IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 7 May 2024, accepted 5 June 2024, date of publication 11 June 2024, date of current version 21 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3412655

RESEARCH ARTICLE

Distinguishing Kidney Tumor Types Using Radiomics Features and Deep Features

ROBERTO MAGHERINI[1], MICHAELA SERVI[1], YARY VOLPE[1], RICCARDO CAMPI[2,3], AND FRANCESCO BUONAMICI[1]

[1]Department of Industrial Engineering, University of Florence, 50139 Florence, Italy
[2]Unit of Urological Robotic Surgery and Renal Transplantation, Careggi University Hospital, University of Florence, 50134 Florence, Italy
[3]Department of Experimental and Clinical Medicine, University of Florence, 50134 Florence, Italy

Corresponding author: Roberto Magherini (roberto.magherini@unifi.it)

This work was supported in part by the Spoke 3 within the Italian National Research Programme (NRP) TUSCANY HEALTH ECOSYSTEM (THE) under Grant CUP: B83C22003920001, and in part by Spoke 6 BIG DATA E QUANTUM COMPUTING (HPC) under Grant CUP: B83C22002830001.

ABSTRACT Despite technological advances in diagnostic imaging, to distinguish the type of renal tumor without performing a biopsy is still an unsolved challenge. In particular, this is even more striking in the case of clear cell renal cell carcinoma and small oncocytomas. To tackle this problem, a fully automated tool is proposed that can provide decision support for physicians to distinguish between these two types of masses in the most critical cases. In this work three approaches for the development of this tool are implemented and compared, specifically two approaches are based on the use of radiomic features and one on the use of deep features. The nnU-net is exploited to achieve tumor segmentation necessary to obtain the different types of features. The architectures are trained and tested by combining two different datasets, the public dataset KiTS2019 and data from the Careggi University Hospital. The best method is able to obtain 73.77% balanced accuracy, 94.59% sensitivity, 52.94% specificity and 86.84% accuracy.

INDEX TERMS Cancer classification, clear cell renal cell carcinoma, computer aided diagnosis, deep learning, oncocytoma, radiomics.

I. INTRODUCTION

According to GLOBOCAN 2020 estimations, renal tumors are among the most prevalent cancers in the world and reach a mortality rate of 42% [1]. Renal tumor is a particular type of kidney mass that can be either malignant or benign. The majority of malignant renal tumors are RCC, a type of malignant tumor (carcinoma) that originates from the glandular epithelial cells of the kidney and has the potential to invade the surrounding tissues tending to metastasize to other anatomical sites [2], being the cause of 80% of renal cancer deaths [3]. As for benign renal tumors, one of the most common is oncocytoma, which is characterized by the presence of large cells with abundant eosinophilic granular cytoplasm [4]. Although oncocytoma does not entail long-term risks [5], it accounts for approximately 16% of surgically removed renal masses [6] because of its high similarity to clear cell RCC [7]. In fact, currently the standard procedure for the treatment of renal tumor involves the use of diagnostic imaging techniques to determine the presence of suspicious renal masses and their characterization by analyzing the tissues with visual inspection, comparing the mass-related parameters obtainable from the diagnostic image (e.g., size, shape of the mass, etc.) [8]. Unfortunately, distinguishing the nature of a renal mass is a very complex task even for an experienced physician due to the similarity between oncocytomas and RCCs at the radiological imaging level [7]. Therefore, the best current method to diagnose the nature of the tumor is based on histological analysis of the tissue of the mass [9] collected by biopsy after a radical nephrectomy of the kidney containing the mass or a partial nephrectomy of the tumor [10].

Renal tumors are abnormal growths that originate within the kidney tissues and can be classified as either malignant

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Zia Ur Rahman.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

(cancerous) or benign (non-cancerous). Malignant tumors, such as renal cell carcinoma (RCC), have the ability to invade surrounding tissues and metastasize, posing a significant threat to the patient's health. On the other hand, benign tumors like oncocytomas are typically non-invasive and do not spread to other organs. Distinguishing between these tumor types solely based on diagnostic imaging techniques can be challenging, as certain malignant and benign tumors may exhibit similar radiological characteristics. This highlights the importance of developing reliable methods for classifying renal tumors based on imaging data, which could provide valuable decision support for clinicians and lead to more effective treatment strategies. Being able to detect a renal tumor in advance and being able to classify it correctly would be a crucial step, as it would allow the introduction and use of medical procedures that can safeguard the patient's life and renal function [11]. In fact, when a mass is identified with certainty as benign, the surgical solution can be avoided and treatment involves continuous monitoring and control. Therefore, it becomes strategic to be able to implement an analysis tool, which can overcome these limitations. Given the wide use of ML techniques in image analysis and their evolution in the analysis of diagnostic images [12], [13], [14], [15] the idea behind this work is to rely on these techniques for the realization of this tool. ML procedures generally change according to the type of available data. In particular, for medical ML applications usually the available data can be of two types: (i) clinical information, derived from routine clinical examinations, e.g. blood tests and medical history, (ii) information derived from diagnostic images, from specific clinical examinations, such as CT and MRI. With regard to diagnostic images, there are various approaches and the most common is based on identifying, by manual segmentation, the areas of interest within the CT and extracting from it specific features (e.g. texture, size, volume, etc.), named radiomic features. In addition, there is also the possibility of directly using diagnostic images with deep learning algorithms to extract characteristic features, called deep features. While the radiomic features are more interpretable, the effort required in terms of segmentation is considerable and the results can vary depending on how this is performed, as proved in Kocak et al. [16].

This work aims to create a machine learning based tool that can distinguish between benign and malignant tumors by comparing radiomics and deep learning. The data used for this purpose consists of a set of CTs from renal cancer patients. This paper proposes two approaches for the classification of renal tumors, with a special focus on the distinction between oncocytomas and clear cell RCCs, to provide an assisted diagnosis tool. The two approaches were developed and compared, with the goal of creating a tool that is generalizable, updateable with new data, and fully automatic (without requiring manual segmentation of regions of interest). The analysis and comparison of the results were performed with great attention, trying to eliminate bias caused by unbalanced data and incorrect readings of output values. Moreover, the focus was on building an algorithm that can extract both radiomic and deep features in a fully automated framework, exploiting automatic CT segmentation using a Convolutional Neural Network (CNN). The results show that automatic segmentation can be used as a starting point for the extraction of radiomic features for this type of task, and that the deep features are sufficiently representative. In particular, comparable results were obtained with the two types of features, achieving an accuracy above 85% for both methods.

II. RELATED WORKS

Various techniques have been developed to be able to distinguish malignant and benign renal tumors through the use of ML-based techniques that have evolved over time along with ML algorithms and their image processing capabilities [17], [18], [19].

