



## Original Article

Int Neurourol J 2024;28(1):33-43  
<https://doi.org/10.5213/inj.2346332.166>  
pISSN 2093-4777 · eISSN 2093-6931



# Prediction of Prostate Cancer Risk Stratification Based on A Nonlinear Transformation Stacking Learning Strategy

Xinyu Cao<sup>1,\*</sup>, Yin Fang<sup>1,\*</sup>, Chunguang Yang<sup>2</sup>, Zhenghao Liu<sup>2</sup>, Guoping Xu<sup>1</sup>, Yan Jiang<sup>1</sup>, Peiyan Wu<sup>1</sup>, Wenbo Song<sup>1</sup>, Hanshuo Xing<sup>1</sup>, Xinglong Wu<sup>1</sup>

<sup>1</sup>School of Computer Science & Engineering, Wuhan Institute of Technology, Wuhan, China

<sup>2</sup>Department of Urology, Tongji Hospital Affiliated Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

**Purpose:** Prostate cancer (PCa) is an epithelial malignancy that originates in the prostate gland and is generally categorized into low, intermediate, and high-risk groups. The primary diagnostic indicator for PCa is the measurement of serum prostate-specific antigen (PSA) values. However, reliance on PSA levels can result in false positives, leading to unnecessary biopsies and an increased risk of invasive injuries. Therefore, it is imperative to develop an efficient and accurate method for PCa risk stratification. Many recent studies on PCa risk stratification based on clinical data have employed a binary classification, distinguishing between low to intermediate and high risk. In this paper, we propose a novel machine learning (ML) approach utilizing a stacking learning strategy for predicting the tripartite risk stratification of PCa.

**Methods:** Clinical records, featuring attributes selected using the lasso method, were utilized with 5 ML classifiers. The outputs of these classifiers underwent transformation by various nonlinear transformers and were then concatenated with the lasso-selected features, resulting in a set of new features. A stacking learning strategy, integrating different ML classifiers, was developed based on these new features.

**Results:** Our proposed approach demonstrated superior performance, achieving an accuracy of 0.83 and an area under the receiver operating characteristic curve value of 0.88 in a dataset comprising 197 PCa patients with 42 clinical characteristics.

**Conclusions:** This study aimed to improve clinicians' ability to rapidly assess PCa risk stratification while reducing the burden on patients. This was achieved by using artificial intelligence-related technologies as an auxiliary method for diagnosing PCa.

**Keywords:** Prostatic neoplasms; Risk stratification; Machine learning; Stacking learning; Artificial intelligence

- **Grant/Fund Support:** This work was supported by the graduate innovation fund of Wuhan Institute of Technology (No. CX2022330) and the graduate innovation fund of Wuhan Institute of Technology (No. CX2022334).
- **Research Ethics:** The study was approved by the Research Ethics Commission of Wuhan Tongji Hospital and the requirement for informed consent was waived by the Ethics Commission (IRB ID: TJ- IRB20211246).
- **Conflict of Interest:** No potential conflict of interest relevant to this article was reported.

### • HIGHLIGHTS

- This article introduces a novel machine learning approach utilizing a stacking strategy for prostate cancer (PCa) risk stratification, overcoming the limitations of prostate-specific antigen levels. Through the integration of clinical records with a novel set of features derived from various machine learning classifiers, this approach realized substantial improvements in accuracy and area under the receiver operating characteristic curve values. The method aims to enhance clinical PCa risk assessment, reducing unnecessary biopsies and patient burden.

**Corresponding author:** Xinglong Wu  <https://orcid.org/0000-0001-9778-0864>  
School of Computer Science and Engineering, Hubei Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan, China  
Email: xwu@wit.edu.cn

\*Xinyu Cao and Yin Fang contributed equally to this study as co-first authors.

**Submitted:** December 4, 2023 / **Accepted after revision:** January 4, 2024



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## INTRODUCTION

Prostate cancer (PCa) is the second most common cancer among men worldwide [1]. The incidence and mortality rates of PCa increase with age [2]. The gold-standard method for diagnosing PCa in patients is a pathological examination of tissue obtained via biopsy. The Gleason score (GS) is recognized as a critical predictor in many PCa studies [3]. Risk stratification for PCa typically involves assessing serum prostate-specific antigen (PSA) levels, GS, and a range of clinical factors. This stratification system classifies PCa into low, intermediate, or high risk, which helps clinicians make informed treatment choices and prognostic evaluations.

The treatment of patients with PCa varies significantly according to risk stratification. For instance, very low and low-risk patients may undergo active surveillance, intermediate-risk patients often receive radical treatment, and those with high-risk or metastatic disease typically undergo comprehensive androgen deprivation therapy-based treatment [4]. Inaccurate risk assessment of PCa patients can result in severe psychological distress and substantial financial burdens. Consequently, there is a critical need to develop a non-invasive approach that can efficiently and accurately determine the malignancy level in PCa patients and predict their risk stratification [5].

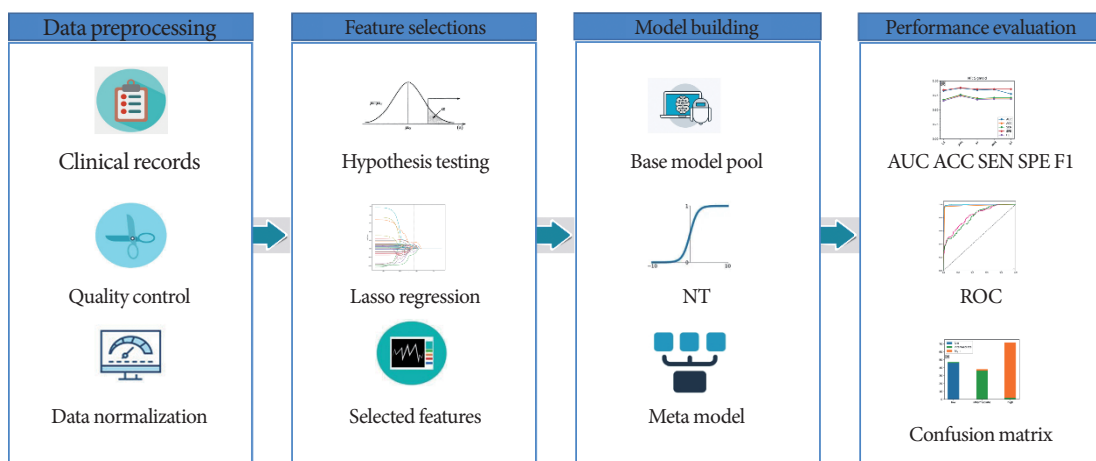
Current efforts to stratify PCa risk are broadly categorized into 2 approaches: traditional machine learning (ML) methods that rely on clinical records and deep learning methods that utilize multimodal data, incorporating both clinical records and medical images. In 2019, Hood et al. [6] selected 8 features

