ORIGINAL ARTICLE

The effect of well-known burn-related features on machine learning algorithms in burn patients' mortality prediction

Hilmi Yazıcı, M.D.,[1] Onur Uğurlu, Asst. Prof.,[2] Yeşim Aygül, MSc.,[3] Mehmet Yıldırım, M.D.,[4] Ahmet Deniz Uçar, M.D.,[4]

[1]Department of General Surgery, Marmara University, Pendik Training and Research Hospital, İstanbul-*Türkiye*
[2]Faculty of Engineering and Architecture, Izmir Bakircay University, İzmir-*Türkiye*
[3]Department of Mathematics, Ege University, İzmir-*Türkiye*
[4]Department of General Surgery, University of Health Sciences, Izmir Bozyaka Training and Research Hospital, İzmir-*Türkiye*

ABSTRACT

BACKGROUND: Burns is one of the most common traumas worldwide. Severely injured burn patients have an increased risk for mortality and morbidity. This study aimed to evaluate well-known risk factors for burn mortality and comparison of six machine learning (ML) Algorithms' predictive performances.

METHODS: The medical records of patients who had burn injuries treated at Izmir Bozyaka Training and Research Hospital's Burn Treatment Center were examined retrospectively. Patients' demographics such as age and gender, total burned surface area (TBSA), Inhalation injury (II), full-thickness burns (FTBSA), and burn types (BT) were recorded and used as input features in ML models. Patients were analyzed under two groups: Survivors and Non-Survivors. Six ML algorithms, including k-Nearest Neighbor, Decision Tree, Random Forest, Support Vector Machine, Multi-Layer Perceptron, and AdaBoost (AB), were used for predicting mortality. Several different input feature combinations were evaluated for each algorithm.

RESULTS: The number of eligible patients was 363. All six parameters (TBSA, Gender, FTBSA, II, Age, BT) that were included in ML algorithms showed a significant difference ($p<0.001$). The results show that AB algorithm using all input features had the best prediction performance with an accuracy of 90% and an area under the curve of 92%.

CONCLUSION: ML algorithms showed strong predictive performance in burn mortality. The development of an ML algorithm with the right input features could be useful in the clinical practice. Further investigations are needed on this topic.

Keywords: Burn; machine learning; mortality; prediction.

INTRODUCTION

Burn injuries are one of the most common traumas worldwide. According to the latest data, almost 500.000 people affected from burn injuries in the USA in 2016, and over 3000 of them died.[1] Especially for severe burn patients, early estimation of mortality helps clinicians during hospitalization. To determine the increased risk of mortality in the early period of burn treatment, we had to know well-described risk factors. Age and total burned surface area (TBSA) are the most important risk factors for burn mortality, as shown in recent studies.[2-4] In addition to these, inhalation injury, burn depth, and gender are also related to mortality.[5,6] However, these parameters' role in mortality is still controversial. Nevertheless, mortality prediction models were developed by using some or all of these parameters. Abbreviated burn severity index, revised Baux, Fatality by Longevity, APACHE (II) score, Measured Extent of burn, and Sex (FLAMES) are some of the

Cite this article as: Yazıcı H, Uğurlu O, Aygul Y, Yıldırım M, Uçar AD. The effect of well-known burn-related features on machine learning algorithms in burn patients' mortality prediction. Ulus Travma Acil Cerrahi Derg 2023;29:1130-1137.

Address for correspondence: Hilmi Yazıcı, M.D.
Marmara University, Pendik Training and Research Hospital, İstanbul, Türkiye
E-mail: hilmiyazici@hotmail.com

Ulus Travma Acil Cerrahi Derg 2023;29(10):1130-1137 DOI: 10.14744/tjtes.2023.79968 Submitted: 25.05.2023 Revised: 26.05.2023 Accepted: 14.08.2023

most well-known.[7-9] Although these parameters have a robust estimation of mortality, most of them are hard to use during clinical practice. Therefore, simple instruments are needed to use in clinical practice.

Machine learning (ML) methods, a branch of artificial intelligence, can be used to predict the non-linear interactions between features. Unlike the traditional techniques, it does not require a priori knowledge about the data. Instead, ML models use algorithms to extract model-like "structure" information from a given data. In recent years, these algorithms were started to be used in different parts of medical sciences.[10,11] Many researchers have used ML for burn care.[12] Especially in the last decade, researchers began to use ML algorithms for the survival prediction of burns patients.[13]

This study evaluates well-known risk factors in a single tertiary burn center's patient population. It investigates the predictive performances of ML algorithms on severe burn patients' mortality, including K-Nearest Neighbors (K-NN), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), and AdaBoost (AB).

MATERIALS AND METHODS

Data regarding patients hospitalized at Izmir Bozyaka Training and Research Hospital burn treatment center between January 2019 and December 2021 were retrospectively analyzed. Patients with burns and who were over age 18 years were included in the study. Patients with isolated carbon monoxide inhalation and patients below age 18 were excluded from the study.

