IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 4 July 2024, accepted 28 July 2024, date of publication 8 August 2024, date of current version 23 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3440259

RESEARCH ARTICLE

Fully Automated Deep Learning-Based Renal Mass Detection on Multi-Parametric MRI

ROHINI GAIKAR[1], AZAR AZAD[2], NICOLA SCHIEDA[3], AND ERANGA UKWATTA[1], (Senior Member, IEEE)

[1]School of Engineering, University of Guelph, Guelph, ON N1G 2W1, Canada
[2]A.I. VALI Inc., Toronto, ON M4N 3N1, Canada
[3]Department of Radiology, The University of Ottawa, Ottawa, ON K1H 8L6, Canada

Corresponding author: Rohini Gaikar (rgaikar@uoguelph.ca)

The work of Rohini Gaikar was supported by Ontario Graduate Scholarship (OGS) for 2021–2023. The work of Eranga Ukwatta was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant for the research work.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the University of Guelph Research Ethics Board (CUREB) and the Ottawa Hospital Research Institute Research Ethics Board (REB).

ABSTRACT Due to superior soft tissue contrast afforded by magnetic resonance imaging (MRI), there is great potential for multi-parametric MRI (mpMRI) for the detection and eventual classification of renal masses (RMs). In this study, we investigated fully automated deep learning methods for RMs detection using T2-Weighted (T2W) spin-echo and two contrast-enhanced T1-Weighted gradient-echo-corticomedullary (T1W-CM), nephrographic-phase (T1W-NG), T1-Weighted In-phase (T1W-IP) and opposed-phase (T1W-OP) images. The dataset contained mpMRI images of 108 kidney cancer patients with an average size of renal mass of 24 ± 7.8 cm. In the first stage, kidneys were segmented using a 2D attention U-Net model, which was reported in a previous study. In the second stage, we tested five different state-of-the-art methods for RMs detections on mpMRI sequences. The model predictions were compared to manual annotations using precision, recall, specificity, and Dice Similarity Coefficient (DSC). The best-performing deep learning models were U-Net, U-Net++, and attention U-Net on the T2W, T1W-CM, and T1W-NG sequences respectively. Of the 5 mpMRI sequences, we also demonstrated that the T1W-CM is the most suitable for RMs detection. This automated detection of RMs in mpMRI sequences may be useful for the subsequent characterization of RMs in a fully automated artificial intelligence-based pipeline.

INDEX TERMS Multi-parametric MRI, kidney cancer, computer-aided detection of cancer, deep learning.

I. INTRODUCTION

The incidence of renal cell carcinoma (RCC), has doubled over the past fifty years [1]. It was estimated that 87,100 adults will be diagnosed with kidney cancer in the North American continent in 2022 [2], [3]. The increase in kidney cancer diagnosis is due to the incidental discovery of renal masses (RMs) in patients undergoing cross-sectional imaging studies for other reasons [4]. This has increased surgeries to treat RMs but it has not decreased kidney cancer-specific mortality in the same period and thus, there may currently be an overdiagnosis and overtreatment of RMs in clinical practice [5], [6]. Moreover, in cystic masses, benign diagnosis is common and in solid RMs $<$ 4cm in size, approximately 20% are benign [7]. A desire to diagnose the specific subtype of RMs before treatment is therefore desired. For cystic masses, the Bosniak Classification provides an accurate estimate of the probability of malignancy [7]; however, no such system exists for solid RMs. Indeed, pre-operative diagnosis of the histology of a solid RM can only be achieved by percutaneous renal mass biopsy which is invasive, associated with low risk of complications, not possible in all patients, and can be non-diagnostic in up to 20% of cases [7].

Imaging diagnosis of solid RMs subtypes is desirable. Computed tomography (CT) is commonly used as the clinical standard for imaging RMs; however, provides limited

The associate editor coordinating the review of this manuscript and approving it for publication was Jason Gu.

information. The multi-parametric (defined as more than one parametric pulse sequence) nature of magnetic resonance imaging (MRI), has shown promise for the characterization of solid RMs [8], [9]. Comprehensive mpMRI of the kidneys includes: T2-weighted (T2W), T1-weighted (T1W) chemical shift imaging, and, dynamic contrast-enhanced (DCE) T1W imaging [10]. Nevertheless, studies evaluating the use of CT or mpMRI scoring systems for solid RMs subtyping have shown only modest accuracy and inter-observer agreement [11], [12].

Artificial intelligence (AI) has the potential to not only speed up but also improve the diagnostic capabilities of CT and MRI for renal mass diagnosis [13], [14]. Machine learning-based quantitative texture analysis of CT images has been extensively studied [15], [16]. Although RMs detection on a single MRI sequence may have been investigated [17], but different mpMRI sequences have not been compared.

