HHS Public Access

Author manuscript

Urology. Author manuscript; available in PMC 2024 October 01.

Published in final edited form as:

Urology. 2023 October ; 180: 160–167. doi:10.1016/j.urology.2023.07.017.

AI-generated R.E.N.A.L.+ score surpasses human-generated score in predicting renal oncologic outcomes.

Nour Abdallah[1,*], **Andrew Wood**[1,*], **Tarik Benidir**[1], **Nicholas Heller**[2], **Fabian Isensee**[3], **Resha Tejpaul**[2], **Dillon Corrigan**[4], **Chalairat Suk-ouichai**[5], **Griffin Struyk**[2], **Keenan Moore**[2], **Nitin Venkatesh**[2], **Onuralp Ergun**[6], **Alex You**[7], **Rebecca Campbell**[1], **Erick M. Remer**[1,8], **Samuel Haywood**[1], **Venkatesh Kirshnamurthi**[1], **Robert Abouassaly**[1], **Steven Campbell**[1], **Nikolaos Papanikolopoulos**[2], **Christopher J Weight**[1]

[1]Glickman Urological and Kidney Institute, Cleveland, Ohio, USA

[2]Department of Computer Science and Engineering, University of Minnesota, Minneapolis, Minnesota, USA

[3]German Cancer Research Center (DKFZ) Heidelberg, University of Heidelberg, Heidelberg, Germany

[4]Department of quantitative health sciences, Lerner Research Institute, Cleveland, Ohio, USA

[5]Department of Surgery, Siriraj Hospital, Mahidol University, Bangkok, Thailand

[6]Cleveland Clinic Research Institute, Cleveland, Ohio, USA

Corresponding author: Nour Abdallah MD, Research Fellow, Glickman Urological and Kidney Institute, Cleveland Clinic Foundation, 9500 Euclid Ave, Cleveland, OH, 44195, 216 444-1105, abdalln@ccf.org.

*Equal first-authorship contribution

Declarations of interest: None.

Conflict of interest:
1) Nour Abdallah: no conflict.
2) Andrew Wood: no conflict.
3) Tarik Benidir: no conflict.
4) Nicholas Heller: no conflict.
5) Fabian Isensee: no conflict.
6) Resha Tejpaul: no conflict.
7) Dillon Corrigan: no conflict.
8) Chalairat Suk-ouichai: no conflict.
9) Griffin Struyk: no conflict.
10) Keenan Moore: no conflict.
11) Nitin Venkatesh: no conflict.
12) Onuralp Ergun: no conflict.
13) Alex You: no conflict.
14) Rebecca Campbell: no conflict.
15) Erick M. Remer: no conflict.
16) Samuel Haywood: no conflict.
17) Venkatesh Kirshnamurthi: no conflict.
18) Robert Abouassaly: no conflict.
19) Steven Campbell: no conflict.
20) Nikolaos Papanikolopoulos: no conflict.
21) Christopher J Weight: no conflict.

[7]Case Western Reserve University, Cleveland, Ohio, USA

[8]Department of Diagnostic Radiology, Imaging Institute, Cleveland Clinic, Cleveland, Ohio, USA

Abstract

Objectives: To determine whether we can surpass the traditional R.E.N.A.L. nephrometry score (H-score) prediction ability of pathologic outcomes by creating artificial intelligence(AI)-generated R.E.N.A.L.+ score(AI+score) with continuous rather than ordinal components. We also assessed the AI+ score components' relative importance with respect to outcome odds.

Methods: This is a retrospective study of 300 consecutive patients with preoperative CT scans showing suspected renal cancer at a single institution from 2010–2018. H-score was tabulated by three trained medical personnel. Deep neural network approach automatically generated kidney segmentation masks of parenchyma and tumor. Geometric algorithms were used to automatically estimate score components as ordinal and continuous variables. Multivariate logistic regression of continuous R.E.N.A.L. components was used to generate AI+score. Predictive utility was compared between AI+, AI, and H-scores for variables of interest, and AI+score components' relative importance was assessed.

Results: Median age was 60 years(IQR 51–68), and 40% were female. Median tumor size was 4.2 cm(2.6–6.12), and 92% were malignant, including 27%, 37%, and 23% with high-stage, high-grade, and necrosis, respectively. AI+score demonstrated superior predictive ability over AI and H-scores for predicting malignant(AUC 0.69 vs.0.67 vs.0.64, respectively), high-stage(AUC 0.82 vs.0.65 vs.0.71, respectively), high-grade(AUC 0.78 vs.0.65 vs.0.65, respectively), pathologic tumor necrosis(AUC 0.81 vs.0.72 vs.0.74, respectively), and partial nephrectomy approach(AUC 0.88 vs.0.74 vs.0.79, respectively). Of AI+score components, the maximal tumor diameter ("R") was the most important outcomes predictor.