from 32 clinical characteristics using a genetic algorithm and a statistical test to construct a feature subset. They then assembled 3 K-nearest neighbors models to classify PCa malignancy and to predict low/intermediate and high-risk dichotomies. In 2021, Liang et al. [7] reported good performance in predicting PCa malignancy using a multiparametric radiomic model and a combined clinical-radiomic model. More recently, in 2023, Yang et al. [8] demonstrated good performance in predicting PCa risk stratifications based on functional subsets of peripheral lymphocytes. However, these studies were limited by the relatively small size of the clinical datasets used and did not offer a tripartite risk stratification.

This study introduces a novel stacking learning approach designed to predict low, intermediate, and high-risk stratifications for patients with PCa. The performance of the algorithm was evaluated and benchmarked against widely-used ML models as well as the conventional stacking algorithm.

The contributions of this work are summarized as follows.

- A novel stacking learning classification algorithm was developed by integrating nonlinear transformation (NT) strategies with an ML model pool.
- The proposed algorithm was validated on a dataset of 197 patients with 42 clinical characteristics, demonstrating superior performance for PCa risk stratification. Additionally, our algorithm was validated with the Kaggle Challenge (PI-CAI [9]) dataset.
- The proposed algorithm is publicly available as the “stackingNT” Python package, making it easily accessible to the research community.



**Fig. 1.** Workflow for the proposed approach for prostate cancer risk stratification. NT, nonlinear transformation; AUC, area under the curve; ACC, accuracy; SEN, sensitivity; SPE, specificity; ROC, receiver operating characteristic.

## MATERIALS AND METHODS

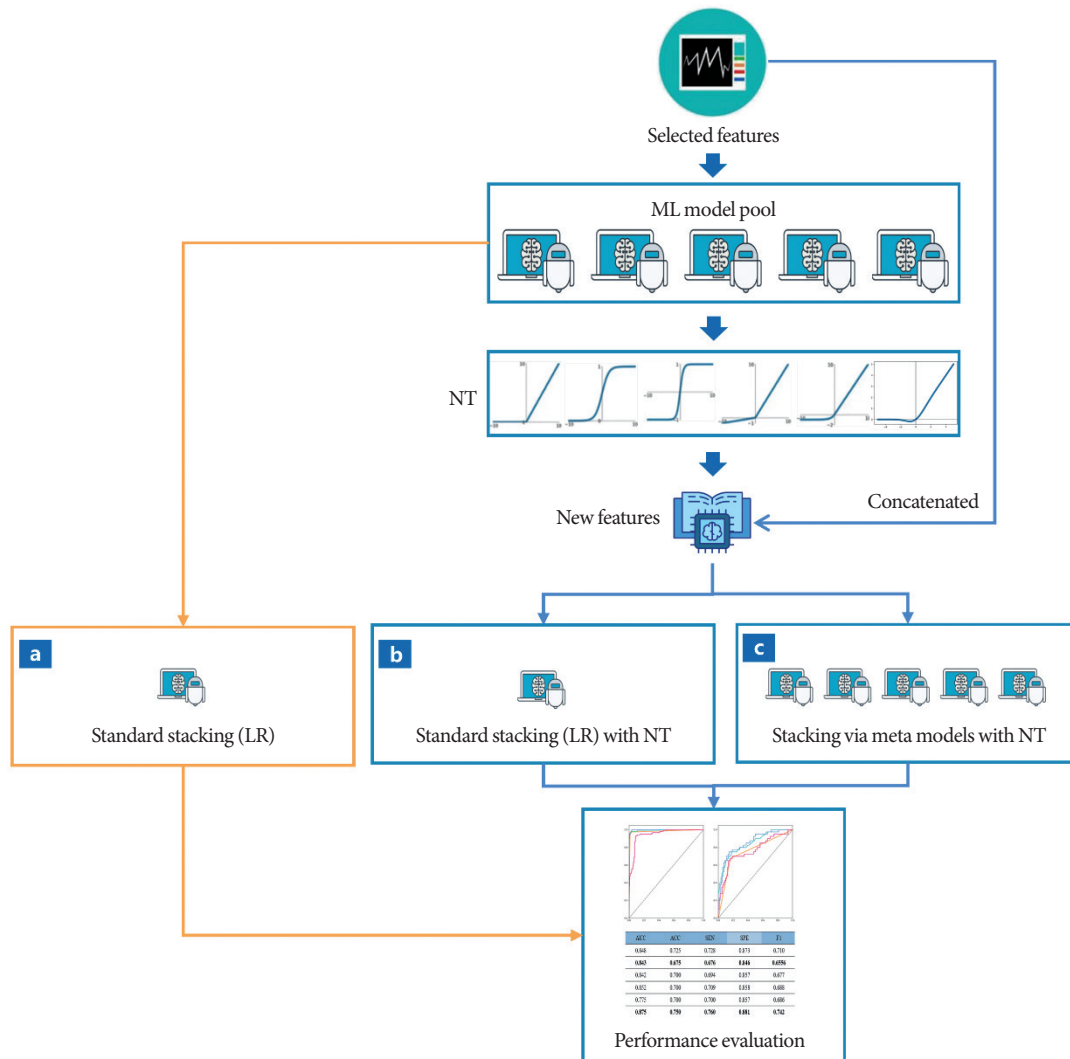
First, clinical records were preprocessed using the lasso algorithm to identify a refined subset of features, and the statistical significance of these selected features was then verified. Second, we trained 5 different ML classifiers as base models [10]. Based on the prediction results from these base models, various NT strategies were applied to derive a new set of features. Third, these newly generated features were combined with the initially selected clinical features to form the input for an ML model, referred to as the meta-model [11]. Concurrently, we evaluated the performance of different ML classifiers when used as meta-

models, which influenced the experimental outcomes. Finally, we selected the combination of the highest-performing meta-model and NT strategy for the predictive outcome. Additionally, we compared our proposed method with popular ML algorithms and a stacking algorithm based on logistic regression (LR) [12].

