

# The Construction and Validation of a new Predictive Model for Overall Survival of Clear Cell Renal Cell Carcinoma Patients with Bone Metastasis Based on Machine Learning Algorithm

Technology in Cancer Research & Treatment  
Volume 22: 1-8  
© The Author(s) 2023  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/15330338231165131  
journals.sagepub.com/home/tct  


Yijun Le, MD<sup>1</sup> , Wen Xu, MD<sup>2</sup>, and Wei Guo, PhD<sup>1</sup>

## Abstract

**Background:** This study aimed to develop and validate predictive models based on machine learning (ML) algorithms for patients with bone metastases (BM) from clear cell renal cell carcinoma (ccRCC) and to identify appropriate models for clinical decision-making. **Methods:** In this retrospective study, we obtained information on ccRCC patients diagnosed with bone metastasis (ccRCC-BM), from the Surveillance, Epidemiology, and End Results (SEER) database from 2010 to 2015 ( $n = 1490$ ), and collected clinicopathological information on ccRCC-BM patients at our hospital ( $n = 42$ ). We then applied four ML algorithms: extreme gradient boosting (XGB), logistic regression (LR), random forest (RF), and Naive Bayes model (NB), to develop models for predicting the overall survival (OS) of patients with bone metastasis from ccRCC. In the SEER dataset, 70% of the patients were randomly divided into training cohorts and the remaining 30% were used as validation cohorts. Data from our center were used as an external validation cohort. Finally, we evaluated the model performance using receiver operating characteristic curves (ROC), area under the ROC curve (AUC), accuracy, specificity, and FI-scores. **Results:** The mean survival times of patients in the SEER and Chinese cohort were 21.8 months and 37.0 months, respectively. Age, marital status, grade, T stage, N stage, tumor size, brain metastasis, liver metastasis, lung metastasis, and surgery were included in the ML model. We observed that all four ML algorithms performed well in predicting the 1-year and 3-year OS of patients with ccRCC-BM. **Conclusion:** ML is useful in predicting the survival of patients with ccRCC-BM, and ML models can play a positive role in clinical applications.

## Keywords

clear cell renal cell carcinoma, bone metastasis, SEER, overall survival, machine learning

## Abbreviations

AUC, area under the ROC curve; BM, bone metastases; ccRCC, clear cell renal cell carcinoma; LR, logistic regression; ML, machine learning; NB, Naive Bayes model; OS, overall survival; RF, random forest; ROC, receiver operating characteristic curve; SEER, Surveillance, Epidemiology, and End Results; SREs, skeletal-related events; TNM, tumor, nodes, and metastases; TRIPOD, transparent reporting of a multivariable prediction model for individual prognosis or diagnosis; XGB, extreme gradient boosting

## Introduction

Renal cell carcinoma is one of the most common malignant tumors; it is the sixth most common cancer among men and the tenth most common cancer among women worldwide.<sup>1</sup> Clear cell renal cell carcinoma (ccRCC) is the most common pathological type of kidney cancer, accounting for 80%–90% of all renal cell carcinomas<sup>2</sup> along with papillary renal cell carcinoma and chromophobe renal cell carcinoma. ccRCC tends to have worse prognosis than other pathological types of RCC.<sup>3,4</sup> At the time of the initial diagnosis, approximately 20%–30% of patients have a combination of distant metastases, of which bone metastasis (BM) is one of the most common sites.<sup>5,6</sup>

The occurrence of bone metastases from renal cell carcinoma has become increasingly common as patients with renal cell carcinoma live longer due to the rapid development of targeted

<sup>1</sup> Musculoskeletal Tumor Center, Peking University People's Hospital, Beijing, China

<sup>2</sup> Department of Dermatology, People's Hospital of Peking University, Beijing, People's Republic of China

## Corresponding Author:

Wei Guo, Musculoskeletal Tumor Center, Peking University People's Hospital, No. 11 Xizhimen South Street, Xicheng District, Beijing 100044, China.  
Email: bonetumor@163.com



therapies and immunotherapy.<sup>7</sup> The International Kidney Cancer Working Group states that overall survival (OS) is significantly worse for bone and liver metastases compared to other metastatic sites.<sup>8</sup> Although major advances have been made in the treatment of primary renal cell carcinoma, the overall survival of patients with RCC who develop bone metastases is only 19.7 months.<sup>9</sup> Bone metastases from kidney cancer mainly occur in the mid-axis bones, and common metastatic sites include the pelvis, spine, and limbs. Altogether, 71% of patients with bone metastases from kidney cancer had multiple bone metastases. Osteolytic lesions (approximately 80%) were predominant in kidney cancer bone metastases, with a small number of osteogenic lesions (7%), while the rest were mixed lesions (13%).<sup>10</sup> BM is usually associated with skeletal-related events (SREs) such as pain, hypercalcemia, pathological fractures, spinal cord compression, and even paraplegia, which greatly reduce the patient's quality of life and result in a worse prognosis.<sup>11</sup>

Machine learning (ML) is the science of artificial intelligence (AI). It has been used in many different scientific and medical applications in recent years because of its powerful data mining and analysis capabilities.<sup>12–14</sup> Marostica et al.<sup>15</sup> used deep convolutional neural networks to detect and diagnose kidney cancer with good results. Zhu et al.<sup>17</sup> used convolutional neural networks for blood cell classification<sup>16</sup> and malaria image classification, and their models were superior to existing methods. ML has been proven to have more accurate predictive power than traditional statistical methods.<sup>18</sup> Several studies have found that the prognosis of patients with bone metastases from renal cell carcinoma is significantly correlated with age, tumor, nodes, and metastases (TNM) stage, other organ metastases, and targeted therapy.<sup>19,20</sup> Models using traditional statistical methods to predict the prognosis of patients with clear cell renal cell carcinoma bone metastases (ccRCC-BM) have been reported,<sup>21,22</sup> however, there have been no studies on ML algorithms.

