BMC Medical Imaging

RESEARCH Open Access

Computed tomography-based radiomics predicts prognostic and treatment-related levels of immune infiltration in the immune microenvironment of clear cell renal cell carcinoma

Shiyan Song[1†], Wenfei Ge[1†], Xiaochen Qi[1†], Xiangyu Che[1], Qifei Wang[1*] and Guangzhen Wu[1*]

Abstract

Objectives The composition of the tumour microenvironment is very complex, and measuring the extent of immune cell infiltration can provide an important guide to clinically significant treatments for cancer, such as immune checkpoint inhibition therapy and targeted therapy. We used multiple machine learning (ML) models to predict differences in immune infiltration in clear cell renal cell carcinoma (ccRCC), with computed tomography (CT) imaging pictures serving as a model for machine learning. We also statistically analysed and compared the results of multiple typing models and explored an excellent non-invasive and convenient method for treatment of ccRCC patients and explored a better, non-invasive and convenient prediction method for ccRCC patients.

Methods The study included 539 ccRCC samples with clinicopathological information and associated genetic information from The Cancer Genome Atlas (TCGA) database. The Single Sample Gene Set Enrichment Analysis (ssGSEA) algorithm was used to obtain the immune cell infiltration results as well as the cluster analysis results. ssGSEA-based analysis was used to obtain the immune cell infiltration levels, and the Boruta algorithm was further used to downscale the obtained positive/negative gene sets to obtain the immune infiltration level groupings. Multifactor Cox regression analysis was used to calculate the immunotherapy response of subgroups according to Tumor Immune Dysfunction and Exclusion (TIDE) algorithm and subgraph algorithm to detect the difference in survival time and immunotherapy response of ccRCC patients with immune infiltration. Radiomics features were screened using LASSO analysis. Eight ML algorithms were selected for diagnostic analysis of the test set. Receiver

[†]Shiyan Song, Wenfei Ge and Xiaochen Qi contributed equally to this work.

*Correspondence:
Qifei Wang
wangqifei6008@hotmail.com
Guangzhen Wu
wuguang0613@hotmail.com

Full list of author information is available at the end of the article

operating characteristic (ROC) curve was used to evaluate the performance of the model. Draw decision curve analysis (DCA) to evaluate the clinical personalized medical value of the predictive model.

Results The high/low subtypes of immune infiltration levels obtained by optimisation based on the Boruta algorithm were statistically different in the survival analysis of ccRCC patients. Multifactorial immune infiltration level combined with clinical factors better predicted survival of ccRCC patients, and ccRCC with high immune infiltration may benefit more from anti-PD-1 therapy. Among the eight machine learning models, ExtraTrees had the highest test and training set ROC AUCs of 1.000 and 0.753; in the test set, LR and LightGBM had the highest sensitivity of 0.615; LR, SVM, ExtraTrees, LightGBM and MLP had higher specificities of 0.789, 1.000, 0.842, 0.789 and 0.789, respectively; and LR, ExtraTrees and LightGBM had the highest accuracy of 0. 719, 0.688 and 0.719 respectively. Therefore, the CT-based ML achieved good predictive results in predicting immune infiltration in ccRCC, with the ExtraTrees machine learning algorithm being optimal.

Conclusion The use of radiomics model based on renal CT images can be noninvasively used to predict the immune infiltration level of ccRCC as well as combined with clinical information to create columnar plots predicting total survival in people with ccRCC and to predict responsiveness to ICI therapy, findings that may be useful in stratifying the prognosis of patients with ccRCC and guiding clinical practitioners to develop individualized regimens in the treatment of their patients.

Keywords Tumour immune infiltration, Immune checkpoint inhibitors, Prognosis, Clear cell renal cell carcinoma, Machine learning, Radiomics

Introduction

Renal cell carcinoma (RCC) has a high incidence in human urinary system tumors, accounting for about 3% of all cancers, and is also the most common ten cancers, and the incidence is increasing [1–3]. RCC accounts for about 90% of kidney cancers, and the main types include clear cell RCC (ccRCC). Surgical treatment has a good clinical benefit for patients with localised renal cell carcinoma, but surgery may not be the best choice for metastatic ccRCC [4, 5]. Surgical treatment combined with targeted therapy or immunotherapy can effectively prolong the survival of RCC patients, among which immune checkpoint inhibitor (ICI) plays a crucial role in RCC treatment [6, 7].

The outcome of tumor is related to immune infiltration, which is heterogeneous among different cancer types and cancer patients [8]. The tumor environment composed of immune infiltrating cells has become an important research field [9, 10], and the difference of immune infiltration affects the treatment outcomes in cancer patients. Renal cancer is a highly immunogenic tumour, and a high level of immune infiltration is one of its characteristics, which can mediate immune dysfunction by inducing immunosuppressive cells to infiltrate the tumour microenvironment [11, 12]. Immune infiltration may influence the response of ccRCC to immune checkpoint inhibitor therapy, and high levels of immune infiltration may enhance the efficacy of ICI in the treatment of ccRCC [13]. Therefore, predicting the level of immune infiltration in ccRCC may provide a more optimal treatment regimen for ccRCC patients. Currently, results of immune infiltration levels usually require invasive procedures such as percutaneous kidney biopsy and tissue sampling after surgery, however, these may not be ideal for patients who require dynamic assessment. How to predict the difference in immunoinfiltration of ccRCC by non-invasive methods needs to be developed.

Radiomics is a study that has received increasing attention from researchers in recent years, analysing data to create models that can help predict outcomes of interest [14]. Researchers are now using them to predict cancer diagnosis, progression and outcome and outcome with promising results [15–18]. 'Radiogenomics' is exploring the relationship between imaging data and disease genotypes [19, 20]. At present, most researchers use cancer genome information, clinical information and medical images provided by The Cancer Genome Atlas (TCGA) and The Cancer Imaging Archive (TCIA) for radiogenomics studies [21–24]. Computed tomography (CT) is used to examine renal masses because it is easier to use and less expensive than enhanced CT, and more studies have been conducted to apply the imaging genomics of enhanced CT to predict renal cancer grading, prognosis and other related information. But fewer studies have been conducted to apply the imaging genomics of CT to predict differences in immune infiltration in ccRCC. Here, we used the data of patients from both the TCGA and TCIA databases to develop a CT image-based radiomics to predict the level of immune infiltration in RCC, which makes it possible to noninvasively predict the survival status of ccRCC patients and the degree of response to ICI therapy.

Materials and methods

Genetic data collection

The sample inclusion criteria are as follows: [1] A total of 539 samples of the TCGA-KIRC project were obtained from the TCGA database, including tumor and para-cancer mRNA-seq TPM data, clinical data (such as age, survival time, survival state, etc.) and pathological stage and grade data, of which 72 had para-cancer tissue (tumor = 539, normal = 72), and subsequent prognostic analysis was performed to include samples with a survival time of more than 30 days ($n = 470$).

Immune cell correlation analysis

The results of the immune infiltration and clustering analysis were obtained for 539 samples based on the SINGLE SAMPLE GSEA algorithm and presented in a heat map, and the correlation coefficients of the co-expression relationship between different immune cells as well as the P-value of the risk ratio and prognostic correlation of these cells were calculated based on the quantitative differences of immune infiltration in the samples, and the co-expression relationships of the immune cells were based on the P-value of these cells. clustering.

Grouping of samples based on immune cell infiltration levels

Based on the immune infiltration levels obtained from the ssGSEA analysis, clustering analysis was used to group the samples (R package: ClassDiscovery, clustering method: ward.D). Then, based on the expression levels of more than 700 genes used to adjust the immune cells, we used unsupervised consensus clustering to classify these genes into sets of positively/negatively correlated genes. The obtained positive/negative gene set was further reduced using Boruta's algorithm and based on Equation (The ICI score quantifies the extent of immune infiltration that).

$$\text{ICI score} = \sum \text{PC1A} - \sum \text{PC1B} \quad (1)$$

Multivariate Cox regression analysis was used to calculate the survival time of ccRCC patients with immune infiltration. We then performed subgroup immunotherapy response calculations based on the Tumor Immune Dysfunction and Exclusion (TIDE) and Submap algorithms to assess the difference in immunotherapy response by level of immune infiltration.

CT data collection

Plain CT data of renal cancer patients from TCIA were initially registered for a total of 268 samples. Exclusion criteria for study samples were [1] no TCGA transcriptome data [2], low quality TCIA CT data, and [3] preoperative plain CT of non-renal cancer patients. After screening, we selected cases co-existing in the TCIA database and the TCGA database for the next radiomics study, and finally screened 163 cases.

