

OPEN ACCESS

EDITED BY
Enrico Capobianco,
Jackson Laboratory, United States

REVIEWED BY
Wenle Li,
Xiamen University, China
Wubin Ding,
Children's Hospital of Philadelphia,
United States
Wencai Liu,
Shanghai Jiao Tong University, China

*CORRESPONDENCE
Qing Bi
bqzjsrmyy@163.com
Yao Kang
kyc0730@163.com

†These authors have contributed
equally to this work

SPECIALTY SECTION
This article was submitted to
Digital Public Health,
a section of the journal
Frontiers in Public Health

RECEIVED 11 August 2022
ACCEPTED 02 November 2022
PUBLISHED 17 November 2022

CITATION
Ji L, Zhang W, Huang J, Tian J,
Zhong X, Luo J, Zhu S, He Z, Tong Y,
Meng X, Kang Y and Bi Q (2022) Bone
metastasis risk and prognosis
assessment models for kidney cancer
based on machine learning.
Front. Public Health 10:1015952.
doi: 10.3389/fpubh.2022.1015952

COPYRIGHT
© 2022 Ji, Zhang, Huang, Tian, Zhong,
Luo, Zhu, He, Tong, Meng, Kang and
Bi. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Bone metastasis risk and prognosis assessment models for kidney cancer based on machine learning

Lichen Ji^{1,2,3,4†}, Wei Zhang^{2,3,5†}, Jiaqing Huang^{2,6†},
Jinlong Tian^{1,2,3,7}, Xugang Zhong^{2,3,5}, Junchao Luo^{1,2,3,4},
Senbo Zhu^{1,2,3,4}, Zeju He^{1,2,3,4}, Yu Tong^{1,2,3}, Xiang Meng^{1,2,3,7},
Yao Kang^{1,2,3*} and Qing Bi^{1,2,3*}

¹Cancer Center, Department of Orthopedics, Zhejiang Provincial People's Hospital, Affiliated People's Hospital, Hangzhou Medical College, Hangzhou, Zhejiang, China, ²Department of Laboratory Medicine, Zhejiang Provincial People's Hospital, Affiliated People's Hospital, Hangzhou Medical College, Hangzhou, Zhejiang, China, ³Center for Rehabilitation Medicine, Osteoporosis Center, Zhejiang Provincial People's Hospital (Affiliated People's Hospital, Hangzhou Medical College), Hangzhou, Zhejiang, China, ⁴Department of Orthopedics, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou, China, ⁵Department of Orthopedics, Zhejiang Provincial People's Hospital, Qingdao University, Qingdao, China, ⁶The Second Clinic Medical College, Zhejiang Chinese Medicine University, Hangzhou, China, ⁷The First Affiliated Hospital of Bengbu Medical University, Bengbu, Anhui, China

Background: Bone metastasis is a common adverse event in kidney cancer, often resulting in poor survival. However, tools for predicting KCBM and assessing survival after KCBM have not performed well.

Methods: The study uses machine learning to build models for assessing kidney cancer bone metastasis risk, prognosis, and performance evaluation. We selected 71,414 kidney cancer patients from SEER database between 2010 and 2016. Additionally, 963 patients with kidney cancer from an independent medical center were chosen to validate the performance. In the next step, eight different machine learning methods were applied to develop KCBM diagnosis and prognosis models while the risk factors were identified from univariate and multivariate logistic regression and the prognosis factors were analyzed through Kaplan–Meier survival curve and Cox proportional hazards regression. The performance of the models was compared with current models, including the logistic regression model and the AJCC TNM staging model, applying receiver operating characteristics, decision curve analysis, and the calculation of accuracy and sensitivity in both internal and independent external cohorts.

Results: Our prognosis model achieved an AUC of 0.8269 (95%CI: 0.8083–0.8425) in the internal validation cohort and 0.9123 (95%CI: 0.8979–0.9261) in the external validation cohort. In addition, we tested the performance of the extreme gradient boosting model through decision curve analysis curve, Precision-Recall curve, and Brier score and two models exhibited excellent performance.

Conclusion: Our developed models can accurately predict the risk and prognosis of KCBM and contribute to helping improve decision-making.

KEYWORDS

kidney cancer, bone metastasis, diagnosis, prognosis, machine learning, predicting model

Introduction

It is estimated that kidney cancer (KC) accounts for the 16th most common malignant tumor in the world (1). Based on the NCCN Clinical Practice Guidelines in Oncology, 76,080 Americans were diagnosed with KC, and 13,780 died in 2021 because of these diseases (2). KC incidence displayed a continuous upward trend before the 1990s but has remained stable or declined in many countries (3, 4). Clear cell renal cell carcinoma (ccRCC), papillary, and chromophobe are the most common subtypes, accounting for 85–90% of all primary KC (5, 6). A study found that 75% of KC patients survived more than 5 years after diagnosis. However, once metastases were found, the 5-year survival rate of the patient dropped to 12% (2).

As a result, cancer metastasis to vital distant organs is considered the final phase of cancer progression, which involves a series of stochastic events known as the metastatic cascade (7, 8). A study by M. Bianchi found the proportions of the most common metastasis site in KC are 45.2% in lung, 29.5% in bone, 21.8% in lymph node, 20.3% in liver, 8.9% in adrenal, and 8.1% in brain (9). Instead of primary tumors, we focus on bone metastases (BM) because they are responsible for 90% of all cancer death (10). BM also causes complications, including pain, hypercalcemia, anemia, inflammation, skeletal fractures, spinal cord compression, instability, and decreased mobility (11–13).

Each of these complications compromises the quality of life and the functional status. The prognosis for kidney cancer with bone metastases (KCBM) remains uncertain, despite advances in surgery, radiation therapy, and targeted medical therapy developed to treat BM (14). In KCBM, most therapies are used to improve skeletal adverse events, not to lower BM rates (15).

There is a dire need to find a standard treatment guideline to reduce the occurrence of KCBM and improve the survival rate among KCBM patients. Furthermore, we need reliable predicting tools to assess the probability of events. Numerous works of literature have developed several nomograms to predict prognosis or evaluate the diagnosis risk of KCBM (16, 17). Nevertheless, we believe the two risk estimation models can improve performance in new ways. Recently, some scholars have used new algorithms to establish bone metastasis diagnosis

models of kidney cancer. Due to the huge data in the SEER database and the scientific nature of the new algorithm, the model performance has been improved. However, the rationality and completeness of the included variables still need to be improved, and the performance evaluation also needs to be supplemented (18).

A growing number of these methods named, artificial intelligence, or machine learning, have now been put in use for biomedical research (19). The application of machine learning can promote the overall quality of prediction in a wide range of practical applications through high-throughput training and taking an ensemble learning approach. Several models can be used to predict the risk of KCBM, including logistic regression (LR), naive Bayes BS classifier (NBC), decision tree (DT), extreme gradient boosting (XGBoost), multilayer perceptron (MLP), random forest (RF), support vector machine (SVM), and k-nearest neighbor (KNN) (20–24). Our study aims to develop several models and compare their functions using different methods of operation. Then we apply the models with the highest performance to clinical valuation and they should provide a more accurate diagnosis and prognosis of KCBM and can help develop treatment guidelines and standard treatment for KCBM.