Demographics including age, gender, TBSA, burn types (BT), full-thickness burns (FTBSA), partial thickness burned surface area (PTBSA), II, mean arterial pressure (MAP), and creatinine levels were recorded. BT was examined under four types: Flame, Scald, Electrical, and, Chemical burns. Laryngeal Endoscopic Imaging with a flexible fiberoptic endoscope and physical examination were used to determine the presence of inhalation injury. In addition, Vitaly parameters and laboratory values at the time of hospitalization were noted. TBSA was calculated according to the Lund–Browder chart.[14]

This study was approved by the Ethics Committee of the Izmir Bozyaka Training and Research Hospital (decision date: March 09, 2022 no: 2022/44)

All methods were performed in accordance with the relevant guidelines and regulations.

Primary Outcomes

Comparison of different ML algorithms on prediction of mortality in a tertiary burn center's patient population.

Secondary Outcomes

Adaptation of ML algorithms in the clinical medical practice of burn centers.

ML Prediction Models

This section summarizes the ML algorithms used in this study and describes the tuning parameters of each algorithm. Besides, we explained the performance metrics used to evaluate these algorithms.

The experimental dataset of this study includes 363 samples with six input features (Age, Gender, TBSA, BT, FTB, and II) and one output parameter (status). Six ML algorithms were evaluated in the burn injury dataset. These ML algorithms were included: k-NN, DT, RF, SVM, MLP, and AB. Scikit-learn, one of the widely used Python libraries for ML, were used to perform these algorithms. In addition, cross-validation studies were also performed for k-NN, DT, RF, SVM, MLP, and AB algorithms using k-fold method.

One of the most crucial factors in the predictive performance of ML algorithms is the hyperparameters of the algorithms. To determine the optimal hyperparameters of the algorithms, a grid search technique was performed within the Scikit-learn framework. This technique searches stored parameters and gives the best parameters for high predictive accuracy. Using this approach, we empirically assessed and compared all ML algorithms' performance.

Performance Evaluation Metrics

Different metrics based on the confusion matrix are used to evaluate the prediction performance of the ML algorithms. (Abbreviations: TP, true positive; TN, true negative; FP, false positive; FN, false negative)

Accuracy is the ratio of the correct predictions of the model to the whole dataset. It defines how the model performs in all classes.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Sensitivity is the proportion of true positives. It allows detection of how many positive samples it can correctly identify.

$$Sensitivity = \frac{TP}{TP + FN}$$

Specificity measures the rate of correctly identifying true negatives. High specificity means that the model correctly identifies most negative outcomes.

$$Specificity = \frac{TN}{TN + FP}$$

Positive predictive value (PPV) is the ratio of true positives versus all positives identified by the test.

$$PPV = \frac{TP}{TP + FP}$$

In addition to these parameters, we also used the receiver operator characteristic (ROC) curve to compare the performance of the ML algorithms by evaluating the area under the curve (AUC).

ML Algorithms

k-NN is a well-known ML algorithm that can be used for both classification and regression. In classification, every point is assigned to the data class with the most representatives among its nearest neighbors. The classification performance of k-NN is based on the number of neighbors k and the predefined metric. The setting parameters of the k-NN are the number of neighbors optimized in {1, 2, 3, ..., 20} and the distance metrics optimized in {minkowski, euclidean, manhattan}.

A DT, which consists of decision nodes and leaf nodes, represents a segmentation of the data. Decision nodes have two or more branches, each representing the tested attribute values, and leaf nodes decide the class of the point. The setting parameters of the DT are the maximum depth of the tree optimized in {3, 4, 5, …, 8}, the function to measure the quality of a split optimized in {gini, entropy}.

SVM aims to separate data by a hyperplane with a maximum gap between the two outcome classes. While computationally challenging, it can be useful in the clinical domain. The tuning parameter of the kernel type optimized in {linear, poly, sigmoid, rbf}. In addition, the SVM model incorporated a linear basis function kernel technique that could give a better classification.

MLP is a layered neural network with feedforward. In MLP, information flows unidirectionally from the input layer to the output layer, passing through the hidden layers. The tuning parameters of the MLP are the number of neurons in the hidden layer optimized in {1, 2, 3, ..., 20}, the activation function of the hidden layer optimized in {identity, logistic, tanh, relu}, and learning rate schedule for weight updates optimized in {constant, invscaling, adaptive}.

RF is an ensemble learning method that constructs several DTs to arrive at a solution. RF works efficiently on large datasets, and the risk of overfitting is relatively low. However, it trains the dataset slowly. The setting parameters of the RF are the number of trees optimized in {100, 200, 300, …, 800}, the function to measure the quality of a split optimized in {gini, entropy}, and the maximum depth optimized in {3, 4, 5, 6, 7}.

AB is another ensemble learning method that adjusts the weights of incorrectly classified samples so that subsequent classifiers focus on more complex cases. AB is less prone to overfitting as the input parameters are not jointly optimized. The tuning parameters of the AB are the maximum number of estimators optimized in {10, 20, 30, ..., 120} with the Logistic Regression as the base estimator.

Figure 1 compares the six ML algorithms and illustrates conceptual drawings serving as examples.

