IJAS

Prediction Of Transfusion Based On Machine Learning

Praveen Kumar Donepudi[1], Naresh Babu Bynagari[2]

[1]*Enterprise Architect, Information Technology, UST-Global, Inc., Ohio, USA*
[2]*Director of Sales, Career Soft Solutions Inc, 145 Talmadge rd Edison NJ 08817, Middlesex, USA*

Abstract: The capacity to anticipate transfusions during a hospital stay may allow for more efficient blood supply management, as well as increased patient safety by assuring a sufficient supply of red blood cells (RBCs) for a specific patient. As a result, we tested the accuracy of four machine learning–based prediction algorithms for predicting transfusion, large transfusion, and the number of transfusions in hospitalized patients. Between January 2008 and June 2017, researchers conducted a retrospective observational study at three adult tertiary care institutions in Western Australia. The area under the curve for the receiver operating characteristics curve, the F1 score, and the average precision of the four machine learning algorithms used: artificial neural networks (NNs), logistic regression (LR), random forests (RFs), and gradient boosting (GB) trees were the primary outcome measures for the classification tasks. Transfusion of at least 1 unit of RBCs could be predicted quite correctly using our four prediction models (sensitivity for NN, LR, RF, and GB: 0.898, 0.894, 0.584, and 0.872, respectively; specificity: 0.958, 0.966, 0.964, 0.965). The four approaches were less successful in predicting large transfusion (sensitivity: 0.780, 0.721, 0.002, and 0.797 for ANN, LR, RF, and GB, respectively; specificity: 0.994, 0.995, 0.993, 0.995). As a result, the total number of packed RBCs transfused was likewise very inaccurately predicted. This study shows that the need for intra-hospital transfusion can be predicted with reasonable accuracy, but the number of RBC units transfused throughout a hospital stay is more difficult to predict.

Keywords: Transfusions, Artificial neural networks (ANNs), logistic regression (LR), random forests (RFs), gradient boosting (GB) trees, Support Vector Machines (SVMs)

1. INTRODUCTION

Transfusions have decreased in most developed nations as a result of the deployment of patient blood management (Shander et al., 2016; Ellingson et al., 2017) in recent years. Although most, if not all, patients benefit from PBM (Leahy et al., 2017) it is possible that particular patient populations benefit more from using PBM concepts. Patients who are most likely to require transfusions during their hospital stay are likely to benefit the most from a PBM program (Hofmann et al., 2012). These are people who have experienced bleeding or anemia as a result of their hospitalization. If the trinity of anemia, bleeding, and transfusion could be prevented or handled properly, many of them may have a better outcome (Leahy et al., 2017).

Transfusion can thus be seen as a result of the combination of anemia and blood loss in many circumstances, and can thus be used retrospectively as an indicator to identify those patients who have the greatest need to focus on bleeding, anemia, and transfusion in order to avoid each of these (Shander and Goodnough, 2010). It is feasible to use PBM for a specific purpose, and

more attention can be given to this patient group in order to avoid the trio of anemia, bleeding, and transfusion.

Moreover, accurate identification of patients who require RBC transfusions may aid blood providers in improving the hospital's blood supply (O'Donnell et al., 2018). Both too much and too little blood are either costly and unneeded, or risky for the patient's safety. As a result, anticipating transfusion needs could aid in the cost-cutting of the blood supply chain (Shih and Rajendran, 2019).

2. OBJECTIVES OF THE STUDY

Several publications (Gombotz et al., 2007; Meier et al., 2016; Vincent and Jaschinski, 2018; Bynagari, 2019) detail the parameters that influence perioperative transfusion, but only a few modest studies (Hayn et al., 2017; Donepudi, 2020b; Behrendt et al., 2020; Bynagari, 2020; Klein et al., 2017) have evaluated a multimodal, machine learning–based prediction model in a large jurisdictional cohort. As a result, the goal of this research is to evaluate a machine learning–based prediction model based on a large cohort and to cross-validate it. We believe that modern machine learning technologies can accurately forecast the need for transfusions and hence aid in identifying which patients would benefit the most from a PBM program.

3. LITERATURE REVIEW

Logistic Regression
Logistic regression, invented by David Cox in 1958, is a prominent method for solving binary and multivariate classification issues. It's called after the logistic function, as the name implies (Mahadevan et al., 2019).

$$f(x) = \frac{1}{1 + e^{-x}}$$

The sigmoid function accepts any real-valued number and returns a numeric value between 0 and 1 as its output. A TRUE value is returned if the value exceeds a specified numeric threshold (often 0.5 for probability calculations). A FALSE value is displayed if it is below the threshold. The following is the cost function that was used to change the expected output value: -

$$cost(f(x), y) = \begin{cases} -long(f(x)), & if\ y = 1 \\ -\log(1 - f(x)), & if\ y = 0 \end{cases}$$

In the medical field, logistic regression is used to predict patient severity, illness risk depending on available factors, and fatality rate. Boyde CR, for instance, used this method to develop the TRISS (Trauma and Injury Severity Score). Predicting public voting trends, determining the probability of a product failing, and gambling on mortgage defaulters are just a few of the numerous applications for this method.

