0.9888)
Pre HR	1.0170	(1.0044, 1.0298)
Pre SpO2	0.9498	(0.8745, 1.0317)
Pre RR	1.0322	(0.9945, 1.0712)
WBC	1.0274	(0.9664, 1.0924)
Hgb	0.9030	(0.8114, 1.0051)
Platelet	0.9995	(0.9976, 1.0015)
Sodium	1.2227	(0.0000, 9.96892E+06)
Potassium	0.8205	(0.5257, 1.2806)
Chloride	0.8121	(0.0000, 6.62277E+06)
Pre CO2	0.8356	(0.0000, 6.81352E+06)
Anion Gap	0.7398	(0.0000, 6.03206E+06)
Glucose	0.9982	(0.9933, 1.0032)
BUN	1.0123	(0.9631, 1.0640)
Creatinine	1.1050	(0.6480, 1.8844)
eGFR Af	1.0031	(0.9580, 1.0503)
eGFR nonAf	1.0101	(0.9481, 1.0760)
BUN/Creatinine	0.9994	(0.9293, 1.0747)
Calcium	0.8602	(0.5521, 1.3403)
Protein	0.6792	(0.3496, 1.3196)
Albumin	2.2148	(0.8606, 5.7002)
Globulin	1.4250	(0.7200, 2.8201)
A/G	1.2228	(0.2682, 5.5744)
Billirubin	0.9802	(0.6054, 1.5869)
Alk Phos	1.0031	(0.9971, 1.0091)
AST	0.9966	(0.9852, 1.0082)
ALT	0.9983	(0.9838, 1.0129)

From the ANOVA age, preoperative diastolic blood pressure, preoperative heart rate, hypertension and congestive heart failure were significant factors with a 95% confidence interval. The odds ratios indicate the odds of hypotension increases by the odds ratio for each unit increase in the variables. Odds ratios close to 1 minimally affect hypotension. For example, for each unit increase in albumin, the odds of hypotension go up 2.21 times.

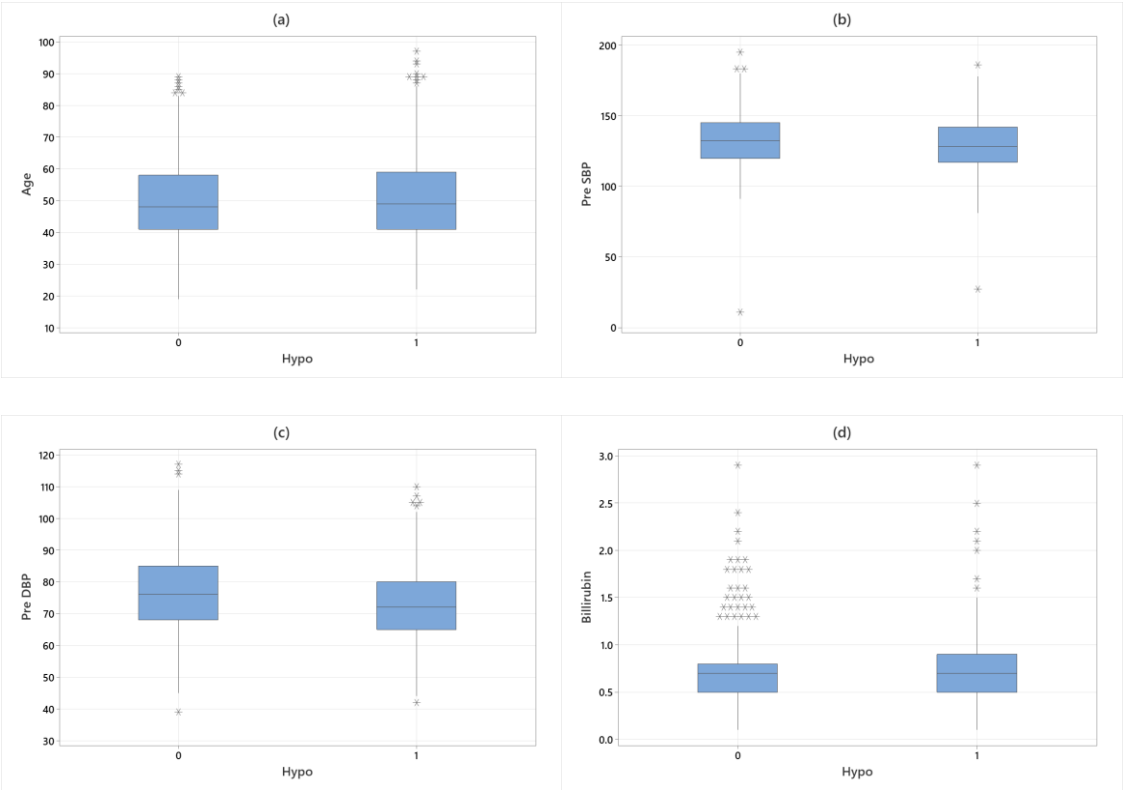


Figure 1: Boxplots of Hypotension vs. (a) Age, (b) Pre SBP, (c) Pre DBP, and (d) Bilirubin

Boxplots are shown above in Figure 1 for hypotension and the selected input variables. The median age is slightly higher for the hypotensive class and the preoperative SBP and DBP are slightly lower. Bilirubin has a larger spread in the hypotensive class. Overall, it is shown that it is difficult to classify hypotension because

the variables have the same attributes and overlap with the hypotensive and non-hypotensive classes.

Scatterplots and distributions were also analyzed using Python and included in Appendix A.

From the scatterplots and distributions, there was no separation of hypotensive and non-hypotensive patients for any of the variables. This indicates that these variables alone are not enough to predict hypotension. Additionally, a correlation matrix indicated no factors even moderately correlated with hypotension. Therefore, intraoperative data had to be combined with the preoperative data.

IV. METHODOLOGY

Motivated by the practical challenges of missing values in covariates, heterogeneous data distributions, and mixed responses which follow different distributions, this thesis proposes a machine learning pipeline, following the definition of the AdaPipe System (Chen & Jin, 2020), to provide interpretable and accurate one-step ahead forecasting for perioperative hypotension. In this section, I will first provide an overview for the pipeline and assumptions, then introduce each component of the pipeline in details.

A. Overview

The proposed pipeline efficiently and accurately predicts a mixed response using a neural network with interpretable decisions using large, multivariate datasets with random and blocks of missing values. The pipeline assumes that: heterogeneous underlying distributions can be reduced to unimodal normal distributions with the GMM; hidden associations exist among mixed responses; and missing values can be replaced by predictions derived by the observable portion of the dataset.

The pipeline in Figure 2 below, processes input data with random or blocks of missing values to predict a mixed response with interpretability.

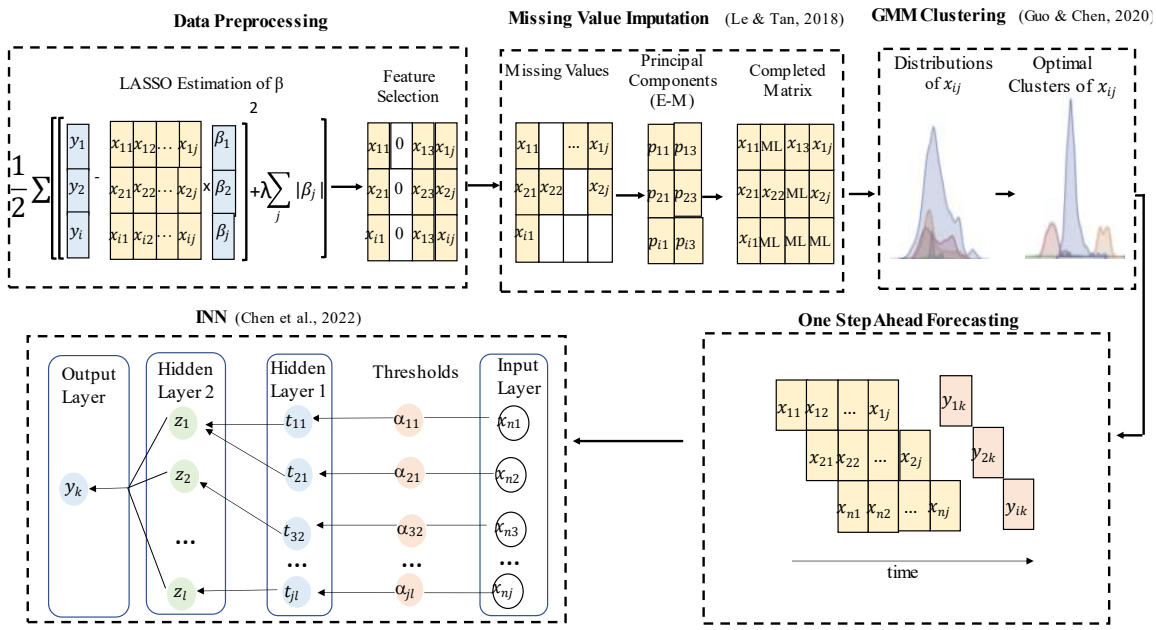


Figure 2: Overview of Pipeline to Predict Mixed Response with Interpretability

B. Pipeline Details

First the notations are defined and summarized in Table 5.

Table 5: Summary of Notations

Notations	Definitions
x_{ij}	Data at time i for feature j
β_j	Coefficient for feature j
y_i	Actual value for y at time i
λ	Tuning parameter for L1 regularization
p_{ij}	Principal component data at time i for feature j

ML	Maximum Likelihood Estimate
y_{ik}	Response variable k for time i
x_{nj}	Summarized statistics where n = time i-1 minus a window size for feature j
α_{jl}	Threshold for feature j for rule l
t_{jl}	Rule based conclusion where t_{jl}=1 if $x_j > \alpha_{jl}$
z_l	Conclusions for rule l
y_k	Prediction for each response variable k
k	Binary hypotension, systolic blood pressure and diastolic blood pressure

i. Data Preprocessing

To model the relationship between the input data and the response variable, a linear regression model was used to reveal their correlation. To estimate this model, each blood pressure reading was considered one sample. While our feature size did not exceed the sample size, the proposed pipeline was developed to accommodate large datasets where the feature size is greater than the sample size. Therefore, a Least Absolution Shrinkage and Selection Operator (LASSO) variable selection method was used to identify a smaller set of predictors (Chen & Jin 2020).

The data for the linear regression model is represented by X which is a matrix of time i and j features. The response variable Y represents the systolic and diastolic blood pressure readings for time i. The relationship between X and Y can be modeled as $Y=X\beta+\varepsilon$ where β is the model coefficients and ε is the model error which is independently an identically distributed and

follows a normal distribution. The LASSO estimation of the model coefficients is formulated as follows (Chen & Jin 2020),

$$\beta=\text{argmin}_{\beta}\{\frac{1}{2}\sum_i(y_i - \beta_0 - \sum_j x_{ij}\,\beta_j)^2 + \lambda\sum_j|\beta_j|\} \quad (1)$$

where λ is a tuning parameter. In this thesis, λ was selected by LASSO cross validation on the training data. The linear regression model was applied to the intraoperative features in the prediction of binary and mixed responses.

ii. Missing Value Imputation

Since missing values permeated the data randomly and in blocks, filling these values with interpolation or other methods like means filling could not be used. Missing data occurs randomly in electronic healthcare records due to monitoring devices that become disconnected or stop working. They also occur in blocks due to tests not ordered (i.e., a group of bloodwork not ordered because a patient is low risk). To impute these missing values, PCA was used because it has demonstrated best imputation for these latent variables (Dray & Josse, 2015).

The data, x_{ij} is reduced to p_{ij} by reducing the Euclidean distance between the original data points and the estimated data using a set number of components. The Expectation–Maximization (EM) algorithm is then used to iteratively calculate the Maximum Likelihood Estimates (MLE) of missing values.

Because the INN cannot accept null values, this method was employed. The number of components was determined to explain 95% of the variance and standardization was done prior to using PCA to fill the missing values.

iii. Gaussian Mixture Model Clustering

Using large, multivariate datasets with the INN can yield large computational time. Since imputing missing values with PCA followed by Gaussian Mixture Model (GMM) for clustering can improve accuracy and computation time, GMM was used to identify the optimal number of clusters (Guo & Chen, 2020).