Conclusions: AI+ score was superior to AI-score and H-score in predicting oncologic outcomes. Time-efficient AI+score can be used at the point of care, surpassing validated clinical scoring systems.

Keywords

Artificial Intelligence; Computed Tomography; Machine Learning; Partial nephrectomy; Renal tumors

INTRODUCTION

The R.E.N.A.L. (radius, endophytic/exophytic, nearness to the collecting system, anterior/posterior, location relative to the polar lines) nephrometry score aims to convey the surgical complexity of a renal mass in a single quantifiable measure [1]. Over the past decade, this nephrometry score has been prospectively validated as a meaningful predictor of perioperative, pathologic, and survival outcomes [1–4]. However, its widespread clinical adoption has been limited by interobserver variability and substantial required time investment by busy clinicians [5,6]. Any technology that allows for mitigating these two barriers would provide substantial value to Urologists.

The application of Deep Learning (DL), a subfield of Machine Learning (ML), continues to expand in the healthcare space to help solve problems based on high-dimensional data [7–9]. In urology, many of the most impressive applications to date have involved renal cancer. For instance, DL has demonstrated an ability to reliably differentiate renal tumor subtypes and predict functional postoperative outcomes [10–13]. A novel DL process for fully automated segmentation of kidneys and kidney tumors was previously described, a process that, when done manually, is often an onerous rate-limiting step in radiomic studies [13,14]. Termed semantic segmentation (SS), this fully automated segmentation process promises paradigm-shifting benefits to researchers looking to investigate radiomic features on larger scales not conducive to manual input. However, the potential advantages of automation in kidney cancer imaging research go beyond computer-generated tumor segmentation. For example, the ability of artificial intelligence (AI)-generated R.E.N.A.L. nephrometry scores to predict survival and oncologic outcomes was assessed in patients undergoing extirpative surgery for renal mass. AI-generated scores were highly correlated with scores calculated by human experts and could deduce the presence of malignancy, grade, stage, and necrosis as reliably as human experts [14].

Another potential advantage to automated AI scoring systems lies in the potential to overcome an inherent limitation of most nephrometry scores: namely, the categorical nature of the captured variables. Grouping ordinal variables (e.g., a renal mass diameter of 4–7 cm gets a score of 2) was intended to simplify, standardize, and reduce the clinician time needed for generating these scores. However, valuable information is lost when categorizing continuous variables [15], thus potentially reducing the predictive accuracy of scoring systems. Additionally, scoring systems such as R.E.N.A.L. have traditionally weighed each component of the score similarly and combined them with a simple addition to make obtaining a total score easy for clinicians. Automated AI nephrometry scoring would not be bounded by this constraint and could combine continuous nephrometry score components using complex multivariate models.

With this in mind, we aimed to generate a modified R.E.N.A.L. score (AI+ score) using AI-generated segmentations and continuous rather than ordinal variables combined using multivariate logistic regression. We hypothesize that this score will improve predictive accuracy for pathologic and operative outcomes while reducing the need for human capital.

MATERIAL AND METHODS

1. Cohort

Following ethics board approval of protocol code 1611M00821 at the University of Minnesota-Twin Cities, 300 patients with preoperative arterial phase computed tomography (CT) scans were identified among 544 consecutive patients undergoing extirpative surgery for a renal tumor at a single institution between 2010 and 2018. Patients with a tumor thrombus were excluded, and if a patient had more than one tumor, the largest tumor was used to determine the nephrometry score. Overall inclusion criteria for this cohort are based on a previously published KiTS19 (kidney and kidney tumor segmentation challenge) protocol [13]. This cohort, with scans, segmentations, clinical details, and outcomes, is now publicly available at https://kits21.kits-challenge.org/. Tumors of the 300 selected patients

were subsequently assigned an AI-generated continuous score for each component of the R.E.N.A.L. score (AI+ score), an AI-generated categorical R.E.N.A.L. score (AI-score), and a human-calculated traditional R.E.N.A.L. score (H-score) based on the following process.

2. Semantic Segmentation

a. Creating the "ground truth" for Deep Learning—Each voxel of included CT scans was manually segmented as "kidney", "kidney tumor", or "background" by specialized medical personnel. The resulting dataset of almost 50,000 individual axial slices is referred to as the "ground truth", providing the training material for the DL algorithms and the basis for extracting R.E.N.A.L. scores. These ground truth manual CT segmentations were then used to host an international segmentation challenge [13], with a single algorithm [16] declared the winner by demonstrating the highest level of fidelity to the ground truth segmentations. Subsequently, this algorithm was used as the basis for the generation of fully automated segmentations masks. From these masks, we extracted the components for nephrometry scores for the 300 patients described above. A more detailed methodology was previously published [14].

b. Collecting Ordinal R.E.N.A.L. score components from segmented CT scans

1) Radius: The distance between the two furthest "tumor" voxels was recorded. The numerical continuous value of the distance was noted in millimeters. A score of 1, 2, or 3 was also allocated as ordinal variables based on the standardized cutoffs (< 40 mm, 40 – 70 mm, and > 70 mm, respectively).