Artificial Neural Networks
Artificial neural networks are modeled after the way biological brains work, which are made up of a network of interconnected neurons (Bynagari & Fadziso, 2018). The input layer feeds the model's initial input, while the output layer provides the final result. One or more hidden layers could exist between those two. Every neuron on one level is linked to every other neuron on the next (Donepudi, 2020c). When a neuron receives an input, it applies an activation function to the signal, causing the model to become nonlinear. Backpropagation is used to train the network by modifying a weight matrix, which is applied to the input signals between layers.

The diagram shows how a typical artificial neural network works at its most basic level (Mahadevan et al., 2019).

One of the most basic ANN implementations is the feedforward neural network (Neogy et al., 2018). The input signal is sent from the input layer to the output layer in a single direction in this network. A different activation function would be employed depending on the type of network used (single or multi-layer perceptron). A single layer perceptron's activation function is usually the logistic or sigmoid function.

Decision Trees

Both classification and regression issues can be solved with Decision Trees (Ganapathy et al., 2021a). It evaluates decisions and their consequences using a tree structure while taking utility costs into consideration. It's mostly utilized in decision-making challenges to find the most likely path to a positive result. It's usually depicted as a flowchart, with the goal of illustrating all of the potential implications of a given decision (Mahadevan et al., 2019).

In decision trees, there are three sorts of nodes:

i. Decision nodes – these are the nodes where the tree's flow is compared to a condition and branches are generated based on the answer.

ii. Chance nodes – These are used to depict the various outcomes that can result from a decision node. It is used to track the various aspects involved in a decision rather than evaluating the flow of the tree with a condition. Ex. High and normal humidity levels are depicted in the diagram above.

iii. End nodes – these are the tree's last terminating nodes, with no additional child nodes. They reflect the decision tree's final states.

There are two types of decision trees:

i. classification trees and

ii. Decision trees with multiple levels of decision making.

Trees of regression

CHAID (Chi squared automatic interaction detector), C4.5, CART (Classification and Regression Tree), MARS, and ID3 are some of the most prominent algorithms for decision trees. The CART method, which was created by Brieman in 1984 and can be used to generate both classification and regression trees, is one of the most prominent decision tree algorithms. The Gini Index is used to pick the splitting attribute for the binomial splitting of a feature. Only binary splits are handled by this index (Mahadevan et al., 2019).

Making and comprehending decision trees is simple. They aid in the provision of hard data-based insight into circumstances. Trees provide an expert representation of a problem statement due to the probabilities, branches, and utility costs connected with them. However, they are exceedingly costly to produce, especially for datasets with a large number of outcomes and values that are uncertain (Bynagari & Amin, 2019).

Support Vector Machines

SVMs (Mahadevan et al., 2019) are a type of supervised learning technique that can be used to solve classification and regression issues. They are officially characterized by a hyperplane that divides all of the data points in an N-dimensional space. An SVM creates a non-probabilistic binary linear classifier when it's trained using a set of instances. Multiple hyperplanes are viable alternatives for an SVM to split the data points in space (Khan et al., 2021). By increasing the distance between the data points of both classes, the best hyperplane is determined (Ahmed et al., 2021). This is done to improve the accuracy of future data point

mapping. The hyperplane's dimension depends on the number of features in the model (Ganapathy et al., 2021b). The output of logistic regression is squashed and limited to the range [0, 1] using the sigmoid function. The proper class label is assigned if the squashed value is less than or equal to a predetermined threshold. SVM, unlike logistic regression, tests if the output of a linear function is in the range [-1, 1]. This range serves as a buffer between data points that should be maximized.

The following loss function is used to optimize the margin between data points and the:

$$- c(x, y, f(x)) = (1 y f(x)) +$$

To balance the margin maximization and loss, a regularization parameter is used. The loss function is given with regularization parameter:-

$$min_{\omega} \lambda \|\omega\|^2 + \sum_{i=1}^{n}(1 - y_1 (x_1, \omega)) +$$

4. METHODS

There is only one cohort in this multicentric retrospective investigation. The PBM data system in Western Australia was used to collect the data used in this investigation. This system combines information from five key hospital information systems: patient administration, laboratory, transfusion medicine, theater management, and emergency department. The specifics of the linkage have been made public (Leahy et al., 2017). Between January 2008 and June 2017, all emergency and elective multiday stay inpatients aged 18 and older were admitted to Western Australia's three adult tertiary care facilities. Hematology, tracheostomy, general surgery, gastroenterology, orthopedics, cardiothoracic surgery, trauma, vascular surgery, urology, and cardiology were among the top ten specializations in the study, which included 73 distinct specialties.

Table 1 provides an overview of the most essential demographic and preoperative data provided. The mean values for numerical features are displayed in this table, along with standard deviations in parentheses. The data included 233,576 hospital stays from 144,419 different patients.

The following is how the data was preprocessed and cleaned:

1. Patients whose hemoglobin (Hb) levels were low at the time of admission were excluded. Patients of unknown sex and those with strange admission dates were also eliminated from the study. We now have 206, 270 stays from 131, 040 different patients.