The workflow of this study is shown in Fig. 1. Fig. 2 illustrates the construction pipeline of the NT stacking learning strategy.

### Data Collection and Preprocessing

The data for this study were collected at Wuhan Tongji Hospital between August 1, 2020 and October 20, 2022. We recruited a



**Fig. 2.** Pipeline of the proposed NT stacking strategy (c). For the performance comparison, the pipelines of standard stacking (LR) without NT (a) and standard stacking (LR) with NT (b) are also shown. ML, machine learning; NT, nonlinear transformation; LR, logistic regression.

total of 197 PCa patients, which included 59 low-risk, 48 intermediate-risk, and 90 high-risk patients. The clinical records were preprocessed to remove outliers and duplicate entries. We capped the values for PSA at 1,000 ng/mL for readings above this threshold, set alanine transaminase values below 5 U/L to 5 U/L, and adjusted interleukin-6 values below 1.5 pg/mL to 1.5 pg/mL. Interleukin-10 was removed from the records due to an excessive number of duplicate values. After this preprocessing, 41 features were retained for analysis. It is important to note that the stratification of PCa patients into 3 risk groups followed the European Association of Urology guidelines [13]. The dataset was subsequently divided in a 4:1 ratio, with 80% of the data being used for training, employing a 10-fold cross-validation approach, and the remaining 20% reserved for performance evaluation.

### Data Normalization and Lasso for Feature Screening

Data normalization ensures that preprocessed data are confined within a specific range, thereby mitigating the adverse effects of anomalous sample data [14]. We employed z-score normalization as our method of choice, as follows.

$$Z_i = \frac{X_i - U}{\sigma} \tag{1}$$

Where X is the individual sample value, U is the total sample mean, and  $\sigma$  is the total sample standard deviation.

When using ML techniques to process large datasets, it is often necessary to reduce the dimensionality of the data. The lasso algorithm operates on the principle of minimizing the sum of squared residuals, subject to the condition that the sum of the absolute values of the regression coefficients remains below a specified constant. This approach allows for some regression coefficients to be exactly zero, resulting in a more interpretable model. The equation for this model is presented as Equation (2) [15].

$$B_{Lasso} = arg_B min \{ |Y - \sum_{j=1}^p X_j B_j|^2 + \alpha \sum_{j=1}^p |B_j| \} \tag{2}$$

Where Y is the target variable, X is the eigenmatrix, B is the parameter vector, P is the number of variables, and  $\alpha > 0$  is a hyperparameter. The compression of the overall regression coefficient can be achieved by adjusting the hyperparameter  $\alpha$ . The value of  $\alpha$  can be estimated using the cross-validation method [16].

### ML Algorithms and Standard Stacking Algorithm

This study utilized 5 popular ML algorithms — support vector machine (SVM) [17], decision tree (DT) [18], random forest (RF) [19], XGBoost (XGB) [20], and LR—to predict the risk stratification of PCa patients. These models were trained using 10-fold cross-validation on a training set that contained 8 clinical features, and then evaluated for performance on a test set. We then selected the best-performing ML model for the subsequent performance assessment.

For the standard stacking (LR) algorithm, this study utilized the same 5 ML classifiers, including SVM, DT, RF, XGB, and LR, as the base model [21]. After lasso and hypothesis testing, 8 selected features were entered into the base model, the prediction results were obtained. To better preserve the information from the original clinical features, we were inspired by shortcut connections in the ResNet network [22] and concatenated the ML models' predicted results with the selected features of the clinical records to form a new dataset, which was used as the training data for the meta-model. We used LR as the meta-model and evaluated its performance on the test set. The algorithmic process is shown in Fig. 2a.

### Standard Stacking (LR) With NT

Various NT strategies, including the cosine function (cos), Gaussian error linear units (Gelu) [23], logistic sigmoid (sigmoid), sine function (sin), Softplus function (Softplus) [24], and hyperbolic tangent function (Tanh), were utilized with the outputs of 5 ML models to generate NT features. We concatenated these transformed features with the selected clinical features to form a set of new features, which were then fed them into the standard stacking (LR) algorithm (Fig. 2b).

### Development and Validation of the Proposed Stacking Learning Strategy

The proposed stacking learning strategy was developed using different NTs with various meta-models (Fig. 2c), searching for the optimal combination of an NT strategy and a meta-model. In the experiment, a total of 6 NT strategies and 5 ML models were utilized.

### Statistical Analysis

The implementation of ML algorithms, lasso regression, and receiver operating characteristic curve analysis was carried out using the Scikit-learn package in Python 3.8. Hypothesis testing was conducted using IBM SPSS Statistics ver. 26.0 (IBM Co., Ar-

