

Enhancing heatstroke prediction accuracy with interpretable machine learning: a multi-center data-driven approach

Qingbo Zeng^{1,2,*}, Xingping Deng^{1,*}, Longping He¹, Lincui Zhong¹,
Qingwei Lin¹, Nianqing Zhang² and Jingchun Song^{1,3}

¹ The 908th Hospital of Chinese PLA Logistic Support Force, Nanchang, China

² Nanchang Hongdu Hospital of Traditional Chinese Medicine, Nanchang, China

³ Nanchang Key Laboratory of Thrombosis and Hemostasis, Nanchang, China

* These authors contributed equally to this work.

ABSTRACT

Background: Heatstroke poses a significant threat to public health, frequently culminating in fatal outcomes. This study aimed to develop and validate an interpretable machine learning (ML) model to forecast heatstroke using clinical and laboratory data.

Methods: Data were collated from 24 hospitals spanning the years 2021 to 2023, with data from 2021 and 2022 comprising the training datasets and data from 2023 designated for validation. Model efficacy was quantified *via* the area under the receiver operating characteristic curve (AUROC) and calibration plots. Furthermore, the SHapley Additive exPlanations (SHAP) methodology was employed to elucidate the interpretability of the final model.

Results: The study encompassed 691 patients, with 176 in the training datasets and 80 in the testing datasets diagnosed with heatstroke. Among the nine ML models assessed, the gradient boosting machine (GBM) model demonstrated superior performance, achieving an AUROC of 0.971 in the training datasets and 0.836 in the testing datasets, and exhibiting substantial net benefits in decision curve analysis. Creatine kinase (CK)-MB was identified as the most impactful variable influencing the GBM model's efficacy.

Conclusion: The ML model we developed demonstrates robust predictive capabilities for heatstroke, potentially aiding clinicians in the identification and management of patients at elevated risk.

Subjects Emergency and Critical Care, Public Health, Sports Medicine

Keywords Machine learning, Heatstroke, Interpretability, Prediction model

INTRODUCTION

Global climate change has caused a dramatic rise in air temperatures, leading to an increased incidence of heat-related illnesses and posing a significant threat to human health (*Gauer & Meyers, 2019; Shimazaki et al., 2020*). Heatstroke is a severe clinical syndrome and medical emergency, which is considered as the most severe form of heat-related illnesses (*Hoek et al., 2023; Bouchama et al., 2022; Asmara, 2020*). It is marked by a core temperature with exceeding 40 °C and abnormal manifestations of central

Submitted 19 May 2025

Accepted 20 October 2025

Published 14 November 2025

Corresponding author

Jingchun Song,

songjingchun@126.com

Academic editor

Luigi Di Biasi

Additional Information and
Declarations can be found on
page 10

DOI 10.7717/peerj.20377

© Copyright

2025 Zeng et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

nervous system. Additionally, heatstroke can easily progress to disseminated intravascular coagulation and multi-organ dysfunction, both of which typically result in poor prognosis. Prior researches have demonstrated that the mortality rate of heatstroke can reach 40–50% (Epstein & Yanovich, 2019; Glaser et al., 2016). Thus, social attention to heatstroke has been growing year by year and accurate prediction of heatstroke is becoming essential.

It is regrettable that the diagnosis of heatstroke continues to be heavily dependent on core body temperature and central nervous system symptoms (Lin et al., 2024). This conventional diagnostic approach may lead to misdiagnoses in certain cases. Previous study has reported that the patient admitted to a hospital with elevated biochemical indicators and consciousness disorder but normal body temperature was diagnosed with heatstroke (Yeoh & Law, 2023). As a result, several predictive models have been developed to identify heat-related illnesses, including heatstroke (Wei et al., 2022; Wan et al., 2023). However, these models have not been widely implemented in clinical settings due to the absence of comprehensive data and sophisticated algorithms. There is an urgent need for a novel diagnostic method that can accurately diagnose heatstroke, thereby enhancing diagnostic accuracy and reducing the rate of missed diagnoses. Recently, predictive models that integrate machine learning (ML) algorithms and extensive clinical data have attracted significant attention. ML predictive models for heatstroke and its prognosis demonstrate exceptional performance and generalizability (Wang et al., 2019, 2023). Nonetheless, the inherent opacity of ML algorithms presents a challenge to their clinical application (Watson et al., 2019). At present, algorithms designed to elucidate ML models are emerging. We are optimistic that these explanatory algorithms will enable a deeper comprehension of the underlying decision-making processes within ML models.

In this study, we developed and validated an ML model for heatstroke prediction using a multi-center database. Additionally, we employed the SHapley Additive exPlanations (SHAP) method to interpret the outputs of the optimal ML model.

METHODS

Data collection and population

In this multi-center retrospective study, 24 hospitals across China provided data from 2021 to 2023 in Excel sheet format. The training set comprises heat-related disease data from 2021 and 2022, while the 2023 data serve as the testing set (Fig. 1). The criteria for inclusion were: (1) age ≥ 18 years, (2) patients admitted to the hospital with a confirmed diagnosis of heat-related illness (Bouchama & Knochel, 2002). The following were excluded: (1) patients younger than 18 years, (2) those with congenital coagulopathy or severe chronic liver or kidney disease, (3) more than 30% missing data. This study was approved by the ethics committee of the 908th Hospital of Chinese PLA Logistic Support Forces, which waived the requirement for informed consent because of the study's retrospective design (approval document ID 908yyLL028).