2) Endophycity: In each axial slice representing tumor voxels, the convex hull of the "kidney" region excluding the tumor was identified, then the location of every "tumor" voxel was documented as being either inside (endophytic) or outside (exophytic) of this convex hull. Endophytic proportions were recorded as numerical continuous values (0–100%), and an ordinal score of 1, 2, or 3 was assigned, corresponding to endophytic proportions of <50%, <100%, and 100%, respectively.

Extracting the "N" (nearness to the urinary collecting system (UCS)) and "L" (location relative to polar lines) components required a slight variation in the score to overcome challenges with AI recognition of these structures. First, renal sinus fat was identified and labeled as central voxels within the kidney segmentation with attenuation below a certain Hounsfield Unit (HU) threshold. Next, we defined polar lines as the highest (cranial) and lowest (caudal) axial slices showing sinus fat voxels in close proximity to normal kidney parenchyma voxels. Finally, as the UCS is generally bordered by sinus fat, we defined the UCS as voxels within the predetermined area of sinus fat with attenuation exceeding the previously mentioned HU threshold (i.e., significantly brighter than the sinus fat).

3) Nearness to the Collecting System: The smallest distance between a "tumor" voxel and a "UCS" voxel was computed in millimeters. The numerical continuous value of the distance was extracted, and an ordinal score of either 1, 2, or 3 was assigned when the distance was > 7 mm, 4 – 7 mm, and < 4 mm, respectively.

4) Longitudinal Location: The tumor voxels were classified based on their location relative to the polar lines. The numerical proportion (0 to 100%) of tumor voxels present between the polar lines was calculated. Also, an ordinal score of either 1, 2, or 3 was assigned according to the following schema: 1 point if 0% of tumor voxels were located between the polar lines, 2 points if 0–50% of tumor voxels were located between the polar lines, 3 points if >50% of tumor voxels were located between polar lines.

The "A" component, referring to the anterior/posterior location, traditionally has not been included in predictive models and was not included in this study as it does not numerically contribute to the composite R.E.N.A.L. score.

Six patients (2%) had limitations in the segmentation masks, which prevented the computation of their corresponding AI-generated score.

3. Ordinal AI Scoring (AI-score)

The AI-score was computed as the sum of the ordinal variables generated by the AI model.

4. Multivariate Logistic Regression Using Continuous R.E.N.A.L. variables (AI+ score)

The AI+ score was generated using logistic regression utilizing the automatically generated R.E.N.A.L. continuous components. SUPPLEMENTARY TABLE 1 represents the equations used to generate the AI+ score and the respective predicted probability for each outcome of interest.

5. Human scoring (H-score)

Traditional human-calculated nephrometry scores were assigned to each tumor by a team of three trained medical personnel blinded to AI+ score results. All personnel was trained in the R.E.N.A.L. scoring system by a fellowship-trained Urologic Oncologist. All values were recorded in an ordinal fashion as originally described and currently implemented in clinical practice [1].

6. Statistical Analysis

Multivariate logistic regression was performed using the generated R.E.N.A.L. components to create a multivariate model to predict each of the outcomes of interest (AI+ score). Receiver operating characteristic (ROC) curves were then developed from logistic regression models to evaluate the discriminatory ability of the AI+ (multivariate regression) versus the AI-score (univariate) and H-score (univariate) for each outcome of interest, including the presence of malignancy, high stage, high grade, tumor necrosis, and surgery type. Areas under the ROC curves (AUC) were calculated. JMP © Pro 14.2 statistical software (SAS Institute Inc., Cary, North Carolina) was used for all statistical analyses.

RESULTS

1. Characteristics of the cohort

The detailed baseline characteristics of the cohort are represented in TABLE 1. The median age was 60 years (Interquartile range (IQR) 51–68), with 120 (40%) females and 180 (60%)

males. The median body mass index was 29 Kg/m^2 (26–35). Concerning the tumors, 275 (92%) were malignant, of which 75 (27%) were high-stage (pathological stage T3-T4), 92 (37%) were high-grade (Fuhrman grade 3–4), and 69 (23%) had tumor necrosis. The median diameter was 4.2 cm (2.6–6.1). Surgically, a minimally invasive approach was used in 221 patients (73%), and 188 patients (63%) had a partial nephrectomy. The median AI-score was 8 (6–9), and the median H-score was 8 (6.25–9).

