Accuracy of CT texture analysis for differentiating low-grade and high-grade renal cell carcinoma: systematic review and meta-analysis

Wei Yu, Gao Liang, Lichuan Zeng, Yang Yang, Yinghua Wu

ABSTRACT

Objectives This study aimed to assess the accuracy of CT texture analysis (CTTA) for differentiating low-grade and high-grade renal cell carcinoma (RCC).

Design Systematic review and meta-analysis.

Data sources PubMed, Cochrane Library, Embase, Web of Science, OVID Medline, Science Direct and Springer were searched to identify the included studies.

Eligibility criteria for including studies Clinical studies that report about the accuracy of CTTA in differentiating low-grade and high-grade RCC.

Methods Multiple databases were searched to identify studies from their inception to 20 October 2021. Two radiologists independently extracted data from the primary studies. The pooled sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR) and diagnostic OR (DOR) were calculated to assess CTTA performance. The summary receiver operating characteristic (SROC) curve was plotted, and the area under the curve (AUC) was calculated to evaluate the accuracy of CTTA in grading RCC.

Results This meta-analysis included 11 studies, with 1603 lesions observed in 1601 patients. Values of the pooled sensitivity, specificity, PLR, NLR, DOR were 0.79 (95% CI 0.73 to 0.84), 0.84 (95% CI 0.81 to 0.87), 5.1 (95% CI 4.0 to 6.4), 0.24 (95% CI 0.19 to 0.32) and 21 (95% CI 13 to 33), respectively. The SROC curve showed that the AUC was 0.88 (95% CI 0.84 to 0.90). Deeks’ test found no significant publication bias among the studies (p=0.42).

Conclusions The findings of this meta-analysis suggest that CTTA has a high accuracy in differentiating low-grade and high-grade RCC. A standardised methodology and large sample-based study are necessary to certain the diagnostic accuracy of CTTA in grading RCC.

INTRODUCTION

Renal cell carcinoma (RCC) is one of the most common malignancies of the urinary system. However, among the main determinants of RCC prognosis, nuclear grading of carcinoma is widely recognised as an important independent factor. According to the Fuhrman grading system (FGS), RCC is divided into grades I–IV. Previous studies have shown that the FGS can be considered to be an independent factor for predicting the prognosis of patients with RCC. In addition, the simplified two-tier FGS has the same accuracy as the four-tier FGS in predicting the prognosis of RCC, with grades I–II being considered low-grade and grades III–IV being considered high-grade. This simplified grading system reduces inter-observer variability. Low-grade RCC has a high 5-year survival rate, while high-grade RCC has a high metastatic rate and low survival rate. However, recent studies have revealed that loopholes in this grading system result in poor reproducibility of tumour grading. The International Society of Urologic Pathology (ISUP) standard proposed a new grading system for RCC at the 2012 ISUP conference, namely the WHO/ISUP grading system, which was recommended by the WHO in 2016. This grading system can accurately distinguish grades, and it has been shown to be valuable for predicting biological behaviour for clear cell and papillary RCC. The FGS is based on the simultaneous assessment of nuclear size, nuclear shape and nucleolar prominence. The grading is based on the highest-grade area, even if it is focal. Thus, minute foci of higher-grade RCC, as well as carcinoma adjacent to the foci of necrosis, should be taken into account for grading purposes. It is likely that these problems will result in limited
systematically assess the accuracy of CTTA in differentiating low-grade and high-grade RCCs.

**MATERIALS AND METHODS**

**Patient and public involvement**

Patient and the public were not involved in this study.

**Searching strategies**

A literature search was independently performed by two radiologists. The databases were searched from their inception to 20 October 2021 including PubMed, Cochrane Library, Embase, Web of Science, OVID Medline, ScienceDirect, and Springer. The search terms were “renal cell carcinoma”, “renal cancer”, “nephroid carcinoma”, “texture analysis”, “radiomics”, “computed tomography”, “CT” and so on. The titles and abstracts were searched for their relevance. The search strategy is presented in detail in online supplemental file 1. Studies were included according to inclusion and exclusion criteria.

