



Mortality Prediction in Cerebral Hemorrhage Patients Using Machine Learning Algorithms in Intensive Care Units

Ximing Nie^{1,2†}, Yuan Cai^{1,2,3†}, Jingyi Liu^{1,2}, Xiran Liu^{1,2}, Jiahui Zhao^{1,2}, Zhonghua Yang^{1,2}, Miao Wen^{1,2} and Liping Liu^{1,2*}

¹ Department of Neurology, Beijing Tiantan Hospital, Capital Medical University, Beijing, China, ² China National Clinical Research Center for Neurological Diseases, Beijing, China, ³ Department of Medicine and Therapeutics, Prince of Wales Hospital, Chinese University of Hong Kong, Hong Kong, China

OPEN ACCESS

Edited by:

Long Wang,
University of Science and Technology
Beijing, China

Reviewed by:

Chao Huang,
University of Macau, China
Wang Zhongjiu,
University of Science and Technology
Beijing, China

*Correspondence:

Liping Liu
lipingsister@gmail.com

†These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Stroke,
a section of the journal
Frontiers in Neurology

Received: 26 September 2020

Accepted: 11 December 2020

Published: 20 January 2021

Citation:

Nie X, Cai Y, Liu J, Liu X, Zhao J,
Yang Z, Wen M and Liu L (2021)
Mortality Prediction in Cerebral
Hemorrhage Patients Using Machine
Learning Algorithms in Intensive Care
Units. *Front. Neurol.* 11:610531.
doi: 10.3389/fneur.2020.610531

Objectives: This study aims to investigate whether the machine learning algorithms could provide an optimal early mortality prediction method compared with other scoring systems for patients with cerebral hemorrhage in intensive care units in clinical practice.

Methods: Between 2008 and 2012, from Intensive Care III (MIMIC-III) database, all cerebral hemorrhage patients monitored with the MetaVision system and admitted to intensive care units were enrolled in this study. The calibration, discrimination, and risk classification of predicted hospital mortality based on machine learning algorithms were assessed. The primary outcome was hospital mortality. Model performance was assessed with accuracy and receiver operating characteristic curve analysis.

Results: Of 760 cerebral hemorrhage patients enrolled from MIMIC database [mean age, 68.2 years (SD, ± 15.5)], 383 (50.4%) patients died in hospital, and 377 (49.6%) patients survived. The area under the receiver operating characteristic curve (AUC) of six machine learning algorithms was 0.600 (nearest neighbors), 0.617 (decision tree), 0.655 (neural net), 0.671 (AdaBoost), 0.819 (random forest), and 0.725 (gcForest). The AUC was 0.423 for Acute Physiology and Chronic Health Evaluation II score. The random forest had the highest specificity and accuracy, as well as the greatest AUC, showing the best ability to predict in-hospital mortality.

Conclusions: Compared with conventional scoring system and the other five machine learning algorithms in this study, random forest algorithm had better performance in predicting in-hospital mortality for cerebral hemorrhage patients in intensive care units, and thus further research should be conducted on random forest algorithm.

Keywords: intracerebral hemorrhage, machine learning, mortality prediction, ICU, mimic

INTRODUCTION

Intracerebral hemorrhage (ICH) is a common neurological emergency, accounts for ~6.5–19.6% of all strokes, and is associated with higher morbidity and mortality rates compared with ischemic strokes (1). It affects ~2 million people in the world every year (2, 3).

ICH is characterized by high mortality rate, and previous studies reported that ~35% patients die within 7 days, and 50% would die within 30 days (3, 4). The global burden of care for ICH

patients is huge, especially for patients in intensive care units (ICUs). Early prediction of mortality in ICH patients is crucial for the assessment of severity of illness and adjudication of the value of novel treatments, interventions, and healthcare policies.