GMM clustering can identify multiple distributions within a dataset and assign a probability of each sample belonging to each cluster. The GMM algorithm uses the EM algorithm iteratively to determine the best mean and variance for a specified number of clusters. The minimum Bayesian Information Criterion (BIC) determines the optimal number of clusters (Wang & Liu, 2006).

The GMM was applied to the dataset output from PCA. The BIC and BIC gradient versus the number of clusters were graphed. The optimal number of clusters was chosen for the minimum BIC and before the BIC gradient upward trend levels off.

iv. One Step Ahead Forecasting

To make predictions on data where measurements do not occur at regular time intervals, a one step ahead forecasting model was used. Multivariate irregular time series data cause random missing values in data where imputation is not appropriate. A common technique used in prediction models is one step ahead forecasting which forecasts the next time period from previous feature values (Kantardzic, 2011).

Summarized statistics are represented by x_{nj} and used to predict y_{ik} where n is time of i-1 minus a window size or lag of previous readings and k is the response. In this way a set of features is used to predict the next time period, t(i+1).

The one step ahead forecasting approach was applied to the intraoperative data which had irregular time series for all features. The window size was set to 6 representing approximately 30 minutes. This summary statistics used were min, max, mean, mode, median, range, standard deviation, variance and entropy. Additionally, two variables to measure the change in response from the previous reading were used.

v. INN

As stated in the literature review, neural networks are widely used in machine learning but do not offer insight into how decisions are made. The INN created by Chen et al. (2022) addresses this lack of transparency. It takes rules established using human domain knowledge and optimizes their thresholds for better prediction performance. The INN provides these optimized thresholds in an Excel file (Chen et al., 2022).

The input data for the INN is represented by x_{nj} where *n* is the summarized statistics for time i-1 through the window size and *j* is the feature. Each feature *j* is associated with rule *l* in the architecture. Each rule *l* has a threshold associated with feature *j* represented by α_{jl}. From the input layer, hidden layer 1 makes a rule based conclusion, t_{jl}, which is 1 if $x_j > \alpha_{jl}$. From there, hidden layer 2 has combination logic that states if $x_1 > \alpha_{11}$ AND $x_2 > \alpha_{21}$, then z_l is 1. In other words, if both input variables associated with a rule are greater than the corresponding thresholds, then the rule conclusion in hidden layer 2 is 1. The response variable in the output layer is then determined to be 1 if any z_l is 1. For example, if z_1 =1 OR z_2 =1, then the response variable would be 1. Hidden layer 1initializes α_{jl} while hidden layer 2 optimizes α_{jl}.

The INN was customized for the number of input variables, rules, activation functions and rule assignments. The INN was used to predict a mixed response with each input variable assigned to a rule.

V. COMPUTATIONAL RESULTS

A case study for perioperative hypotension forecasting was adopted to evaluate the performance of the proposed machine learning pipeline by comparing with several well-known benchmark models. In this section, the computational results will be introduced following the sequence of the pipeline.

A. Pipeline Implementation Details

i. Input Data

The intraoperative data added to the existing preoperative data included oxygen saturation, pulse rate, heart rate, respiratory rate and temperature readings at 5-minute intervals throughout surgery. Due to limitations of the data available, 588 of the 1,463 patients were used for the remainder of this research. Of these patients, 499 out of 7,233 readings, or 7%, met the criteria for hypotension.

Within the intraoperative data, there were inaccurate readings caused by the arterial line or non-invasive line not connected to patients, the arterial line or non-invasive line was working before and now it is not working, the patient is having severe hypotension or cardiac arrest or artifacts. These inaccurate readings were removed from the data per the consulting anesthesiologist and included systolic arterial blood pressure less than 30, difference between systolic and diastolic arterial blood pressure in the same measurement less than 15mmHg, respiratory rate values less than 5 or more than 40, pulse rate less than 30 or more than 150, and temperature values less than 34 degrees Celsius. All values of oxygen saturation were kept.

ii. Data Preprocessing

A cross validation technique (Kirkland et al., 2015) was used in Python to optimize the tuning parameter, λ, of the LASSO model. And then the best λ associated with the lowest cross validation error was used to estimate the LASSO model coefficients. The top twenty features were selected using 5-fold cross validation in order of coefficient magnitude and are shown in Table 6 below. These features were used in all testing and represent the intraoperative data.

Table 6: Top 20 Features Selected by Coefficient Magnitude

Feature	Coefficients
DBP Delta	7.28
Temperature_ mean	-1.64
Pulse Rate _max	1.60
Respiratory Rate_entropy	-0.64
Temperature _var	-0.59
Pulse Rate _var	-0.58
Pulse Rate _entropy	0.58
Temperature_entropy	0.56
SBP Delta	0.52
Respiratory Rate_stdev	0.48
Temperature_range	0.47
Temperature_median	-0.32
Pulse Rate _min	0.28
Respiratory Rate_mode	0.26
SPO2_max	-0.19
Temperature_min	-0.13
SPO2_min	-0.11
Respiratory Rate_max	0.11
SPO2_mode	0.09
Temperature_stdev	0.07

iii. Missing Value Imputation

In order for PCA to impute the missing values, the number of features must be determined first. This thesis used a threshold of 95% to determine the number of features

necessary to explain the variance. Figure 3 below plots the number of features versus the cumulative variance. As shown, the number of components or features that explains 95% of the variance was 34.

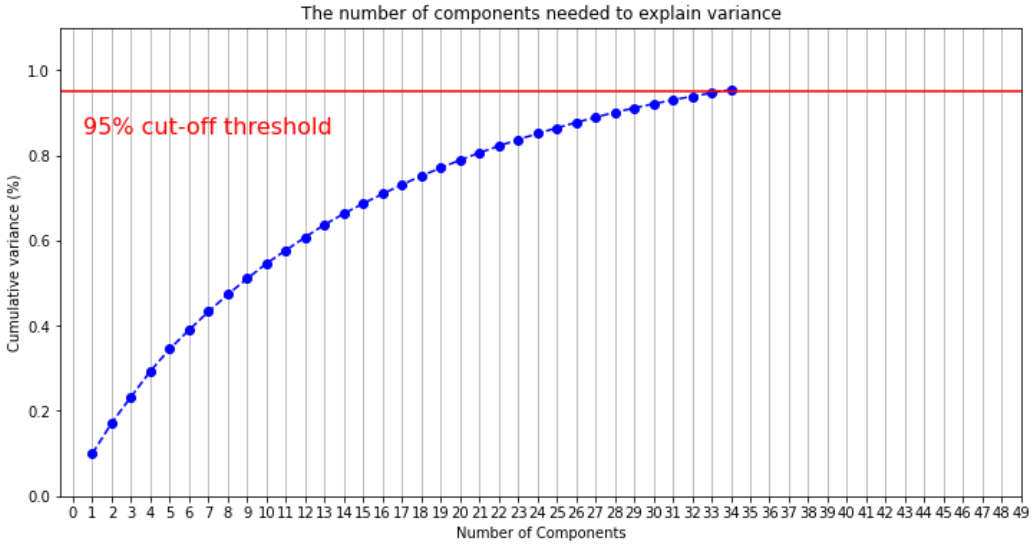


Figure 3: Number of Components to Explain 95% of the Variance

Within the preoperative data, there were distinct groups of patients where bloodwork was not collected. This is primarily due to their ASA Class but could be due to other factors as well. Three distinct groups were identified by uncollected data and labeled as low, medium and high risk. For low risk patients, all preoperative vitals were taken along with white blood count, hemoglobin and platelets. The medium risk patients had all the information from the low risk group but had additional bloodwork of sodium, potassium, chloride, carbon dioxide, anion gap, glucose, BUN, creatinine, eGFR Af, eGFR non-Af, BUN to creatinine ratio and calcium. The high risk group had everything the medium risk group had along with protein, albumin, globulin, albumin to globulin ratio, bilirubin, alkaline phosphatase, AST and ALT. All of the coagulation bloodwork,

magnesium, phosphorus, glucose POC2 and CRP were eliminated due to high numbers of missing values. Within these groups, missing values occurred randomly. Each risk group had the same proportion of hypotensive patients within each group at 7-8%. Figure 4 below shows the percentage of total hypotensive patients for each risk group.

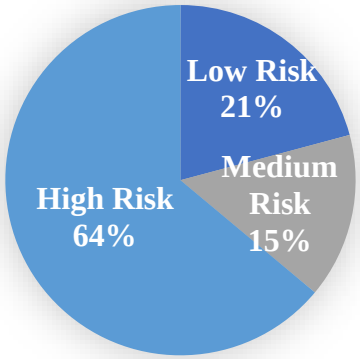


Figure 4: Hypotension by Risk Group

This graph shows that the High Risk group has the highest percentage of hypotensive readings followed by the Low Risk group and the Medium Risk group.

To account for the groups of missing values by risk in preparation for Principal Component Analysis (PCA), two indicator variables were added to the data. The low risk group patients would have zeroes for both indicator variables, the medium risk group patients had a one for the first indicator variable and a zero for the second and the high risk patients would have ones for both variables. This allowed the algorithm to fill the values differently depending on the indicator variable values.

iv. Gaussian Mixture Model Clustering

Using this data set, the Gaussian Mixture Model was used to determine the number of clusters to use, the Bayesian Information Criterion (BIC) and BIC gradient

were graphed versus the number of clusters ranging from 2 to 12 shown below in Figure 6a and 6b.

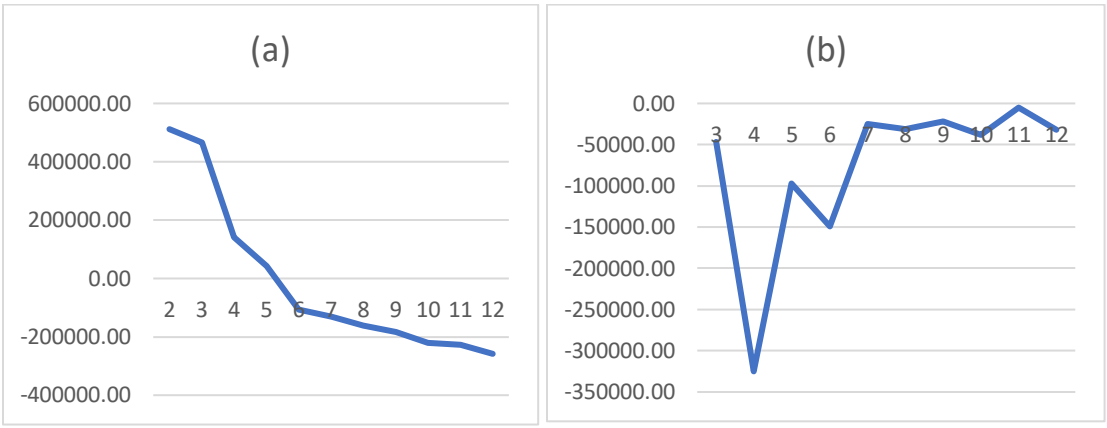


Figure 5: Number of Clusters vs. (a) BIC and (b) BIC Gradient

In Figure 5a, a BIC closest to one indicates it is the best model. In evaluating the gradient in Figure 5b, the optimal number of clusters is right before the upward trend levels off. In both graphs, the optimal number of clusters was five.

The separation of clusters can be visualized in Figure 6 comparing preoperative bilirubin and preoperative heart rate.

