Heliyon 10 (2024) e30198

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Prediction analysis of TBI 24-h survival outcome based on machine learning

Yang Yang[a,b,1], Liulei Zhou[a,1], Jinhua Luo[c,1], Jianhua Xue[a], Jiajia Liu[a], Jiajia Zhang[d], Ziheng Wang[d,e,f,g,h,**], Peipei Gong[d,***], Tianxi Chen[i,*]

[a] *Department of Trauma Center, Affiliated Hospital of Nantong University, No.20 Xisi Road, Chongchuan District, Nantong City, Jiangsu Province, 226001, China*
[b] *Department of Chemistry, School of Science, China Pharmaceutical University, Nanjing, 211198, China*
[c] *Department of Anesthesia Surgery, Affiliated Hospital of Nantong University, No.20 Xisi Road, Chongchuan District, Nantong City, Jiangsu Province, 226001, China*
[d] *Department of Neurosurgery, Affiliated Hospital of Nantong University, No.20 Xisi Road, Chongchuan District, Nantong City, Jiangsu Province, 226001, China*
[e] *Clinical and Translational Research Center, Affiliated Hospital of Nantong University, No.20 Xisi Road, Chongchuan District, Nantong City, Jiangsu Province, 226001, China*
[f] *Suzhou Industrial Park Monash Research Institute of Science and Technology, Suzhou, China*
[g] *The School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia*
[h] *Centre for Precision Medicine Research and Training, Faculty of Health Sciences, University of Macau, Macau, China*
[i] *Department of Emergency Medicine, Affiliated Hospital of Nantong University, No.20 Xisi Road, Chongchuan District, Nantong City, Jiangsu Province, 226001, China*

ARTICLE INFO

Keywords:
survival
Trauma
Machine learning
RF
KNN
LR
DNN

ABSTRACT

Background: Traumatic brain injury (TBI) is the major reason for the death of young people and is well known for its high mortality and morbidity. This paper aim to predict the 24h survival of patients with TBI.
Methods: A total of 1224 samples were involved in this analysis, and the clinical indicators involved included age, gender, blood pressure, MGAP and other fields, among which the target variable was "outcome", which was a binary variable. The methods mainly involved in this paper include data visualization analysis, single factor analysis, feature engineering analysis, random forest model (RF), K-Nearst Neighbors (KNN) model, and so on. Logistic regression model (LR) and deep neural network model (DNN). We will oversample the training set using the SMOTE method because of the very unbalanced labeling of the sample itself.
Results: Although the accuracy of all models is very high, the recall rate is relatively low. The DNN model with the best performance only reaches 0.17, and the corresponding AUC is 0.80. After resampling, we find that the recall rate of positive samples of all models has increased a lot, but the AUC of some models has decreased. Finally, the optimal model is LR, whose positive sample recall rate is 0.67 and AUC is 0.82.

* Corresponding author.
** Corresponding author. Department of Neurosurgery, Affiliated Hospital of Nantong University, No.20 Xisi Road, Chongchuan District, Nantong City, Jiangsu Province, 226001, China.
*** Corresponding author.
E-mail addresses: wang.ziheng@connect.um.edu.mo (Z. Wang), ntgpp@ntu.edu.cn (P. Gong), tdfytianxi@163.com (T. Chen).
[1] These Authors contribute equally.

https://doi.org/10.1016/j.heliyon.2024.e30198
Received 21 September 2023; Received in revised form 19 April 2024; Accepted 22 April 2024
Available online 25 April 2024
2405-8440/

Conclusion: Through resampling, we obtained that the best model is the RF model, whose recall rate and AUC are the best, and the AUC level is about 0.87, indicating that the accuracy performance of the model is still good.

1. Introduction

TBI is a common and serious neurological disorder whose results can have a profound impact on patients' lives and health [1]. While acknowledging the gravity of TBI, it is crucial to delve deeper into the current landscape of TBI research, existing challenges, and identified gaps in predicting outcomes. The early prognosis prediction model (EPPM) demonstrated an accuracy of 80 %, with a sensitivity of 78.8 % and a specificity of 80.8 % in the training set [2]. Machine learning models allow for customization based on individual patient data, enhancing the ability to tailor predictions to specific characteristics and risk factors [3]. But TBI encompasses a wide range of injuries with varying degrees of severity and different underlying mechanisms. This heterogeneity poses a challenge in developing a unified predictive model [4].

