Prediction of Acute Kidney Injury in Hospitalized COVID-19 Patients Using Machine Learning

By: David Carbonello

Thesis Advisors: Christopher Schmid PhD, Pankaj Sarin MD

PHP 2560 Final Project

1. Introduction

Originating from Wuhan, China in 2019, the highly infectious COVID-19 virus has become a global pandemic with 69M confirmed cases and 1.57M deaths worldwide (1). The virus is caused by a newly identified coronavirus, SARS-CoV-2 (Severe Acute Respiratory Syndrome), and is known to have many physiological manifestations. Initially thought of as a lung infection, the virus is now also understood to impact the cardiovascular system causing complications in the kidneys, brain, and other organs (3). Recent studies show that acute kidney injury (AKI), a sudden episode of kidney failure that happens within a few hours or days (18), is associated with increased disease severity and mortality in hospitalized COVID-19 patients. Latest data provides evidence of AKI occurrence in 37.5% of fatal COVID-19 cases(23). Additionally, up to 20% of COVID-19 cases experiencing AKI require dialysis, and has led to shortages in dialysis supplies and medical equipment during spikes of the pandemic (10). Therefore, identifying COlVID-19 patients that are at higher risk of AKI is of great importance, and could help hospitals allocate resources more effectively and improve patient outcomes moving forward.

Although many predictive models of AKI in hospitalized COVID-19 patients have had success, few implement machine learning algorithms that incorporate demographics, vitals, lab results, ICU admission and pre-existing medical conditions in conjunction (2). Additionally, no known models have been developed using data from the Greater Boston Area. With this in mind, the following study reports the implementation of Logistic Regression, Random Forest and Neural Network models developed using data from electronic medical records of hospitalized COVID-19 patients at Brigham & Women's Hospital in Boston, MA.

Objective:

The aim of this study is to develop a tool to predict the risk of AKI among hospitalized COVID-19 patients within their first 48 hours of hospital admission using a variety of risk factors. A strong prediction tool will aid in the appropriate triage of individuals with high risk of AKI, and help hospitals be better prepared with medical equipment moving forward.

2. Methods

2.1 Study Population

Study population included all positive COVID-19 cases that were admitted to Brigham and Women's Hospital in Boston, MA between January 2020-September 2020. Electronic Medical Records (EMRs) were investigated retrospectively to identify occurrences of AKI. A positive COVID-19 case was defined using a positive reverse transcriptase polymerase chain reaction (RT-PCR). Presumptive positive cases were also included as positive COVID-19 cases in the study population. Patients that had pre-existing end stage kidney disease were excluded. Additionally, patients that were missing BMI or serum creatinine lab values were also excluded. 2,002 patients were hospitalized with COVID-19 and were included in the model building process. Of these patients included, 35.6% experienced AKI based on guidelines set by KIGDO.

2.2 Data Collection and Preparation

All data on hospitalized COVID-19 patients from January 2020-September 2020 was queried from QAQI database of Brigham & Women's Hospital in Boston, MA. Specific information on demographics, vitals, lab values, ICU admission, and pre-existing conditions were extracted from patient's Electronic Medical Records (EMR).

Demographics included patient's age, sex, and race. Patient race was divided into 4 categories including Black or African, White (non-Hispanic), Hispanic or Latino, and Asian (Non-Hispanic) defined using United States Census Bureau categories.

Vitals collected included patient's Systolic Blood Pressure, Diastolic Blood Pressure, Body Temperature (Fahrenheit), Heart Rate, Body Mass Index, and Respiration Rate.

Lab values included as predictors were potassium, calcium, albumin, chloride and white blood cell count. Lab values from the first 48 hours after admission were averaged for each patient and used as predictors. All lab values of Serum Creatinine were used identify occurrence of Acute Kidney Injury at any point during a patient's hospitalization.

ICU admission status was defined as a 0,1 variable based on whether a patient was admitted directly to the ICU or within the first 48 hours of their hospital stay.

Pre-existing conditions for patients were defined using the latest International Classification Disease Codes (ICD-10 codes). Comorbidities and their corresponding ICD-10 code included as predictors were Hypertension (I10),Congestive Heart Failure(I50.9), Liver Disease (K76.9) and Peripheral Vascular Disease(I73.9). (Note I50.9 code corresponds to Unspecified Heart Failure, so additional filters were applied in SQL to retrieve specifically Congestive Heart Failure).