Tumor segmentation and pyradiomic feature extraction

Image voxels were first resampled to $1 \times 1 \times 1$ mm^3, followed by normalisation of image signal intensity and z-score normalisation of grey scale (The layer thickness of the CT images is all 5 mm). Regions of interest (ROIs) were selected from tumor regions in CT images of ccRCC patients, and segmentation of ROIs was done independently by an experienced urologist and reviewed by another physician. They were not aware of the clinical and pathologic information of the patients. One of the urologists (urologist A, with 9 years of experience) manually plotted the ROIs on the CT images using 3D Slicer 4.11 software, and the ROIs in all cases were first done independently by this physician, which reduced the errors generated by plotting by different physicians. Another urologic surgeon (Urologist B, with 12 years of experience) reviewed the ROIs manually plotted by Physician A. The ROIs were reviewed by another urologic surgeon (Urologist B, with 12 years of experience). (Fig. 1). We then used the pyradiomicv3.0 toolkit in Python 3. 4 to perform feature extraction on the acquired ROIs, which includes First-order features, Shape features

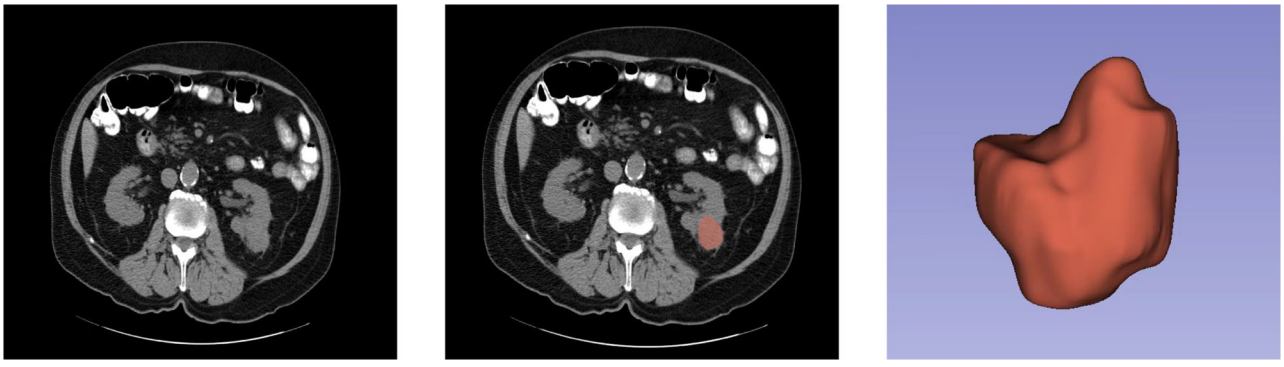

Fig. 1 The ROIs were manually drawn layer by layer in ccRCC plain CT images using 3D Slicer 4.11 software

and texture features. Shape features describe the geometric properties of ROI based on the shape, and texture features describe the second-order and higher-order spatial distribution of the pattern [25]. The yaml files used to extract the features are described in the attached material.

Normalisation of eigenvalues using Z-score. The features were filtered using Pearson's correlation coefficient ($p < 0.05$) and the features were filtered based on the correlation coefficient, with one of the two features having a correlation coefficient > 0.9. In order to select the model with the most accurate prediction and the best generalization ability, the dataset was divided into a training set and a test set, each accounting for 50%.

Predictive model construction based on radiomics

To investigate the correlation between differences in radiomics and immune infiltration and disease outcomes and immunotherapy, LASSO regression analysis was used to construct a model, and the selected features were further selected (coefficient > 0). Based on the selected tumor radiomics features as the independent variable and immune cell infiltration level as the dependent variable. A total of eight machine learning algorithms including Logistic Regression (LR), Support Vector Machine (SVM), k-Nearest Neighbour (KNN), Random Forest (RF), ExtraTrees, XGBoost, LightGBM and Multilayer Perceptron (MLP) were used to select features with high classification performance on the test set to perform diagnostic analyses. Cross-validation was then performed, each time using 80% training and a random 20% for testing, to find the best model, and the corresponding best division of the data, so as to select the one with a high level of classification performance. To evaluate the diagnostic performance and clinical utility of the prediction model, we drew the receiver operating characteristic (ROC) curve of each model, calculated and compared the area under the curve (AUC), and drew the decision curve analysis (DCA). (Fig. 2 shows the flowchart of the study).

Statistical analyses

Comparison between groups was done by Kruskal-Wallis test Non-parametric test Survival analysis was done by Kaplan-Meier test p-value correction was done by Bonferroni test.

Results

Fitting score of immune infiltration level and sample clustering

The results of immune cell infiltration of 539 samples obtained based on the SINGLE SAMPLE GSEA algorithm as well as the results of clustering analysis are shown in a heat map. The results show that the immune infiltration level of C1 clusters is low, while the immune

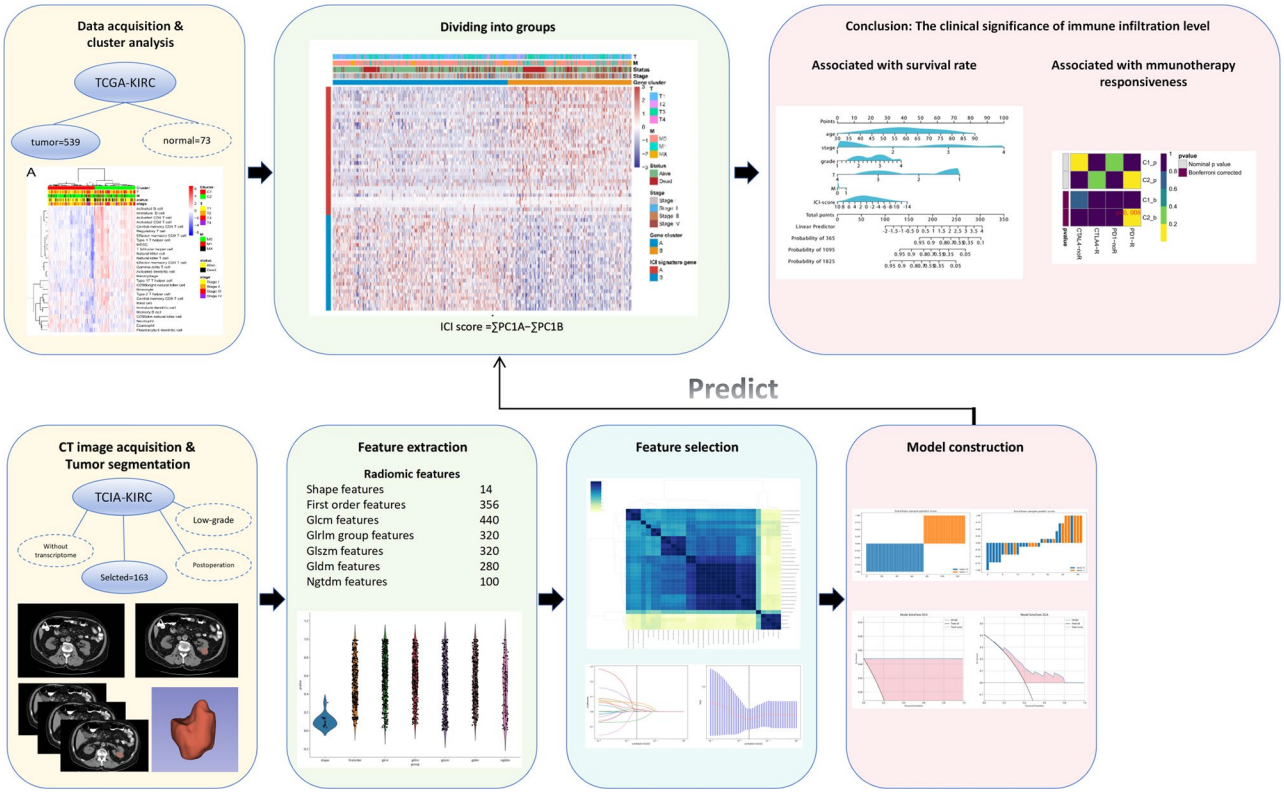

Fig. 2 Flow chart of the study

infiltration level of C2 clusters is high (Fig. 3A). Based on the quantitative differences of immune infiltration in the samples, we calculated the correlation coefficients of the co-expression relationship between different immune cells, as well as the P-value of the risk ratio and prognostic correlation of these cells, while clustering these immune cells according to their co-expression relationship (Fig. 3B). The results showed that some immune cells were strongly correlated with the prognosis of ccRCC patients. The immune cells are mainly activated dendritic cell, activated CD4 T cell, Neutrophil, macrophage, mast cell and other cells directly involved in the immune process. However, some auxiliary and regulatory immune cells did not have a strong correlation. The former is mainly distributed in Cell cluster-B, while the latter is mostly distributed in Cell cluster A/C.

Clustering of samples by degree of immune cell infiltration

However, the clustering results we obtained for the two cohorts based on this approach still did not reflect the large variability in survival. Therefore, we decided to further improve the cohort clustering method. Clustering the cohort based on the gene expression levels of the matched immune cells is a feasible way. By principal component analysis, we distinguished between genes with positive and negative correlation to the survival of the samples, and Boruta's algorithm was used to filter these genes and calculate ICI score values for each sample based on Eq. 1 (Fig. 4A, B). After these analyses, we obtained more prognostically differentiated subgroups, as well as an ICI score with a stronger prognostic correlation that could demonstrate the differences of immune infiltration in the samples (Fig. 4C-E).

Based on the 29 immune cell infiltration scores obtained from the ssGSEA analysis, clustering analysis was used to group the samples (R package: ClassDiscovery, clustering method: ward.D). Subsequently, based on the expression levels of more than 700 genes used to adjust the immune cells, we used unsupervised consensus clustering to classify these genes into sets of positively/negatively correlated genes. The obtained positive/negative gene set was further downscaled using Boruta's algorithm and based on Eq. 1. Subsequently, to construct an imaging histological model for prediction, we grouped the samples according to the degree of immune cell infiltration. We calculated the ICI score value of all samples ($n = 539$), took the median as the boundary, set the samples whose ICI score value was less than the median as group B, and the remaining samples as group A, and integrated the group with the cases in the TCIA database. nd finally we classified 163 cases as label 0 (95 cases) and label 1 (68 cases).Among them, label 1 was the group with higher level of immune infiltration, which was positive for immune infiltration.