2. Because we sought to create a predictive model at the time of admission to the hospital, we excluded variables that contained information obtained after admission.
   Additionally, any variables containing free text were deleted because they were difficult to employ in our forecast. As a result, the following 21 characteristics remained: age at admission, sex, elective/non-elective, primary diagnosis code, secondary diagnosis code, diagnosis-related group (DRG) code, Charlson Comorbidity Index (CCI), admission Hb level (g/L), had any transfusion (RBCs, fresh frozen), Charlson Comorbidity Index (CCI), Charlson Comorbidity Index (CCI), Charlson Comorbidity Index (CCI), Charlson Comorbidity Index total number of cryoprecipitate units transfused, total number of plasma units transfused, total number of platelet units transfused, total number of RBC units transfused, in hospital mortality, total number of cryoprecipitate units transfused, total number of plasma units transfused, total number of platelet units transfused, total number of RBC units transfused.

3. As transfusion consumption decreased with time, we encoded the year of admission as an extra number characteristic.

4. All numerical variables (e.g., age at admission, admission Hb, etc.) were standard normalized.
5. The primary diagnosis code, which was supplied as an International Classification of Diseases, Revision 10 code, was broken into two parts, with just the first half being utilized as an added feature.
6. The primary or secondary diagnosis code was assigned to category "minor" if it appeared in less than 0.1 percent of the records (fewer than 205 occurrences in the data set).
7. All categorical features were then one-hot encoded after that. One-hot encoding is a popular encoding method. It works by making a column for each category in the feature and assigning a 1 or 0 to indicate whether the category is present in the data. The total number of columns in the final data collection was 1357.

Table 1: Demographic data

Variable	All patients N = 206,270 (100%)	Had no RBC transfusion n = 180,614 (87.6%)	Had RBC transfusion n = 25,655 (12.4%)	Had transfusion but not mass transfusion n = 24,686 (12.0%)	Had massive transfusion n = 967 (0.4%)
Patients, n (%)					
Hospital 1	60, 245 (29.1)	53, 084 (29.3)	7, 160 (2.7)	6, 942 (28.0)	217 (22.4)
Hospital 2	29, 831 (14.4)	25, 985 (14.3)	3,845 90.4)	3, 751 (15.1)	93 (9.6)
Hospital 3	116,192 (55.3)	101, 543 (56.1)	14,648 (5.6)	13, 992 (56.6	656 (67.7)
Specialty, n (%)					
General surgery	47, 470 (22.9)	43, 410 (24.1)	4,060 (15.8)	3,736 (15.0)	322 (33.3)
General medicine	37, 509 (18.1)	32, 285 (17.8)	5, 223 (20.3)	5, 108 (20.6)	114 (11.8)
Orthopedics	29, 448 (14.2)	25, 347 (13.9)	4, 100 (15.9)	4, 053 (16.3)	46 (4.8)
Cardiology	27, 698 (13.3)	26, 158 (14.4)	1,539 (5.9)	1,444 (5.8)	94 (9.7)
Other	64, 141 (31.0)	53, 410 (29.5)	10, 730 (41.7)	10, 342 (41.9)	387 (40.0)
Age, y, median (range)	64 (47-77)	69 (55-80)	64 (46-77)	70 (55-80)	59 (75-42)
Sex, n (%)					
Female	90, 213 (43.6)	78, 624 (43.4)	11,316 (45.1)	11,587 (45.1)	270 (28.9)
Male	116, 057 (55.2)	101, 989 (56.4)	14, 067 (54.7)	13, 370 (54.1)	696 (71.9)
Charlson comorbidity	0 (0/1)	0 (0/1)	1 (0/3)	1 (0/3)	1 (0/3)

index					
Length of stay, d, median (range)	4 (3-7)	4 (2-6)	9 (2-18)	9 (5-18)	19 (9-34)
Length of stay ICU, h, median (range)	0 (0-0)	0 (0-0)	0 (0-0)	0 (0-0)	90 (24-215)
Anemia at admission					
None	120, 424	116, 192	4, 231	3, 962	268
Mild	39, 870	36, 065	3,804	3,597	338
Severe	7, 353	861	6, 491	6,338	152
Hemoglobin concentration admission, g/dL, median (range)	12.8 (11.1-14.2)	13.2 (11.6-14.4)	9.4 (7.8-11.4)	9.3 (7.8-11.4)	10.9 (8.8-12.7)
RBC transfusion, median (range)	0 (0-0)	0 (0-0)	2 (2-3)	2 (2-3)	11 (4-16)
Cryo transfusion, median (range)	0 (0-0)	0 (0-0)	0 (0-0)	0 (0-0)	4 (0-9)
FFP transfusion, median (range)	0 (0-0)	0 (0-0)	0 (0-0)	0 (0-0)	6 (2-10)
Platelet transfusion, median (range)	0 (0-0)	0 (0-0)	0 (0-0)	0 (0-0)	1 (0-3)
Complications					
Postprocedural	14.352	9, 219	5, 132	4, 654	477
Infections	2, 203	1, 277	1, 425	1, 317	107
Cardiovascular	7, 814	9, 219	4, 871	4, 506	364
Respiratory	7, 357	4, 563	4, 563	2, 555	237
Gastrointestinal	7, 814	7, 814	2, 793	2, 483	173
Genitourinary	7, 276	5, 156	2, 657	2, 373	190
Hematological	4, 349	865	3, 483	3, 252	230
Mortality (%)	2.2	1.5	6.3	5.8	18.0

Note: Demographic parameters of patients included.