**Inclusion and exclusion criteria**

Studies were selected according to the following criteria: (1) Clinical studies of CTTA for evaluating the accuracy of differentiating low-grade and high-grade RCC, including diagnostic case-control studies, (2) Data were available and could be extracted for calculating the true positive (TP), false positive (FP), true negative (TN) and false negative (FN) values, (3) Histopathological results were used as the gold standard and (4) English literature. The studies were excluded according to the following criteria: (1) Case reports, reviews, abstracts, meta-analyses or animal studies, (2) The data could not be extracted sufficiently or used to calculate estimates in the study and (3) The grade of RCC was assessed only by medical imaging without pathological confirmation.

**Quality assessment of included studies**

The quality of each study was evaluated according to the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2),11 which is recommended by the Cochrane collaboration web.

**Data extraction**

Some studies were deemed irrelevant after reading the titles and abstracts and were excluded. The included studies were selected after reading the full texts based on the inclusion and exclusion criteria. Information extracted from the primary study was as follows: the first author, year of publication, country and language, sample size, research type, model used, gold standard, age of patients, TP, FP, FN, TN, CT slice thickness, contrast, speed of injection and segmentation software. Low-grade RCC (grade I–II) was considered positive, while high-grade RCC (grades II–IV) was considered negative.

**Meta-analysis**

Meta-analysis was conducted by Review Manager V.5.3, Meta-DiSc V.1.4 (Meta disc, Unit of Clinical Biostatistics...
### Table 1 Characteristics of included studies in the meta-analysis