2. Prediction of Clinical and Pathologic Outcomes

The AI+ score demonstrated superior AUC values than models using either AI-score or H-score for predicting all measured outcomes. FIGURE 1 represents the relative performance of the AI+ score, AI-score, and H-score in predicting pathologic outcomes. For predicting malignancy, AI+ score, AI-score, and H-score demonstrated AUCs of 0.69, 0.67, and 0.64, respectively. For predicting high-stage disease: 0.82, 0.65, and 0.71, respectively. For predicting high-grade disease: 0.78, 0.65, and 0.65, respectively. For predicting tumor necrosis, 0.81, 0.72, and 0.74, respectively. FIGURE 2 represents the relative performance of the AI+ score, AI-score, and H-score in predicting partial (vs. radical) nephrectomy. AI+ score, AI-score, and H-score demonstrated AUCs of 0.88, 0.74, and 0.79, respectively.

3. Analysis of the relative importance of the AI+ score components

Examining the AI+ score components using multivariable logistic regression modeling, represented in TABLE 2, revealed that the “R” component was a statistically significant predictor of high-stage disease (per cm, Odds ratio (OR) 2.84 (1.34–4.45; p<0.001), high-grade disease (per cm, OR 1.80 (0.33–3.30); p=0.01), and the presence of tumor necrosis (per cm, OR 4.41 (2.60–6.10); p<0.001). The “E” component was a statistically significant predictor for high-grade disease (OR 0.10 (0.02–0.61); p=0.008). All other components did not predict outcomes of interest in a statistically significant fashion when combined into a multivariate model.

COMMENT

In this study, we explored the ability of multivariate logistic regression using continuous versions of R.E.N.A.L. components (AI+ score) to surpass validated clinical scoring systems to accurately predict pathologic outcomes in patients with renal mass. Importantly, the calculation of the AI+ score and AI score can be performed without human intervention at any step, as the generation of segmentation masks of kidneys and tumors from CT scans, measurement of continuous R.E.N.A.L. components, and weighting and combination of components into a multivariate model can all be accomplished in an automated fashion. The AI+ score surpassed the AI-generated and the human expert-generated R.E.N.A.L. scores (which both use ordinal variables) in predicting clinical and pathologic outcomes of renal tumors. While it is clearly possible that a human expert could also quantify all these variables in a continuous fashion and calculate the probability of the outcomes using equations in SUPPLEMENTARY TABLE 1, we believe that this will never happen since it is too complicated.

A recent meta-analysis of R.E.N.A.L. score studies showed that not all the components of the R.E.N.A.L. score were significantly correlated with surgical strategies [17]. Because of this finding, we aimed to assess the relative contribution of each component of the R.E.N.A.L. score in its ability to predict postoperative pathologic outcomes. We found that the “R” component, reflecting the tumor size, was the most influential factor in predicting pathologic outcomes, including a high-stage disease, high-grade disease, and tumor necrosis. This was concordant with the results of previous studies showing that the “R” component was the most important predictor of malignant and high-grade tumors [18,19].

In addition to our own internal comparisons, the predictive ability of the AI+ score compares favorably to other previously published nomograms incorporating the R.E.N.A.L. score. For example, Deng et al. were able to achieve a comparable but inferior AUC of 0.66 (AI+ AUC 0.69) for predicting malignant tumors using a nomogram based on R.E.N.A.L. components [20]. Similarly, Wang et al. were able to achieve a similar but inferior AUC of 0.73 (AI+ AUC of 0.78) for predicting high-grade tumors using a R.E.N.A.L.-based nomogram [21]. Importantly, neither of these studies included continuous versions of R.E.N.A.L. variables, and neither utilized an automated AI-driven process for the generation of kidney and tumor segmentations and measurement of R.E.N.A.L. variables. Because of these differences, the only way previous studies could generate nephrometry score-based nomograms was through an extensive human effort to manually measure ordinal variables.

By replacing ordinal variables with continuous variables, and combining them in a multivariate model that allowed for differential weighting of components, we were able to demonstrate a modified R.E.N.A.L. score with improved predictive utility. Continuous variables are often grouped into categories in order to simplify the analysis and interpretation of results [15]. However, it is well recognized that categorizing continuous variables leads to loss of information and, thus, decreased predictive utility in the analysis [22, 23]. Within the context of nephrometry scores, the clearest disadvantage of this approach is the blurring of information within the same category, as all tumors belonging to the same category are treated equally, yet prognosis may vary considerably [24]. There is a significant difference, for instance, between a 6.9 cm 95% endophytic renal mass and a 4.1 cm 50% endophytic renal mass. Despite this difference, the current ordinal R.E.N.A.L. score system describes these tumors identically. Additionally, the current R.E.N.A.L. scoring system weighs all components the same, another concession to simplicity that reduces prognostic rigor.