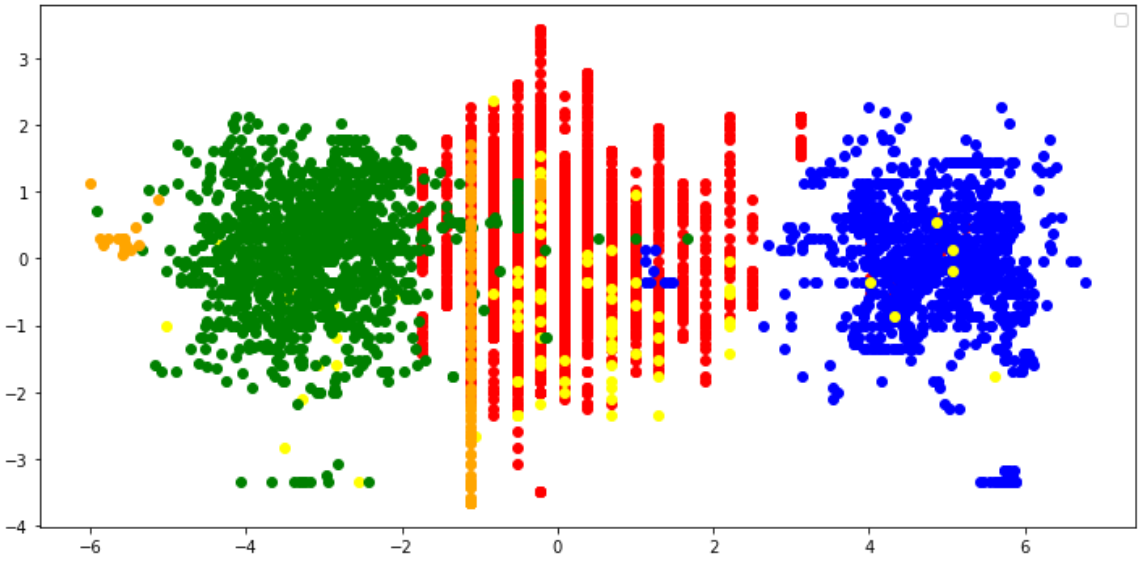


Figure 6: Preoperative Bilirubin vs. Preoperative Heart Rate. Cluster 1: Red; Cluster 2: Blue; Cluster 3: Yellow; Cluster 4: Green, and Cluster 5: Orange.

The five clusters had the best separation among variables in reviewing the scatterplots for clusters. Comparing the Clusters to the risk groups, Clusters 1, 3 and 5 were equivalent to the high risk Group. Cluster two was equivalent to the medium risk group. Cluster four made up the low risk group with a little bit of Clusters 5 and 3. So, the main difference was that the high risk group was split into 3 clusters. The basic statistics of risk groups and clusters are located in Appendix B.

v. One Step Ahead Forecasting

One step ahead forecasting was then used to consolidate the data by the response variable. In a typical anesthesia record, there are missing values that occur before, during and after surgery as show in Table 7 below.

Table 7: Sample Intraoperative Data

SPO2	Temperature - Celsius	Heart Rate	Pulse Rate	Respiratory Rate	Systolic	Diastolic	Mean
100		110	111	12			
97		104	109	12			
89		92	82	15	107	58	
99		110	104	15			
98	36	78	78	17			
96	35.9	87	87	15	127	81	
99	35.8	106	107	17	157	90	
99	35.9	84	84	17			
97	35.9	79	78	17	112	62	
99	36	79	79	17	123	75	
98	36	74	74	17	119	63	
97	36	72	73	17	119	67	
98	36.1	67	67	17			
98	36.1	71	71	17	113	58	
97	36.1	71	71	17			

This occurs in surgery due to the timing of when the arterial or non-invasive line was connected or disconnected. In order to preserve this data format without artificially filling it, one step ahead forecasting was used. Figure 7 below illustrates the technique.

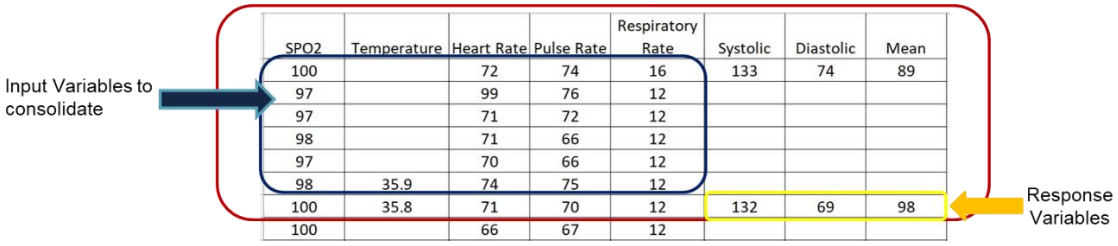


Figure 7: One Step Ahead Forecasting

This consolidation started with the first blood pressure reading after the first six input variable readings and continued through the last blood pressure reading. If there was a blood pressure reading in the first six rows, it was disregarded. A total of 47 input variables were created for each blood pressure reading, although heart rate and pulse rate were duplicitous and heart rate variables would later be removed. The blood pressure readings represent approximately 30 minutes, although not all the readings were every 5 minutes. There were some irregular time interval readings where the intervals could be

10 minutes. The benefit of this approach is that all anesthesia records can be consolidated as is and the window size can be modified as well as the statistics generated.

B. INN Classification

Only 7% of the patient population had hypotensive readings which can lead to misleading accuracy. If the model predicted all patient readings as non-hypotensive, the accuracy would still be 93%. Therefore, stratified k-fold and Synthetic Minority Oversampling Technique (SMOTE) were used to balance the dataset. Stratified k-fold splits the data based to preserve the percentage of samples for each class. Within each fold, SMOTE was applied to the training data. Using the risk groups and clusters, a binary response variable was tested using an ANN and the INN for the purpose of classification using stratified 5-fold cross validation. Both the ANN and INN had one input layer, two hidden layers and one output layer. The ANN was fully connected between the input layer and the hidden layers with sigmoid activation functions.

The INN architecture was modified to incorporate the number of input variables and rule assignments. Each rule was assigned no more than four input variables. The risk groups are shown in Tables 8 through 10 below.

Table 8: Low Risk Group Rules for the INN

Rule	Variables	Number of Variables
1	Age	1
2 & 3	Preop Vitals	6
4	Bloodwork - WBC up to Hgb	3
5	Oxygen Saturation – min, max, mode	3
6 & 7	Temperature – min, mean, median, range, standard deviation, variance, entropy	7
8	Pulse Rate – min, max, variance, entropy	4
9	Respiratory Rate - max, mode, standard deviation, entropy	4
10	SBP and DBP Delta	2

Table 9: Medium Risk Group Rules for the INN

Rule	Variables	Number of Variables
1	Age	1
2 & 3	Preop Vitals	6
4 to 7	Bloodwork - WBC up to Calcium	15
8	Oxygen Saturation – min, max, mode	3
9 & 10	Temperature – min, mean, median, range, standard deviation, variance, entropy	7
11	Pulse Rate – min, max, variance, entropy	4
12	Respiratory Rate - max, mode, standard deviation, entropy	4
13	SBP and DBP Delta	2

Table 10: High Risk Group Rules for the INN

Rule	Variables	Number of Variables
1	Age	1
2 & 3	Preop Vitals	6
4 to 9	Bloodwork - WBC through ALT	23
10	Oxygen Saturation – min, max, mode	3
11 & 12	Temperature – min, mean, median, range, standard deviation, variance, entropy	7
13	Pulse Rate – min, max, variance, entropy	4
14	Respiratory Rate - max, mode, standard deviation, entropy	4
15	SBP and DBP Delta	2

The low risk group had 30 input variables and 10 rules, the medium risk group had 42 input variables and 13 rules, and the high risk group had 50 input variables and 15 rules. The five clusters used the same rules as the high risk group since each cluster had 50 input variables. Any missing values were filled with the mean for each variable for each risk group.

The input and response variables were read from files and then the data was split using stratified kfold. After splitting the data into training and test sets, SMOTE was applied to the training data and then scaled using the standard scaler. Prior to fitting the model, the AdamWarmup optimizer was used for a warmup set to 0.01 and a decay set to 0.0001. Total steps and warmup steps for the optimizer was done using a warmup proportion of 0.1 and 1,000 epochs. The batch size was set to 10% of the number of samples in each group or cluster. The model for the neural networks were compiled using this optimizer with a loss set to binary cross entropy and metrics set to accuracy. The model was then fit to the training data with 3,000 epochs. The testing accuracy was averaged over the five folds as well as the standard error and the confusion matrix was displayed. In the case of the INN, the threshold values for the variables within each rule were written to an Excel file.

The results of testing the ANN and the INN are shown in Table 11 below.

Table 11: Classification Accuracy by Risk Group and Cluster

Neural Network	Cluster	Number of Rows	Avg Test Accuracy	Avg Std Error	False Positives	False Negatives
ANN	Low Risk	1,253	85.1%	0.01	11%	4%
	Medium Risk	1,096	95.2%	0.01	3%	1%
	High Risk	4,884	90.2%	0.01	7%	3%
	1	4,620	**89.5%**	0.004	10%	1%
	2	1,090	**99.2%**	0.01	0.4%	0.5%
	3	96	**99.0%**	0.01	0%	1%
	4	1,264	**97.3%**	0.01	1%	1%
	5	163	**97.0%**	0.03	2%	1%
INN	Low Risk	1,253	80.1%	0.02	17%	3%
	Medium Risk	1,096	93.0%	0.01	6%	1%

	High Risk	4,884	80.2%	0.01	17%	3%
	1	4,620	78.1%	0.02	19%	3%
	2	1,090	92.9%	0.01	5%	2%
	3	96	93.8%	0.01	2%	4%
	4	1,264	78.8%	0.01	17%	4%
	5	163	95.7%	0.02	2%	2%

The clusters show better performance than the risk groups and the ANN shows better performance than the INN in both the clusters and groups. The false positives and false negatives were also lower with the ANN, even though cluster 1 had a high percentage of false positives.

For comparison purposes, classifiers were run with default parameter settings for all five clusters. The accuracy results in Table 12 show that Random Forest and Extra Trees have the highest accuracy.

Table 12: Classifier Accuracy by Cluster

		Logistic Regression	KNN	SVC	SGD	Decision Tree	Random Forest	Extra Trees
Cluster 1	Train	78.8%	96.0%	97.8%	77.4%	98.3%	99.6%	99.7%
	Test	74.5%	86.3%	91.6%	72.1%	91.0%	**94.6%**	94.2%
Cluster 2	Train	94.0%	97.4%	98.8%	91.8%	99.1%	99.8%	99.9%
	Test	88.6%	91.1%	95.2%	89.7%	93.7%	96.4%	**96.5%**
Cluster 3	Train	79.6%	74.1%	79.3%	76.8%	78.3%	78.7%	79.7%
	Test	73.3%	59.2%	73.3%	73.3%	71.7%	**76.7%**	**76.7%**
Cluster 4	Train	80.0%	94.0%	97.0%	74.8%	97.7%	99.7%	99.7%
	Test	76.7%	80.5%	87.3%	73.7%	86.6%	91.8%	**91.9%**
Cluster 5	Train	99.6%	98.5%	98.7%	99.3%	99.9%	99.3%	99.2%
	Test	96.3%	93.9%	95.7%	97.5%	**97.0%**	95.7%	95.7%

In comparison classifiers to the ANN, the ANN performs the same or better among the clusters except in cluster 1 where Random Forest has better accuracy.

C. INN Mixed Response

A mixed response variable was tested for improved accuracy for both the ANN and INN. Given that the clusters performed better than the risk groups, only the clusters were tested for the mixed response.

The output layer was customized to predict 3 response variables or a mixed response. The mixed response consists of a binary response of hypotension as well as the continuous response variables for systolic and diastolic blood pressure. The INN activation function was changed in hidden layer 2 to Rectified Linear Unit (ReLU) from sigmoid. In addition to the activation and output changes, the loss function was changed to include a binary cross entropy loss for the binary response variable and the mean squared error loss for the continuous response variables. This new loss function had weighting to account for lack of binary hypotensive readings as represented by Equation 2 below.

$$\text{Loss} = \text{binary cross entropy loss} + \text{weight}*(\text{MSE for SBP and DBP}) \qquad (2)$$

The weight was set to 100 for testing. The results are shown in Table 13 below.