Level of immune infiltration associated with survival in renal cancer patients

Multifactorial Cox regression analysis was then used to calculate response to immunotherapy based on TIDE and subgroup submap algorithms to calculate the prognostic value of immune infiltration levels and differences in response to immunotherapy in BRCA patients. The results of multifactorial COX regression showed a significant effect ($p < 0.05$) on the survival time of ccRCC patients based on ICI.score as well as a variety of clinical factors including age, grade, stage, T and M. The

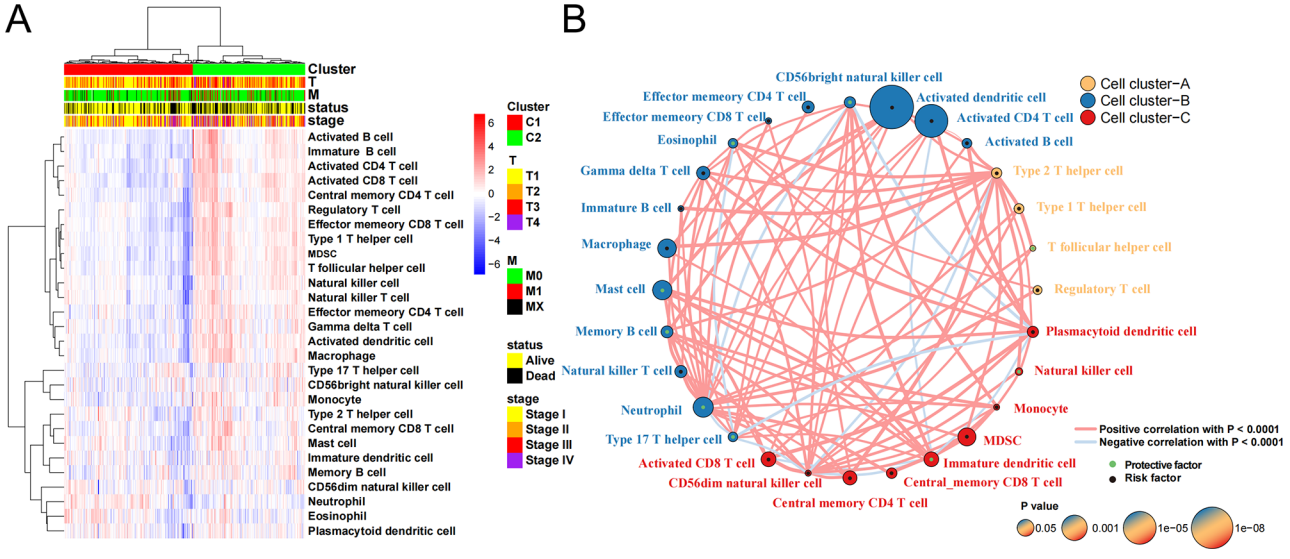

Fig. 3 Fitting scores of immune infiltration levels and clustering of samples. (**A**) Results of immune cell infiltration. The top panel of the heat map shows the clinical data of the samples, and the inner panel shows the infiltration of each immune cell; (**B**) Communication between immune cells

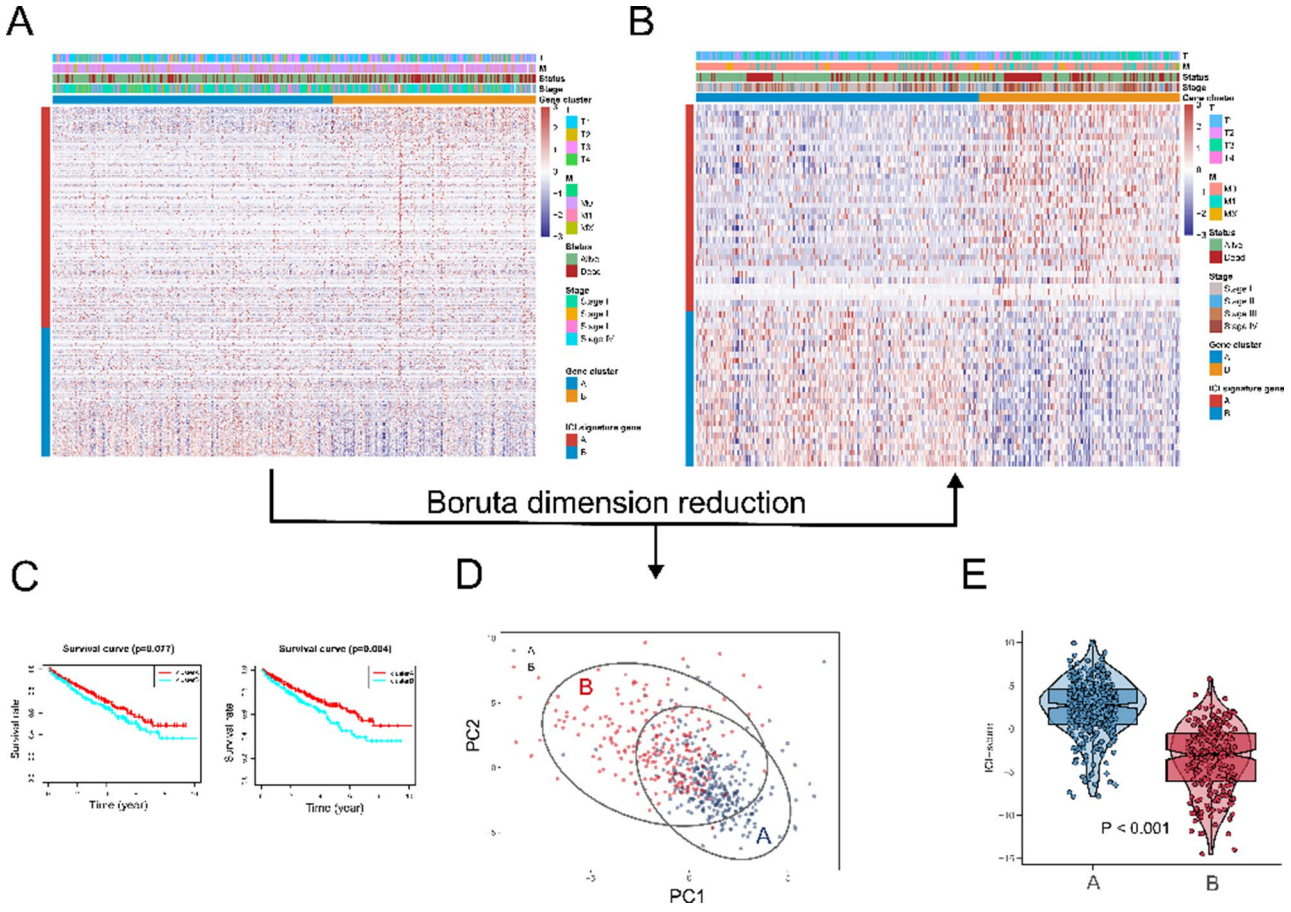

Fig. 4 Cluster analysis of the cohort based on gene expression levels of the fitted immune cells. (**A** is before fitting, **B** is after fitting); **C**. Survival curves of the two clusters before and after fitting; **D**. Dimensionality reduction based on PCA principal component analysis; **E**. Violin plots showing the ICI score values of the samples in the two clusters

clinical factor HR > 1 was positively correlated with the final death status of the patients, and the ICI.score was negatively correlated (see Fig. 5A). We then evaluated the predictive power of ICI.score and clinical information using ROC (Fig. 5C and D), which showed that the AUC of 1,3,5-year survival for a single ICI.score was 0.71, 0.62 and 0.58, and the AUC of 1,3,5-year survival of ICI.score combined with clinical factors were 0.84, 0.74 and 0.78, suggesting that the multifactorial of ICI.score combined with clinical factors better predicts the survival time of renal cancer patients. Finally, we used Nomogram to present the multifactorial Cox regression results visually and effectively (Fig. 5B).

Level of immune infiltration correlates with response to immunotherapy in patients with ccRCC

To further investigate the likelihood of our subgroups responding to immunotherapy, after correcting for the p-value, we performed immunotherapy responsiveness calculations based on the TIDE and submap algorithms for label0 and label1, and the results showed (Fig. 5E) that label1 was more promising for responding to anti-PD-1 therapy, reinforcing the idea that our grouping of samples based on the level of immune cell infiltration has clinical and prognostic significance.

