

Application of Machine Learning Techniques to Predict Bone Metastasis in Patients with Prostate Cancer

Wen-Cai Liu ^{1,2}
Ming-Xuan Li²
Wen-Xing Qian³
Zhi-Wen Luo^{1,4}
Wei-Jie Liao^{1,4}
Zhi-Li Liu^{1,4}
Jia-Ming Liu ^{1,4}

¹Department of Orthopaedic Surgery, The First Affiliated Hospital of Nanchang University, Nanchang, 330006, People's Republic of China; ²The First Clinical Medical College of Nanchang University, Nanchang, 330006, People's Republic of China; ³School of Computer and Information Technology, Beijing Jiaotong University, Beijing, 100044, People's Republic of China; ⁴Institute of Spine and Spinal Cord, Nanchang University, Nanchang, 330006, People's Republic of China

Objective: This study aimed to develop and validate a machine learning model for predicting bone metastases (BM) in prostate cancer (PCa) patients.

Methods: Demographic and clinicopathologic variables of PCa patients in the Surveillance, Epidemiology and End Results (SEER) database from 2010 to 2017 were retrospectively analyzed. We used six different machine learning algorithms, including Decision tree (DT), Random forest (RF), Multilayer Perceptron (MLP), Logistic regression (LR), Naive Bayes classifiers (NBC), and eXtreme gradient boosting (XGB), to build prediction models. External validation using data from 644 PCa patients of the First Affiliated Hospital of Nanchang University from 2010 to 2016. The performance of the models was evaluated using the area under receiver operating characteristic curve (AUC), accuracy score, sensitivity (recall rate) and specificity. A web predictor was developed based on the best performance model.

Results: A total of 207,137 PCa patients from SEER were included in this study. Of whom, 6725 (3.25%) developed BM. Gleason score, Prostate-specific antigen (PSA) value, T, N stage and age were found to be the risk factors of BM. The XGB model offered the best predictive performance among these 6 models (AUC: 0.962, accuracy: 0.884, sensitivity (recall rate): 0.906, and specificity: 0.879). An XGB model-based web predictor was developed to predict BM in PCa patients.

Conclusion: This study developed a machine learning model and a web predictor for predicting the risk of BM in PCa patients, which may help physicians make personalized clinical decisions and treatment strategy for patients.

Keywords: prostate cancer, bone metastasis, machine learning, prediction model, SEER

Introduction

Prostate cancer (PCa) is the most common non-skin cancer among men globally, with approximately 1.6 million cases and 366,000 deaths reported each year.^{1,2} Metastatic prostate cancer has important clinical implications, and metastatic disease may occur on the occasion of the initial clinical diagnosis.^{3,4} Bone metastases (BM) take up a great proportion of patients with metastases, accounting for approximately 16.7% of all metastatic cases, and these patients had significantly reduced 5-year survival and quality of life.^{5,6}

Recently, prostate-specific membrane antigen (PSMA) ligands have presented good results in the diagnosis and treatment of BM from PCa.⁷ However, this method is still at the stage of clinical trial and is not acceptable for most patients

Correspondence: Jia-Ming Liu
Department of Orthopaedic Surgery, The First Affiliated Hospital of Nanchang University, No. 17 Yongwaizheng Street, Donghu District, Nanchang, Jiangxi Province, People's Republic of China
Tel/Fax +86-791-86319815
Email liujiamingdr@hotmail.com

due to the radiation injury and the high cost.^{8,9} The current method for detecting BM in PCa patients is mainly bone scan. But it just suggested for patients with suspected skeletal-related events (SRE) due to the severe radiation injury. In addition, the median time to SRE has been reported to be 5 months after bone metastases.¹⁰ An assisted decision-making system that can help determining which cancer patients should receive bone scan will provide help for addressing these issues. And with the development of precision medicine, the diagnosis and treatment of cancer should be individualized. The Surveillance, Epidemiology and End Results (SEER) database is a publicly obtainable, federally funded cancer reporting system that brings raw material to our insights into complex diseases.¹¹ With the increasing level of computer hardware, machine learning can facilitate the diagnosis of cancer metastasis by processing and analyzing large, heterogeneous and complex clinical data and building predictive models. And there are already studies that have been conducted in this area with good results.^{12–14}

Therefore, in this study, we aim to build a prediction model to evaluate the risk of BM in PCa patients based on machine learning techniques, and develop a web-based predictor that can be easily manipulated by physicians and patients. This study may provide some help for clinicians to make personalized decisions for the treatment of patients with PCa BM.

Materials and Methods

Study Population

The study is carried out based on the SEER database. Patients' data were obtained from the "SEER Research Plus Data, 18 Registries, Nov 2019 Sub (2000–2017)" and downloaded using SEER*stat 8.3.9 software. Patients who were diagnosed with PCa from 2010 to 2017 were included in this study. Exclusion criteria were as follows:¹ PCa was not the first tumor.² The information of race, grade, Prostate-specific antigen (PSA), Gleason score, T, N stage, metastatic status and marital status missed or unknown. Additionally, 644 PCa patients from the First Affiliated Hospital of Nanchang University from 2010 to 2016 were included for external validation. The case screening process is shown in Figure 1.

Data Selection

Eight variables from the SEER database that may affect BM in PCa patients were selected in this study, including age at

diagnosis, race, grade, PSA value, Gleason score, T, N stage and marital status. Besides, we only included patients with PSA values between 0.1 and 98.0ng/mL. Because no specific values over 98.0 ng/mL were provided by the SEER database. All patients enrolled in this study were staged using the 7th edition of the AJCC TNM staging system and relevant guidelines of the SEER program.

Statistical Analyses and Model Establishment

All statistical analysis was performed in Python (version 3.8, Python Software Foundation) and SPSS (version 26, IBM, USA).¹⁵ All machine learning algorithms were built based on scikit-learn (version 0.24.1). The patient data were randomly sliced into training and internal test sets in a ratio of 7:3 using python. Training set is used to build the model, and the internal test set is used for model validation and evaluation.

In order to determine the variables included in the machine learning model, we conducted a univariate analysis to compare these variables between patients with and without BM. The Chi-square test was utilized for categorical data, and the Wilcoxon rank-sum test was used for continuous non-normally distributed data. Variables with a $P < 0.05$ in univariate analysis were enclosed within the construction of machine learning models and multivariate logistic regression was performed to identify the risk factor for BM. Meanwhile, based on the Permutation Importance principle,¹⁶ we performed feature importance analysis on the variables in each machine learning model. Correlation analysis was performed on the screened variables to test whether the variables would affect each other.