Statistical Analysis

SPSS version 24.0 (Spss Inc., IBM, Chicago, US) was used for

Table 1. Patient demographics

n=363	Survivors (n=260)	Non-survivors (n=103)	P-value
Age (median) (Min–Max)	37 (18–82)	58 (18–89)	<0.001
Gender (%)			
Male	221 (85%)	66 (64%)	<0.001
Female	39 (15%)	37 (36%)	
TBSA (%)(Median)	20% (1–78)	51% (5–100)	<0.001
Burn types			
Flame	162	89	
Scald	62	8	<0.001
Electrical	25	6	
Chemical	11	0	
Full-thickness burn	123 (47%)	89 (86%)	<0.001
PTBSA (%) (Median) n=301	13% (1–59) N: 225	13.5% (1–78) N: 76	0.245
FTBSA (%) (Median) n=223	9% (1–45) N: 124	30% (4–90) N: 89	<0.001
Inhalation Injury	22 (8%)	45 (44%)	<0.001
MAP (mmHg)	95 (47–141)	90 (35–145)	0.001
Creatinine (mg/dL)	0.9	1.1	<0.001

TBSA: Total burned surface area; PTBSA: Partial-thickness burned surface area; FTBSA: Full-thickness burned surface area; MAP: Mean arterial pressure.

statistical analysis. Data on quantitative variables are presented as median and minimum–maximum and frequencies for qualitative variables. We used the Mann–Whitney U-test for continuous data and the $\chi 2$ test or Fisher's exact test for categorical data. Univariate analysis was performed to compare patients who survived with non-survivors. ROC analysis was also performed to compare different ML algorithms' determination performances.

RESULTS

A total number of 363 patients were included in the study. The median age of the entire cohort was 41 (18–89). The majority of the patients had male gender (260–72%). The median TBSA of patients was 26 (1–100). The patient population was examined under two groups: Survivors and Non-Survivors. Patient demographics among the two groups were summarized in Table 1. The Non-Survivor Group had a higher median age than the Survivor Group (respectively; 58 [18–89] vs. 37 [18–82], P<0.001). Female patients had statistically significantly higher mortality rates than male patients (P<0.001). The median TBSA of the Non-Survivor Group was higher than the Survivor Group as expected (respectively; 51% [5–100] vs. 20% [1–78], P<0.001). BT, presence of FTB, FTBSA, II, MAP, and creatinine level were significantly different between the two groups. PTBSA had no statistically significant difference between the two groups.

All ML algorithms were performed with the 10-fold cross-validation technique. In addition, hyper-parameters of the algorithms were optimized through the grid search process. For k-NN, the best performance was achieved for k = 9 with the Manhattan distance metric. For DT, the entropy function with a maximum depth of 5 gave the highest accuracy. Linear kernel type had the best performance for SVM. For MLP, the highest accuracy was obtained when the number of neurons equals 15 with an identity activation function and constant learning rate. For RF, an optimal tree number of 100 and a depth of 7 were identified with the entropy function. Finally, the best-performing AB was comprised of 10 estimators. Note that different hyperparameters were determined for all different feature combinations. Here, we reported the hyperparameters of the model, which used all the features since it achieved the best prediction performance.

Table 2 summarizes the mean accuracy for the ML algorithms with differing combinations of the features. The results show that all algorithms have comparable performance on the dataset. The highest prediction accuracy and AUC were noted within the AB algorithm using all features (90% and 92%, respectively). The average accuracy and AUC of all algorithms using all features are 88% and 92%, respectively. Excluding BT and retaining the other features did not change the average scores of the algorithms. Similar to BT, excluding gender, FTB, and inhalation injury separately did not significantly change the average scores of algorithms. However, in the absence of age, the average scores of the algorithms reduced average accuracy to 85% and average AUC to 86%. Similar

Table 2. Mean accuracies of algorithms

Mean (SD) accuracy (%) feature combination	k - NN	DT	RF	SVM	MLP	AB	Mean
Age, Gender, TBSA, BT, FTB, II	88 (4)	89 (5)	86 (4)	89 (4)	88 (6)	90 (4)	88.3
Gender, TBSA, BT, FTB, II	83 (5)	85 (6)	87 (6)	85 (8)	86 (8)	84 (9)	85
Age, TBSA, BT, FTB, II	88 (4)	87 (6)	88 (4)	89 (4)	87 (5)	88 (4)	87.8
Age, Gender, TBSA, FTB, II	88 (4)	88 (6)	87 (5)	90 (5)	89 (5)	88 (4)	88.3
Age, Gender, BT, FTB, II	76 (12)	83 (6)	85 (7)	82 (8)	83 (7)	83 (6)	82
Age, Gender, TBSA, BT, II	88 (4)	87 (5)	88 (5)	89 (5)	86 (7)	88 (6)	87.6
Age, Gender, TBSA, BT, FTB	88 (4)	88 (4)	88 (5)	87 (5)	88 (5)	89 (4)	88
Gender, BT, FTB, II	83 (10)	83 (9)	82 (10)	83 (9)	83 (10)	83 (10)	82.8
BT, TBSA, FTB, II	84 (4)	84 (6)	86 (5)	86 (5)	84 (6)	85 (6)	84.8
Age, FTB, II	79 (10)	84 (6)	84 (8)	83 (6)	85 (8)	82 (5)	82.8
Age, BT, TBSA	89 (4)	88 (4)	88 (5)	88 (5)	88 (5)	88 (5)	88.2
Age, TBSA, II	89 (3)	88 (5)	88 (3)	89 (5)	89 (5)	88 (5)	88.5
Age, BT, FTB	77 (12)	79 (5)	79 (8)	79 (9)	80 (7)	80 (8)	79
Age, TBSA	89 (3)	88 (5)	89 (4)	87 (4)	82 (17)	88 (5)	87.1
Age, TBSA, FTB	88 (3)	88 (4)	88 (4)	87 (4)	88 (4)	89 (4)	88
Gender, TBSA, BT, II	84 (4)	86 (6)	86 (7)	85 (8)	85 (7)	83 (7)	84.8
Age, Gender, TBSA	88 (4)	88 (4)	89 (4)	88 (5)	87 (5)	88 (5)	88