This study aimed to investigate the application of ML to predict the prognosis of patients with bone metastases from renal cell carcinoma. This will help identify high-risk patients and guide clinical interventions.

## Materials and Methods

### Data Source and Selection of Variables

The original data in this study were extracted from the Surveillance, Epidemiology, and End Results (SEER) database (SEER Research Plus Data 18 Registries, Nov 2020 Sub (2000–2018)) and our hospital. The SEER database agreement was signed and permission was obtained to access the information (accession username: 12906-Nov 2021). This retrospective study did not include personal information; therefore, institutional review board approval or informed consent were not required. All the methods used in our study complied with the rules of the SEER database. This study conformed to transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidelines.<sup>23</sup>

All patients with ccRCC diagnosed between 2010 and 2015 were considered for this study, and the exclusion criteria were as follows: (1) No bone metastasis; (2) Tumor size unknown; (3) Unknown TNM stage unknown; (4) Survival month unknown or < 1 month; (5) Presence of brain/liver/lung metastasis unknown; (6) Age at diagnosis < 18 years. Altogether, 1490 patients from the SEER database were included. These patients were randomly divided into validation and training cohorts. The external validation cohort consisted of patients with a confirmed diagnosis of ccRCC-BM at our hospital between 2010 and 2016 ( $n = 42$ ). These patients were contacted via telephone to obtain verbal informed consent.

To take full advantage of the ML algorithm, several demographic and clinical variables were selected for data analysis. Demographic variables included sex, age, marital status, and insurance purchase, whereas clinical variables included tumor size, laterality, TNM stage, pathological grade, brain metastasis, liver metastasis, lung metastasis, surgery of primary site, radiotherapy, and chemotherapy. The endpoints of the predictive models were the 1- and 3-year overall survival rates.

### Statistics Analysis

All statistical analyses were performed using R software version 4.1.3 (<https://www.r-project.org/>) and Python (version 3.10; Python Software Foundation). Pearson correlation tests were performed between all the variables, and the results are presented in the form of a heat map. To compare baseline characteristics, Pearson's chi-square test was used for categorical variables and Mann–Whitney U test was used for continuous variables. The results were considered statistically significant when the P value was less than 0.05 (two-sided).

### Establishment and Validation of Models

Patients from the SEER database were randomly divided into training and validation cohorts in a 7:3 ratio. The chi-square test was used to compare the association between the two groups. Cox proportional hazards model was used for univariate and multivariate analyses to identify independent risk factors for ccRCC-BM. These independent predictors were used to construct machine-learning models to predict the 1- and 3-year overall survival rates. We used four machine-learning algorithms to build the models: logistic regression (LR), random forest (RF), extreme gradient boosting (XGB), and Naive Bayes (NB). Receiver operating characteristic (ROC) curve, area under the curve (AUC), accuracy, specificity, and F1-scores were then used to compare the predictive power of the models.