This study used six different machine learning algorithms to model the data, including Decision tree (DT), Random Forest (RF), Multilayer Perceptron (MLP), Logistic regression (LR), Naive Bayes classifiers (NBC) and eXtreme gradient boosting (XGB).^{17–22} The ML algorithms were trained by using Python software to predict the BM in PCa patients. Model parameter settings are detailed in the [Supplementary Materials](#). Parameter settings link is accessible from <https://share.streamlit.io/liu-wencaincu/prostate-cancer/main/prostate.py>. Then, the predictive power of the machine learning models was evaluated in internal ten-fold cross-validation of the train set, internal test set and external test set. The area under the receiver operating characteristic curve (AUC), the

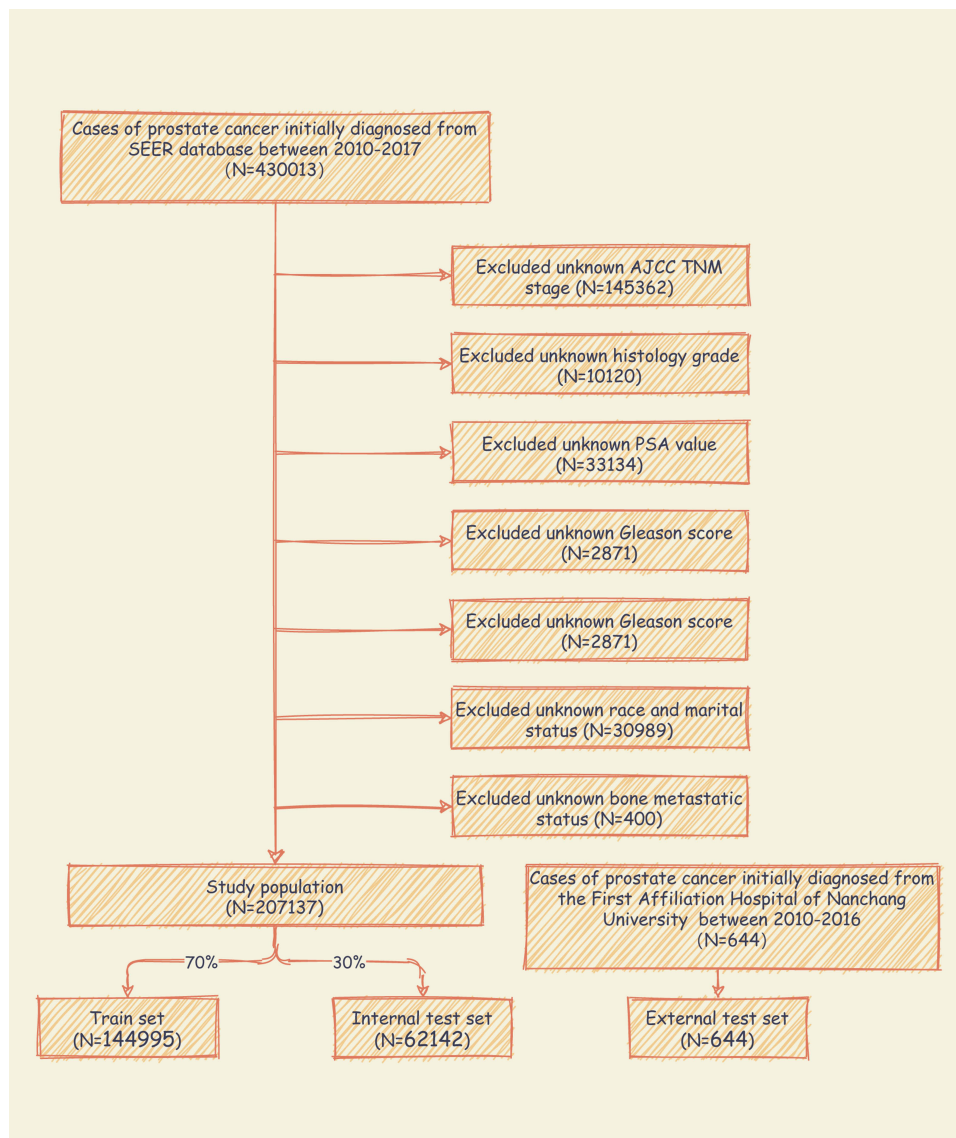


Figure 1 Flow diagram of the study population selected from the Surveillance, Epidemiology, and End Results (SEER) database and the First Affiliated Hospital of Nanchang University. According to the inclusion and exclusion criteria, a total of 207,137 patients of SEER were included in this study, and they were randomly cut into the training and internal test sets in a 7:3 ratio. Data from the First Affiliated Hospital of Nanchang University as an external test set.

sensitivity (recall rate), specificity, and accuracy score were calculated. The best-performing model was selected to build a web predictor.

Results

Demographic Baseline Characteristics

A total of 207,137 PCa patients from SEER database were included in the study. Of whom, 6725 (3.25%) developed BM and 200,412 (96.75%) had no BM. All demographic and clinicopathological characteristics of these patients were demonstrated in detail in Table 1. All patients were

randomly cut into a training set ($n = 144,995$) and an internal test set ($n = 62,142$) in a ratio of 7:3. External validation was conducted by using data of 644 PCa patients from the First Affiliated Hospital of Nanchang University. The details of the training and test sets are shown in Table 2.

Univariate and Multivariate Logistic Regression Analysis

Based on the univariate analysis, age, race, grade, T, N stage, PSA value, Gleason score and marital status

Table 1 Clinical and Pathological Characteristics of Study Population

Variables	ALL	NBM	BM
	N=207137	N=200412	N=6725
Age			
≤39	100 (<0.1%)	97 (<0.1%)	3 (<0.1%)
40–49	5902 (2.8%)	5760 (2.9%)	142 (2.1%)
50–59	46,532 (22.5%)	45,478 (22.7%)	1054 (15.7%)
60–69	92,231 (44.5%)	89,924 (44.9%)	2307 (34.3%)
≥70	62,372 (30.1%)	59,153 (29.5%)	3219 (47.9%)
Race			
American Indian/Alaska Native	813 (0.4%)	778 (0.4%)	35 (0.5%)
Asian or Pacific Islander	10,032 (4.8%)	9673 (4.8%)	359 (5.3%)
Black	33,822 (16.3%)	32,574 (16.3%)	1248 (18.6%)
White	162,470 (78.4%)	157,387 (78.5%)	5083 (75.6%)
Grade			
Grade I	20,130 (9.7%)	20,071 (10.0%)	59 (0.9%)
Grade II	254 (0.1%)	212 (0.1%)	42 (0.6%)
Grade III	102,729 (49.6%)	96,633 (48.2%)	6096 (90.6%)
Grade IV	84,024 (40.6%)	83,496 (41.7%)	528 (7.9%)
T stage			
T1	87,738 (42.4%)	85,198 (42.5%)	2540 (37.8%)
T2	90,974 (43.9%)	88,612 (44.2%)	2362 (35.1%)
T3	26,494 (12.8%)	25,520 (12.7%)	974 (14.5%)
T4	1931 (0.9%)	1082 (0.5%)	849 (12.6%)
N stage			
N0	199,991 (96.6%)	195,282 (97.4%)	4709 (70.0%)
N1	7146 (3.4%)	5130 (2.6%)	2016 (30.0%)
PSA (ng/mL)	6.5 [4.8;10.4]*	6.4 [4.8;9.9]*	77.4 [20.6;98.0]*
Gleason score			
≤6	82,688 (39.9%)	82,508 (41.2%)	180 (2.7%)
7	83,693 (40.4%)	82,688 (41.3%)	1005 (14.9%)
8	21,937 (10.6%)	20,265 (10.1%)	1672 (24.9%)
≥9	18,819 (9.1%)	14,951 (7.5%)	3868 (57.5%)
Marital status			
Married	181,913 (87.8%)	176,389 (88.0%)	5524 (82.1%)
Unmarried	25,224 (12.2%)	24,023 (12.0%)	1201 (17.9%)

Note: *Median [interquartile range, IQR].

Abbreviations: BM, bone metastasis; NBM, no bone metastasis; PSA, prostate specific antigen.

were significantly associated with the BM in PCa patients ($P < 0.05$) (Table 3). Variables with a P values <0.05 between these two groups were selected for multivariate logistic regression analysis. Based on the analysis, T, N stage, Gleason score and PSA value were found to be the independent risk factors for BM in PCa patients (Table 3).