TBSA: Total burned surface area; II: Inhalation ınjury; BT: Burn type; FTB: Full-thickness burn.

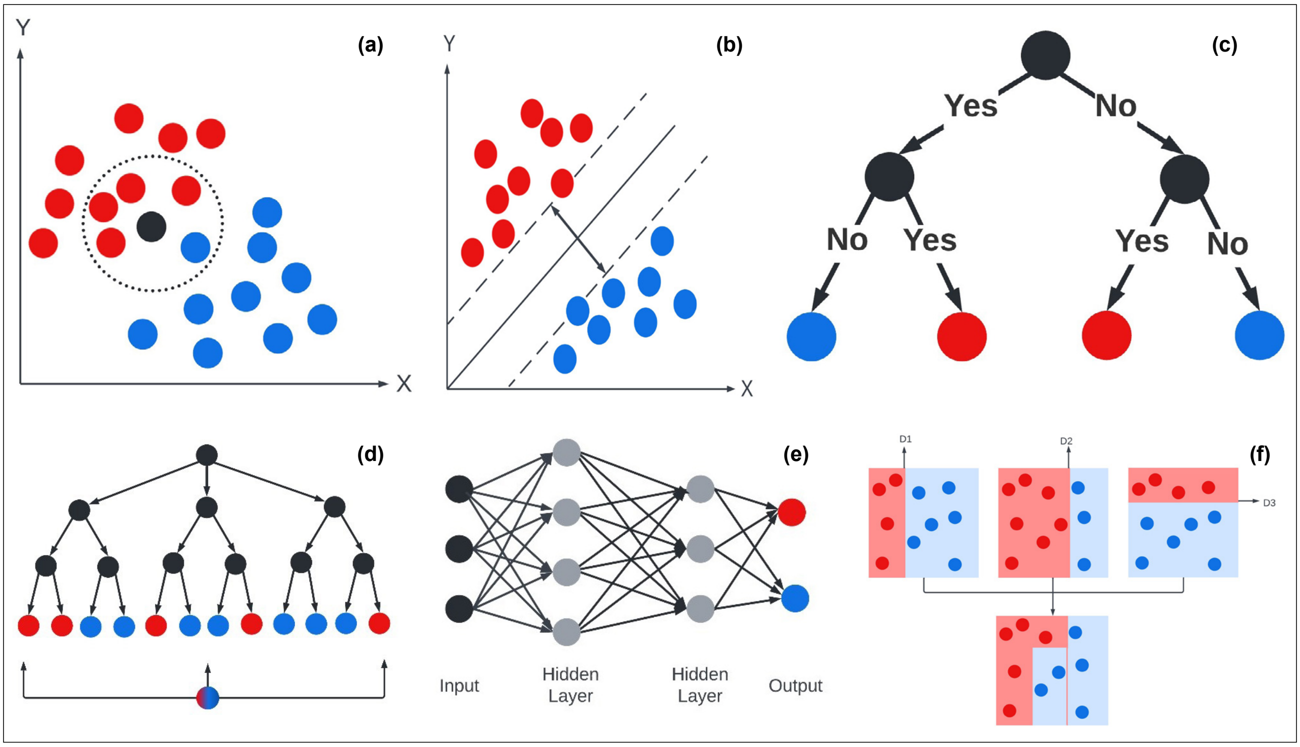

Figure 1. Comparison of k-NN **(a)**, DT **(b)**, RF **(c)**, SVM **(d)**, MLP **(e)**, and AB **(f)**, respectively. Red circles indicate burn injury patients, blue circles indicate non-burn injury patients, and black circles indicate unclassified patients.