Early works in the literature with significantly relevant results use radiomic features extracted directly from manually segmented areas within diagnostic images such as CT and MRI. Relying on the use of radiomic features, in Raman et al. [20] starting from the CT slice on which the size of the mass is largest, 5 to 10 slices were selected to extract the features to be used with a random forest classifier with the purpose of classifying various solid renal masses (oncocytomas, ccRCC, cysts, and papillary RCC). In order to identify the actual value of texture analysis, in Yu et al. [21] radiomic features are extracted from 10 consecutive CT slices to be used with a support vector machine classifier to analyze the radiomic features related to each of the 10 slices with a majority voting algorithm to decide the actual tumor class. Kunapuli et al. [22] aim to investigate the usefulness of relational statistical machine learning algorithms to differentiate benign and malignant tumors, using radiomic features extracted from CT, analyzing the slice in which the tumor diameter is largest to obtain the two-dimensional features and the entire volume to derive the three-dimensional ones. Sun et al. [23] is among the first research for this specific task to compare the ability in differentiating benign and malignant solid renal masses of experienced radiologists and machine learning algorithms trained on radiomic features extracted from tumor CT by MCNemar test [24]. Unlike other works, Hoang et al. [25], extracts radiomic features from contrast-enhanced MRI images related to 3 consecutive slices to differentiate various types of renal tumors, selected from the section where the tumor area is largest, using them with a random forest classifier. Other research [20], [21], [22], [23], [25] follow a similar pipeline: manual segmentation of the tumor within the diagnostic images by one or more experts with an internal validation by another expert; feature extraction; selection of the most relevant features using specific algorithms (e.g., recursive feature elimination); feature processing, with the aim of normalizing the values (e.g., z-score normalization); classifier training; and finally, validation with algorithms specific to the identified method (k-fold cross validation). Finally, one of the most recent works [26]

TABLE 1. Comparison of state-of-the-art algorithms to distinguish malignant and benign renal tumors through the use of ML-based techniques.

Paper	Model	Data Type	Classification Task	Metrics						
				Sensitivity	*Specificity*	*Accuracy*	*F Score*	*AUC*	*PPV*	*NPV*
[20]	RF	CT Arterial Phase	Onco	81%	97.80%	-	-	-	-	-
			ccRCC	93.50%	95.10%	-	-	-	-	-
	Database	25 ccRCC, 24 Onco								
	Summary	The authors consider individual slices containing the element of interest as a single input. The percentages in the results are relative to the number of total slices and not on the actual number of masses studied. Results refers to the slices of the testing set composed by 4 oncocytomas and 5 ccRCC.								
[21]	SVM (linear)	CT Portal venous phase	ccRCC vs all	-	-	-	-	91%	-	-
			Onco vs all	-	-	-	-	86%	-	-
	Database	46 ccRCC, 63 other RCC, 10 Onco								
	Summary	The authors considered 10 slices containing tumor per CT. From these they extracted 43 texture features to be used with a classifier. The results are reported on all cases despite using 5-fold-cross validation.								
[22]	RFGB	CT - 4 phases	Benign vs Malignant	-	-	82%	87%	83%	-	-
	Database	70 ccRCC, 30 other RCC, 30 Onco, 20 Angio								
	Summary	The authors use the slice where the tumor diameter is largest to obtain 2D features and use the entire volume for 3D ones. This process is repeated for each CT stage, and the obtained features are concatenated to form a single descriptor.								
[23]	SVM	CT Arterial phase	ccRCC vs (Angio and Onco)	86.3%	83.3%	85.8%	91.1%	-	96.5%	53.6%
	Database	190 ccRCC, 64 other RCC, 10 Onco, 26 Angio								
	Summary	The authors analyze radiomic features extracted from tumor CT by MCNemar test. Metrics are considered on the full dataset.								
[25]	RF	MRI	Onco vs ccRCC	67.3%	88.9%	79.3%	-	-	-	-
	Database	90 ccRCC, 22 other RCC, 30 Onco								
	Summary	The authors analyze features extracted from 3 consecutive slices of an MRI starting with the slice where the tumor has the largest diameter								
[27]	Inception V3	CT multiphases	ccRCC vs Onco	88.3%	52.9%	75.4%	-	-	80.3%	-
	Database	128 ccRCC, 51 Onco								
	Summary	The authors create RGB images to be used as network inputs by inserting one slice per phase into the R-G-B channels. Different combinations have been tested varying the phases and the type of slice, but the results refer only to the best ones.								
[29]	LR + 2x ResNet50	MRI	Benign vs Malignant	40.81%	92.06%	69.64%	54%	-	80%	66.7%
	Database	425 ccRCC, 230 others RCC, 92 Onco, 406 Angio								
	Summary	The authors use the 3 slices in which the tumor is largest on the axial, coronal, and sagittal axis, fitting them into the three R-G-B channels to create the images to be used as input. An ensemble model is built using 2 Resnet50s, one trained with MRI T1C and the other with MRI T2WI, and a logistic regression classifier trained with radiomic features extracted from the lesion present in the MRI.								
[32]	DenseAUX Net201	CT	ccRCC vs Onco	98.98%	18.75%	92.92%	96.3%	-	93.7%	60%
	Database	196 ccRCC, 16 Onco								
	Summary	The authors, from the other things done in their work, identify the ROI of the tumors and kidneys and use it as input for a deep learning model to classify the tumors.								
[26]	RF	CT	Malignant vs Benign	87%	61%	-	-	79%	80%	72%
	Database	51 ccRCC, 93 Onco								
	Summary	The authors use radiomics to extract 6 texture features from the ROI manually created of the tumor and use it with a random forest classifier.								
[33]		MRI	ccRCC	84.1%	84.3%	-	-	87.1%	-	-

TABLE 1. *(Continued.)* **Comparison of state-of-the-art algorithms to distinguish malignant and benign renal tumors through the use of ML-based techniques.**

	3D Res-UNet + ResNet50	Onco	76.9%	77.7%	-	-	79.9%	-	-
	Database	203 ccRCC, 16 Onco							
	Summary	After using a deep learning model to identify the ROI of the kidney and the tumor, the authors use the identify region to isolate the kidney and the tumor. Then, they use two parallel model to extract the global features from the ROI image of the kidney and the local features from the tumor. Finally the two type of features are combined and classified into the specific tumor class.							

extracts six radiomics texture-based features (mean, standard deviation, mean value of positive pixels, entropy, asymmetry, and kurtosis) from the ROI of a single-phase CT and use them to train a random forest classifier to distinguish between oncocytoma and ccRCC.

In recent years, thanks to the increase in computing power due to the large-scale deployment and affordability of GPUs, studies began to introduce the use of deep learning algorithms using CNN capable of performing direct image processing. Coy et al. [27] study the diagnostic value and feasibility of a renal tumor classifier based on deep learning, taking advantage of a CNN (Inception v3 [28]), and the four phases of tumor CT. [29] introduces a mixed classifier that is divided into three main branches: one using a logistic regression classifier with radiomic features, and the other two using a CNN (ResNet50 [30]) by taking as input an image related to a slice of the tumor in which the three channels R, G, and B are matched to the axial, sagittal, and coronal gray-scale slice in which the tumor had the largest diameter, following the 2.5D model [31]. Mahmud et al. [32] propose an approach that also involves the use of deep learning models for automatic classification of tumor type, through use of the DenseAUXNet201 model that uses CT relative to the ROI of the kidney with the tumor as input. According to one of the most recent research [33], the authors apply a 3D deep learning network (Res-UNet) to segment regions of the kidney and kidney mass, followed by a dual-pathway classification network that exploits local and global features to classify tumors by focusing on the ROI of the whole kidney and tumor. A compact view of the performance and methods found in the state of the art can be seen in TABLE 1; the performances are given only as a reference since the results are not directly comparable as the used databases are different. For each paper is indicated the used model or classifier, the addressed task, the type of input, the amount of data, and a brief summary of how the data were used and how the results were obtained (in cases where a standard procedure was not followed). In addition, the metrics obtained for each classifier are provided. For consistency, only the results related to classifications that involved oncocytomas and ccRCC are reported.