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Country</th>
<th>Study type</th>
<th>n (all)</th>
<th>n (HG)</th>
<th>n (LG)</th>
<th>Age (mean or range)</th>
<th>Machine learning model</th>
<th>Segmentation software</th>
<th>Grading system</th>
<th>CT slicer thinner (mm)</th>
<th>Contrast</th>
<th>Injection speed (mL/s)</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feng</td>
<td>2018</td>
<td>China</td>
<td>Re</td>
<td>131</td>
<td>54</td>
<td>77</td>
<td>25-81</td>
<td>NA</td>
<td>MATLAB</td>
<td>Fuhrman</td>
<td>3</td>
<td>Iodine contrast</td>
<td>3</td>
<td>63</td>
<td>12</td>
<td>14</td>
<td>42</td>
</tr>
<tr>
<td>Cui</td>
<td>2019</td>
<td>China</td>
<td>Re</td>
<td>347</td>
<td>131</td>
<td>216</td>
<td>22-88</td>
<td>SVM</td>
<td>ITK-SNAP</td>
<td>WHO/ISUP</td>
<td>3</td>
<td>Iopamidol</td>
<td>3</td>
<td>168</td>
<td>22</td>
<td>48</td>
<td>109</td>
</tr>
<tr>
<td>Shu</td>
<td>2019</td>
<td>China</td>
<td>Re</td>
<td>108</td>
<td>34</td>
<td>74</td>
<td>54.1</td>
<td>RF</td>
<td>Radclou platform</td>
<td>WHO/ISUP</td>
<td>5</td>
<td>Non-ionic contrast</td>
<td>3.5</td>
<td>69</td>
<td>2</td>
<td>5</td>
<td>32</td>
</tr>
<tr>
<td>Bektas</td>
<td>2018</td>
<td>Turkey</td>
<td>Re</td>
<td>54</td>
<td>23</td>
<td>31</td>
<td>57.5</td>
<td>SVM</td>
<td>3D-Slicer</td>
<td>Fuhrman</td>
<td>1-2</td>
<td>Non-ionic contrast</td>
<td>/</td>
<td>25</td>
<td>2</td>
<td>6</td>
<td>21</td>
</tr>
<tr>
<td>Kocak</td>
<td>2019</td>
<td>Turkey</td>
<td>Re</td>
<td>47</td>
<td>33</td>
<td>14</td>
<td>59.7</td>
<td>CNN</td>
<td>PyRadiomics</td>
<td>Fuhrman</td>
<td>5</td>
<td>Non-ionic contrast</td>
<td>/</td>
<td>11</td>
<td>6</td>
<td>3</td>
<td>27</td>
</tr>
<tr>
<td>Coy</td>
<td>2019</td>
<td>US</td>
<td>Re</td>
<td>132</td>
<td>43</td>
<td>89</td>
<td>62</td>
<td>LR</td>
<td>FDA-approved in-house software</td>
<td>Fuhrman</td>
<td>3</td>
<td>Iodixanol</td>
<td>3</td>
<td>67</td>
<td>9</td>
<td>22</td>
<td>34</td>
</tr>
<tr>
<td>Lin</td>
<td>2019</td>
<td>China</td>
<td>Re</td>
<td>232</td>
<td>43</td>
<td>189</td>
<td>54.9</td>
<td>ML</td>
<td>ITK-SNAP</td>
<td>Fuhrman</td>
<td>1/3</td>
<td>Iopamidol</td>
<td>3</td>
<td>163</td>
<td>5</td>
<td>26</td>
<td>38</td>
</tr>
<tr>
<td>Shu</td>
<td>2018</td>
<td>China</td>
<td>Re</td>
<td>260</td>
<td>99</td>
<td>161</td>
<td>57.1</td>
<td>LR</td>
<td>Radcloud</td>
<td>Fuhrman</td>
<td>5</td>
<td>Iopromid</td>
<td>3.5</td>
<td>109</td>
<td>16</td>
<td>52</td>
<td>83</td>
</tr>
<tr>
<td>Luo</td>
<td>2021</td>
<td>China</td>
<td>Re</td>
<td>230</td>
<td>53</td>
<td>177</td>
<td>56.3</td>
<td>RF</td>
<td>ITK-SNAP</td>
<td>Fuhrman</td>
<td>/</td>
<td>Non-ionic contrast</td>
<td>/</td>
<td>119</td>
<td>7</td>
<td>58</td>
<td>46</td>
</tr>
<tr>
<td>Hussain</td>
<td>2021</td>
<td>Canada</td>
<td>Re</td>
<td>30</td>
<td>15</td>
<td>15</td>
<td>61.2</td>
<td>CNN</td>
<td>PyRadiomics</td>
<td>Fuhrman</td>
<td>/</td>
<td>Unenhance CT</td>
<td>/</td>
<td>12</td>
<td>3</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>Wang</td>
<td>2021</td>
<td>China</td>
<td>Re</td>
<td>32</td>
<td>16</td>
<td>16</td>
<td>58.78</td>
<td>LR</td>
<td>ITK-SNAP</td>
<td>WHO/ISUP</td>
<td>5</td>
<td>Unenhance CT</td>
<td>/</td>
<td>13</td>
<td>3</td>
<td>3</td>
<td>13</td>
</tr>
</tbody>
</table>

CNN, Convolutional Neural Networks; FN, false negative; FP, false positive; HG, high-grade; ISUP, International Society of Urologic Pathology; ITK-SNAP, open-source software; LG, low-grade; LR, logistic regression; M, machine learning; n, number; NA, not available; RF, random forest; SVM, support vector machines; TN, true negative; TP, true positive.
of Ramón y Cajal Hospital, Madrid, Spain) and Stata V.15.1. Based on our opinion of the heterogeneity in the extracted data, we adopted a bivariate random effects model to calculate the pooled estimates in advance. The Cochran-Q method and inconsistency index ($I^2$) were used to investigate heterogeneity among the studies. If $I^2 > 50\%$, $p<0.05$, the observed heterogeneity was significant. If $I^2 < 50\%$, $p>0.05$, the observed heterogeneity was not significant. Pooled sensitivity (Sen), specificity (Spec), PLR, NLR, and diagnostic OR (DOR) were calculated to assess the diagnostic performance of CTTA. The summary receiver operating characteristic (SROC) curve was plotted, and the area under the curve (AUC) was calculated. Deeks’ test was used to evaluate publication bias, and $p>0.05$, indicating that there was no significant bias.\(^{12}\)