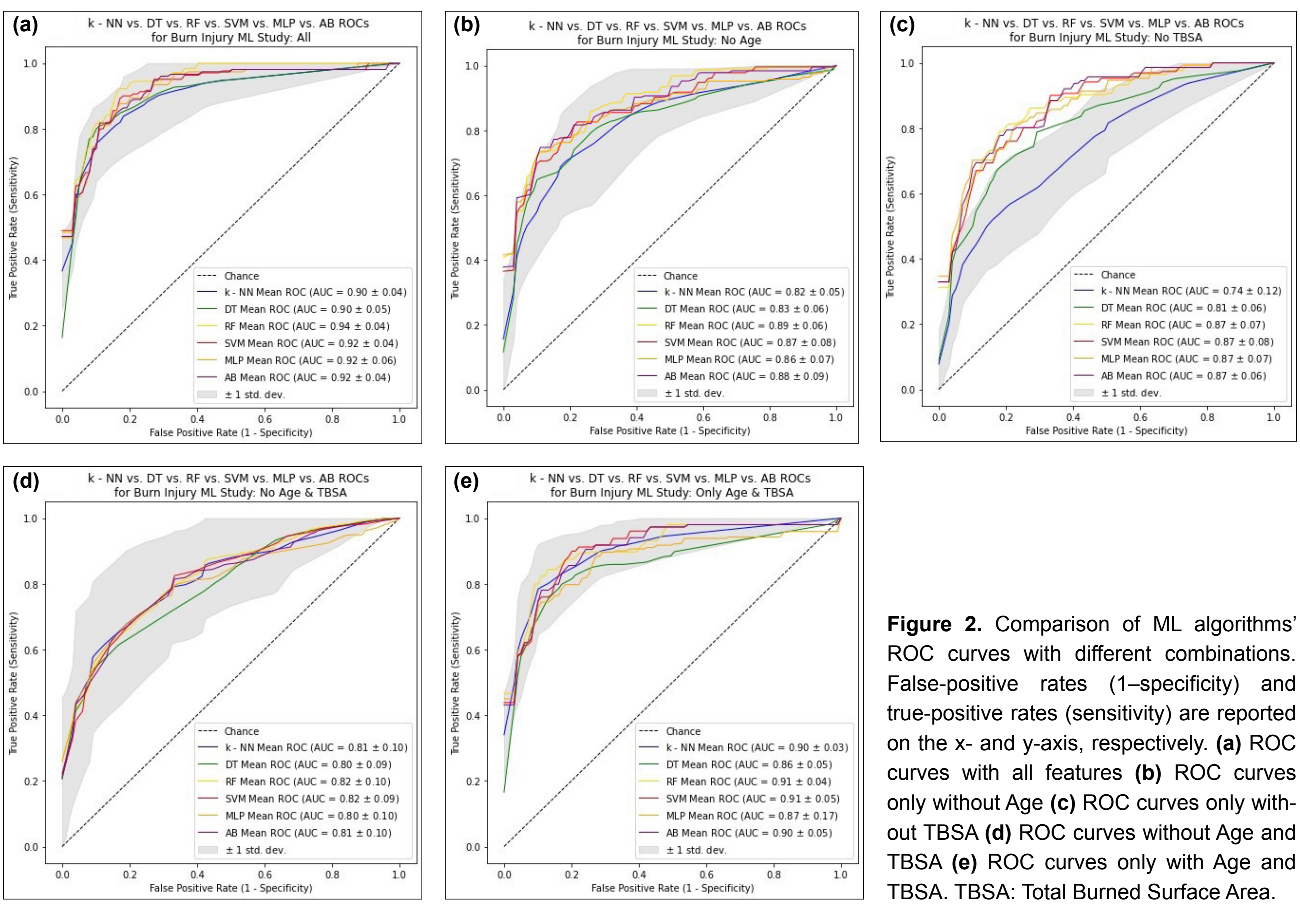

Figure 2. Comparison of ML algorithms' ROC curves with different combinations. False-positive rates (1–specificity) and true-positive rates (sensitivity) are reported on the x- and y-axis, respectively. **(a)** ROC curves with all features **(b)** ROC curves only without Age **(c)** ROC curves only without TBSA **(d)** ROC curves without Age and TBSA **(e)** ROC curves only with Age and TBSA. TBSA: Total Burned Surface Area.