Data frames containing information on demographics, vitals, lab values, ICU and pre-existing conditions were cleaned and pre-processed separately by removing unneeded columns and defining variables of interest. Ultimately, each individual data frame was joined on specific patientencounterIDs to build a final data frame that was used for the model building process.

2.3 Definitions of Outcomes:

The outcome variable of Acute Kidney Injury was defined using the following criteria from Kidney Disease Improving Global Outcomes (KIDGO):

- Increase in serum creatinine by .3mg/dL or more within 48 hours **or**
- Increase in serum creatinine by 1.5 times baseline or more within last 7 days

Because many patients did not have a reliable baseline serum creatinine on record, baseline serum creatinine was estimated using a back-calculation for various age, sex, and race categories as suggested by KDIGO (appendix Table 1).

2.4 Data Summary and Visualization

Standard summary statistics of continuous predictors were calculated and tabulated in Table 2 of the appendix**.** Histograms of continuous variables were produced to observe distributions and potential outliers in data and are shown in Figure 1 of the appendix. Distributions of continuous predictors were approximately normally distributed with few observed outliers. Variables that did appear to have outliers were Temperature (Fahrenheit), Respiration Rate, and White Blood Cell Count. Data validation has yet to be performed for extreme values and no outliers were removed before fitting models. Incidence of AKI within specific subgroups is shown in appendix Table 3. The overall proportion of AKI in the study population was 35.6%. Greater proportions of AKI were observed for patients with pre-existing conditions such as hypertension (43%) liver disease (54%), and peripheral vascular disease(48%). Additionally, 67% of patients admitted to the ICU experienced AKI.

2.5 Model Development

Methods considered to produce the best fitting model and develop the strongest prediction tool for classifying patients with AKI included Logistic Regression, Random Forest, and Neural Network. Logistic Regression was chosen for its probabilistic interpretation, popularity in COVID-19 research, and competitive performance results when compared to other machine learning algorithms used in classification tasks (12). Random Forest and Neural Networks have been chosen as alternative modeling approaches to allow more flexible relationships between prediction features and outcome, and to see if a higher prediction accuracy can be attained. For the Logistic Regression and Neural Network, a probability threshold of .5 was used to classify predictions of AKI Status, where probability >.5 was classified as experiencing AKI, and <.5 was classified as not experiencing AKI. All models were fit using response variable *AKI_Status,* which was a 0,1 variable corresponding to AKI diagnosis. To increase generalizability of the study, data was randomly split 75% into a training set (1,501 observations) for model development, and the remaining 25% (501 observations) into a test set to assess model performance. All of the statistical analysis was carried out in python.

2.5.1 Logistic Regression

A stepwise selection algorithm was implemented to perform variable selection when fitting the logistic regression on the training data. The algorithm used both forward selection and backward elimination examining at each step for variable inclusion or exclusion. A p-value threshold of less than .01 was used for variable inclusion, and a p-value threshold of greater than .05 was used for variable exclusion. All predictors were considered in the initial list of variables that passed through the stepwise selection algorithm. The resulting model from stepwise selection were then evaluated on the holdout test set.

2.5.2 CART

A random forest classifier was implemented as an extension of a single classification tree for its proven ability to make significant improvements to traditional CART methods. The algorithm uses a bootstrap aggregation (bagging) procedure and a random subset of features to fit multiple decision trees, resulting in decreased variance and decorrelation of trees. Classification predictions from random forest are determined using a majority vote rule in terminal lead nodes (22). The random forest classifier was trained using 10-fold cross validation to determine the optimal combination of number of trees in the forest and criterion from a prespecified list of parameter grid values based on highest area under the roc curve produced on the holdout test set. The parameter grid values considered for number of trees were 10, 25, 50, and 100, and criterion of either gini or entropy. The resulting best combination of criterion and number of trees determined by cross validation was the gini criterion and 100 trees, and this model structure was used to make predictions on the holdout test set.

2.5.3 Neural Network

The Neural Network was developed with architecture of a Multilayer Perceptron Neural Network (MPNN), meaning it contained one input and output layer, and one or more hidden layers where each node is fully connected. This structure can lead to more effective decision making by detecting and using hidden information that may be present in high volumes of data. An example of a MPNN structure with two hidden layers is shown in Figure 2. The blue circles serve as the input layer and are the predictors included in the model. The gray circles are the neurons that form the hidden layers and connect successive layers. The red node is the output layer corresponding to the predicted response (20).

