

SCIENTIFIC REPORTS



OPEN

Comparison of machine learning models for the prediction of mortality of patients with unplanned extubation in intensive care units

Meng Hsuen Hsieh¹, Meng Ju Hsieh², Chin-Ming Chen^{3,4}, Chia-Chang Hsieh⁵, Chien-Ming Chao⁶ & Chih-Cheng Lai⁶

Unplanned extubation (UE) can be associated with fatal outcome; however, an accurate model for predicting the mortality of UE patients in intensive care units (ICU) is lacking. Therefore, we aim to compare the performances of various machine learning models and conventional parameters to predict the mortality of UE patients in the ICU. A total of 341 patients with UE in ICUs of Chi-Mei Medical Center between December 2008 and July 2017 were enrolled and their demographic features, clinical manifestations, and outcomes were collected for analysis. Four machine learning models including artificial neural networks, logistic regression models, random forest models, and support vector machines were constructed and their predictive performances were compared with each other and conventional parameters. Of the 341 UE patients included in the study, the ICU mortality rate is 17.6%. The random forest model is determined to be the most suitable model for this dataset with F_1 0.860, precision 0.882, and recall 0.850 in the test set, and an area under receiver operating characteristic (ROC) curve of 0.910 (SE: 0.022, 95% CI: 0.867–0.954). The area under ROC curves of the random forest model was significantly greater than that of Acute Physiology and Chronic Health Evaluation (APACHE) II (0.779, 95% CI: 0.716–0.841), Therapeutic Intervention Scoring System (TISS) (0.645, 95% CI: 0.564–0.726), and Glasgow Coma scales (0.577, 95% CI: 0.497–0.657). The results revealed that the random forest model was the best model to predict the mortality of UE patients in ICUs.

Acute respiratory failure is a common clinical condition in the intensive care unit (ICU), and most patients with acute respiratory failure require endotracheal intubation with mechanical ventilation (MV) support. For these patients, both endotracheal tube and mechanical ventilator are essential for life support devices. However, unplanned extubation (UE) – an accidental removal of an endotracheal tube (ETT), may develop in about 2–16% of acute respiratory failure patients requiring MV^{1–8}. Although some UE patients may have successful extubation without re-intubation, other patients experiencing UE may have severe complications, such as aspiration pneumonia, unstable hemodynamic status, airway obstruction, bronchospasm, respiratory failure, prolonged MV uses, and a prolonged length of stay in ICUs and hospitals^{9–13}. Most unfortunately, some UE patients may present a fatal outcome; the mortality rate can even be as high as 25%^{7,13,14}.

Several studies reported that tachypnea before UE, underlying uremia or liver cirrhosis, severe conditions with higher APACHE II scores, not undergoing the weaning process, reintubation, chronic neurological disease,

¹Department of Electrical Engineering and Computer Science, University of California, Berkeley, Berkeley, California, USA. ²Department of Medicine, Poznan University of Medical Science, Poznan, Poland. ³Department of Recreation and Health Care Management, Chia Nan University of Pharmacy and Science, Tainan, Taiwan. ⁴Department of Intensive Care Medicine, Chi Mei Medical Center, Tainan, Taiwan. ⁵Department of Pediatrics, China Medical University Children's Hospital, China Medical University, Taichung, Taiwan. ⁶Department of Intensive Care Medicine, Chi Mei Medical Center, Liouying, Tainan, Taiwan. Meng Hsuen Hsieh and Meng Ju Hsieh contributed equally. Correspondence and requests for materials should be addressed to C.-M.Chen. (email: chenm3383@yahoo.com.tw) or C.-C.Lai. (email: dtmed141@gmail.com)

and emergency surgery were found to be significantly associated with high mortality among UE patients^{4,15}. Since clinical assessment of illness severity is especially important for the prediction of the mortality and morbidity of critically ill patients, several researchers have attempted to design various scoring systems to determine illness severity for the prediction of patient prognosis. The Acute Physiology, Age, Chronic Health Evaluation (APACHE-II) score predicts the mortality risk for critically ill patients in the ICU¹⁶, whereas the Therapeutic Intervention Scoring System (TISS) assesses illness severity. Both systems are often used in conjunction to assess patient prognosis¹⁷. The usage of these scales, however, comes with limitations. The TISS-28 includes 28 items, divided into seven groups: basic activities, ventilatory support, cardiovascular support, renal support, neurological support, metabolic support, and specific interventions¹⁸. However, its accuracy in predicting actual mortality, alongside the APACHE II system, is limited¹⁹.