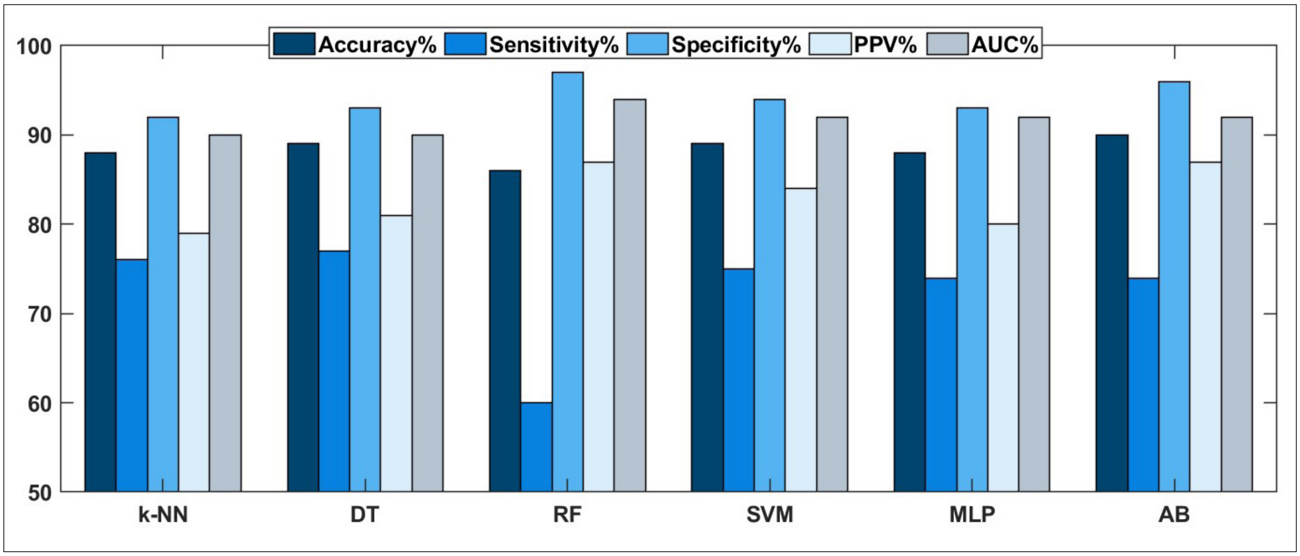

Figure 3. Performance metrics of algorithms based on accuracy, sensitivity, specificity, PPV, and AUC.

to age, excluding total burn, the average scores of the algorithms reduced average accuracy to 82% and average AUC to 84%.

Figure 2 (a-e) shows comparison of the average area under the ROC curve of ML algorithms with different feature combinations. All algorithms had an AUC of over 90% when all features were used (Figure 2a). Excluding age and TBSA, retaining the other features reduced AUC values to 81%, 80%, 82%, 82%, 80%, and 81% for k-NN, DT, SVM, MLP, RF, and AB, respectively (Figure 2d).

Figure 3 gives the five performance metrics: accuracy, sensitivity, specificity, PPV, and AUC for all ML algorithms. The results indicate that the AB algorithm had the best prediction performance with an accuracy of 90%, sensitivity of 74%, specificity of 96%, PPV of 97%, and an AUC of 92%.

DISCUSSION

To the best of our knowledge, this is the first study that evaluated ML algorithms' predictive performances among severe burn patients in a tertiary burn treatment center.

In the current study, six well-known parameters were used to calculate the predictive performances of ML algorithms. The results of ML algorithms showed that, even if the best result was obtained with all six parameters, the results of ML with only age and TBSA also showed excellent and comparable predictive performance. According to these results, TBSA is the most important predictive factor on burn mortality, followed by age. This finding is similar to recent studies.[15,16] Excluding age and total burn, retaining the other features, the best results were obtained using SVM with an accuracy of 83%, sensitivity of 62%, specificity of 91%, and an AUC of 82%. The RF algorithm that relied on a combination of age and TBSA only could achieve an accuracy of 89%, sensitivity of 76%, specificity of 94%, and an AUC of 81%. These results indicate that age and total burn are the strongest predictors of mortality and should be included in the ML models dataset to maintain the highest predictive performance. However, adding the other features (gender, BT, FTB, and II) can improve the success rates of predictions. Although some studies showed that the female gender was a significant risk factor for burn mortality, some others showed different results.[17,18] As a result, it is thought that gender is not a clearly determining factor. In some previous studies, II, FTBSA and BT were shown as factors affecting burn mortality.[6,19,20] However, ML algorithm results did not support this in the current study.

Cobb et al. used tree-based ensemble models such as stochastic gradient boosting and RF.[21] The study reported the area under the ROC curves ranging from 62% to 82%. Although these values seem relatively lower than our results, considering both patient and hospital-level data for the prediction models may affect ML models' performance.

Fransén et al. used DT, extreme boosting, RF, SVM, and a generalized linear regression model and compared the results of the ML techniques with Baux scores using the DeLong test.[13] The first feature selection included 17 different input features for the ML models. The study reported the area under the ROC curves ranging from 82% to 92%. In the second feature selection, the exclusion of SAPS III significantly reduced AUC. However, the authors thought that the cause of this reduction is that SAPS includes ages and comorbidities. Nevertheless, excluding laboratory and clinical features (without age and burn extent) showed minimal improvement or no change in AUC. Although the study included comorbidity parameters to ML algorithms, the results are similar to this study.

Patil et al. studied data from the records of 180 patients with burn injuries.[22] They included the percentage of burns received for eight different body parts in four different ML models (Naive Bayes, SVM, Backpropagation, DT), along with the patient's age and sex. All models achieved over 95% AUC values. However, although ML models had high prediction performance, the patient population consisted of 91 alive and 89 dead patients. Although it might cause to increase in the success rate of ML, it brings some questions about the homogenization of the cohort in a retrospective study, as the authors mentioned at the end of the study.

Stylianou et al. used data from 66,611 burn patients from 2003 to 2011 and evaluated the predictive performance of several ML methods, including RF, SVM, artificial neural network, logistic regression, and Naive Bayes.[23] The AUC values were obtained as 97% for the ANN model and 95% for the RF model. The TBSA median in this study was 1.50%, whereas, in the current study, this value was 26%. Since our patient population included more severe burn patients, it is thought that the differences in the ML results may be due to this distinction. Moreover, the cross-validation technique of ML methods was not used in this study which reduces the reliability of the results.

Our study is unique in that it evaluates ML models in severe burn patients and investigates the different combinations of input features. Moreover, six ML algorithms with unique hyperparameter combinations were used to determine which model provides the highest accuracy across the burn patient population. Specifically, our data showed that ML algorithms were able to enhance the predictive capability of mortality in burn patients. The algorithm with the best accuracy was obtained using all input features. AB algorithm showed an accuracy of 90%, sensitivity of 74%, specificity of 96%, and an AUC of 92%. There were some studies that compared the predictive performances of well-known scoring systems for burn patients.[24,25] Although these systems' prediction of mortality was comparable with ML algorithm's results, most of them were not easy to calculate during clinical practice. It was thought that the development of an ML algorithm for burn patients would be useful for clinicians.