Predicting the degree of immune infiltration by constructing radiomics model

We extracted 1832 features using pyradiomicv3.0 program in Python 3.8 (Fig. 6), and used Pearson's correlation coefficient for feature screening to find 31 relevant features, and the correlation coefficient between features is shown in Fig. 6, and filtered the features according to the correlation coefficient to get 18 features, and we used LASSO to select 7 highly weighted features, as shown in Fig. 7. Among these 7 features, we found 3 Gy Level Size Zone Matrix (GLSZM) feathers, 2 Gy Level Co-occurence Matrix (GLCM) features and 2 (Neighbouring Gray Tone Difference Matrix) NGTDM features. These features are second-order features, which describe the second-and higher-order spatial distribution of patterns or intensities, and involve the description of intensity distributions and the spatial relationship between various intensity levels [26]. These characteristics reflect the

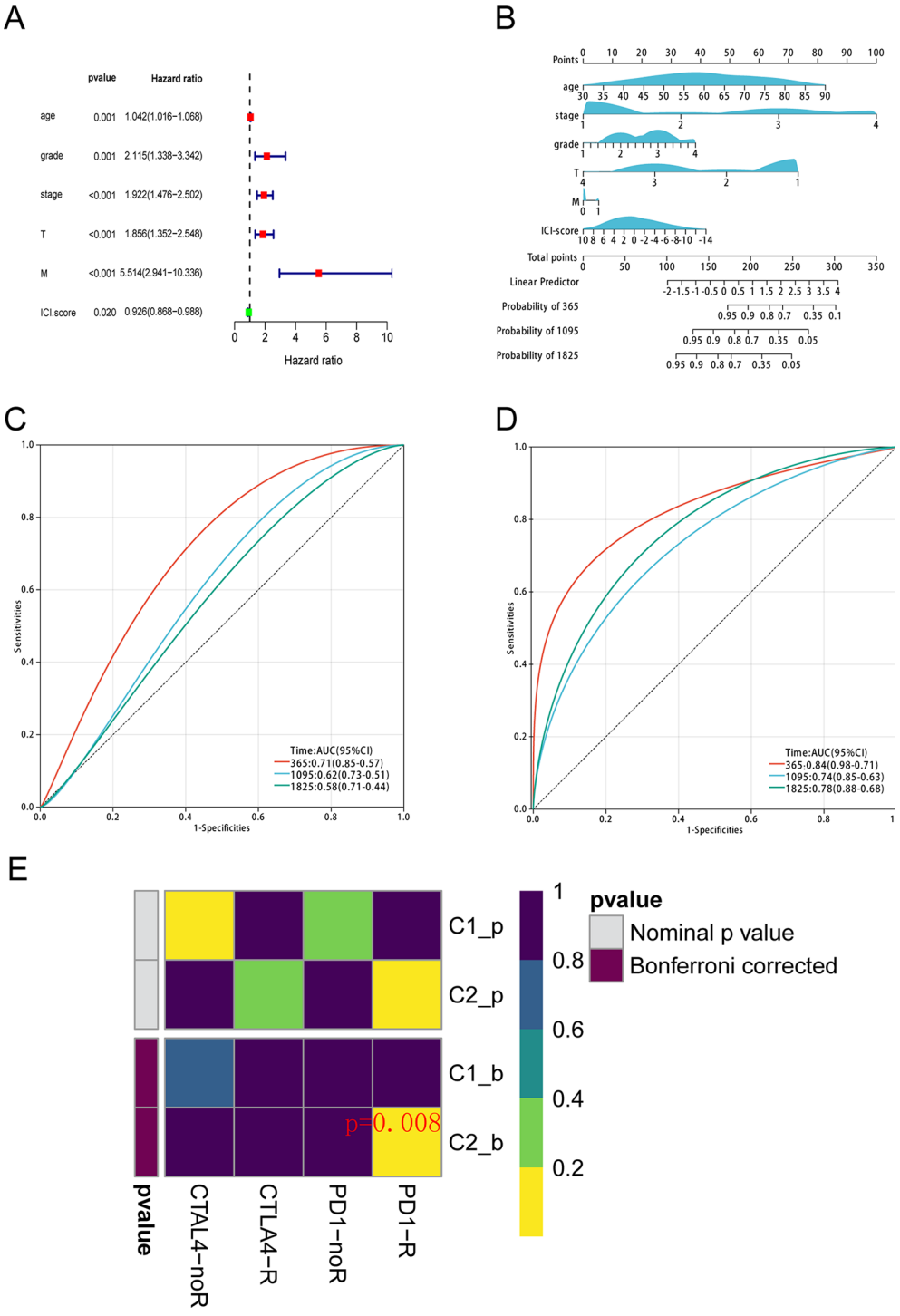

Fig. 5 Multivariate Cox regression analysis calculated the survival time of ccRCC patients with immune infiltration. (**A**) Hazard ratio of ICI score as well as multiple clinical factors on the survival time of patients with renal cancer; (**B**) Prediction of patient's survival probability based on ICI. score and multiple clinical factors including age, grade, stage, T, M) survival probability at 1,3,5 years; (**C**) ROC curve for single ICI-score predicting patient survival; (**D**) ROC curve of ICI combined with clinical information for predicting survival; (**E**) Correlation of our subgroups with immunotherapies

biological characteristics such as intra-tumor and inter-tumor heterogeneity, and may be related to intratumoral immune infiltration.

We used 8 machine learning algorithms to build radiomics models and used ROC curves to assess the performance of the radiomics models in predicting immune infiltration levels. Subsequently, cross-validation

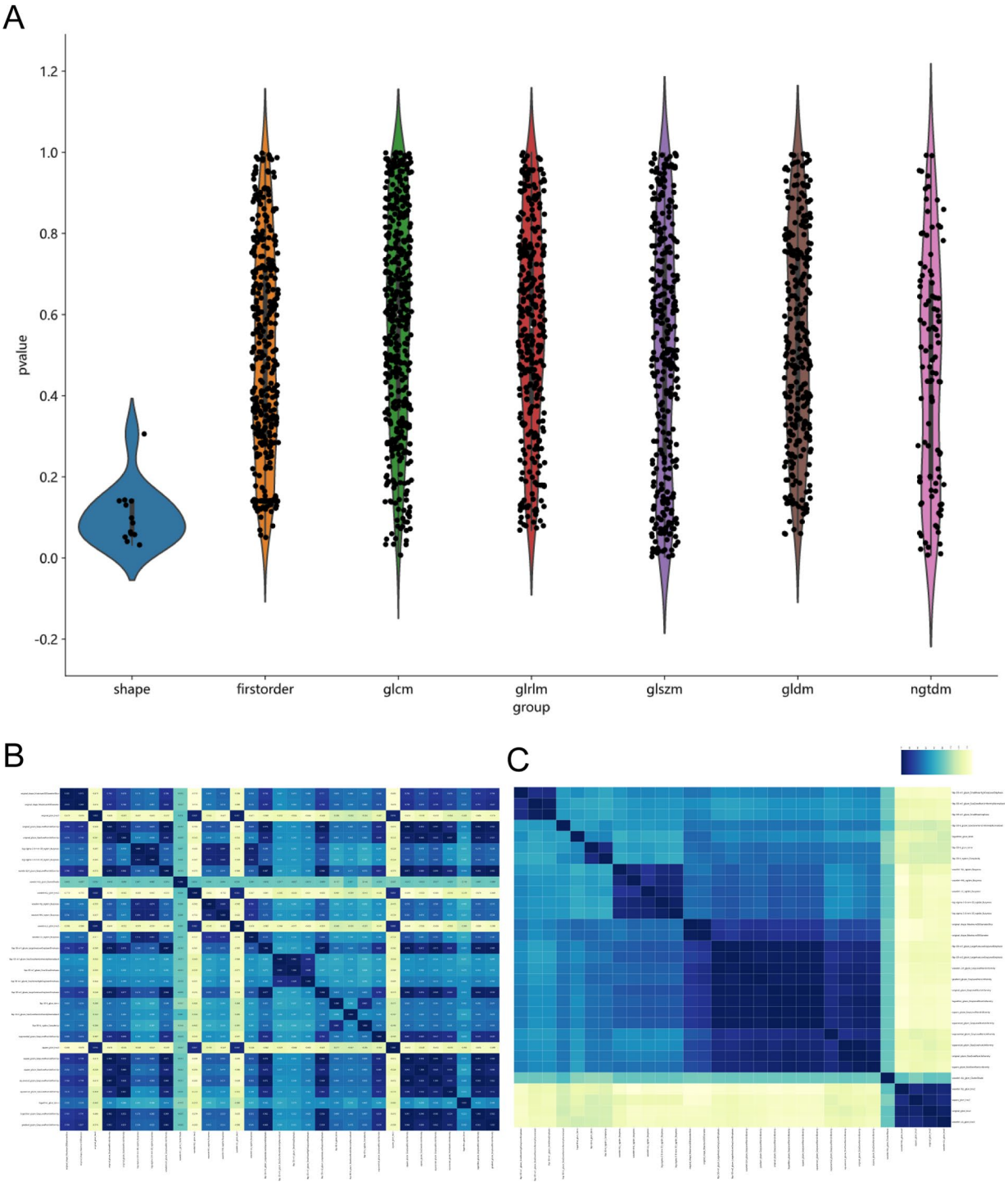

Fig. 6 (**A**) Distribution of features after feature extraction. A total of 1820 features, including 14 Shape features, 356 First order features, 440 GLCM features,320 GLRLM features,320 GLSZM features, 280 GLDM features and 100 NGTDM features; (**B**) Correlation coefficients between 31 relevant features after feature screening using Pearson's correlation coefficient; (**C**) Cluster analysis of 31 features

was performed to find the best primary division method and saved in the results. Among the eight machine learning classifiers, ExtraTrees produced the highest test and training set ROC AUCs of 0.753 and 1.000 (Figs. 8 and 9A). Moreover, Fig. 9B-E shows the confusion matrix predicted by the ExtraTrees model in the test set (9B), the weights of the 7 features (9 C), and the histogram of sample prediction of the ExtraTrees model in the training set (9D) and the test set (9E), respectively. LR and LightGBM had the highest sensitivity in the test set of 0. 615; and LR,