Chen et al. [18] implemented a three-layer perceptron model to discriminate between manually segmented low-grade and high-grade clear cell renal cell carcinomas (cc-RCC) with an accuracy of 86.2% using texture features of renal masses on contrast-enhanced MRI images. Dwivedi et al. [19] performed radiomics analysis on manual tumor segmentation of T2W images and dynamic contrast-enhanced (DCE) MRI images were superior to tumor size for the prediction of high-grade histology in cc-RCC. Said et al. [20] implemented MRI-based radiomics feature classification using a random forest ML model to characterize solid renal neoplasms on T2W, T1W pre- and post-contrast, and DWI. Canvasser et al. [21] evaluated the diagnostic performance of MRI in cT1a renal masses prediction of size less than 4 cm using regression models on 11 MR imaging features achieved accuracies of 81% and 91% in the diagnosis of cc-RCC and papillary RCC (p-RCC) respectively. Xi et al. [22] studied a deep learning residual convolutional neural network (CNN) model and radiomics feature-based bagging classifier for noninvasive differentiation of renal tumors from RCC on T2W and T1-post contrast MRI sequences. A study on the differentiation of RCC subtypes by Wang et al. [23] suggested that multiple MRI sequences could help to differentiate the three subtypes of RCC. These renal mass characterization studies [18], [19], [20] and RCC subtyping [21], [22], [23] were semi-automated and required manual labeling by an expert.

A deep learning-based 2D U-Net designed by Aslam et al. [24] was used to auto-segment the renal cortex and medulla of chronic kidney disease (CKD) allograft kidney T1 maps. Daniel et al. [25] developed a CNN model for automated renal segmentation in healthy and chronic kidney disease subjects to calculate total kidney volume. Automatic semantic segmentation of kidney cysts in MR images of patients affected by autosomal-dominant polycystic kidney disease was developed by Kline et al. [26] using a CNN model. Kim et al. [27] developed an automated segmentation of kidneys from MRI in patients with autosomal dominant polycystic kidney disease. Recently, Anush et al. [17] proposed an automated RMs detection method for a T1W-nephrographic contrast-enhanced (T1W-NG) mpMRI sequence using a mixture of ensembles of deep learning models. Although kidney segmentation on MRI has been explored using deep-learning models [28] and cascaded CNN [29], fully-automated RMs segmentation and detection on MRI images is still a clinically open challenge.

In this paper, we presented the first comprehensive study of automated RMs detection on different MRI sequences, including T2W, T1W-CM, T1W-NG, T1W-IP, and T1W-OP mpMRI. We also identified T1W-CM as the most suitable mpMRI for the detection of RMs. We included a diverse spectrum of RMs including renal cell carcinoma (RCC) subtypes, benign fat-poor angiomyolipoma (fpAML), and oncocytoma. The models were evaluated based on precision, recall, and specificity in the detection of RMs along with the average Dice Similarity Coefficient (DSC) of RMs segmentation.

II. MATERIALS AND METHODS

The algorithm developed for the detection of RMs consists of two stages as shown in Figure 1. First, we segmented the kidneys using separately trained attention U-Net models for three mpMRI images: T2W, T1W-CM, and T1W-NG [28]. In the second stage, a separate CNN model was trained on each mpMRI sequence for RMs detection within kidney ROIs. The segmented kidney mask was used as a second channel to prepare the training images for the detection of the RMs using a deep-learning model. This is called a 'spatial prior' to guide the CNN model to predict within the kidney ROIs which excludes false positive detections in non-kidney tissues [30]. This 2-stage RMs detection was implemented separately on T2W, T1W-CM, T1W-NG, T1W-IP, and T1W-OP mpMRI images.

A. STUDY SUBJECTS AND IMAGE ACQUISITION

This study was approved by the joint review boards of the University of Guelph Research Ethics Board (CUREB) and Ottawa Hospital Research Institute Research Ethics Board (REB). The mpMRI scans were acquired using Siemens, GE Healthcare, and Phillips clinical MRI systems with magnetic field strengths of 1.5T and 3T. The imaging protocols for the T2W, T1W-CM, and T1W-NG, mpMRI sequences are provided in Appendix. Our dataset comprised 108 patients who underwent partial or total nephrectomy for RMs between January 1, 2015, and December 31, 2017. We excluded RMs $<$ 1cm in size because at this size threshold imaging characterization becomes difficult and RMs of this size are generally shrivelled [30]. We further excluded patients with observed genetic predisposition to develop cystic or solid RMs namely Autosomal Dominant Polycystic Kidney Disease, Tuberous Sclerosis, Von Hippel Lindau disease, Birt Hogg Dube Syndrome, Hereditary Papillary RCC syndrome who are not representative of the general population in which incidental renal masses are detected [31]. The selected 108 patients with 119 histologically confirmed RMs of mean volume $12.20 \pm 11.92\ cm^3$ had diagnoses of 64 clear cell RCC,