Variables extraction

Variable extraction was performed starting from the point of hospital admission. Variables contained demographic information, clinical parameters and laboratory results.

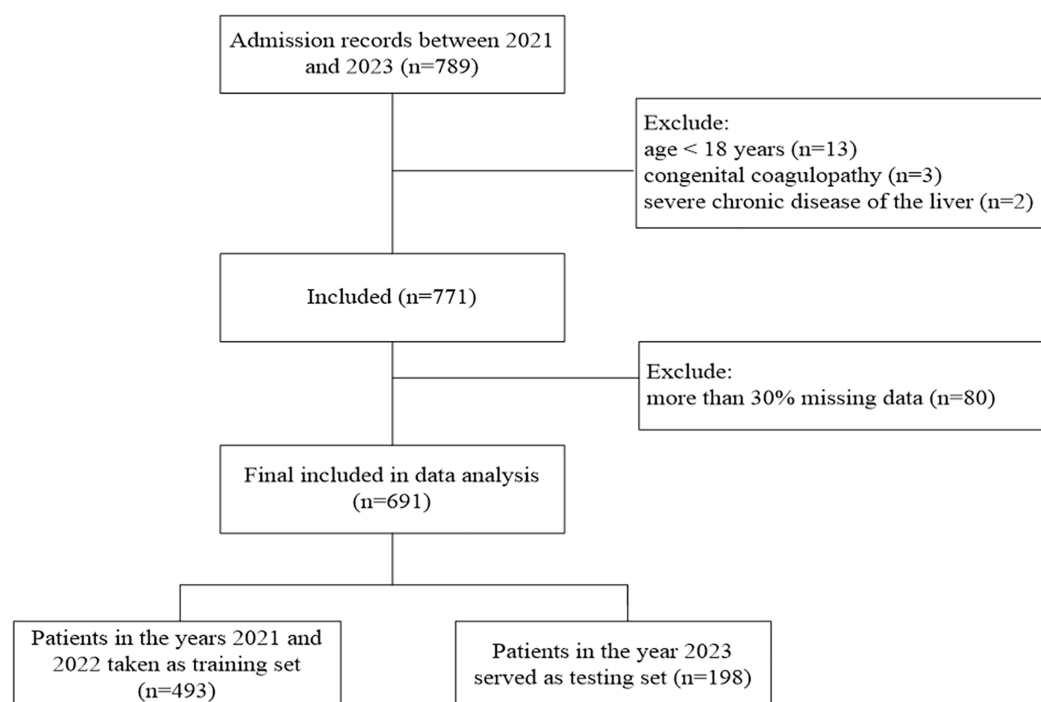


Figure 1 Flowchart of patients' selection.

Full-size  DOI: [10.7717/peerj.20377/fig-1](https://doi.org/10.7717/peerj.20377/fig-1)

Laboratory results: blood samples were collected from all patients with heat-related illnesses upon admission for laboratory testing. The selected variables were: demographic information including age, and gender; clinical characteristics including mean arterial pressure, core temperature, heart rate, and SPO₂; biochemical indices including prothrombin time (PT), activated partial thrombin time (APTT), fibrinogen, international normalized ratio (INR), D-dimer, hemoglobin (HB), platelets count, white blood cells (WBC), Hematocrit (HCT), alanine transaminase (ALT), aspartate transaminase (AST), Total bilirubin (Tbil), creatinine (Cr), Creatine Kinase (CK), creatine kinase MB (CK MB), and lactate (Lac). Each data record represents a patient and includes detailed admission time information. It should be specifically noted that variables such as core temperature, thrombin time, and D-dimer were obtained from data collected immediately upon admission.

Outcome definition

This study's main outcome was whether hospital admission resulted in a confirmed diagnosis of heatstroke, represented by the total number of heatstroke cases. The definition of heatstroke is as follows: patients who had previously experienced hot and humid conditions or participated in high-intensity activities were eligible for inclusion if they fulfilled at least one of the criteria detailed in the Chinese Expert Consensus on the Diagnosis and Treatment of Heatstroke ([Liu et al., 2020](#)): (1) neurological issues like coma, seizures, delirium, or unusual behavior; (2) core temperature of 40 °C or above; (3) impairment in the function of at least two organs; or (4) severe coagulopathy or disseminated intravascular coagulation (DIC).

Feature selection, model development, and evaluation

Support vector machine-recursive feature elimination was employed for feature selection to improve model discrimination and eliminate redundant variables. The procedure was strictly performed on the training set, and the resulting feature set was directly applied to the 2023 test set without any refitting. Nine well-known ML algorithms, including K-nearest neighbors (KNN), logistic regression (LR), naive Bayes (NB), support vector machine (SVM), gradient boosting machine (GBM), multiplayer perception (MLP), extreme gradient boosting machine (XGB), neural network (NN), Gaussian Process (GP), was utilized to establish the heatstroke model. Their performance was evaluated using the testing set. The machine-learning algorithms retained their default parameters without any parameter optimization. we introduced the area under the receiver operating characteristic, sensitivity, specificity, accuracy, and F1 score to assess the model performance. The concordance between predictions and actual observations was evaluated with a calibration plot.