**Table 1.** Clinical characteristics of patients

Characteristic	Low risk	Intermediate risk	High risk	P-value
No. of patients	59 (29.95)	48 (24.37)	90 (45.69)	
Activated T cells (CD3+HLA-DR+) (/μL)	17.28 ± 7.31	18.36 ± 6.27	17.81 ± 6.23	0.721
Activated Ts cells (CD3+CD8+HLA-DR+)/Ts (%)	40.85 ± 13.05	45.27 ± 11.88	45.01 ± 12.55	0.448
Age (yr)	63.12 ± 7.96	65.58 ± 6.92	67.92 ± 7.57	0.18
Alkaline phosphatase (U/L)	74.53 ± 28.39	67.15 ± 15.51	218.12 ± 520.38	0.002*
ALT (U/L)	21.02 ± 12.5	18.5 ± 11.29	20.52 ± 16.13	0.225
B cells (CD3-CD19+) (%)	12.74 ± 5.59	12.36 ± 6.31	12.05 ± 5.57	0.566
B cells (CD3-CD19+) (/μL)	219.53 ± 164.27	220.12 ± 175.31	177.59 ± 91.37	0.877
Hemoglobin (g/L)	135.68 ± 15.15	136.98 ± 10.81	129.52 ± 16.5	0.043*
IFN-γ+ CD4+ T cells/Th (%)	21.84 ± 7.95	21.39 ± 8.68	20.38 ± 7.67	0.283
IFN-γ+ CD8+ T cells/Ts (%)	60.86 ± 16.97	61.84 ± 13.24	62.45 ± 14.81	0.686
IFN-γ+ NK cells/NK (%)	75.29 ± 13.97	74.28 ± 13.78	74.21 ± 15.38	0.208
Induced regulatory T cells	2.8 ± 0.92	3.26 ± 0.98	3.2 ± 0.98	0.805
Interleukin-1β (pg/mL)	6.54 ± 3.92	6.41 ± 4.92	8.12 ± 7.35	0.008*
Interleukin-2R (U/mL)	418.14 ± 174.59	435.9 ± 139.03	599.64 ± 501.18	0.000*
Interleukin-6 (pg/mL)	3.9 ± 4.92	4.45 ± 11.66	10.07 ± 14.35	0.000*
Interleukin-8 (pg/mL)	22.16 ± 34.68	28.19 ± 38	34.14 ± 47	0.000*
Lactate dehydrogenase	166.59 ± 34.48	157.27 ± 31.93	195.16 ± 117.37	0.040*
Lymphocyte percentage (%)	24.26 ± 9.43	30.42 ± 9.17	26.58 ± 7.97	0.604
Lymphocytes (× 10 <sup>9</sup> /L)	1.58 ± 0.7	1.69 ± 0.6	1.55 ± 0.44	0.552
Memory Th cells (CD3+CD4+CD45RO+)/Th (%)	67.29 ± 13.78	65.11 ± 13.21	67.7 ± 14.31	0.030*
Naïve regulatory T cells	0.66 ± 0.38	0.9 ± 0.65	0.76 ± 0.4	0.066
Naïve Th cells (CD3+CD4+CD45RA+)/Th (%)	32.71 ± 13.78	34.89 ± 13.21	32.41 ± 14.18	0.000*
Neutrophil percentage (%)	64.88 ± 11.14	57.49 ± 10.75	60.45 ± 9.43	0.168
Neutrophils (× 10 <sup>9</sup> /L)	4.48 ± 1.98	3.41 ± 1.67	3.76 ± 1.52	0.030*
NK cells (CD3-/CD16+CD56+) (%)	18.51 ± 8.32	18.99 ± 9.01	19.03 ± 10.31	0.887
NK cells (CD3-/CD16+CD56+) (/μL)	293.95 ± 160.09	316.33 ± 176.38	293.26 ± 220.92	0.001*
PSA	7.46 ± 6.79	12.03 ± 8.79	202.43 ± 308.62	0.524
Regulatory T cells (CD3+CD4+CD25+CD127low+)	3.47 ± 1.18	4.15 ± 1.38	3.95 ± 1.17	0.047*
Serum creatinine (mmol/L)	88.58 ± 57.71	84.08 ± 13.08	94.47 ± 69.27	0.927
T cells (CD3+CD19-) (%)	67.86 ± 8.37	67.91 ± 8.71	68.17 ± 10.14	0.611
T cells (CD3+CD19-) (/μL)	1,093.15 ± 416.45	1,154.75 ± 378.38	1,003.7 ± 282.78	0.262
T cells+B cells+NK cells (%)	99.11 ± 1.01	99.26 ± 0.56	99.25 ± 0.57	0.938
T cells+B cells+NK cells (/μL)	1,606.63 ± 600.92	1,691.21 ± 517.19	1,474.54 ± 412.02	0.397
Th cells (CD3+CD4+) (%)	43.32 ± 7.84	43.58 ± 9.07	44.74 ± 8.66	0.223
Th cells (CD3+CD4+) (/μL)	710.27 ± 327.36	740.71 ± 277.75	658.14 ± 201.11	0.157
Th cells+CD28+(CD3+CD4+CD28+)/Th	94.02 ± 8.22	93.51 ± 8.39	94.86 ± 6.64	0.155
Th/Ts	2.26 ± 0.95	2.42 ± 1	2.55 ± 1.15	0.563
Ts cells (CD3+CD8+) (%)	21.78 ± 7.49	19.82 ± 5.7	19.91 ± 6.39	0.309
Ts cells (CD3+CD8+) (/μL)	340.36 ± 134.31	335.19 ± 142.71	294.52 ± 124.24	0.812
Ts cells+CD28+(CD3+CD8+CD28+)/Ts	60.04 ± 20.98	56.25 ± 16.55	59.44 ± 15.48	0.000*
Tumor necrosis factor-α (pg/mL)	17.93 ± 21.93	19.01 ± 27.68	22.76 ± 37.41	0.004*

Values are presented as number (%) or mean ± standard deviation.

HLA, human leukocyte antigen; Ts, suppressor T cells; IFN, interferon; NK, Natural Killer; Th, T helper; ALT, alanine transaminase; IFN, interferon; NK, natural killer; PSA, prostate-specific antigen.

\*P < 0.05, statistically significant differences.

monk, NY, USA). A 2-sided P-value of less than 0.05 was considered statistically significant. Analysis of variance was used for continuous variables conforming to normal distribution and homogeneity of variance [25]. Other continuous and categorical variables were tested using the Kruskal-Wallis H test [26].