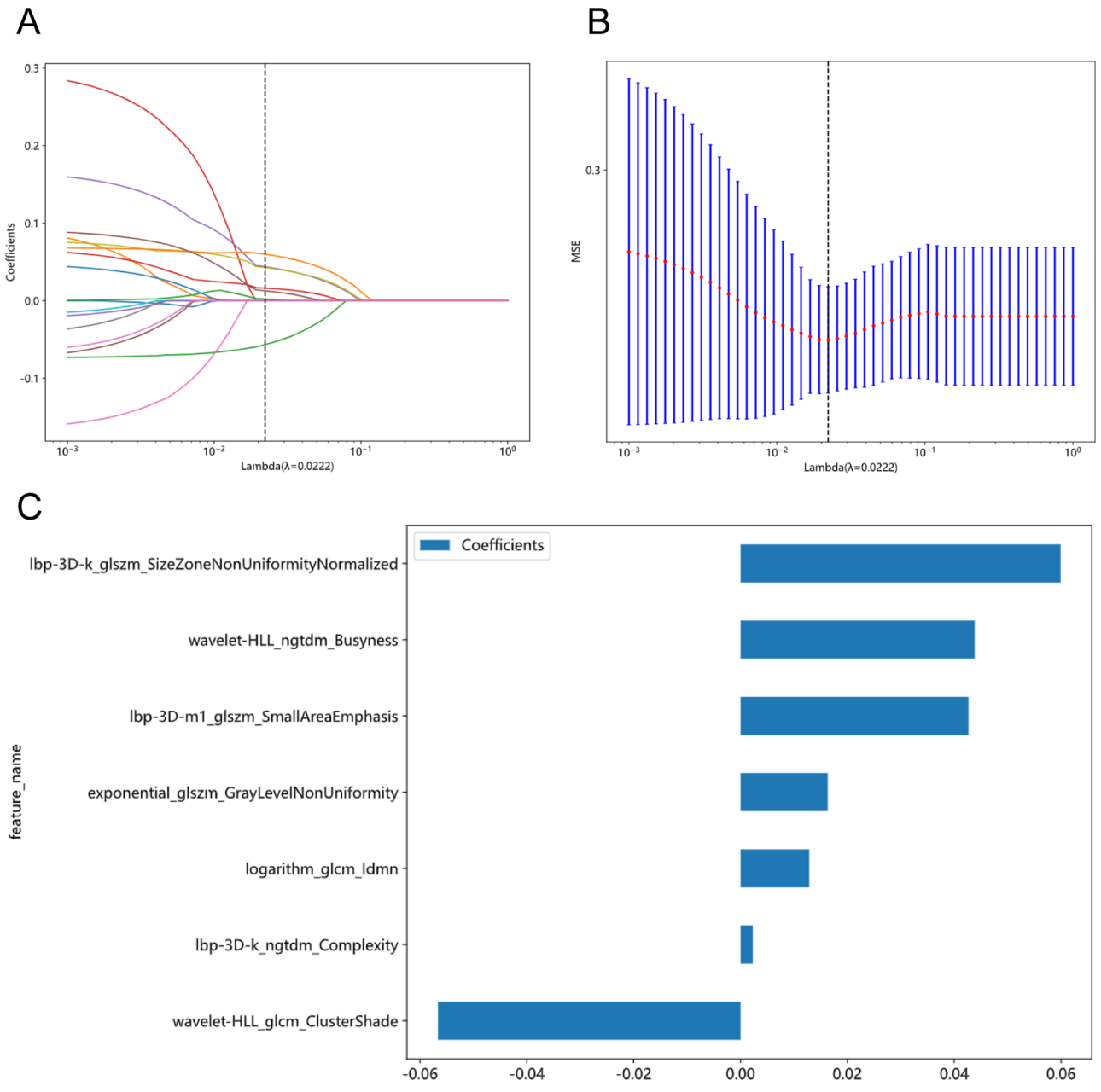

Fig. 7 **A**, **B**: 7 highly weighted features selected for building our model using the LASSO method; **C**. Weights occupied by each feature

SVM, ExtraTrees, LightGBM and MLP had higher specificities of 0.789, 1.000, 0.842, 0.789 and 0.789 respectively, and LR, ExtraTrees and LightGBM had the highest accuracy of 0.719, 0.688 and 0.719 respectively (Table 1). From the results, it can be seen that the ExtraTrees machine learning algorithm has a better overall performance than the other algorithms in predicting the level of immune infiltration. Meanwhile, decision curve analysis (DCA) was performed to assess the clinical utility of the prediction model. The results are shown in Fig. 9 (F, G), where the ExtraTrees model provided a high net benefit (10-80%). In summary, ExtraTrees demonstrated comprehensive and strong predictive ability.

Discussion

In the present study, we developed and validated CT-based radiomics models for predicting the level of immune infiltration in RCC, and thus the prognostic survival of patients as well as their responsiveness to ICI therapy, and compared their performance. Our study demonstrated that ct-based multi-modelling achieved good results in predicting the prognostic

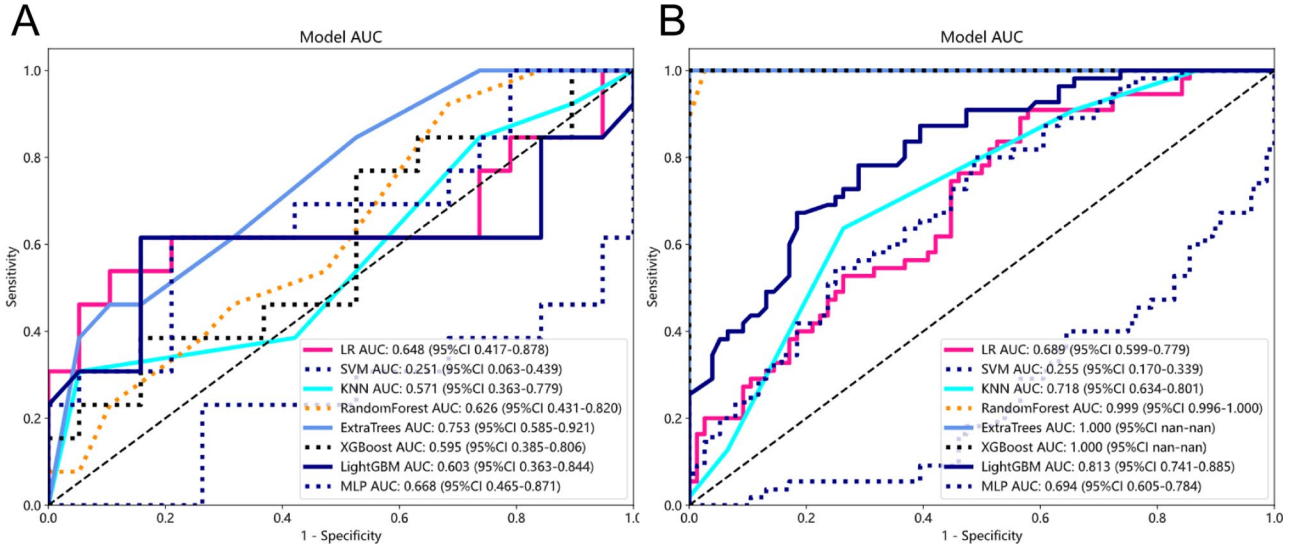

Fig. 8 (**A**) ROC plots of 8 models for predicting immune infiltration subgroups in the test set; (**B**) ROC plots of 8 models for predicting immune infiltration subgroups in the training set

and treatment-related levels of immune infiltration in the immune microenvironment of clear cell renal cell carcinoma.

With the gradual development of radiomics and deep learning, more and more researchers are committed to predicting certain properties of Rcc, such as pathological grading, tumour heterogeneity, immune microenvironment, etc., in a non-invasive manner, and ultimately reap the benefits in terms of patients' survival [27–29] et al. reported that radiomics model created by them was used to predict the prognosis of patients with clear cell renal cell carcinoma by reflecting tumour heterogeneity and microenvironment. prognosis, and the AUCs for 1, 3, and 5 year-OS of this model reached more than 0.83 [29]. Wang et al.'s model also had good predictive performance in predicting pathological grading and prognosis of ccRCC [28]. The model of Fan et al. predicts the extent of macrophage infiltration in gliomas [30]. All of the above mentioned studies provide effective aids for clinical decision making for ccRCC patients in a non-invasive situation. In addition, patients may be more interested in knowing about the effect of drug treatment, so we constructed a multi-model combining radiomics, deep learning, and genomic features to predict the prognostic and treatment-related levels of immune infiltration in the immune microenvironment of ccRCC. Compared with other radiomics models, our experiment quantifies the immune infiltration by making it possible to predict the survival and responsiveness to ICI treatment of renal cancer patients by predicting the degree of immune infiltration. It provides new ideas related to the prognosis of renal cancer patients, and can also determine the efficacy of ICI drugs to a certain extent, and better help patients to decide their treatment. It is a supplement and extension of the previous radiomics model, and, as the study of immune infiltration and tumour continues to deepen, we will have more evidence and direction to improve this model in the future.

In our study, we used plain CT rather than enhanced CT because plain CT is more universal and economical for patients and the use of enhanced CT is limited in patients with impaired renal function [31]. A procedure may be more necessary to assist the physician in diagnosis with a plain CT than with enhanced CT images. For those patients who have only had a plain CT, this model may also provide a means of quickly understanding the condition. Fewer studies have focused on radiomics based on scanning CT for ccRCC, so this study is more valuable in terms of clinical utility and economy. Regarding ROI outlining, at the beginning of the study, we tried several automated or semi-automated means, including AI-assisted and completed procedures, but after our review, the results at junctions (e.g. between tumour and normal tissue) were not as good as we would have liked, which would likely affect the accuracy of feature extraction and machine learning, so after weighing up the situation, we decided to still use manual outlining of ROIs, the and reduce the error by reviewing it. We selected a total of eight machine learning models. LR, SVM, KNN, ExtraTrees, and MLP are single models, which are trained and validated independently of each other, and RF, XGBoost, and LightGBM are integrated models, which combine multiple single models to form a single model that takes the best of all the single models and achieves a relative best performance. These algorithms are supervised models. From the results, it can be seen that ExtraTrees, a machine learning algorithm has a better overall performance than other algorithms.