Table 13: Cluster Accuracy with Mixed Response

Neural Network	Cluster	Number of Rows	Avg Test Accuracy	SBP Avg RMSE	DBP Avg RMSE
ANN	1	4,620	96.2%	3.00	2.20
	2	1,090	97.0%	2.60	2.00
	3	96	93.8%	22.75	15.75
	4	1,264	94.1%	7.60	6.80
	5	163	96.3%	3.40	1.60
INN	1	4,620	93.8%	16.40	11.80
	2	1,090	94.9%	13.60	9.60

	3	96	92.7%	25.80	19.40
	4	1,264	91.6%	15.20	12.20
	5	163	93.9%	20.00	9.40

The mixed response resulted in improved accuracy for the INN and ANN compared to their classification accuracies. The ANN still performed better than the INN, however, the gap in accuracy was much smaller. The root mean square error is smaller than the INN in all clusters, although cluster 3 had the largest for both systolic and diastolic blood pressure.

In comparing the responses of the clusters, Table 14 shows that the ANN has higher accuracy with classification and the INN has higher accuracy with the mixed response.

Table 14: Cluster Accuracy by Response

Neural Network	Cluster	Binary Response Avg Test Accuracy	Mixed Response Avg Test Accuracy
ANN	1	89.5%	**96.2%**
	2	**99.2%**	97.0%
	3	**99.0%**	93.8%
	4	**97.3%**	94.1%
	5	**97.0%**	96.3%
INN	1	78.1%	**93.8%**
	2	92.9%	**94.9%**
	3	**93.8%**	92.7%
	4	78.8%	**91.6%**
	5	**95.7%**	93.9%

The ANN had the best prediction accuracy overall with classification. Comparing this to regression with decision tree, random forest and extra trees, Table 15 shows the R^2 value

for training and testing as well as the root mean squared error for systolic and diastolic pressure.

Table 15: Fit of Regressors by Cluster

		Decision Tree	Random Forest	Extra Trees
Cluster 1	Train R2	100.0%	97.5%	100.0%
	Test R2	62.9%	81.7%	85.1%
	Avg SBP RMSE	12.6	8.8	8.2
	Avg DBP RMSE	9.8	7.0	6.2
Cluster 2	Train R2	100.0%	97.5%	100.0%
	Test R2	63.4%	82.3%	87.8%
	Avg SBP RMSE	11.4	8.0	6.6
	Avg DBP RMSE	9.6	6.6	5.6
Cluster 3	Train R2	100.0%	90.7%	100.0%
	Test R2	1.9%	39.9%	42.2%
	Avg SBP RMSE	21.0	16.0	15.4
	Avg DBP RMSE	14.6	12.4	11.8
Cluster 4	Train R2	100.0%	95.9%	100.0%
	Test R2	42.3%	69.7%	76.0%
	Avg SBP RMSE	14.0	10.2	9.0
	Avg DBP RMSE	12.4	8.8	8.0
Cluster 5	Train R2	100.0%	98.9%	100.0%
	Test R2	89.9%	93.2%	95.0%
	Avg SBP RMSE	10.2	8.8	7.0
	Avg DBP RMSE	6.2	5.0	4.4

Extra trees had the best testing fit for all clusters except for 3. In addition to superior classification accuracy, the ANN had better RMSE values than any of these regressors.

Even though the INN did not perform as well as the ANN, the accuracy for the mixed response was still high. The main difference between the architecture of the ANN and INN is the number of nodes in the hidden layers: the ANN has 50 nodes while the

INN has 15 nodes. Therefore, the mapping of input variables to rules in the INN contributed to the difference in accuracy.

The INN offers the advantage of interpretability which is worth the tradeoff in accuracy. The optimized threshold values by cluster from the INN are listed in Appendix C. The interpretation with Rule 1 for Cluster 1 is age is greater than 48, then the patient will go hypotensive. For Rule 10 and Cluster 1, the logic would be if SBP Delta is greater than 26.71 and DBP Delta is less than 9.1 or if SBP Delta is less than 26.71 and DBP Delta is greater than 9.11 then the patient will go hypotensive. These thresholds can help anesthesiologists refine their current logic to predict which patients will experience hypotension in surgery.

VI. CONCLUSIONS AND FUTURE RESEARCH

This thesis proposed a pipeline for data with missing values and heterogeneous distributions of input variables for the prediction of a mixed response that provides insight into how decisions were made. It includes the use of LASSO for feature selection, PCA for feature reduction and missing value imputation along with GMM for clustering, one step ahead forecasting and the INN with a mixed response. The ANN had the highest accuracy in classification and in the mixed response as compared the INN. Comparing the response of each neural network, the ANN performed the best overall in classification whereas the INN performed the best with the mixed response.

For the purposes of future research, the INN should be tested with the mixed response and classification using a fully connected architecture. Each input variable would comprise one rule and the optimized thresholds would be consistent with the logic that anesthesiologists use today.

In addition to changing the rule mapping for the INN, acquiring intraoperative data in one minute intervals would lead to better training and accuracy. It would also provide a better prediction window in the first 15 minutes of surgery that would be beneficial to anesthesiologists. It would also provide more insight into the subtle hemodynamic changes that precede hypotension given the additional readings.

Finally, a validation dataset, unseen to the neural networks, would provide a better test of their accuracy. The neural networks in this study were trained and tested on

five different splits of the data for cross validation. With only 7,233 readings, it was best to train the models with the most data. A validation dataset would better reflect the accuracy of both the ANN and the INN.

REFERENCES

Briganti, G., & Le Moine, O. (2020). Artificial intelligence in medicine: today and tomorrow. Frontiers in medicine, 7, 27.

Budiarto, A., Mahesworo, B., Hidayat, A. A., Nurlaila, I., & Pardamean, B. (2021). Gaussian Mixture Model Implementation for Population Stratification Estimation from Genomics Data. *Procedia Computer Science*, *179*, 202-210.

Chen, X., & Jin, R. (2020). Adapipe: A recommender system for adaptive computation pipelines in cyber-manufacturing computation services. *IEEE Transactions on Industrial Informatics*, *17*(9), 6221-6229.

Chen, X., Zeng, Y., Kang, S., & Jin, R. (2022). INN: An Interpretable Neural Network for AI Incubation in Manufacturing. *ACM Transactions on Intelligent Systems and Technology (TIST)*.

Dray, S., & Josse, J. (2015). Principal component analysis with missing values: a comparative survey of methods. *Plant Ecology*, *216*(5), 657-667.

Guo, Y., & Chen, H. (2020). Fault diagnosis of VRF air-conditioning system based on improved Gaussian mixture model with PCA approach. *International journal of refrigeration*, *118*, 1-11.

Hatib, F., Jian, Z., Buddi, S., Lee, C., Settels, J., Sibert, K., ... & Cannesson, M. (2018). Machine-learning algorithm to predict hypotension based on high-fidelity arterial pressure waveform analysis. Anesthesiology, 129(4), 663-674.

Hegde, H., Shimpi, N., Panny, A., Glurich, I., Christie, P., & Acharya, A. (2019). MICE vs PPCA: Missing data imputation in healthcare. Informatics in Medicine Unlocked, 17, 100275.

Hwang, B. S., & Pennell, M. L. (2014). Semiparametric Bayesian joint modeling of a binary and continuous outcome with applications in toxicological risk assessment. Statistics in medicine, 33(7), 1162-1175.

Johnson, D. (2020). Artificial Neural Network Tutorial with TensorFlow ANN Examples. Www.guru99.com. https://www.guru99.com/artificial-neural-network-tutorial.html?msclkid=4736ed90b5d911ec94b628384a7eee67

Kang, A. R., Lee, J., Jung, W., Lee, M., Park, S. Y., Woo, J., & Kim, S. H. (2020). Development of a prediction model for hypotension after induction of anesthesia using machine learning. PloS one, 15(4), e0231172.

Kang, L., Kang, X., Deng, X., & Jin, R. (2018). A Bayesian hierarchical model for quantitative and qualitative responses. Journal of Quality Technology, 50(3), 290-308.

Kantardzic, M. (2011). *Data mining : concepts, models, methods, and algorithms* (Second). IEEE Press.

Kendale, S., Kulkarni, P., Rosenberg, A. D., & Wang, J. (2018). Supervised machine-learning predictive analytics for prediction of postinduction hypotension. Anesthesiology, 129(4), 675-688.

Kirkland, L. A., Kanfer, F., & Millard, S. (2015, January). LASSO tuning parameter selection. In *Annual Proceedings of the South African Statistical Association Conference* (Vol. 2015, No. con-1, pp. 49-56). South African Statistical Association (SASA).

Le, T. D., Beuran, R., & Tan, Y. (2018, November). Comparison of the most influential missing data imputation algorithms for healthcare. In 2018 10th International Conference on Knowledge and Systems Engineering (KSE) (pp. 247-251). IEEE.

Lin, C. S., Chang, C. C., Chiu, J. S., Lee, Y. W., Lin, J. A., Mok, M. S., ... & Li, Y. C. (2011). Application of an artificial neural network to predict postinduction hypotension during general anesthesia. Medical Decision Making, 31(2), 308-314.

Saugel, B., Kouz, K., Hoppe, P., Maheshwari, K., & Scheeren, T. W. (2019). Predicting hypotension in perioperative and intensive care medicine. Best Practice & Research Clinical Anaesthesiology, 33(2), 189-197.

Wang, Y., & Liu, Q. (2006). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of stock–recruitment relationships. *Fisheries Research*, *77*(2), 220-225.

Wijnberge, M., Geerts, B. F., Hol, L., Lemmers, N., Mulder, M. P., Berge, P., ... & Veelo, D. P. (2020). Effect of a machine learning–derived early warning system for intraoperative hypotension vs standard care on depth and duration of intraoperative hypotension during elective noncardiac surgery: the HYPE randomized clinical trial. Jama, 323(11), 1052-1060.