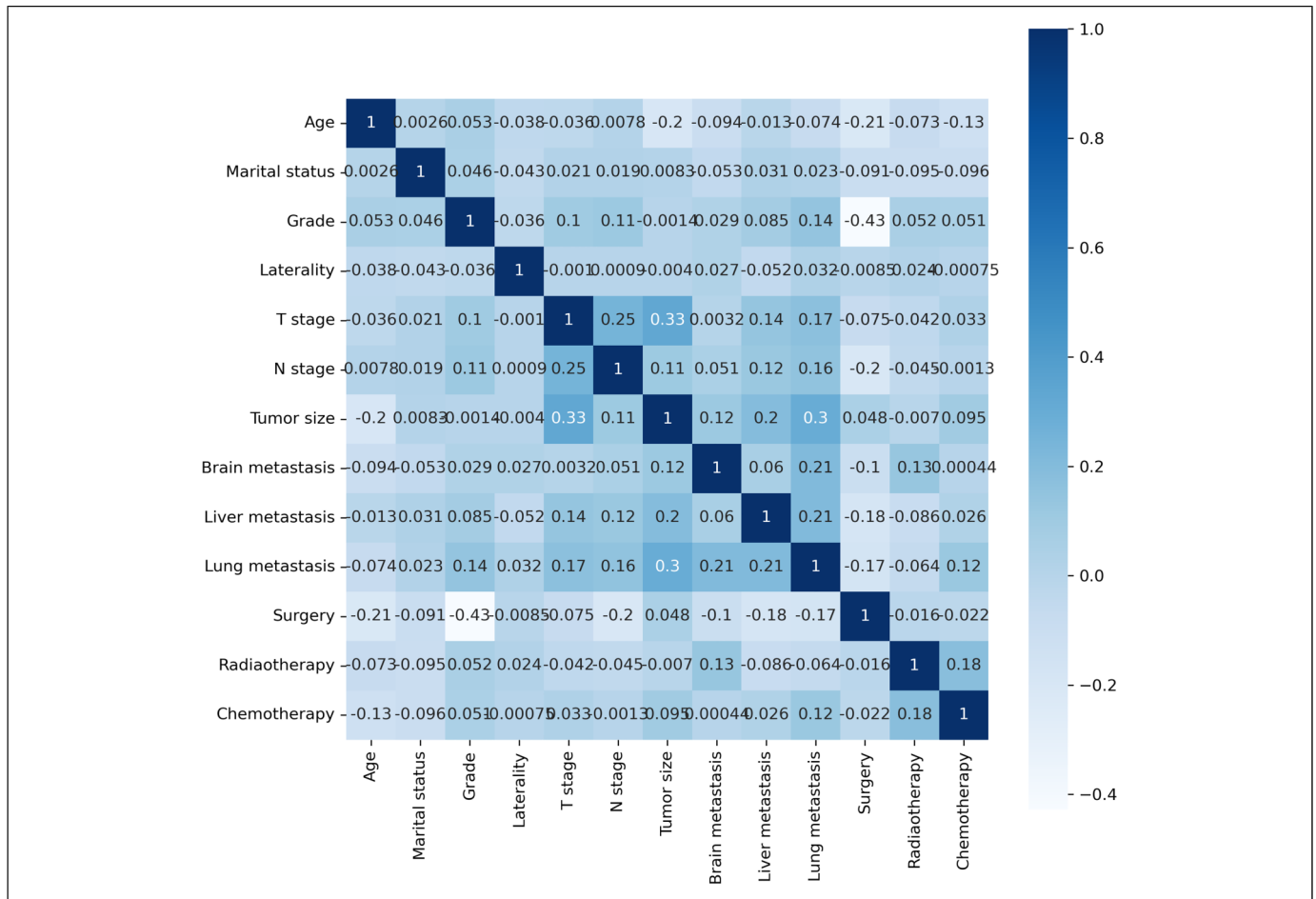
## Results

### Demographic and Clinicopathological Characteristics

After screening, 1532 patients diagnosed with ccRCC-BM were included in our study, including 1490 patients from the SEER

**Table 1.** Demographic Characteristics and Clinicopathological Features of ccRCC-BM Patients.

|                  | Training Group<br>(N = 1043) | Validation Group<br>(N = 447) | External Group<br>(N = 42) | Overall<br>(N = 1532) | P-value |
|------------------|------------------------------|-------------------------------|----------------------------|-----------------------|---------|
| Age              |                              |                               |                            |                       | 0.658   |
| Mean (SD)        | 63.4 (11.1)                  | 63.2 (10.7)                   | 61.7 (11.2)                | 63.3 (11.0)           |         |
| Sex              |                              |                               |                            |                       | 0.788   |
| Female           | 310 (29.7%)                  | 122 (27.3%)                   | 11 (26.2%)                 | 443 (28.9%)           |         |
| Male             | 733 (70.3%)                  | 325 (72.7%)                   | 31 (73.8%)                 | 1089 (71.1%)          |         |
| Race             |                              |                               |                            |                       | <0.001  |
| Other            | 75 (7.2%)                    | 21 (4.7%)                     | 42 (100%)                  | 138 (9.0%)            |         |
| Black            | 77 (7.4%)                    | 37 (8.3%)                     | 0 (0%)                     | 114 (7.4%)            |         |
| White            | 891 (85.4%)                  | 389 (87.0%)                   | 0 (0%)                     | 1280 (83.6%)          |         |
| Marital status   |                              |                               |                            |                       | 0.042   |
| Married          | 642 (61.6%)                  | 277 (62.0%)                   | 35 (83.3%)                 | 954 (62.3%)           |         |
| Unmarried        | 401 (38.4%)                  | 170 (38.0%)                   | 7 (16.7%)                  | 578 (37.7%)           |         |
| Insurance        |                              |                               |                            |                       | <0.001  |
| No               | 770 (73.8%)                  | 314 (70.2%)                   | 7 (16.7%)                  | 1091 (71.2%)          |         |
| Yes              | 273 (26.2%)                  | 133 (29.8%)                   | 35 (83.3%)                 | 441 (28.8%)           |         |
| Grade            |                              |                               |                            |                       | 0.001   |
| I                | 26 (2.5%)                    | 7 (1.6%)                      | 1 (2.4%)                   | 34 (2.2%)             |         |
| II               | 152 (14.6%)                  | 75 (16.8%)                    | 19 (45.2%)                 | 246 (16.1%)           |         |
| III              | 249 (23.9%)                  | 100 (22.4%)                   | 9 (21.4%)                  | 358 (23.4%)           |         |
| IV               | 134 (12.8%)                  | 51 (11.4%)                    | 2 (4.8%)                   | 187 (12.2%)           |         |
| Unknown          | 482 (46.2%)                  | 214 (47.9%)                   | 11 (26.2%)                 | 707 (46.1%)           |         |
| Laterality       |                              |                               |                            |                       | 0.970   |
| Left             | 524 (50.2%)                  | 228 (51.0%)                   | 19 (45.2%)                 | 771 (50.3%)           |         |
| Right            | 511 (49.0%)                  | 217 (48.5%)                   | 23 (54.8%)                 | 751 (49.0%)           |         |
| Bilateral        | 8 (0.8%)                     | 2 (0.4%)                      | 0 (0%)                     | 10 (0.7%)             |         |
| Tumor size       |                              |                               |                            |                       | 0.878   |
| Mean (SD)        | 7.91 (3.67)                  | 8.11 (3.73)                   | 5.52 (2.56)                | 7.91 (3.68)           |         |
| T stage          |                              |                               |                            |                       | 0.700   |
| T1               | 285 (27.3%)                  | 113 (25.3%)                   | 19 (45.2%)                 | 417 (27.2%)           |         |
| T2               | 194 (18.6%)                  | 86 (19.2%)                    | 4 (9.5%)                   | 284 (18.5%)           |         |
| T3               | 370 (35.5%)                  | 166 (37.1%)                   | 14 (33.3%)                 | 550 (35.9%)           |         |
| T4               | 95 (9.1%)                    | 41 (9.2%)                     | 2 (4.8%)                   | 138 (9.0%)            |         |
| TX               | 99 (9.5%)                    | 41 (9.2%)                     | 3 (7.1%)                   | 143 (9.3%)            |         |
| N stage          |                              |                               |                            |                       | 0.273   |
| N0               | 691 (66.3%)                  | 304 (68.0%)                   | 24 (57.1%)                 | 1019 (66.5%)          |         |
| N1               | 255 (24.4%)                  | 103 (23.0%)                   | 9 (21.4%)                  | 367 (24.0%)           |         |
| NX               | 97 (9.3%)                    | 40 (8.9%)                     | 9 (21.4%)                  | 146 (9.5%)            |         |
| Brain metastasis |                              |                               |                            |                       | 0.341   |
| No               | 932 (89.4%)                  | 405 (90.6%)                   | 41 (97.6%)                 | 1378 (89.9%)          |         |
| Yes              | 111 (10.6%)                  | 42 (9.4%)                     | 1 (2.4%)                   | 154 (10.1%)           |         |
| Liver metastasis |                              |                               |                            |                       | 0.461   |
| No               | 914 (87.6%)                  | 382 (85.5%)                   | 39 (92.9%)                 | 1335 (87.1%)          |         |
| Yes              | 129 (12.4%)                  | 65 (14.5%)                    | 3 (7.1%)                   | 197 (12.9%)           |         |
| Lung metastasis  |                              |                               |                            |                       | 0.178   |
| No               | 570 (54.7%)                  | 240 (53.7%)                   | 30 (71.4%)                 | 840 (54.8%)           |         |
| Yes              | 473 (45.4%)                  | 207 (46.3%)                   | 12 (28.6%)                 | 692 (45.2%)           |         |
| Surgery          |                              |                               |                            |                       | <0.001  |
| No               | 557 (53.4%)                  | 229 (51.2%)                   | 9 (21.4%)                  | 795 (51.9%)           |         |
| Yes              | 486 (46.6%)                  | 218 (48.8%)                   | 33 (78.6%)                 | 737 (48.1%)           |         |
| Radiotherapy     |                              |                               |                            |                       | <0.001  |
| No               | 416 (39.9%)                  | 175 (39.1%)                   | 36 (85.7%)                 | 627 (40.9%)           |         |
| Yes              | 627 (60.1%)                  | 272 (60.9%)                   | 6 (14.3%)                  | 905 (59.1%)           |         |
| Chemotherapy     |                              |                               |                            |                       | <0.001  |
| No               | 413 (39.6%)                  | 182 (40.7%)                   | 38 (90.5%)                 | 633 (41.3%)           |         |
| Yes              | 630 (60.4%)                  | 265 (59.3%)                   | 4 (9.5%)                   | 899 (58.7%)           |         |
| Survival months  |                              |                               |                            |                       | <0.001  |
| Mean (SD)        | 21.3 (22.6)                  | 22.9 (23.2)                   | 37.0 (39.5)                | 22.2 (23.5)           |         |