Materials and methods

Study design and participants

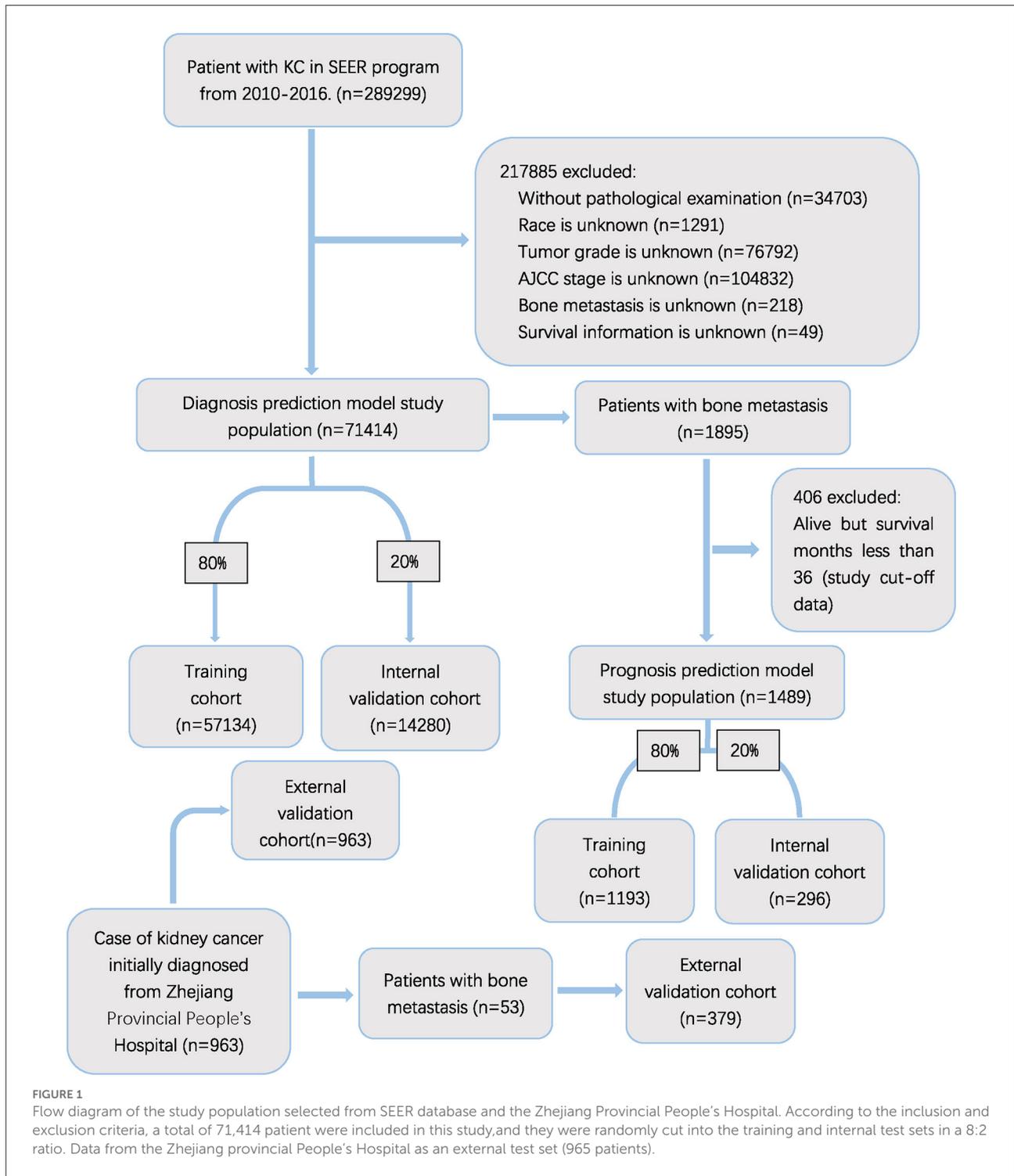
We extracted patients with KC diagnosed from 2010 to 2016 in the SEER database, considered the most common and authoritative cancer database in the USA. The inclusion criteria included: (1) kidney cancer patients with complete survival data; (2) the ensured effectiveness of follow-up; (3) the source of the case should exclude all cases obtained through necropsy and maintain those determined on the death report; (4) KC diagnosed by pathology, alone with BM observed by imaging or pathology.

The exclusion criteria involve: (1) unavailable clinical or relevant examination information, (2) unknown survival information. The flow chart of the study is shown in [Figure 1](#).

Data collection

The selected baseline characteristics were age at diagnosis, race, gender, primary site, grade, histology subtype, marital status, insurance recode, stage, TNM stage, surgery, lymph node surgery, radiation recode, chemotherapy, brain metastasis, liver metastasis, lung metastasis, laterality, and tumor size. Furthermore, histological type code were divided into seven group based on the International Classification of Disease for

Abbreviations: KCBM, kidney cancer bone metastasis; KC, kidney cancer; BM, bone metastasis; ccRCC, clear cell renal cell carcinoma; SEER, Surveillance, Epidemiology, and End Results; AJCC, American Joint Committee on Cancer; ROC, receiver operating characteristic; DCA, Decision Curve Analysis; PR, Precision-Recall; LR, logistic regression; NBC, naive Bayes BS classifier; DT, decision tree; XGB, extreme gradient boosting; MLP, multilayer perceptron; RF, random forest; SVM, support vector machine; KNN, k-nearest neighbor; ICD-O, International Classification of Disease for Oncology; OS, overall survival; AUC, Area under the curve; TP, true positives; TN, true negatives; FP, false positives; FN, false negatives.



Oncology (ICD-O): transitional cell carcinoma (8120), papillary transitional cell carcinoma (8130), papillary adenocarcinoma (8260), clear-cell adenocarcinoma (8310), renal cell carcinoma (8312), renal cell carcinoma chromophobe type (8317), other types (8000, 8005, 8010, 8012, 8013, 8020, 8022, 8031–8033,

8035, 8041, 8045, 8046, 8050, 8052, 8070–8072, 8074, 8082, 8083, 8121, 8122, 8130, 8131, 8140, 8210, 8211, 8240, 8246, 8249, 8255, 8250, 8263, 8270, 8280, 8290, 8310, 8312, 8313, 8315–8319, 9320, 8323, 8330, 8342, 8480, 8481, 8490, 8510, 8522, 8560, 8574, and 8980). Oncology staging was determined on the basis of

the 7th TNM classification of the AJCC. Regarding the survival model of KCBM, overall survival (OS) was seen as the primary endpoint event.

Statistical analysis

In this study, R (Version 4.1.3), IBM SPSS Statistics (Version 22), and Python (Version 3.9.7) were utilized to complete all statistical analyses. We used the following package: “foreign,” “survival,” “caret,” “rms,” “survminer,” “sklearn.linear_model,” “sklearn.ensemble,” “sklearn.tree,” “sklearn.svm,” “sklearn.neural_network,” “sklearn.tree.” The specific code has been uploaded to github (<https://github.com/JiLichen/Kidney-diagnosis-and-prognosis>). All KC patients were divided into a training and a validation cohort at random according to the proportion of 8:2. Continuous data was compared by independent sample *t*-tests or Mann-Whitney U tests and categorical data was compared by chi-square tests or Fisher exact test. All variables we included in the analysis were analyzed by univariate logistic regression. Values of $p < 0.05$ were regarded statistically significant. Multivariate logistic regression was utilized to test whether or not these significant factors were associated with BM in patients with KC. To determine independent prognostic factors of KCBM, the cohort with BM were grouped by an 8:2 ratio as training and validation cohorts. Then, Kaplan-Meier analysis and Cox proportional hazard regression analysis were conducted on the factors above and $p < 0.05$ were regarded statistically significant. Aiming to develop a model to predict the risk and overall survival of KCBM accurately, we used LR, NBC, DT, XGBoost, MLP, RF, SVM, and KNN algorithms based on the risk factors to establish diagnosis models. Additionally, we applied the above algorithms to independent prognostic factors to build surviving models at the 3-year observation point. After testing the various performances of the above two types of models, we selected the most representative models as clinical recommendations (25, 26).

We use the following formula to calculate the performance of the model:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Sensitivity = \frac{TP}{TP + FN} = recall \quad (3)$$

$$F1 = \frac{2 * P * R}{P + R} \quad (4)$$

$$Brierscore = \frac{1}{N} \sum_{T=1}^N (f_t - o_t) \quad (5)$$

Model visualization

We used web pages to establish risk assessment tools for diagnosis and prognosis in kidney cancer patients. Clinicians can log into the website to utilize the risk assessment tools.

Results

Cohort description

Our study included 71,414 patients from SEER database after above screening. As a result, 1,895 (2.6%) patients were observed to have KCBM. In the cohort of patients with BM, 406 patients who was alive but follow-up time < 36 were excluded. 1,385 (93.0%) of the remaining 1,489 patients died during an average of 35.74 months (Std = 22.33) of follow-up. The training cohort comprised 1,193 and remaining 296 patients formed an internal validation cohort according to the grouping ratio of 8:2, respectively. 71,414 patients of diagnose cohort was grouped in the same way. The randomness of the grouping was verified by the chi-square test and *t*-test (Supplementary Tables 1, 4). The 963 patients from Zhejiang Provincial People’s Hospital were selected as an external validation cohort, with 53 patients have KCBM. Details of variables including sociodemographic characteristics, clinical features and treatment regimens of kidney cancer patients are demonstrated in Supplementary Table 2.