To better understand and predict the outcome of TBI, we conducted a study that combined relevant patient data, such as patient metrics such as gender, age, blood pressure, and Glasgow Coma Scale(GCS), to build a predictive model [5]. In this study, we used machine learning algorithms such as LR, RF, ANN, and KNN to predict the outcome of TBI. These algorithms have different characteristics and applicability, and by comparing their performance, we can choose the best model to make predictions.

However, we face a challenge when dealing with raw data: the number of resulting labels in TBI is extremely unevenly distributed. This means that the sample size of the minority category is much smaller than the sample size of the majority category. To solve this problem, we sampled the training set using a Technique called SMOTE (Synthetic Minority Over sampling Technique). SMOTE improves the performance and predictive power of the model by generating synthetic samples to balance the uneven class distribution [6].

In this study, we used the collected patient data, including gender, age, blood pressure, GCS and other indicators as predictors, and the results of TBI as target variables. We will begin by conducting exploratory data analysis (EDA) on the data to understand the distribution, correlation, and missing values of the data [7]. We will then apply the SMOTE sampling technique to deal with the uneven label distribution and divide the data set into a training set and a test set. Next, we will train the models using machine learning algorithms such as logistic regression, random forests, artificial neural networks, and KNN, and evaluate their performance using test sets. We will compare the accuracy, recall, accuracy and F1 scores of the various models to select the best model to predict the outcome of the TBI.

Through this study, we hope to be able to provide an accurate and reliable method for predicting TBI outcomes, providing valuable information for clinical decision-making and patient management. This is of great significance for improving the treatment and rehabilitation process of TBI patients.

2. Material and methods

2.1. Preliminary understanding of data

This study was approved by the Ethics Committee of Nantong University Affiliated Hospital(Grand No. 2021-K084-01). A total of 1244 patients were involved in this modeling data, and the patient indicators mainly included 15 indicators such as age, gender, heart rate, blood oxygen, blood pressure, and MGAP.

2.2. Data exploration and analysis

First of all, we will show the missing part of the data, and set that if the missing value of the field exceeds 20 %, it is recommended to delete this field. "Diastolic blood pressure" has missing value, and the missing value is only about 0.3 %, so it is recommended to save this variable. Secondly, for variables with very sparse composition (for example, the number of 0 values in the variable accounts for more than 90 %), such fields are also recommended to be deleted (except target variables), otherwise these variables will be used as noise to interfere with data modeling.

2.3. Feature screening

Through the above steps, the data is initially normalized and individual extreme fields are eliminated. Next, we will use the machine learning method: Xgboost model performs feature importance screening, and removes variables with importance of 0, so as to conduct feature depth screening. Through the establishment of the model, we get the following ranking information of all fields with feature importance greater than 0.

3. Handling outliers

For outlier processing, here we will perform outlier processing based on the final input module data. The appearance of outliers will inevitably have a negative impact on the data, so this paper will start to do outlier processing on the data. Isolation Forest is a machine learning algorithm for detecting and removing outliers. The general steps for modeling an isolated forest model are as follows:

1 First, import the necessary libraries, such as the IsolationForest class and other data processing libraries from scikit-learn. Then, load the dataset and perform necessary data preprocessing, such as handling missing values and converting data types [7]
2 For anomaly detection, create an IsolationForest object and set relevant parameters like n_estimators (the number of trees in the forest) and contamination (the estimated proportion of outliers). Fit the IsolationForest model to the dataset using the fit() method. Next, use the predict() method to assign an outlier value score to each sample (−1 for abnormal, 1 for normal). Sort or select thresholds based on the scores to determine which samples are considered outliers.
3 In the outlier processing step, based on the result from Step 2, you have the option to handle outliers accordingly. It is important to exercise caution when deciding to delete outliers from the dataset, as it may impact the model's performance and results.
4 Finally, utilize the datasets processed for outliers to build machine learning models. Choose suitable model types and algorithms such as regression, classification, or clustering based on the specific requirements of the problem.

4. Results

4.1. Feature screening

From Fig. 1, we find that for the model, none of the fields involved in evaluation and classification have entered the ranks of feature importance, and the index "pupil reactivity" ranks first, and its importance is very high, and it is worth noting.

4.1.1. Handling outliers

Through threshold screening (the theoretical threshold is 0), we get 1181 normal samples and 63 abnormal samples, so these 63 variables need to be eliminated(Fig. 2).