### RESULTS

#### Research and selection of studies

A total of 730 relevant articles were initially identified, and 239 duplicate articles were excluded. Additionally, 444 articles were removed after reading their titles and abstracts and being deemed irrelevant. Subsequently, after reading the full texts, 30 articles were found to be reviews or not related to the grade of RCC, and 6 articles were unavailable for data extraction. Ultimately, after checking for relevant studies of the reference in each review or meta analysis, 11 articles were included.\(^{13–23}\) The literature search process is shown in figure 1. There were six studies in which detailed data were unavailable.\(^{24–29}\)

#### Quality assessment and publication bias

The quality of the included studies was evaluated according to the QUADAS-2 checklist, and the results are shown in figures 2 and 3. It was observed that ‘index test’ in ‘risk of bias’ and ‘applicability concerns’ revealed high shortcomings (2/11), which may suggest bias regarding inclusion. Overall, the quality of included studies was satisfactory. Deeks’ funnel plot asymmetry test was used to assess the

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Country</th>
<th>AUC</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deng(^{24})</td>
<td>2019</td>
<td>China</td>
<td></td>
<td>High Fuhrman grade cancers were associated with larger tumour diameter and an increased entropy value(texture analysis) at coarse filter correlated with high Fuhrman grade tumour.</td>
</tr>
<tr>
<td>Lubner(^{25})</td>
<td>2016</td>
<td>USA</td>
<td></td>
<td>Entropy, the SD of the pixel distribution histogram, and the mean of positive pixels were associated with nuclear grade.</td>
</tr>
<tr>
<td>Scrima(^{26})</td>
<td>2019</td>
<td>USA</td>
<td></td>
<td>Entropy and mean of the positive pixels also showed an association with nuclear grade.</td>
</tr>
<tr>
<td>Ding(^{27})</td>
<td>2018</td>
<td>China</td>
<td>0.771</td>
<td>Texture-score based models can facilitate the preoperative discrimination of the high from low grade clear cell RCC.</td>
</tr>
<tr>
<td>Sun(^{28})</td>
<td>2019</td>
<td>China</td>
<td>0.91</td>
<td>The SVM model constructed by CT-based radiomic features can effectively identify the ISUP grades of clear cell RCC.</td>
</tr>
<tr>
<td>Haji-Momenian(^{29})</td>
<td>2020</td>
<td>USA</td>
<td>0.97</td>
<td>The histologic grade of small clear cell RCC can be accurately predicted with machine learning algorithms using histogram features.</td>
</tr>
</tbody>
</table>

AUC, area under the curve; ISUP, International Society of Urologic Pathology; RCC, renal cell carcinoma; SVM, support vector machines.
potential publication bias. The results shown in figure 4 indicate no significant bias (p=0.42).

**Pooled results**

The results of the meta-analysis are presented in figures 5 and 6. Pooled sensitivity and specificity were 0.79 (95% CI 0.73 to 0.84) and 0.84 (95% CI 0.81 to 0.87), respectively (figure 5). Values of PLR, NLR, and DOR were 5.1 (95% CI 4.0 to 6.4), 0.24 (95% CI 0.19 to 0.32), and 21 (95% CI 13 to 33), respectively. The AUC of SROC was 0.88 (95% CI 0.84 to 0.90) (figure 6). These findings indicate that CTTA has a high diagnostic performance in differentiating low-grade and high-grade RCC.

**Heterogeneity test**

Spearman correlation analysis was applied to test the threshold effect, which was caused by the use of different diagnostic cut-off values in a single diagnostic test. The Spearman correlation coefficient was −0.191 (p=0.574), indicating that no significant threshold effect was produced. Heterogeneity was tested using Cochran-Q and $I^2$. In figure 5, the p value of the Cochran-Q test was 0.00 (p<0.05), and $I^2$ was 76.39% in pooled sensitivity. And there was no significant heterogeneity in the pooled specificity (p>0.05, $I^2=0.00$%). These results indicated that there was high heterogeneity in pooled sensitivity among the included studies. Thereafter, we used a bivariate random effect meta-regression to explore the potential association of heterogeneity. The results are presented in online supplemental file 2.