The distribution and characteristics of KC group and KCBM group are shown in Supplementary Tables 2, 4. Elderly people aged 60–69 are the most common in training cohort (31.1%). Males with KC nearly twice as many as women (63.7%). Clear-cell adenocarcinoma accounts for the largest proportion among all histology types (59.7%). The commonest grade, T and *n* stage are grade II (47.4%), T1 (64.2%) and N0 (93.8%). The primary lesions are basically equal to the left and right in laterality (49.4 in left and 50.4% in right). In terms of treatment, the vast majority of patients have undergone different types of surgery. Among them, radical nephrectomy account for about half, reaching 44.5%. A small number of patients of patients underwent lymph node removed surgery (13.3%). Additionally, 1,476 (2.5%) patients were treated with radiation therapy and 4,299 (7.5%) patients received chemotherapy. Regarding the distance metastasis of KC, 2,607 (4.5%) patients were detected to have lung metastasis, 786 (1.3%) patients had liver metastasis and 392 (0.6%) had brain metastasis.

In external validation cohort, due to geographical restriction, the included cases are all Asian. Elder people age 50–69 accounts for 51.4% of all age groups. There are also about twice as many male patients as female patients. Clear-cell adenocarcinoma is the most common histology type (85.7%). Radical nephrectomy (30.7%) and partial/ subtotal nephrectomy/ partial ureterectomy (45.6%) have the largest proportion and 56 patients (5.8%) accept lymph nodes removed surgical treatment. Regarding other treatments, 15 patients (1.5%) received radiation therapy, 64 patients (6.6%) were treated with chemotherapy. The metastatic status of distant organs is shown as follows: lung metastasis (3.2%), liver metastasis (1.4%), brain metastasis (0.5%).

Independent risk factors for KCBM

According to [Supplementary Table 3](#), we selected variables about sociodemographic characteristics, clinical features, treatment regimens and performed a univariate logistic regression analysis on them. The result demonstrated that nineteen variables met the requirement of $p < 0.05$. Furthermore, the variables mentioned above were selected through multivariable logistic regression. Independent predictors of KCBM contained age, primary site, grade, histology, n stage, surgery, radiation therapy, chemotherapy, brain metastasis, liver metastasis, lung metastasis and laterality. The correlation analysis was performed with Spearman correlation coefficient test and revealed no significant correlation between variables we included in diagnosis models ([Supplementary Figure 1](#)).

Diagnosis machine learning model construction

Based on the independent predictors screened through logistic regression, we developed eight models using machine learning methods in the training cohort, such as decision tree (DT), random forest (RF), multilayer perceptron (MLP), logistic regression (LR), naive Bayes BS classifier (NBC), extreme gradient boosting (XGB), support vector machine (SVM) and k-nearest neighbor (KMN).

Diagnostic machine learning model validation

As illustrated in [Figure 2A](#), the XGB model performed well in ten-cross validations with an average AUC of 0.96 (Std = 0.01) while LR model, DT model, RF model, NBC model, MLP model, SVM model, KMN model indicated average AUC of 0.91 (Std = 0.01), 0.92 (Std = 0.01), 0.93 (Std =

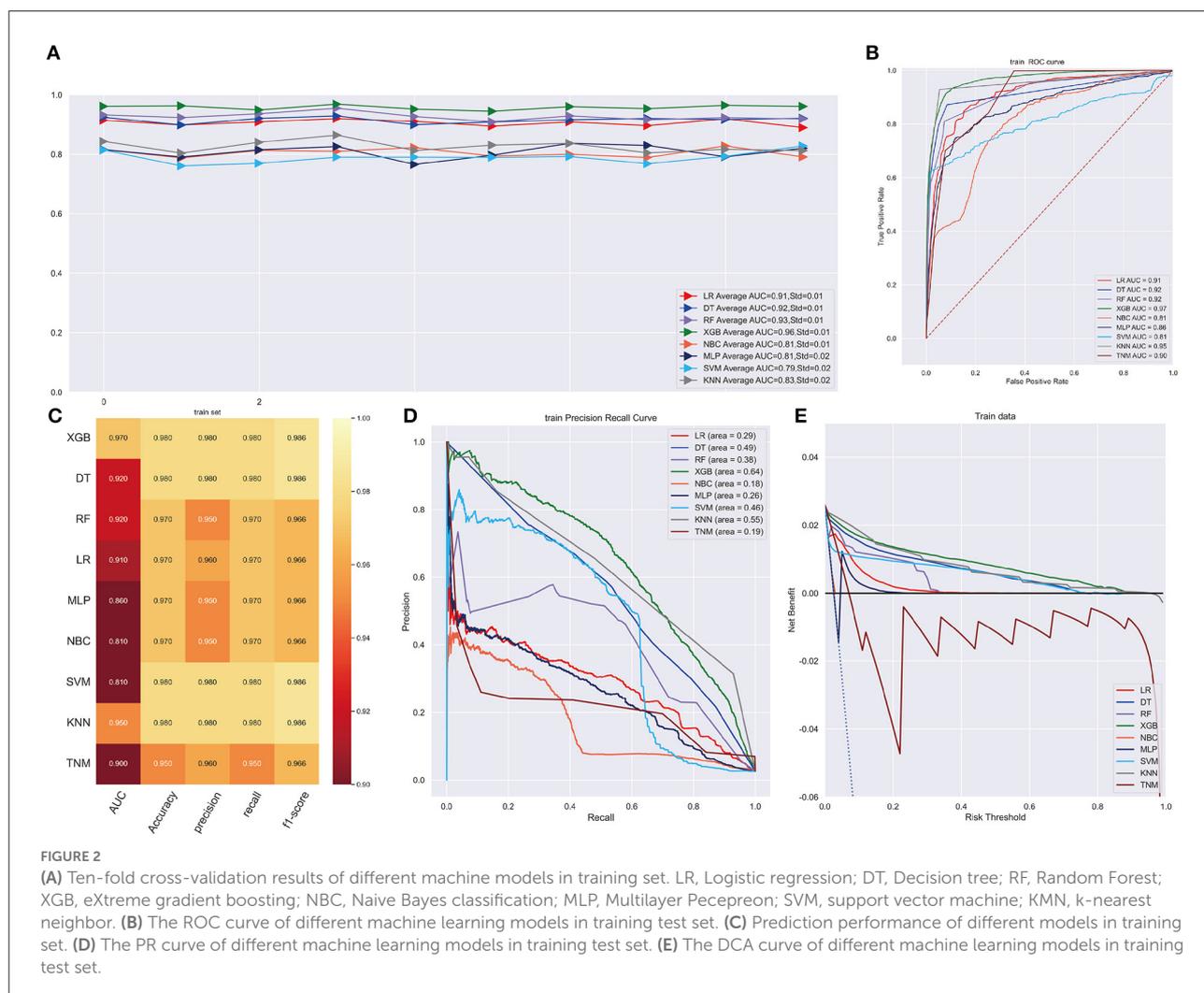
0.01), 0.81 (Std = 0.01), 0.81 (Std = 0.02), 0.79 (Std = 0.02), 0.83 (Std = 0.02). The discrimination performance of different machine learning models was evaluated by receiver operating characteristic (ROC) curve analysis and XGB model had a highest AUC of 0.97 (95%CI: 0.9469–0.9817) ([Figure 2B](#)). Additionally, XGB model achieve a Brier score of 0.014, which was lower than that Brier scores of LR model (0.022), DT model (0.016), RF model (0.019), NBC model (0.025), MLP model (0.025), SVM model (0.018), KMN model (0.015). Other performances were shown in [Figure 2C](#). PR curve and DCA curve were performed to test the models in training cohort and XGB model was proved to be highly reliable ([Figures 2D,E](#)).

XGB model achieved the best AUC of 0.960 (accuracy of 0.980, precision of 0.980, sensitivity of 0.980, f1-score of 0.981) in internal validation cohort and got an AUC of 0.980 (accuracy of 0.950, precision of 0.950, sensitivity of 0.950, f1-score of 0.940) in external validation cohort ([Figures 3A,B,E,F](#)). Moreover, [Figures 3C,G](#) shows that the area under the PR curve for the XGB is larger than any other model, including the TNM staging model. To further validate the potency of the model in clinical practice, the DCA curve depicted that the XGB model still performed well ([Figures 3D,H](#)).