Several scores have been developed with the objective of predicting hospital mortality from baseline ICH patient characteristics. ICH score is one of the most commonly used scores for predicting the mortality of ICH patients (5). The score ranges from 0 to 6 and includes both clinical and radiological factors, such as Glasgow Coma Scale (GCS) score, age, infratentorial origin, ICH volume, and intraventricular hemorrhage. However, the ICH score has some limitations when used in clinical practice. The image part of ICH score needs to be assessed by experienced radiologists and neurologists; thus, it could be tricky, time consuming, and tedious for non-professional users. Acute Physiology and Chronic Health Evaluation (APACHE) II system is a widely used disease classification system in the ICU (6). It has been proven that the APACHE II can effectively predict the mortality of general ICU patients (7–10), and limited data showed that, for the ICH patients, the area under the receiver operating characteristic curve (AUC) is ~ 0.8 (11, 12).

Same as other clinical modules, these scores use conventional statistical analysis to identify the most relevant covariates from a set of features preselected by domain experts (13). However, in order to make it more convenient for clinical manual calculation, these models are usually simplified, which means that the weight of the model is discretized, and the number of covariates is artificially reduced, leading to the deterioration of the model performance.

By contrast, machine learning method allows the discovery of important variables and empirical patterns in data through automatic algorithms. Starting from the observation of the patient, the algorithm selects a large number of variables to identify the combination that can reliably predict the outcomes (14). With a variety of algorithms, machine learning can deal with variables with complex interactions without linear assumptions. In addition, another highlight of machine learning is that it can process a large number of predicted values, which enables the exploration of big data in a more comprehensive and in-depth way.

The current study was conducted in order to compare the results of multiple machine learning algorithms and conventional clinical scores for early prediction of mortality after ICH, based on initial clinical parameters, and attempts to optimize the model by improving algorithms.

MATERIALS AND METHODS

Data Resources, Patient Selection, and Variables

A retrospective multicenter study was conducted using a high-quality intensive care database, Medical Information Mart for Intensive Care (MIMIC-III) (15). MIMIC-III is a large, multicenter database containing data on patients admitted to critical care units at large tertiary care hospitals, including vital

signs, medications, laboratory measurements, observations and notes charted by care providers, fluid balance, procedure codes, diagnostic codes, imaging reports, hospital length of stay, survival data, etc. Part of the MIMIC-III database is extracted from the MetaVision system, and the other part is extracted from the CareVue system. The current analysis using data recorded within the first 24 h after ICU admission from the database was performed for the part extracted from the MetaVision system only so as to enhance data comparability. The study included patients aged ≥ 18 years and treated for ICH in the ICU during 2008–2012. The patients were identified by their ICU admission diagnosis as ICH, one of the diagnostic classifications used in the MIMIC database. Additionally, data on in-hospital mortality were obtained from the variables in the database. More details about the database can be found on the MIMIC-III website (<https://mimic.physionet.org/>).

Scores

In order to determine the performance of the proposed machine learning method, APACHE II score was used as the benchmark, which provides a general measure of disease severity based on 12 conventional physiological measurements, age, and initial values of previous health conditions. APACHE II scores were collected when the database was established.

Machine Learning Algorithms

Navicat for MySQL was used for description and visualization of MIMIC III database. A method that combines automatic algorithms and artificial selection aimed at dimension reduction was used for feature extraction from thousands of variables in this analysis. All features were selected by clinicians based on their experience in diagnosis before automatic analysis. The random forest algorithm was used for final extraction. According to the descending order of importance, the feature score higher than 0.0005 was selected for final analysis.

Multiple algorithms were chosen to improve the probability of good discrimination performance. This study used the following classifiers: nearest neighbors (NN), decision tree, neural net, AdaBoost, random forest, and gcForest, as they are the most successful and widely used models for clinical data.