**Figure 1.** The heat map shows that all variables are independent of each other.

database and 42 patients from our center. We divided the 1490 patients from the SEER database into training ( $n = 1043$ ) and validation ( $n = 477$ ) cohorts in a 7:3 ratio, with the remaining 42 patients in the external validation group. The demographic and clinical characteristics of the patients are shown in Table 1. There were significant differences between the three groups in terms of race, insurance, pathological grade, surgery, radiotherapy, chemotherapy, and survival time ( $P < 0.05$ ). In the external validation cohort, the race was classified as “other,” and a greater proportion of patients had insurance, a greater propensity for pathological grade II, more patients who underwent surgery for the primary site, and a smaller proportion of patients who received radiotherapy or chemotherapy. This may have led to a longer survival time in these patients. The remaining variables did not differ significantly among the three groups. The mean age was 63.3 years, patients were predominantly elderly, 1089 (71.1%) were male, 954 (62.3%) were married, and the mean tumor size was 7.91 cm. The number of patients with T1, T2, T3, T4, and TX stages were 417 (27.2%), 284 (18.5%), 550 (35.9%), 138 (9.0%), and 143 (9.3%), respectively. Regarding N stage, the number of patients with N0, N1, and NX were 1019 (66.5%), 367 (24.0%), and 146 (9.5%), respectively. Most patients had no brain metastases ( $n = 1378$ ,

89.9%) or liver metastases ( $n = 1335$ , 87.1%), while 840 (54.8%) patients had no lung metastases.

### Independent Risk Factors of BM in ccRCC Patients

The Pearson correlation analysis heat map (Figure 1) showed that all variables existed independently of each other, did not affect each other, and could be included in the subsequent statistical analysis. In univariate analysis, the following variables were significantly associated with the prognosis of ccRCC-BM patients ( $P < 0.05$ ) (Table 2): age, marital status, grade, laterality, T stage, N stage, tumor size, brain metastasis, lung metastasis, liver metastasis, and surgery. Multivariate analysis based on these significant variables demonstrated that age, marital status, grade, T stage, N stage, tumor size, brain metastasis, lung metastasis, liver metastasis, and surgery were independent predictors of overall survival (Table 3).