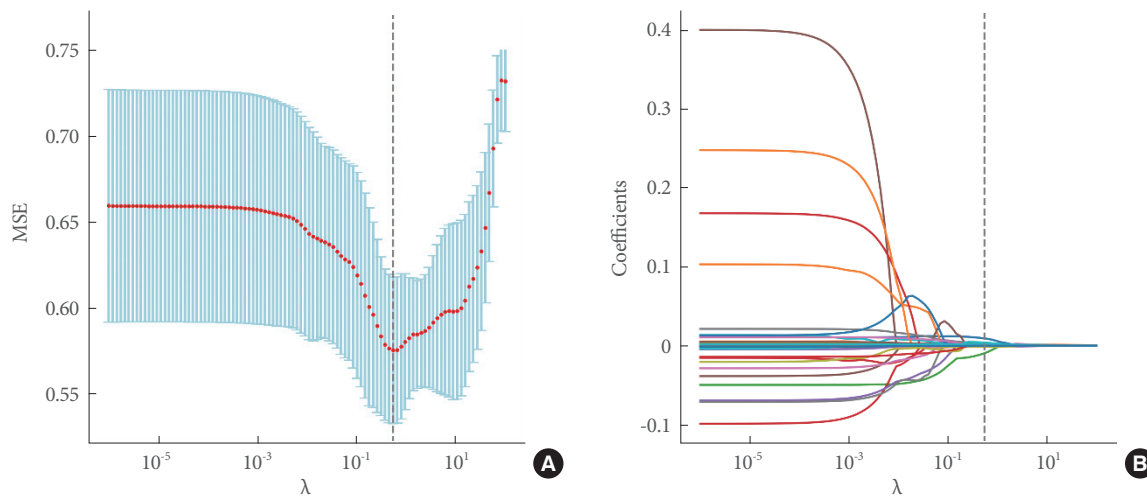
## RESULTS

### Characteristics of Patients

The study examined 41 features in patients with low, intermedi-

ate, and high-risk profiles and identified statistically significant differences in 14 clinical characteristics ( $P < 0.05$ ). These characteristics include age, PSA levels, neutrophil counts, hemoglobin levels, alkaline phosphatase activity, lactate dehydrogenase activity, T-cell counts (CD3+CD8+), regulatory T-cell counts (CD3+CD4+CD25+CD127low+), induced regulatory T-cell counts (CD45RO+CD3+CD4+CD25+CD127low+), and levels of interleukin-1, interleukin-2R, interleukin-6, interleukin-8, and tumor necrosis factor (refer to Table 1).

We then performed a feature selection process using the lasso



**Fig. 3.** (A) Five-fold cross-validation for tuning the hyperparameter  $\lambda$  selection in lasso regression. (B) Lasso coefficient profiles of the 41 variables. As the value of  $\lambda$  decreased, the degree of model compression increased, and 14 features were selected. MSE, mean squared error.

**Table 2.** Performance evaluation of ML models, the standard stacking (LR) algorithm, and the standard stacking (LR) algorithm with NT

NT	ML strategy	AUC	Accuracy	Sensitivity	Specificity	F1
N/A	SVM	0.757	0.550	0.520	0.783	0.503
	LR	0.764	0.575	0.544	0.792	0.537
	RF	0.807	0.675	0.661	0.845	0.643
	XGB	<b>0.789</b>	<b>0.700</b>	<b>0.710</b>	<b>0.860</b>	<b>0.688</b>
	DT	0.756	0.675	0.661	0.839	0.642
	Stacking (LR)	<b>0.839</b>	<b>0.817</b>	<b>0.728</b>	<b>0.869</b>	<b>0.711</b>
Tanh	Stacking (LR)	0.846	0.783	0.676	0.846	0.656
Sigmoid		0.820	0.783	0.676	0.843	0.656
Sin		0.823	0.783	0.676	0.843	0.656
Cos		0.847	0.816	0.733	0.869	0.717
Softplus		0.846	0.816	0.733	0.870	0.717
Gelu		<b>0.856</b>	<b>0.816</b>	<b>0.733</b>	<b>0.870</b>	<b>0.717</b>

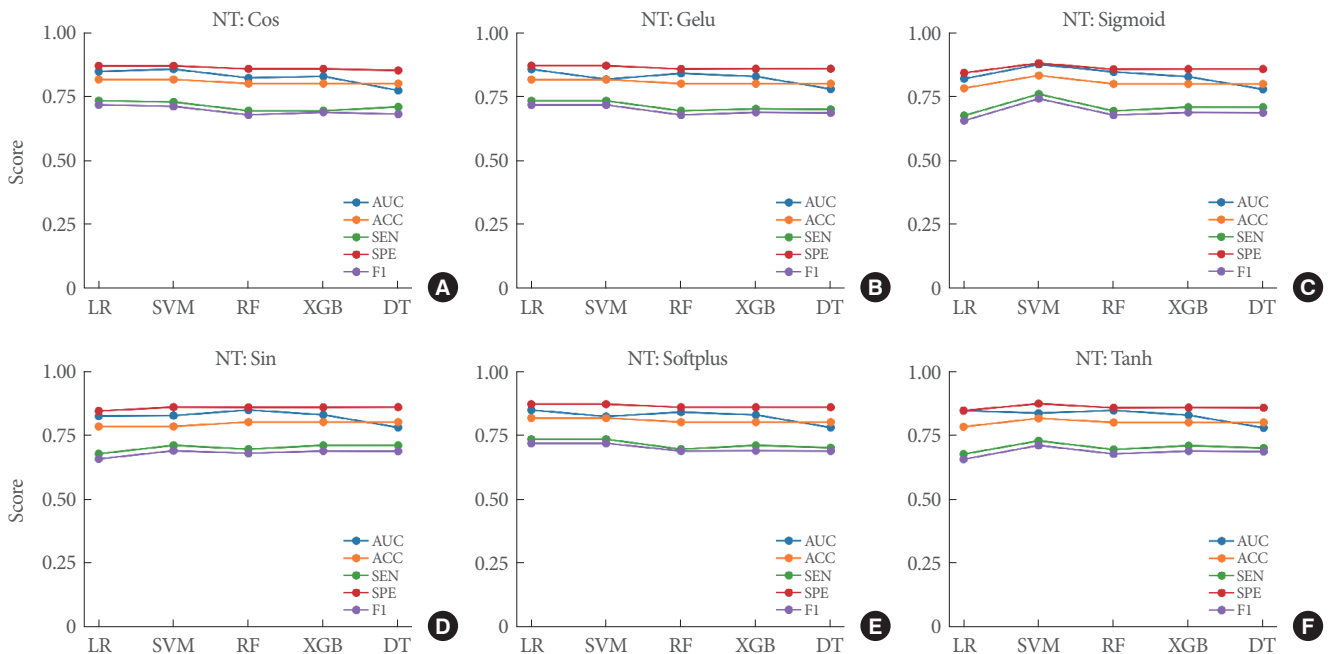
ML, machine learning; LR, logistic regression; NT, nonlinear transformation; AUC, area under the curve; SVM, support vector machine; RF, random forest; XGB, XGBoost; DT, decision tree; Tanh, hyperbolic tangent; Sigmoid, logistic sigmoid; Sin, sine function; Cos, cosine; Softplus, Softplus function; Gelu, Gaussian error linear units.

algorithm and identified an optimal set of 8 clinical features for the subsequent analysis. These features, which are presented as the final trial features in Fig. 3, include age, PSA, alkaline phosphatase, lactate dehydrogenase, regulatory T cells (CD3+CD4+CD25+CD127low+), interleukin-1 $\beta$ , interleukin-2R, and interleukin-6.