In recent years, multivariate outcome prediction models such as artificial neural networks (ANN), logistic regression models (LR), random forest models (RF), and support vector machines (SVM) have been developed in many areas of health care research^{20–25}. In this study, we aim to construct several machine learning models to predict the mortality of UE patients and compare their predicting performance with other conventional parameters.

Materials and Methods

Patients and setting. This study was conducted in eight adult ICUs of Chi-Mei Medical Center from December 1, 2008 through July 31, 2017. This is a 1288-bed tertiary medical center with 96 ICU beds: 48 medical ICU beds, 9 cardiac beds, and 39 surgical beds for adults. Every year, an average of more than 5,000 patients are admitted to the ICU. The ICU is covered by intensivists, senior residents, nurses, respiratory therapists, dietitians, physical therapists, and clinical pharmacists. Each shift had the same workload and the patient-to-nursing staff ratio is of 2:1. There were no differences in nursing experience by shift. Each respiratory therapist was responsible for fewer than 10 patients at the same time on every shift. The ICU team made rounds at least once daily, and respiratory therapists were responsible for all the weaning processes and spontaneous breathing trials of all MV patients. An UE was defined as the dislodgement or removal of the ETT from the trachea in a patient undergoing invasive MV at a time that was not specifically planned for or ordered by the physicians in charge of the patient. During the study period, a total of 341 patients experiencing UE were enrolled in this study, and their demographic and clinical information, laboratory results, comorbidities, severity scores, mortality, and length of stays for both ICU and hospital were collected for analysis. Our elderly patients (≥ 65 years) were about 58.7% (201/341) of all UE patients. The data were retrospectively collected and then analyzed. Therefore, informed consent was specifically waived and the study was approved by the Institutional Review Board of Chi Mei Medical Center (IRB: 10705–011). All methods were performed in accordance with the relevant guidelines and regulations.

Constructing data sets. All features are extracted from the original dataset. The categorical data is cleaned and one-hot encoded in RStudio. The age data, which is continuous, is unity-based normalized and standardized. After data processing, there are 16 input features. The features include subject sex, age, APACHE II scores, Glasgow Coma Scale (GCS), TISS scale, comorbidities, ICU length of stay in days, and hospital length of stay in days. These features are chosen due to their wide availability in ICUs.

Data description. The entire data set is comprised of 341 data points. The data is split into training and test sets at an approximate 9:1 ratio. 307 subjects are placed into the train set while 34 subjects are placed in the test set. In the overall dataset, the data distribution is unbalanced: ratio between patient death and survival is 1:4.683. To ensure that the output of the prediction model does not overfit the data, the data is weighted according to their outcome ratios when training the models.

Algorithm and training. The configuration of each model was reached through a hyperparameter selection process using the k -fold cross-validation accuracy ($k = 10$). We decided to use k -fold cross-validation instead of holdout cross-validation due to the limited number of subjects. The hyperparameter selection process was independently performed for each model type.

Artificial Neural Network Model. We began the model construction process by testing a three-layer perceptron network with a hidden layer that has half the number of neurons compared to the input layer. Using this configuration, we tested various activation functions in the hidden layer (Rectified Linear Unit (ReLU), Scaled Exponential Linear Unit (SeLU), and inverse tangent (\tanh)) and the output layer (Softmax and sigmoid). If the model was under-fitting, we increased the number of hidden layers and neurons. If the model was over-fitting, we added regularization or decreased the number of hidden layers and neurons. Once the model achieved high precision and recall, we applied various optimizers (Root Mean Square Propagation (RMSProp), Adam, Adadelta, Adam with Nesterov Momentum, and PowerSign) to the same architecture and determined if the model would perform better. We would decrease the learning rate if the model did not converge, but this was not found necessary.