Nearest Neighbors

NN classifier classifies unlabeled observations by assigning them to the most similar labeled sample class. Both the training data set and the test data set collected the characteristics of the observation data (16). In the feature space, if most of the k nearest (i.e., the nearest) samples near a sample belonged to a certain category, then the sample would be classified to that category. When it is impossible to determine which category the current point to be classified should belong to, after taking a look at its location characteristics according to the theory of statistics and measuring the weight of its neighbors, the researchers would classify (or assign) it to the category with greater weight.

Random Forest

Random forest algorithm is a machine learning method widely used in classification and regression, especially when the

TABLE 1 | Baseline characteristics of participants.

Characteristics	Overall population (n = 760)	Dead at hospital discharge (n = 383)	Alive at hospital discharge (n = 377)	P-value
Age, mean ± SD (years)	68.2 ± 15.5	68.0 ± 15.4	68.4 ± 15.5	0.67
Male, n (%)	430 (56.6)	209 (54.6)	221 (58.7)	0.26
GCS score, median (IQR)	8 (4-14)	8 (4-15)	12 (7-14)	<0.01
Eye opening	2 (1-4)	1 (1-4)	3 (1-4)	<0.01
Verbal response	1 (1-5)	1 (1-3.5)	4 (1-5)	<0.01
Motor response	5 (2-6)	4 (1-6)	6 (4-6)	<0.01
APACHE II score, median (IQR)	27 (21-33)	25 (19-32)	29 (23-34)	<0.01
Baseline systolic blood pressure, mean ± SD (mm Hg)	123.1 ± 53.7	120.5 ± 54.1	125.8 ± 53.1	0.25
Heart rate, mean ± SD (bpm)	72.1 ± 30.6	72.5 ± 30.6	71.6 ± 30.6	0.85
Respiratory rate, mean ± SD (bpm)	15.4 ± 7.1	15.7 ± 7.3	15.2 ± 7.0	0.44
Temperature, mean ± SD (F)	85.7 ± 32.6	84.9 ± 33.4	86.4 ± 31.9	0.06
WBC, mean ± SD (K/ μ L)	10.7 ± 5.7	10.6 ± 4.3	10.8 ± 6.8	0.31
Hematocrit, mean ± SD (%)	33.5 ± 10.0	31.0 ± 12.2	36.0 ± 6.1	<0.01
Hemoglobin, mean ± SD (mg/dL)	11.3 ± 3.5	10.5 ± 4.3	12.2 ± 2.2	<0.01
Chloride, mean ± SD (mEq/L)	98.2 ± 25.2	93.1 ± 33.2	103.4 ± 10.3	0.02
BUN, mean ± SD (mg/dL)	18.6 ± 15.2	20.1 ± 18.8	17.1 ± 10.2	0.01
Creatinine, mean ± SD (mg/dL)	1.1 ± 1.5	1.1 ± 1.9	1.0 ± 0.6	0.80
Glucose, mean ± SD (mg/dL)	138.1 ± 68.4	137.7 ± 72.6	138.4 ± 63.9	0.04
Sodium, mean ± SD (mEq/L)	131.0 ± 33.1	124.2 ± 43.8	138.0 ± 12.9	0.01
Potassium, mean ± SD (mEq/L)	3.7 ± 1.2	3.5 ± 1.4	3.9 ± 0.8	0.32
Troponin-T, mean ± SD (ng/mL)	0.1 ± 0.4	0.1 ± 0.5	0.1 ± 0.2	0.64
CK-MB, mean ± SD (ng/mL)	3.8 ± 10.8	4.3 ± 13.3	3.2 ± 7.5	0.41
CK, mean ± SD (IU/L)	242.1 ± 970.2	275.6 ± 1,180.2	208.2 ± 694.8	0.06

APACHE, Acute Physiology and Chronic Health Evaluation; GCS, Glasgow Coma Scale; WBC, white blood cell; BUN, blood urea nitrogen; CK, creatine kinase; SD, standard deviation.

number of potential explanatory variables is far more than the observed values (17, 18). The decision tree was built by using the method in the second section, and the set of these decision trees was random forest. Each tree would get the result of classification when predicting the data, $P_i = \{A_1, A_2, A_3, \dots, A_i \ (i = A)\}$. The prediction results of each decision tree in I would be voted, and the one with the largest number of results would be selected as the random forest prediction value.