III. MATERIALS

In this work, a new framework for the study and analysis of diagnostic images for the classification of renal tumors through the use of ML techniques was developed and validated. Specifically, CTs of renal tumor patients with proven histological findings were used, no other data types are included in the study, and all data was provided or retrieved anonymized. It is important to point out that CTs performed with contrast agents are multiphase, i.e., multiple acquisitions are performed, up to a maximum of 4 phases for the kidney, depending on the time elapsed since contrast agent injection: i) before the use of contrast agent unenhanced phase, ii) after 30 seconds corticomedullary (or arterial) phase, iii) after 90 seconds nephrographic phase, iv) excretory phase. Given the possibility of having multiple phases, it is necessary to keep in mind that it is not always possible to have them all available, as they are acquired according to the clinics' individual protocols.

The information was retrieved from two different sources: a private dataset from the (1) ''Careggi'' Florence Hospital (Azienda Ospedaliera Universitaria Careggi - AOUC) [34]; a public dataset from the (2) the 2019 Kidney Tumor Segmentation Challenge (kits2019) [35]. The data from AOUC are related to 61 patients of which 29 with ccRCC and 34 having an oncocytoma, with one or more acquisition phases and histological result. On the other hand, as for the kits2019 data, these are publicly accessible [36], and are related to 300 patients with at least one renal tumor, in particular there is the CT of the arterial phase for each patient, the histological result and some information about the patients, e.g. if they are smokers, if they have other diseases, if they have already had surgery, gender, more information can be found online [37]. Details of the data available for this study can be seen in Table 2.

To standardize the type of available data, the following choices were made: (1) use only diagnostic images, discarding any registry data, and the corresponding histologic result as the ground truth. In particular for data from AOUC, for patients in whom more than one acquisition phase was present, the corticomedullary phase was chosen, being the one known to be most used for diagnostic imaging in the case of CT contrast enhanced; (2) only patients with a single renal tumor of ccRCC or oncocytoma type were considered, discarding any patients with multiple renal tumors or renal tumors of different nature at the same time. As a result, the final database comprises 271 patients. Fig 1 shows the

TABLE 2. Dataset included in the study.

Data source	Number of patients	Data	Description
AOUC	61	CT multiphase	The data are referred to patients with only one kidney tumor that can be an oncocytoma or a ccRCC. In both cases the size is limited to a few millimeters. The data have been acquired using different scanners and different protocols, for this reason for each patient there could be from 1 to 4 phases.
KITS 2019	300	CT corticomedullary	The data are referred to patients with at least one kidney tumor of any kind. The size is heterogeneous, but all the CT are referred to the same phase. The data in this dataset have been acquired from two different clinics and with different scanners.

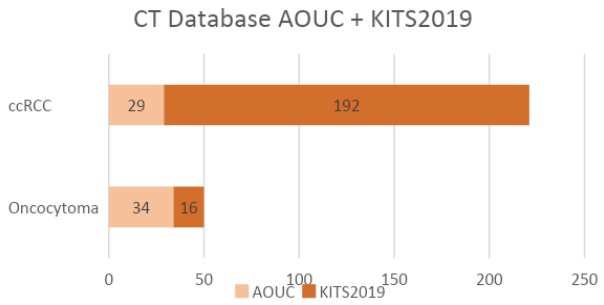

FIGURE 1. CT dataset composition derived from the union of AOUC and KITS2019 data.

database information used in this study after appropriate exclusions.

To help results reproducibility, a GitHub repository has been created (https://github.com/maghero/KidneyTumor Radiomics-DeepFeauters) containing the scripts used for data processing, feature extraction, and model training and evaluation. Instructions and documentation are provided in the repository.

IV. METHODS

At a general level, the structure of the implemented application can be divided into two modules: (i) the feature generator that is in charge of creating, from the patients' CTs, a database of features extracted with specific algorithms; (ii) the classifier in charge of distinguishing the type of the tumor by the analysis of the extracted features.

In particular, in the feature generation process the extraction of two different types of features was implemented, one based on statistical methods (radiomic features) and the other based on deep learning (deep features). Finally, classification algorithms of different nature were examined with the aim of identifying the best performance for feature type.

In the following sections, the radiomic and deep feature extraction methods are explained in detail, specifying with which algorithm and techniques they were obtained and how they were processed before being included in the final database.

A. RADIOMIC FEATURES EXTRACTION – TYPE 1 FEATURES

The main goal of radiomics is to extract quantitative and ideally reproducible features from diagnostic images, thereby including complex patterns that are difficult to recognize or quantify by the human eye [38]. Radiomic features represent properties related to the characteristics of tissues and lesions present in diagnostic images e.g., heterogeneity, shape, etc., and within a sufficiently large dataset these data turn out to be minable, i.e., usable to uncover hidden patterns to discover diseases or implement specific therapies.

Broadly speaking, radiomic features can be divided into statistical features, e.g., histograms, texture analysis, analysis by parameterized models, analysis by transformations (Fourier, Gabor, Haar, etc.), and shape-derived features, which can be extracted either two- or three-dimensionally.

The commonly used procedure for radiomic feature extraction consists of the following steps: (1) deciding which is the region or volume of interest (ROI) (2) segmentation – manual or automated – of the area of interest within each slice under consideration; (3) feature extraction using specific algorithms; and (4) feature post-processing comprising one or more of the existing feature processing techniques, e.g., feature harmonization, selection, and reduction [39]. After these steps the features are ready to be used with a statistical model to accomplish the required task.

Several strategies have been proposed to define the portion of anatomy from which to extract radiomic features. In this work the standard procedure that takes into account the whole region is considered along with the strategy that uses only the slice with maximum tumor size on the three CT planes.

In order to automatically identify and segment ROIs, avoiding time-consuming and non-reproducible processes, it was decided to use a deep learning algorithm. More in detail using the nnU-Net framework [40] it is possible to use a 3D U-Net with pre-trained weights able to obtain a dice score of 0.97 and 0.85 in the task of segmenting the kidney and the kidney tumor.

Regardless of the procedure used to obtain radiomic features, they must go through a pre-processing step before they can be used. This consists of a normalization process according to the following equation

$$X_{new} = \frac{X - \mu}{\sigma} \tag{1}$$

in which the normalized value is obtained by dividing the subtraction between the original value (X) and the mean value of the data (μ) by the value of the standard deviation of the data (σ). This is equivalent to performing a z-score normalization.