The final ANN model consists of one input layer of 16 dimensions, a hidden layer of 24 dimensions, and an output layer of 2 dimensions. The network is trained using stochastic gradient descent and optimized using Adam with default parameters outlined by Kingma *et al.*²⁶. The neural network is trained for 200 epochs. We used the SeLU activation function at each layer and the Softmax at the output layer²⁷. Dropout regularization of 20% is applied at the input layer and 50% at the output layer²⁸. The categorical cross entropy error function for binary classification is used as the loss function. The ANN model is implemented using the Tensorflow framework (version 1.9.0)²⁹.

Logistic Regression Model. First, we used the default configuration to determine the best optimizer. We tested different solvers, which include the Newton-conjugate gradient method (Newton-CG), Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm, stochastic average gradient descent, and the liblinear solver, and compared their performances. Then, we trained models with different regularization strengths on a linear scale.

The final LR model used L_2 regularization with primal formulation. Primal formulation was used because there are more samples than features. Stochastic average gradient descent was used as the optimizer. The one-vs-rest scheme was used as the loss function. The regularization strength was set to 1.0, and the model was trained for 100 iterations before convergence. The LR software was implemented using the scikit-learn library (version 0.19.1)³⁰ and the LIBLINEAR library (version 3.21)³¹.

Random Forest Model. First, we trained the data on a single decision tree model to determine the optimal depth. We started with a depth of one and increased the depth until the model began to overfit, or when the precision and recall of the train and test sets began to diverge. Then, we trained random forest models with various number of trees. We started with one tree and increased the number of trees until the out-of-bag error did not decrease further.

The final RF model used ten separate decision tree estimators. Each decision tree used Gini impurity to measure the quality of split. The minimum number of samples required to split a node was set to two, and the minimum samples per leaf is set to one. All trees had a maximum depth of four; this was done to prevent the model from overfitting the training set. Probability estimates were used to plot the ROC curve. The RF model was implemented with the scikit-learn framework (version 0.19.2)³⁰.

Support Vector Machine Model. The SVM model is a C-support vector classification (C-SVC) model. We began by testing out various kernel types (linear, polynomial, sigmoid, and radial-basis function kernels) using the default kernel coefficient (gamma) and C value. We tested the polynomial kernel with degree three. Then, we trained different models with varying gamma and C values on a logarithmic scale.

The final SVM model used a radial basis function (RBF) as its kernel with the shrinking heuristic enabled. The model used a C value of one and a gamma value of the reciprocal of the number of features. Additionally, probability estimates were calculated in order to plot a ROC curve for the model. The SVM model was implemented using the LIBSVM library (version 3.21)³².

Statistical analyses. Mean values, standard deviations, and group sizes were used to summarize the results for continuous variables. The differences between the survival and non-survival group at hospital discharge were examined by univariate analysis with a Student *t* test and a Chi-square test. A *p* value < 0.05 was considered statistically significant. Statistical analysis of the data was done with SPSS 13.0 for Windows (SPSS, Inc., IL, USA).

Since the data distribution is unbalanced, accuracy is not a reliable measurement of prediction model performance³³. Instead, we used the weighted averaged F_1 , precision and recall values to measure model performance. These three metrics are calculated for the train set, the test set, and all data.

The Receiving Operating Characteristic (ROC) curve is also used as a metric to measure prediction model performance. The area under ROC curve (AUROC) of each prediction model was pairwise-compared using the DeLong test³⁴. The area under ROC curve of the prediction models were also compared to the those of the control predictors of the original data set, which include the APACHE II score, GCS, and TISS scale.

Results

Clinical features of UE patients. Table 1 shows the demographic and clinical characteristics of the study population. Of the 341 patients included in the study, 67.1% were male and 32.8% were female. The mean age of survivors was 64.96 years, while the mean age of non-survivors was 66.85 years. The ICU mortality rate is 17.6%. The mean APACHE II score among non-survivors is 24.23, which is significantly higher than that of the survivors, at 15.92 ($p < 0.001$). The non-survivor group had higher risks of underlying cancer, liver cirrhosis and uremia than the survivor group. The mean length of stay in ICU and hospital were 13.7 days and 36.12 days respectively for the survivor group, and 21.66 and 44.55 days respectively for the non-survivor group.