The ordinal treatment of variables and equal weighting in the original R.E.N.A.L. model was clearly done to help facilitate ease of implementation. Despite this effort, the largest impediment t widespread clinical use remains clinician time investment. For a score that is already rarely used outside of research efforts, asking clinicians to measure variables in individual millimeters and then combine them using complex coefficients is unrealistic. Fortunately, once validated, the utilization of a fully automated modified R.E.N.A.L. score promises to obviate this concern and improve upon predictive utility. It is easy to envision a world where every radiology report for a CT scan that identifies a renal mass includes estimates of risk for pathologic variables of interest based on this automated modified R.E.N.A.L. score without requiring clinician or radiologist time investment. This automation

would also allow for the democratization of the benefits of nephrometry scoring to clinicians who are not renal cancer or statistical experts.

While the demonstrated improvements in the prediction of adverse pathologic features will not always result in changes to management decisions, they carry important prognostic implications. With recent developments in the adjuvant treatment of renal cell carcinoma using immunotherapy, the ability to counsel patients accurately on the possibility of adverse pathology is more important than ever. In addition, with the use of active surveillance for renal masses growing steadily, we envision a day when a fully automated system is able to provide information about which patients are safe to surveil. Our data provide evidence that clinicians can trust fully automated versions of previously understood scoring systems (such as R.E.N.A.L.) to generate useful information. This is an important intermediary step in the development and implementation of more robust machine learning-based radiomic scoring systems, where the complexity of the calculations may not be intelligible or reproducible by humans (such as radiomics terms like kurtosis, entropy and others yet to be described).

In addition to benefits to clinical implementation, automation of a continuous R.E.N.A.L. score also carries important implications for continued renal mass and renal cell carcinoma research investigation. Although one of the originally stated goals of developing the R.E.N.A.L. score [2], widespread standardization of renal mass complexity measurement remains a vital, unattained goal. This is particularly important in the current environment of AI and ML-based investigation. The unique value of neural network approaches relates to their ability to identify nuances of variation between cases. This ability can only be realized with a sufficient sample size to capture how these nuances in variation relate to outcomes of interest [25]. While there is no single answer to the quantity of training data required for a given problem (i.e., predicting kidney cancer oncologic outcomes), a higher volume of high-quality data generally produces superior results. Even outside of the AI and ML research space, a fully automated continuous variable-based nephrometry score has important implications for the efficiency and uniformity of kidney cancer research. With current methods, calculating nephrometry scores on a multi-thousand-patient renal mass dataset requires hundreds of person-hours, and that’s for standard ordinal variables. The use of an automated system would also prevent issues with interobserver variability in interpretation and measurement. In summary, the ability to generate superior R.E.N.A.L. nephrometry scores in an automated fashion for large-volume data sets promises to improve the consistency, quality, and overall volume of renal mass complexity data for the field as a whole.

Our study is not without limitations. First, our algorithm could not generate an AI-score for 6 cases out of the 300 (2%). This was likely due to poor image quality, motion artifact, or significant renal atrophy, and segmentation algorithms that don’t work properly can make this approach not possible. While it is important to recognize the technological limitations of an automated system, we believe a 2% failure rate is acceptable at this stage. Another potential limitation was our single-center experience, as surgery was done at a single institution. However, the CT scan images were collected in more than 70 medical facilities, providing a solid base for the generalizability and external validity of the AI segmentation model and AI+ score. We also utilized consecutive patients and thus accepted all tumor

sizes, locations, and types, which reflects real-world practice. Other segmentation-related limitations were previously reported in detail [14].

CONCLUSIONS

An AI-generated continuous variable-based modified R.E.N.A.L. score is able to provide meaningful predictions of clinical and pathologic outcomes of interest superior to those generated in an ordinal fashion. We propose the transition to continuous variables within nephrometry scoring systems as the use of automated systems becomes more common. While fully automated segmentation and calculation of nephrometry scores are in their infancy, they promise paradigm-shifting benefits to both clinical practice and research investigation for patients with renal mass.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding source:

National Institute of Health (RO1-CA225435)

ABBREVIATIONS

AI	artificial intelligence AUC area under the curve
CT	computed tomography
DL	deep learning
HU	Hounsfield Unit IQR interquartile range ML machine learning OR odds ratio
R.E.N.A.L.	radius, endophytic/exophytic, nearness to the collecting system, anterior/posterior, location relative to the polar lines
ROC	receiver operating characteristic
SS	semantic segmentation
UCS	urinary collecting system

REFERENCES

1. Joshi SS, Uzzo RG. Renal Tumor Anatomic Complexity: Clinical Implications for Urologists. Urol Clin North Am 2017;44(2):179–187. [PubMed: 28411910]
2. Kutikov A, Uzzo RG. The R.E.N.A.L. nephrometry score: a comprehensive standardized system for quantitating renal tumor size, location and depth. J Urol 2009;182(3):844–853. [PubMed: 19616235]
3. Kutikov A, Smaldone MC, Egleston BL, et al. Anatomic features of enhancing renal masses predict malignant and high-grade pathology: a preoperative nomogram using the RENAL Nephrometry score. Eur Urol 2011;60(2):241–248. [PubMed: 21458155]