Model interpretability and utility

We employed the SHAP approach using the SHAP package to conduct additional analysis on the ML model. This facilitated comprehension of the underlying reasoning mechanism behind the final model and enhanced its interpretability. Decision curve analysis was introduced to evaluate the clinical relevance of the final model.

Statistical analyses

For continuous variables, the current dataset was presented as median and interquartile range (IQR), while categorical variables were described using counts and percentages. Statistical analyses and predictive model construction were performed using the R software (version 4.1.3, The R Foundation, Vienna). Variables with more than 30% missing values were excluded from the analysis. For the remaining variables, those with more than 5% missing values were imputed using the K-nearest neighbors (KNN) method. The KNN imputation model was trained exclusively on the training set data and then directly applied to the 2023 test set without refitting.

RESULTS

Baseline characteristics of the study populations

The study involved 691 participants, with 493 assigned to the training group and 198 in the testing group. The incidence of heatstroke was 35.7% in the training group and 40.4% in the validation group, without any significant statistical differences. Further analysis of various variables across the two groups showed no significant differences in the baseline characteristics observed ([Table 1](#)).

Building and calibrating machine learning models for heatstroke

We compiled a total of 23 clinical and biological features, and through applying recursive feature elimination, six features for heatstroke were ultimately selected. They were TT, CKMB, Lac, core temperature, mean arterial pressure (MAP) and D-dimer. [Table 2](#) and

Table 1 Baseline characteristics of the training and validation datasets.

Variables	All (n = 691)	Training set (n = 493)	Testing set (n = 198)	p value
Men	657 (95.1%)	469 (95.1%)	188 (94.9%)	1.000
Age, yr	22.0 [20.0; 24.0]	22.0 [20.0; 24.0]	22.0 [20.0; 24.0]	0.440
Core temperature	38.2 [37.5; 39.6]	38.3 [37.5; 39.7]	38.1 [37.5; 39.5]	0.391
HR, min ⁻¹	80.0 [68.0; 98.0]	80.0 [68.0; 98.0]	80.0 [68.0; 99.0]	0.682
SPO ₂ , %	98.9 [98.0; 99.1]	98.8 [98.0; 99.1]	99.0 [98.0; 99.2]	0.298
MAP, mmHg	87.3 [81.8; 93.3]	87.3 [82.0; 93.3]	86.7 [81.7; 93.1]	0.429
WBC, ×10 ¹² /L	10.1 [7.85; 13.7]	10.0 [7.76; 13.7]	10.7 [8.20; 14.2]	0.250
HB, g/L	139 [131; 149]	139 [130; 149]	139 [131; 149]	0.852
HCT, %	41.6 [38.6; 44.0]	41.6 [38.6; 44.0]	41.5 [38.6; 43.9]	0.710
PLT, ×10 ⁹ /L	194 [152; 230]	193 [152; 230]	194 [155; 226]	0.886
PT, s	13.2 [12.3; 14.5]	13.2 [12.3; 14.5]	13.3 [12.4; 14.4]	0.825
INR	1.12 [1.05; 1.22]	1.12 [1.04; 1.23]	1.12 [1.06; 1.21]	0.904
APTT, s	30.3 [27.1; 34.4]	30.2 [27.1; 34.4]	30.4 [27.2; 34.4]	0.889
FIB, g/L	2.28 [1.83; 2.78]	2.31 [1.83; 2.79]	2.23 [1.83; 2.69]	0.398
TT, s	16.6 [14.8; 18.1]	16.5 [14.6; 18.0]	17.0 [15.0; 18.3]	0.121
DD, µg/L	0.33 [0.16; 1.00]	0.34 [0.16; 1.05]	0.30 [0.15; 0.95]	0.356
AST, U/L	29.8 [20.6; 53.0]	29.0 [20.3; 52.9]	30.8 [21.7; 55.2]	0.273
ALT, U/L	26.0 [18.0; 50.1]	25.0 [18.0; 50.0]	26.0 [18.6; 50.0]	0.579
TBIL	14.5 [9.74; 20.6]	14.5 [9.70; 20.6]	14.4 [9.79; 21.0]	0.867
Cr, µmol/ml	84.0 [69.0; 108]	85.0 [69.6; 108]	82.0 [67.2; 108]	0.329
CK, U/L	268 [147; 708]	277 [146; 721]	254 [49; 666]	0.533
CKMB, U/L	114 [44.9; 263]	112 [47.0; 261]	123 [42.2; 265]	0.974
Lac, mmol/L	1.47 [1.10; 2.20]	1.50 [1.10; 2.28]	1.44 [1.07; 2.16]	0.334
Heatstroke	256 (37.0%)	176 (35.7%)	80 (40.4%)	0.284

Notes:
 Values are n (%), mean ± standard deviation or median (interquartile range), unless otherwise noted.
 Abbreviations: HR, heart rate; SPO₂, oxygen saturation of blood; MAP, mean arterial pressure; WBC, white blood cells; HB, hemoglobin; HCT, Hematocrit; PT, prothrombin time; INR, international normalized ratio; APTT, activated partial thrombin time; FIB, fibrinogen; TT, thrombin time; DD, D-dimer; PLT, platelets; ALT, alanine transaminase; AST, aspartate transaminase; TBIL, total bilirubin; Cr, creatinine; CK, Creatine kinase; CKMB, Creatine kinase isozyme; Lac, lactate.