AdaBoost

AdaBoost algorithm is a popular ensemble method, which combines several weak learners to improve the generalization performance (19). The mechanism is to first train a base learner from the training set and then to adjust the sample distribution according to the performance of the base learner so that more attention is paid to the previously divided samples and, at last, to train the next base learner based on the adjusted distribution. This process would be repeated until the number of learners reached the specified number, or the generalized error rate reaches certain requirements. Finally, the T learners were weighted and combined. The generation of the $T + 1$ learner depends on the T learner; thus, it is a serialization method of serial generation.

Decision Tree

Decision tree is a kind of tree structure, in which each internal node represents a judgment on an attribute, each branch represents the output of a judgment result, and each leaf node represents a classification result (20, 21). Decision tree is a top-down, no backtracking, and continuous search for important split variables of inductive learning algorithm. Its basic goal is to construct a concise and intuitive tree structure from a group of unordered and irregular cases under the guidance of specific learning tasks.

Training and Algorithm Optimization

The learning parameter of each algorithm was adjusted adaptively with the affinity to promote the global search ability. Accuracy and receiver operating characteristic (ROC) curve analysis were used to evaluate model identification performance. The differences were evaluated with mean AUC among the machine learning models to identify the best algorithm. The off-the-shelf methods were adopted in the Python module Scikit-Learn for the implementation of all machine learning algorithms. In order to train and validate the model, GridSearchCV method was used to optimize the performance of the model by traversing the given combination of parameters. The parameter values within the grid range were selected as the model input, and the k-fold cross-validation method was adopted to verify the accuracy

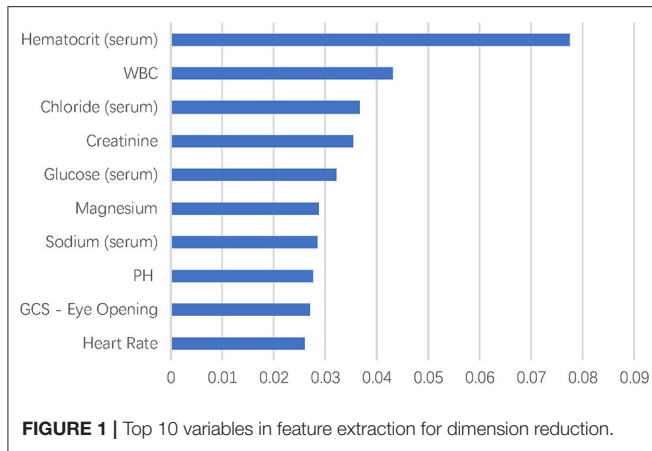


FIGURE 1 | Top 10 variables in feature extraction for dimension reduction.

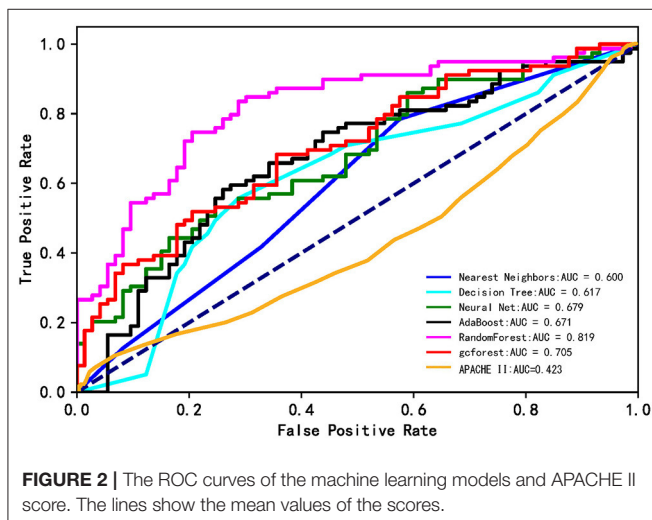


FIGURE 2 | The ROC curves of the machine learning models and APACHE II score. The lines show the mean values of the scores.