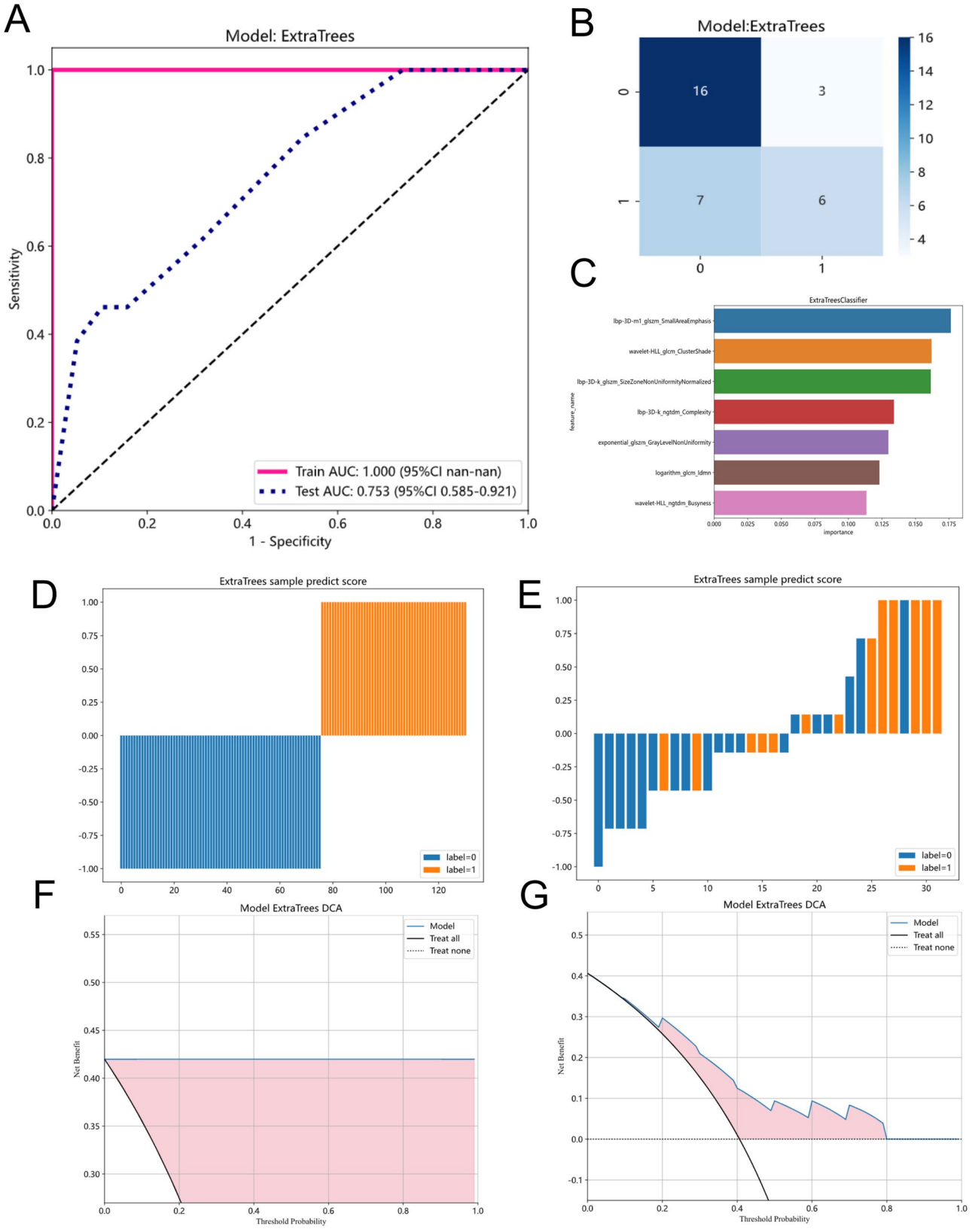

Fig. 9 (**A**) ROC plots of the predictions of the ExtraTrees models in the training set and the test set; (**B**) Confusion matrix of the predictions of the ExtraTrees models in the test set; (**C**) Weights assigned to the 7 features of the ExtraTrees models; D, E: Histograms of the predictions of the samples in the training set (**D**) and the test set (**E**); **F**, **G**: DCA evaluation of the clinical usefulness of the ExtraTrees prediction model

Table 1 Prediction results of the eight machine learning models in the test set. PPV: positive prediction value; NPV: negative predictive value; F1: F1 score

Model name	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Precision	Recall	F1	Threshold
LR	0.719	0.648	0.4175–0.8781	0.615	0.789	0.667	0.75	0.667	0.615	0.64	0.537
SVM	0.594	0.251	0.0628–0.4392	0	1	0	0.594	0	0	NaN	0.285
KNN	0.5	0.571	0.3627–0.7790	0.385	0.579	0.385	0.579	0.385	0.385	0.385	0.8
RF	0.594	0.626	0.4308–0.8202	0.385	0.737	0.5	0.636	0.5	0.385	0.435	0.3
ExtraTrees	0.688	0.753	0.5850–0.9210	0.462	0.842	0.667	0.696	0.667	0.462	0.545	0.7
XGBoost	0.531	0.595	0.3845–0.8058	0.462	0.579	0.429	0.611	0.429	0.462	0.444	0.263
LightGBM	0.719	0.603	0.3626–0.8438	0.615	0.789	0.667	0.75	0.667	0.615	0.64	0.521
MLP	0.625	0.668	0.4650–0.8710	0.385	0.789	0.556	0.652	0.556	0.385	0.455	0.485

There are more types of immune cells, and their distribution and heterogeneity are large and vary significantly among different tumors [32]. It is well known that RCC is considered one of the solid tumors with the greatest degree of immune infiltration [11, 33]. Studies have found that ccRCC has a high immune infiltration score as well as a t cell infiltration score [34]. However, more advanced tumors may exhibit a stronger immunosuppressive phenotype [35]. This suggests that immune cells in the tumor microenvironment are dynamic and that the level of immune infiltration may be a factor in clinical prognosis [36]. In this study, we could find a strong correlation between some immune cells and prognosis of ccRCC patients. In order to reflect the stronger survival differentiation, we distinguished the genes with positive and negative correlation on the survival of the samples by principal component analysis, and after the above analysis we obtained more prognostically differentiated subgroups, as well as the ICIscore demonstrating the level of immune cell infiltration of the samples.Subsequently, in order to evaluate the prognostic value of the immune cell infiltration level of the patients with BRCA, we used used a multifactorial Cox regression using analysis, and the results proved our conjecture that the ICI.score combined with the clinical factors of the multifactorial more well predicted the kidney cancer patients' survival time.

ICI is crucial in the treatment of RCC, in which PD-1/PD-L1 is the major pathway and plays an essential role in tumor immunity tumor progression, the binding of PD-1 and PD-L1 leads to tumor immune escape, and PD-L1 is overexpressed in about 30% of ccRCC [37]. On the basis of this mechanism, a variety of types of anti-pd -1/PD-L1 antibodies have been used for the treatment of ccRCC [38–40]. In addition to predicting patient prognosis and survival status, immune infiltration of tumors is closely related to ICI therapeutic responsiveness. It has been found that tumor cells infiltrated with high immune cells (e.g., CD8 + T vs. MHCII cells, etc.) are more susceptible to receive benefit from ICI therapy [41, 42]. In RCC, Braun et al. noted that infiltrated tumors depleted of favorable PBRM1 mutations and enriched for unfavorable chromosomal losses of 9p21.3. These chromosomal alterations associated with response or resistance to PD-1 blockade [43]. In this experiment, we performed immunotherapy responsiveness calculations based on TIDE and submap algorithms, and the outcome surfaced that the highly immune infiltrated group was more promising to respond to antiPD-1 therapy, which makes our sample grouping clinically and prognostically significant.

In the process of building the radiomics model, we first selected 18 features using the Pearson correlation coefficient feature screening method. The Pearson correlation coefficient mainly measures the linear correlation between the features and the target variables, which is quick and easy to understand [44], but it is only sensitive to the linear relationship, and if the relationship between the variables is non-linear, it may not accurately reflect the relationship between the variables. We then used lasso regression to select seven highly weighted features that prevent overfitting and are able to handle high-dimensional data. We used multiple machine learning methods to compare and plot DCA to assess the clinical utility of prediction models. Metrics such as sensitivity, specificity and area under the ROC curve measure the predictive model's diagnostic accuracy, whereas the advantage of DCA is that it integrates patient or decision maker preferences into the analysis which is in line with the current trend of personalized medicine. This new approach provides insights into clinical outcomes based on the probability of thresholds, from which there can be a net gain [45–47].