Table 2 Performance of ML models for predicting heatstroke.

Model	AUC	Accuracy	Balanced accuracy	Precision	F1-scores	Sensitivity	Specificity
GBM	0.966	0.879	0.858	0.938	0.833	0.750	0.966
SVM	0.672	0.671	0.547	0.679	0.802	0.887	0.981
XGB	0.789	0.702	0.605	0.723	0.610	0.733	0.943
MLP	0.772	0.726	0.643	0.746	0.521	0.648	0.934
KNN	0.867	0.775	0.707	0.822	0.699	0.528	0.943
NN	0.710	0.663	0.582	0.553	0.615	0.705	0.868
GP	0.702	0.700	0.587	0.783	0.676	0.795	0.968
LR	0.594	0.655	0.533	0.594	0.182	0.109	0.959
NB	0.770	0.722	0.636	0.747	0.517	0.665	0.937

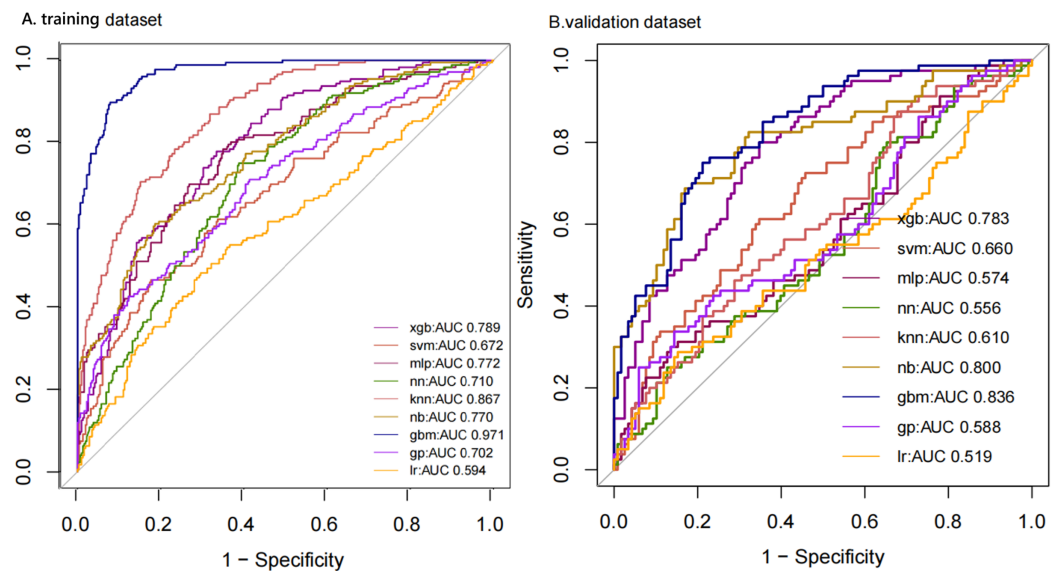


Figure 2 ROC curves of models with different algorithms in the training and testing datasets. ROC, receiver operating characteristic; KNN, k-nearest neighbors; LR, logistic regression; NB, naive Bayes; SVM, support vector machine; GBM, gradient boosting machine; MLP, multilayer perceptron; XGB, extreme gradient boosting machine; NN, neural network; GP, Gaussian Process.

Full-size [DOI: 10.7717/peerj.20377/fig-2](https://doi.org/10.7717/peerj.20377/fig-2)

Fig. 2 summarize the performance of the ML models. In the training dataset, the GBM model exhibited the highest performance, achieving an area under the receiver operating characteristic curve (AUROC) of 0.971. In comparison, the performance metrics for KNN, NN, MLP, SVM, GP, NB, XGB, and LR were 0.867, 0.710, 0.772, 0.672, 0.702, 0.770, 0.789, and 0.594, respectively. The testing datasets was employed to confirm the model's effectiveness in predicting heatstroke, and GBM model showed good performance with an AUROC up to 0.836 (Fig. 2). We also incorporated area under the curve (AUC), accuracy, balanced accuracy, kappa, sensitivity, specificity, precision, and F1-scores for a comprehensive evaluation of the model's performance. Those of GBM were higher than other models (Table 2). The calibration curves of the GBM model in the validation set exhibited robust concordance between observed and predicted probabilities, with a calibration intercept of 0.15 (95% confidence interval (CI) [-0.17 to 0.48]) and a calibration slope of 1.23 (95% CI [0.86–1.60]) (Fig. 3).