4.2. Group divided

Once the data preprocessing is finished, the dataset is divided into a training set and a test set using an 80:20 split. However, to obtain a better evaluation of the model's generalization ability, random cross-validation is employed. This process involves dividing the dataset into multiple folds, such as K-folds. During each iteration, one fold is used as the validation set while the remaining folds serve as the training set. This training and evaluation process is repeated multiple times to obtain more objective results in evaluating the model.

4.3. Algorithm choose

Logistic regression (LR), a supervised learning algorithm in machine learning, is also known as the generalized linear regression

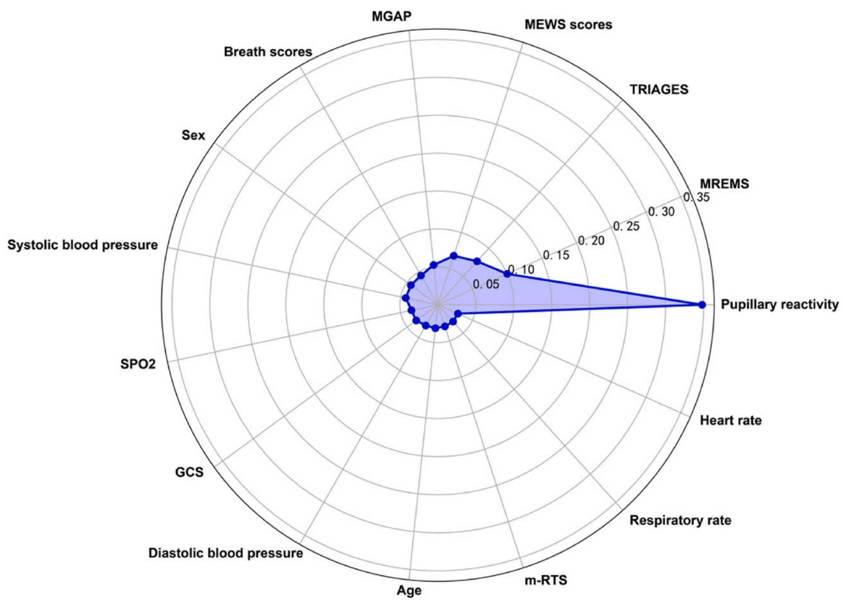

Fig. 1. Feature importance ranking.

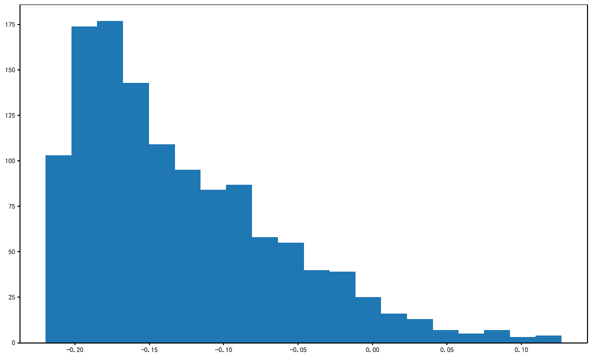

Fig. 2. Outlier test results of isolated forest.

model. LR shares many similarities with linear regression. In fact, logistic regression can be seen as a linear regression algorithm when the Sigmoid mapping function is removed. Thus, it can be said that logistic regression is theoretically supported by linear regression. Logistic regression can handle dependent variables that are binary or multi-categorical, although binary variables are more common and easier to interpret. For multi-categorical variables, the softmax method is often used for processing. In practice, binary logistic regression is the most commonly used [8].

The random forest (RF) model can be considered an extension of the decision tree algorithm, suitable for both regression and classification tasks. The basic idea behind random forest is as follows: First, bootstrap sampling is performed to extract samples from the original training set, with each sample having the same size as the original training set. Then, a decision tree model is built for each sample, producing a classification result. Finally, the classification results from each decision tree are aggregated through voting to determine the final classification of each record [9].

KNN (K-nearest neighbors) is a versatile algorithm applicable to both regression and classification tasks. Similar to other machine learning algorithms, KNN uses distance metrics to define proximity among data points based on their characteristics. For each test instance, the algorithm estimates the value of the response variable using its K nearest neighbors. The hyperparameter K controls the learning mode of the algorithm by specifying how many neighbors are considered. The estimated value is not derived from the training data itself but rather is based on the selected K nearest neighbors using a distance function [10].