Results of Prediction Models and Control Predictors. The F_1 , precision, and recall values of all models are shown in Table 2. In the test set, the RF model has the greatest F_1 value among all models, followed by the SVM, ANN, and LR models. The RF model also has the greatest recall and precision values among all models in the test set.

The area under ROC curves of the control predictors are outlined in Table 3 and Fig. 1. The area under ROC curves of all control predictors are significantly greater than the null hypothesis area of 0.5. The area under ROC curve for all prediction models is summarized in Table 4 and Fig. 2. The RF model had the highest area under ROC curve among all prediction models. There was no significant difference between any of the prediction models using the standard 95% confidence interval criteria. However, the *p* values between the RF model and all other models ($p \approx 0.08$) were lower than the other comparisons.

The area under ROC curves were also pair-wise compared with those of the control variables. The RF model was the only model that had a significantly better prediction power than APACHE II, GCS, and TISS according to the area under ROC curve ($p < 0.0001$) (Table 5). The ANN, LR, and SVM models have a significantly better prediction power than GCS and TISS and does not have a significantly better prediction power than APACHE II (Table 6).

Variable	Survivor n = 281	Non-survivor n = 60	P value
Sex (male/female)	183/98	46/14	
Age, year	64.96 ± 16.86	66.85 ± 14.36	0.420
BMI	23.56 ± 4.33	23.47 ± 5.02	0.894
APACHE II scores	15.92 ± 7.83	24.23 ± 8.55	<0.001
Glasgow Coma scales	10.46 ± 3.73	9.50 ± 4.13	0.076
TISS scales	26.75 ± 8.00	31.00 ± 9.21	0.001
Comorbidities			
Cancer	18 (6)	12 (20)	0.001
COPD	41 (15)	7 (12)	0.554
Coronary artery disease	69 (25)	14 (23)	0.841
Liver cirrhosis	5 (2)	10 (17)	<0.001
Uremia	14 (5)	13 (22)	<0.001
Stroke	92 (33)	13 (22)	0.092
Diabetes	67 (24)	20 (33)	0.126
ICU days	13.70 ± 11.64	21.66 ± 20.79	<0.001
Hospital days	36.12 ± 26.64	44.55 ± 41.99	0.049

Table 1. The demographic and clinical characteristics of ICU patients. Data are presented as mean ± standard deviation or n (%). BMI = Body Mass Index (kg/m²). APACHE = Acute Physiology and Chronic Health Evaluation. TISS = Therapeutic Intervention Scoring System. COPD = Chronic obstructive lung disease. ICU = intensive care unit.

Dataset	Metric	ANN	LR	RF	SVM
All	F ₁	0.819	0.819	0.863	0.831
	Precision	0.846	0.848	0.888	0.848
	Recall	0.805	0.805	0.853	0.821
Train	F ₁	0.829	0.706	0.824	0.775
	Precision	0.840	0.788	0.824	0.790
	Recall	0.824	0.676	0.824	0.765
Test	F ₁	0.820	0.808	0.860	0.825
	Precision	0.844	0.842	0.882	0.842
	Recall	0.806	0.792	0.850	0.815

Table 2. F₁, Precision, and Recall values for all prediction models.

Control Variable	AUROC	SE	95% CI
APACHE II scores	0.779	0.032	0.716–0.841
Glasgow Coma Scales	0.577	0.041	0.497–0.657
TISS scales	0.645	0.041	0.564–0.726

Table 3. Area under ROC of Control Variables. AUROC = Area under ROC curve. SE = Standard Error of area under ROC curve. CI = Confidence interval.

Discussion

In terms of the F₁ value across all data, the RF model achieves the best performance among all models. The RF model is also the only model that performed significantly better than APACHE II, GCS, and TISS according to the area under ROC curve. The RF model has F₁ 0.860, precision 0.882, and recall 0.850 in the test set, and an area under ROC curve of 0.910 (SE: 0.022, 95% CI: 0.867–0.954). Therefore, in this study, we demonstrated that a random forest model is a good predictor of UE patient mortality.