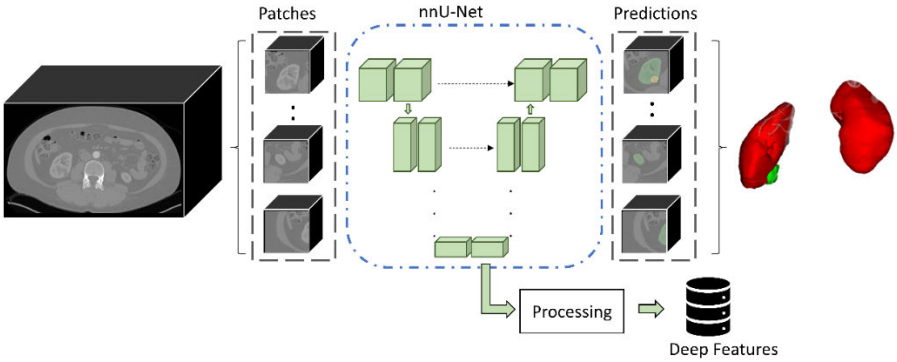

FIGURE 2. Deep feature extraction scheme: the input CT image is elaborated in patches using a nnU-Net pretrained on the task of segmenting kidney and tumors, if in a patch is found the tumor then the corresponding deep features are processed and saved.

To extract the radiomic features it has been exploited the pyradiomics library [41], with which it was possible to extract up to 120 different features given the specific ROI. These features go from first order features, to gray level-based, passing through the shape-based ones. In addition, given the high number of extractable features, a feature selection algorithm, namely recursive feature elimination (RFE), was applied to reduce the number of features to be analyzed and consider only the most significant ones. Specifically a variant of RFE was used which exploits an SVM and has been shown to be more efficient when analyzing biomedical data [42]. Forty-one different sets of radiomics features were created selecting from each case a number of features ranged from a minimum of 5 to a maximum of 45.

B. DEEP FEATURES EXTRACTION - TYPE 2 FEATURES

Deep features correspond to the output of a layer belonging to a deep neural network. The ''depth'' of these features will depend solely on the depth level of the layer from which they are extracted. ''Depth'' is directly proportional to the level of detail and inversely proportional with generalizability; therefore, the deeper the extracted features the greater the level of detail of these features and the lower the level of generalizability.

Taking into consideration the definition of deep features and how they are obtained, it was considered as one of the main strategic elements the choice of the deep neural network to be used for feature creation. A network trained in a task similar or at least related to the one addressed was selected for this purpose. Considering the selection made to obtain the radiomics features, it has been decided to use the same framework and the 3D U-Net model that is already able to identify and segment the kidney and the kidney tumors. This capability allows to obtain deep features that should be able to provide characteristic and specific details useful in classifying the type of tumor. To use this model, it is necessary to use the overall CT of the patient but do not need to be trained.

The framework uses a patch-based approach to handle inputs of any size, by dividing the input CT into multiple patches. Each patch goes through several convolutional blocks, each consisting of a $3 \times 3 \times 3$ convolution with a stride of 2 for downsampling, instance normalization, and a leaky ReLU activation function. The patches reach a minimum size of $4 \times 4 \times 4$ with 320 feature maps at the lowest point of the nnU-Net, which corresponds to the dimensions of the depth, width, height, and number of feature maps respectively. With the aim of identifying the most specific features, the level chosen for feature extraction is the one in which the achieved dimension is the smallest possible and consequently the number of features is also the smallest possible. By doing so, it is possible to obtain features for each one of the patches and consequently for the entire CT. Finally, by reverse mapping the patches using the final output of the network (the segmentation of the tumor and kidney) all features related to patches in which no parts of the tumor are present are discarded, obtaining a final deep features matrix labelled based on the type of the tumor contained in the origin patch. The deep feature extraction procedure just described can be seen in Fig 2.

After obtaining the deep feature matrices, there is still a problem that needs to be solved, which stems from the fact that the number of patches from which the features are extracted can vary depending on the size of the tumor. Therefore, the matrices have to be processed to obtain a single vector of a fixed size that is independent of the size of the CT input to the network. This processing consists of the following steps for each CT: all the feature maps obtained are flattened and concatenated into a single matrix of N rows, where N is the number of valid patches. Then, a statistical operation is applied to reduce the matrix into a single feature vector of 20480 elements, which is used as input for a final classifier.

C. CLASSIFICATION

At this stage, the previously generated data is divided to form two sets, the training set (66,79%), and the testing set

(33,21%). The most common types of classifiers (random forest (RF), k nearest neighbor (K-NN), support vector machine (SVM), artificial neural network (ANN)) were tested.

Concerning the parameters and variances of the classifiers, for RF it was set an internal decision tree number of 1,000, with a minimum number of splits of 2 and Gini impurity [43] as the criterion for measuring the quality of splits. Considering K-NN the number of k neighbors varied between 2 and 8. For SVM it was decided to test the possible variants of the kernels, namely linear, polynomial (by varying the degree from 2 to 5), radial basis function, and sigmoid. Finally, with regard to the artificial neural network an approach based on the following thumb rule on the number of hidden nodes to be used for this type of classifier was used:

$$Hidden_{nodes} \leq \sqrt{Input_{nodes} \times Output_{nodes}} \tag{2}$$

With $Input_{nodes}$ equal to the number of input features and $Output_{nodes}$ equal to 1. The number of internal nodes in the network is limited to reduce the risk of overfitting the model on the training set. Regarding the number of layers and the distribution of these nodes, a trial-and-error strategy was implemented, thanks to which it was possible to identify the best configuration taking into account the trade-off between sensitivity and specificity while looking at the balanced accuracy:

$$Sensitivy = \frac{TP}{TP + FN} \tag{3}$$

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

$$Balanced\ Accuracy = \frac{Sensitivity + Specificity}{2} \tag{5}$$

With TP being the malignant tumors correctly classified, TN the benign tumors correctly classified, FP the benign tumors incorrectly classified, and FN the malignant tumors incorrectly classified.

Table 3 provides a schematic of the strategies used for each one of the classifiers considered. The selection of classifiers and their parameter settings was based on common practices and recommendations from the literature. By exploring a diverse set of classifiers, including Random Forest, k-Nearest Neighbors, Support Vector Machines, and Artificial Neural Networks, we aimed to identify the most suitable approach for the task of renal tumor classification based on the available features (radiomic and deep features). The specific parameter settings for each classifier were chosen to cover a range of reasonable values and configurations, while maintaining computational feasibility.

The training and evaluation workflow followed the standard procedure. Once the various predictors were trained, a statistical analysis of the metrics of interest was performed. Specifically, the most common metrics, i.e. accuracy and precision, were studied, and in addition, sensitivity and specificity were analyzed, as the former allows to get a sense of the predictor's ability to be able to correctly identify malignant tumor [44], the latter refers to the classifier's ability to correctly discard benign tumors. A sensitivity of 100 percent ensures certainty in the case of negativity that a patient is "healthy", conversely a specificity of 100 percent in the case of positivity ensures certainty that the patient is "sick". The goal to be pursued in this study is to seek the best possible trade-off between sensitivity and specificity while trying to obtain a sensitivity that is as close to 100% as possible. This means that the solution sought will have the fewest number of malignant cases misclassified as benign at the expense of benign cases identified as malignant. In other words, the goal is not to misclassify a malignant case, since on a practical level it would bring the greatest disadvantage to the patient.