**Performance Evaluation of ML Models, the Standard Stacking (LR) Algorithm, and the Standard Stacking (LR) Algorithm With NT**

The 8 selected features were learned with 5 popular ML models,

the standard stacking (LR) algorithm and the standard stacking (LR) algorithm with the NT strategy. The performance is shown in Table 2. Among the ML models, the XGB model achieved the best overall performance, with an area under the curve (AUC) of 0.789 and an accuracy of 0.700, and the standard stacking (LR) algorithm achieved an AUC of 0.839 and an accuracy of 0.817. The standard stacking (LR) algorithm with additional Gelu transformations achieved an AUC of 0.856 and an accuracy of 0.816. The NT strategy introduced in the standard stacking (LR) algorithm significantly improved the performance of the meta-model.



**Fig. 4.** Performance evaluation of meta-models with different NT strategies. Panels A–F indicate the performance of NT: Cos, NT: Gelu, NT: Sigmoid, NT: Sin, NT: Softplus, NT: Tanh in different meta-models, respectively, for the risk stratification of prostate cancer. NT, nonlinear transformation; Cos, cosine; Gelu, Gaussian error linear units; Sigmoid, logistic sigmoid; Tanh, hyperbolic tangent; LR, logistic regression; SVM, support vector machine; RF, random forest; XGB, XGBoost; DT, decision tree.

**Table 3.** Performance comparison of XGB, standard stacking (LR) algorithm, Gelu+standard stacking (LR) algorithm, and sigmoid+stacking (SVM) algorithm

NT	ML strategy	AUC	Accuracy	Sensitivity	Specificity	F1
N/A	XGB	0.790	0.700	0.710	0.860	0.690
	Stacking (LR)	0.839	0.817	0.728	0.869	0.711
Gelu	Stacking (LR)	0.856	0.816	0.733	0.870	0.717
Sigmoid	Stacking (SVM)	<b>0.876</b>	<b>0.833</b>	<b>0.760</b>	<b>0.881</b>	<b>0.742</b>

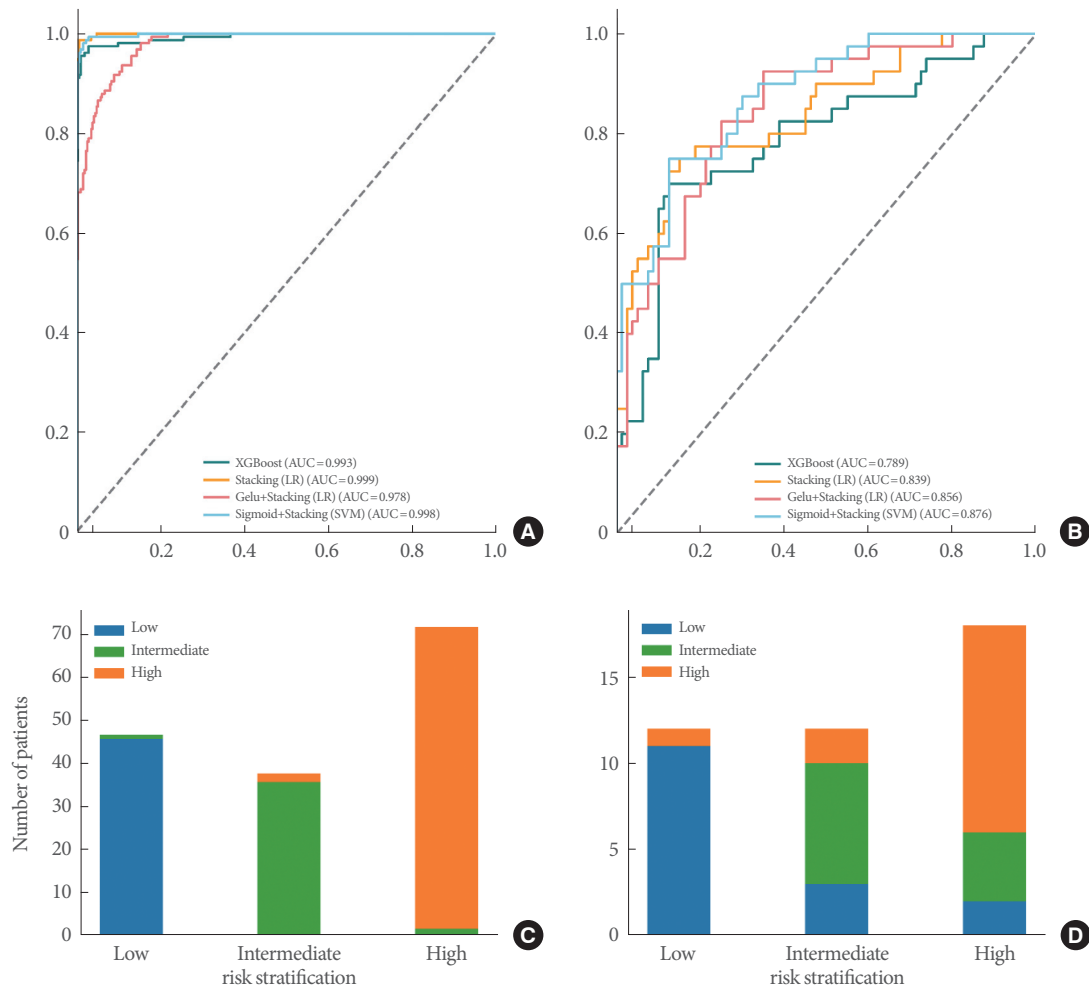
XGB, XGBoost; LR, logistic regression; Gelu, Gaussian error linear units; SVM, support vector machine; NT, nonlinear transformation; ML, machine learning; AUC, area under the curve; Sigmoid, logistic sigmoid.

### Performance of Different NT and ML Strategy Combinations in Stacking

The NT strategy, as introduced in the standard stacking (LR) algorithm, can significantly enhance the performance of the meta-model. We experimented with various NT combinations across different meta-models, as shown in Fig. 4. Our extensive comparative experiments revealed that the pairing of the sigmoid NT with SVM yielded the most superior results, achieving an AUC of 0.876 and an accuracy of 0.833, as detailed in Table 3. This optimal combination resulted in a 6.82% increase in the AUC and a 6.38% increase in accuracy when compared to the sigmoid+stacking (LR) combination.