Figure 2: Multilayer Neural Network example structure

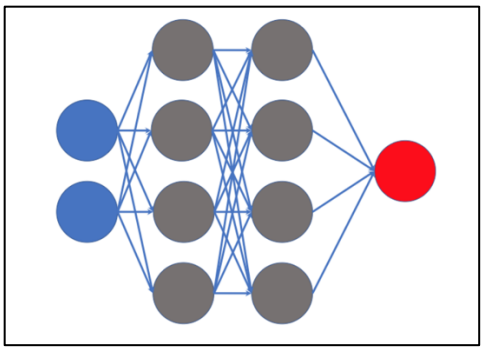
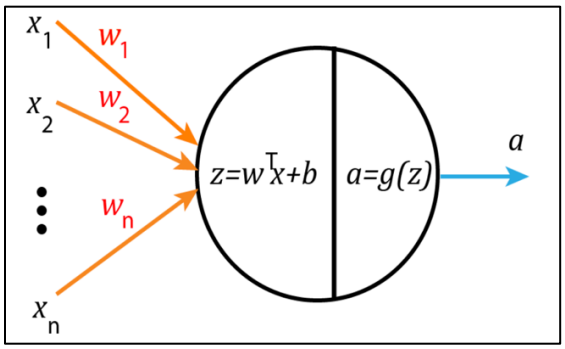


Figure 3: Mathematical representation of 1 neuron



The mathematical structure of an individual neuron is shown in Figure 3 where $X_1 \dots X_n$ are a specific vector of inputs that are multiplied by weights $w_1 \dots w_n$. A bias, b is then added to the weighted sum of inputs as shown in Equation 1:

$$z = \sum_{i=1}^{n} x_i w_i + b \qquad (1)$$

Here z is the net input fed through a specific that is passed through an activation function $g(z)$ as shown in Equation 2:

$$a = g(z) = g(\sum_{i=1}^{n} x_i w_i + b) \qquad (2)$$

where a is the output that is passed to other neurons (21). The neural network developed to predict Acute Kidney Injury contained 25 features in the input layer, 2 hidden layers, and 1 output layer. The first hidden layer contained 220 neurons, second hidden layer contained 20 neurons, and output layer contained a single neuron. The input layer and first hidden layer both used Rectified Linear Unit (ReLU) (Figure 4) activation function, and second hidden layer used sigmoid activation function (Figure 5).

Figure 4

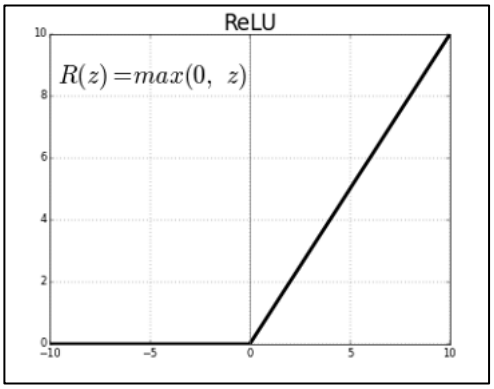
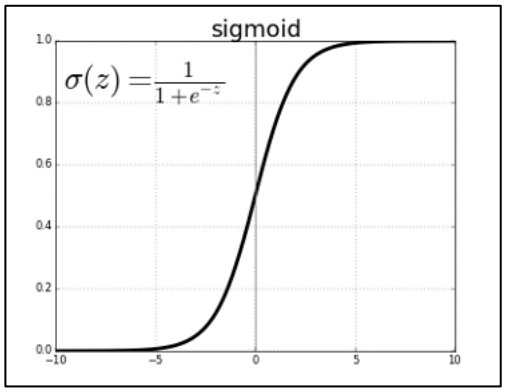


Figure 5



The ReLU function was used for its computational efficiency on account that some neurons will be deactivated during the model fitting process if the net input to the neuron is less than 0 (16). This is because the ReLU activation function sets negative values to 0 as shown in Figure 4. The sigmoid activation function was used to restrict the continuous predicted outcome between 0 and 1 for a probabilistic interpretation. The Adam optimizer algorithm was used as an alternative to stochastic gradient decent to minimize the binary loss function and recursively update weights when fitting the neural network. All input features in the Neural Network were standardized to improve model performance and efficiency.