TABLE 3. Classifier parameters for the training phase.

Classifier	Description
RF	Up to 1000 predictors, with a minimum number of splits of 2, based on Gini impurity criterion for split.
KNN	K ∈ [2,8], Euclidean distance used as weight for the query result.
SVM	Linear, polynomial, radial basis function, and sigmoid kernels used, for the polynomial the grade varied from 2 up to 5.
ANN	Number of nodes decided using the Formula above, the input nodes depends on the number of input features and the number of output nodes is 1, because it is a binary classification task.

V. RESULTS

The results obtained by the use of the described approaches are presented below. All the numerical results refer to the testing set. For the radiomic features, using the total tumor segmentation, TABLE 4 shows the best results achieved by each classifier, specifying the number of features selected by RFE. It can be seen that K-NN, SVM, and RF have higher sensitivity than ANN, reaching about 96%, but lower specificity, where only ANN reaches 58.82% with a sensitivity of 87.84%. The accuracy of the classifiers is similar, ranging from 81% to 85%, and the precision is highest for ANN with 90%. The balanced accuracy is also highest for ANN with 73.31%.

Using only the segmentation of the largest slice in the three dimensions, Table 5 shows the best results achieved by each classifier. As in the previous case, K-NN, SVM, and RF have higher sensitivity than ANN, with results from 87% to 95%. For specificity, the maximum is 56.25% achieved by ANN. The other metrics are not very different from each other, and compared to the previous approach, there is a general decline in performance, except for RF which improves slightly, but still remains lower than the best model in the previous case.

For the deep features, Table 6 shows the results obtained by each classifier. In this case, the sensitivity is above 94% for KNN, RF, and ANN, the specificity peaks with ANN at 52.94%, and the accuracy and precision are between 78% and

TABLE 4. Performance of the classifiers used with the radiomic features of the testing set, considering the approach with complete tumor segmentation.

Classifier	RFE	Sensitivity	Specificity	Accuracy	Precision	Balanced Accuracy
K-NN (K=4)	20	94.59%	41.18%	**84.62%**	87.50%	67.89%
SVM (Linear)	43	94.52%	29.41%	82.22%	85.19%	61.97%
RF	7	**95.95%**	17.65%	81.32%	83.53%	56.80%
ANN	40	87.84%	**58.82%**	82.42%	**90.28%**	**73.31%**

TABLE 5. Performance of the classifiers used with the radiomic features of the testing set related to the approach with single slices of maximum area in the three axes.

Classifier	RFE	Sensitivity	Specificity	Accuracy	Precision	Balanced Accuracy
K-NN (K=2)	9	87.01%	43.75%	79.57%	88.16%	65.38%
SVM (Linear)	39	93.51%	12.50%	79.57%	83.72%	53.00%
RF	31	**94.81%**	31.25%	**83.87%**	86.91%	63.03%
ANN	32	80.26%	**56.25%**	76.09%	**89.71%**	**68.26%**

TABLE 6. Performance of classifiers using deep features obtained from the proposed approach.

Classifier	Sensitivity	Specificity	Accuracy	Precision	Balanced Accuracy
K-NN (K=6)	95.95%	31.25%	84.44%	86.59%	63.60%
SVM (Linear)	86.49%	43.75%	78.89%	87.67%	65.12%
RF	**97.26%**	29.41%	84.44%	85.54%	63.34%
ANN	94.59%	**52.94%**	**86.84%**	**89.74%**	**73.77%**

TABLE 7. Best performance for each of the tested methods, selecting the classifiers with the best-balanced accuracy for the three types of features considered: for all three cases this was found to be the ANN classifier.

Method	Sensitivity	Specificity	Accuracy	Precision	Balanced Accuracy
Radiomic feature Standard procedure	87.84%	**58.82%**	82.42%	**90.28%**	73.31%
Radiomic feature maximum tumor size	80.26%	56.25%	76.09%	89.71%	68.26%
Deep feature	**94.59%**	52.94%	**86.84%**	89.74%	**73.77%**

87% and between 85% and 90%, respectively. The balanced accuracy is also highest for ANN with 73.77%.

VI. DISCUSSION

To further discuss the proposed solutions, TABLE 7 shows the best performance metrics of the classifiers, considering the trade-off between sensitivity and specificity, and selecting the best classifiers based on the maximum value of balanced accuracy.

From the table, we can compare the two solutions that use radiomics: the use of all spatial information results in increased performance for all metrics, especially sensitivity and accuracy, with more than 6 percentage points for both metrics. This result highlights the importance of features derived from the entire tumor mass, which is lost in the second radiomic approach where only the planar information is considered. Comparing the results obtained by radiomic features and deep features, we can note some important aspects: i) considering the trade-off between sensitivity and specificity, the best-balanced accuracy, of 73.77%, is obtained by the deep features, with a high sensitivity of 94.59% and a moderate specificity of 52.94%. This is not the case in the two previous cases, where the best sensitivity is 87.84% and the best specificity is 58.82%. Therefore, considering the balanced accuracy as a tool to balance the trade-off between sensitivity and specificity, we have that the best approaches, with comparable results, are the use of deep features and the use of radiomic features extracted with the first approach.

Objective comparisons with other state-of-the-art approaches are difficult to make due to the diversity of the datasets used, both in terms of size and representation of the tumor types. Another factor to consider is that the numerical results in the state of the art, Table 1, can vary depending on how the data were used. However, a recent study [32] used the KiTS19 dataset so a comparison would be more valid, given that a part of the dataset is the same, but it is also

TABLE 8. Metrics of the best approaches in literature and of the best approach realized in this work.

Method	Metrics						
	Sensitivity	Specificity	Accuracy	F-Score	AUC	PPV	NPV
[27]	88.30%	52.90%	75.40%	-	-	80.30%	-
[32]	98.98%	18.75%	92.92%	96.28%	-	93.72%	60.00%
[26]	87.00%	61.00%	-	-	79.00%	80.00%	72.00%
Our	94.59%	52.94%	86.84%	93.58%	71.34%	89.74%	69.23%

possible to compare the results by checking if the final metrics follow a similar behavior, especially with the solutions using a similar workflow to the one proposed. In Table 8, the metrics for the most similar approaches are reported [26], [27], [32] with the proposed solution ones to make a simpler comparison. The table shows all the available metrics, and it is possible to see how the behavior is similar between the several methods, with sensitivity being the highest metric and specificity the lowest. It is possible to see that looking to the overall metrics the most difficult part of this task is to identify correctly the benign cases, and when more effort is spent in that way, like in the case of Miskin et al [26], the other metrics are affected negatively. Also checking the case that uses a subset of our dataset [32], it is possible to see that if the focus is exclusively on the ccRCC cases the final specificity is strongly penalized becoming quite useless for a real scenario application. If we focus on the metrics that our approach obtains it is possible to see how the sensitivity is the second highest of the four, but it keeps a specificity that is still above the 50% being also the second highest of the shown solutions. Considering the idea of choosing one of the possible alternatives as the best solution, it must be selected one given the trade-off of the several pros and cons, given the high rate of malignant cases, our solution could be selected as the best one, but considering a more uniform distribution the solution proposed by Miskin et al [26] could be picked, but must be considered that it is directly comparable with our results obtained in the case of using radiomics.