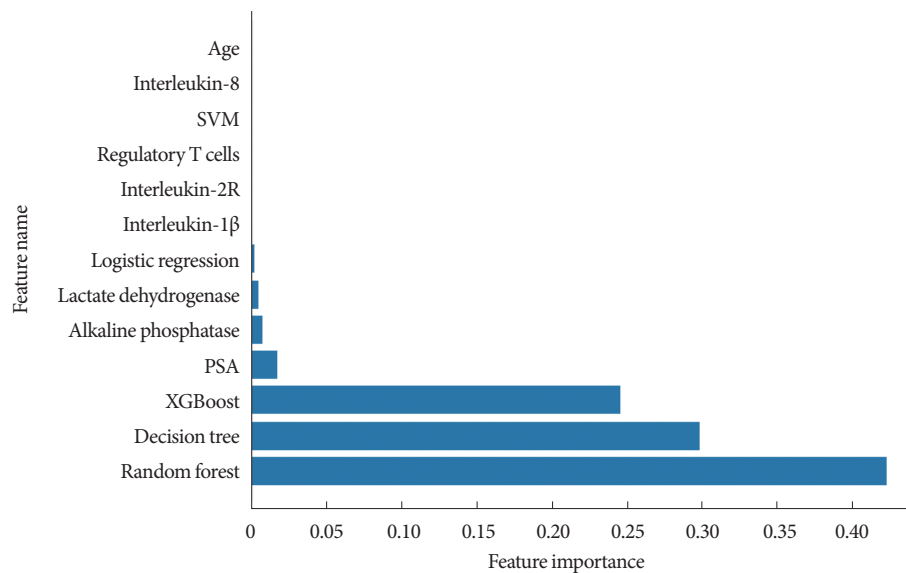
### Performance Evaluation of the NT Stacking Strategy

This section compares the performance of the optimal model combination (sigmoid+stacking [SVM]) with that of the XGB, standard stacking (LR), and Gelu+stacking (LR) models. The proposed optimal combination algorithm achieved a 10.88% improvement in AUC compared to the popular ML model XGB, a 4.41% improvement compared to the standard stacking (LR) algorithm, and a 2.33% improvement compared to the sin+stacking (LR) algorithm (Fig. 5A). The figure demonstrates that our method substantially improves risk stratification for active surveillance in PCa (Fig. 5B). Furthermore, the proposed NT stacking algorithm outperformed other methods across all performance metrics, including accuracy of 0.833, sensitivity of 0.760, specificity of 0.881, and F1 of 0.742 (Table 3). These ex-



**Fig. 5.** Receiver operating characteristic curves for XGBoost (XGB), standard stacking (LR), Gelu+standard stacking (LR), and sigmoid+stacking (SVM) in the training set (A) and testing set (B). Number of prostate cancer patients in low-, intermediate-, and high-risk groups according to the sigmoid+stacking (SVM) predictive scores in the training (C) and test set (D). AUC, area under the curve; LR, logistic regression; Gelu, Gaussian error linear units; Sigmoid, logistic sigmoid; SVM, support vector machine.





**Fig. 6.** Feature importance of the NT: sigmoid+stacking (SVM) strategy. NT, nonlinear transformation; SVM, support vector machine; PSA, prostate-specific antigen.

perimental results demonstrate the superiority of the proposed optimal stacking algorithm combination in achieving improved performance.

### Analysis of Feature Importance in NT Stacking

Analyzing feature importance in ML models is essential for their interpretability, especially for models applied for auxiliary diagnoses. Therefore, it is necessary to investigate how clinical features and NT features from various ML models contribute to predictions made by NT stacking models. For the best-performing NT stacking model (i.e., sigmoid+stacking [SVM]), the analysis revealed that the top 3 most important features (NT of the outputs of XGB, DT, and RF) contributed nearly 95% (Fig. 6). This result demonstrates the effectiveness of the proposed NT strategy in stacking learning. In addition, PSA, as a clinical feature, was ranked the fourth most important feature.

## DISCUSSION

The study analyzed a cohort of 197 patients, from whom 42 clinical characteristics across functional subsets of peripheral lymphocytes were collected. After conducting Lasso regression and hypothesis testing, we identified an optimal subset of 8 clinical features. We then introduced a shortcut connection within the standard stacking (LR) algorithm by concatenating NT features with those selected by the lasso algorithm to create

a new training set. Additionally, we investigated the optimal combination of various NT strategies and popular ML models. The analysis of feature importance underscored the effectiveness of our proposed stacking strategy. When compared with 5 other ML algorithms and the standard stacking (LR), our proposed stacking algorithm demonstrated superior performance, achieving an accuracy of 0.833 and an AUC of 0.876.

Previous studies have primarily concentrated on the binary classification of PCa patients into low/intermediate and high-risk groups, often with a relatively small sample size. Related research has typically addressed the benign versus malignant binary classification, as well as the low/intermediate versus high-risk binary classification in PCa patients [6, 27]. In contrast, our study categorized 197 PCa patients into low, intermediate, and high-risk groups, providing a classification that more closely aligns with clinical treatment decisions. Additionally, to promote the practical use and reproducibility of our stacking algorithm, we have created an open-source Python package called “stackingNT.” This package, which includes an implementation of our algorithm, is accessible at <https://pypi.org/project/stackingNT/>. The package includes an implementation of our proposed algorithm. The installation instructions and user guide can be found at [stackingNT · PyPI](https://pypi.org/project/stackingNT/).

This study has several limitations. First, all clinical data were collected from a single center, which suggests that a multi-center cross-validation study would be valuable in the future. Sec-

ond, investigating additional ML models and NT techniques could improve the study. Finally, imaging data are crucial for the diagnosis of PCa, and incorporating imaging data modalities is expected to further improve the diagnostic efficiency of the proposed approach.

In summary, we have proposed a novel stacking algorithm that integrates an NT strategy with a pool of ML models to improve the predictive performance of risk stratification in PCa. We believe that this approach could be applied to a broader range of clinical modality data analyses.