Variable Influence and Feature Importance on Prediction

Correlation tests were performed among different variables identified from the univariate analysis. The correlation heat map showed that variables were not significantly correlated with each other (Figure 2), which indicated that they were mutually independent.

Table 2 Clinical and Pathological Characteristics of Training Set and Test Set

Variables	Training Set	BM(%) (n=4735)	Internal Test Set	BM(%) (n=1990)	External Test Set	BM(%) (n=117)
	NBM(%) (n=140,260)		NBM(%) (n=60,152)		NBM(%) (n=527)	
Age						
≤39	67 (<0.1)	2 (<0.1)	30 (0.1)	1 (<0.1)	0 (0.0)	0 (0.0)
40–49	4028 (2.8)	101 (2.1)	1732 (2.9)	41 (2.1)	8 (1.5)	2 (1.7)
50–59	31,862 (22.7)	762 (16.1)	13,616 (22.6)	292 (14.7)	91 (17.3)	19 (16.2)
60–69	62,724 (44.7)	1626 (34.3)	27,200 (45.2)	681 (34.2)	211 (40.0)	27 (23.1)
≥70	41,579 (29.6)	2244 (47.4)	17,574 (29.2)	975 (49.0)	217 (41.1)	69 (59.0)
Race						
American Indian/Alaska Native	557 (0.4)	24 (0.5)	221 (0.4)	11 (0.6)	0 (0)	0 (0)
Asian or Pacific Islander	6720 (4.8)	246 (5.2)	2953 (4.9)	113 (5.7)	527 (100)	117 (100)
Black	22,724 (16.2)	868 (18.3)	9850 (16.4)	380 (19.1)	0 (0)	0 (0)
White	110,259 (78.6)	3597 (76.0)	47,128 (78.3)	1486 (74.7)	0 (0)	0 (0)
Grade						
Grade I	14,050 (10.0)	39 (0.8)	6021 (10.0)	20 (1.0)	10 (1.9)	2 (1.7)
Grade II	152 (0.1)	29 (0.6)	60 (0.1)	13 (0.7)	1 (0.2)	1 (0.9)
Grade III	67,619 (48.2)	4291 (90.6)	29,014 (48.2)	1805 (90.7)	310 (58.8)	108 (92.3)
Grade IV	58,439 (41.7)	376 (7.9)	25,057 (41.7)	152 (7.6)	206 (39.1)	6 (5.1)
T stage						
T1	59,594 (42.5)	1773 (37.4)	25,604 (42.6)	767 (38.5)	292 (55.4)	41 (35.0)
T2	62,041 (44.2)	1675 (35.4)	26,571 (44.2)	687 (34.5)	193 (36.6)	34 (29.1)
T3	17,844 (12.7)	684 (14.4)	7676 (12.8)	290 (14.6)	41 (7.8)	26 (22.2)
T4	781 (0.6)	603 (12.7)	301 (0.5)	246 (12.4)	1 (0.2)	16 (13.7)
N stage						
N0	136,654 (97.4)	3325 (70.2)	58,628 (97.5)	1384 (69.5)	521 (98.9)	85 (72.6)
N1	3606 (2.5)	1410 (29.8)	1524 (2.5)	606 (30.5)	6 (1.1)	32 (27.4)
PSA (ng/mL)	6.4 (5.1)*	77.0 (77.9)*	6.4 (5.1)*	79.6 (76.3)*	7.4 (6.8)*	98.0 (62.9)*
Gleason score						
≤6	57,814 (41.2)	124 (2.6)	24,694 (41.1)	56 (2.8)	224 (42.5)	3 (2.6)
7	57,791 (41.2)	772 (15.2)	24,897 (41.4)	283 (14.2)	195 (37.0)	14 (12.0)
8	14,120 (10.1)	1153 (24.4)	6145 (10.2)	519 (26.1)	64 (12.1)	27 (23.1)
≥9	10,535 (7.5)	2736 (57.8)	4416 (7.3)	1132 (56.9)	44 (8.3)	73 (62.4)
Marital status						
Married	123,372 (88.0)	3925 (82.9)	53,017 (88.1)	1599 (80.4)	489 (92.8)	15 (12.8)
Unmarried	16,888 (12.0)	810 (17.1)	7135 (11.9)	391 (19.6)	38 (7.2)	102 (87.2)

Note: *Median (interquartile range, IQR).

Abbreviations: BM, bone metastasis; NBM, no bone metastasis; PSA, prostate specific antigen.

The importance of features in each machine learning model for predicting BM is shown in Figure 3. Although the importance of features varied slightly among different machine learning algorithms, PSA value, Gleason score and N stage ranked at the top three of five models, which had similarities with the results of multivariate

logistic analysis. In contrast, marital status ranked at the last place in most algorithms, but it also made some contributions to the models. In the XGB model, the importance of features were sorted in descending order by PSA value, Gleason score, N stage, T stage, age, grade, race and marital status.

Table 3 Univariate Analysis and Multivariate Logistic Regression Analysis of Variables

Variables	Univariate Analysis	Multivariate Logistic Analysis	P value
	P value	OR (95% CI)	
Age ≤39 40–49 50–59 60–69 ≥70	<0.001*	Reference 0.555 (0.076–4.060) 0.508 (0.071–3.661) 0.518 (0.072–3.370) 0.632 (0.088–4.555)	0.562 0.502 0.514 0.650
Race American Indian/Alaska Native Asian or Pacific Islander Black White	<0.001*	Reference 0.938 (0.548–1.603) 0.896 (0.533–1.506) 1.214 (0.727–2.030)	0.814 0.677 0.459
Grade Grade I Grade II Grade III Grade IV	<0.001*	Reference 0.709 (0.368–1.365) 0.773 (0.519–1.150) 0.840 (0.576–1.226)	0.303 0.204 0.367
T stage T1 T2 T3 T4	<0.001*	Reference 0.917 (0.843–0.997) 0.511 (0.456–0.572) 2.015 (2.641–3.231)	0.043* <0.001* <0.001*
N stage N0 N1	<0.001*	Reference 2.921 (2.641–3.231)	<0.001*
Gleason score ≤6 7 8 ≥9	<0.001*	Reference 4.502 (3.563–5.689) 15.828 (12.238–20.472) 32.566 (25.243–42.014)	<0.001* <0.001* <0.001*
PSA	<0.001*	1.039 (1.038–1.040)	<0.001*
Marital status Unmarried Married	<0.001* <0.001	Reference 0.965 (0.869–1.072)	0.511

Note: *P < 0.05.

Abbreviations: BM, bone metastasis; NBM, no bone metastasis; PSA, prostate specific antigen.

Model Performance

The predictive performance of different models was compared using the internal ten-fold cross-validation of training set, internal and external test sets, which were detailed in Figures 4–6 and Table 4. Among these models, the XGB model showed the best performance with an average AUC of 0.951 in the internal ten-fold cross-validation (Figure 4). In the internal test set, the XGB model gained

the best score with an AUC of 0.955, an accuracy of 0.881, a sensitivity (recall rate) of 0.905 and a specificity of 0.880. In the external test set, the XGB model also showed excellent performance with an AUC of 0.962, an accuracy of 0.884, a sensitivity (recall rate) of 0.906 and a specificity of 0.879 (Figures 5 and 6 and Table 4). Meanwhile, the prediction results of different models were presented with a heat map in Figure 7.