Despite our novel findings, this study has some limitations. Due to the strict inclusion and exclusion criteria, the overall small size of the dataset in this study and the relatively small number of cases in the label 1 group (68 cases) compared to the number of cases in the label 0 group (95 cases) may lead to the presence of selection bias as well as underfitting of the results; therefore, future studies should have larger sample sizes, multi-vendor images, and external test sets [48]. It is worth mentioning that among all the models, the SVM model shows a less favourable AUC (0.251), we speculate that the reason is due to the mismatch of the test set caused by overfitting, and the second reason is due to the impact of the

unbalanced number of samples in the two data sets. At the same time, this study lacks an external validation set to further validate the prediction results of the model. However, as mentioned above, we verified the diagnostic performance of the model by drawing ROC curves and other methods. In future studies, external validation sets need to be considered for inclusion in relevant studies to strengthen the reliability of the explanatory results. And considering the generality of application, we chose plain CT as the image source for the study of imaging genomics, which is easy to operate and less expensive compared to enhanced CT, but for diagnostic accuracy, it is not as good as enhanced CT. Therefore, in the future, we can apply based on enhanced CT combined with genomics to predict the relevant information. Due to the constraints of the TCIA database, the number of samples that aligned with the bioinformatics analysis might not fulfill the requirements for sample size calculation. Nonetheless, in our machine learning hyperparameter optimization process, we managed to maintain the highest possible model accuracy by appropriately reducing the sample size. We are also cognizant of our limitations, as the integration of genomics/transcriptomics with radiomics research necessitates the preservation of both the genetic data integrity and the availability of imaging data. The absence of either in these two omics could result in the exclusion of our research samples. Consequently, we must confront the loss of a significant number of samples (from $n = 539$ to $n = 163$) and the potential negative effects on the final predictive model. Such effects may include selection bias leading to bias in effect estimation and waste of resources. It can be solved by more stringent matching design in the future. We also aspire to address these challenges in future research, leveraging larger datasets from more comprehensive platforms to enhance our samples, thereby making our work more scientifically rigorous and open to discussion.

Conclusion

The use of radiomics model based on renal CT images can be noninvasively used to predict the immune infiltration level of ccRCC as well as combined with clinical information to create columnar plots predicting total survival in people with ccRCC and to predict responsiveness to ICI therapy, findings that may be useful in stratifying the prognosis of patients with ccRCC and guiding clinical practitioners to develop individualized regimens in the treatment of their patients.

Abbreviations

ML	Machine learning
CT	Computed tomography
RCC	Renal cell carcinoma
ccRCC	Clear cell renal cell carcinoma
TCGA	The Cancer Genome Atlas
GSEA	Gene Set Enrichment Analysis
TCIA	The Cancer Imaging Archive
ssGSEA	Single Sample Gene Set Enrichment Analysis
TIDE	Tumor Immune Dysfunction and Exclusion
LASSO	Least absolute shrinkage and selection operator
ICI	Immune checkpoint inhibitor
ROI	Regions of interest
LR	Logistic Regression
SVM	Support Vector Machine
KNN	K-Nearest Neighbour
RF	Random Forest
MLP	Multilayer Perceptron
ROC	Receiver operating characteristic
AUC	Area under the curve
DCA	Decision curve analysis
PPV	Positive prediction value
NPV	Negative predictive value

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12880-025-01749-3.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

Acknowledgements

We are grateful to all those who contributed to this manuscript, and to the pioneers of research in related fields. On this basis, we are also grateful to OnekeyAI platform for providing us with computational support for radiomics and training methods for machine learning models.

Author contributions

Wang set the formulation of overarching research goals and aims. Qi conducted data collection and provided analytical tools. Song and Ge conducted data analysis and first draft writing. Wu and Che reviewed the manuscript. All authors contributed to this manuscript.

Funding

JYTMS20230577 (Education Department of Liaoning Province, 50000CNY), 2024ZDJH01PT069 (Dalian Life and Health Guidance Program Project, 100000CNY).

Data availability

The datasets examined in this study are available in the Supplementary Materials or can be obtained by contacting the corresponding author. Experiment design and samples inclusion criteria are provided in the Supplementary Material (CLEAR.docx, METRICS.docx). Relevant codes for image feature extraction are also available in the Supplementary material (exampleCT.yaml.docx, Radiomics code.docx, supplymentary material).

Declarations

Ethical approval

All datasets in this study were downloaded from public databases, including the TCGA (https://www.cancer.gov/ccg/research/genome-sequencing/tcga) and TCIA (https://www.cancerimagingarchive.net/) databases. These public databases allow researchers to download and analyse public datasets for scientific purposes.

Consent for publication

Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Urology, The First Affiliated Hospital of Dalian Medical University, No.222 Zhongshan Road, Dalian, Liaoning 116011, PR China

Received: 27 January 2025 / Accepted: 28 May 2025
Published online: 01 July 2025

References

1. Bukavina L, Bensalah K, Bray F, Carlo M, Challacombe B, Karam JA, Kassouf W, Mitchell T, Montironi R, O'Brien T, et al. Epidemiology of renal cell carcinoma: 2022 update. Eur Urol. 2022;82:529–42. https://doi.org/10.1016/j.eururo.2022.08.019.
2. Miller KD, Goding Sauer A, Ortiz AP, Fedewa SA, Pinheiro PS, Tortolero-Luna G, Martinez-Tyson D, Jemal A, Siegel RL. Cancer statistics for Hispanics/Latinos, 2018. CA Cancer J Clin. 2018;68:425–45. https://doi.org/10.3322/caac.21494.
3. Ferlay J, Colombet M, Soerjomataram I, Dyba T, Randi G, Bettio M, Gavin A, Visser O, Bray F. Cancer incidence and mortality patterns in Europe: estimates for 40 countries and 25 major cancers in 2018. Eur J Cancer. 2018;103:356–87. https://doi.org/10.1016/j.ejca.2018.07.005.
4. Arai T, Sazuka T, Sato H, Kato M, Kamada S, Katsura S, Seito A, Miyamoto S, Wakai K, Takeuchi N, et al. A clinical investigation of recurrence and lost follow-up after renal cell carcinoma surgery: a single-center, long-term, large cohort, retrospective study. Int J Clin Oncol. 2022;27:1467–76. https://doi.org/10.1007/s10147-022-02204-x.
5. Marconi L, Dabestani S, Lam TB, Hofmann F, Stewart F, Norrie J, Bex A, Bensalah K, Canfield SE, Hora M, et al. Systematic review and Meta-analysis of diagnostic accuracy of percutaneous renal tumour biopsy. Eur Urol. 2016;69:660–73. https://doi.org/10.1016/j.eururo.2015.07.072.
6. Lee WS, Yang H, Chon HJ, Kim C. Combination of anti-angiogenic therapy and immune checkpoint Blockade normalizes vascular-immune crosstalk to potentiate cancer immunity. Exp Mol Med. 2020;52:1475–85. https://doi.org/10.1038/s12276-020-00500-y.
7. Ge W, Song S, Qi X, Chen F, Che X, Sun Y, Wang J, Li X, Liu N, Wang Q, et al. Review and prospect of immune checkpoint Blockade therapy represented by PD-1/PD-L1 in the treatment of clear cell renal cell carcinoma. Oncol Res. 2023;31:255–70. https://doi.org/10.32604/or.2023.027942.
8. Becht E, Giraldo NA, Germain C, de Reyniès A, Laurent-Puig P, Zucman-Rossi J, Dieu-Nosjean M-C, Sautès-Fridman C, Fridman WH. Immune contexture, immunoscore, and malignant cell molecular subgroups for prognostic and theranostic classifications of cancers. Adv Immunol. 2016;130:95–190. https://doi.org/10.1016/bs.ai.2015.12.002.
9. Liu Y, Cao X. Characteristics and significance of the Pre-metastatic niche. Cancer Cell. 2016;30:668–81. https://doi.org/10.1016/j.ccell.2016.09.011.
10. Yoshida T, Nakamoto T, Atsumi N, Ohe C, Sano T, Yasukochi Y, Tsuta K, Kinoshita H. Impact of LAG-3/FGL1 pathway on immune evasive contexture and clinical outcomes in advanced urothelial carcinoma. J Immunother Cancer. 2024;12:e009358. https://doi.org/10.1136/jitc-2024-009358.
11. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautès-Fridman C, Fridman WH, et al. Erratum to: estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. 2016;17:249. https://doi.org/10.1186/s13059-016-1113-y.
12. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell. 2015;160:48–61. https://doi.org/10.1016/j.cell.2014.12.033.
13. Wu Z, Zhou J, Xiao Y, Ming J, Zhou J, Dong F, Zhou X, Xu Z, Zhao X, Lei P, et al. CD20 + CD22 + ADAM28 + B cells in tertiary lymphoid structures promote immunotherapy response. Front Immunol. 2022;13:865596. https://doi.org/10.3389/fimmu.2022.865596.
14. Gardin I, Grégoire V, Gibon D, Kirisli H, Pasquier D, Thariat J, Vera P. Radiomics: principles and radiotherapy applications. Crit Rev Oncol Hematol. 2019;138:44–50. https://doi.org/10.1016/j.critrevonc.2019.03.015.
15. Shao J, Wang C, Shu K, Zhou Y, Cheng N, Lai Z, Li K, Xu L, Chen J, Du F, et al. A contrast-enhanced CT-based radiomic nomogram for the differential diagnosis of intravenous leiomyomatosis and uterine leiomyoma. Front Oncol. 2023;13:1239124. https://doi.org/10.3389/fonc.2023.1239124.
16. Khodabakhshi Z, Amini M, Mostafaei S, Haddadi Avval A, Nazari M, Oveisi M, Shiri I, Zaidi H. Overall survival prediction in renal cell carcinoma patients using computed tomography radiomic and clinical information. J Digit Imaging. 2021;34:1086–98. https://doi.org/10.1007/s10278-021-00500-y.
17. Shayesteh S, Nazari M, Salahshour A, Sandoughdaran S, Hajianfar G, Khateri M, Yaghobi Joybari A, Jozian F, Fatehi Feyzabad SH, Arabi H, et al. Treatment response prediction using MRI-based pre-, post-, and delta-radiomic features and machine learning algorithms in colorectal cancer. Med Phys. 2021;48:3691–701. https://doi.org/10.1002/mp.14896.
18. Arabi H, AkhavanAllaf A, Sanaat A, Shiri I, Zaidi H. The promise of artificial intelligence and deep learning in PET and SPECT imaging. Phys Med. 2021;83:122–37. https://doi.org/10.1016/j.ejmp.2021.03.008.
19. Wu J, Li B, Sun X, Cao G, Rubin DL, Napel S, Ikeda DM, Kurian AW, Li R. Heterogeneous enhancement patterns of Tumor-adjacent parenchyma at MR imaging are associated with dysregulated signaling pathways and poor survival in breast Cancer. Radiology. 2017;285:401–13. https://doi.org/10.1148/radiol.2017162823.
20. Fan M, Xia P, Liu B, Zhang L, Wang Y, Gao X, Li L. Tumour heterogeneity revealed by unsupervised decomposition of dynamic contrast-enhanced magnetic resonance imaging is associated with underlying gene expression patterns and poor survival in breast cancer patients. Breast Cancer Res. 2019;21:112. https://doi.org/10.1186/s13058-019-1199-8.
21. Arefan D, Hausler RM, Sumkin JH, Sun M, Wu S. Predicting cell invasion in breast tumor microenvironment from radiological imaging phenotypes. BMC Cancer. 2021;21:370. https://doi.org/10.1186/s12885-021-08122-x.
22. He H, Jin Z, Dai J, Wang H, Sun J, Xu D. Computed tomography-based radiomics prediction of CTLA4 expression and prognosis in clear cell renal cell carcinoma. Cancer Med. 2023;12:7627–38. https://doi.org/10.1002/cam4.5449.
23. Jiang W, Wu R, Yang T, Yu S, Xing W. Profiling regulatory T lymphocytes within the tumor microenvironment of breast cancer via radiomics. Cancer Med. 2023;12:21861–72. https://doi.org/10.1002/cam4.6757.
24. Qian H, Ren X, Xu M, Fang Z, Zhang R, Bu Y, Zhou C. Magnetic resonance imaging-based radiomics was used to evaluate the level of prognosis-related immune cell infiltration in breast cancer tumor microenvironment. BMC Med Imaging. 2024;24:31. https://doi.org/10.1186/s12880-024-01212-9.
25. Kayi Cangir A, Orhan K, Kahya Y, Özakıncı H, Kazak BB, Konuk Balcı BM, Karasoy D, Uzun Ç. CT imaging-based machine learning model: a potential modality for predicting low-risk and high-risk groups of thymoma: impact of surgical modality choice. World J Surg Oncol. 2021;19:147. https://doi.org/10.1186/s12957-021-02259-6.
26. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, Zegers CML, Gillies R, Boellard R, Dekker A, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48:441–6. https://doi.org/10.1016/j.ejca.2011.11.036.
27. Şenbabaoğlu Y, Gejman RS, Winer AG, Liu M, Van Allen EM, de Velasco G, Miao D, Ostrovnaya I, Drill E, Luna A, et al. Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. Genome Biol. 2016;17:231. https://doi.org/10.1186/s13059-016-1092-z.
28. Wang S, Zhu C, Jin Y, Yu H, Wu L, Zhang A, Wang B, Zhai J. A multi-model based on radiogenomics and deep learning techniques associated with histological grade and survival in clear cell renal cell carcinoma. Insights Imaging. 2023;14:207. https://doi.org/10.1186/s13244-023-01557-9.
29. Wu J, Li J, Huang B, Dong S, Wu L, Shen X, Zheng Z. Radiomics predicts the prognosis of patients with clear cell renal cell carcinoma by reflecting the tumor heterogeneity and microenvironment. Cancer Imaging. 2024;24:124. https://doi.org/10.1186/s40644-024-00768-7.
30. Fan X, Li J, Huang B, Lu H, Lu C, Pan M, Wang X, Zhang H, You Y, Wang X, et al. Noninvasive radiomics model reveals macrophage infiltration in glioma. Cancer Lett. 2023;573:216380. https://doi.org/10.1016/j.canlet.2023.216380.
31. Sodagari F, Davenport MS, Asch D, Cavallo JJ, Cohan RH, Ellis JH, Pahade JK. A survey of practicing radiologists on the use of premedication before intravenous iodinated contrast medium administration. J Am Coll Radiol. 2024;21:795–804. https://doi.org/10.1016/j.jacr.2023.04.024.
32. Fridman WH, Pagès F, Sautès-Fridman C, Galon J. The immune contexture in human tumours: impact on clinical outcome. Nat Rev Cancer. 2012;12:298–306. https://doi.org/10.1038/nrc3245.
33. Jonasch E, Walker CL, Rathmell WK. Clear cell renal cell carcinoma ontogeny and mechanisms of lethality. Nat Rev Nephrol. 2021;17:245–61. https://doi.org/10.1038/s41581-020-00359-2.