of the method. The results of K different training groups were averaged to reduce the variance; therefore, the performance of the model would be less sensitive to the partition of data. After training on each training set, the model was tested with the corresponding test set. The final parameter values were selected for the optimal model after all steps of GridSearchCV method were completed. More details of algorithms were described in **Supplementary Material**.

RESULTS

Between 2008 and 2012, 760 ICH patients who were treated in ICU, with a mean age of 68.2 years (SD, ± 15.5 years), were enrolled from MIMIC III database based on the diagnostic classifications. Among them, 583 (50.4%) patients died in hospital, and 377 (49.6%) patients survived. The median APACHE II score was 27. In the group of in-hospital death, the GCS score and APACHE II score were higher compared with the alive group. More details of baseline characteristics of the two groups are shown in **Table 1**.

TABLE 2 | Comparison of sensitivity, specificity, accuracy, and AUC of each model.

Model	Sensitivity	Specificity	Accuracy	AUC
gcForest	0.544	0.658	0.632	0.725
Nearest neighbors	0.418	0.671	0.539	0.600
Decision tree	0.557	0.712	0.610	0.617
Neural net	0.518	0.644	0.605	0.655
AdaBoost	0.556	0.753	0.645	0.671
Random forest	0.468	0.794	0.704	0.819

AUC, area under the receiver operating characteristic curve.

Feature Extraction of the Machine Learning Models

There were more than 10,000 variables in MIMIC III database. After selection by two ICU physicians, 2,023 variables were analyzed in feature extraction for machine learning. At last, 72 variables within the first 24 h after ICU admission were used for the training of the model. The top 10 most important variables for in-hospital mortality prediction are shown in **Figure 1**. The most important factor for prediction was serum hematocrit level. In addition, the top eight important variables were laboratory test results of blood biochemistry and routine blood examination of patients, including white blood cell, chloride, creatinine, glucose, magnesium, sodium, and pH. The GCS–Eye Opening score and heart rate were also affective predicting factors in the machine learning models.

Comparison Between the Models and APACHE II Score

Figure 2 shows ROC curves for hospital mortality prediction in external validation. The AUC for each modified model was as follows: 0.600 (nearest neighbors), 0.617 (decision tree), 0.655 (neural net), 0.671 (AdaBoost), 0.819 (random forest), and 0.725 (gcForest). However, the AUC was 0.423 for APACHE II score in the study population, which was much lower than that for machine learning models. All the machine learning algorithms showed better prediction efficiency compared with APACHE II score in the study population.

Table 2 shows the details of sensitivity, specificity, accuracy, and AUC of each model. Among the six machine learning algorithms, the random forest had the highest specificity and accuracy and the greatest AUC, showing the best ability to discriminate in-hospital mortality and survival.

DISCUSSION

The current study showed that compared with APACHE II score, all machine learning algorithms used for prediction showed much better prediction efficiency for ICU hospital mortality in this ICH population. Among the six machine learning algorithms used in this study, the random forest

was the best model for predicting the mortality of ICH patients treated in ICU. In addition, all the other five models showed moderate classification ability (ranging from 0.60 to 0.71). The current study could be considered an entirely novel exploration on the modified machine learning approach for hospital mortality prediction in ICH patients.