We showed that when a model with a combination of multiple existing physiological scores, including APACHE II, GCS, TISS scores and eight comorbidities, are analyzed using a random forest model, the outcome of death or survival can be predicted reasonably. The usage of random forests in this study opens the feasibility for an aggregate index for a reliable patient prognosis modelling using existing scoring systems. Furthermore, the model used in this study includes features that are not included in standard APACHE II scoring such as chronic liver disease, which in Table 1 shows a significant difference between survivors and non-survivors. Ultimately, the incorporation of multiple ICU scoring systems in our proposed model shows its flexibility to extend and enhance existing frameworks for better prognosis prediction.

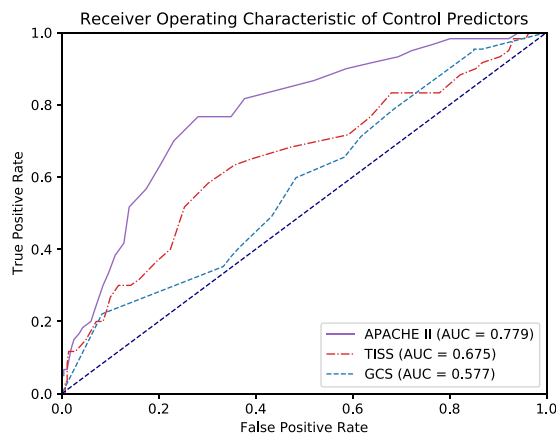


Figure 1. ROC curve of control variables.

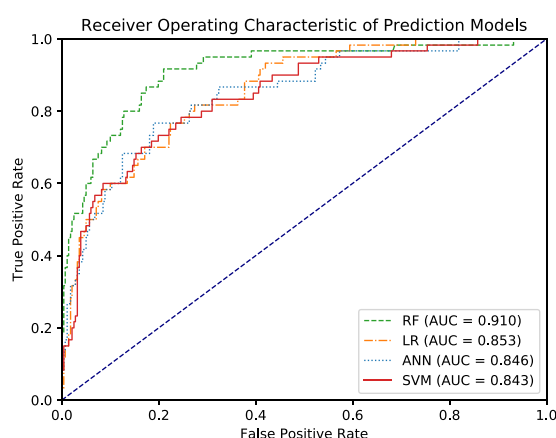


Figure 2. ROC curve of ANN, LR, RF, and SVM models.

Model	AUROC	SE	95% CI
ANN	0.846	0.028	0.791–0.902
LR	0.853	0.025	0.804–0.903
RF	0.910	0.022	0.867–0.954
SVM	0.843	0.028	0.787–0.898

Table 4. Area under ROC curve for all prediction models. AUROC = Area under ROC curve. SE = Standard Error of area under ROC curve. CI = Confidence interval.

AUROC	ANN	LR	RF	SVM
ANN	1	0.844	0.085	0.912
LR	0.844	1	0.086	0.792
RF	0.085	0.086	1	0.074
SVM	0.912	0.792	0.074	1

Table 5. Pairwise *p* values of area under ROC curves of prediction models using the DeLong test.

Several studies had tried to compare the predictive power in ICU mortality between different machine learning models^{35–37}. In a medical-neurological Indian ICU, Nimgaonkar *et al.*³⁵ showed that an ANN using 15 variables was superior to APACHE II in predicting hospital mortality ($p < 0.001$). Another study using the University of Kentucky Hospital’s data showed that the performance of ANN was as good as APACHE II in predicting ICU mortality³⁶. Though this study had similar findings in that this ANN model had a slight edge over APACHE II, TISS, and GCS, for predicting ICU mortality, this study further compared among different machine learning

Control Variable	ANN	LR	RF	SVM
APACHEII	0.085	0.063	0.002	0.101
GCS	<0.0001	<0.0001	<0.0001	<0.0001
TISS	<0.0001	<0.0001	<0.0001	<0.0001

Table 6. Pairwise p values of area under ROC curves of prediction models and the control variables using the DeLong test.

models including ANN, LR, RF, and SVM models and found that the RF model results in the highest AUROC, hence the best predictive power. It indicates that RF model may be a better machine learning method for prediction of the outcome of UE patients. In summary, while all of the tentative models can help in predicting the outcome of ICU patients, even for a specific group – UE patients; the RF model may even outperform conventional predicting tools, such as APACHE scores.