A major challenge that we faced during the training of the classifiers was the imbalance of the available dataset, with the number of ccRCC cases being about four times more than the oncocytomas. We solved this problem by using appropriate weights for the elements belonging to the different classes. One of the main future directions of this work is to reduce this imbalance and increase the number of cases as much as possible to make the results more statistically significant. Moreover, with the introduction of new cases, we expect that the deep features-based solution, which relies on feature extraction exclusively based on deep learning, will be able to generalize better across the two tumor types and achieve higher performance than the current one. Other future directions include extending this approach to a larger number of renal tumor types, or using other variants of deep neural networks that can identify other tumors. Additionally, we can also consider combining radiomics with deep learning approaches, or using approaches that require manual delineation of the region of interest to apply convolutional neural networks only on the marked input with the aim of directly classifying the tumor type.

Another important aspect is the clinical focus. The radiomic features are more suitable for clinical applications, because they use well-established and interpretable methods that can reveal the tumor characteristics and behavior. They can also be extracted from any CT scanner, without needing specific hardware or software, which makes them more accessible and adaptable. The deep features, however, depend on the data quality and quantity for training the neural network, and they are less transparent and explainable, which may affect their clinical acceptance and trust. But this does not mean that the deep features are not useful or promising. They have performed similarly or better than the radiomic features in some metrics, and they have the benefit of being fully automated and data-driven, without requiring any prior knowledge or assumptions. They can also capture more complex and subtle patterns that the radiomic features may miss, and they can be adjusted to different tasks and modalities by changing the neural network. Therefore, we believe that the deep features have great potential for clinical applications, especially with more data and explainable methods.

While the obtained specificity of 52.94% is relatively low, it is important to consider the trade-off between sensitivity and specificity in the context of this problem. A high sensitivity, such as the 94.59% achieved by our method, is crucial to minimize the risk of misclassifying malignant tumors as benign, which could have severe consequences for the patient. In a clinical setting, the proposed method could be used as a decision support tool to assist radiologists and clinicians in the initial assessment of suspected renal tumors. Cases classified as malignant could be prioritized for further investigation, while cases classified as benign could undergo additional diagnostic procedures or closer monitoring. Ultimately, the final diagnosis and treatment plan would still rely

on a comprehensive evaluation of all available clinical data, imaging findings, and potentially histopathological analysis.

It is important also to notice that combining datasets from different sources could potentially create a larger benchmark dataset, there are several challenges that make this task difficult in the context of medical imaging data. These include strict data protection regulations, patient privacy concerns, varying imaging protocols and acquisition parameters, and differences in inclusion/exclusion criteria and patient demographics across datasets. Combining datasets with such variations could introduce bias and inconsistencies, potentially negatively impacting the performance of machine learning models trained on the combined dataset. Given the practical challenges and potential risks associated with combining datasets from different sources, we decided to focus on our institutional and public datasets in this study.

VII. CONCLUSION

This work aimed to build a machine learning based tool that can provide support for medical diagnosis by comparing two distinct types of features, radiomic and deep, obtained by automated procedures. For the radiomic features, two state-of-the-art methods were tested. For the deep features, a new method of extraction from patient CTs was proposed, based on the exclusive use of a deep neural network (nnU-net) that can directly segment diagnostic images, minimizing the human intervention required to identify regions of interest. The comparison of the features was useful to understand the feasibility and validity of the proposed method. The performance of the features was evaluated using the most common classification methods (RF, KNN, SVM, ANN). The results showed that the radiomic features achieved better sensitivity than ANN with RF, KNN, and SVM, but at the cost of lower specificity. Therefore, considering balanced accuracy, ANN was the best classifier for both types of features. Moreover, when all the features were considered, ANN achieved the best results with 73.77% balanced accuracy, 94.59% sensitivity, 52.94% specificity, and 86.84% accuracy. Based on these results and considering that the deep features performed similarly or better than the radiomic features, we can conclude that the deep features obtained by the method presented in this article are valid for differentiating renal tumors, especially when only the CT of a patient with renal cancer is available. This can be considered a very important aspect for future works given the possibility and the potential to use only deep features without the necessity of choosing specific radiomics features, reducing the time needed to perform segmentation of the ROI and avoiding the step of features selection.

REFERENCES

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, A Cancer J. Clinicians*, vol. 71, no. 3, pp. 209–249, May 2021, doi: 10.3322/caac.21660.

[2] *Renal Cell Carcinoma | Monarch Initiative*. Accessed: Oct. 25, 2023. [Online]. Available: https://monarchinitiative.org/MONDO

[3] B. D. Ballard and N. Guzman. (2022). *Renal Mass in StatPearls*. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK567761/

[4] *Kidney Oncocytoma | Monarch Initiative*. Accessed: Oct. 25, 2023. [Online]. Available: https://monarchinitiative.org/MONDO

[5] *Oncocytoma | Risk Factors, Signs, Symptoms, Diagnosis and Staging*. Accessed: Oct. 25, 2023. [Online]. Available: https://www.knowcancer.com/oncology/oncocytoma/

[6] F. U. Kay and I. Pedrosa, "Imaging of solid renal masses," *Urologic Clinics North Amer.*, vol. 45, no. 3, pp. 311–330, Aug. 2018, doi: 10.1016/j.ucl.2018.03.013.

[7] *Renal Oncocytoma | Radiopaedia.org*. Accessed: Mar. 13, 2024, doi: 10.53347/rID-1969.

[8] S. Garg and S. R. Bhagyashree, "Detection and classification of tumors using medical imaging techniques: A survey," in *Intelligent Communication Technologies and Virtual Mobile Networks*, vol. 33. Cham, Switzerland: Springer, 2020, pp. 363–372.

[9] S. S. Shen and J. Y. Ro, "Histologic diagnosis of renal mass biopsy," *Arch. Pathol. Lab. Med.*, vol. 143, no. 6, pp. 705–710, Jun. 2019, doi: 10.5858/arpa.2018-0272-ra.

[10] Amer. Cancer Soc. *Kidney Cancer Surgery | American Cancer Society*. Accessed: Oct. 25, 2023. [Online]. Available: https://www.cancer.org/cancer/types/kidney-cancer/treating/surgery.html

[11] C. L. Vendrami, C. P. Villavicencio, T. J. DeJulio, A. Chatterjee, D. D. Casalino, J. M. Horowitz, D. T. Oberlin, G.-Y. Yang, P. Nikolaidis, and F. H. Miller, "Differentiation of solid renal tumors with multiparametric MR imaging," *Radiographics*, vol. 37, no. 7, pp. 2026–2042, Nov. 2017.