An artificial neural network (ANN) is composed of multiple layers of nodes arranged in a directed graph, with each layer fully connected to the next. Each node, except for the input nodes, functions as a neuron or processing unit with a nonlinear activation function. Training ANN often involves using a supervised learning method called the backpropagation algorithm [11].

4.4. Modeling results before optimization

In this modeling, we will make use of the underlying understanding of data parameters to conduct manual fine-tuning. For example, for the KNN model, the increase of K will reduce the overfitting of the model, and the increase of max_depth of the RF model will

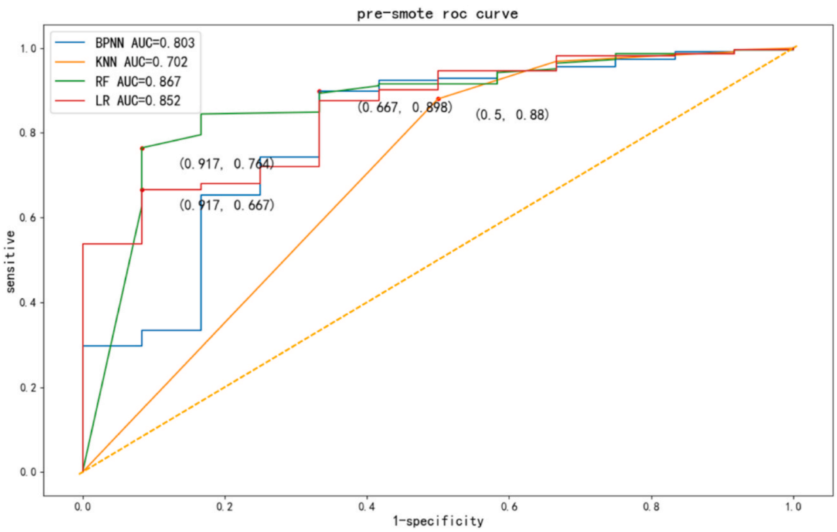

Fig. 3. ROC performance results of each model test set.

increase the overfitting of the model, and so on. Because fine-tuning data requires us to have a certain understanding of the underlying model and parameters, although we may not get the optimal parameters, we can get approximately the optimal parameters, and most importantly, we can save a lot of parameter tuning time.

After many experiments, we determined the optimal parameters for fine-tuning and obtained the following test set AUC results for each model(Fig. 3).

In addition, other indicators of the model, such as F1 and positive sample recall rate, will be presented as follows:

From Fig. 3 and Table 1, we find that the accuracy of each model is very consistent. However, in the data exploration stage, it is found that there are a lot of negative data samples (0 label), so most of the models are predicted to be 0. When the threshold is 0.5, the recall rate of positive samples of the models is very low, and the highest one is only 0.17. After comparing different prediction probability values, the model obtained the ROC curve, and we found that the best result was RF, whose AUC was 0.87. In other words, by constantly adjusting the prediction threshold, the best accuracy rate of the model was 0.87, and the overall performance of the model was good.

5. Modeling results after optimization

In order to make the results comparable, we will only balance the samples of the training set, and the test set will remain unchanged. SMOTE method was used in this paper for sample balance. After sample balancing, the same output modeling result is shown in Fig. 4.

From Fig. 4 and Table 2, we find that the loss of positive sample prediction ability of all models has been greatly improved. When the threshold value is 0.5, the positive sample recall rate of LR model can reach 0.67, and its AUC level is about 0.82, which is still good. By comparison, the AUC value of RF is the highest, reaching 0.87, but its positive sample recall rate is only 0.33 under the threshold of 0.5. In other words, although the accuracy of the model can be improved by adjusting the threshold value, the prediction degree of the positive sample is not very high. In contrast, the optimal accuracy of the LR model is indeed inferior to that of RF. However, the prediction degree of the positive sample is higher than that of RF, in other words, it can make the recognition ability of the positive sample more obvious. Here, it can be found from the sensitive optimal value of the ROC curve that the prediction accuracy of the positive sample may be more meaningful in reality. In summary, we believe that LR is the best model.