In the SHAP graph, each point represents a single patient for each feature in the XGB model. The location of a point along the actual SHAP value corresponds the effect a variable had on output of the model for that case ([Figure 3I](#)). Moreover, [Supplementary Figure 2](#) showed the feature importance in each algorithm and [Supplementary Figure 3](#) illustrated the prediction results of the models as a heatmap.

Characteristics and survival analyses of KCBM

[Supplementary Figure 4A](#) displays that the overall survival curve for 1489 KCBM patients we selected from the SEER database declined rapidly before the 3-year cut-off while the curve declined slowly after the 3-year time point. As a result, selecting 3 years as the predicting node has a high clinical value for treatment planning. Clinical features and treatment regimens information of KC patients with BM are displayed in [Supplementary Table 5](#). Using a ratio of 8:2, the study population was randomly split into a training set and a validation set. The Chi-square test and Fisher's exact test results showed that there were no significant differences in the characteristics between the training cohort and validation cohort ([Supplementary Table 4](#)). When the Kaplan-Meier survival curves and log-rank tests were used on categorical



variables, it was discovered that characteristics including race ($p = 0.47$), insurance ($p = 0.980$), and lymph node surgery ($p = 0.44$) were not thought to have a sufficient influence on survival. Age ($p < 0.001$), sex ($p = 0.037$), primary site ($p < 0.001$), grade ($p < 0.001$), histology ($p < 0.001$), marital status ($p < 0.001$), T stage ($p < 0.001$), n stage ($p < 0.001$), surgery ($p < 0.001$), radiation therapy ($p = 0.001$), chemotherapy ($p = 0.002$), brain metastasis ($p < 0.001$), liver metastasis ($p < 0.001$), lung metastasis ($p < 0.001$), and laterality ($p=0.002$) affected survival significantly (Supplementary Figures 4B–P). The continuous variable such as tumor size was evaluated using Cox proportional hazard regression analysis and it was closely related to patient survival (HR 1.001, 95%CI 1.000–1.001, $p < 0.001$). Lastly, it was found that age, sex, primary site, grade, histology subtype, marital status, T stage, n stage, surgery, chemotherapy, radiation therapy, brain metastasis, liver metastasis, lung metastasis, laterality and tumor size were tested as independent prognostic

factors for OS and these variables were selected into model construction (Supplementary Table 6). A correlation analysis was used to interpret the independence of each selected variable (Supplementary Figure 5).

Prognostic machine learning model development and validation

Following data splitting, patients were used for training and ten-fold cross-validations of the algorithms while validation was performed using the remaining 296 patients. Figure 4A shows that XGB got a highest average AUC of 0.84 (Std = 0.06), predicting 3-year OS in the training cohort. The AUC values were 0.87 (95%CI: 0.8499–0.8824) for the prediction of 3-year OS indicating the model had a superior predictive ability. AUC of other machine learning models and TNM staging is listed in Figure 4B. In internal and external validation

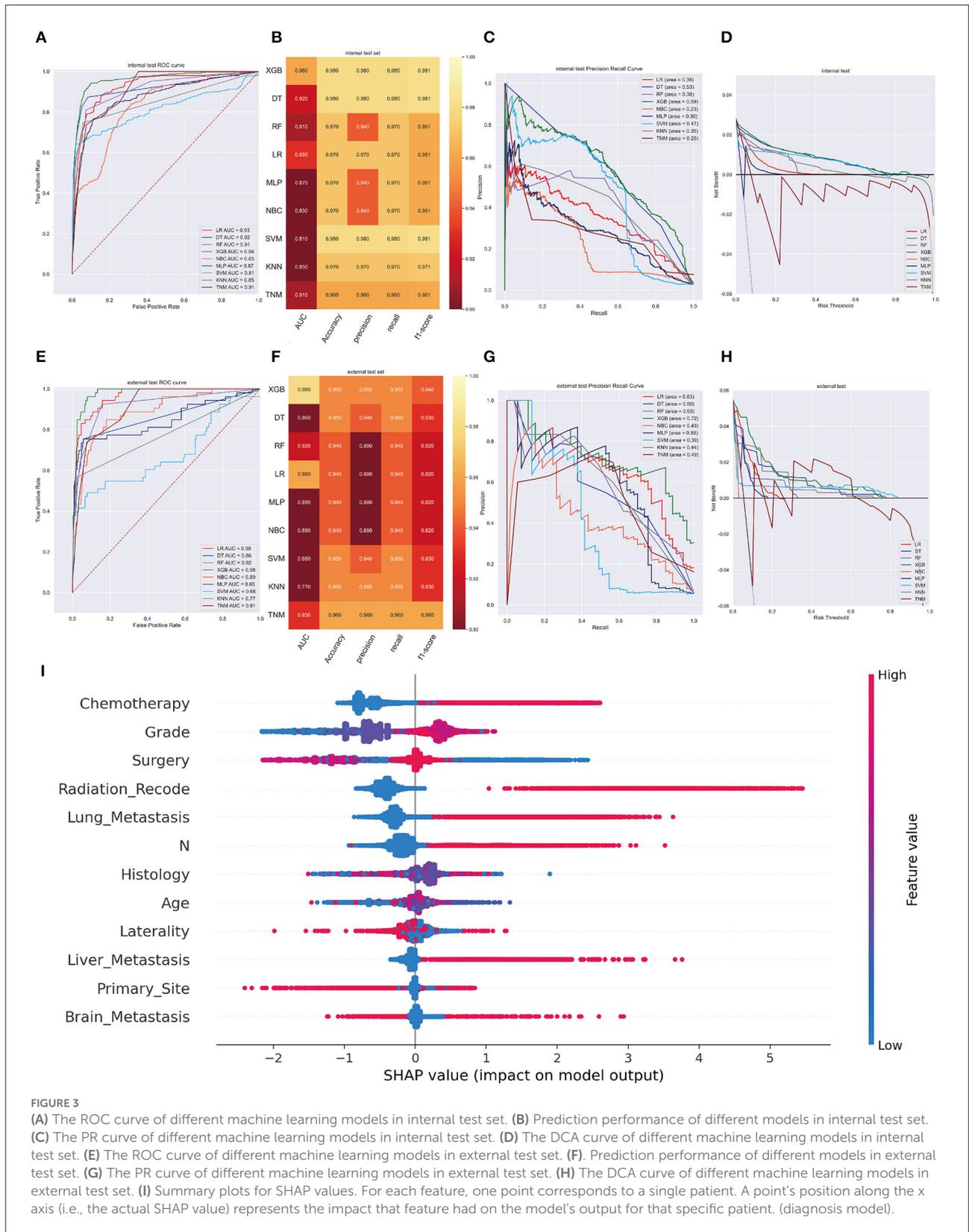
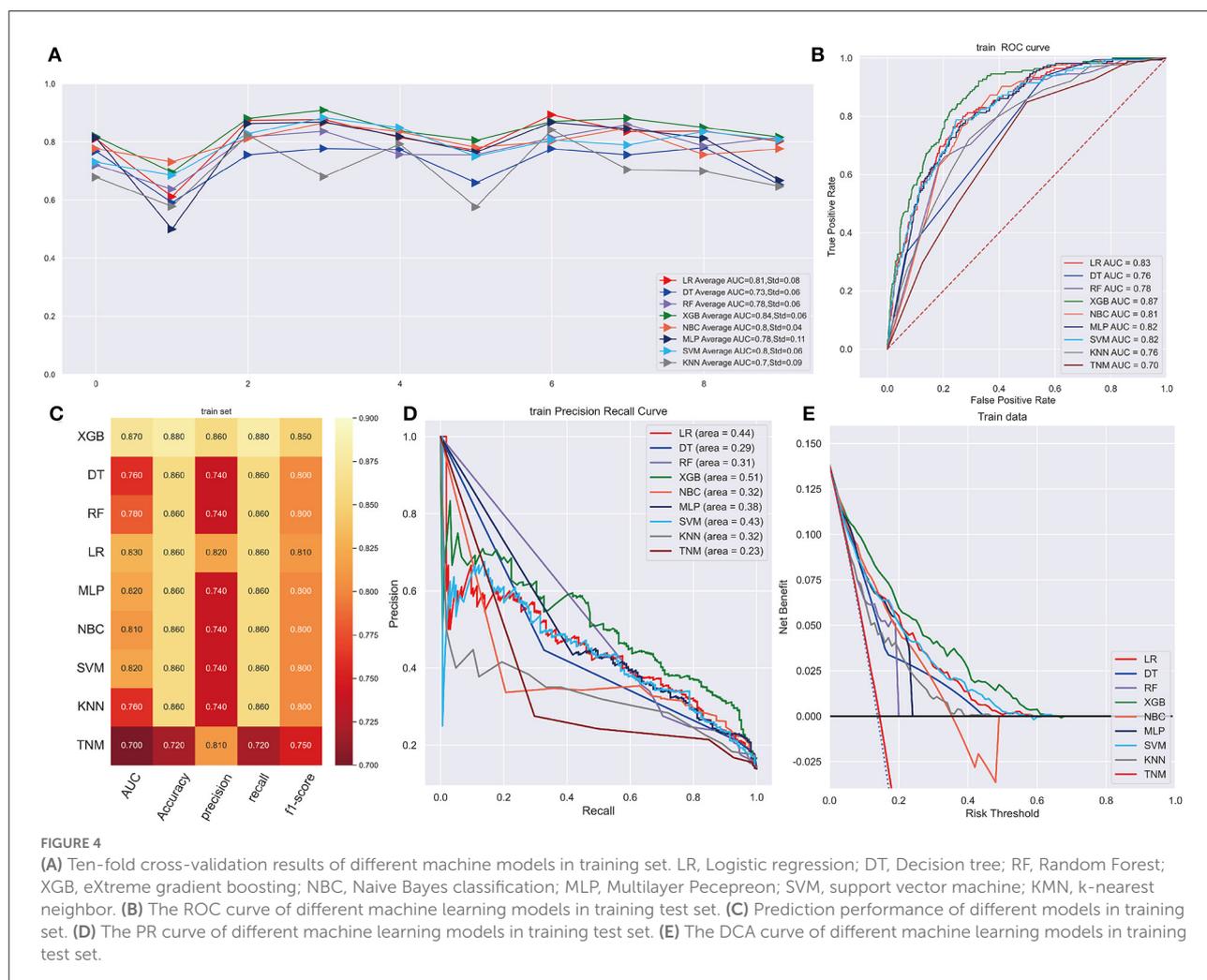