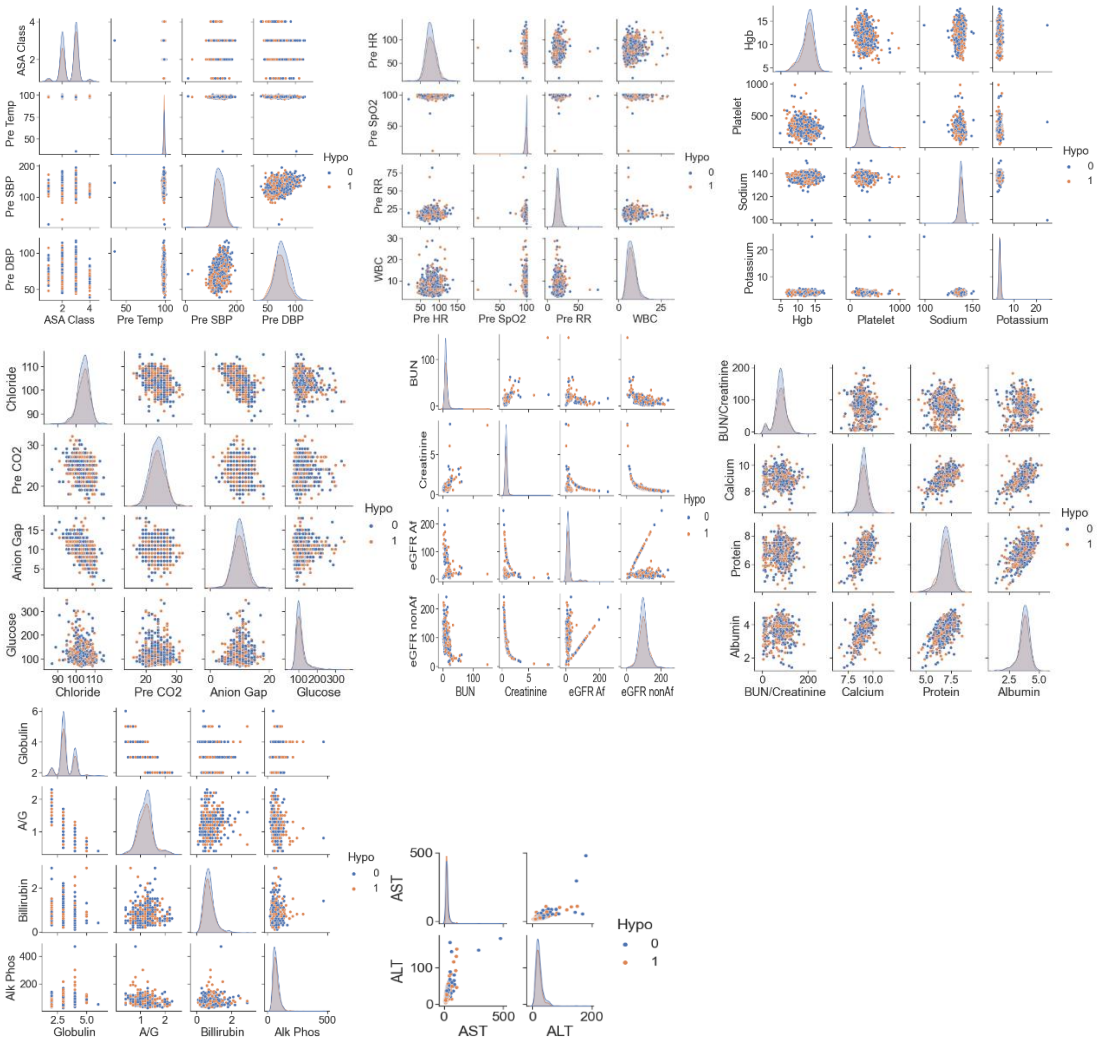
APPENDICES

Appendix A: Variable List for 1,463 Patients

	Variable	Type of Variable
Patient Information	MRN	String
	Name	String
	Age	Continuous
	Gender	Categorical
	Admission Date	Date
	Discharge Date	Date
	Surgery Date	Date
	Surgery Start	Datetime
	Surgery End	Datetime
	ASA Class	Categorical
	Taking Anti-hypertensive	Binary
	Hypertensive Class	Binary
	ACEI	Binary
	ARBs	Binary
	BB	Binary
	Heart pathologies	Binary
	Abnormal EKG	Binary
Preoperative Vitals and Bloodwork	Temp	Continuous
	Systolic Blood Pressure	Continuous
	Diastolic Blood Pressure	Continuous
	Heart Rate	Continuous
	Oxygen Sat. (SpO2)	Continuous
	Respiratory Rate	Continuous
	White Blood Count (WBC)	Continuous
	Hemoglobin	Continuous
	Platelet	Continuous
	Coag PT	Continuous
	Coag INR	Continuous
	Coag PITT	Continuous
	Coag Fib. Level	Continuous

	Sodium	Continuous
	Potassium	Continuous
	Chloride	Continuous
	Pre CO2	Continuous
	Anion Gap	Continuous
	Glucose	Continuous
	BUN	Continuous
	Creatinine	Continuous
	eGFR Af	Continuous
	eGFR nonAf	Continuous
	BUN/Creatinine	Continuous
	Calcium	Continuous
	Protein	Continuous
	Albumin	Continuous
	Globulin	Continuous
	A/G	Continuous
	Bilirubin	Continuous
	Alk Phos	Continuous
	AST	Continuous
	ALT	Continuous
	Magnesium	Continuous
	Phosphorus	Continuous
	Glucose POC2	Continuous
	CRP	Continuous
Response	Hypotension	Binary

Appendix B: Scatterplots and Distribution for 1,463 Patients



Appendix C: Risk Group and Cluster Statistics

Low Risk Statistics

Variable	N	Mean	SE Mean	St Dev	Min	Q1	Median	Q3	Max
Age	1253	48.45	0.35	12.31	27.00	39.00	46.00	59.00	75.00
ASA Class	1253	2.34	0.02	0.62	1.00	2.00	2.00	3.00	4.00
Pre Temp	1253	97.98	0.02	0.73	95.40	97.60	98.00	98.50	99.60
Pre SBP	1253	129.67	0.48	17.06	81.00	117.00	128.00	141.00	174.00
Pre DBP	1253	76.62	0.32	11.35	48.00	68.00	76.00	84.00	110.00
Pre HR	1253	79.46	0.37	13.19	42.00	68.00	77.00	90.00	110.00
Pre SpO2	1253	99.05	0.03	1.23	95.00	98.00	100.00	100.00	100.00
Pre RR	1253	18.35	0.23	8.07	8.00	14.00	18.00	20.00	82.00
WBC	1253	6.88	0.05	1.68	3.40	5.80	6.88	7.20	15.10
Hgb	1253	12.58	0.04	1.31	8.10	12.00	12.58	13.40	15.60
Platelet	1253	262.86	1.79	63.48	99.00	227.00	262.86	288.00	530.00
SPO2_min	1253	97.71	0.06	2.29	85.00	97.00	98.00	99.00	100.00
SPO2_max	1253	99.13	0.03	1.20	94.00	99.00	100.00	100.00	100.00
SPO2_mode	1253	98.60	0.04	1.59	89.00	98.00	99.00	100.00	100.00
Temperature_min	1253	36.17	0.02	0.59	34.00	35.80	36.20	36.60	37.50
Temperature_mean	1253	36.27	0.02	0.56	34.20	35.90	36.32	36.65	37.50
Temperature_median	1253	36.27	0.02	0.56	34.20	35.90	36.30	36.65	37.50
Temperature_range	1253	0.18	0.01	0.19	0.00	0.10	0.10	0.20	2.00
Temperature_stdev	1253	0.07	0.00	0.08	0.00	0.04	0.05	0.10	0.90
Temperature_var	1253	0.01	0.00	0.04	0.00	0.00	0.00	0.01	0.82
Temperature_entropy	1253	0.84	0.01	0.50	0.00	0.69	0.69	1.10	1.79
Pulse Rate _min	1253	72.01	0.39	13.83	41.00	61.00	71.00	81.00	131.00
Pulse Rate _max	1253	84.47	0.46	16.27	54.00	73.00	83.00	95.00	146.00
Pulse Rate _var	1253	36.17	1.85	65.38	0.00	4.57	13.10	37.37	691.47
Pulse Rate _entropy	1253	1.53	0.01	0.27	0.00	1.39	1.61	1.78	1.79
Respiratory Rate_max	1253	14.56	0.10	3.45	7.00	12.00	14.00	16.00	32.00
Respiratory Rate_mode	1253	13.47	0.07	2.57	5.00	12.00	13.00	15.00	26.00
Respiratory Rate_stdev	1253	0.83	0.04	1.32	0.00	0.00	0.41	1.10	8.73
Respiratory Rate_entropy	1253	0.45	0.01	0.50	0.00	0.00	0.59	0.69	1.78
SBP Delta	1253	2.89	0.54	19.17	-55.00	-8.00	2.00	14.00	182.00
DBP Delta	1253	0.74	0.46	16.38	-58.00	-8.50	0.00	9.00	192.00
Systolic	1253	119.19	0.55	19.57	70.00	105.00	118.00	131.00	307.00
Diastolic	1253	69.73	0.42	14.96	22.00	60.00	68.00	78.00	278.00
Mean	1249	89.34	0.45	15.98	38.00	79.00	88.00	98.00	293.00
Hypo	1253	0.08	0.01	0.28	0.00	0.00	0.00	0.00	1.00

Medium Risk

Variable	N	Mean	SE Mean	St Dev	Min	Q1	Median	Q3	Max
Age	1096	50.28	0.35	11.64	27.00	40.00	49.00	56.00	86.00
ASA Class	1096	2.67	0.02	0.55	2.00	2.00	3.00	3.00	4.00
Pre Temp	1096	97.74	0.02	0.57	96.30	97.40	97.70	98.30	98.70
Pre SBP	1096	134.06	0.65	21.43	95.00	114.00	135.00	151.00	173.00
Pre DBP	1096	75.47	0.39	12.97	47.00	66.25	74.00	85.00	105.00
Pre HR	1096	76.56	0.37	12.18	47.00	69.00	73.00	81.00	113.00
Pre SpO2	1096	98.29	0.07	2.19	80.00	97.00	99.00	100.00	100.00
Pre RR	1096	16.81	0.09	2.98	10.00	14.00	17.00	19.00	24.00
WBC	1096	7.75	0.07	2.34	4.10	6.10	7.40	9.10	16.00
Hgb	1096	12.65	0.06	1.96	7.20	11.50	12.60	13.90	17.00
Platelet	1096	287.50	2.21	73.03	115.00	262.00	288.00	319.00	565.00
Sodium	1096	136.89	0.07	2.40	131.00	136.00	137.00	138.00	142.00
Potassium	1096	3.79	0.01	0.36	2.50	3.60	3.80	4.10	4.40
Chloride	1096	103.20	0.10	3.25	95.00	102.00	104.00	105.00	109.00
Pre CO2	1096	23.21	0.07	2.17	18.00	22.00	23.00	25.00	29.00
Anion Gap	1096	10.43	0.06	2.11	6.00	9.00	10.00	12.00	15.00
Glucose	1096	112.01	1.64	54.42	74.00	88.00	96.00	113.00	347.00
BUN	1096	13.84	0.29	9.52	4.00	9.00	12.00	14.00	56.00
Creatinine	1096	0.89	0.01	0.39	0.31	0.70	0.83	0.95	3.17
eGFR Af	1096	41.80	1.23	40.81	6.30	12.50	18.50	86.75	248.00
eGFR nonAf	1096	88.68	0.88	29.12	20.00	72.00	86.00	103.00	213.00
BUN/Creatinine	1096	57.19	1.06	35.19	5.80	16.90	61.00	85.00	176.00
Calcium	1096	9.04	0.02	0.51	8.00	8.80	9.00	9.30	11.40
SPO2_min	1096	97.59	0.07	2.41	83.00	96.00	98.00	100.00	100.00
SPO2_max	1096	98.86	0.05	1.58	93.00	98.00	100.00	100.00	100.00
SPO2_mode	1096	98.37	0.06	1.94	92.00	97.00	99.00	100.00	100.00
Temperature_min	1096	36.07	0.02	0.62	34.20	35.80	36.10	36.50	37.50
Temperature_mean	1096	36.14	0.02	0.61	34.20	35.85	36.19	36.53	37.58
Temperature_median	1096	36.14	0.02	0.61	34.20	35.85	36.20	36.50	37.60
Temperature_range	1096	0.15	0.00	0.14	0.00	0.10	0.10	0.20	1.50
Temperature_stdev	1096	0.06	0.00	0.06	0.00	0.04	0.05	0.08	0.66
Temperature_var	1096	0.01	0.00	0.02	0.00	0.00	0.00	0.01	0.43
Temperature_entropy	1096	0.75	0.01	0.47	0.00	0.69	0.69	1.10	1.79
Pulse Rate _min	1096	71.67	0.41	13.47	46.00	61.00	70.00	80.00	110.00
Pulse Rate _max	1096	84.76	0.50	16.69	48.00	74.00	82.00	94.75	142.00
Pulse Rate _var	1096	42.39	2.74	90.60	0.17	4.57	15.60	40.90	1034.40
Pulse Rate _entropy	1096	1.55	0.01	0.23	0.69	1.39	1.61	1.78	1.79

Respiratory Rate_max	1096	14.28	0.11	3.66	8.00	12.00	13.00	16.00	31.00
Respiratory Rate_mode	1096	13.37	0.09	3.01	6.00	12.00	12.00	16.00	28.00
Respiratory Rate_stdev	1096	0.77	0.04	1.21	0.00	0.00	0.41	1.03	9.81
Respiratory Rate_entropy	1096	0.44	0.01	0.47	0.00	0.00	0.63	0.69	1.77
SBP Delta	1096	-0.33	0.72	23.85	-79.00	-14.00	0.00	12.00	187.00
DBP Delta	1096	-1.95	0.48	16.04	-44.00	-12.00	-2.00	7.00	79.00
Systolic	1096	123.53	0.64	21.22	77.00	110.00	121.00	133.00	310.00
Diastolic	1096	71.46	0.43	14.35	26.00	62.00	71.00	80.00	127.00
Mean	1093	91.34	0.52	17.25	44.00	81.00	90.00	100.00	302.00
Hypo	1096	0.07	0.01	0.25	0.00	0.00	0.00	0.00	1.00