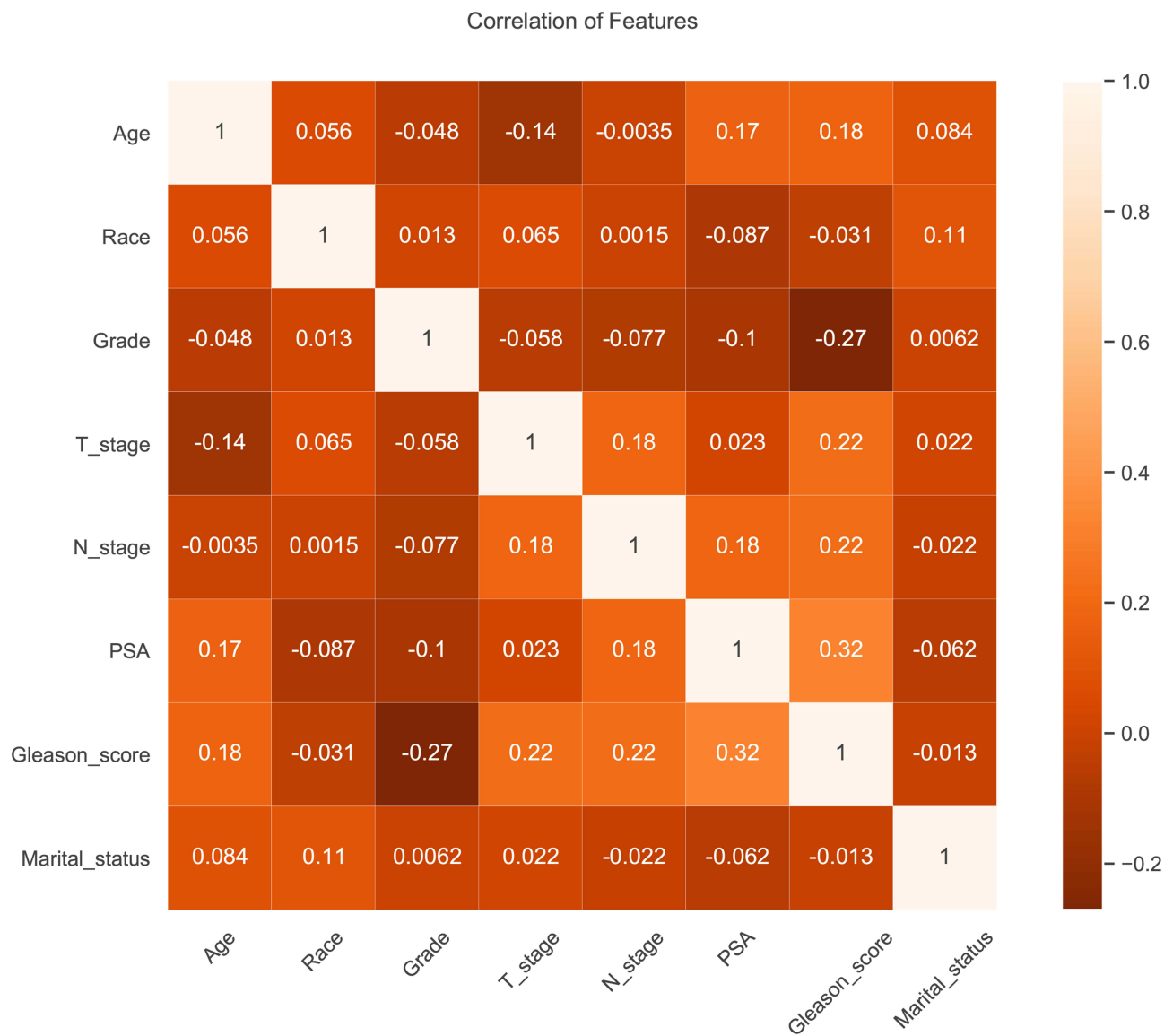


Figure 2 Results of correlation analysis between all variables. The heat map shows the correlation between the variables.

Web Predictor

A web predictor based on the best predictive performance of the XGB model was developed to predict BM in PCa patients. The risk of BM from PCa could be easily predicted by simply setting the variables in the sidebar of the web page (<https://share.streamlit.io/liuwencaincu/prostate-cancer/main/prostate.py>) (Figure 8).

Discussion

Prostate cancer is the second leading cause of cancer-related death among men in the world. Bone-related events usually occur after bone metastases, which will

result in reduced patient quality of life and survivorship.²³ The current diagnostic method for BM from PCa is bone scan and prostate-specific membrane antigen examination combined with nuclear imaging. However, there is radiation injury and high cost for the bone scan, and not all PCa patients are recommended for BM screening.^{24–26} Tissue biopsy is another method for the diagnosis. But it may increase the risk of further tumor invasion. Although skeletal-related events are considered to be a sign of BM, it would not be reasonable for screening BM from PCa because it may delay the treatment. Thus, it makes sense to develop

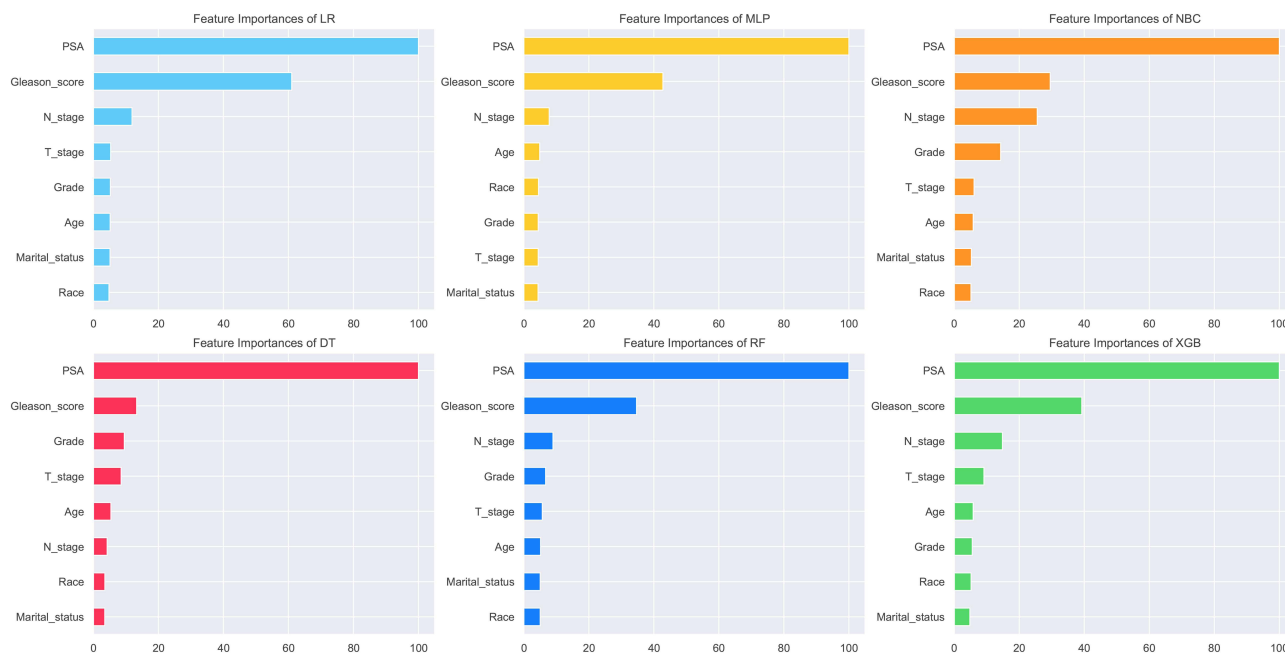


Figure 3 Feature importance of different models. The plot shows the ranking of the relevant importance of features in all models.

a model to provide early attention and screening PCa patients at a high risk of BM. In this study, we built a predictive model using machine learning technologies to predict BM in PCa patients and identify patients at a high risk of BM.