3. Results

Models were evaluated primarily based on prediction accuracy and AUROC curve produced on the holdout test set. Additional metrics that were returned included Balanced Accuracy, Sensitivity, AUPRC, and F1 scores as shown in Table 4.

Table 4

Model	Dataset	Outcome Measure	Accuracy	Balanced Accuracy	Sensitivity	AUROC	AUPRC	F1
Logistic Regression	test	AKI_status	0.7005988	0.638943765	0.39572193	0.75240982	0.66666667	0.4966443
Random Forest	test	AKI_status	0.71856287	0.655437856	0.40641711	0.79135018	0.71698113	0.51877133
Neural network	test	AKI_status	0.71257485	0.663638067	0.47058824	0.76066964	0.66165414	0.55

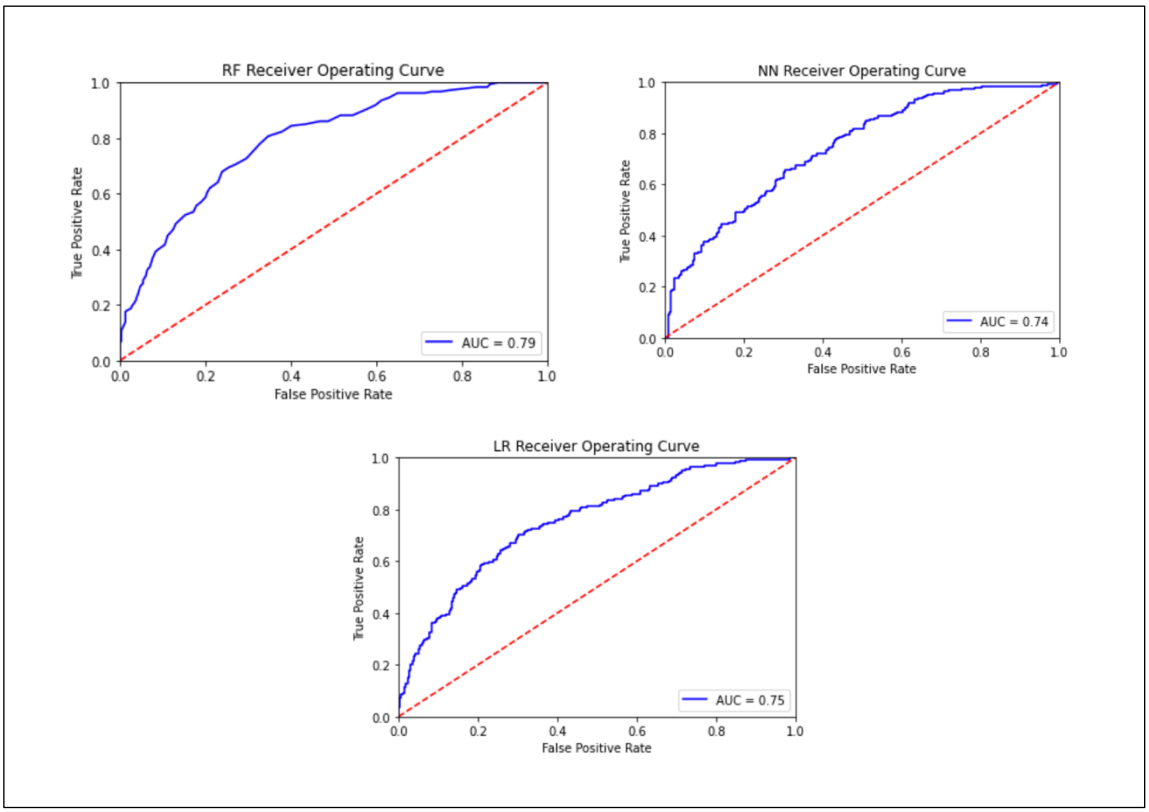
The random forest produced the highest prediction accuracy, correctly identifying 71.85% of patient's AKI disease status in the test set. The random forest also yielded the highest AUROC (.7913), AUCPRC (.7170), and F1 scores (.512). The top 10 most important features included in the random forest model were Albumin, Potassium, White Blood Cell Count, Calcium Systolic Blood Pressure, Heart Rate, Temperature (Fahrenheit), and Body Mass Index.