As classification targets, we chose the incidence of RBC transfusion (binary result, classification) and the occurrence of large transfusion (binary outcome, classification). Massive transfusion was defined as receiving at least 10 RBCs in less than 24 hours while in the hospital. We used the model selection technique for four state-of-the-art machine learning methods in the two classification scenarios: logistic regression (LR), random forests (RFs) (Breiman, 2001) artificial neural networks (NNs) (Hopfield, 1988) and gradient boosting (GB) (Ahmed & Ganapathy, 2021). In addition, the number of RBC transfusions (an integer outcome) was predicted.

The following model selection technique was used to see if machine learning could learn these outcomes: we took 10% of the data from the cohort (20,461 rows; 13,104 unique patients) and used it to tune the parameters (20,461 rows; 13,104 unique patients). The remaining 90% of the data was utilized to create a training and test set, with fivefold cross validation used to evaluate each method's ability to generalize to previously encountered cases. If a patient had multiple hospitalizations, the training or the test set were completely to blame. To prevent overestimating the model's ability to generalize to previously unseen data, this is necessary (Donepudi, 2020a). The operation took place.

The models were assessed using the following criteria for the classification task: balanced accuracy, area under the receiver operating characteristics [ROC] curve (AUC), precision (also known as positive predictive value, recall, F1 score, and average precision (AP; the area under the precision recall curve), and average precision (AP; the area under the precision recall curve). The decision criterion was set at the harmonic mean of precision and recall that maximized the F1 score The F1 score was chosen because of its resistance to class imbalance (assigning all patients to the larger class would yield a high accuracy but a poor sensitivity). RFs, artificial NNs, Huber regression (Sun et al., 2018), and GB were also used to predict the number of RBC transfusions. For those patients who underwent at least one RBC transfusion, we forecasted the number of transfusions for each patient separately. A linear regression between the number of RBC transfusions predicted and the number of RBC transfusions actually administered was done to determine the quality of the techniques (Bynagari, 2018). The root mean squared error (RMSE) as well as the R^2 score were determined.

5. RESULTS AND DISCUSSION

Results:

The final data set of our 10-year study includes 206, 271 inpatient hospitalizations. Table 1 presents a summary of demographic statistics. The median patient age was 65 years (interquartile range: 30 years), with women accounting for 43.7 percent of the total. At least one unit of RBCs was given to 12.4 percent of the patients admitted. RBCs, cryoprecipitate, FFP, and platelets totaled 93, 375 units, 24, 662 units, and 19, 384 units, respectively. The median number of RBCs transfused was two in the group of patients who received at least one unit of RBCs. All other components (cryoprecipitate, FFP, platelets) in this group had a median of 0, showing that in most cases (Ellingson et al., 2017), RBCs were given with the goal of increasing Hb concentration rather than treating severe bleeding with accompanying coagulopathy. Massive transfusions were performed on only 0.5 percent of all patients. A median of 11 RBC units, 4 cryoprecipitate units, 6 FFP units, and 1 platelet unit were transfused in these patients.

At least 1 unit of RBC transfusion could be predicted rather reliably using our four predictive models. The AUCs were 0.966, 0.965, 0.963, and 0.966, respectively, while using NNs, LR, RFs, and GB. F1 scores were 0.749, 0.748, 0.743, and 0.755, with average precision values of 0.828, 0.820, 0.821, and 0.835. (For details see Figure 1). Overall, the Hb at admission, the patient's age, and the CCI were the most important features for predicting transfusion of at least 1 unit of RBCs. The CCI estimates a patient's one-year mortality based on a variety of comorbid illnesses.
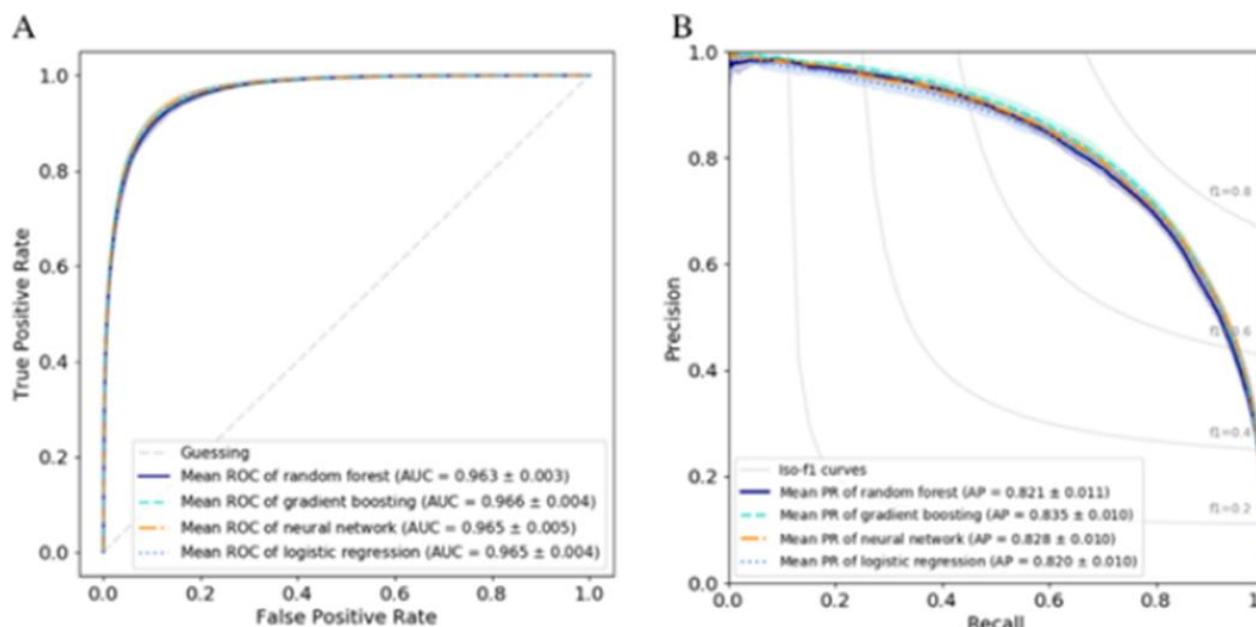
Figure 1: Transfusion of at least 1 RBC unit. Transfusion of at least 1 RBC unit. A, ROC curves for the different methods. B, Precision-recall curve for the different methods