With the development of the computer technique, machine learning technology has been widely used in different fields. And it also shows great promise for application in the biomedical science.^{13,27} Studies have already used machine learning technology to predict the development of diseases.^{12,28} In this study, several widely used machine learning algorithms were developed and validated to predict the risk of BM in PCa patients. After the comparison of algorithms with several evaluation indicators, the XGB algorithm-based prediction model showed the best performance among these models. The XGB model achieved better performance than others probably because it uses a number of strategies to prevent overfitting, exploits the second-order derivatives of the loss function and supports parallelization, and has a fast data processing speed.²⁹ These results can provide clinicians with more accurate prediction outcomes and help them to make personalized decisions for the treatment of PCa.

In the present study, four risk factors, including T, N stage, PSA and Gleason Score, were screened out of the pre-selected eight factors by univariate and multivariate analyses. They were in high agreement with the models' feature importance ranking (Table 4, Figure 3). However, machine learning algorithms consider that other variables that are not statistically significant can also make some contribution to the prediction. This may be because the algorithms can make predictions by exploring intrinsic connections between data that cannot be discovered through traditional statistical methods.

Studies have found that PCa patients with high PSA values have a higher likelihood of developing BM. It is recommended that PCa patients with PSA >20ng/mL need a bone scan to check for BM.^{30,31} In this study, we also found that PSA values were important in predicting BM from PCa. It has been shown that advanced age and clinical stage were both correlated with the BM from PCa and poor prognosis.^{32,33} Chen et al³⁴ found that age greater than 70 years was the threshold for significantly higher risk of BM for PCa patients. However, Stolzenbach et al³⁵ found that age and race were not informative in predicting the progression of PCa metastasis, which was in

Table 4 Comparison Prediction Performances of Different Models for Bone Metastasis

	Models	AUC	Accuracy	Sensitivity (Recall Rate)	Specificity
Internal test set	DT	0.938	0.833	0.883	0.831
	LR	0.903	0.849	0.867	0.848
	MLP	0.947	0.876	0.898	0.875
	NBC	0.941	0.880	0.885	0.879
	RF	0.950	0.879	0.902	0.879
	XGB	0.955	0.881	0.905	0.880
External test set	DT	0.944	0.877	0.846	0.884
	LR	0.905	0.849	0.867	0.848
	MLP	0.950	0.874	0.906	0.867
	NBC	0.934	0.869	0.914	0.859
	RF	0.949	0.874	0.880	0.873
	XGB	0.962	0.884	0.906	0.879

Abbreviations: DT, Decision tree; LR, Logistic regression; MLP, Multilayer Perceptron; NBC, Naive Bayes classification; RF, Random Forest; XGB, eXtreme gradient boosting.

line with the results of our study. Patients diagnosed at the T4 stage had the highest risk of BM. The same trend was seen in Grade. These results were consistent with the findings of Lu et al and Guo et al.^{32,36} Patients with

regional lymphoma metastases tend to be more likely to develop BM than those with N0 stage.³⁷ The Gleason score also played an important role in the prediction model. The degree of BM in patients varies greatly with

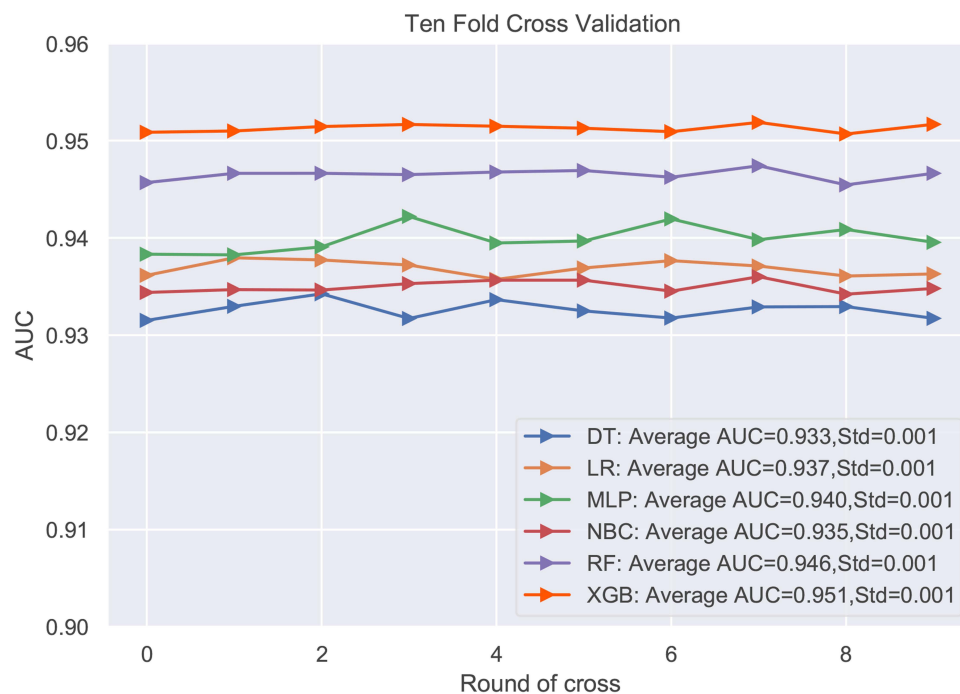


Figure 4 Ten-fold cross-validation results of different machine learning models in the training set.

Abbreviations: DT, Decision tree; LR, Logistic regression; MLP, Multilayer Perceptron; NBC, Naive Bayes classification; RF, Random Forest; XGB, eXtreme gradient boosting.