34. Şenbabaoğlu Y, Gejman RS, Winer AG, Liu M, Van Allen EM, de Velasco G, Miao D, Ostrovnaya I, Drill E, Luna A, et al. Erratum to: tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. Genome Biol. 2017;18:46. https://doi.org/10.1186/s13059-017-1180-8.
35. Braun DA, Street K, Burke KP, Cookmeyer DL, Denize T, Pedersen CB, Gohil SH, Schindler N, Pomerance L, Hirsch L, et al. Progressive immune dysfunction with advancing disease stage in renal cell carcinoma. Cancer Cell. 2021;39:632–e6488. https://doi.org/10.1016/j.ccell.2021.02.013.
36. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, Coussens LM, Gabrilovich DI, Ostrand-Rosenberg S, Hedrick CC, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med. 2018;24:541–50. https://doi.org/10.1038/s41591-018-0014-x.
37. Rodriguez-Vida A, Hutson TE, Bellmunt J, Strijbos MH. New treatment options for metastatic renal cell carcinoma. ESMO Open. 2017;2:e000185. https://doi.org/10.1136/esmoopen-2017-000185.
38. Msaouel P, Goswami S, Thall PF, Wang X, Yuan Y, Jonasch E, Gao J, Campbell MT, Shah AY, Corn PG, et al. A phase 1–2 trial of sitravatinib and nivolumab in clear cell renal cell carcinoma following progression on antiangiogenic therapy. Sci Transl Med. 2022;14:eabm6420. https://doi.org/10.1126/scitranslmed.abm6420.
39. Gu L, Peng C, Liang Q, Huang Q, Lv D, Zhao H, Zhang Q, Zhang Y, Zhang P, Li S, et al. Neoadjuvant Toripalimab plus axitinib for clear cell renal cell carcinoma with inferior Vena Cava tumor thrombus: NEOTAX, a phase 2 study. Signal Transduct Target Ther. 2024;9:264. https://doi.org/10.1038/s41392-024-01990-2.
40. Motzer RJ, Tannir NM, McDermott DF, Arén Frontera O, Melichar B, Choueiri TK, Plimack ER, Barthélémy P, Porta C, George S, et al. Nivolumab plus ipilimumab versus Sunitinib in advanced Renal-Cell carcinoma. N Engl J Med. 2018;378:1277–90. https://doi.org/10.1056/NEJMoa1712126.
41. Lopez de Rodas M, Nagineni V, Ravi A, Datar IJ, Mino-Kenudson M, Corredor G, Barrera C, Behlman L, Rimm DL, Herbst RS, et al. Role of tumor infiltrating lymphocytes and Spatial immune heterogeneity in sensitivity to PD-1 axis blockers in non-small cell lung cancer. J Immunother Cancer. 2022;10:e004440. https://doi.org/10.1136/jitc-2021-004440.
42. Carlisle JW, Jansen CS, Cardenas MA, Sobierajska E, Reyes AM, Greenwald R, Del Balzo L, Prokhnevska N, Kucuk O, Carthon BC, et al. Clinical outcome following checkpoint therapy in renal cell carcinoma is associated with a burst of activated CD8 T cells in blood. J Immunother Cancer. 2022;10:e004803. https://doi.org/10.1136/jitc-2022-004803.
43. Braun DA, Hou Y, Bakouny Z, Ficial M, Sant' Angelo M, Forman J, Ross-Macdonald P, Berger AC, Jegede OA, Elagina L, et al. Interplay of somatic alterations and immune infiltration modulates response to PD-1 Blockade in advanced clear cell renal cell carcinoma. Nat Med. 2020;26:909–18. https://doi.org/10.1038/s41591-020-0839-y.
44. Chang S, Li D, Qi Y. Pearson's goodness-of-fit tests for sparse distributions. J Appl Stat. 2023;50:1078–93. https://doi.org/10.1080/02664763.2021.2017413.
45. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than Meets the eye. Lancet Oncol. 2015;16:e173–180. https://doi.org/10.1016/S1470-2045(14)71116-7.
46. Van Calster B, Vickers AJ. Calibration of risk prediction models: impact on decision-analytic performance. Med Decis Mak. 2015;35:162–9. https://doi.org/10.1177/0272989X14547233.
47. Alhulaili ZM, Linnemann RJ, Dascau L, Pleijhuis RG, Klaase JM. A transparent reporting of a multivariable prediction model for individual prognosis or diagnosis analysis to evaluate the quality of reporting of postoperative pancreatic fistula prediction models after pancreatoduodenectomy: A systematic review. Surgery. 2023;174:684–91. https://doi.org/10.1016/j.surg.2023.04.058.
48. Bluemke DA, Moy L, Bredella MA, Ertl-Wagner BB, Fowler KJ, Goh VJ, Halpern EF, Hess CP, Schiebler ML, Weiss CR. Assessing radiology research on artificial intelligence: A brief guide for authors, reviewers, and Readers-From the radiology editorial board. Radiology. 2020;294:487–9. https://doi.org/10.1148/radiol.2019192515.

Publisher's note