### AUTHOR CONTRIBUTION STATEMENT

- Conceptualization: XC, CY, ZL, YJ, PW, WS, HX, XW
- Data curation: XC, YF, CY, ZL, YJ, PW, WS, HX
- Formal analysis: XC, WS, HX
- Funding acquisition: XW
- Methodology: XC, YF, GX, YJ, PW, XW
- Project administration: XC, GX, XW
- Visualization: XC, YF
- Writing - original draft: XC, YF, XW
- Writing - review & editing: XC, YF, XW

### ORCID

Xinyu Cao	0009-0005-0568-2279
Yin Fang	0000-0001-9040-7071
Chunguang Yang	0000-0002-3882-4940
Zhenghao Liu	0000-0003-1185-598X
Guoping Xu	0000-0003-1431-7191
Yan Jiang	0009-0006-1219-7770
Peiyan Wu	0009-0005-6538-7167
Wenbo Song	0009-0007-8694-0923
Hanshuo Xing	0009-0001-2521-4828
Xinglong Wu	0000-0001-9778-0864

### REFERENCES

1. Surveillance, Epidemiology, and End Results. Cancer Stat Facts: Prostate Cancer [Internet]. Bethesda (MD): National Cancer Institute; 2023 [cited 2024 Mar 16]. Available from: <https://seer.cancer.gov/statfacts/html/prost.html>.
2. Pienta KJ, Esper PS. Risk factors for prostate cancer. *Ann Intern Med* 1993;118:793-803.
3. Bulten W, Kartasalo K, Chen PC, Ström P, Pinckaers H, Nagpal K, et al. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. *Nat Med* 2022;28:154-63.
4. Selvadurai ED, Singhera M, Thomas K, Mohammed K, Woode-Amissah R, Horwich A, et al. Medium-term outcomes of active surveillance for localised prostate cancer. *Eur Urol* 2013;64:981-7.
5. Sanda MG, Cadeddu JA, Kirkby E, Chen RC, Crispino T, Fontanarosa J, et al. Clinically localized prostate cancer: AUA/ASTRO/SUO guideline. Part I: risk stratification, shared decision making, and care options. *J Urol* 2018;199:683-90.
6. Hood SP, Cosma G, Foulds GA, Johnson C, Reeder S, McArdle SE, et al. Identifying prostate cancer and its clinical risk in asymptomatic men using machine learning of high dimensional peripheral blood flow cytometric natural killer cell subset phenotyping data. *Elife* 2020;9:e50936.
7. Liang L, Zhi X, Sun Y, Li H, Wang J, Xu J, et al. A nomogram based on a multiparametric ultrasound radiomics model for discrimination between malignant and benign prostate lesions. *Front Oncol* 2021;11:610785.
8. Yang C, Liu Z, Fang Y, Cao X, Xu G, Wang Z, et al. Development and validation of a clinic machine-learning nomogram for the prediction of risk stratifications of prostate cancer based on functional subsets of peripheral lymphocyte. *J Transl Med* 2023;21:465.
9. Sunoqrot MRS, Saha A, Hosseinzadeh M, Elschot M, Huisman H. Artificial intelligence for prostate MRI: open datasets, available applications, and grand challenges. *Eur Radiol Exp* 2022;6:35.
10. Chiu PK, Shen X, Wang G, Ho CL, Leung CH, Ng CF, et al. Enhancement of prostate cancer diagnosis by machine learning techniques: an algorithm development and validation study. *Prostate Cancer Prostatic Dis* 2022;25:672-6.
11. Kim C, You SC, Reys JM, Cheong JY, Park RW. Machine-learning model to predict the cause of death using a stacking ensemble method for observational data. *J Am Med Inform Assoc* 2021;28:1098-107.
12. Domínguez-Almendros S, Benítez-Parejo N, Gonzalez-Ramirez AR. Logistic regression models. *Allergol Immunopathol (Madr)* 2011;39:295-305.
13. Mottet N, van den Bergh RCN, Briers E, Van den Broeck T, Cumberbatch MG, De Santis M, et al. EAU-EANM-ESTRO-ESUR-SIOG guidelines on prostate cancer-2020 update. Part 1: screening, diagnosis, and local treatment with curative intent. *Eur Urol* 2021;79:243-62.
14. Liu X, Li N, Liu S, Wang J, Zhang N, Zheng X, et al. Normalization methods for the analysis of unbalanced transcriptome data: a review. *Front Bioeng Biotechnol* 2019;7:358.
15. Wang T, Dai L, Shen S, Yang Y, Yang M, Yang X, et al. Comprehen-

- sive molecular analyses of a macrophage-related gene signature with regard to prognosis, immune features, and biomarkers for immunotherapy in hepatocellular carcinoma based on WGCNA and the LASSO algorithm. *Front Immunol* 2022;13:843408.
16. Pi L, Halabi S. Combined performance of screening and variable selection methods in ultra-high dimensional data in predicting time-to-event outcomes. *Diagn Progn Res* 2018;2:21.
  17. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics Proteomics* 2018;15:41-51.
  18. Szeghalmy S, Fazekas A. A comparative study of the use of stratified cross-validation and distribution-balanced stratified cross-validation in imbalanced learning. *Sensors (Basel)* 2023;23:2333.
  19. Hanko M, Grendár M, Snopko P, Opšenač R, Šutovský J, Benčo M, et al. Random forest-based prediction of outcome and mortality in patients with traumatic brain injury undergoing primary decompressive craniectomy. *World Neurosurg* 2021;148:e450-8.
  20. Hou N, Li M, He L, Xie B, Wang L, Zhang R, et al. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. *J Transl Med* 2020;18:462.
  21. Hwangbo L, Kang YJ, Kwon H, Lee JI, Cho HJ, Ko JK, et al. Stacking ensemble learning model to predict 6-month mortality in ischemic stroke patients. *Sci Rep* 2022;12:17389.
  22. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27-30; Las Vegas (NV), USA; 2016. p. 770-8.
  23. Ahmed B, Haque MA, Iquebal MA, Jaiswal S, Angadi UB, Kumar D, et al. DeepAProt: deep learning based abiotic stress protein sequence classification and identification tool in cereals. *Front Plant Sci* 2023; 13:1008756.
  24. Tang Z, Luo L, Xie B, Zhu Y, Zhao R, Bi L, et al. Automatic sparse connectivity learning for neural networks. *IEEE Trans Neural Netw Learn Syst* 2023;34:7350-64.
  25. Hu D, Wang C, O'Connor AM. A likelihood ratio test for the homogeneity of between-study variance in network meta-analysis. *Syst Rev* 2021;10:310.
  26. Sherwani RAK, Shakeel H, Awan WB, Faheem M, Aslam M. Analysis of COVID-19 data using neutrosophic Kruskal Wallis H test. *BMC Med Res Methodol* 2021;21:215.
  27. Cosma G, McArdle SE, Foulds GA, Hood SP, Reeder S, Johnson C, et al. Prostate cancer: early detection and assessing clinical risk using deep machine learning of high dimensional peripheral blood flow cytometric phenotyping data. *Front Immunol* 2021;12:786828.