The result showed that all the machine learning algorithms had better ability to predict the mortality of ICH patients in this study, which may be explained by the fact that the patients in this study were treated in ICU and had severe neurological deficits. In the current study, most of the patients had conscious disturbance (median GCS, 8), and their condition is usually complicated and changes rapidly; thus, the traditional scores that are based on several preselected covariates by domain experts would be too simple to have enough power to make the correct mortality prediction (14), whereas the machine learning allows the analysis on a large number of variables simultaneously and can process the non-linear relations and the complex interactions among the variables (22). Another possible explanation would be that APACHE II is more suitable for the systemic failure patients, such as those who suffered from sepsis/septic shock (8, 9, 23), whereas for patients with abnormal vital signs and metabolic disorders caused by intracranial lesions, the variables in the scale do not have good predictive value.

In the current study, the random forest showed the best ability to discriminate in-hospital mortality and survival. The random forest had been widely used in the data analysis in neuroscience. It exhibited great ability to produce the best accuracy in many diseases and biological information analyses, such as genomic profiling, the corticospinal tract profile in amyotrophic lateral sclerosis, or the classification of neuroimaging data in Alzheimer disease (24–27). One of the merits of random forest is that it can avoid overfitting in the analysis of small sample size data (17, 18), which may explain why it had better performance in the current study. In addition, random forest has an important advantage that it has an intrinsic feature selection step applied before the classification task, which reduces the variable space by assigning an important value to each feature (28).

The model in the current study included several biochemical indicators that are considered to be related to prognosis in clinical practice but are difficult to quantify in traditional models. In the machine learning algorithm, the interaction among variables is considered in the selection of the important variables, to efficiently extract prediction patterns from data. It provides a solution different from traditional statistical screening of variables for the establishment of prediction model (17, 18). In the top 10 variables, most of the factors reflected the blood biochemistry and blood routine examination indexes, which was consistent with previous studies that showed electrolyte disorder, such as hypernatremia, is associated with clinical prognosis of cerebral hemorrhage (29). However, some of the important variables found in the current analysis were not explored in the ICH study before. Therefore, this study

could also provide some insights for further hypothesis-based research.

One of the highlights of the current research is that, in the machine learning model, all the selected variables were initial clinical data and electronic monitoring data that can be automatically obtained by the monitor or can be simply evaluated (such as age, gender, GCS score), which could be automatically and dynamically assessed after simple operation by users. Although compared with the ICH scores that include radiological predictors, the prediction performance in certain aspect may be compromised, this model can be completed by nurses or assistants, thereby greatly reducing the burden of clinical work for doctors. In the future, this can even be completed by artificial intelligence monitoring instruments, achieving full automation.

To interpret the findings, it must be admitted that the current research has certain limitations. The multiple-imputation method was used to handle the missing values in the analysis, which might reduce the authenticity of the data and the accuracy of both conventional scoring system and machine learning models. Complete data will be needed in follow-up studies to improve model accuracy. The data were from the MIMIC-III database, which is a non-specialist ICU database that collects data provided by not only neurologists but also other specialists, leading to the lack of some neurologic evaluation scale data. However, its simplification is conducive to its extensive promotion among non-neurologists. In addition, the prediction model in the current study contained more variables compared with the traditional scale and showed better predictive value. In the next step, the further training of the model should be conducted with expanded sample size and improved algorithm, with the aim to further improve the prediction efficiency and reduce the required variables, so as to prepare for the next step of clinical application.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: <https://archive.physionet.org/works/MIMICIIIClinicalDatabase/files/>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Health Insurance Portability and Accountability Act (HIPAA) safe harbor provision. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

LL conceived and designed the research. XN and YC acquired the data and drafted the manuscript. JL helped

to analyze the data. XL and JZ helped to perform the statistical analysis. MW and ZY made critical revision of the contribution for important intellectual content. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (8200071080) and National Natural Science Foundation of China (81820108012).