Interpretability and clinical benefit analysis

Finally, we utilized the SHAP method to calculate and display feature importance. Figure 4 provides an extensive overview of how importance is distributed among all features, making it easier to understand the significance of each feature. TT, CKMB, Lac, core temperature, MAP and D-dimer were confirmed as the top six influencing variables for heatstroke. CKMB is the most influential factor in the model predictions. Figure 5 depicts the clinical effectiveness of ML models across different risk thresholds. The GBM model showed a significant advantage in decision curve analysis (DCA) compared to the “treat-

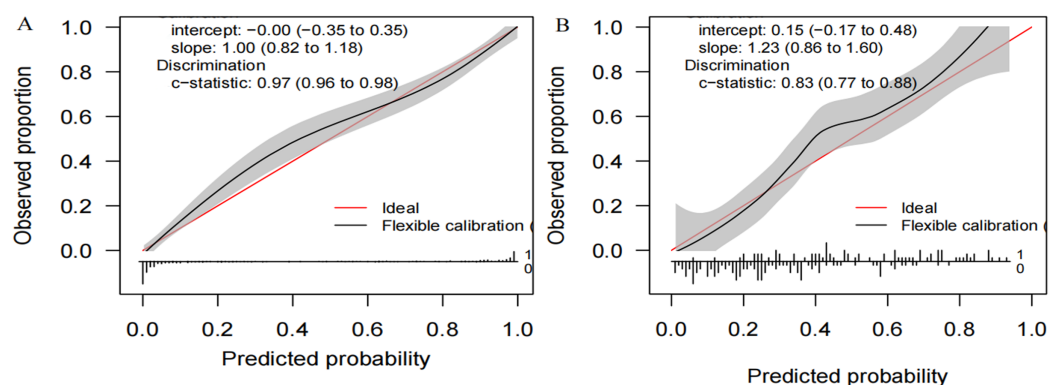


Figure 3 Calibration plots of the GB M model in the (A) training and (B) the testing set. GBM: gradient boosting machine. [Full-size](#) DOI: 10.7717/peerj.20377/fig-3

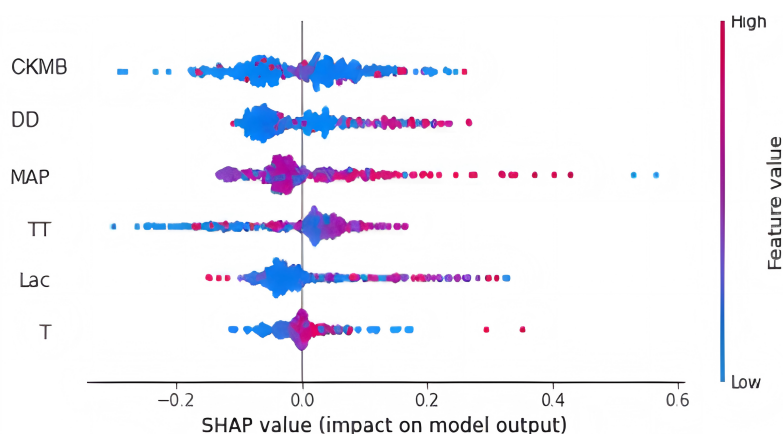


Figure 4 Feature importance derived from gradient boosting machine model. This figure is the result of the SHAP package. Abbreviations: TT, thrombin time; CKMB, Creatine kinase isozyme; Lac, Lactate; MAP, mean arterial pressure; DD, D-dimer. [Full-size](#) DOI: 10.7717/peerj.20377/fig-4

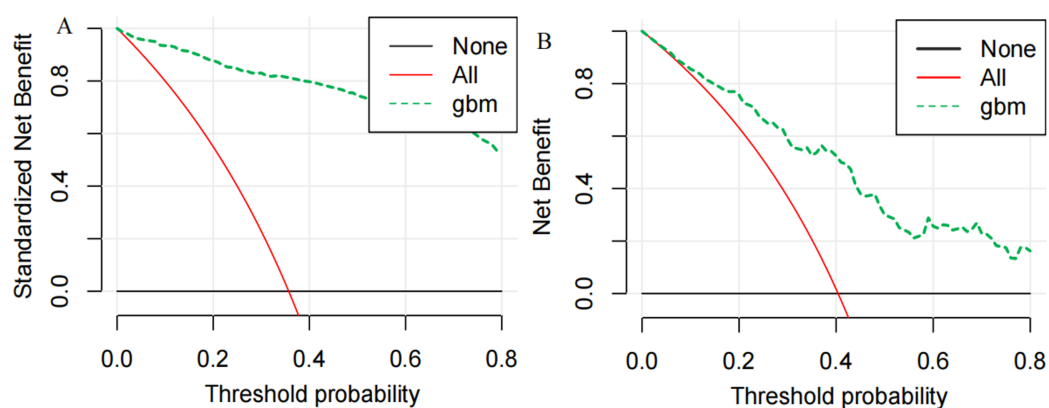


Figure 5 DCA curve analysis of the GBM model in the (A) training and (B) the testing set. GBM, gradient boosting machine. [Full-size](#) DOI: 10.7717/peerj.20377/fig-5