[12] C. Martin-Isla, V. M. Campello, C. Izquierdo, Z. Raisi-Estabragh, B. Baeßler, S. E. Petersen, and K. Lekadir, "Image-based cardiac diagnosis with machine learning: A review," *Frontiers Cardiovascular Med.*, vol. 7, pp. 1–15, Jan. 2020, doi: 10.3389/fcvm.2020.00001.

[13] J. Latif, C. Xiao, A. Imran, and S. Tu, "Medical imaging using machine learning and deep learning algorithms: A review," in *Proc. 2nd Int. Conf. Comput., Math. Eng. Technol.*, Jan. 2019, pp. 1–5, doi: 10.1109/ICOMET.2019.8673502.

[14] A. Saha, S. Tso, J. Rabski, A. Sadeghian, and M. D. Cusimano, "Machine learning applications in imaging analysis for patients with pituitary tumors: A review of the current literature and future directions," *Pituitary*, vol. 23, no. 3, pp. 273–293, Jun. 2020.

[15] P. Windisch, C. Koechli, S. Rogers, C. Schroder, R. Forster, D. R. Zwahlen, and S. Bodis, "Machine learning for the detection and segmentation of benign tumors of the central nervous system: A systematic review," *Cancers*, vol. 14, no. 11, p. 2676, Jun. 2022.

[16] B. Kocak, E. Ates, E. S. Durmaz, M. B. Ulusan, and O. Kilickesmez, "Influence of segmentation margin on machine learning–based high-dimensional quantitative CT texture analysis: A reproducibility study on renal clear cell carcinomas," *Eur. Radiol.*, vol. 29, no. 9, pp. 4765–4775, Sep. 2019, doi: 10.1007/s00330-019-6003-8.

[17] R. Magherini, E. Mussi, Y. Volpe, R. Furferi, F. Buonamici, and M. Servi, "Machine learning for renal pathologies: An updated survey," *Sensors*, vol. 22, no. 13, p. 4989, Jul. 2022, doi: 10.3390/s22134989.

[18] B. Kocak, E. A. Kus, A. H. Yardimci, C. T. Bektas, and O. Kilickesmez, "Machine learning in radiomic renal mass characterization: Fundamentals, applications, challenges, and future directions," *Amer. J. Roentgenology*, vol. 215, no. 4, pp. 920–928, Oct. 2020, doi: 10.2214/ajr.19.22608.

[19] A. Bhandari, M. Ibrahim, C. Sharma, R. Liong, S. Gustafson, and M. Prior, "CT-based radiomics for differentiating renal tumours: A systematic review," *Abdominal Radiol.*, vol. 46, no. 5, pp. 2052–2063, May 2021, doi: 10.1007/s00261-020-02832-9.

[20] S. P. Raman, Y. Chen, J. L. Schroeder, P. Huang, and E. K. Fishman, "CT texture analysis of renal masses," *Academic Radiol.*, vol. 21, no. 12, pp. 1587–1596, Dec. 2014, doi: 10.1016/j.acra.2014.07.023.

[21] H. S. Yu, J. Scalera, M. Khalid, A. S. Touret, N. Bloch, B. Li, M. M. Qureshi, J. A. Soto, and S. W. Anderson, "Texture analysis as a radiomic marker for differentiating renal tumors," *Abdominal Radiol.*, vol. 42, no. 10, pp. 2470–2478, Oct. 2017.

[22] G. Kunapuli, B. A. Varghese, P. Ganapathy, B. Desai, S. Cen, M. Aron, I. Gill, and V. Duddalwar, "A decision-support tool for renal mass classification," *J. Digit. Imag.*, vol. 31, no. 6, pp. 929–939, Dec. 2018, doi: 10.1007/s10278-018-0100-0.

[23] X.-Y. Sun, Q.-X. Feng, X. Xu, J. Zhang, F.-P. Zhu, Y.-H. Yang, and Y.-D. Zhang, "Radiologic-radiomic machine learning models for differentiation of benign and malignant solid renal masses: Comparison with expert-level radiologists," *Amer. J. Roentgenology*, vol. 214, no. 1, pp. W44–W54, Jan. 2020.

[24] H. I. Brunner and E. H. Giannini, "Trial design, measurement, and analysis of clinical investigations-," in *Textbook Pediatric Rheumatology*. Saunders, 2011, pp. 127–156.

[25] U. N. Hoang, S. M. Mirmomen, O. Meirelles, J. Yao, M. Merino, A. Metwalli, W. M. Linehan, and A. A. Malayeri, "Assessment of multiphasic contrast-enhanced MR textures in differentiating small renal mass subtypes," *Abdominal Radiol.*, vol. 43, no. 12, pp. 3400–3409, Jun. 2018.

[26] N. Miskin, L. Qin, S. G. Silverman, and A. B. Shinagare, "Differentiating benign from malignant cystic renal masses: A feasibility study of computed tomography texture-based machine learning algorithms," *J. Comput. Assist. Tomogr.*, vol. 47, no. 3, pp. 376–381, May 2023, doi: 10.1097/rct.0000000000001433.

[27] H. Coy, K. Hsieh, W. Wu, M. B. Nagarajan, J. R. Young, M. L. Douek, M. S. Brown, F. Scalzo, and S. S. Raman, "Deep learning and radiomics: The utility of Google TensorFlowTM inception in classifying clear cell renal cell carcinoma and oncocytoma on multiphasic CT," *Abdominal Radiol.*, vol. 44, no. 6, pp. 2009–2020, Jun. 2019.

[28] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Dec. 2015, pp. 2818–2826.

[29] I. L. Xi, Y. Zhao, R. Wang, M. Chang, S. Purkayastha, K. Chang, R. Y. Huang, A. C. Silva, M. Vallieres, P. Habibollahi, Y. Fan, B. Zou, T. P. Gade, P. J. Zhang, M. C. Soulen, Z. Zhang, H, X. Bai, and S. W. Stavropoulos, "Deep learning to distinguish benign from malignant renal lesions based on routine MR imaging," *Clin. Cancer Res.*, vol. 26, no. 8, pp. 1944–1952, Apr. 2020.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[31] K. Chang, H. X. Bai, H. Zhou, C. Su, W. L. Bi, and E. Agbodza, "Residual convolutional neural network for the determination of IDH status in low- and high-grade gliomas from mr imaging," *Clin. Cancer Res.*, vol. 24, no. 5, pp. 1073–1081, Mar. 2018.

[32] S. Mahmud, T. O. Abbas, A. Mushtak, J. Prithula, and M. E. H. Chowdhury, "Kidney cancer diagnosis and surgery selection by machine learning from CT scans combined with clinical metadata," *Cancers*, vol. 15, no. 12, p. 3189, Jun. 2023, doi: 10.3390/cancers15123189.