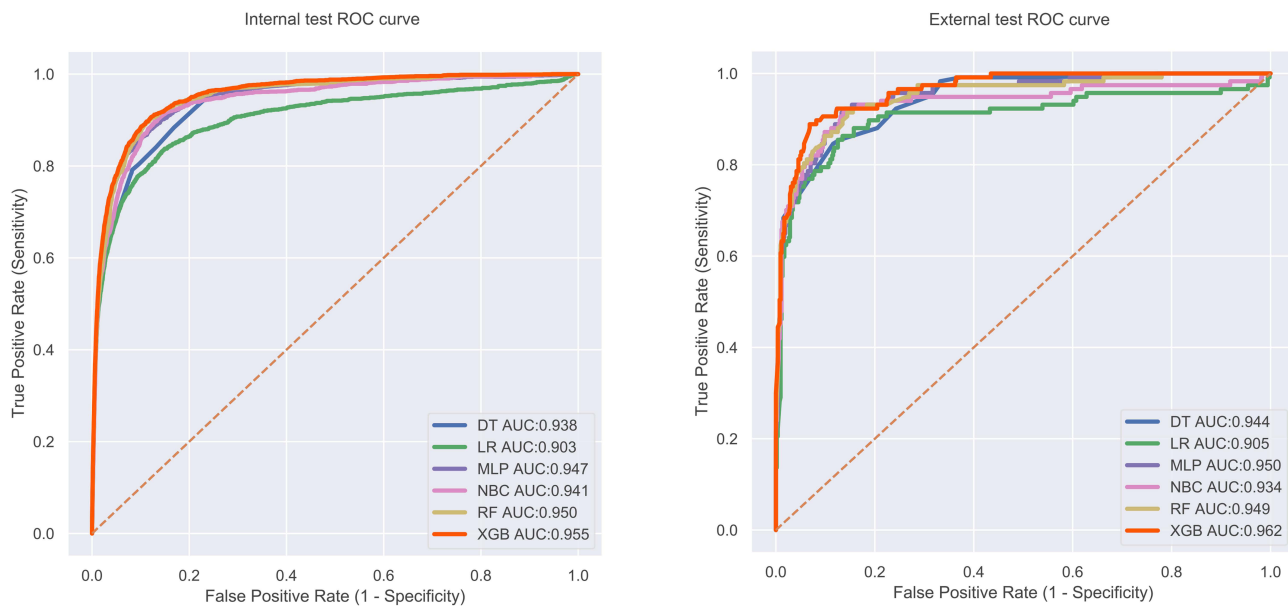


Figure 5 The roc curves of different machine learning models in internal test set and external test set.

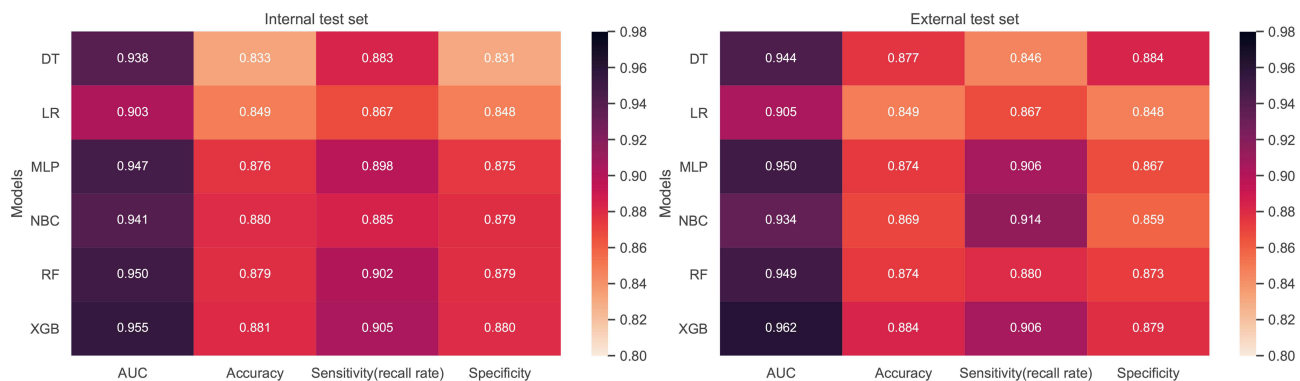


Figure 6 Prediction performances of different models.

the change of Gleason score.³⁸ It was reported that Gleason score greater than 6 had a high specificity (88.9%) for the diagnosis of BM.³⁹ In addition, unmarried male patients had a higher risk of BM from PCa, which was in line with previous studies.^{36,40} Our model adequately incorporates various risk factors that may affect BM in PCa patients and achieves excellent predictive performance.

Based on machine learning algorithms and the huge amount of data in the SEER database, a model was

constructed to predict BM in PCa patients and a web page predictor was developed. However, there are still some limitations in this study. First, the population in this study was obtained from the SEER database and externally validated using data just from a single center, which will be limitations for the application. Second, due to the inherent black-box properties of machine learning algorithms, it may pose some difficulties for the interpretation of the model. Third, the SEER database just reports the initial diagnostic information of PCa patients, and

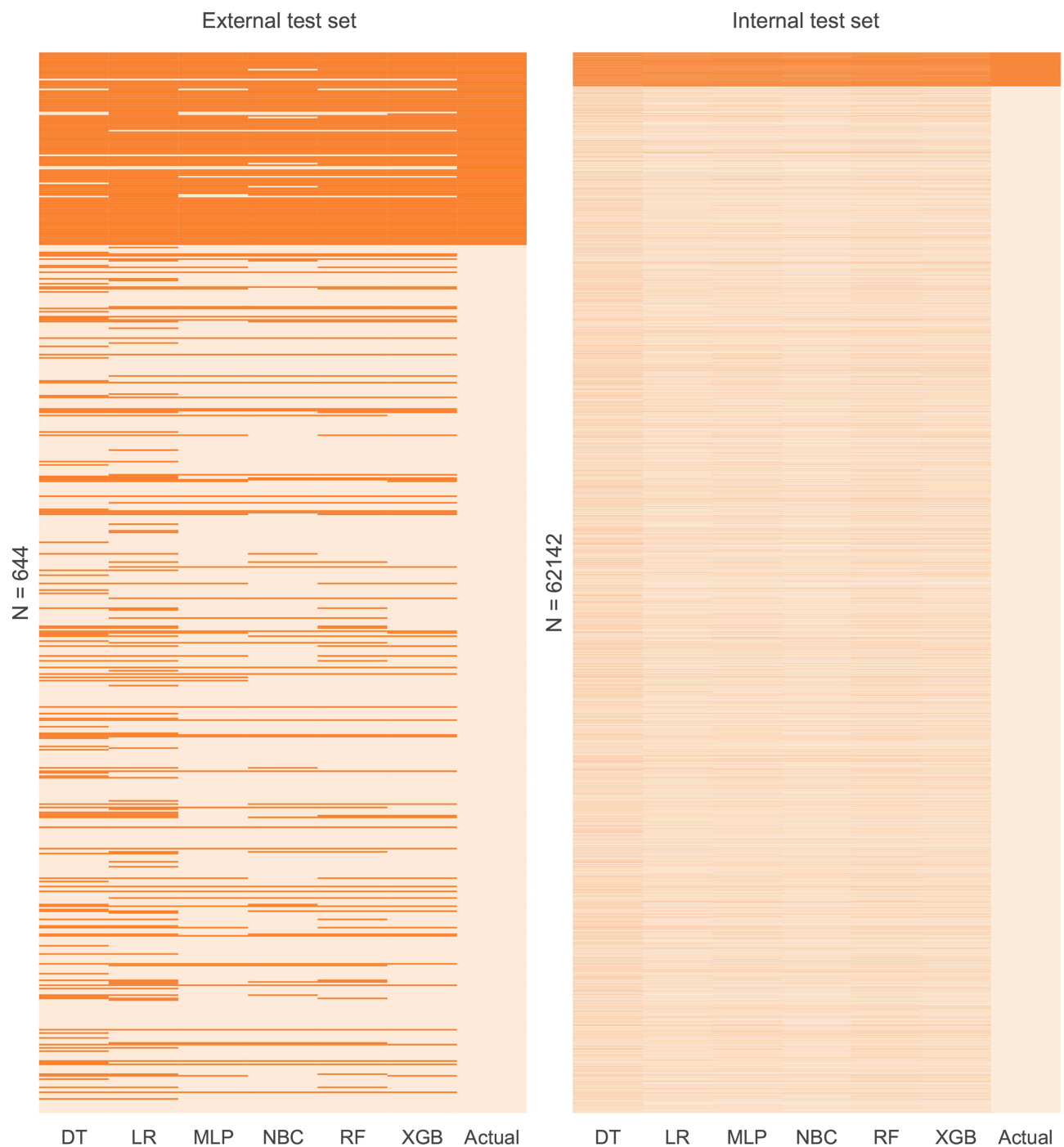


Figure 7 Prediction results of the different models. The heat map shows the predicted results of all models versus the actual situation in internal test set and external test set. Each column in the heat map represents the models' predicted results of bone metastases for all patients in the dataset. Dark colors represent bone metastases cases and light colors are non-bone metastases.

further therapeutic information is missing. We cannot access this information for further analysis.