Of the 25 predictors considered for logistic regression model, 10 predictors were selected from the stepwise selection algorithm. Selected variables included Albumin, Potassium, Calcium, Hypertension , ICU , WhiteBloodCellCNT, Chloride, White (non-hispanic), sex, BloodPressureSystolicNBR. Corresponding coefficients, OR and p-values of these variables are shown in Table 7 in the appendix. Factors associated with a higher odds of AKI were elevated levels of Potassium (OR=2.03), pre-existing Hypertension (OR=1.83), being admitted to the ICU(OR=2.38), having sex=1 (male) (OR=1.436), and elevated systolic blood pressure (OR=1.014). Factors associated with a lower odds of AKI were elevated levels of Albumin (OR=.703), Calcium (OR=.485), and being in the White (non-hispanic) race category (OR=.71752). Age was initially added to the logistic regression model and then removed later in stepwise algorithm based on p-value of .06. The accuracy of the logistic regression was .701 and produced AUROC of .7524.

The Neural Network produced the second highest prediction accuracy (.7125) and AUROC (.761), and produced the highest balanced accuracy, sensitivity and F1 score. The ROC curve produced for each prediction model are shown below in Figure 6, which indicates the

model performance across all possible classification thresholds. Effects plots of specific features for all three models are included in the appendix in Figure 7**.** The effect plots show the relationship between predictor variables and the probability of AKI predicted by the three models holding other variables constant.

Figure 6 : Shows ROC curves for RF, NN, and LR respectively



4. Discussion

The random forest was the most successful learning algorithm of the three models considered correctly classifying 71.85% of patient's AKI disease status in the test set. Additionally, the AUROC of the random forest (accuracy=.79) was the greatest of any of the models indicating the best fit considering all different classification thresholds. However, the sensitivity of the random forest (sensitivity=.40) is less than that of the neural network (sensitivity=.47), indicating that neural network performed better than the random forest in terms of correctly identifying patients

that experienced AKI. The overall accuracy of the neural network (accuracy=.717) was only slightly less than the random forest suggesting this method could be a favorable alternative. Although the logistic regression had the lowest accuracy and AUROC, it should continue to be investigated for its interpretability. Additionally, this model was produced without using cross-validation and with an algorithm that was very restrictive to the number of predictors included. Considering the logistic regression only included 10 predictors, where the random forest and neural network contained all 25 predictors, the results of the logistic regression are very competitive and alternative model development procedures should be further investigated.

Based on the results of the stepwise logistic regression (appendix Table 7), the odds of developing AKI are 2.38 times greater for a patient admitted to the intensive care unit compared to a patient who is not holding other predictors constant. Additionally, the odds of developing AKI are 1.833 times greater for a patient with pre-existing hypertension holding other variables constant. Furthermore, increases in potassium, white blood cell count, chloride and systolic blood pressure were associated with increases in the odds of AKI. Similar associations between predictors and predicted probabilities of AKI were shown from the effects plots (appendix Figure 7), suggesting the random forest, logistic regression, and neural network are finding similar relationships between predictors and response.

There are many factors that should be considered to improve prediction performance of these models. In terms of data included, there are additional predictors such as Lactate dehydrogenase (LDH), C-reactive protein (CRP) and glucose that have been shown to have a strong association with Acute Kidney injury and should be incorporated into the models(2). Additionally, no data validation has been performed thus far and no extreme values have been removed that could be effecting model performance. The issue of multicollinearity could be presenting itself in the logistic regression model, as correlated predictors have not yet been investigated and removed. For model development, other variable selection methods should be considered such as a penalized lasso or ridge regression to see if higher prediction accuracy can be attained. Furthermore, no interactions between variables were considered in the logistic regression model causing potentially significant relationships between variables to go undetected. For the random forest, more robust methods should be considered such as boosting, where each tree is grown using information from previously grown trees. The neural network was developed using a user specified number of nodes within each hidden layer. The number of hidden layers and nodes should be investigated further to maximize prediction accuracy on the test set.

5. Conclusion

The various prediction models developed suggest that Acute Kidney Injury can be identified in COVID-19 patients within their first 48 hours of hospitalization, although the results still leave much room for improvement. Overall, the Random Forest classifier produced the strongest prediction results when evaluated on the test set with an accuracy of .712 and AUROC of .79. However, both the Logistic Regression and Neural Network produced competitive results to that of the Random Forest and should be investigated further once more variables are included and more robust methods have been implemented.