The four techniques for forecasting big transfusions were shown to be less effective. Because of the asymmetric nature of this prediction job (Figure 2), the AUC values of the ROC analysis were relatively high (0.945, 0.949, 0.932, 0.947, respectively), but the AP values were quite low (0.162, 0.176, 0.174, 0.184, respectively). The most crucial features for transfusion prediction were no longer as evident. It's not surprise that the quantity of RBCs transfused prediction was disappointing, given that huge transfusion prediction isn't trustworthy with the features available. The R^2 score for NNs was 0.152, 0.122 for Huber regression, 0.137 for ordinal regression, 0.135 for RFs, and 0.176 for GB, respectively, with RMSE of 16.549, 17.140, 16.890, and 16.094 for each of these models, indicating that the actual number of transfusions cannot be predicted accurately with the features used.
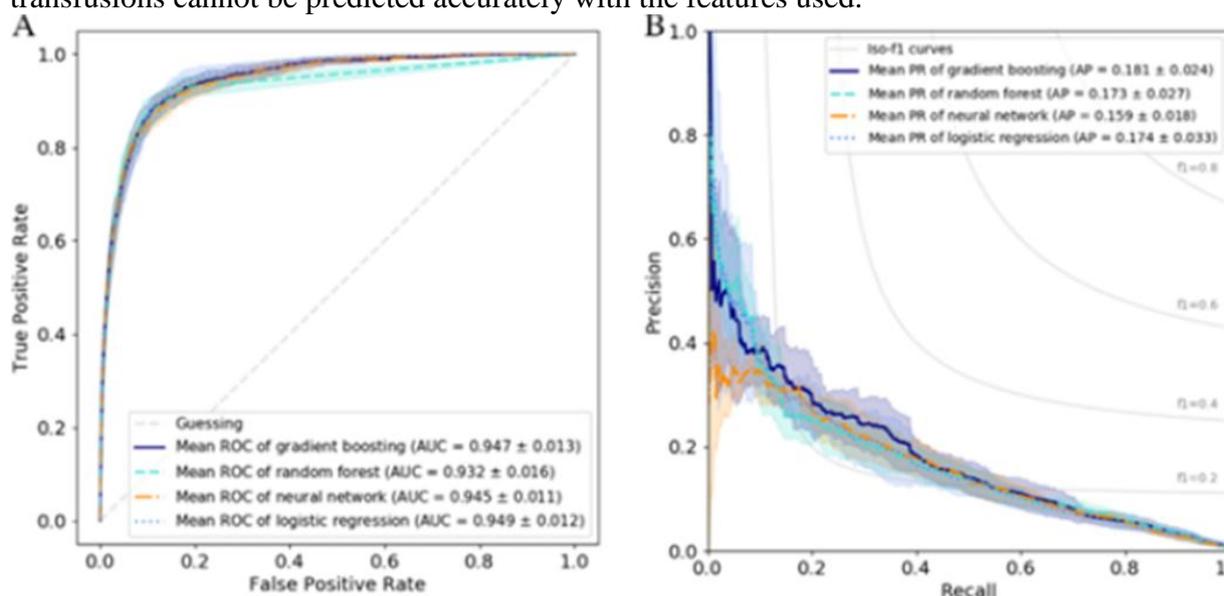

Figure 2: Massive transfusion. Prediction of massive transfusion. A, ROC curves for the different methods. B, Precision-recall curve for the different method

Discussion

Predicting the number of patients who will need blood transfusions during their stay in the hospital will be important for two reasons. First and foremost, it allows for reliable management of the allogeneic blood supply chain (O'Donnell et al., 2018; Shih and Rajendran, 2019); however, it also has the potential to help classify the risk profile of an individual patient requiring transfusion (Hayn et al., 2017), highlighting the need to implement PBM measures as thoroughly as possible in this specific patient. However, only a small number of characteristics are available at the time of hospital admission that could aid in predicting the need for transfusion later on. With a manageable number of variables provided upon hospital admission, we were able to predict which patients would require RBC transfusions using modern machine learning technologies.