6. Discussion

Traumatic brain injury (TBI) affects approximately 70 million people globally, causing severe disruptions in their lives [12]. Prediction models can facilitate early detection of TBI, enabling timely intervention and potentially improving patient outcomes [13]. TBI is particularly concerning due to its association with high mortality rates and permanent disabilities, resulting in around 1.5 million fatalities and several million emergency procedures annually. This distressing trend is apparent in the Middle East as well, with an average TBI rate of 45 per 100,000 individuals. In emergency rooms, TBI patients face a fatality rate of 10 %, while those treated in intensive care units (ICU) have a mortality rate of 25 % [14]. Accurate prediction of the risk of death following a TBI in the early stages of treatment is of paramount importance, as it enables clinical decisions to be made with precision and healthcare resources to be allocated with efficiency [4]. Integrating predictive models into routine clinical practice requires overcoming barriers such as physician acceptance, system integration, and addressing concerns about model reliability and validity. While prediction models for Traumatic Brain Injury offer significant advantages in terms of early detection and customization, they also face challenges related to data quality, ethical considerations, and the complex nature of TBI. Ongoing research and advancements in technology are essential to address these limitations and enhance the effectiveness of predictive models in clinical settings.

In order to solve the problem of predictive analysis of TBI, the data is preprocessed and exploratory analysis is carried out. Then, Xgboost model is used for feature screening, and then outliers of data are screened based on isolated forests. Finally, comparative analysis before and after sample sampling is conducted based on various machine learning methods. Finally, based on the technical indicators and actual business indicators, the optimal model is LR, whose F1 value is 0.67 and AUC is 0.82.

This modeling is relatively complete and scientific in both the data analysis stage and the modeling analysis stage, but there are still the following points worth looking forward to: There is a serious imbalance of sample distribution in this modeling, which interferes with the modeling itself. In the future modeling, in order to improve the differentiation of each label of the model, we can appropriately increase some small number of category samples. And more hospital will be invited to improve the quality of the model.

Table 1
Main technical indicators of each model test set.

	LR	KNN	RF	ANN
accuracy	0.945	0.945	0.945	0.941
recall	0.083	0.083	0.083	0.167
precision	0.333	0.333	0.333	0.333
F1	0.133	0.133	0.133	0.222
AUC	0.852	0.702	0.867	0.803

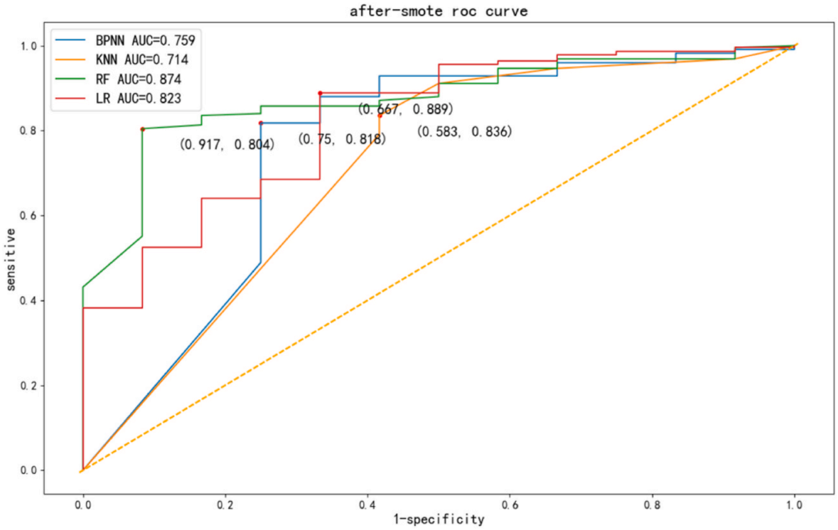

Fig. 4. ROC performance results of each model test set after training set sampling.

Table 2
Main technical index performance of each model test set after training set sampling.

	LR	KNN	RF	BPNN
accuracy	0.865	0.89	0.937	0.907
recall	0.667	0.5	0.333	0.333
precision	0.222	0.231	0.364	0.222
F1	0.333	0.316	0.348	0.267
AUC	0.823	0.714	0.874	0.759

Funding

This work was supported by the Science and Technology Project of Nantong Municipal Health Commission (NO. MS2023016) and (NO.MS2022010) and Jiangsu Provincal Research Hospital(YJXYY202204-YSB20).

Data availability statement

All data can be accessed via correspondence authors.