4. Weight CJ, Atwell TD, Fazzio RT, et al. A multidisciplinary evaluation of inter-reviewer agreement of the nephrometry score and the prediction of long-term outcomes. J Urol 2011;186(4):1223–1228. [PubMed: 21849200]
5. Chapin BF, Wood CG. The RENAL nephrometry nomogram: statistically significant, but is it clinically relevant?. Eur Urol 2011;60(2):249–252. [PubMed: 21514040]
6. Spaliviero M, Poon BY, Aras O, et al. Interobserver variability of R.E.N.A.L., PADUA, and centrality index nephrometry score systems. World J Urol 2015;33(6):853–858. [PubMed: 25149471]
7. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. N Engl J Med 2019;380(14):1347–1358. [PubMed: 30943338]
8. Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. JAMA 2018;319(13):1317–1318. [PubMed: 29532063]
9. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436–444. [PubMed: 26017442]
10. Kocak B, Yardimci AH, Bektas CT, et al. Textural differences between renal cell carcinoma subtypes: Machine learning-based quantitative computed tomography texture analysis with independent external validation. Eur J Radiol 2018;107:149–157. [PubMed: 30292260]
11. Feng Z, Rong P, Cao P, et al. Machine learning-based quantitative texture analysis of CT images of small renal masses: Differentiation of angiomyolipoma without visible fat from renal cell carcinoma. Eur Radiol 2018;28(4):1625–1633. [PubMed: 29134348]
12. Sharma N, Zhang Z, Mir MC, et al. Comparison of 2 Computed Tomography-based Methods to Estimate Preoperative and Postoperative Renal Parenchymal Volume and Correlation With Functional Changes After Partial Nephrectomy. Urology 2015;86(1):80–86. [PubMed: 26142589]
13. Heller N, Isensee F, Maier-Hein KH, et al. The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 challenge. Med Image Anal 2021;67:101821. [PubMed: 33049579]
14. Heller N, Tejpaul R, Isensee F, et al. Computer-Generated R.E.N.A.L. Nephrometry Scores Yield Comparable Predictive Results to Those of Human-Expert Scores in Predicting Oncologic and Perioperative Outcomes [published correction appears in J Urol. 2022 Oct;208(4):939]. J Urol 2022;207(5):1105–1115. [PubMed: 34968146]
15. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ 2006;332(7549):1080. [PubMed: 16675816]
16. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 2021;18(2):203–211. [PubMed: 33288961]
17. Shi N, Zu F, Shan Y, et al. The value of renal score in both determining surgical strategies and predicting complications for renal cell carcinoma: A systematic review and meta-analysis. Cancer Med 2020;9(11):3944–3953. [PubMed: 32281277]
18. Thompson RH, Kurta JM, Kaag M, et al. Tumor size is associated with malignant potential in renal cell carcinoma cases. J Urol 2009;181(5):2033–2036. [PubMed: 19286217]
19. Chen SH, Wu YP, Li XD, et al. R.E.N.A.L. Nephrometry Score: A Preoperative Risk Factor Predicting the Fuhrman Grade of Clear-Cell Renal Carcinoma. J Cancer 2017;8(18):3725–3732. [PubMed: 29151960]
20. Deng X, Liu X, Hu B, et al. Pathological diagnostic nomograms for predicting malignant histology and unfavorable pathology in patients with endophytic renal tumor. Front Oncol 2022;12:964048. [PubMed: 36212405]
21. Wang HK, Zhu Y, Yao XD, et al. External validation of a nomogram using RENAL nephrometry score to predict high grade renal cell carcinoma. J Urol 2012;187(5):1555–1560. [PubMed: 22425078]
22. Zhao LP, Kolonel LN. Efficiency loss from categorizing quantitative exposures into qualitative exposures in case-control studies. Am J Epidemiol 1992;136(4):464–474. [PubMed: 1415166]
23. Selvin S Statistical Power and Sample Size Calculations: Statistical Analysis of Epidemiological Data New York: Oxford University Press; 2004:75–92.

24. Naggara O, Raymond J, Guilbert F, Roy D, Weill A, Altman DG. Analysis by categorizing or dichotomizing continuous variables is inadvisable: an example from the natural history of unruptured aneurysms. AJNR Am J Neuroradiol 2011;32(3):437–440. [PubMed: 21330400]

25. Rasmussen R, Sanford T, Parwani AV, Pedrosa I. Artificial Intelligence in Kidney Cancer. Am Soc Clin Oncol Educ Book 2022 Apr;42:1–11.