REFERENCES

- Feigin VL, Lawes CM, Bennett DA, Anderson CS. Stroke epidemiology: a review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. *Lancet Neurol.* (2003) 2:43–53. doi: 10.1016/S1474-4422(03)00266-7
- Cordonnier C, Demchuk A, Ziai W, Anderson CS. Intracerebral haemorrhage: current approaches to acute management. *Lancet.* (2018) 392:1257–68. doi: 10.1016/S0140-6736(18)31878-6
- van Asch CJ, Luitse MJ, Rinkel GJ, van der Tweel I, Algra A, Klijn CJ. Incidence, case fatality, and functional outcome of intracerebral haemorrhage over time, according to age, sex, and ethnic origin: a systematic review and meta-analysis. *Lancet Neurol.* (2010) 9:167–76. doi: 10.1016/S1474-4422(09)70340-0
- Carolei A, Marini C, Di Napoli M, Di Gianfilippo G, Santalucia P, Baldassarre M, et al. High stroke incidence in the prospective community-based L'Aquila registry (1994–1998). First year's results. *Stroke.* (1997) 28:2500–6. doi: 10.1161/01.STR.28.12.2500
- Hemphill JC 3rd, Bonovich DC, Besmertis L, Manley GT, Johnston SC. The ICH score: a simple, reliable grading scale for intracerebral hemorrhage. *Stroke.* (2001) 32:891–7. doi: 10.1161/01.STR.32.4.891
- Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med.* (1985) 13:818–29. doi: 10.1097/00003246-198510000-00009
- Basile-Filho A, Lago AF, Meneguetti MG, Nicolini EA, Rodrigues LAB, Nunes RS, et al. The use of APACHE II, SOFA, SAPS 3, C-reactive protein/albumin ratio, and lactate to predict mortality of surgical critically ill patients: a retrospective cohort study. *Medicine.* (2019) 98:e16204. doi: 10.1097/MD.00000000000016204
- Lee H, Lim CW, Hong HP, Ju JW, Jeon YT, Hwang JW, et al. Efficacy of the APACHE II score at ICU discharge in predicting post-ICU mortality and ICU readmission in critically ill surgical patients. *Anaesthesia Intensive Care.* (2015) 43:175–86. doi: 10.1177/0310057X1504300206
- Godinjak A, Igllica A, Rama A, Tančica I, Jusufović S, Ajanović A, et al. Predictive value of SAPS II and APACHE II scoring systems for patient outcome in a medical intensive care unit. *Acta Med Acad.* (2016) 45:97–103. doi: 10.5644/ama2006-124.165
- Huang J, Xuan D, Li X, Ma L, Zhou Y, Zou H. The value of APACHE II in predicting mortality after paraquat poisoning in Chinese and Korean population: a systematic review and meta-analysis. *Medicine.* (2017) 96:e6838. doi: 10.1097/MD.0000000000006838
- Fallenius M, Skrifvars MB, Reinikainen M, Bendel S, Raj R. Common intensive care scoring systems do not outperform age and glasgow coma scale score in predicting mid-term mortality in patients with spontaneous intracerebral hemorrhage treated in the intensive care unit. *Scand J Trauma Resuscit Emerg Med.* (2017) 25:102. doi: 10.1186/s13049-017-0448-z
- Pan K, Panwar A, Roy U, Das BK. A comparison of the intracerebral hemorrhage score and the acute physiology and Chronic Health Evaluation II score for 30-day mortality prediction in spontaneous

ACKNOWLEDGMENTS

The authors thank all study quality coordinators for their meticulous work on data quality control, and are grateful for the participation and engagement of all the subjects and investigators of the MIMIC-III database.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2020.610531/full#supplementary-material>