FIGURE 3

(A) The ROC curve of different machine learning models in internal test set. (B) Prediction performance of different models in internal test set. (C) The PR curve of different machine learning models in internal test set. (D) The DCA curve of different machine learning models in internal test set. (E) The ROC curve of different machine learning models in external test set. (F) Prediction performance of different models in external test set. (G) The PR curve of different machine learning models in external test set. (H) The DCA curve of different machine learning models in external test set. (I) Summary plots for SHAP values. For each feature, one point corresponds to a single patient. A point's position along the x axis (i.e., the actual SHAP value) represents the impact that feature had on the model's output for that specific patient. (diagnosis model).



cohort, our prognosis model got AUCs of 0.83 (95%CI: 0.8083–0.8425) and 0.91 (95%CI: 0.8979–0.9261), respectively (Figures 5A,E). We calculated the integrated Brier score to assess the accuracy of the established models and the results were as follows, XGB: 0.088; DT: 0.103; RF: 0.110; LR: 0.097; MLP: 0.104; NBC: 0.104; SVM: 0.097; KMN: 0.107. The metrics of each machine learning algorithm on these datasets are demonstrated in Figures 4C, 5B,F. Furthermore, the area under PR curve reached 0.51 in training cohort, 0.49 in the internal validation cohort, and 0.61 in the external validation cohort (Figures 4D, 5C,G). DCA suggests net benefits of the eight different models and TNM staging as displayed in Figures 4E, 5D,H. We estimated each feature’s impact on predicting prognosis in the XGB model by observing the SHAP values (Figure 5I). The feature importance of each model was shown in Supplementary Figure 6. The predicted results of the XGB model and the TNM staging in training and validation cohort are listed on the heatmap (Supplementary Figure 7).

Web predictor

To contribute to clinical decision-making, we developed two web-based calculators based on XGB machine learning algorithm for KCBM diagnosis and prognosis prediction. The website addresses were as follows:

- <https://share.streamlit.io/lry4000/sa/main>;
- https://share.streamlit.io/lryoxidkghwqls/survival_three_years/main.

Users can directly enter variable values and estimate the probability of occurrence and survival of KCBM. A snapshot of the online calculator is demonstrated in Figure 6.

Discussion

One of the most frequent locations for the spread of hemogenous tumor cells in KC is the bone. At advanced illness stages, patients with KC are more prone to develop painful and