### Model Construction and Performance of ML Algorithms

Based on the training set, we applied four ML algorithms to predict the 1-year and 3-year OS rates. The validation set was applied to examine and compare the predictive performance

**Table 2.** Univariate Analysis for the Prognosis Factors of ccRCC-BM Patients.

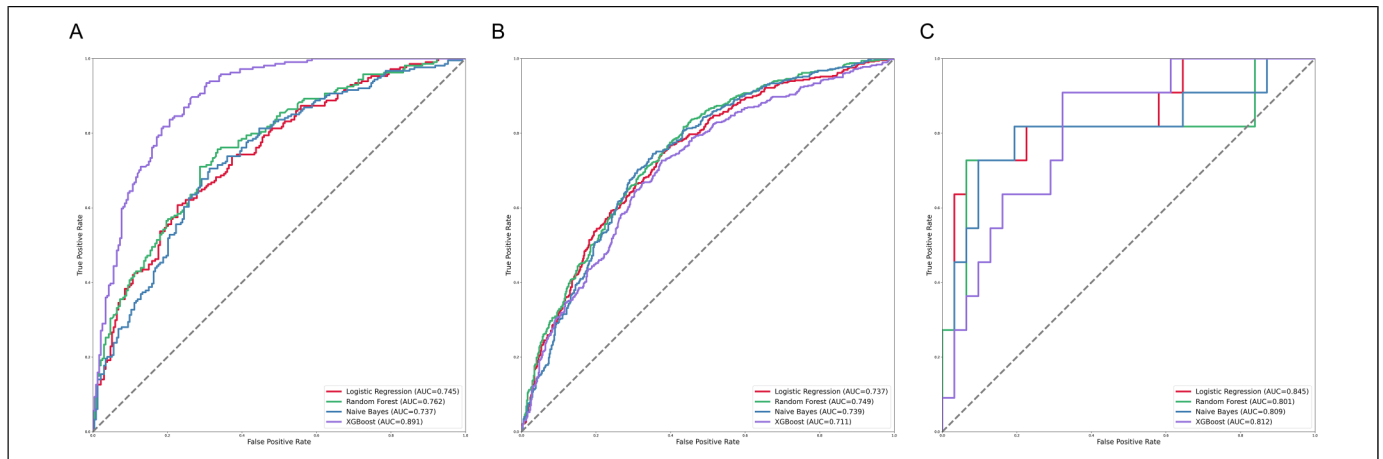
| Characteristics  | HR        | 95%CI     | P-value |
|------------------|-----------|-----------|---------|
| Age              | 1.01      | 1.01–1.02 | <0.001  |
| Race             |           |           |         |
| White            | Reference |           |         |
| Black            | 1.11      | 0.9–1.36  | 0.340   |
| Other            | 1.05      | 0.84–1.32 | 0.649   |
| Sex              |           |           |         |
| Female           | Reference |           |         |
| Male             | 0.92      | 0.82–1.04 | 0.206   |
| Marital Status   |           |           |         |
| Married          | Reference |           |         |
| Unmarried        | 1.22      | 1.09–1.37 | <0.001  |
| Insurance        |           |           |         |
| No               | Reference |           |         |
| Yes              | 1.05      | 0.93–1.19 | 0.450   |
| Grade            |           |           |         |
| I                | Reference |           |         |
| II               | 0.75      | 0.49–1.16 | 0.195   |
| III              | 0.94      | 0.62–1.42 | 0.761   |
| IV               | 1.25      | 0.81–1.92 | 0.308   |
| Unknown          | 1.52      | 1.01–2.28 | 0.045   |
| Laterality       |           |           |         |
| Bilateral        | Reference |           |         |
| Left             | 0.49      | 0.25–0.94 | 0.031   |
| Right            | 0.53      | 0.27–1.03 | 0.059   |
| T-stage          |           |           |         |
| T1               | Reference |           |         |
| T2               | 1.1       | 0.93–1.3  | 0.252   |
| T3               | 1.04      | 0.9–1.19  | 0.635   |
| T4               | 1.71      | 1.39–2.1  | <0.001  |
| TX               | 2.06      | 1.69–2.52 | <0.001  |
| N-stage          |           |           |         |
| N0               | Reference |           |         |
| N1               | 1.69      | 1.48–1.92 | <0.001  |
| NX               | 1.8       | 1.5–2.17  | <0.001  |
| Tumor size       | 1.04      | 1.02–1.05 | <0.001  |
| Brain metastasis |           |           |         |
| No               | Reference |           |         |
| Yes              | 1.86      | 1.56–2.22 | <0.001  |
| Liver metastasis |           |           |         |
| No               | Reference |           |         |
| Yes              | 1.95      | 1.66–2.28 | <0.001  |
| Lung metastasis  |           |           |         |
| No               | Reference |           |         |
| Yes              | 1.97      | 1.76–2.2  | <0.001  |
| Surgery          |           |           |         |
| No               | Reference |           |         |
| Yes              | 0.39      | 0.35–0.44 | <0.001  |
| Radiotherapy     |           |           |         |
| No               | Reference |           |         |
| Yes              | 1.01      | 0.91–1.14 | 0.814   |
| Chemotherapy     |           |           |         |
| No               | Reference |           |         |
| Yes              | 0.9       | 0.8–1     | 0.060   |