References:

1.) Coronavirus Cases: *Worldmeter,* www.worldometers.info/coronavirus/.
2.) Vaid1, A., Somani1*, S., Russak1, A., Freitas1, J., Chaudhry1, F., Paranjpe1, I., . . . Authors..., L. (n.d.). Machine Learning to Predict Mortality and Critical Events in a Cohort of Patients With COVID-19 in New York City: Model Development and Validation. Retrieved December 11, 2020, from https://www.jmir.org/2020/11/e24018
3.) Fornell, Dave. "The Cardiovascular Impact of COVID-19." *DAIC,* 28 July 2020, www.dicardiology.com/article/cardiovascular-impact-covid-19.

4.) "Coronavirus." *World Health Organization*, World Health Organization, www.who.int/health-topics/coronavirus.
5.) Sauer, Lauren M. "What Is Coronavirus?" *What Is Coronavirus? | Johns Hopkins Medicine*, www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus.
6.) Palevsky, Paul M., Radhakrishnan, Jai, Townsend, Raymond R. "Coronavirus Disease 2019 (COVID-19): Issues Related to Kidney Disease and Hypertension." *UpToDate*, www.uptodate.com/contents/coronavirus-disease-2019-covid-19-issues-related-to-kidney-disease-and-hypertension.
7.) Cheng Y, Luo R, Wang K, et al. Kidney disease is associated with in-hospital death of patients with COVID-19. 2020 March 5
8.) Diao, Bo, et al. "Human Kidney Is a Target for Novel Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infection." *MedRxiv*, Cold Spring Harbor Laboratory Press, 1 Jan. 2020, www.medrxiv.org/content/10.1101/2020.03.04.20031120v4.. Further clinical and autopsy reports of COVID-19 patients suggest there is evidence of a pathophysiology of AKI in COVID-19
9.) Batlle, Daniel, et al. "Acute Kidney Injury in COVID-19: Emerging Evidence of a Distinct Pathophysiology." *American Society of Nephrology*, American Society of Nephrology, 1 July 2020, jasn.asnjournals.org/content/31/7/1380.
10.) Chan, Lili, et al. "Acute Kidney Injury in Hospitalized Patients with COVID-19." *MedRxiv*, Cold Spring Harbor Laboratory Press, 1 Jan. 2020, www.medrxiv.org/content/10.1101/2020.05.04.20090944v1.
11.) PolicyLab at Children's Hospital of Philadelphia. "New COVID-19 Projections Show Risk for Resurgence Continues Spread to Northeast." *PR Newswire: News Distribution, Targeting and Monitoring*, 22 July 2020, www.prnewswire.com/news-releases/new-covid-19-projections-show-risk-for-resurgence-continues-spread-to-northeast-301098283.html.
12.) Kirasich, Kaitlin; Smith, Trace; and Sadler, Bivin (2018) "Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets," *SMU Data Science Review*: Vol. 1 : No. 3 , Article 9.
13.) LIPPMANN, R. P., 1987, An introduction to computing with neural nets. IEEE ASSP Magazine, April, pp. 4-22.
14.) KAVZOGLU, T., and MATHER, P.M., 1998, Assessing artificial neural network pruning algorithm. Proceedings of the 24th Annual Conference and Exhibition of the Remote Sensing Society (RSS'98), pp. 603-609.
15.) Riley, Richard D, et al. "Calculating the Sample Size Required for Developing a Clinical

Prediction Model." *The BMJ*, British Medical Journal Publishing Group, 18 Mar. 2020, www.bmj.com/content/368/bmj.m441.

17.) Dishashree GuptaDishashree is passionate about statistics and is a machine learning enthusiast. She has an experience of 1.5 years of Market Research using R. "Activation Functions: Fundamentals Of Deep Learning." *Analytics Vidhya*, 19 July 2020, www.analyticsvidhya.com/blog/2020/01/fundamentals-deep-learning-activation-functions-when-to-use-them/.

18.) "Acute Kidney Injury (AKI)." *National Kidney Foundation*, 30 Oct. 2020, www.kidney.org/atoz/content/AcuteKidneyInjury.

19.) (n.d.). Retrieved from https://kdigo.org/wp-content/uploads/2016/10/KDIGO-2012-AKI-Guideline-English.pdf

20.) Alex. (2020, May 12). Feedforward Neural Networks and Multilayer Perceptrons. Retrieved December 11, 2020, from https://boostedml.com/2020/04/feedforward-neural-networks-and-multilayer-perceptrons.html)

21.)Bagheri, R. (2020, August 28). An Introduction to Deep Feedforward Neural Networks. Retrieved December 11, 2020, from https://towardsdatascience.com/an-introduction-to-deep-feedforward-neural-networks-1af281e306cd

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

23.) Https://www.medrxiv.org/content/10.1101/2020.04.06.20055194v1. (n.d.).