In this study, the overall mortality rate of UE patients was 17.6%, and the mortality cases had more underlying cancer, liver cirrhosis, and uremia. All these findings are consistent with the previous study⁴, and remind intensivists that they should be aware of the high mortality of UE patients, especially for the patients with multiple co-morbidities. In contrast to mortality cases, more than 80% of patients had successful extubation and survival-to-discharge. This finding might hint that the extubation was delayed in these cases with successful UE. The possible cause of delayed extubation may be due to that the final decision regarding extubation should be made by intensivists even though hospitals have standard weaning and extubation protocols.

Conclusion

The results revealed that the random forest model was the best model for predicting the mortality of UE patients in the ICUs. Such a model will be helpful for predicting ICU patients' mortality.

Limitations of our study include the lack of more patient data and features. Goodfellow *et al.* recommends that a supervised deep learning algorithm will generally achieve acceptable performance with more than 5000 data points³⁸. However, this study shows that we can still develop a good prediction model using a limited data set. Future studies can be performed to determine whether similar datasets with a larger number of samples will produce comparable results. In addition, clinical lab data, such as liver and renal function, are omitted because these features fluctuate frequently during the patient's stay. We chose features that were consistent for each patient and were known to predict patient mortality, which could yield good results according to this study.

References

- Betbese, A. J., Perez, M., Bak, E. & Mancebo, J. A prospective study of unplanned endotracheal extubation in intensive care unit patients. *Crit. Care Med.* **26**, 1180–1186 (1998).
- Boulain, T. Unplanned extubations in the adult intensive care unit: a prospective multicenter study. Association des Reanimateurs du Centre-Ouest. *Am. J. Respir. Crit. Care Med.* **157**, 1131–1137 (1998).
- Chao, C. M. *et al.* Multidisciplinary interventions and continuous quality improvement to reduce unplanned extubation in adult intensive care units: A 15-year experience. *Medicine (Baltimore)* **96**, e6877 (2017).
- Chao, C. M. *et al.* Prognostic factors and outcomes of unplanned extubation. *Sci. Rep.* **7**, 8636 (2017).
- Christie, J. M., Dethlefsen, M. & Cane, R. D. Unplanned endotracheal extubation in the intensive care unit. *J. Clin. Anesth.* **8**, 289–293 (1996).
- Coppolo, D. P. & May, J. J. Self-extubations. A 12-month experience. *Chest* **98**, 165–169 (1990).
- de Groot, R. I., Dekkers, O. M., Herold, I. H., de Jonge, E. & Arbous, M. S. Risk factors and outcomes after unplanned extubations on the ICU: a case-control study. *Crit. Care* **15**, R19 (2011).
- Vassal, T. *et al.* Prospective evaluation of self-extubations in a medical intensive care unit. *Intensive Care Med.* **19**, 340–342 (1993).
- Pandey, C. K. *et al.* Self-extubation in intensive care and re-intubation predictors: a retrospective study. *J. Indian Med. Assoc.* **100**(11), 14–16 (2002).
- Birkett, K. M., Southerland, K. A. & Leslie, G. D. Reporting unplanned extubation. *Intensive Crit. Care Nurs.* **21**, 65–75 (2005).
- Epstein, S. K., Nevins, M. L. & Chung, J. Effect of unplanned extubation on outcome of mechanical ventilation. *Am. J. Respir. Crit. Care Med.* **161**, 1912–1916 (2000).
- de Lassence, A. *et al.* Impact of unplanned extubation and reintubation after weaning on nosocomial pneumonia risk in the intensive care unit: a prospective multicenter study. *Anesthesiology* **97**, 148–156 (2002).
- Krinsley, J. S. & Barone, J. E. The drive to survive: unplanned extubation in the ICU. *Chest* **128**, 560–566 (2005).
- Phoa, L. L., Pek, W. Y., Syap, W. & Johan, A. Unplanned extubation: a local experience. *Singapore Med. J.* **43**, 504–508 (2002).
- Lee, J. H. *et al.* Clinical outcomes after unplanned extubation in a surgical intensive care population. *World J. Surg.* **38**, 203–210 (2014).
- Knaus, W. A., Draper, E. A., Wagner, D. P. & Zimmerman, J. E. APACHE II: a severity of disease classification system. *Crit. Care Med.* **13**, 818–829 (1985).
- Vincent, J. L. & Moreno, R. Clinical review: scoring systems in the critically ill. *Crit. Care* **14**, 207 (2010).
- Miranda, D. R., de Rijk, A. & Schaufeli, W. Simplified Therapeutic Intervention Scoring System: the TISS-28 items—results from a multicenter study. *Crit. Care Med.* **24**, 64–73 (1996).
- Saleh, A. A. M., Sultan, I. & Abdel-Lateif, A. comparison of the mortality prediction of different ICU scoring systems (APACHE II and III, SAPS II, and SOFA) in a single-center ICU subpopulation with acute respiratory distress syndrome. *Egyptian. J. Chest Dis. Tuberc.* **64**, 843–848 (2015).
- DiRusso, S. M., Sullivan, T., Holly, C., Cuff, S. N. & Savino, J. An artificial neural network as a model for prediction of survival in trauma patients: validation for a regional trauma area. *J. Trauma* **49**, 212–220, discussion 220–213 (2000).
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
- Hsieh, M. H. *et al.* An artificial neural network model for predicting successful extubation in intensive care units. *J. Clin. Med.* **7**, 240 (2018).
- Tu, J. V. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J. Clin. Epidemiol.* **49**, 1225–31 (1996).