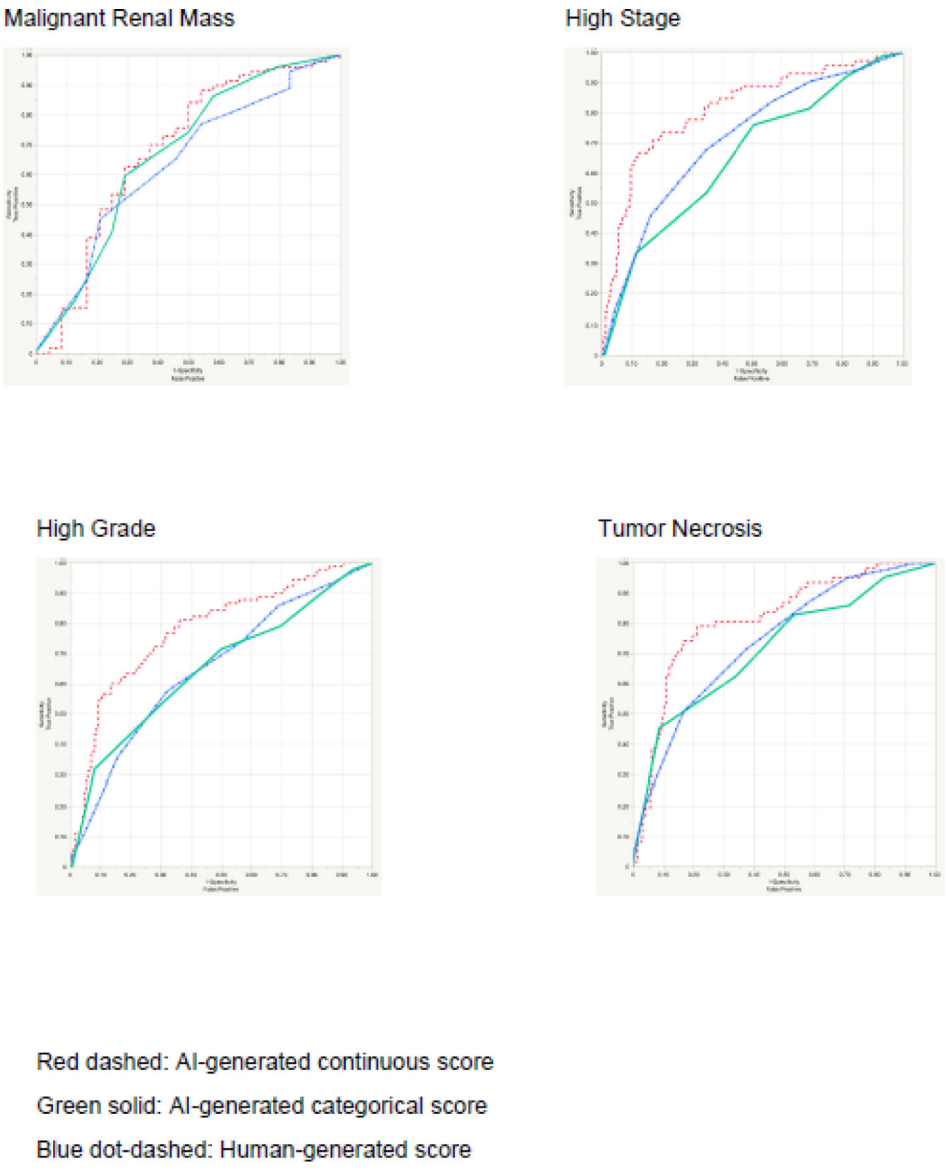

Figure 1.

Receiver operating characteristic (ROC) Curves Assessing the Predictive ability of Pathologic Outcomes for the different R.E.N.A.L. scores.