High Risk

Variable	N	Mean	SE Mean	St Dev	Min	Q1	Median	Q3	Max
Age	4884	52.71	0.19	13.49	19.00	42.00	52.00	61.00	90.00
ASA Class	4884	2.71	0.01	0.61	1.00	2.00	3.00	3.00	4.00
Pre Temp	4884	98.05	0.01	0.59	96.00	97.70	98.10	98.50	99.70
Pre SBP	4884	133.81	0.24	16.52	91.00	123.00	135.00	146.00	186.00
Pre DBP	4884	75.23	0.18	12.84	39.00	67.00	75.00	85.00	114.00
Pre HR	4884	80.41	0.19	13.24	42.00	71.00	78.00	88.00	123.00
Pre SpO2	4884	98.07	0.04	2.69	70.00	97.00	98.00	100.00	100.00
Pre RR	4884	18.96	0.13	8.99	10.00	15.00	17.00	20.00	69.00
WBC	4884	8.05	0.04	3.12	2.80	6.00	7.40	9.30	26.00
Hgb	4884	12.45	0.03	1.90	7.10	11.30	12.90	13.70	17.60
Platelet	4884	279.53	1.32	91.95	105.00	212.00	269.00	324.00	806.00
Sodium	4884	137.22	0.06	4.32	99.00	136.00	138.00	139.00	143.00
Potassium	4884	3.99	0.03	1.97	2.60	3.60	3.80	4.00	25.00
Chloride	4884	103.86	0.04	3.11	93.00	102.00	104.00	106.00	114.00
Pre CO2	4884	23.03	0.03	2.11	17.00	22.00	23.00	24.00	29.00
Anion Gap	4884	10.67	0.03	2.17	4.00	9.00	10.00	12.00	18.00
Glucose	4884	109.98	0.45	31.61	69.00	92.00	102.00	116.00	301.00
BUN	4884	11.04	0.15	10.73	2.00	7.00	9.00	13.00	147.00
Creatinine	4884	0.85	0.01	0.60	0.43	0.67	0.78	0.89	8.63
eGFR Af	4884	30.12	0.54	37.75	3.30	10.30	13.90	19.60	196.00

eGFR nonAf	4884	94.28	0.39	27.35	6.00	76.00	95.00	108.00	195.00
BUN/Creatinine	4884	66.82	0.48	33.39	3.30	50.00	73.00	88.00	161.00
Calcium	4884	8.97	0.01	0.52	6.80	8.60	9.00	9.30	10.30
Protein	4884	6.86	0.01	0.66	4.30	6.60	6.90	7.30	8.30
Albumin	4884	3.61	0.01	0.56	1.50	3.40	3.70	4.00	4.90
Globulin	4884	3.26	0.01	0.67	2.00	3.00	3.00	4.00	5.00
A/G	4884	1.15	0.00	0.30	0.40	1.00	1.10	1.30	2.30
Bilirubin	4884	0.77	0.00	0.33	0.20	0.60	0.70	0.90	2.50
Alk Phos	4884	70.42	0.42	29.47	23.00	54.00	66.00	78.00	469.00
AST	4884	25.79	0.31	21.72	12.00	18.00	21.00	28.00	478.00
ALT	4884	23.07	0.23	16.16	4.00	14.00	19.00	26.00	180.00
SPO2_min	4884	97.43	0.04	2.45	61.00	96.00	98.00	99.00	100.00
SPO2_max	4884	98.93	0.02	1.50	91.00	98.00	100.00	100.00	100.00
SPO2_mode	4884	98.35	0.03	1.89	83.00	97.00	99.00	100.00	100.00
Temperature_min	4884	36.13	0.01	0.63	34.00	35.74	36.10	36.50	38.20
Temperature_mean	4884	36.22	0.01	0.61	34.00	35.85	36.23	36.60	38.28
Temperature_median	4884	36.22	0.01	0.61	34.00	35.85	36.20	36.60	38.30
Temperature_range	4884	0.18	0.00	0.20	0.00	0.10	0.10	0.20	2.20
Temperature_stdev	4884	0.07	0.00	0.08	0.00	0.04	0.05	0.09	1.08
Temperature_var	4884	0.01	0.00	0.05	0.00	0.00	0.00	0.01	1.17
Temperature_entropy	4884	0.81	0.01	0.49	0.00	0.69	0.69	1.10	1.79
Pulse Rate _min	4884	71.29	0.18	12.21	32.00	63.00	70.00	79.00	116.00
Pulse Rate _max	4884	83.91	0.20	14.05	50.00	73.00	83.00	93.00	141.00
Pulse Rate _var	4884	35.93	0.89	62.38	0.00	5.37	14.97	40.30	1032.57
Pulse Rate _entropy	4884	1.56	0.00	0.24	0.00	1.39	1.61	1.78	1.79
Respiratory Rate_max	4884	14.47	0.05	3.49	8.00	12.00	14.00	16.00	40.00
Respiratory Rate_mode	4884	13.34	0.04	2.52	5.00	12.00	12.00	14.00	37.00
Respiratory Rate_stdev	4884	0.86	0.02	1.36	0.00	0.00	0.41	1.10	13.58
Respiratory Rate_entropy	4884	0.48	0.01	0.49	0.00	0.00	0.68	0.69	1.78
SBP Delta	4884	4.71	0.39	27.38	-120.00	-9.00	2.00	20.00	219.00
DBP Delta	4884	1.97	0.26	18.00	-87.00	-8.00	1.00	12.00	89.00

Systolic	4884	126.84	0.33	23.04	55.00	111.00	125.00	140.00	318.00
Diastolic	4884	73.43	0.22	15.17	13.00	63.00	73.00	83.00	167.00
Mean	4835	94.07	0.24	16.98	26.00	82.00	93.00	105.00	224.00
Hypo	4884	0.07	0.00	0.25	0.00	0.00	0.00	0.00	1.00

Cluster Statistics

Variable	C	N	Mean	SE Mean	St Dev	Min	Q1	Median	Q3	Max
Age	1	4620	52.02	0.20	13.34	19.00	42.00	51.00	59.00	90.00
	2	1090	50.29	0.35	11.66	27.00	40.00	49.00	56.00	86.00
	3	96	55.72	1.61	15.80	30.00	42.00	53.50	71.00	86.00
	4	1264	48.95	0.35	12.29	27.00	41.00	47.00	59.00	75.00
	5	163	66.87	0.65	8.33	41.00	70.00	70.00	70.00	70.00
ASA Class	1	4620	2.67	0.01	0.57	1.00	2.00	3.00	3.00	4.00
	2	1090	2.67	0.02	0.55	2.00	2.00	3.00	3.00	4.00
	3	96	2.58	0.07	0.66	1.00	2.00	3.00	3.00	4.00
	4	1264	2.36	0.02	0.62	1.00	2.00	2.00	3.00	4.00
	5	163	3.91	0.02	0.29	3.00	4.00	4.00	4.00	4.00
Pre Temp	1	4620	98.03	0.01	0.58	96.00	97.70	98.10	98.40	99.70
	2	1090	97.75	0.02	0.58	96.30	97.40	97.70	98.30	98.70
	3	96	97.74	0.08	0.78	96.00	97.00	97.85	98.28	99.10
	4	1264	97.97	0.02	0.72	95.40	97.70	98.00	98.50	99.60
	5	163	98.73	0.03	0.43	98.20	98.60	98.60	98.60	99.70
Pre SBP	1	4620	133.75	0.25	16.70	91.00	123.00	134.00	146.00	186.00
	2	1090	134.10	0.65	21.45	95.00	114.00	135.00	151.00	173.00
	3	96	136.22	1.94	18.97	103.00	126.00	133.00	140.00	186.00
	4	1264	129.35	0.48	17.00	81.00	117.00	128.00	141.00	174.00
	5	163	136.47	0.50	6.43	123.00	140.00	140.00	140.00	140.00
Pre DBP	1	4620	75.44	0.19	12.90	39.00	67.00	74.00	85.00	114.00
	2	1090	75.52	0.39	12.98	47.00	67.00	74.00	85.00	105.00

	3	96	76.14	0.91	8.89	58.00	71.00	78.00	82.00	107.00
	4	1264	76.11	0.32	11.24	48.00	68.00	76.00	84.00	110.00
	5	163	72.26	1.07	13.71	42.00	76.00	76.00	76.00	92.00
Pre HR	1	4620	80.48	0.20	13.45	42.00	71.00	78.00	88.00	123.00
	2	1090	76.59	0.37	12.20	47.00	69.00	73.00	81.00	113.00
	3	96	74.98	0.94	9.25	42.00	72.00	74.00	78.75	105.00
	4	1264	79.98	0.38	13.34	42.00	68.00	79.00	91.00	110.00
	5	163	77.18	0.31	3.90	75.00	75.00	75.00	75.00	85.00
Pre SpO2	1	4620	98.19	0.04	2.65	70.00	97.00	98.07	100.00	100.00
	2	1090	98.29	0.07	2.19	80.00	97.00	99.00	100.00	100.00
	3	96	98.21	0.18	1.78	94.00	97.00	99.00	100.00	100.00
	4	1264	98.93	0.04	1.42	95.00	98.00	100.00	100.00	100.00
	5	163	95.29	0.18	2.32	94.00	94.00	94.00	94.00	100.00
Pre RR	1	4620	17.64	0.06	3.93	10.00	15.00	17.00	20.00	37.00
	2	1090	16.82	0.09	2.98	10.00	14.00	17.00	19.00	24.00
	3	96	17.96	0.87	8.49	10.00	14.00	16.50	18.96	69.00
	4	1264	17.56	0.11	3.96	8.00	14.00	18.00	20.00	28.00
	5	163	62.99	1.40	17.89	22.00	69.00	69.00	69.00	82.00
WBC	1	4620	8.06	0.05	3.12	2.80	6.10	7.50	9.30	26.00
	2	1090	7.75	0.07	2.34	4.10	6.10	7.40	9.10	16.00
	3	96	7.11	0.21	2.09	3.50	5.85	6.65	9.25	11.20
	4	1264	7.04	0.05	1.85	3.40	6.00	6.88	7.40	15.10
	5	163	7.13	0.25	3.15	5.80	5.80	5.80	5.80	14.50
Hgb	1	4620	12.54	0.03	1.82	7.50	11.60	12.90	13.60	17.60
	2	1090	12.66	0.06	1.95	7.20	11.50	12.60	13.90	17.00
	3	96	12.64	0.19	1.85	7.50	11.00	13.30	14.30	15.40
	4	1264	12.67	0.04	1.29	8.10	12.10	12.58	13.50	15.60
	5	163	9.00	0.06	0.82	7.10	9.30	9.30	9.30	9.70
Platelet	1	4620	277.93	1.37	93.01	105.00	211.00	265.00	325.00	806.00
	2	1090	286.96	2.21	72.83	115.00	261.50	288.00	319.00	565.00