CRediT authorship contribution statement

Yang Yang: Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Liulei Zhou:** Writing – original draft, Validation, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Jinhua Luo:** Writing – review & editing, Writing – original draft, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Jianhua Xue:** Writing – review & editing, Writing – original draft, Validation. **Jiajia Liu:** Writing – review & editing, Writing – original draft, Formal analysis, Data curation, Conceptualization. **Jiajia Zhang:** Writing – review & editing, Writing – original draft, Conceptualization. **Ziheng Wang:** Writing – review & editing, Writing – original draft, Data curation. **Peipei Gong:** Writing – review & editing, Supervision, Software, Resources, Data curation, Conceptualization. **Tianxi Chen:** Writing – review & editing, Writing – original draft, Software, Project administration, Methodology, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:Ziheng Wang is the assocaited editor of Heliyon If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Z. Wang, Z. Lu, Y. Chen, et al., Targeting the AKT-P53/CREB pathway with epicatechin for improved prognosis of traumatic brain injury, CNS Neurosci. Ther. 30 (2) (Feb 2024) e14364, https://doi.org/10.1111/cns.14364.

[2] J.P. Hayes, M.W. Logue, N. Sadeh, et al., Mild traumatic brain injury is associated with reduced cortical thickness in those at risk for Alzheimer's disease, Brain 140 (3) (Mar 1 2017) 813–825, https://doi.org/10.1093/brain/aww344.

[3] S. Kar, R. Chawla, S.P. Haranath, et al., Multivariable mortality risk prediction using machine learning for COVID-19 patients at admission (AICOVID), Sci. Rep. 11 (1) (Jun 17 2021) 12801, https://doi.org/10.1038/s41598-021-92146-7.

[4] S. Pugazenthi, M.A. Hernandez-Rovira, R. Mitha, et al., Evaluating the state of non-invasive imaging biomarkers for traumatic brain injury, Neurosurg. Rev. 46 (1) (2023/09/08 2023) 232, https://doi.org/10.1007/s10143-023-02085-2.

[5] P.E. Rapp, B.M. Rosenberg, D.O. Keyser, et al., Patient Characterization protocols for psychophysiological studies of traumatic brain injury and post-TBI psychiatric disorders, Front. Neurol. 4 (2013) 91, https://doi.org/10.3389/fneur.2013.00091.

[6] M.U. Gul, Azman MH. Kamarul, K.A. Kadir, J.A. Shah, S. Hussen, Supervised machine learning based noninvasive prediction of atrial flutter mechanism from P-to-P interval variability under imbalanced dataset conditions, Comput. Intell. Neurosci. 2023 (2023) 8162325, https://doi.org/10.1155/2023/8162325.

[7] B. Lencha, G. Ameya, G. Baresa, Z. Minda, G. Ganfure, Intimate partner violence and its associated factors among pregnant women in Bale Zone, Southeast Ethiopia: a cross-sectional study, PLoS One 14 (5) (2019) e0214962, https://doi.org/10.1371/journal.pone.0214962.

[8] C.-Y.J. Peng, K.L. Lee, G.M. Ingersoll, An introduction to logistic regression analysis and reporting, J. Educ. Res. 96 (1) (2002/09/01 2002) 3–14, https://doi.org/10.1080/00220670209598786.

[9] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001/10/01 2001) 5–32, https://doi.org/10.1023/A:1010933404324.

[10] P. Cunningham, S.J. Delany, K-nearest neighbour classifiers - a tutorial, ACM Comput. Surv. 54 (6) (2021) 128, https://doi.org/10.1145/3459665.

[11] Q. Zheng, L. Yang, B. Zeng, et al., Artificial intelligence performance in detecting tumor metastasis from medical radiology imaging: a systematic review and meta-analysis, EClinicalMedicine 31 (Jan 2021) 100669, https://doi.org/10.1016/j.eclinm.2020.100669.

[12] G.M. S, B. Terefe, M.G. Asfaw, B. Liyew, Outcomes and associated factors of traumatic brain injury among adult patients treated in Amhara regional state comprehensive specialized hospitals, BMC Emerg. Med. 23 (1) (Sep 19 2023) 109, https://doi.org/10.1186/s12873-023-00859-x.

[13] R. Raj, T. Luostarinen, E. Pursiainen, et al., Machine learning-based dynamic mortality prediction after traumatic brain injury, Sci. Rep. 9 (1) (2019/11/27 2019) 17672, https://doi.org/10.1038/s41598-019-53889-6.

[14] M.A.K.B. Rached, J.G. Gaudet, C. Delhumeau, B. Walder, Comparison of two simple models for prediction of short term mortality in patients after severe traumatic brain injury, Injury 50 (1) (2019) 65–72, https://doi.org/10.1016/j.injury.2018.08.022.