- intracerebral hemorrhage. *J Stroke Cerebrovasc Dis.* (2017) 26:2563–9. doi: 10.1016/j.jstrokecerebrovasdis.2017.06.005
- Monteiro M, Fonseca AC, Freitas AT, Pinho EMT, Francisco AP, Ferro JM, et al. Using machine learning to improve the prediction of functional outcome in ischemic stroke patients. *IEEE/ACM Trans Comput Biol Bioinformatics.* (2018) 15:1953–9. doi: 10.1109/TCBB.2018.2811471
- Obermeyer Z, Emanuel EJ. Predicting the future - big data, machine learning, and clinical medicine. *N Engl J Med.* (2016) 375:1216–9. doi: 10.1056/NEJMp1606181
- Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation.* (2000) 101:E215–20. doi: 10.1161/01.CIR.101.23.e215
- Zhang Z. Introduction to machine learning: k-nearest neighbors. *Ann Transl Med.* (2016) 4:218. doi: 10.21037/atm.2016.03.37
- Breiman L. Random forests. *Mach Learn.* (2001) 45:5–32. doi: 10.1023/A:1010933404324
- Yao C, Spurlock DM, Armentano LE, Page CD Jr, VandeHaar MJ, Bickhart DM, et al. Random Forests approach for identifying additive and epistatic single nucleotide polymorphisms associated with residual feed intake in dairy cattle. *J Dairy Sci.* (2013) 96:6716–29. doi: 10.3168/jds.2012-6237
- Breiman L. Prediction games and arcing algorithms. *Neural Comput.* (1999) 11:1493–517. doi: 10.1162/089976699300016106
- Chern CC, Chen YJ, Hsiao B. Decision tree-based classifier in providing telehealth service. *BMC Med Informatics Decision Making.* (2019) 19:104. doi: 10.1186/s12911-019-0825-9
- Honigfeld G, Klein DF, Feldman S. Prediction of psychopharmacologic effect in man: development and validation of a computerized diagnostic decision tree. *Comput Biomed Res Int J.* (1969) 2:350–61. doi: 10.1016/0010-4809(69)90020-2
- Nishi H, Oishi N, Ishii A, Ono I, Ogura T, Sunohara T, et al. Predicting clinical outcomes of large vessel occlusion before mechanical thrombectomy using machine learning. *Stroke.* (2019) 50:2379–88. doi: 10.1161/STROKEAHA.119.025411
- Choi JY, Jang JH, Lim YS, Jang JY, Lee G, Yang HJ, et al. Performance on the APACHE II, SAPS II, SOFA and the OHCA score of post-cardiac arrest patients treated with therapeutic hypothermia. *PLoS ONE.* (2018) 13:e0196197. doi: 10.1371/journal.pone.0196197
- Chen X, Wang M, Zhang H. The use of classification trees for bioinformatics. *Wiley Interdiscipl Rev Data Mining Knowl Discov.* (2011) 1:55–63. doi: 10.1002/widm.14
- Calle ML, Urrea V, Boulesteix AL, Malats N. AUC-RF: a new strategy for genomic profiling with random forest. *Human Heredity.* (2011) 72:121–32. doi: 10.1159/000330778
- Sarica A, Cerasa A, Valentino P, Yeatman J, Trotta M, Barone S, et al. The corticospinal tract profile in amyotrophic lateral sclerosis. *Human Brain Map.* (2017) 38:727–39. doi: 10.1002/hbm.23412
- Dimitriadis SI, Liparas D. How random is the random forest? Random forest algorithm on the service of structural imaging biomarkers for Alzheimer's disease: from Alzheimer's disease neuroimaging initiative (ADNI)

- database. *Neural Regen Res.* (2018) 13:962–70. doi: 10.4103/1673-5374.233433
28. Caruana R, Niculescu-Mizil A. An empirical comparison of supervised learning algorithms. In: *Proceedings of the 23rd International Conference on Machine Learning*. New York, NY: Association for Computing Machinery (2006). p. 161–8. doi: 10.1145/1143844.1143865
29. Boland T, Henderson GV, Gibbons FK, Brouwers HB, Greenberg SM, Raffeld M, et al. Hypernatremia at hospital discharge and out of hospital mortality following primary intracerebral hemorrhage. *Neurocrit Care.* (2016) 25:110–6. doi: 10.1007/s12028-015-0234-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Nie, Cai, Liu, Liu, Zhao, Yang, Wen and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.