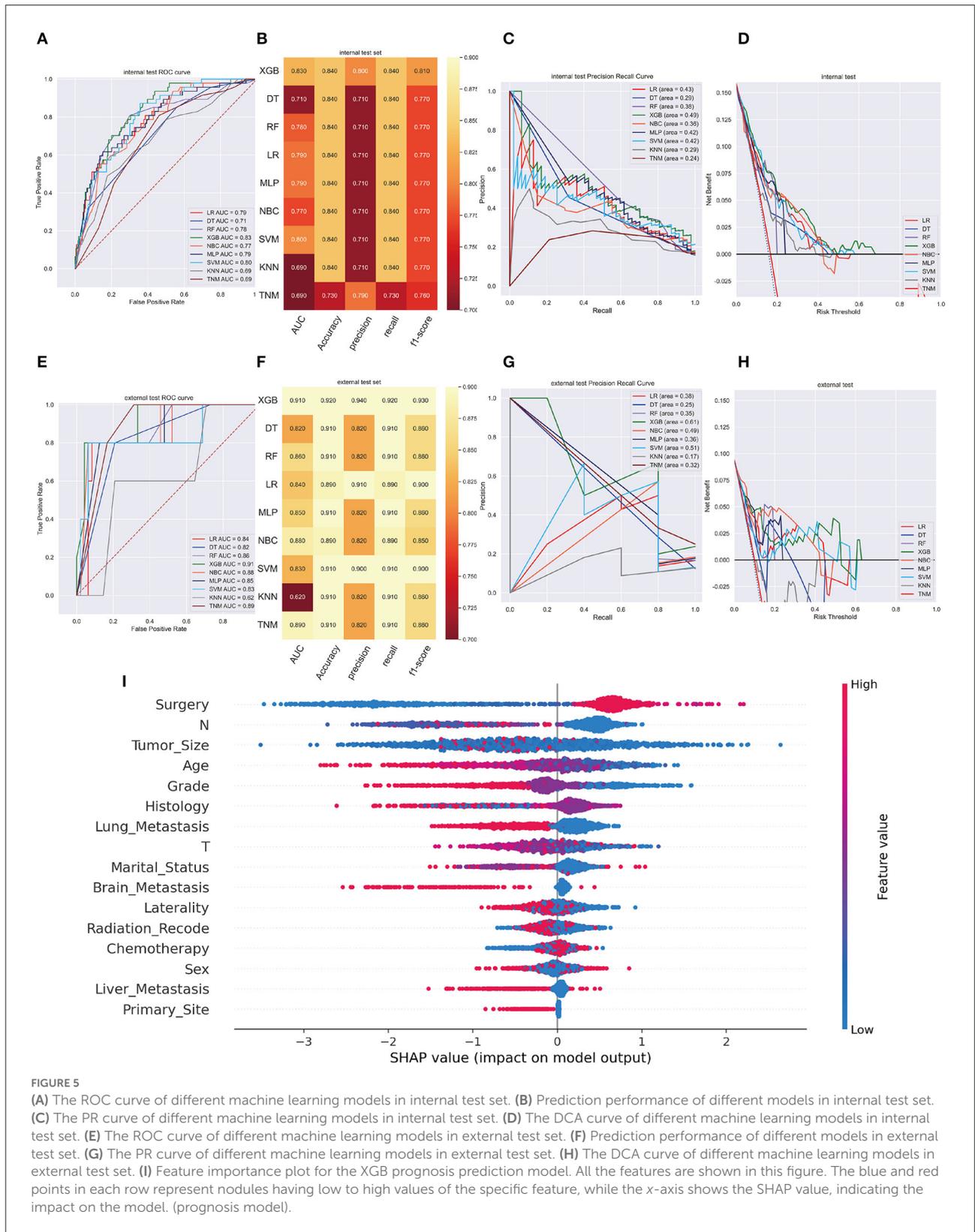
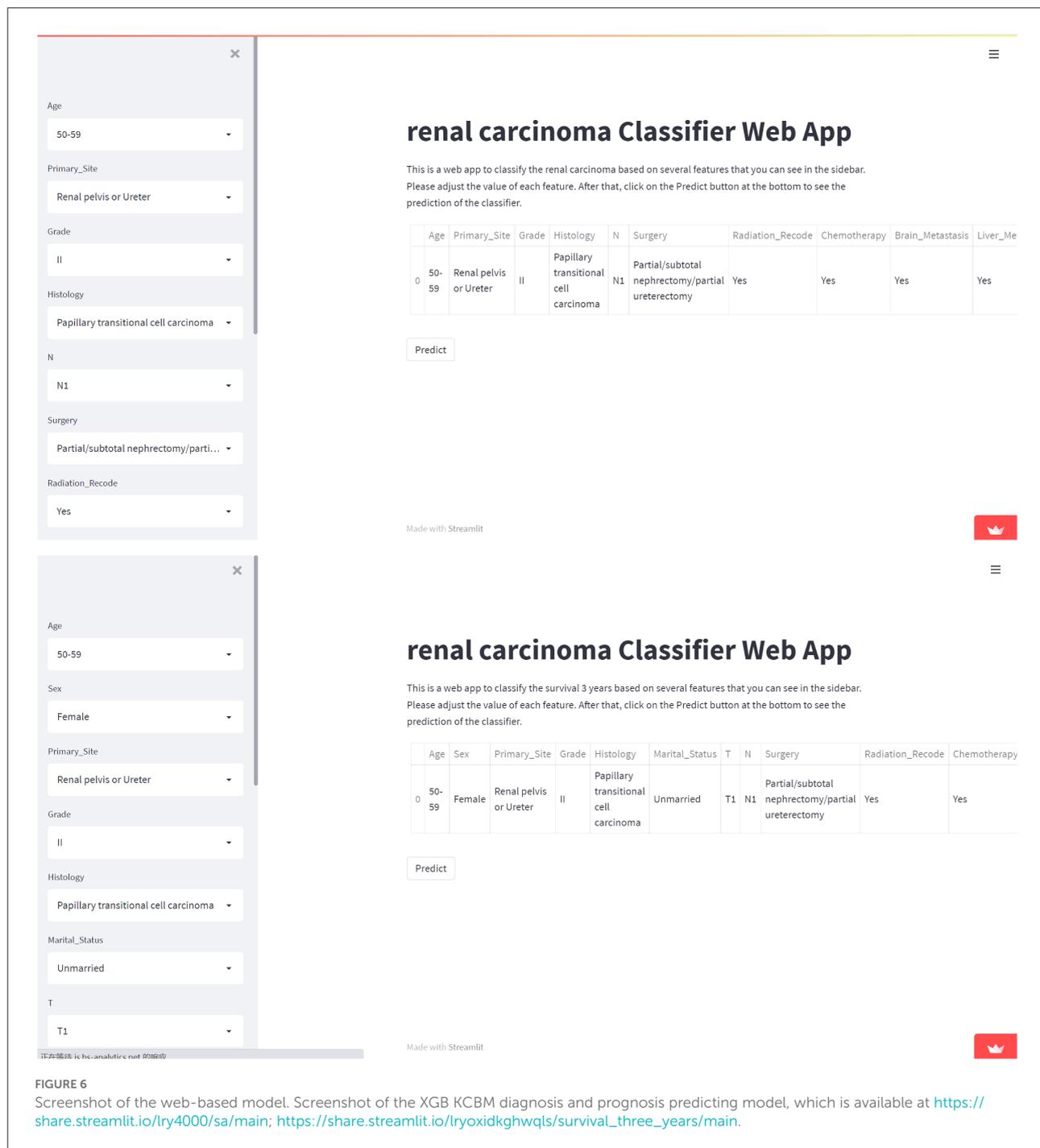


FIGURE 5

(A) The ROC curve of different machine learning models in internal test set. (B) Prediction performance of different models in internal test set. (C) The PR curve of different machine learning models in internal test set. (D) The DCA curve of different machine learning models in internal test set. (E) The ROC curve of different machine learning models in external test set. (F) Prediction performance of different models in external test set. (G) The PR curve of different machine learning models in external test set. (H) The DCA curve of different machine learning models in external test set. (I) Feature importance plot for the XGB prognosis prediction model. All the features are shown in this figure. The blue and red points in each row represent nodules having low to high values of the specific feature, while the x-axis shows the SHAP value, indicating the impact on the model. (prognosis model).



functionally incapacitating bone metastases (27). Furthermore, approximately one-third of patients with metastasis KC have already been diagnosed with bone metastasis, and another third of patients will develop them during their course of the disease (28). In our study, we found that the survival rate of patients with KCBM decreased sharply at 3 years, illustrating the poor diagnosis and the high rate of skeletal-related events (SREs), including pathological fractures, radiotherapy, surgery, neural compression, or hypercalcemia could reach 74–84%

(27). However, the treatment aims at palliation that includes interferon-alpha interleukin-2 and targeted therapy with drugs based on tyrosine kinase inhibitors, TKIs, and mTOR inhibitors (29–31). We must intervene early in patients who are at high risk of KCBM and SREs to reduce the incidence of these disease and the occurrences of SREs. Historically, a nomogram was often used to establish a diagnosis evaluation model, but it has several limitations. There was no way to compensate, resulting in the removal of too many cases reducing the accuracy of

the model. Through logistic analysis, traditional nomograms can only reached an AUC of 0.83 (16). A variety of machine learning algorithms and artificial intelligence systems have been developed as a result of advances in computer technology, and they are now being used more frequently in the field of medical biology to create diagnostic or prognosis models, offer solutions to automated decision support systems for personalized treatment, and perform other tasks that increase the effectiveness of the healthcare system (32).

We used descriptive statistics and logistic regression analysis to investigate variables related to KCBM at the time of diagnosis and exported the SHAP values to evaluate the impact of each factor. We found that younger people are more likely than elders to develop KCBM. Elderly patients are likely to have a worse prognosis. Evidence shows advanced age increases breast cancer's risk of bone metastasis. The fact that age is considered a protective factor for other cancers such as pancreatic cancer, may imply that age is type-specific as an independent risk factor for diagnosis (33, 34). Mitchell Fane and Ashani T. Weeraratna believe that age affects metastasis in several ways including changes in the immune microenvironment (inflaming, immunosenescence, tumor-associated macrophages, myeloid-derived suppressor cells, regulatory T cell), that result in loss of tissue extracellular matrix integrity (35). Paradoxically, many factors involved in the evolution of age tissue that promote malignant transformation and hyperplastic growth contribute to the arrest of growth, apoptosis, and degradation of other cells and component of structural tissue components (36). We hypothesized that suppressive metastasis factors are stronger than prometastasis factors in KC's immune microenvironment and ECM of KC. Furthermore, most studies still agree that patients with KCBM will have a poorer prognosis of KCBM patients with aging (17, 37).

Treatment-wise, according to article written by George S. Karagiannis et al. on the residual breast cancer of patients treated with neoadjuvant paclitaxel after receiving doxorubicin plus cyclophosphamide, the density and activity of TMEM sites and Mena expression (a significant locus for tumor microenvironment of metastasis) increased. This suggests that chemotherapy, despite shrinking tumor size, increases the chance of metastatic (38). The effect of chemotherapy on BM in KC may be similar to that in breast cancer. In the SHAP graph, the characteristics of the factors confirmed that the application of chemotherapy plays a significant role in KCBM. In contrast, the prognosis SHAP graph illustrates the significance of chemotherapy treatment for the 5-year OS. Our study also depicted that the implementation of radiotherapy did not reduce BM rate and in KC patients or prolong survival time. International recommendations do not support the use of adjuvant radiation following nephrectomy. According to the Copenhagen Renal Cell Cancer Study Group's randomized experiment, radiation had no observable toxic effects and

did not improve survival after 2 years (39). Nevertheless, to treat bone and brain metastases, radiotherapy, particularly stereotactic radiotherapy, can significantly relieve local systems (40). The mode, dose and mechanism of radiotherapy are complex. Radiotherapy's effects on controlling bone metastasis and prolong renal cancer prognosis depend on the mode and dose of delivery (41). In our study, surgery was deemed necessary to prevent metastasis and improve the prognosis.