Limitations

This study has some limitations. First, this study was performed on a patient population in a single center. Hence, further investigations are needed on different patient populations and larger case scales. Second, the retrospective design of this study might be affected the results. ML algorithms' predictive performances should be performed in a prospective study and on a different patient population. Further investigations are needed on this topic.

CONCLUSION

ML algorithms are an actual and developing topic as a result of technological developments. Over time, it will continue to gain more space, especially in medical sciences. The use of technological developments in clinical practice is essential in spending less time and convenience for healthcare professionals.

Ethics Committee Approval: This study was approved by the University of Health Sciences Izmir Bozyaka Training and Research Hospital Ethics Committee (Date: 09.03.2022, Decision No: 2022/44).

Peer-review: Externally peer-reviewed.

Authorship Contributions: Concept: H.Y., O.U.; Design: H.Y., Y.A.; Supervision: A.D.U., M.Y.; Resource: H.Y., M.Y.; Materials: A.D., M.Y.; Data collection and/or processing: O.U., Y.A.; Analysis and/or interpretation: H.Y., A.D.U.; Literature search: O.U., Y.A.; Writing: A.D.U., O.U.; Critical review: A.D.U., M.Y.

Conflict of Interest: None declared.

Financial Disclosure: The author declared that this study has received no financial support.

REFERENCES

1. Burn Incidence and Treatment in the United States: 2016. American Burn Association; 2016. Available from: https://ameriburn.org/who-we-are/media/burn-incidence-fact-sheet Accessed Nov 29, 2019.
2. Hefny AF, Idris K, Eid HO, Abu-Zidan FM. Factors affecting mortality of critical care trauma patients. Afr Health Sci 2013;13:731–5. [CrossRef]
3. Colohan SM. Predicting prognosis in thermal burns with associated inhalational injury: A systematic review of prognostic factors in adult burn victims. J Burn Care Res 2010;31:529–39. [CrossRef]
4. Dries DJ, Endorf FW. Inhalation injury: Epidemiology, pathology, treatment strategies. Scand J Trauma Resusc Emerg Med 2013;21:31. [CrossRef]
5. Endorf FW, Gamelli RL. Inhalation injury, pulmonary perturbations, and fluid resuscitation. J Burn Care Res 2007;28:80–3. [CrossRef]
6. Moore EC, Pilcher DV, Bailey MJ, Stephens H, Cleland H. The Burns Evaluation and Mortality Study (BEAMS): Predicting deaths in Australian and New Zealand burn patients admitted to intensive care with burns. J Trauma Acute Care Surg 2013;75:298–303. [CrossRef]
7. Tobiasen J, Hiebert JM, Edlich RF. The abbreviated burn severity index. Ann Emerg Med 1982;11:260–2. [CrossRef]
8. Osler T, Glance LG, Hosmer DW. Simplified estimates of the probability of death after burn injuries: Extending and updating the baux score. J Trauma 2010;68:690–7. [CrossRef]
9. Gomez M, Wong DT, Stewart TE, Redelmeier DA, Fish JS. The FLAMES score accurately predicts mortality risk in burn patients. J Trauma 2008;65:636–45. [CrossRef]
10. Marcinkevics R, Reis Wolfertstetter P, Wellmann S, Knorr C, Vogt JE. Using machine learning to predict the diagnosis, management and severity of pediatric appendicitis. Front Pediatr 2021;9:662183. [CrossRef]
11. Daunhawer I, Kasser S, Koch G, Sieber L, Cakal H, Tütsch J, et al. Enhanced early prediction of clinically relevant neonatal hyperbilirubinemia with machine learning. Pediatr Res. 2019;86:122–7. [CrossRef]
12. Liu NT, Salinas J. Machine learning in burn care and research: A systematic review of the literature. Burns 2015;41:1636–41. [CrossRef]
13. Fransén J, Lundin J, Fredén F, Huss F. A proof-of-concept study on mortality prediction with machine learning algorithms using burn intensive care data. Scars Burn Heal 2022;8:20595131211066585. [CrossRef]
14. Murari A, Singh KN. Lund and Browder chart-modified versus original: A comparative study. Acute Crit Care 2019;34:276–81. [CrossRef]
15. Jeschke MG, Pinto R, Kraft R, Nathens AB, Finnerty CC, Gamelli RL, et al. Morbidity and survival probability in burn patients in modern burn care. Crit Care Med 2015;43:808–15. [CrossRef]
16. Seo DK, Kym D, Yim H, Yang HT, Cho YS, Kim JH, et al. Epidemiological trends and risk factors in major burns patients in South Korea: A 10-year experience. Burns 2015;41:181–7. [CrossRef]