In conclusion, we developed a prediction model to predict the risk of BM in PCa patients based on the XGB algorithm with machine learning techniques, and

developed a web predictor in this study. And those at high risk of BM were recommended for further detailed screening based on the web predictor. This may help physicians to individualize the treatment of BM in patients with PCa.

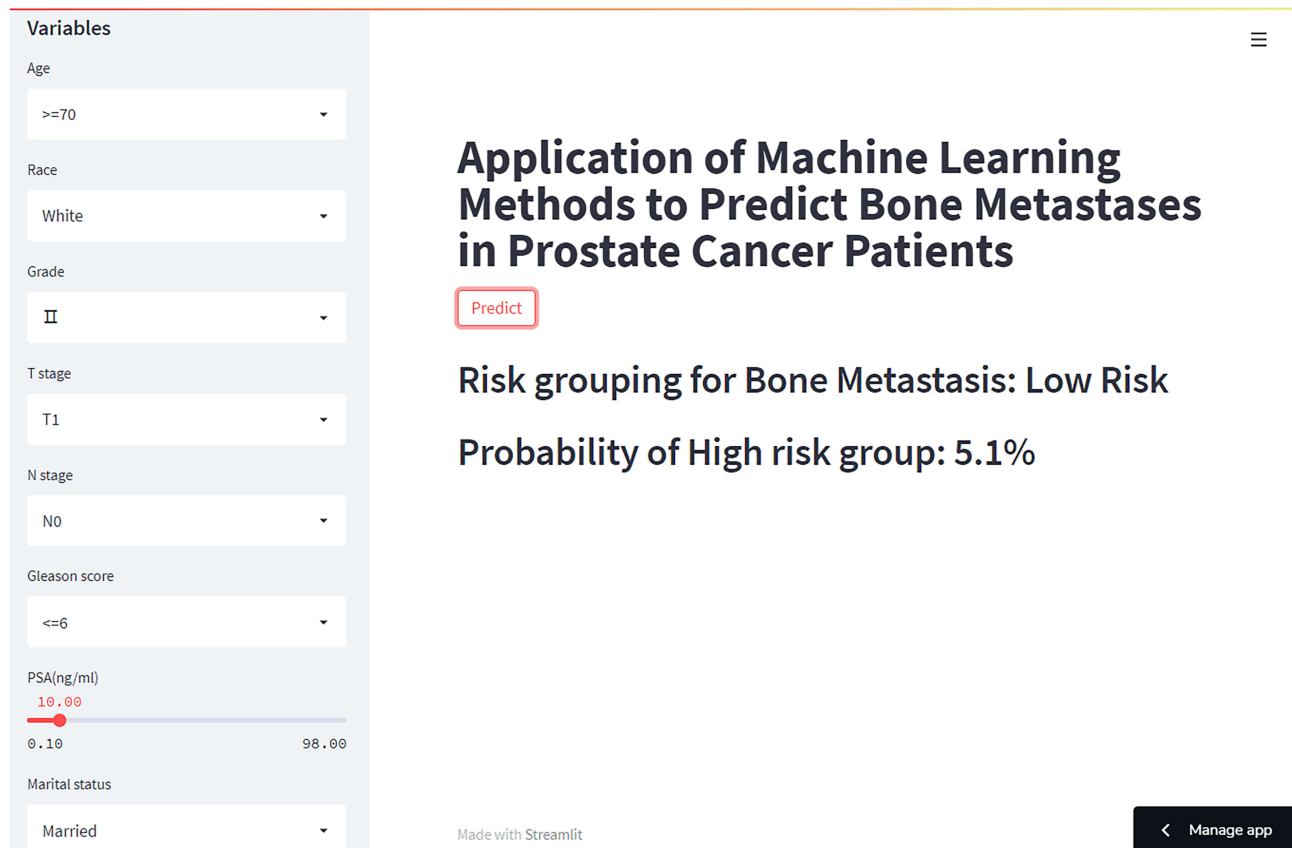


Figure 8 The machine learning model-based web predictor for predicting bone metastases in prostate cancer patients.

Data Sharing Statement

The datasets generated and/or analyzed during the current study are available in the SEER database (<https://seer.cancer.gov/>).

Ethics Approval and Consent to Participate

We received permission to access the research data file in the SEER program from the National Cancer Institute, US. Approval was waived by the local ethics committee, as SEER data is publicly available and de-identified. This study was approved by the Ethics Committee of the First Affiliated Hospital of Nanchang University, and cases from the First Affiliated Hospital of Nanchang University signed written informed consent form. This study followed the guidelines outlined in the Declaration of Helsinki.

Acknowledgments

This work is supported by the Department of Science and Technology Program of Jiangxi Province, China

(No. 20202BBGL73015, 20203BBG73045) and the project of Jiangxi Provincial Health Commission (No. 20161024).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020;70(1):7–30. doi:10.3322/caac.21590
2. Steele CB, Li J, Huang B, Weir HK. Prostate cancer survival in the United States by race and stage (2001-2009): findings from the concord-2 study. *Cancer.* 2017;123:5160–5177. doi:10.1002/encr.31026
3. McDougall JA, Bansal A, Goulart BH, et al. The clinical and economic impacts of skeletal-related events among medicare enrollees with prostate cancer metastatic to bone. *Oncologist.* 2016;21(3):320. doi:10.1634/theoncologist.2015-0327
4. Nørgaard M, Jensen AØ, Jacobsen JB, Cetin K, Fryzek JP, Sørensen HT. Skeletal related events, bone metastasis and survival of prostate cancer: a population based cohort study in Denmark (1999 to 2007). *J Urol.* 2010;184(1):162–167. doi:10.1016/j.juro.2010.03.034
5. Sanjaya I, Mochtar CA, Umbas R. Correlation between low Gleason score and prostate specific antigen levels with incidence of bone metastases in prostate cancer patients: when to omit bone scans? *Asian Pac J Cancer Prev.* 2013;14(9):4973–4976. doi:10.7314/APJCP.2013.14.9.4973