SREs induced by KCBM can lead to reduce quality of life and an increase in health-care burdens (42). Hence, preventing SRE and KCBM is important for the management of patient with KC. In addition, the ESMO clinical practice guidelines suggest that treatment with anti-SRE drugs such as BMAs, denosumab is recommended for those who have a life expectancy >3 months (43). Therefore, accurate prediction of the prognosis and patients and identification of the predictive risk factors for BM are also important to guild the early initiation of anti-SREs treatment.

DT is commonly utilized for highly accurate tumor categorization and image screening (21, 44). A machine learning classifier called RF utilizes various trees to train and forecast variables that could reduce training variance and enhance integration and generalization (23, 45). MLP can use cross-entropy loss along with the stochastic gradient descent optimization with a momentum algorithm to improve the model's performance (46). LR models are commonly used to validate the influence of trait variables on end events. LR models are seen as binary classifiers (47). NBC is a model based on applying of the Bayes' theorem whose basic assumption is conditional independence of predictors based on the outcome (48). A machine learning algorithm called XGB that uses the gradient boosting framework (22, 49). SVM is often used to process gene expression profiles from tumor samples or peripheral blood for diagnosis or prognosis (20). Non-parametric classification methods like KMN are widely used; however, they can be impractical to implement with large databases because of memory consumption. Several techniques have been developed recently to improve these method (24). Using 10-fold cross-validation, optimal model hyperparameters were selected and fine-tuned by grid research. Overfitting was controlled by using the early-stop method (50).

Machine learning algorithms were suitable for observing associations between data beyond one-dimensional statistical methods such as logistic regression or Cox proportional hazard modeling. As computing power and storage space increase, machine learning algorithms can analyze more complex data and output instantaneously. Since traditional nomogram models must delete a large number of incomplete information cases, improving their prediction performance is always challenging. We maintained a large sample size, further enhanced by ten-fold cross-validation during model construction (51, 52). In recent years, XGB has become one of the most popular and innovative algorithms and has won the machine learning competition (49).

Comparisons between our machine learning models with the TNM staging model and the other seven algorithms, including logistic regression, suggested that the XGB model incorporating clinical characteristics and treatment information input can effectively predict KCBM diagnosis and prognosis. A network calculator based on the XGB algorithm has been developed to visualize diagnostic and prognostic models and increase their speed and efficiency.

Conclusion

To diagnose and prognosticate KCBM, we develop a variety of diagnostic and prognostic prediction models using machine learning and artificial intelligence technology. The XGB model was selected, and network tools were established after performance comparison. Using these models, clinicians can identify people individuals at high risk of BM and predict the prognostic of for patients with BM so that early treatment can improve prognosis and quality of life.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary material](#), further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving human participants were reviewed and approved by Medical Ethics Committee of Zhejiang Provincial People's Hospital Zhejiang Provincial People's Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

QB and YK designed the project, reviewed, and edited the manuscript. LJ, WZ, and JH wrote the manuscript. JT

and JL contributed to the literature retrieval. XZ and SZ carried out the research selection, data extraction, and statistical analysis. YT, ZH, and XM prepared the tables and figures. All authors contributed to this article and approved the submitted version.

Funding

This study was supported by grants from National Science Foundation of China (Grant No. 81672769) and Major Science and Technology Projects of Zhejiang Province (2021C03078) as well as Medical Health Science and Technology Project of Zhejiang Provincial Health Commission, No. 2022ky583.

Conflict of interest

The authors 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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpubh.2022.1015952/full#supplementary-material>

References

1. Sung H, Ferlay J, Siegel R, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2021) 71:209–49. doi: 10.3322/caac.21660
2. Motzer R, Jonasch E, Boyle S, Carlo M, Manley B, Agarwal N, et al. NCCN guidelines insights: kidney cancer, version 1.2021. *J Natl Compr Canc Netw.* (2020) 18:1160–70. doi: 10.6004/jnccn.2020.0043
3. Chow W, Dong L, Devesa S. Epidemiology and risk factors for kidney cancer. *Nat Rev Urol.* (2010) 7:245–57. doi: 10.1038/nrurol.2010.46
4. Curado M-P, Edwards B, Shin HR, Storm H, Ferlay J, Heanue M, et al. Cancer incidence in five continents, Volume IX: IARC Press, International Agency for Research on Cancer (2007).
5. Xu W, Atkins M, McDermott D. Checkpoint inhibitor immunotherapy in kidney cancer. *Nat Rev Urol.* (2020) 17:137–50. doi: 10.1038/s41585-020-0282-3