17. O'Keefe GE, Hunt JL, Purdue GF. An evaluation of risk factors for mortality after burn trauma and the identification of gender-dependent differences in outcomes. J Am Coll Surg 2001;192:153–60. [CrossRef]
18. Arshi S, Sadeghi-Bazargani H, Mohammadi R, Ekman R, Hudson D, Djafarzadeh H, et al. Prevention oriented epidemiologic study of accidental burns in rural areas of Ardabil, Iran. Burns 2006;32:366–71. [CrossRef]
19. Walker PF, Buehner MF, Wood LA, Boyer NL, Driscoll IR, Lundy JB, et al. Diagnosis and management of inhalation injury: An updated review. Crit Care 2015;19:351. [CrossRef]
20. Al Ibran E, Mirza FH, Memon AA, Farooq MZ, Hassan M. Mortality associated with burn injury-a cross sectional study from Karachi, Pakistan. BMC Res Notes 2013;6:545. [CrossRef]
21. Cobb AN, Daungjaiboon W, Brownlee SA, Baldea AJ, Sanford AP, Mosier MM, et al. Seeing the forest beyond the trees: Predicting survival in burn patients with machine learning. Am J Surg 2018;215:411–6. [CrossRef]
22. Patil BM, Joshi RC, Toshniwal D, Biradar S. A new approach: Role of data mining in prediction of survival of burn patients. J Med Syst 2011;35:1531–42. [CrossRef]
23. Stylianou N, Akbarov A, Kontopantelis E, Buchan I, Dunn KW. Mortality risk prediction in burn injury: Comparison of logistic regression with machine learning approaches. Burns 2015;41:925–34. [CrossRef]
24. Halgas B, Bay C, Foster K. A comparison of injury scoring systems in predicting burn mortality. Ann Burns Fire Disasters 2018;31:89–93.
25. Salehi SH, As'adi K, Abbaszadeh-Kasbi A, Isfeedvajani MS, Khodaei N. Comparison of six outcome prediction models in an adult burn population in a developing country. Ann Burns Fire Disasters 2017;30:13–7.

ORİJİNAL ÇALIŞMA - *ÖZ*

Yanık hastalarının mortalite öngörüsünde makine öğrenimi algoritmalarının kullanılması

Dr. Hilmi Yazıcı,[1] Dr. Onur Ugurlu,[2] Dr. Yesim Aygul,[3] Dr. Mehmet Yildirim,[4] Dr. Ahmet Deniz Uçar[4]

[1]Marmara Üniversitesi Pendik Eğitim ve Araştırma Hastanesi, Genel Cerrahi Kliniği, İstanbul, Türkiye
[2]İzmir Bakırçay Üniversitesi, Mimarlık ve Mühendislik Fakültesi, İzmir, Türkiye
[3]Ege Üniversitesi, Matematik Bölümü, İzmir, Türkiye
[4]Sağlık Bilimleri Üniversitesi, İzmir Bozyaka Eğitim ve Araştırma Hastanesi, Genel Cerrahi Kliniği, İzmir, Türkiye

AMAÇ: Yanıklar dünya çapında en yaygın travmalardan biridir. Ciddi yanık hastalarında, mortalite ve morbidite riski oldukça yüksektir. Bu çalışmada, yanık mortalitesi için risk faktörlerini değerlendirmeyi ve bu faktörler için altı farklı Makine Öğrenimi (MÖ) Algoritmasının tahmin performanslarını karşılaştırılması amaçlandı.

GEREÇ VE YÖNTEM: Yanık tedavi merkezinde tedavi edilen yanık hastalarının tıbbi kayıtları retrospektif olarak incelendi. Hastaların yaş ve cinsiyet, Toplam Yanık Yüzey Alanı, İnhalasyon Yaralanması, tam kat yanıkları, yanık tipleri gibi verileri kaydedildi ve MÖ modellerinde girdi özelliği olarak kullanıldı. Hastalar Yaşayanlar ve Yaşamayanlar olarak iki grup altında incelendi. Mortaliteyi tahmin etmek için k-Nearest Neighbor, Decision Tree, Random Forest, Support Vector Machine, Multi-Layer Perceptron, ve AdaBoost olmak üzere altı MÖ algoritması kullanıldı. Her algoritma için birçok farklı öğreti kombinasyonu değerlendirildi.

BULGULAR: Çalışmaya dahil edilen hasta sayısı 363 idi. MÖ algoritmalarına dahil edilen altı parametrenin tümü tek değişkenli analizde anlamlı bir fark gösterdi ($p<0.001$). Sonuçlar, tüm girdi özelliklerini kullanan AdaBoost algoritmasının %90 Doğruluk ve %92 AUC ile en iyi tahmin performansına sahip olduğunu göstermektedir.

SONUÇ: Doğru verilerin öğretildiği MÖ algoritmaları, yanık mortalitesinde yüksek tahmin performansı göstermiştir. Yanık hastaları için uygun verilerin öğretildiği MÖ modellerinin, klinik pratikte kullanışlı olabileceğini düşünmekteyiz. Bu nedenle, bu konuda daha ileri araştırmalara ihtiyaç vardır.

Anahtar sözcükler: Makine öğrenimi; sağkalım; tahminleme; yanık.

Ulus Travma Acil Cerrahi Derg 2023;29(10):1130-1137 DOI: 10.14744/tjtes.2023.79968