**Table 3.** Multivariate Analysis for the Prognosis Factors of ccRCC-BM Patients.

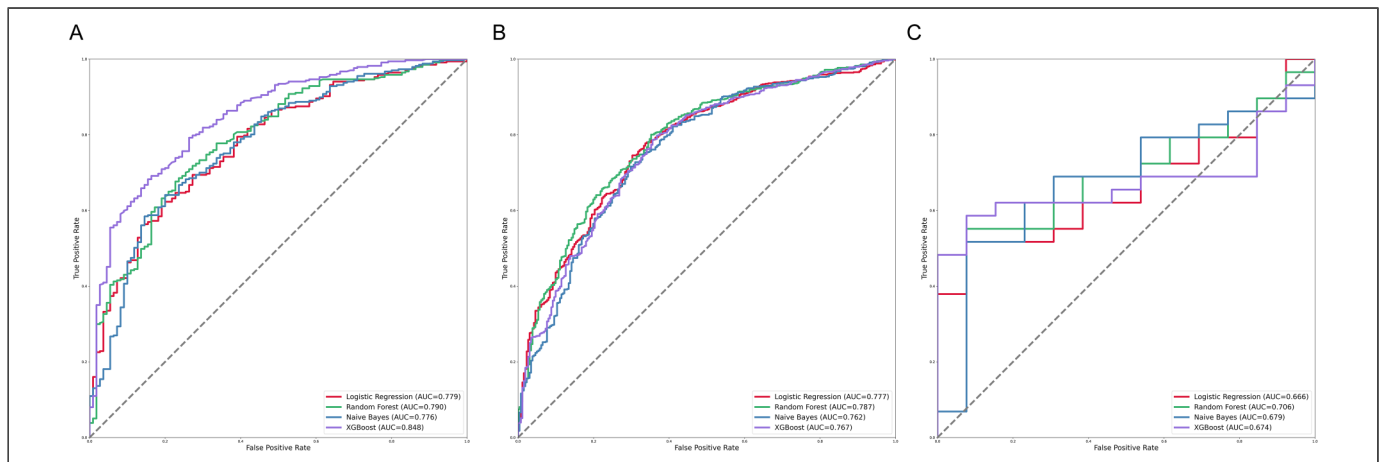
| Characteristics  | HR        | 95%CI     | P-value |
|------------------|-----------|-----------|---------|
| Age              | 1.01      | 1–1.02    | <0.001  |
| Marital Status   |           |           |         |
| Married          | Reference |           |         |
| Unmarried        | 1.22      | 1.09–1.36 | <0.001  |
| Grade            |           |           |         |
| I                | Reference |           |         |
| II               | 1.24      | 0.8–1.9   | 0.335   |
| III              | 1.76      | 1.14–2.71 | 0.010   |
| IV               | 2.2       | 1.41–3.44 | <0.001  |
| Unknown          | 1.34      | 0.89–2.01 | 0.167   |
| Tumor size       | 1.02      | 1–1.03    | 0.033   |
| Laterality       |           |           |         |
| Bilateral        | Reference |           |         |
| Left             | 0.66      | 0.34–1.3  | 0.233   |
| Right            | 0.71      | 0.36–1.4  | 0.324   |
| T-stage          |           |           |         |
| T1               | Reference |           |         |
| T2               | 0.85      | 0.7–1.02  | 0.076   |
| T3               | 0.96      | 0.81–1.13 | 0.624   |
| T4               | 0.99      | 0.79–1.24 | 0.908   |
| TX               | 1.24      | 1–1.53    | 0.043   |
| N-stage          |           |           |         |
| N0               | Reference |           |         |
| N1               | 1.36      | 1.19–1.55 | <0.001  |
| NX               | 1.27      | 1.05–1.53 | 0.016   |
| Brain metastasis |           |           |         |
| No               | Reference |           |         |
| Yes              | 1.49      | 1.24–1.79 | <0.001  |
| Liver metastasis |           |           |         |
| No               | Reference |           |         |
| Yes              | 1.24      | 1.05–1.46 | 0.012   |
| Lung metastasis  |           |           |         |
| No               | Reference |           |         |
| Yes              | 1.5       | 1.33–1.7  | <0.001  |
| Surgery          |           |           |         |
| No               | Reference |           |         |
| Yes              | 0.39      | 0.33–0.46 | <0.001  |

of all the models. We used ROC, AUC (Figures 2 and 3), accuracy scores, specificity, and F1-scores to evaluate and compare the performance of the models. Finally, the models were externally validated using our dataset. In predicting the 1-year OS,

the XGB algorithm model performed best in the training set (AUC value of 0.891), in the validation set (AUC value of 0.711), and in the external validation cohort (AUC value of 0.812). For the LR algorithm model, the AUC value in the training set was 0.745, and the AUC values in the validation set and external validation cohort were 0.737 and 0.845, respectively. The other two algorithm models also performed well. The AUC values in the training, validation, and external validation sets of the RF model were 0.762, 0.749, and 0.801, respectively, while those of the NB model were 0.737, 0.739, and 0.809, respectively. For predicting the 3-year OS, the AUC value of the XGB algorithm model in the training set was 0.848, and the AUC values of the other three algorithms, LR, RF, and NB were 0.779, 0.790, and 0.776, respectively. The AUC values in the validation set were 0.777 for LR, 0.787 for RF, 0.762 for NB, and 0.767 for XGB. In the external validation cohort, the RF model performed best, with an AUC value of



**Figure 2.** ROC curves of the four models: lr, rf, nb and XGB. (A) training group, (B) validation group and (C) external validation group of 1-year OS predictive model.



**Figure 3.** ROC curves of the four models: lr, rf, nb and XGB. (A) training group, (B) validation group and (C) external validation group of 3-year OS predictive model.

0.706, while the other three algorithms LR, NB, and XGB had AUC values of 0.666, 0.679, and 0.674, respectively.

Finally, the accuracy, specificity, and F1 scores of the four algorithms were determined (Table 4). Compared to the other three models, the XGB model had the highest accuracy, specificity, and F1 score of 0.792, 0.806, 0.854, respectively. The LR model had an accuracy of 0.680, specificity of 0.805, and F1 score of 0.756. The RF model had an accuracy of 0.698, specificity of 0.777, and F1 score of 0.775. The accuracy, specificity, and F1 score of the NB model were 0.752, 0.680, and 0.827, respectively.