We did not limit prediction to one patient group, as some other studies have done (Jo et al., 2020; Huang et al., 2018), but instead produced a model that may be utilized across a wide range of indications. However, despite the fact that this prediction has proven to be very reliable for the classification of "transfusion" vs. "no transfusion," we discovered in our database that the total number of RBC transfusions per patient and the occurrence of massive transfusion in a specific patient cannot be reliably predicted, a phenomenon that has been described in liver transplantations by other groups (Cywinski et al., 2014). This is (a) primarily due to the task's severe asymmetry, and (b) most certainly due to the fact that the characteristics carrying the necessary information are unlikely to be included in our data set. It is possible that the grounds for huge transfusion do not materialize until after the treatment procedure has begun, making it impossible to foresee massive transfusion ahead of time. Furthermore, at the time of prediction, the influencing factors for the number of RBCs required may not be known.

Modern machine learning algorithms have the potential to change prediction tasks in a variety of fields, including medicine (Rush et al., 2018). Whereas only LR models could be used to handle linear classification tasks a few years ago, numerous nonlinear relationships can now be accurately characterized using recent machine learning approaches such as decision trees or NNs. However, only when the underlying data cannot be characterized linearly can the accuracy of these modern methods greatly outperform the old approach (Couronne et al., 2018). The use of an LR model for the classification task of transfusion vs. no transfusion yielded remarkably good results in our data set.

GB had the best prediction performance, as evidenced by its high AUC, high precision recall values, and best F1 score, depending on the outcome parameters. This makes GB an important tool in our clinical scenario, albeit it should be noted that the other ways were only marginally poorer and that there is no obvious winner when it comes to the other machines method of education. GB had the best prediction performance, as evidenced by its high AUC, high precision recall values, and best F1 score, depending on the outcome parameters. This makes GB an important tool in our clinical scenario, albeit it should be noted that the other ways were only marginally poorer and that there is no obvious winner when it comes to the other machines method of education.

This mathematical selectivity can be applied to a variety of clinical cases. Initial and foremost, using our GB model, the chance of a transfusion could be estimated for every hospital stay following the first blood sample. Patients are currently assigned to a PBM program mostly in the weeks leading up to a surgical operation, whereas medical patients are frequently overlooked (Franchini et al., 2019). This strategy, however, overlooks the possibility of identifying all patients who potentially benefit from PBM beginning (including those who aren't in the midst of a surgical operation).

Surprisingly, admission type (elective or non-elective) and DRG code do not play a major impact in our model, indicating that surgery is simply one factor that influences transfusion need. Although PBM should not be viewed as a measure implemented only in a specific group of (surgical) patients, but rather as a general paradigm to treat all patients, it appears prudent to identify patients at high risk for transfusion prior to their hospital stay in order to take the necessary precautions. The Hb at admission, the patient's age, and the CCI score were revealed to be the three most relevant features in our model. Although it is well recognized that these three characteristics can play a significant impact in the clinical prediction of transfusion (Gombotz and Knotzer, 2013), we were able to acquire the best classification results utilizing recent machine learning technologies. It's worth noting that all previous articles on transfusion prediction have focused on specific clinical scenarios, primarily total hip or complete knee replacement. Despite the fact that more particular features could be employed in these publications due to the uniform preparation of these specific patient groups, the classification quality of these models is often lower than our model. We also trained LR models using the most critical features (Hb at admission, age, and CCI) to see if adding more features helped or not.

The large number of patients that might be used in the training procedure is one of our prediction model's strengths. To our knowledge, this is one of the largest data sets used in a general hospital population for such a task. We also used information from three additional sources. The majority of the other studies rely on smaller data sets from a single center in specific clinical scenarios. . As a result, our prediction model may be applied to a wide range of clinical circumstances without requiring adaptation to a new training set. Our training set is made up of hospitals in Western Australia that used PBM during the data collection period (Leahy et al., 2017). As a result, PBM measurements may have a limited role at the start of data collecting before becoming more important afterwards. As a result, we can't rule out the possibility of some time series effects over the data collection period from 2008 to 2017. Because every newly added patient will have a higher year of inclusion than all of the patients in the training set, we employed the year of admission as a feature (Donepudi, 2021). This statistic, however, most likely reflects current clinical practice, as practically all hospitals are in the midst of implementing PBM at the moment, and as a result, transfusion practices may alter over time.