**DISCUSSION**

TA technology was first applied to assess the heterogeneity of tumours, and it was considered to have great potential for the evaluation of renal masses. In recent years, TA based on CT has been gradually applied to differentiate RCC grades. However, studies have demonstrated different diagnostic performances of TA in the diagnosis of RCC. The aim of this meta-analysis was to assess the accuracy of CTTA in differentiating between low-grade and high-grade RCC. The values of pooled sensitivity, specificity, PLR, NLR and AUC were 0.79, 0.84, 5.1, 0.24 and 0.88, respectively. The results indicated that CTTA had excellent diagnostic performance in differentiating low-income and high-grade RCC, which could be considered a reliable method for diagnosing the grade of RCC in clinical practice.

The gold standard for the diagnosis of RCC is histopathological biopsy. However, this method is invasive and may be inaccurate because of the samples being collected from different biopsy sites. The field of CTTA, owing to the ability of this technology to quantitatively extract texture features, has attracted the attention of researchers. It avoids the subjective influence of image processing and reduces the possibility of errors. Currently, the assessment of RCC based on traditional imaging methods primarily includes the overall outline of RCC, such as size, shape and degree of contrast enhancement. However, these parameters can only define the outline and anatomical sites of tumours and are unable to provide vital information regarding the grade of the carcinomas. The differences between low-grade and high-grade RCC involve changes in the pixel intensity of the images. CTTA can detect subtle changes in pixel intensity caused by heterogeneity between low-grade and high-grade RCC. In addition, CTTA performs a comprehensive evaluation of lesions. Compared with biopsy, this technique evaluates the mass through an integrated rather than a focal analysis, which avoids the influence of sampling site error. Different grades of RCC require different therapies. Low-grade RCC patients may undergo partial nephrectomy,
whereas high-grade RCC patients may require more invasive and extensive surgery. As an important prognostic factor, it is important to preoperatively differentiate the grade of RCC in clinical practice and provide more valuable guidance for clinicians.

Image preprocessing is an essential step in TA, and the methods used for image preprocessing differed greatly in the included studies. Image segmentation and quantitative analysis were conducted after the image preprocessing step, followed by the establishment of a diagnostic predictive model. Lastly, the SROC curve and AUC were calculated to evaluate diagnostic performance. Previous studies have indicated that texture features, such as entropy, SD and mean of the positive pixels, were associated with nuclear grade. An increased entropy value correlated with high Fuhrman grade tumours. The radiomics model with texture features constructed by some scholars indicated a high prediction accuracy in identifying the grading of RCC.

There were some limitations in this meta-analysis: all included studies in this meta-analysis were retrospective in design, which has a higher bias risk than prospective studies; there was high heterogeneity between the included studies, which may be due to the age of patients, tools of TA and image preprocessing; some of the included studies were conducted with a small number of samples; and the sample size varied greatly, which may affect the accuracy of the results.

**CONCLUSION**

This study suggested that CTTA has high accuracy in differentiating low-grade and high-grade RCC, which could be considered as a non-invasive method to provide crucial information for the grading of RCC. However, a standard methodology and large sample-based study are necessary to certain the diagnostic accuracy of CTTA in RCC grading.

**Contributors** WY and GL came up with this meta-analysis and completed the work of design. LZ and YY implemented this systematic review with the guidance of WY. LZ and YY performed the literature search, data extraction and quality assessment. Statistical analysis was completed by WY and YW. YW is the guarantor.
for this article, WY and GL wrote the first draft, then all authors participated in the modification of the manuscript.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient consent for publication** Not applicable.

**Ethics approval** This study does not involve human participants.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** All data relevant to the study are included in the article or uploaded as online supplemental information.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

**ORCID iD**

Yinghua Wu http://orcid.org/0000-0001-9116-1624

**REFERENCES**