Appendix:

Table 1: Estimated Baseline Serum Creatinine Based on patient demographics

Age (years)	Black males mg/dl (μmol/l)	Other males mg/dl (μmol/l)	Black females mg/dl (μmol/l)	Other females mg/dl (μmol/l)
20–24	1.5 (133)	1.3 (115)	1.2 (106)	1.0 (88)
25–29	1.5 (133)	1.2 (106)	1.1 (97)	1.0 (88)
30–39	1.4 (124)	1.2 (106)	1.1 (97)	0.9 (80)
40–54	1.3 (115)	1.1 (97)	1.0 (88)	0.9 (80)
55–65	1.3 (115)	1.1 (97)	1.0 (88)	0.8 (71)
>65	1.2 (106)	1.0 (88)	0.9 (80)	0.8 (71)

Estimated glomerular filtration rate=75 (ml/min per 1.73 m^2)=186 × (serum creatinine [S_{Cr}]) − 1.154 × (age) − 0.203 × (0.742 if female) × (1.210 if black)=exp(5.228 − 1.154 × ln [S_{Cr}]) − 0.203 × ln(age) − (0.299 if female) + (0.192 if black).

Reprinted from Bellomo R, Ronco C, Kellum JA *et al.* Acute renal failure - definition, outcome measures, animal models, fluid therapy and information technology needs: the Second International Consensus Conference of the Acute Dialysis Quality Initiative (ADQI) Group. Crit Care 2004; 8: R204-212 with permission from Bellomo R *et al.*[22]; accessed http://ccforum.com/content/8/4/R204

Table 2 : Summary statistics of continuous variables in final data frame

Predictor	count	mean	std	min	25%	50%	75%	max
PotassiumResultAvg	2002	4.05	0.44	2.20	3.75	4.00	4.30	6.40
CalciumResultAvg	2002	8.77	0.57	6.50	8.42	8.77	9.10	12.12
ChlorideResultAvg	2002	100.84	5.26	73.18	98.00	100.67	103.33	136.00
AlbuminResultAvg	2002	3.45	0.55	1.33	3.10	3.50	3.80	5.20
WhiteBloodCellCNT	2002	7.84	6.47	0.13	5.00	6.73	9.14	180.77
age	2002	63.52	17.94	10.31	51.88	64.48	77.28	107.18
BloodPressureSystolicNBR	2002	124.70	20.88	0.00	111.00	123.00	137.00	197.00
BloodPressureDiastolicNBR	2002	69.29	11.81	0.00	61.00	69.00	77.00	109.00
TemperatureFahrenheitNBR	2002	98.08	3.10	7.00	97.60	98.10	98.60	107.20
HeartRateNBR	2002	80.20	20.45	0.00	71.00	81.00	92.00	177.00
BodyMassIndexNBR	2002	28.54	7.85	10.00	24.00	27.00	32.00	135.00
RespirationRateNBR	2002	19.72	6.85	0.00	18.00	18.00	20.00	178.00

Table 3: Proportions of AKI within specific subgroup in final data frame

Subgroup	AKI Status=1	AKI Status=0	Proportion of AKI
Gender			
Female	296	643	0.32
Male	418	645	0.39
Race			
Black or African	111	176	0.39
White (non-hispanic)	282	529	0.35
Asian (Non Hispanic)	24	43	0.36
'Hispanic Or Latino	152	294	0.34
Other_Ethnicity	145	246	0.37
Comorb			
Hypertension	462	603	0.43
Liver Disease	7	6	0.54
PeripheralVascularDisease	29	31	0.48
Admitted to ICU			
ICU	42	21	0.67

Table 7: Coefficients and p-values associated with variables selected from stepwise variable selection algorithm when fit on training data.