	3	96	278.87	7.96	77.99	112.00	218.00	262.86	320.75	473.00
	4	1264	265.42	1.84	65.55	99.00	230.00	262.86	290.00	530.00
	5	163	310.61	3.74	47.69	212.00	320.00	320.00	320.00	398.00
Sodium	1	4620	137.40	0.04	2.45	124.00	136.00	138.00	139.00	143.00
	2	1090	136.89	0.07	2.41	131.00	136.00	137.00	138.00	142.00
	3	96	137.73	0.24	2.34	131.72	136.00	138.00	139.00	142.00
	4	1264	134.45	0.20	6.93	99.00	133.98	135.68	137.09	150.24
	5	163	141.98	0.10	1.24	140.00	142.00	142.00	142.00	146.57
Potassium	1	4620	3.81	0.01	0.35	2.60	3.60	3.80	4.00	5.20
	2	1090	3.79	0.01	0.36	2.50	3.60	3.80	4.10	4.40
	3	96	3.74	0.06	0.59	2.15	3.40	3.70	4.00	5.98
	4	1264	4.49	0.11	3.90	-1.11	3.19	3.92	4.59	25.00
	5	163	3.98	0.02	0.25	2.25	4.00	4.00	4.00	4.20
Chloride	1	4620	103.74	0.04	3.02	93.00	102.00	104.00	106.00	114.00
	2	1090	103.19	0.10	3.25	95.00	102.00	104.00	105.00	109.00
	3	96	104.61	0.30	2.94	98.51	102.41	104.44	107.00	110.00
	4	1264	102.11	0.05	1.70	91.52	101.57	102.32	103.08	106.64
	5	163	107.93	0.15	1.92	101.18	109.00	109.00	109.00	109.00
Pre CO2	1	4620	23.10	0.03	2.06	17.00	22.00	23.00	24.00	29.00
	2	1090	23.21	0.07	2.16	18.00	22.00	23.00	25.00	29.00
	3	96	23.03	0.28	2.70	18.00	21.00	23.00	25.00	28.00
	4	1264	21.82	0.06	2.17	16.10	20.34	21.64	23.20	37.29
	5	163	20.52	0.12	1.58	17.00	21.00	21.00	21.00	23.50
Anion Gap	1	4620	10.56	0.03	2.10	4.00	9.00	10.00	12.00	18.00
	2	1090	10.44	0.06	2.11	6.00	9.00	10.00	12.00	15.00
	3	96	10.14	0.17	1.63	6.00	9.00	10.00	11.00	14.13
	4	1264	11.75	0.06	2.05	1.49	10.47	11.85	13.09	20.33
	5	163	13.47	0.20	2.60	12.00	12.00	12.00	12.00	19.28
Glucose	1	4620	110.84	0.46	31.29	73.00	93.00	102.00	117.00	301.00

	2	1090	112.13	1.65	54.54	74.00	88.00	96.00	115.00	347.00
	3	96	111.21	3.69	36.11	27.81	88.00	99.00	118.90	241.82
	4	1264	132.52	1.82	64.56	-180.75	96.55	131.92	174.79	516.15
	5	163	69.72	5.01	63.98	-170.42	69.00	69.00	69.00	170.00
BUN	1	4620	10.11	0.06	4.32	2.00	7.00	9.00	12.00	38.00
	2	1090	13.87	0.29	9.54	4.00	9.00	12.00	14.00	56.00
	3	96	10.88	0.49	4.81	-8.73	8.00	10.00	13.00	24.00
	4	1264	6.32	0.20	7.00	-35.24	2.53	6.82	10.28	31.18
	5	163	38.24	3.64	46.47	5.30	19.00	19.00	19.00	147.00
Creatinine	1	4620	0.79	0.00	0.20	0.43	0.67	0.77	0.88	2.30
	2	1090	0.89	0.01	0.39	0.31	0.70	0.83	0.95	3.17
	3	96	0.77	0.02	0.21	0.16	0.66	0.80	0.87	1.37
	4	1264	0.60	0.01	0.32	-0.92	0.40	0.61	0.78	1.71
	5	163	2.46	0.21	2.63	0.77	1.37	1.37	1.37	8.63
eGFR Af	1	4620	29.91	0.56	37.87	3.30	10.10	13.90	19.58	196.00
	2	1090	41.90	1.24	40.84	6.30	12.50	18.55	87.00	248.00
	3	96	33.93	4.04	39.58	-111.15	11.90	17.45	54.15	170.00
	4	1264	14.75	1.45	51.63	-361.08	-8.78	17.67	43.74	177.95
	5	163	12.03	0.67	8.49	-24.27	13.90	13.90	13.90	22.27
eGFR nonAf	1	4620	96.18	0.38	25.85	26.00	78.00	96.00	110.00	195.00
	2	1090	88.67	0.88	29.16	20.00	72.00	86.00	103.00	213.00
	3	96	94.10	3.14	30.73	39.35	72.00	92.00	107.50	207.52
	4	1264	89.17	1.52	54.09	-201.98	58.10	91.82	122.51	241.57
	5	163	43.26	1.55	19.73	6.00	46.00	46.00	46.00	109.69
BUN/Creatinine	1	4620	68.48	0.48	32.91	3.30	58.00	74.00	89.00	161.00
	2	1090	57.13	1.07	35.21	5.80	16.90	61.00	85.00	176.00
	3	96	65.30	3.57	34.97	6.50	46.47	72.17	87.00	145.00
	4	1264	72.38	0.91	32.45	-60.13	54.40	75.85	94.17	162.61
	5	163	37.31	1.58	20.14	5.00	38.00	38.00	38.00	108.83

Calcium	1	4620	8.99	0.01	0.51	6.80	8.70	9.00	9.30	10.30
	2	1090	9.04	0.02	0.51	8.00	8.80	9.00	9.30	11.40
	3	96	8.98	0.05	0.45	8.00	8.80	8.90	9.30	10.56
	4	1264	8.99	0.01	0.51	7.20	8.69	8.97	9.31	11.97
	5	163	8.36	0.03	0.44	6.90	8.30	8.30	8.30	9.20
Protein	1	4620	6.89	0.01	0.65	4.30	6.60	7.00	7.30	8.30
	2	1090	7.14	0.04	1.30	4.17	6.17	7.02	7.95	11.01
	3	96	6.83	0.07	0.68	3.92	6.40	6.80	7.34	8.23
	4	1264	6.82	0.02	0.56	4.99	6.48	6.85	7.16	9.24
	5	163	5.69	0.03	0.44	4.55	5.60	5.60	5.60	6.60
Albumin	1	4620	3.66	0.01	0.51	1.50	3.50	3.70	4.00	4.90
	2	1090	3.58	0.02	0.70	0.06	3.19	3.62	4.00	5.27
	3	96	3.65	0.06	0.56	2.40	3.40	3.63	4.08	4.98
	4	1264	3.59	0.02	0.70	1.28	3.14	3.61	4.06	6.76
	5	163	2.16	0.04	0.51	0.27	2.40	2.40	2.40	2.40
Globulin	1	4620	3.25	0.01	0.66	2.00	3.00	3.00	4.00	5.00
	2	1090	3.74	0.05	1.63	-0.30	2.49	3.39	4.74	9.39
	3	96	3.14	0.08	0.80	0.28	3.00	3.00	4.00	5.53
	4	1264	3.18	0.01	0.41	1.31	2.95	3.18	3.43	4.31
	5	163	3.41	0.06	0.75	3.00	3.00	3.00	3.00	5.00
A/G	1	4620	1.17	0.00	0.29	0.40	1.00	1.20	1.30	2.30
	2	1090	0.99	0.02	0.60	-1.38	0.71	1.10	1.42	2.73
	3	96	1.26	0.04	0.42	0.17	0.90	1.19	1.43	2.20
	4	1264	1.18	0.01	0.31	0.17	0.99	1.19	1.38	2.83
	5	163	0.66	0.02	0.27	-0.24	0.80	0.80	0.80	0.80
Bilirubin	1	4620	0.78	0.00	0.33	0.20	0.60	0.70	0.90	2.50
	2	1090	2.37	0.01	0.27	1.14	2.19	2.40	2.59	3.01
	3	96	0.80	0.07	0.64	-0.89	0.60	0.75	1.00	2.63
	4	1264	-0.25	0.01	0.31	-1.18	-0.46	-0.27	-0.07	1.32
	5	163	0.31	0.04	0.46	-1.21	0.40	0.40	0.40	0.70

Alk Phos	1	4620	71.00	0.44	30.02	23.00	55.00	66.00	78.00	469.00
	2	1090	62.37	2.05	67.72	-193.71	21.57	61.62	106.87	404.51
	3	96	69.46	3.04	29.78	-18.21	54.00	67.00	88.95	197.32
	4	1264	90.79	1.33	47.44	-157.37	60.19	89.21	120.78	360.71
	5	163	52.16	0.58	7.39	22.56	50.00	50.00	50.00	88.61
AST	1	4620	26.32	0.33	22.17	12.00	18.00	22.00	29.00	478.00
	2	1090	26.34	1.38	45.72	-107.50	-2.04	25.46	51.42	383.46
	3	96	21.14	1.58	15.47	-53.74	16.00	20.00	24.77	83.94
	4	1264	20.16	0.31	10.99	-16.39	13.65	19.11	25.75	87.58
	5	163	13.92	0.28	3.62	12.00	12.00	12.00	12.00	27.07
ALT	1	4620	23.56	0.24	16.41	4.00	15.00	20.00	26.00	180.00
	2	1090	10.76	1.05	34.55	-70.74	-9.61	7.92	25.94	262.07
	3	96	19.63	1.36	13.29	-46.92	14.37	20.00	22.83	53.51
	4	1264	25.47	0.43	15.40	-42.11	14.69	24.93	35.00	113.44
	5	163	12.04	0.45	5.77	-12.17	11.00	11.00	11.00	23.00
SPO2_min	1	4620	97.38	0.04	2.43	61.00	96.00	98.00	99.00	100.00
	2	1090	97.58	0.07	2.42	83.00	96.00	98.00	100.00	100.00
	3	96	97.74	0.21	2.06	93.00	96.25	98.00	100.00	100.00
	4	1264	97.74	0.06	2.29	85.00	97.00	98.00	99.00	100.00
	5	163	98.64	0.23	2.87	88.00	99.00	100.00	100.00	100.00
SPO2_max	1	4620	98.89	0.02	1.51	91.00	98.00	99.00	100.00	100.00
	2	1090	98.85	0.05	1.59	93.00	98.00	100.00	100.00	100.00
	3	96	99.02	0.15	1.45	95.00	99.00	100.00	100.00	100.00
	4	1264	99.14	0.03	1.20	94.00	99.00	100.00	100.00	100.00
	5	163	99.85	0.06	0.75	95.00	100.00	100.00	100.00	100.00
SPO2_mode	1	4620	98.30	0.03	1.89	83.00	97.00	99.00	100.00	100.00
	2	1090	98.37	0.06	1.95	92.00	97.00	99.00	100.00	100.00
	3	96	98.51	0.17	1.64	95.00	98.00	99.00	100.00	100.00
	4	1264	98.62	0.04	1.58	89.00	98.00	99.00	100.00	100.00

	5	163	99.50	0.11	1.39	94.00	100.00	100.00	100.00	100.00
Temperature_min	1	4620	36.18	0.01	0.55	34.10	35.80	36.20	36.50	38.20
	2	1090	36.07	0.02	0.62	34.20	35.80	36.10	36.50	37.50
	3	96	34.93	0.07	0.67	34.00	34.30	34.90	35.28	36.90
	4	1264	36.20	0.02	0.56	34.20	35.80	36.30	36.60	37.50
	5	163	35.12	0.09	1.20	34.00	34.00	34.60	36.30	37.00
Temperature_mean	1	4620	36.26	0.01	0.54	34.10	35.88	36.23	36.62	38.28
	2	1090	36.14	0.02	0.61	34.20	35.85	36.20	36.53	37.58
	3	96	35.63	0.06	0.58	34.47	35.21	35.58	35.93	37.48
	4	1264	36.29	0.02	0.55	34.20	35.95	36.33	36.65	37.50
	5	163	35.21	0.09	1.21	34.00	34.00	34.80	36.45	37.20
Temperature_median	1	4620	36.26	0.01	0.55	34.10	35.90	36.25	36.60	38.30
	2	1090	36.14	0.02	0.61	34.20	35.85	36.20	36.50	37.60
	3	96	35.70	0.06	0.62	34.20	35.25	35.65	36.00	37.65
	4	1264	36.29	0.02	0.55	34.20	35.95	36.35	36.65	37.50
	5	163	35.22	0.10	1.22	34.00	34.00	34.80	36.40	37.25
Temperature_range	1	4620	0.16	0.00	0.14	0.00	0.10	0.10	0.20	0.90
	2	1090	0.14	0.00	0.12	0.00	0.10	0.10	0.20	0.80
	3	96	1.18	0.04	0.44	0.50	0.90	1.10	1.40	2.20
	4	1264	0.16	0.00	0.15	0.00	0.10	0.10	0.20	1.00
	5	163	0.18	0.02	0.22	0.00	0.00	0.10	0.30	1.20
Temperature_stdev	1	4620	0.07	0.00	0.06	0.00	0.04	0.05	0.09	0.35
	2	1090	0.06	0.00	0.05	0.00	0.04	0.05	0.08	0.30
	3	96	0.49	0.02	0.19	0.18	0.36	0.46	0.59	1.08
	4	1264	0.07	0.00	0.06	0.00	0.04	0.05	0.09	0.37
	5	163	0.07	0.01	0.09	0.00	0.00	0.05	0.12	0.45
Temperature _var	1	4620	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.13
	2	1090	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.09
	3	96	0.27	0.02	0.22	0.03	0.13	0.21	0.35	1.17
	4	1264	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.14