As a result, we cannot rule out the possibility that our model will perform worse in predicting transfusion needs in the three Australian centers in a few years when PBM is fully implemented in all of the hospitals. We only used a small range of features that are available at the time of admission to the hospital. Our model can be transferred to other hospitals using this strategy. The disadvantage of this technique is that our model's precision is limited due to several missing variables that could aid in the identification of patients at risk for transfusion. However, our findings do not indicate that any of the anticipated transfusions may be avoided with alternate treatment. As a result, clinical applicability is limited at this time, as the benefit of any extra attempts to lower the risk of bleeding, anemia, or transfusion is unknown.

6. CONCLUSION

For our research, we used four distinct prediction models. The "gold standard" of classification tasks was LR. The more current competitors, RFs, GBs, and ANNs, were selected since they also allow out-of-the-box nonlinear modeling. In the medical field, these instruments have been shown to be quite effective at predicting outcomes. Other methods, which were not included in our study and potentially outperform our results, are available. Transfusion can be

successfully predicted at the time of hospital admission using advanced machine learning methods. The ability to forecast the number of RBCs transfused or the likelihood of large transfusion was less successful. Knowing whether patients are at risk for anemia, bleeding, and transfusion after admission could assist improve their care and result in the future.

7. REFERENCES

[1] Ahmed, A. A. A., & Ganapathy, A. (2021). Creation of Automated Content with Embedded Artificial Intelligence: A Study on Learning Management System for Educational Entrepreneurship. *Academy of Entrepreneurship Journal, 27*(3), 1-10, https://doi.org/10.5281/zenodo.4973057

[2] Ahmed, A. A. A.; Paruchuri, H.; Vadlamudi, S.; & Ganapathy, A. (2021). Cryptography in Financial Markets: Potential Channels for Future Financial Stability. Academy of Accounting and Financial Studies Journal, 25(4), 1–9. https://doi.org/10.5281/zenodo.4774829

[3] Behrendt, C. A., Debus, E. S., and Schwaneberg, T. (2020). Predictors of bleeding or anemia requiring transfusion in complex endovascular aortic repair and its impact on outcomes in health insurance claims. *J Vasc Surg.*, *71*(2), 382–389.

[4] Breiman, L. (2001). Random forests. *Mach Learn.*, *45*(1), 5–32.

[5] Bynagari , N. B. (2020). The Difficulty of Learning Long-Term Dependencies with Gradient Flow in Recurrent Nets. *Engineering International*, *8*(2), 127-138. https://doi.org/10.18034/ei.v8i2.570

[6] Bynagari, N. B. (2018). On the ChEMBL Platform, a Large-scale Evaluation of Machine Learning Algorithms for Drug Target Prediction. *Asian Journal of Applied Science and Engineering*, *7*, 53–64. Retrieved from https://upright.pub/index.php/ajase/article/view/31

[7] Bynagari, N. B. (2019). GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *Asian Journal of Applied Science and Engineering*, *8*, 25–34. Retrieved from https://upright.pub/index.php/ajase/article/view/32

[8] Bynagari, N. B., & Amin, R. (2019). Information Acquisition Driven by Reinforcement in Non-Deterministic Environments. *American Journal of Trade and Policy*, *6*(3), 107-112. https://doi.org/10.18034/ajtp.v6i3.569

[9] Bynagari, N. B., & Fadziso, T. (2018). Theoretical Approaches of Machine Learning to Schizophrenia. *Engineering International*, *6*(2), 155-168. https://doi.org/10.18034/ei.v6i2.568

[10] Couronné, R., Probst, P., and Boulesteix, A. L. (2018). Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics, 19*(1), 270.

[11] Cywinski, J. B., Alster, J. M., Miller, C., Vogt, D. P., and Parker, B. M. (2014). Prediction of intraoperative transfusion requirements during orthotopic liver transplantation and the influence on postoperative patient survival. *Anesth Analg.*, 118(2), 428–437.

[12] Donepudi, P. K. (2020a). Leveraging Cloud Computing and High Performance Computing (HPC) Advances for Next Generation Projects and Technologies. *American Journal of Trade and Policy*, *7*(3), 73-78. https://doi.org/10.18034/ajtp.v7i3.499

[13] Donepudi, P. K. (2020b). Reinforcement Learning for Robotic Grasping and Manipulation: A Review. *Asia Pacific Journal of Energy and Environment*, *7*(2), 69-78. https://doi.org/10.18034/apjee.v7i2.526

[14] Donepudi, P. K. (2020c). Robots in Retail Marketing: A Timely Opportunity. *Global Disclosure of Economics and Business*, *9*(2), 97-106. https://doi.org/10.18034/gdeb.v9i2.527

[15] Donepudi, P. K. (2021). Impact of Machine Learning in Neurosurgery: A Systematic Review of Related Literature. *Malaysian Journal of Medical and Biological Research*, *8*(1), 13-20. https://doi.org/10.18034/mjmbr.v8i1.520

[16] Ellingson, K. D., Sapiano, M. R. P., and Haass, K. A. (2017). Continued decline in blood collection and transfusion in the United States-2015. *Transfusion.* 57(suppl 2), 1588–1598.