6. Shuch B, Amin A, Armstrong A, Eble J, Ficarra V, Lopez-Beltran A, et al. Understanding pathologic variants of renal cell carcinoma: distilling therapeutic opportunities from biologic complexity. *Eur Urol.* (2015) 67:85–97. doi: 10.1016/j.eururo.2014.04.029
7. Nguyen DX, Bos PD, Massague J. Metastasis: from dissemination to organ-specific colonization. *Nat Rev Cancer.* (2009) 9:274–84. doi: 10.1038/nrc2622
8. Obenauf A, Massagué J. Surviving at a distance: organ-specific metastasis. *Trends in cancer.* (2015) 1:76–91. doi: 10.1016/j.trecan.2015.07.009
9. Bianchi M, Sun M, Jeldres C, Shariat SF, Trinh QD, Briganti A, et al. Distribution of metastatic sites in renal cell carcinoma: a population-based analysis. *Ann Oncol.* (2012) 23:973–80. doi: 10.1093/annonc/mdr362
10. Leber M, Efferth T. Molecular principles of cancer invasion and metastasis (review). *Int J Oncol.* (2009) 34:881–95. doi: 10.3892/ijo_00000214
11. Jimenez-Andrade JM, Mantyh WG, Bloom AP, Ferng AS, Geffre CP, Mantyh PW. Bone cancer pain. *Ann n Y Acad Sci.* (2010) 1198:173–81. doi: 10.1111/j.1749-6632.2009.05429.x
12. Coleman RE. Clinical features of metastatic bone disease and risk of skeletal morbidity. *Clin Cancer Res.* (2006) 12:6243s–9s. doi: 10.1158/1078-0432.CCR-06-0931
13. Mercadante S. Malignant bone pain: pathophysiology and treatment. *Pain.* (1997) 69:1–18. doi: 10.1016/S0304-3959(96)03267-8
14. Chen SC, Kuo PL. Bone metastasis from renal cell carcinoma. *Int J Mol Sci.* (2016) 17:6. doi: 10.3390/ijms17060987
15. Huang Z, Du Y, Zhang X, Liu H, Liu S, Xu T. Clear cell renal cell carcinoma bone metastasis: what should be considered in prognostic evaluation. *Eur J Surg Oncol.* (2019) 45:1246–52. doi: 10.1016/j.ejso.2019.01.221
16. Dong S, Yang H, Tang ZR, Ke Y, Wang H, Li W, et al. Development and validation of a predictive model to evaluate the risk of bone metastasis in kidney cancer. *Front Oncol.* (2021) 11:731905. doi: 10.3389/fonc.2021.731905
17. Wang K, Wu Z, Wang G, Shi H, Xie J, Yin L, et al. Survival nomogram for patients with bone metastatic renal cell carcinoma: a population-based study. *Int Braz J Urol.* (2021) 47:333–49. doi: 10.1590/s1677-5538.ibju.2020.0195
18. Xu C, Liu W, Yin C, Li W, Liu J, Sheng W, et al. Establishment and validation of a machine learning prediction model based on big data for predicting the risk of bone metastasis in renal cell carcinoma patients. *Comput Math Methods Med.* (2022) 2022:5676570. doi: 10.1155/2022/5676570
19. Deo R. Machine learning in medicine. *Circulation.* (2015) 132:1920–30. doi: 10.1161/CIRCULATIONAHA.115.001593
20. Noble W. What is a support vector machine? *Nat Biotechnol.* (2006) 24:1565–7. doi: 10.1038/nbt1206-1565
21. Payabvash S, Aboian M, Tihan T, Cha S. Machine learning decision tree models for differentiation of posterior fossa tumors using diffusion histogram analysis and structural MRI findings. *Front Oncol.* (2020) 10:71. doi: 10.3389/fonc.2020.00071
22. Chen X, Huang L, Xie D, Zhao Q, EGBMMDA. Extreme gradient boosting machine for mirna-disease association prediction. *Cell Death Dis.* (2018) 9:3. doi: 10.1038/s41419-017-0003-x
23. Jiang H, Mao H, Lu H, Lin P, Garry W, Lu H, et al. Machine learning-based models to support decision-making in emergency department triage for patients with suspected cardiovascular disease. *Int J Med Inform.* (2021) 145:104326. doi: 10.1016/j.ijmedinf.2020.104326
24. Salvador-Meneses J, Ruiz-Chavez Z, Garcia-Rodriguez J. kCompressed NN: K-Nearest neighbors with data compression. *Entropy (Basel, Switzerland).* (2019) 21:3. doi: 10.3390/e21030234
25. Adeoye J, Koochi-Moghadam M, Lo A, Tsang R, Chow V, Zheng L, et al. Deep learning predicts the malignant-transformation-free survival of oral potentially malignant disorders. *Cancers.* (2021) 13(23). doi: 10.3390/cancers13236054
26. Liu WC, Li MX, Qian WX, Luo ZW, Liao WJ, Liu ZL, et al. Application of machine learning techniques to predict bone metastasis in patients with prostate cancer. *Cancer Manag Res.* (2021) 13:8723–36. doi: 10.2147/CMAR.S330591
27. Grunwald V, Eberhardt B, Bex A, Florcken A, Gauler T, Derlin T, et al. An interdisciplinary consensus on the management of bone metastases from renal cell carcinoma. *Nat Rev Urol.* (2018) 15:511–21. doi: 10.1038/s41585-018-0034-9
28. Wood SL, Brown JE. Skeletal metastasis in renal cell carcinoma: current and future management options. *Cancer Treat Rev.* (2012) 38:284–91. doi: 10.1016/j.ctrv.2011.06.011
29. Zekri J, Ahmed N, Coleman R, Hancock B. The skeletal metastatic complications of renal cell carcinoma. *Int J Oncol.* (2001) 19:379–82. doi: 10.3892/ijo.19.2.379
30. Szendroi A, Dinya E, Kardos M, Szász A, Németh Z, Áts K, et al. Prognostic factors and survival of renal clear cell carcinoma patients with bone metastases. *Pathol Oncol Res.* (2010) 16:29–38. doi: 10.1007/s12253-009-9184-7
31. Tan TH, Pranavan G, Haxhimolla HZ, Yip D. New systemic treatment options for metastatic renal-cell carcinoma in the era of targeted therapies. *Asia-Pacific J Clin Oncol.* (2010) 6:5–18. doi: 10.1111/j.1743-7563.2010.01277.x
32. Zhu W, Xie L, Han J, Guo X. The application of deep learning in cancer prognosis prediction. *Cancers.* (2020) 12:603. doi: 10.3390/cancers12030603
33. Yao Y, Zheng X, Luo X, Wu A. Incidence, prognosis and nomograms of breast cancer with bone metastases at initial diagnosis: a large population-based study. *Am J Transl Res.* (2021) 13:10248–61.
34. Zhang W, Ji L, Wang X, Zhu S, Luo J, Zhang Y, et al. Nomogram predicts risk and prognostic factors for bone metastasis of pancreatic cancer: a population-based analysis. *Front Endocrinol (Lausanne).* (2021) 12:752176. doi: 10.3389/fendo.2021.752176
35. Fane M, Weeraratna AT. Normal aging and its role in cancer metastasis. *Cold Spring Harb Perspect Med.* (2020) 10(9). doi: 10.1101/cshperspect.a037341
36. Campisi J. Aging, cellular senescence, and cancer. *Annu Rev Physiol.* (2013) 75:685. doi: 10.1146/annurev-physiol-030212-183653
37. Kume H, Kakutani S, Yamada Y, Shinohara M, Tominaga T, Suzuki M, et al. Prognostic factors for renal cell carcinoma with bone metastasis: who are the long-term survivors? *J Urol.* (2011) 185:1611–4. doi: 10.1016/j.juro.2010.12.037
38. Karagiannis G, Pastoriza J, Wang Y, Harney A, Entenberg D, Pignatelli J, et al. Neoadjuvant chemotherapy induces breast cancer metastasis through a TMEM-mediated mechanism. *Science translational medicine.* (2017) 9:397. doi: 10.1126/scitranslmed.aan0026
39. Ljungberg B, Albiges L, Abu-Ghanem Y, Bedke J, Capitanio U, Dabestani S, et al. European association of urology guidelines on renal cell carcinoma: The 2022. *Update Eur Urol.* (2022) 3:6. doi: 10.1016/j.eururo.2022.03.006
40. Powles T, Albiges L, Staehler M, Bensalah K, Dabestani S, Giles RH, et al. Updated European association of urology guidelines: recommendations for the treatment of first-line metastatic clear cell renal cancer. *Eur Urol.* (2018) 73:311–5. doi: 10.1016/j.eururo.2017.11.016
41. De Meerleer G, Khoo V, Escudier B, Joniau S, Bossi A, Ost P, et al. Radiotherapy for renal-cell carcinoma. *Lancet Oncol.* (2014) 15:e170–7. doi: 10.1016/S1470-2045(13)70569-2
42. Antczak C, Trinh VQ, Sood A, Ravi P, Roghmann F, Trudeau V, et al. The health care burden of skeletal related events in patients with renal cell carcinoma and bone metastasis. *J Urol.* (2014) 191:1678–84. doi: 10.1016/j.juro.2013.12.042
43. Coleman R, Body J-J, Aapro M, Hadji P, Herrstedt J, Group EGW. Bone health in cancer patients: ESMO clinical practice guidelines. *Ann Oncol.* (2014) 25:iii124–iii37. doi: 10.1093/annonc/mdu103
44. Ramanujam R, Zhu F, Fink K, Karrenbauer V, Lorscheider J, Benkert P, et al. Accurate classification of secondary progression in multiple sclerosis using a decision tree. *Multiple sclerosis (Houndmills, Basingstoke, England).* (2021) 27:1240–9. doi: 10.1177/1352458520975323
45. Zhong H, Poeran J, Gu A, Wilson L, Gonzalez Della Valle A, Memtsoudis S, et al. Machine learning approaches in predicting ambulatory same day discharge patients after total hip arthroplasty. *Reg Anesth Pain Med.* (2021) 46:779–83. doi: 10.1136/rapm-2021-102715
46. Alkadri S, Ledwos N, Mirchi N, Reich A, Yilmaz R, Driscoll M, et al. Utilizing a multilayer perceptron artificial neural network to assess a virtual reality surgical procedure. *Comput Biol Med.* (2021) 136:104770. doi: 10.1016/j.compbiomed.2021.104770
47. Nick TG, Campbell KM. Logistic regression. *Methods Mol Biol.* (2007) 404:273–301. doi: 10.1007/978-1-59745-530-5_14
48. Bellocchio F, Lonati C, Ion Titapiccolo J, Nadal J, Meiselbach H, Schmid M, et al. Validation of a novel predictive algorithm for kidney failure in patients suffering from chronic kidney disease: the prognostic reasoning system for chronic kidney disease (PROGRES-CKD). *Int J Environ Res Public Health.* (2021) 18:23. doi: 10.3390/ijerph182312649
49. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (2016).
50. Zhang T, Yu B. Boosting with early stopping: Convergence and consistency. *Annals Statistic.* (2005) 33:1538–79. doi: 10.1214/009053605000000255
51. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *New Engl J Med.* (2019) 380:1347–58. doi: 10.1056/NEJMra1814259
52. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J.* (2015) 13:8–17. doi: 10.1016/j.csbj.2014.11.005