	5	163	0.01	0.00	0.03	0.00	0.00	0.00	0.01	0.20
Temperature_entropy	1	4620	0.80	0.01	0.48	0.00	0.69	0.69	1.10	1.79
	2	1090	0.74	0.01	0.47	0.00	0.69	0.69	1.10	1.79
	3	96	1.37	0.03	0.24	0.69	1.10	1.39	1.61	1.79
	4	1264	0.82	0.01	0.50	0.00	0.69	0.69	1.10	1.79
	5	163	0.71	0.05	0.67	0.00	0.00	0.69	1.39	1.79
Pulse Rate _min	1	4620	71.60	0.18	12.21	32.00	63.00	70.50	80.00	116.00
	2	1090	71.71	0.41	13.48	46.00	61.00	70.00	80.00	110.00
	3	96	68.05	1.22	11.96	48.00	58.00	67.00	74.00	94.00
	4	1264	71.77	0.38	13.63	41.00	60.00	70.00	81.00	131.00
	5	163	65.88	0.96	12.25	44.00	58.00	65.00	73.00	98.00
Pulse Rate _max	1	4620	84.24	0.21	14.17	50.00	74.00	83.00	93.00	141.00
	2	1090	84.76	0.51	16.70	48.00	74.00	82.00	94.25	142.00
	3	96	81.22	1.34	13.13	58.00	72.00	78.00	92.00	111.00
	4	1264	84.04	0.45	15.94	54.00	73.00	82.00	95.00	146.00
	5	163	79.55	1.03	13.12	60.00	70.00	77.00	85.00	131.00
Pulse Rate _var	1	4620	36.10	0.93	63.16	0.00	5.37	14.97	40.24	1032.57
	2	1090	42.20	2.74	90.55	0.17	4.57	15.55	40.68	1034.40
	3	96	37.22	5.37	52.62	0.40	5.64	16.57	47.73	277.87
	4	1264	35.21	1.81	64.35	0.00	4.44	12.68	35.77	691.47
	5	163	39.34	4.32	55.12	0.17	4.57	21.47	47.10	376.57
Pulse Rate _entropy	1	4620	1.56	0.00	0.24	0.00	1.39	1.61	1.78	1.79
	2	1090	1.55	0.01	0.23	0.69	1.39	1.61	1.78	1.79
	3	96	1.57	0.02	0.22	1.10	1.39	1.61	1.79	1.79
	4	1264	1.53	0.01	0.27	0.00	1.39	1.61	1.78	1.79
	5	163	1.55	0.02	0.25	0.69	1.39	1.61	1.78	1.79
Respiratory Rate_max	1	4620	14.56	0.05	3.53	8.00	12.00	14.00	16.00	40.00
	2	1090	14.28	0.11	3.67	8.00	12.00	13.00	16.00	31.00
	3	96	14.00	0.31	3.08	9.00	12.00	13.50	16.00	28.00
	4	1264	14.49	0.10	3.45	7.00	12.00	14.00	16.00	32.00
	5	163	12.77	0.15	1.95	8.00	12.00	13.00	14.00	17.00

Respiratory Rate_mode	1	4620	13.42	0.04	2.52	5.00	12.00	12.00	15.00	37.00
	2	1090	13.37	0.09	3.01	6.00	12.00	12.00	16.00	28.00
	3	96	12.74	0.26	2.56	8.00	12.00	12.00	14.00	20.00
	4	1264	13.44	0.07	2.57	5.00	12.00	13.00	15.00	26.00
	5	163	11.85	0.15	1.94	8.00	10.00	12.00	14.00	16.00
Respiratory Rate_stdev	1	4620	0.86	0.02	1.37	0.00	0.00	0.41	1.10	13.58
	2	1090	0.76	0.04	1.20	0.00	0.00	0.41	1.03	9.81
	3	96	0.98	0.14	1.40	0.00	0.00	0.52	1.63	7.11
	4	1264	0.82	0.04	1.30	0.00	0.00	0.41	1.03	8.73
	5	163	0.70	0.08	0.96	0.00	0.00	0.41	1.03	3.67
Respiratory Rate_entropy	1	4620	0.48	0.01	0.49	0.00	0.00	0.68	0.69	1.78
	2	1090	0.44	0.01	0.47	0.00	0.00	0.63	0.69	1.77
	3	96	0.52	0.05	0.50	0.00	0.00	0.68	0.69	1.59
	4	1264	0.46	0.01	0.49	0.00	0.00	0.63	0.69	1.78
	5	163	0.47	0.04	0.51	0.00	0.00	0.63	0.69	1.59
SBP Delta	1	4620	5.46	0.39	26.28	-92.00	-9.00	2.00	20.00	219.00
	2	1090	-0.33	0.72	23.91	-79.00	-14.00	0.00	12.00	187.00
	3	96	5.88	2.95	28.88	-67.00	-10.50	0.00	25.75	75.00
	4	1264	3.46	0.55	19.57	-55.00	-8.00	2.00	15.00	182.00
	5	163	-21.70	3.11	39.69	-120.00	-62.00	-14.00	8.00	120.00
DBP Delta	1	4620	2.00	0.27	18.04	-87.00	-8.00	1.00	12.00	89.00
	2	1090	-1.99	0.49	16.06	-44.00	-12.00	-2.00	7.00	79.00
	3	96	1.57	1.93	18.92	-52.00	-9.00	0.50	15.75	42.00
	4	1264	0.88	0.46	16.34	-58.00	-8.00	0.00	10.00	192.00
	5	163	0.26	1.28	16.30	-42.00	-12.00	1.00	10.00	77.00
Systolic	1	4620	126.15	0.33	22.26	60.00	111.00	124.00	139.00	318.00
	2	1090	123.58	0.64	21.22	77.00	110.00	121.00	133.00	310.00
	3	96	127.26	2.36	23.09	72.00	110.00	123.00	142.00	196.00
	4	1264	119.96	0.57	20.16	70.00	105.00	118.00	133.00	307.00

	5	163	140.48	2.77	35.39	55.00	112.00	135.00	165.00	295.00
Diastolic	1	4620	73.35	0.22	14.97	13.00	63.00	73.00	83.00	167.00
	2	1090	71.48	0.43	14.34	26.00	62.00	71.00	80.00	127.00
	3	96	73.32	1.53	14.98	44.00	63.00	73.50	81.75	122.00
	4	1264	70.05	0.42	15.07	22.00	60.00	69.00	79.00	278.00
	5	163	73.17	1.57	20.07	35.00	63.00	71.00	87.00	157.00
Mean	1	4580	93.85	0.25	16.67	26.00	82.00	93.00	104.00	224.00
	2	1087	91.35	0.52	17.25	44.00	81.00	90.00	100.00	302.00
	3	96	96.10	1.89	18.49	57.00	83.00	93.50	106.50	152.00
	4	1260	89.75	0.46	16.17	38.00	79.00	88.00	99.00	293.00
	5	154	96.27	1.89	23.49	45.00	80.00	92.50	112.00	202.00
Hypo	1	4620	0.06	0.00	0.24	0.00	0.00	0.00	0.00	1.00
	2	1090	0.07	0.01	0.25	0.00	0.00	0.00	0.00	1.00
	3	96	0.04	0.02	0.20	0.00	0.00	0.00	0.00	1.00
	4	1264	0.08	0.01	0.27	0.00	0.00	0.00	0.00	1.00
	5	163	0.17	0.03	0.38	0.00	0.00	0.00	0.00	1.00

Appendix D: INN Threshold Values by Cluster

Rule		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
1	Age	48.32	52.06	51.67	44.92	48.32
2	Pre Temp	97.84	97.84	97.51	97.85	97.84
2	Pre SBP	130.76	136.43	138.80	121.24	130.76
2	Pre DBP	77.21	75.92	76.05	70.19	77.21
3	Pre HR	79.62	76.16	75.16	77.31	79.62
3	Pre SpO2	98.10	97.90	97.85	98.93	98.10
3	Pre RR	18.34	16.37	18.48	16.49	18.34
4	WBC	7.30	7.88	6.69	7.41	7.30
4	Hgb	12.52	12.80	12.09	12.87	12.52
4	Platelet	247.83	263.95	269.41	274.86	247.83
5	Sodium	136.97	136.46	137.99	132.84	136.97
5	Potassium	3.85	3.88	3.84	5.08	3.85
5	Chloride	104.62	102.43	104.45	101.47	104.62
5	Pre CO2	23.42	22.62	23.05	21.29	23.42
6	Anion Gap	10.78	10.79	9.98	12.16	10.78
6	Glucose	110.76	103.68	104.55	93.74	110.76
6	BUN	9.34	12.94	11.46	6.55	9.34
6	Creatinine	0.75	0.89	0.74	0.77	0.75
7	eGFR Af	15.13	51.35	31.58	33.07	15.13
7	eGFR nonAf	87.89	89.10	95.86	121.22	87.89
7	BUN/Creatinine	64.89	69.58	65.76	64.57	64.89
7	Calcium	8.81	8.94	8.89	9.11	8.81
8	Protein	6.40	6.79	6.59	6.85	6.40
8	Albumin	3.59	3.22	3.67	3.24	3.59
8	Globulin	3.18	3.40	3.01	3.19	3.18
8	A/G	1.10	0.97	1.25	1.04	1.10
9	Bilirubin	0.86	2.25	0.73	-0.26	0.86
9	Alk Phos	59.53	53.39	64.92	76.77	59.53
9	AST	18.71	24.94	21.30	21.98	18.71
9	ALT	25.76	17.28	17.58	21.17	25.76
10	SPO2_min	97.94	97.89	98.29	97.97	97.94
10	SPO2_max	98.79	99.29	98.62	99.24	98.79
10	SPO2_mode	97.87	98.22	98.63	98.34	97.87
11	Temperature_min	36.12	35.87	34.85	36.16	36.12
11	Temperature_mean	36.27	35.96	35.67	36.02	36.27
11	Temperature_median	35.81	36.17	35.61	36.43	35.81
11	Temperature_range	0.15	0.10	1.05	0.22	0.15
12	Temperature_stdev	0.05	0.06	0.46	0.08	0.05

12	Temperature_var	0.01	0.01	0.19	0.01	0.01
12	Temperature_entropy	0.70	0.65	1.37	0.65	0.70
13	Pulse Rate _min	72.26	68.16	69.11	71.24	72.26
13	Pulse Rate _max	83.83	86.92	77.58	86.36	83.83
13	Pulse Rate _var	23.89	15.32	37.06	14.51	23.89
13	Pulse Rate _entropy	1.61	1.60	1.62	1.54	1.61
14	Respiratory Rate_max	14.33	14.67	14.78	14.90	14.33
14	Respiratory Rate_mode	13.09	12.43	12.31	12.88	13.09
14	Respiratory Rate_stdev	0.97	0.62	0.91	1.03	0.97
14	Respiratory Rate_entropy	0.75	0.47	0.53	0.63	0.75
15	SBP Delta	26.71	-3.13	6.71	14.79	26.71
15	DBP Delta	9.11	3.72	-0.48	1.28	9.11

CURRICULUM VITA

NAME:	Jodie Ritter
ADDRESS:	5607 Harrods Glen Drive Prospect, KY 40059
DOB:	Sept. 19, 1972
EDUCATION:	M.S., Industrial Engineering University of Louisville 2020-2022