Predictor	Coefficient	Exponetiated Coef	P-value
AlbuminResultAvg	-0.3511909	0.703849376	7.00E-19
PotassiumResultAvg	0.71076228	2.035542321	8.90E-11
CalciumResultAvg	-0.72253519	0.485519809	2.07E-09
Hypertension	0.60627375	1.833586253	4.15E-07
ICU	0.86737037	2.380642407	8.92E-05
WhiteBloodCellCNT	0.05551621	1.057086152	0.000868
ChlorideResultAvg	0.03520191	1.035828832	0.002463
White (non-hispanic)	-0.33194182	0.717529067	0.00605
sex	0.3619527	1.436131011	0.005913
BloodPressureSystolicNBR	0.01469795	1.014806496	0.000946

Figure 1:Histogrrams of continuous predictors

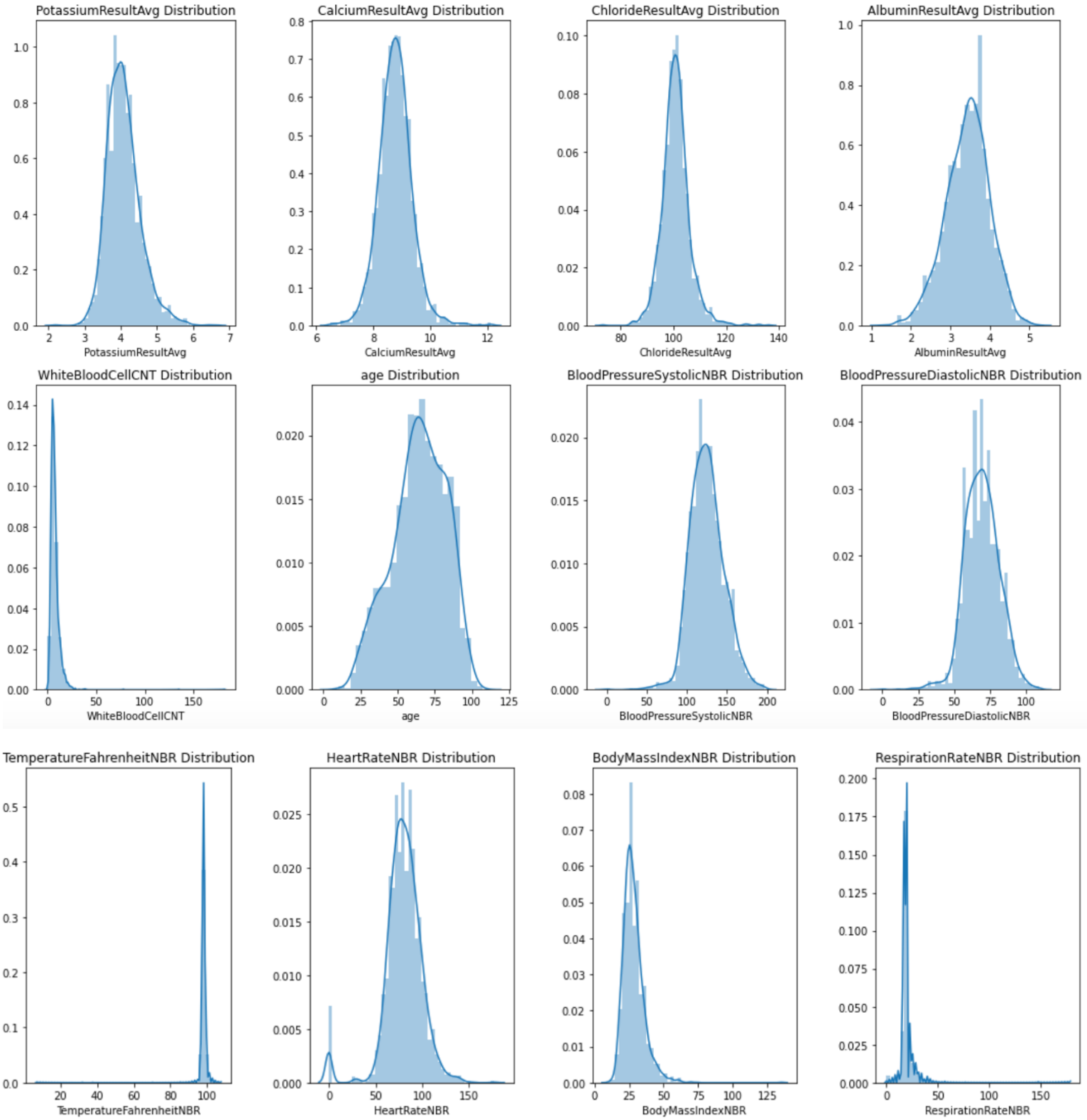


Figure 7: Effects Plots of albumin, calcium, potassium, and ICU for each model respectively