[17] Franchini, M., Marano, G., and Veropalumbo, E. (2019). Patient blood management: a revolutionary approach to transfusion medicine. *Blood Transfus.*, *17*(3), 191–195. https://doi.org/10.2450/2019.0109-19

[18] Ganapathy, A., Ahmed, A. A. A, Siddique, M. NEA., (2021a). Easy URLs in the Content Management System with Crawlers for Added Security. Academy of Marketing Studies Journal, 25(4), 1-10. https://doi.org/10.5281/zenodo.5002945

[19] Ganapathy, A., Vadlamudi, S., Ahmed, A. A. A., Hossain, M. S., Islam, M. A. (2021b). HTML Content and Cascading Tree Sheets: Overview of Improving Web Content Visualization. *Turkish Online Journal of Qualitative Inquiry, 12*(3), 2428-2438. https://www.tojqi.net/index.php/journal/article/view/1724

[20] Gombotz, H., and Knotzer, H. (2013). Preoperative identification of patients with increased risk for perioperative bleeding. *Curr Opin Anaesthesiol.*, *26*(1), 82–90.

[21] Gombotz, H., Rehak, P. H., Shander, A. and Hofmann, A. (2007). Blood use in elective surgery: the Austrian benchmark study. *Transfusion*, *47*(8), 1468–1480.

[22] Hayn, D., Kreiner, K. and Ebner, H. (2017). Development of multivariable models to predict and benchmark transfusion in elective surgery supporting patient blood management. *Appl Clin Inform*. 8(2), 617–631.

[23] Hofmann, A., Farmer, S. and Towler, S. C. (2012). Strategies to preempt and reduce the use of blood products: an Australian perspective. *Curr Opin Anaesthesiol*. 25(1), 66–73.

[24] Hopfield, J. J. (1988). Artificial neural networks. *IEEE Circuits Devices Mag.*, *4*(5), 3–10.

[25] Huang, Z., Huang, C., and Xie, J. (2018). Analysis of a large data set to identify predictors of blood transfusion in primary total hip and knee arthroplasty. *Transfusion*, *58*(8), 1855–1862.

[26] Jo, C., Ko, S., and Shin, W. C. (2020). Transfusion after total knee arthroplasty can be predicted using the machine learning algorithm. *Knee Surg Sports Traumatol Arthrosc*. 28(6), 1757–1764.

[27] Khan, W., Ahmed, A. A. A., Vadlamudi, S., Paruchuri, H., Ganapathy, A. (2021). Machine Moderators in Content Management System Details: Essentials for IoT Entrepreneurs. Academy of Entrepreneurship Journal, 27(3), 1-11. https://doi.org/10.5281/zenodo.4972587

[28] Klein, A. A., Collier, T. and Yeates, J. (2017). The ACTA PORT-score for predicting perioperative risk of blood transfusion for adult cardiac surgery. *Br J Anaesth.* 119(3), 394–401.

[29] Leahy, M. F., Hofmann, A., and Towler, S. (2017). Improved outcomes and reduced costs associated with a health-system-wide patient blood management program: a retrospective observational study in four major adult tertiary-care hospitals. *Transfusion*, *57*(6), 1347–1358.

IJAS

[30] Mahadevan, S., Poornima, S., Tripathi, K. and Pushpalatha, M. (2019). A survey on machine learning algorithms for the blood donation supply chain. International Conference on Physics and Photonics Processes in Nano Sciences. https://doi.org/10.1088/1742-6596/1362/1/012124

[31] Meier, J., Filipescu, D. and Kozek-Langenecker, S. (2016). Intraoperative transfusion practices in Europe. *Br J Anaesth.*, *116*(2), 255–261.

[32] Neogy, T. K., & Bynagari, N. B. (2018). Gradient Descent is a Technique for Learning to Learn. *Asian Journal of Humanity, Art and Literature*, *5*(2), 145-156. https://doi.org/10.18034/ajhal.v5i2.578

[33] O'Donnell, T. F. X., Shean, K. E. and Deery, S. E. (2018). A preoperative risk score for transfusion in infrarenal endovascular aneurysm repair to avoid type and cross. *J Vasc Surg.*, *67*(2), 442–448.

[34] Rush, B., Stone, D. J., and Celi, L. A. (2018). From big data to artificial intelligence: harnessing data routinely collected in the process of care. *Crit Care Med.*, *46*(2), 345–346.

[35] Shander, A. S., and Goodnough, L. T. (2010). Blood transfusion as a quality indicator in cardiac surgery. *JAMA.*, *304*(14), 1610–1611.

[36] Shander, A., Isbister, J., and Gombotz, H. (2016). Patient blood management: the global view. *Transfusion*, 56(suppl 1), S94–S102.

[37] Shih, H., and Rajendran, S. (2019). Comparison of time series methods and machine learning algorithms for forecasting Taiwan Blood Services Foundation's blood supply. *J Healthc Eng.*, 1–6.

[38] Sun, Q., Zhou, W. and Fan, J. (2018). Adaptive Huber Regression. ArXiv170606991 Math Stat. http://arxiv.org/abs/1706.06991

[39] Vincent, J. L., and Jaschinski, U. (2018). ICON Investigators, Worldwide audit of blood transfusion practice in critically ill patients. *Crit Care.* *22*(1), 102. https://doi.org/10.1186/s13054-018-2018-9