## Discussion

Bone is a common site of metastasis in ccRCC,<sup>24</sup> and bone metastases are often osteolytic lesions. SREs, such as pathological fractures, pain, and spinal cord compression caused by bone metastases, not only lead to reduced quality of life and increased medical costs, but also to a worse prognosis. In this

**Table 4.** Comparison Prediction Performances of Different Models.

| Models | Accuracy | Specificity | F1-Score |
|--------|----------|-------------|----------|
| LR     | 0.68     | 0.805       | 0.756    |
| RF     | 0.698    | 0.777       | 0.775    |
| NB     | 0.752    | 0.68        | 0.827    |
| XGB    | 0.792    | 0.806       | 0.854    |

study, we first used traditional univariate and multivariate analyses to identify independent risk factors associated with the prognosis of patients with ccRCC-BM, and then constructed prognostic models using several different ML algorithms. We evaluated the performance of the four algorithms and confirmed the feasibility and accuracy of the ML algorithms in constructing prognostic models for patients with ccRCC-BM.

The role of age as a prognostic risk factor for patients with renal cell carcinoma has been controversial.<sup>25–27</sup> Our study showed that age was an independent risk factor for the

prognosis of patients with ccRCC-BM. Marital status was also an independent prognostic factor, which is consistent with previous findings. Several studies have shown that marriage is a protective factor in the prognosis of oncology patients.<sup>28–30</sup> This may be attributed to the psychological and financial support provided by the spouse as well as daily living assistance. Regarding clinicopathological features, tumor size, T stage, N stage, pathological grade, brain metastasis, liver metastasis, lung metastasis, and surgery of the primary lesion were confirmed as independent risk factors for ccRCC-BM. The effects of different metastatic sites on the prognosis of patients with renal cell carcinoma have also been widely reported. Lin et al.<sup>31</sup> reported that patients with isolated bone metastases had the highest overall survival rate, and that patients with multiple bone metastases had higher survival rates than those with pulmonary metastases. McKay et al.<sup>8</sup> showed that liver metastases negatively impacted the prognosis of patients with renal cell carcinoma receiving targeted therapy. In addition, some studies have suggested that brain metastases are a risk factor for increased risk of death and worse prognosis in patients with renal cell carcinoma.<sup>32,33</sup>

We selected the above factors to construct ML models for predicting the 1-year and 3-year overall survival of patients with ccRCC-BM. Internal validation was performed using the SEER dataset, and external validation was performed using data from our center. The performance of the models was then evaluated using the AUC values, accuracy, specificity, and F1 scores. The models constructed using all four machine learning algorithms performed well, but the XGBoost algorithm had the best performance; this new artificial intelligence algorithm has several advantages in terms of preventing overfitting, cross-validation, and handling missing values. In the medical industry, this new ML algorithm is becoming increasingly popular and is frequently used for disease prediction and early diagnosis. To the best of our knowledge, there are no ML applications for predicting the prognosis of patients with bone metastases from renal cell carcinoma. In this study, we validated the feasibility of ML using four algorithmic models. This facilitates the advancement of ML applications in medicine.

However, this study had some limitations. First, this was a retrospective study; therefore, bias in case selection could not be avoided and further validation with prospective data is needed. Second, the data were obtained from the SEER database, and some clinical features, such as the number of bone metastases, SREs, and time to bone metastases, were not available from the SEER database. Third, immunotherapy and targeted therapies that are recommended for patients with ccRCC, were not recorded in the SEER database. Finally, although we performed external validation, the number of patients was not sufficiently large.

In future studies, we will collect more cases to validate the model and improve the algorithm to increase the stability of the model. In addition, we will attempt to use ML for diagnostic and prognostic studies of other diseases.

## Conclusion

We developed ML models to predict overall survival of patients with ccRCC-BM. All the models performed well. We also identified the risk factors affecting the prognosis of patients with bone metastases from renal clear-cell carcinoma, which will provide relevant information for clinicians and help in clinical decision-making.

## Acknowledgments

We'd like to express our gratitude to the SEER database for allowing us to access free and open data.

## Data Availability Statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Declaration of Conflicting Interests

The author(s) declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Ethical Approval

The study was approved by the ethics committee of Peking University People's Hospital, Approval Number: 2022PHB394-001.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iD

Yijun Le  <https://orcid.org/0000-0001-5730-3350>

## References

1. Capitanio U, Bensalah K, Bex A, et al. Epidemiology of renal cell carcinoma. *Eur Urol*. 2019;75(1):74-84.
2. Ljungberg B, Bensalah K, Canfield S, et al. EAU Guidelines on renal cell carcinoma: 2014 update. *Eur Urol*. 2015;67(5):913-924.
3. Capitanio U, Cloutier V, Zini L, et al. A critical assessment of the prognostic value of clear cell, papillary and chromophobe histological subtypes in renal cell carcinoma: A population-based study. *BJU Int*. 2009;103(11):1496-1500.
4. Keegan KA, Schupp CW, Chamie K, et al. Histopathology of surgically treated renal cell carcinoma: Survival differences by subtype and stage. *J Urol*. 2012;188(2):391-397.
5. Gandaglia G, Ravi P, Abdollah F, et al. Contemporary incidence and mortality rates of kidney cancer in the United States. *Can Urol Assoc J*. 2014;8(7–8):247-252.
6. Sun M, Thuret R, Abdollah F, et al. Age-adjusted incidence, mortality, and survival rates of stage-specific renal cell carcinoma in North America: A trend analysis. *Eur Urol*. 2011;59(1):135-141.
7. Brozovich A, Garnezy B, Pan T, et al. All bone metastases are not created equal: Revisiting treatment resistance in renal cell carcinoma. *J Bone Oncol*. 2021;31:100399.