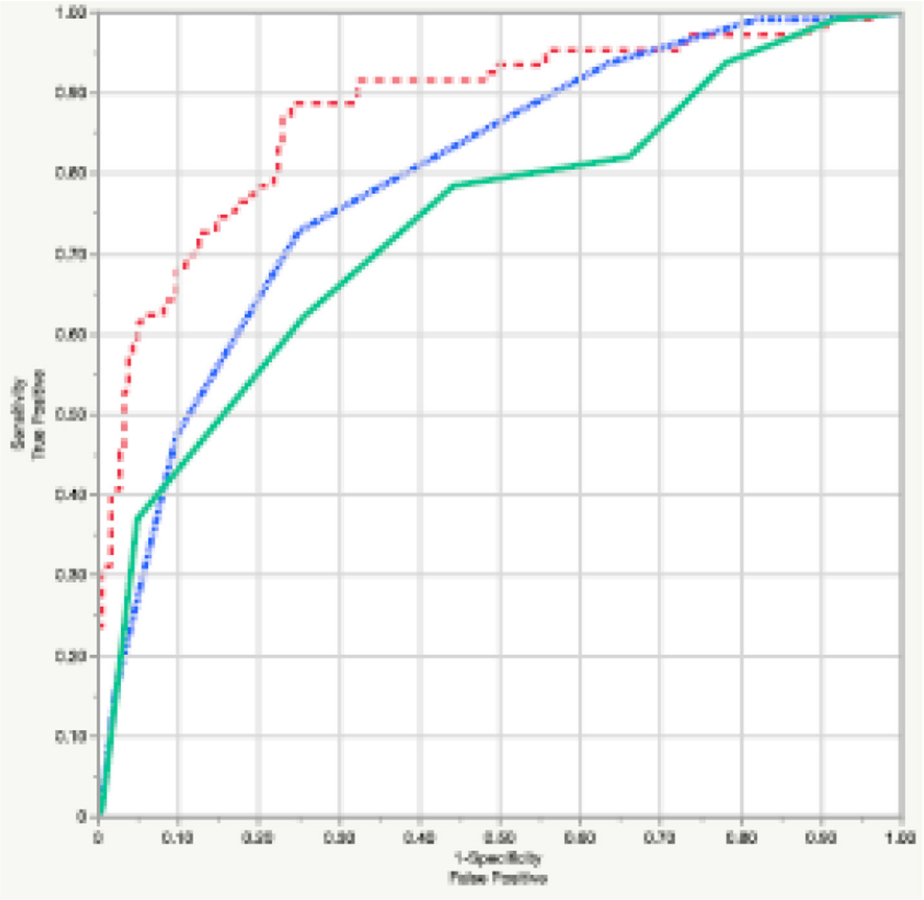

Figure 2.
Receiver operating characteristic (ROC) Curves Assessing the Predictive ability of Surgery Type for the different R.E.N.A.L. scores.

Table 1.

Baseline characteristics (N=300)

Characteristic	**N = 300**
Gender, n (%) Female Male	 120 (40) 180 (60)
Age (yrs.), Median (IQR)	60 (51 – 68)
Tumor diameter (cm), Median (IQR)	4.20 (2.60 – 6.12)
Body Mass Index (kg/m^2), Median (IQR)	29 (26 – 35)
Baseline eGFR (mL/min/1.73m^2), Median (IQR)	72 (60 – 81)
AI-R.E.N.A.L. score, Median (IQR)	8.00 (6.00 – 9.00)
Human-R.E.N.A.L. score, Median (IQR)	8.00 (6.25 – 9.00)
Malignant Renal Mass, n (%)	275 (92)
Pathologic T-stage, n (%) 0 1a 1b 2a 2b 3 3a 4	 24 (8.0) 121 (40) 60 (20) 15 (5.0) 5 (1.7) 8 (2.7) 62 (21) 5 (1.7)
Tumor Necrosis, n (%)	69 (23)
Tumor Grade, n (%) 0 1 2 3 4	 25 (9.3) 33 (12) 119 (44) 66 (25) 26 (9.7)
Surgical Technique, n (%) Laparoscopic Open Robotic	 49 (16) 79 (26) 172 (57)
Nephrectomy Type, n (%) Partial Nephrectomy Radical Nephrectomy	 188 (63) 112 (37)
Estimated Blood Loss (mL), Median (IQR)	200 (100 – 400)
Complications, n (%)	47 (16)
Length of Hospital Stay (Days), Median (IQR)	3.00 (2.00 – 4.00)

Table 2.

Multivariable Logistic Regression Models using Continuous AI+ score Components

	Malignant Mass		High Stage		High Grade		Tumor Necrosis	
R.E.N.A.L. Component	**OR (95% CI)[1]**	**p-value**	**OR (95% CI)[1]**	**p-value**	**OR (95% CI)[1]**	**p-value**	**OR (95% CI)[1]**	**p-value**
R (per cm)	1.02 (0.98 – 1.03)	0.37	2.84 (1.34 – 4.45)	**<0.001**	1.80 (0.33 – 3.30)	**0.01**	4.41 (2.60 – 6.10)	**<0.001**
E	1.83 (0.20 – 1.71)	0.59	0.17 (0.02 – 1.10)	0.06	0.10 (0.02 – 0.61)	**0.008**	1.01 (0.99 – 1.03)	0.22
N (per mm)	0.96 (0.92 – 1.01)	0.16	0.96 (0.90 – 1.02)	0.22	0.97 (0.92 – 1.02)	0.20	0.96 (0.90 – 1.03)	0.29
L	1.01 (0.99 – 1.02)	0.18	1.00 (0.99 – 1.01)	0.95	1.00 (0.99 – 1.01)	0.61	0.96 (0.68 – 1.35)	0.93

[1]OR = Odds Ratio, CI = Confidence Interval