24. Yang, F., Wang, H. Z., Mi, H., Lin, C. D. & Cai, W. W. Using random forest for reliable classification and cost-sensitive learning for medical diagnosis. *BMC Bioinformatics* **10**(Suppl 1), S22 (2009).
25. Furey, T. S. *et al.* Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* **16**, 906–914 (2000).
26. Kingma, D. & Adam, J. B. A method for stochastic optimization. International Conference on Learning Representations (ICLR) (2015).
27. Klambauer, G. *et al.* Self-normalizing neural networks. *Advances in Neural Information Processing Systems* (2017).
28. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Machine Learning. Res.* **15**, 1929–1958 (2014).
29. Abadi, M. *et al.* Tensor Flow: A System for Large-Scale Machine Learning. OSDI. Vol. 16 (2016).
30. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J of Machine Learning res* **12**, 2825–2830 (2011).
31. Fan, R. E., Chang, K. W., Hsieh, C. J., Wang, X. R. & Lin, C. J. LIBLINEAR: A library for large linear classification. *Journal of machine learning research* **9**, 1871–1874 (2008).
32. Chang, C. C. & Lin, C. J. LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)* **2**, 27 (2011).
33. He, H. & Garcia, E. A. Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering* **21**, 1263–1284 (2009).
34. De Long, E. R., De Long, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 837–845 (1988).
35. Nimgaonkar, A. *et al.* Prediction of mortality in an Indian intensive care unit. Comparison between APACHE II and artificial neural networks. *Intensive Care. Med.* **30**, 248–253 (2004).
36. Kim, S., Kim, W. & Park, R. W. A comparison of intensive care unit mortality prediction models through the use of data mining techniques. *Healthc. Inform. Res.* **17**, 232–243 (2011).
37. Wong, L. S. & Young, J. D. A comparison of ICU mortality prediction using the APACHE II scoring system and artificial neural networks. *Anaesthesia* **54**, 1048–1054 (1999).
38. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning*. MIT Press (2016).

Author Contributions

C.M. Chen is the guarantor of this manuscript, M.H. Hsieh, M.J. Hsieh, C.C. Hsieh, C.M. Chao and C.C. Lai contributed to the conception and design of the study, M.H. Hsieh, M.J. Hsieh and C.C. Hsieh analysed and interpreted the data, C.C. Lai, C.M. Chen and M.H. Hsieh drafted the manuscript. All authors reviewed the manuscript.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018