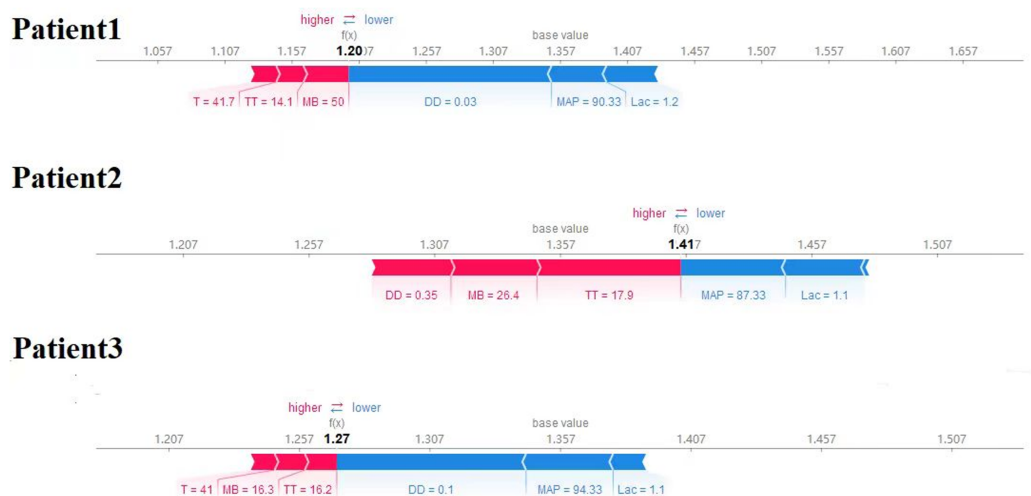


Figure 6 Explaining of patient prediction results. Figure created using SHAP package for explaining GBM model predictions. Full-size [DOI: 10.7717/peerj.20377/fig-6](https://doi.org/10.7717/peerj.20377/fig-6)

all” and “treat-none” strategies when the threshold probability exceeded 20%. Figure 6 showed the individual force plot for three patients. The force plot primarily highlighted the role of each variable in predicting heatstroke for an individual. Variables depicted in red increase the likelihood of heatstroke, whereas those in blue decrease the heatstroke risk.

DISCUSSION

The integration of machine learning into medical and clinical settings represents a significant emerging research trend (Haug & Drazen, 2023; Van Calster & Wynants, 2019). In this study, we built and verified a heatstroke prediction model by employing nine machine learning algorithms. This model was built using routine clinical parameters and laboratory variables collected from 24 hospitals. The results indicated that GBM model had the highest AUC value ($AUC > 0.8$) in both the training and testing datasets, demonstrating the better discriminatory performance. Additionally, other metrics such as accuracy, balanced accuracy, sensitivity, specificity, precision, and F1-scores demonstrated superior performance compared to other machine learning models, highlighting the efficacy of the top-performing model. Decision curve analysis revealed that the GBM model could increase the net benefit for predicting heatstroke. Illustrating the threshold ranges above the prediction-all and -none curves provides insight into how the model can be utilized in real-world clinical settings. Therefore, compared with other ML models, it had better predictability to identify heatstroke in clinical settings. Doctors can readily assess patients’ risk of heatstroke and take proactive measures to prevent it. Additionally, we anticipate that this prediction model could aid in optimizing medical resources by facilitating the early detection and monitoring of high-risk heatstroke patients. However, further investigations are required to validate these assumptions.

Numerous models have been developed to predict heat-related illnesses, including heatstroke, by integrating various weather data (Wan et al., 2023; Takada et al., 2023).

These models exhibit relatively effective prediction capabilities for heat-related illnesses or heatstroke. However, they failed to take into account clinical and laboratory variables, and some were restricted to daily-unit predictions. Consequently, weather data may not be sufficient to accurately predict heatstroke in patients, and a new heatstroke prediction model is needed. A recently developed interpretable machine learning model to predict heatstroke has been found to be effective ([Ogata et al., 2021](#)), but this model still requires further validation and merely include weather relevant variables. In addition, the model based on Japanese society may not be suitable to patients in China because of regional and ethnic differences. Hence, the objective of our study was to build machine learning models specifically for screening heatstroke in the Chinese population.

There has been widespread application of explainable machine learning algorithms to predict the occurrence or outcome of various diseases ([Guan et al., 2023](#); [Nielsen et al., 2024](#); [Deshmukh & Merchant, 2020](#); [Choi et al., 2023](#)). We implemented an explainable analysis to describe how these features influenced the GBM model and how this model achieved individual case predictions. Prior studies showed that machine learning models were highly predictive but poorly interpretable ([Wan et al., 2023](#); [Van Calster & Wynants, 2019](#); [Lundberg & Lee, 2017](#)). When users entered data to acquire outputs, it was hard to understand how the model generated predictions. The lack of transparency in ML models greatly restricts their clinical applicability, particularly in complicated cases with significant medical consequences. The explainable ML model we constructed enable users to better comprehend the model's decision-making process, thus enhancing its reliability and transparency. In terms of the number of variables, [Ogata et al. \(2021\)](#) and [Wan et al. \(2023\)](#) employed 18 and 33 features, respectively, which may be susceptible to collinearity, potentially biasing the model's predictions of heatstroke. Simultaneously, when a prediction model includes a large number of features, its complexity and time-consuming nature make it challenging for clinical physicians to use. We employed recursive feature elimination to identify the optimal features and ultimately chose six indicators for developing ML models. This approach helps mitigate potential biases that can arise from manually selecting predictors without sufficient experience ([Hayashida et al., 2018](#); [Staartjes et al., 2022](#)). Through the interpretative analysis, we discovered that CKMB was the key predictor for the GBM model's decision-making process. Other essential predictors included MAP, D-dimer, lactate, TT, and core temperature. These early indicators may be associated with subsequent diagnostic criteria for heat-related illnesses. To ensure the integrity of the predictive modeling, this study restricted all variables to data collected during the pre-diagnosis admission phase, effectively preventing label leakage that could occur if diagnostic criteria were retroactively incorporated into the predictors. The integration of these common clinical features into machine learning models demonstrates promising potential for practical clinical applications. Moreover, decision-making process of the GBM model was further demonstrated by SHAP summary and force plot, adding clinical credibility and interpretability.