8. McKay RR, Kroeger N, Xie W, et al. Impact of bone and liver metastases on patients with renal cell carcinoma treated with targeted therapy. *Eur Urol*. 2014;65(3):577-584.
9. Heng DY, Xie W, Bjarnason GA, et al. Progression-free survival as a predictor of overall survival in metastatic renal cell carcinoma treated with contemporary targeted therapy. *Cancer*. 2011;117(12):2637-2642.
10. Grunwald V, Eberhardt B, Bex A, et al. An interdisciplinary consensus on the management of bone metastases from renal cell carcinoma. *Nat Rev Urol*. 2018;15(8):511-521.
11. Woodward E, Jagdev S, McParland L, et al. Skeletal complications and survival in renal cancer patients with bone metastases. *Bone*. 2011;48(1):160-166.
12. Stumpo V, Staartjes VA, Esposito G, et al. Machine learning and intracranial aneurysms: From detection to outcome prediction. *Acta Neurochir Suppl*. 2022;134:319-331.
13. Zilcha-Mano S, Roose SP, Brown PJ, et al. A machine learning approach to identifying placebo responders in late-life depression trials. *Am J Geriatr Psychiatry*. 2018;26(6):669-677.
14. Vey J, Kapsner LA, Fuchs M, et al. A Toolbox for Functional Analysis and the Systematic Identification of Diagnostic and Prognostic Gene Expression Signatures Combining Meta-Analysis and Machine Learning. *Cancers (Basel)*. 2019;11(10):1606.
15. Marostica E, Barber R, Denize T, et al. Development of a histopathology informatics pipeline for classification and prediction of clinical outcomes in subtypes of renal cell carcinoma. *Clin Cancer Res*. 2021;27(10):2868-2878.
16. Zhu Z, Wang SH, Górriz JM, et al. BCNet: A novel network for blood cell classification. *Front Cell Dev Biol*. 2021;9:813996.
17. Zhu Z, Wang S, Zhang Y. ROENet: A ResNet-based output ensemble for malaria parasite classification. *Electronics (Basel)*. 2022;11(13):2040.
18. Frizzell JD, Liang L, Schulte PJ, et al. Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: Comparison of machine learning and other statistical approaches. *JAMA Cardiol*. 2017;2(2):204-209.
19. Huang Z, Zhang X, Liu H, et al. Clear cell renal cell carcinoma bone metastasis: What should be considered in prognostic evaluation. *Eur J Surg Oncol*. 2019;45(7):1246-1252.
20. Du Y, Pahernik S, Hadaschik B, et al. Survival and prognostic factors of patients with renal cell cancer with bone metastasis in the era of targeted therapy: A single-institution analysis. *Urol Oncol*. 2016;34(10):433 e1-8.
21. Wang K, Wu Z, Wang G, et al. Survival nomogram for patients with bone metastatic renal cell carcinoma: A population-based study. *Int Braz J Urol*. 2021;47(2):333-349.
22. Zhou H, Yang S, Xie T, et al. Risk factors, prognostic factors, and nomograms for bone metastasis in patients with newly diagnosed clear cell renal cell carcinoma: A large population-based study. *Front Surg*. 2022;9:877653.
23. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Br Med J*. 2015;350:g7594.
24. Chandrasekar T, Klaassen Z, Goldberg H, et al. Metastatic renal cell carcinoma: Patterns and predictors of metastases-A contemporary population-based series. *Urol Oncol*. 2017;35(11):661 e7-661 e14.
25. Fan Z, Huang Z, Huang X. Bone metastasis in renal cell carcinoma patients: Risk and prognostic factors and nomograms. *J Oncol*. 2021;2021:5575295.
26. Kang HW, Seo SP, Kim WT, et al. Impact of young age at diagnosis on survival in patients with surgically treated renal cell carcinoma: A multicenter study. *J Korean Med Sci*. 2016;31(12):1976-1982.
27. Thompson RH, Ordonez MA, Ordonez A, et al. Renal cell carcinoma in young and old patients—is there a difference? *J Urol*. 2008;180(4):1262-1266.
28. Du X, Zhan W, Li X, et al. Marital status and survival in laryngeal squamous cell carcinoma patients: A multinomial propensity scores matched study. *Eur Arch Otorhinolaryngol*. 2022;279(6):3005-3011.
29. Merrill RM, Johnson E. Benefits of marriage on relative and conditional relative cancer survival differ between males and females in the USA. *J Cancer Surviv*. 2017;11(5):578-589.
30. Tang L, Pan Z, Zhang X. The effect of marital status on the survival of patients with multiple myeloma. *Hematology*. 2022;27(1):187-197.
31. Lin PP, Mirza AN, Mirza VO, et al. Patient survival after surgery for osseous metastases from renal cell carcinoma. *J Bone Joint Surg Am*. 2007;89(8):1794-1801.
32. Vornicova O, Bar-Sela G. Do we have a “game changer” in treating patients with brain metastasis from renal cell carcinoma? *Ann Transl Med*. 2019;7(Suppl 8):S360.
33. Vickers MM, Al-Harbi H, Choueiri TK, et al. Prognostic factors of survival for patients with metastatic renal cell carcinoma with brain metastases treated with targeted therapy: Results from the international metastatic renal cell carcinoma database consortium. *Clin Genitourin Cancer* 2013;11(3):311-315.