[33] J. Liu, O. Yildirim, O. Akin, and Y. Tian, "AI-driven robust kidney and renal mass segmentation and classification on 3D CT images," *Bioengineering*, vol. 10, no. 1, p. 116, Jan. 2023, doi: 10.3390/bioengineering10010116.

[34] *Careggi*. Accessed: Oct. 25, 2023. [Online]. Available: https://www.aou-careggi.toscana.it/internet/index.php?lang=it

[35] N. Heller et al., "The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 challenge," 2019, *arXiv:1912.01054*.

[36] (2021). *Neheller/Kits21: The Official Repository of the 2021 Kidney and Kidney Tumor Segmentation Challenge*. Accessed: Oct. 25, 2023. [Online]. Available: https://github.com/neheller/kits21

[37] N. Heller, N. Sathianathen, A. Kalapara, E. Walczak, K. Moore, H. Kaluzniak, J. Rosenberg, P. Blake, Z. Rengel, M. Oestreich, J. Dean, M. Tradewell, A. Shah, R. Tejpaul, Z. Edgerton, M. Peterson, S. Raza, S. Regmi, N. Papanikolopoulos, and C. Weight, "The KiTS19 challenge data: 300 kidney tumor cases with clinical context, CT semantic segmentations, and surgical outcomes," 2019, *arXiv:1904.00445*.

[38] M. E. Mayerhoefer, A. Materka, G. Langs, I. Häggström, P. Szczypiński, P. Gibbs, and G. Cook, "Introduction to radiomics," *J. Nucl. Med.*, vol. 61, no. 4, pp. 488–495, Apr. 2020, doi: 10.2967/jnumed.118.222893.

[39] W. Jia, M. Sun, J. Lian, and S. Hou, "Feature dimensionality reduction: A review," *Complex Intell. Syst.*, vol. 8, no. 3, pp. 2663–2693, Jan. 2022, doi: 10.1007/s40747-021-00637-x.

[40] F. Isensee and K. H. Maier-Hein, "An attempt at beating the 3D U-Net," 2019, *arXiv:1908.02182*.

[41] J. J. M. Van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. H. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, and H. J. W. L. Aerts, "Computational radiomics system to decode the radiographic phenotype," *Cancer Res.*, vol. 77, no. 21, pp. e104–e107, Nov. 2017.

[42] H. Sanz, C. Valim, E. Vegas, J. M. Oller, and F. Reverter, "SVM-RFE: Selection and visualization of the most relevant features through non-linear kernels," *BMC Bioinf.*, vol. 19, no. 1, pp. 1–18, Dec. 2018.

[43] Y. Yuan, L. Wu, and X. Zhang, "Gini-impurity index analysis," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 3154–3169, 2021, doi: 10.1109/TIFS.2021.3076932.

[44] D. G. Altman and J. M. Bland, "Statistics notes: Diagnostic tests 1: Sensitivity and specificity," *BMJ*, vol. 308, no. 6943, p. 1552, Jun. 1994, doi: 10.1136/bmj.308.6943.1552.

ROBERTO MAGHERINI was born in Empoli, Florence, Italy, in 1995. He received the bachelor's degree in computer engineering from the University of Pisa in 2017, the master's degree in 2020, and the Ph.D. degree in industrial engineering from the Department of Industrial Engineering, University of Florence, Florence, in 2024.

Since 2020, his research interests have been focused on the applications of artificial intelligence across various domains. His particular emphasis is on the biomedical and industrial fields, especially in the areas of automatic classification, detection, and segmentation of images using deep learning models.

MICHAELA SERVI received the master's degree in computer engineering from the School of Engineering, University of Florence, Florence, Italy, in 2015, and the Ph.D. degree in industrial engineering from the Department of Industrial Engineering, University of Florence, developing a research project focused on the development of procedures for the realization of patient-specific medical devices. Her master's thesis was titled "Disparity Coherent Stereo Video Watermarking."

She is currently an RTDA Researcher with the Department of Industrial Engineering, University of Florence, within the SSD ING-IND/15. Her research interests include reverse engineering, computer vision, software engineering, CAD, 3D printing, and tools and methods to support personalized medicine. She has authored more than 50 publications in these scientific areas.

YARY VOLPE received the M.Sc. degree in mechanical engineering from the University of Florence, Italy, in 2002, and the Ph.D. degree in machine designs and construction, in 2006.

After working as a Postdoctoral Researcher with the Department of Mechanics and Industrial Technologies, University of Florence (DIEF), in 2020, he assumed the faculty position as an Associate Professor in "Design and Methods of Industrial Engineering." He has coordinated the T3Ddy lab. (joint laboratory between Meyer Children's Hospital and DIEF), since 2016. He is the author and coauthor of more than 150 publications printed in international journals and participated in a number of international conferences. His research interests include computer-aided design, image processing, virtual prototyping, FE simulation, reverse engineering, and rapid prototyping.

RICCARDO CAMPI received the Diploma degree in laparoscopic surgery from the University of Strasbourg, France, and the M.D. and Ph.D. degrees. He was a Research Fellow in urologic oncology with La Pitié-Salpêtrière Academic Hospital, Paris, France. He is currently a Consultant Urologist with the Unit of Urologic Robotic, Minimally-Invasive Surgery and Renal Transplantation, Careggi University Hospital, Florence, Italy, one of the highest-volume tertiary referral Cancer Centre in Italy. He received the National Scientific Qualification for eligibility as an Associate Professor and a Full Professor in urology by the Italian Ministry of University and Research. He has authored numerous editorials, book chapters, and over 280 publications in peer-reviewed journals. He has an active research portfolio in the field of urologic oncology. He is actively involved in several large studies, including international trials in prostate and renal cancer. His main clinical and academic interests are renal cancer, with a special emphasis on both the diagnosis and management of localized renal masses, renal transplantation, and minimally invasive surgery. He is an active member of several national and international societies. He is currently the Chairman of the European Association of Urology (EAU) Young Academic Urologists (YAU) Renal Cancer working group and an Associate Member of the EAU Guidelines Panel on Penile Cancer and Renal Cell Carcinoma. He is also an Associate Member of the EAU Section of Oncological Urology (ESOU) and a member of the Italian Society of Urology (SIU) Research Office. He is a Supporting Editor of *European Urology Oncology*. He serves as a reviewer and an editorial board member for high-ranked journals in the field of urology.

FRANCESCO BUONAMICI received the master's degree in mechanical engineering from the School of Engineering, University of Florence, Italy, in 2014, and the Ph.D. degree in industrial engineering from the Department of Industrial Engineering, University of Florence, developing a research project focused on a methodology for reconstructing CAD models from scanned data. His master's thesis was titled ''Study of an Automatic System to Assist Blind People in the Tactile Exploration of Works of Art.'' He is currently an RTDA Researcher with the Department of Industrial Engineering, University of Florence, within the SSD ING-IND/15. His research interests include reverse engineering, design for additive manufacturing, CAD, 3D printing, topological optimization, and tools and methods to support personalized medicine. He has authored more than 50 publications in these scientific areas.

Open Access funding provided by 'Università degli Studi di Firenze' within the CRUI CARE Agreement