This study makes several significant contributions, including the employment of various sophisticated ML techniques for modeling, the application of the SHAP method for interpreting the decision-making mechanism employed by the ML model, and the use of

decision curve analysis. The inclusion of 691 patients from 24 medical centers enhances the model's generalizability. Nevertheless, there are some limitations to the study. Firstly, it is retrospective and may inherently contain biases. Second, it lacks the design of clinical trials to evaluate the model's performance in real-world settings. Thirdly, we used the KNN method to generate interpolation values, which could lead to estimation errors. Finally, the study cohort consisted exclusively of patients with confirmed heat-related illnesses, with a predominance of males, which limits the broad applicability of the model across different gender groups. Future research should prospectively evaluate the model's performance in real-world settings and explore whether this prediction model can effectively reduce cases of heatstroke.

CONCLUSION

An ML model could serve as a reliable tool for heatstroke identification and possess robust predictive power, assisting doctors in recognizing patients at high risk. This would facilitate early intervention and the effective prevention and management of individuals at high risk of heatstroke.

ACKNOWLEDGEMENTS

We thank the patients for participating in our study.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This research does not receive external funding.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Qingbo Zeng performed the experiments, prepared figures and/or tables, and approved the final draft.
- Xingping Deng performed the experiments, prepared figures and/or tables, and approved the final draft.
- Longping He analyzed the data, prepared figures and/or tables, and approved the final draft.
- Lincui Zhong analyzed the data, prepared figures and/or tables, and approved the final draft.
- Qingwei Lin analyzed the data, prepared figures and/or tables, and approved the final draft.
- Nianqing Zhang analyzed the data, prepared figures and/or tables, and approved the final draft.
- Jingchun Song conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.

Human Ethics

The following information was supplied relating to ethical approvals (*i.e.*, approving body and any reference numbers):

The Ethics Committee of the 908th Hospital of Logistic Support Force (Nanchang, China).

Ethics

The following information was supplied relating to ethical approvals (*i.e.*, approving body and any reference numbers):

The ethics committee of the 908th Hospital of Chinese PLA Logistic Support Forces.

Data Availability

The following information was supplied regarding data availability:

The raw measurements are available in the [Supplemental files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.20377#supplemental-information>.

REFERENCES

- Asmara IGY. 2020. Diagnosis and management of heatstroke. *Acta Medica Indonesiana* 52(1):90–97.
- Bouchama A, Abuyassin B, Lehe C, Laitano O, Jay O, O'Connor FG, Leon LR. 2022. Classic and exertional heatstroke. *Nature Reviews Disease Primers* 8(1):8 DOI 10.1038/s41572-021-00334-6.
- Bouchama A, Knochel JP. 2002. Heat stroke. *New England Journal of Medicine* 346(25):1978–1988 DOI 10.1056/nejmra011089.
- Choi J, Anderson T, Tennakoon L, Spain DA, Forrester JD. 2023. Explainable machine learning to bring database to the bedside: development and validation of the TROUT (Trauma fRailty OUTcomes) index, a point-of-care tool to prognosticate outcomes after traumatic injury based on frailty. *Annals of Surgery* 278(1):135–139 DOI 10.1097/sla.0000000000005649.
- Deshmukh F, Merchant SS. 2020. Explainable machine learning model for predicting gi bleed mortality in the intensive care unit. *American Journal of Gastroenterology* 115(10):1657–1668 DOI 10.14309/ajg.0000000000000632.
- Epstein Y, Yanovich R. 2019. Heatstroke. *The New England Journal of Medicine* 380(25):2449–2459 DOI 10.1056/NEJMr1810762.
- Gauer R, Meyers BK. 2019. Heat-related illnesses. *American Family Physician* 99(8):482–489.
- Glaser J, Lemery J, Rajagopalan B, Diaz HF, García-Trabanino R, Taduri G, Madero M, Amarasinghe M, Abraham G, Anutrakulchai S, Jha V, Stenvinkel P, Roncal-Jimenez C, Lanasa MA, Correa-Rotter R, Sheikh-Hamad D, Burdmann EA, Andres-Hernando A, Milagres T, Weiss I, Kanbay M, Wesseling C, Sánchez-Lozada LG, Johnson RJ. 2016. Climate change and the emergent epidemic of ckd from heat stress in rural communities: the case for heat stress nephropathy. *Clinical Journal of the American Society of Nephrology* 11(8):1472–1483 DOI 10.2215/cjn.13841215.
- Guan C, Ma F, Chang S, Zhang J. 2023. Interpretable machine learning models for predicting venous thromboembolism in the intensive care unit: an analysis based on data from 207 centers. *Critical Care* 27(1):406 DOI 10.1186/s13054-023-04683-4.