6. Saad F, Lipton A, Cook R, Chen YM, Smith M, Coleman R. Pathologic fractures correlate with reduced survival in patients with malignant bone disease. *Cancer*. 2007;110(8):1860–1867. doi:10.1002/encr.22991
7. Haberkorn U, Eder M, Kopka K, Babich JW, Eisenhut M. New strategies in prostate cancer: prostate-specific membrane antigen (PSMA) ligands for diagnosis and therapy. *Clin Cancer Res*. 2016;22(1):9–15. doi:10.1158/1078-0432.CCR-15-0820
8. Fendler WP, Rahbar K, Herrmann K, Kratochwil C, Eiber M. 177Lu-PSMA radioligand therapy for prostate cancer. *J Nucl Med*. 2017;58(8):1196–1200. doi:10.2967/jnumed.117.191023
9. Barrio M, Fendler WP, Czernin J, Herrmann K. Prostate specific membrane antigen (PSMA) ligands for diagnosis and therapy of prostate cancer. *Expert Rev Mol Diagn*. 2016;16(11):1177–1188. doi:10.1080/14737159.2016.1243057
10. Farooki A, Leung V, Tala H, Tuttle RM. Skeletal-related events due to bone metastases from differentiated thyroid cancer. *J Clin Endocrinol Metab*. 2012;97(7):2433–2439. doi:10.1210/jc.2012-1169
11. Doll KM, Rademaker A, Sosa JA. Practical guide to surgical data sets: surveillance, epidemiology, and end results (SEER) database. *JAMA Surg*. 2018;153(6):588–589. doi:10.1001/jamasurg.2018.0501
12. Liu WC, Li ZQ, Luo ZW, Liao WJ, Liu ZL, Liu JM. Machine learning for the prediction of bone metastasis in patients with newly diagnosed thyroid cancer. *Cancer Med*. 2021;10(8):2802–2811. doi:10.1002/cam4.3776
13. Goecks J, Jalili V, Heiser LM, Gray JW. How machine learning will transform biomedicine. *Cell*. 2020;181(1):92–101. doi:10.1016/j.cell.2020.03.022
14. Darcy AM, Louie AK, Roberts LW. Machine learning and the profession of medicine. *JAMA*. 2016;315(6):551–552. doi:10.1001/jama.2015.18421
15. Oliphant TE. Python for scientific computing. *Comput Sci Eng*. 2007;9(3):10–20. doi:10.1109/MCSE.2007.58
16. Altmann A, Tološi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. *Bioinformatics*. 2010;26(10):1340–1347. doi:10.1093/bioinformatics/btq134
17. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H. Xgboost: extreme gradient boosting. *R Package Version 04-2*. 2015;1(4):1–4.
18. Qi Y. Random forest for bioinformatics. In: *Ensemble Machine Learning*. Springer; 2012:307–323.
19. Tang J, Deng C, Huang G-B. Extreme learning machine for multi-layer perceptron. *IEEE Trans Neural Netw Learn Syst*. 2015;27(4):809–821. doi:10.1109/TNNLS.2015.2424995
20. Sperandei S. Understanding logistic regression analysis. *Biochem Med*. 2014;24(1):12–18. doi:10.11613/BM.2014.003
21. Myles AJ, Feudale RN, Liu Y, Woody NA, Brown SD. An introduction to decision tree modeling. *J Chemom*. 2004;18(6):275–285. doi:10.1002/cem.873
22. Rish I, editor. An empirical study of the naive Bayes classifier. In: *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*; 2001.
23. Fornetti J, Welm AL, Stewart SA. Understanding the bone in cancer metastasis. *Bone Miner Res*. 2018;33(12):2099–2113. doi:10.1002/jbmr.3618
24. Kuten J, Fahoum I, Savin Z, et al. Head-to-head comparison of 68Ga-PSMA-11 with 18F-PSMA-1007 PET/CT in staging prostate cancer using histopathology and immunohistochemical analysis as a reference standard. *J Nucl Med*. 2020;61(4):527–532. doi:10.2967/jnumed.119.234187
25. Afshar-Oromieh A, Babich JW, Kratochwil C, et al. The rise of PSMA ligands for diagnosis and therapy of prostate cancer. *J Nucl Med*. 2016;57(Supplement3):79S–89S. doi:10.2967/jnumed.115.170720
26. Lenzo NP, Meyrick D, Turner JH. Review of gallium-68 PSMA PET/CT imaging in the management of prostate cancer. *Diagnostics*. 2018;8(1):16. doi:10.3390/diagnostics8010016
27. Camacho DM, Collins KM, Powers RK, Costello JC, Collins JJ. Next-generation machine learning for biological networks. *Cell*. 2018;173(7):1581–1592. doi:10.1016/j.cell.2018.05.015
28. Zhu J, Zheng J, Li L, et al. Application of machine learning algorithms to Predict central lymph node metastasis in T1-T2, non-invasive, and clinically node negative papillary thyroid carcinoma. *Front Med*. 2021;8. doi:10.3389/fmed.2021.635771
29. Ogunleye A, Wang Q-G. XGBoost model for chronic kidney disease diagnosis. *IEEE/ACM Trans Comput Biol Bioinform*. 2019;17(6):2131–2140. doi:10.1109/TCBB.2019.2911071
30. Greene KL, Albertsen PC, Babaian RJ, et al. Prostate specific antigen best practice statement: 2009 update. *J Urol*. 2013;189(1):S2–S11. doi:10.1016/j.juro.2012.11.014
31. Network NCC. NCCN clinical practice guidelines in oncology. Prostate cancer V. 4; 2011. Available from: http://www.nccn.org/professionals/physician_gls/pdf/prostate.pdf. Accessed November 17, 2021.
32. Guo X, Zhang C, Guo Q, et al. The homogeneous and heterogeneous risk factors for the morbidity and prognosis of bone metastasis in patients with prostate cancer. *Cancer Manag Res*. 2018;10:1639. doi:10.2147/CMAR.S168579
33. Briganti A, Suardi N, Gallina A, et al. Predicting the risk of bone metastasis in prostate cancer. *Cancer Treat Rev*. 2014;40(1):3–11. doi:10.1016/j.ctrv.2013.07.001
34. Chen S, Wang L, Qian K, et al. Establishing a prediction model for prostate cancer bone metastasis. *Int J Biol Sci*. 2019;15(1):208. doi:10.7150/ijbs.27537
35. Stolzenbach LF, Rosiello G, Deuker M, et al. The impact of race and age on distribution of metastases in patients with prostate cancer. *J Urol*. 2020;204(5):962–968. doi:10.1097/JU.0000000000001131
36. Lu YJ, Duan WM. Establishment and validation of a novel predictive model to quantify the risk of bone metastasis in patients with prostate cancer. *Transl Androl Urol*. 2021;10(1):310–325. doi:10.21037/tau-20-1133
37. Wilczak W, Wittmer C, Clauditz T, et al. Marked prognostic impact of minimal lymphatic tumor spread in prostate cancer. *Eur Urol*. 2018;74(3):376–386. doi:10.1016/j.eururo.2018.05.034
38. Battisti V, Maders LD, Bagatini MD, et al. Ectonucleotide pyrophosphatase/phosphodiesterase (E-NPP) and adenosine deaminase (ADA) activities in prostate cancer patients: influence of Gleason score, treatment and bone metastasis. *Biomed Pharmacother*. 2013;67(3):203–208. doi:10.1016/j.biopha.2012.12.004
39. Zaman MU, Fatima N, Sajjad Z. Metastasis on bone scan with low prostate specific antigen (≤ 20 ng/mL) and Gleason's score (< 8) in newly diagnosed Pakistani males with prostate cancer: should we follow Western guidelines. *Asian Pac J Cancer Prev*. 2011;12(6):1529–1532.
40. Vaarala MH, Hirvikoski P, Kauppila S, Paavonen TK. Identification of androgen-regulated genes in human prostate. *Mol Med Rep*. 2012;6(3):466–472. doi:10.3892/mmr.2012.956

Cancer Management and Research

Dovepress

Publish your work in this journal

Cancer Management and Research is an international, peer-reviewed open access journal focusing on cancer research and the optimal use of preventative and integrated treatment interventions to achieve improved outcomes, enhanced survival and quality of life for the cancer patient.

The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/cancer-management-and-research-journal>