- Haug CJ, Drazen JM. 2023. Artificial intelligence and machine learning in clinical medicine. *New England Journal of Medicine* 388(13):1201–1208 DOI 10.1056/nejmra2302038.
- Hayashida K, Kondo Y, Hifumi T, Shimazaki J, Oda Y, Shiraishi S, Fukuda T, Sasaki J, Shimizu K. 2018. A novel early risk assessment tool for detecting clinical outcomes in patients with heat-related illness (J-ERATO score): development and validation in independent cohorts in Japan. *PLOS ONE* 13(5):e0197032 DOI 10.1371/journal.pone.0197032.
- Hoek AE, Dollee N, Prins G, Alsmas J. 2023. Hitteziekte [Heat illness]. *Nederlands Tijdschrift voor Geneeskunde* 167:D7442.
- Lin Q-W, Zhong L-C, He L-P, Zeng Q-B, Zhang W, Song Q, Song J-C. 2024. A newly proposed heatstroke-induced coagulopathy score in patients with heat illness: a multicenter retrospective study in China. *Chinese Journal of Traumatology* 27(2):83–90 DOI 10.1016/j.cjtee.2023.08.001.
- Liu SY, Song JC, Mao HD, Zhao JB, Song Q, Expert Group of Heat Stroke Prevention and Treatment of the People's Liberation Army, and People's Liberation Army Professional Committee of Critical Care Medicine. 2020. Expert consensus on the diagnosis and treatment of heat stroke in China. *Military Medical Research* 7(1):1 DOI 10.1186/s40779-019-0229-2.
- Lundberg S, Lee SI. 2017. A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems*.
- Nielsen RL, Monfeuga T, Kitchen RR, Egerod L, Leal LG, Schreyer ATH, Gade FS, Sun C, Helenius M, Simonsen L, Willert M, Tahrani AA, McVey Z, Gupta R. 2024. Data-driven identification of predictive risk biomarkers for subgroups of osteoarthritis using interpretable machine learning. *Nature Communications* 15(1):2817 DOI 10.1038/s41467-024-46663-4.
- Ogata S, Takegami M, Ozaki T, Nakashima T, Onozuka D, Murata S, Nakaoku Y, Suzuki K, Hagihara A, Noguchi T, Iihara K, Kitazume K, Morioka T, Yamazaki S, Yoshida T, Yamagata Y, Nishimura K. 2021. Heatstroke predictions by machine learning, weather information, and an all-population registry for 12-hour heatstroke alerts. *Nature Communications* 12(1):4575 DOI 10.1038/s41467-021-24823-0.
- Shimazaki J, Hifumi T, Shimizu K, Oda Y, Kanda J, Kondo Y, Shiraishi S, Takauji S, Hayashida K, Moriya T, Yagi M, Yamaguchi J, Yokota H, Yokobori S, Wakasugi M, Yaguchi A, Miyake Y. 2020. Clinical characteristics, prognostic factors, and outcomes of heat-related illness (Heatstroke Study 2017–2018). *Acute Medicine & Surgery* 7(1):e516 DOI 10.1002/ams2.516.
- Staatjes VE, Kernbach JM, Stumpo V, van Niftrik CHB, Serra C, Regli L. 2022. Foundations of feature selection in clinical prediction modeling. *Acta Neurochirurgica Supplement* 134(13):51–57 DOI 10.1007/978-3-030-85292-4_7.
- Takada A, Koderia S, Suzuki K, Nemoto M, Egawa R, Takizawa H, Hirata A. 2023. Estimation of the number of heat illness patients in eight metropolitan prefectures of Japan: correlation with ambient temperature and computed thermophysiological responses. *Frontiers in Public Health* 11:1061135 DOI 10.3389/fpubh.2023.1061135.
- Van Calster B, Wynants L. 2019. Machine learning in medicine. *The New England Journal of Medicine* 380(26):2588 DOI 10.1056/NEJMra1814259.
- Wan L, Shi X, Yang J, Qian J, Wang F, Chen R, Tong H. 2023. Construction and validation of the nomogram based on von willebrand factor predicting mortality in patients with heatstroke. *Therapeutic Hypothermia and Temperature Management* 13(4):191–199 DOI 10.1089/ther.2022.0059.
- Wang Y, Li D, Wu Z, Zhong C, Tang S, Hu H, Lin P, Yang X, Liu J, He X, Zhou H, Liu F. 2023. Development and validation of a prognostic model of survival for classic heatstroke patients: a multicenter study. *Scientific Reports* 13(1):19265 DOI 10.1038/s41598-023-46529-7.

- Wang Y, Song Q, Du Y, Wang J, Zhou J, Du Z, Li T. 2019.** A random forest model to predict heatstroke occurrence for heatwave in China. *Science of the Total Environment* **650**(Pt 2):3048–3053 DOI [10.1016/j.scitotenv.2018.09.369](https://doi.org/10.1016/j.scitotenv.2018.09.369).
- Watson DS, Krutzinna J, Bruce IN, Griffiths CE, McInnes IB, Barnes MR, Floridi L. 2019.** Clinical applications of machine learning algorithms: beyond the black box. *BMJ* **364**:l886 DOI [10.1136/bmj.l886](https://doi.org/10.1136/bmj.l886).
- Wei D, Gu T, Yi C, Tang Y, Liu F. 2022.** A nomogram for predicting patients with severe heatstroke. *Shock* **58**(2):95–102 DOI [10.1097/shk.0000000000001962](https://doi.org/10.1097/shk.0000000000001962).
- Yeoh CW, Law WC. 2023.** A case report of near-missed heat stroke. *Medicine* **102**(51):e36676 DOI [10.1097/md.00000000000036676](https://doi.org/10.1097/md.00000000000036676).