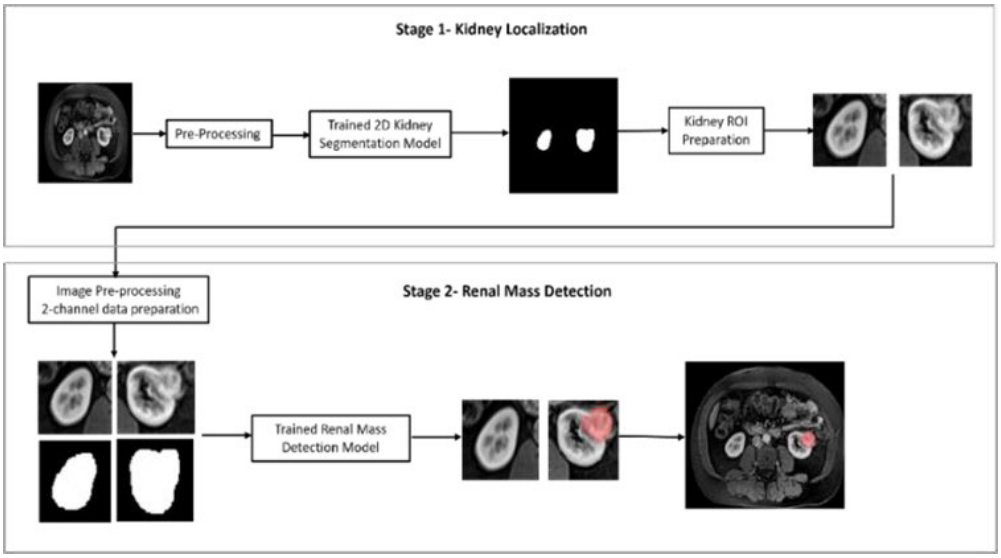

FIGURE 1. An overview of the proposed two-stage methodology for renal mass detections on T2W, T1WCM, T1W-NG, T1W-IP, and T1W-OP mpMRI sequences where stage 1 is kidney localization and stage 2 are renal mass detection.

18 papillary RCC, 2 chromophobe RCC, 13 oncocytomas, 10 fpAMLs, 1 metastasis, 1 metanephric adenoma, 1 epithelioid AML and 9 masses of unknown histology observed during manual segmentation. It is common to have incidental benign simple or minimally complex Bosniak Type 1 or 2 renal cysts in the images of solid renal mass patients [32], therefore in a cohort of 108 patients, 61 patients were identified with synchronous Bosniak Type 1 or 2 renal cysts along with solid RMs.

B. MANUAL SEGMENTATION OF THE KIDNEYS AND RMs

The kidney boundaries in the dataset were manually segmented slice-by-slice by a fellowship-trained abdominal radiologist. The same radiologist identified and segmented the RMs in each MR image correlating the location of each renal mass from surgical and pathological reports. RMs without any histological confirmation were segmented but no diagnosis was made. Benign cystic RMs of size $\geq$10 mm in size were also segmented. Manual segmentation was performed using ITK-SNAP v3.2 [33] and saved in NIFTI format.

C. AUTOMATED KIDNEY SEGMENTATION

Owing to the differences in the acquisition protocols of the mpMRI sequences, there is substantial variation in the appearance of RMs in T2W, T1W-CM, T1W-NG, and T1W-In Phase (T1W-IP) images as shown in Figure 2. Therefore, a generalized deep-learning model for all mpMRI sequences [34], [35] was unavailable. We trained separate 2D attention U-Net-based CNN models for kidney segmentation on the T2W, T1W-CM, and T1W-NG mpMRI images. These images were randomly divided into training (N = 86) and test (N = 22) datasets. The 2D axial slices were resized to 256×256 and preprocessed using histogram equalization to enhance the local contrast in the images. The training images were augmented using an elastic deformation technique to increase the amount of training data. We used binary cross entropy as a loss function along with the ADAM optimizer to train the model. The kidney segmentation model was trained using five-fold cross-validation techniques on the training data of 86 patient cases and validated on unseen test data of 22 patient cases. This automated kidney segmentation work was published earlier [28] and we used these kidney segmentations to prepare the kidneys as an ROI for the detection of RMs.

D. DATA PRE-PROCESSING FOR RMs DETECTION

When the kidney boundaries were segmented on mpMRI images, the left and right kidney ROIs of size 128×128 were cropped from the 2D axial slice of the original images. The training and test data for RMs detection were prepared on T2W, T1W-CM, and T1W-NG mpMRI images. The RMs detection was the second stage of the model implementation, therefore the training (N = 86) and test (N = 22) datasets were prepared using the same patient cases as used in Stage 1 of kidney segmentation. The extracted kidney ROI images were preprocessed using histogram equalization, and the image intensities were normalized to the range of 0 to 1. The training data and test data were prepared such that the extracted kidney ROI image was the first channel, and the corresponding segmented kidney mask was used as the second channel, which was a spatial-prior approach to minimize false predictions in non-kidney tissues. The training data were biased towards healthy kidney slices compared to slices containing RMs. The extracted kidney ROIs containing RMs in them were augmented with random rotation,

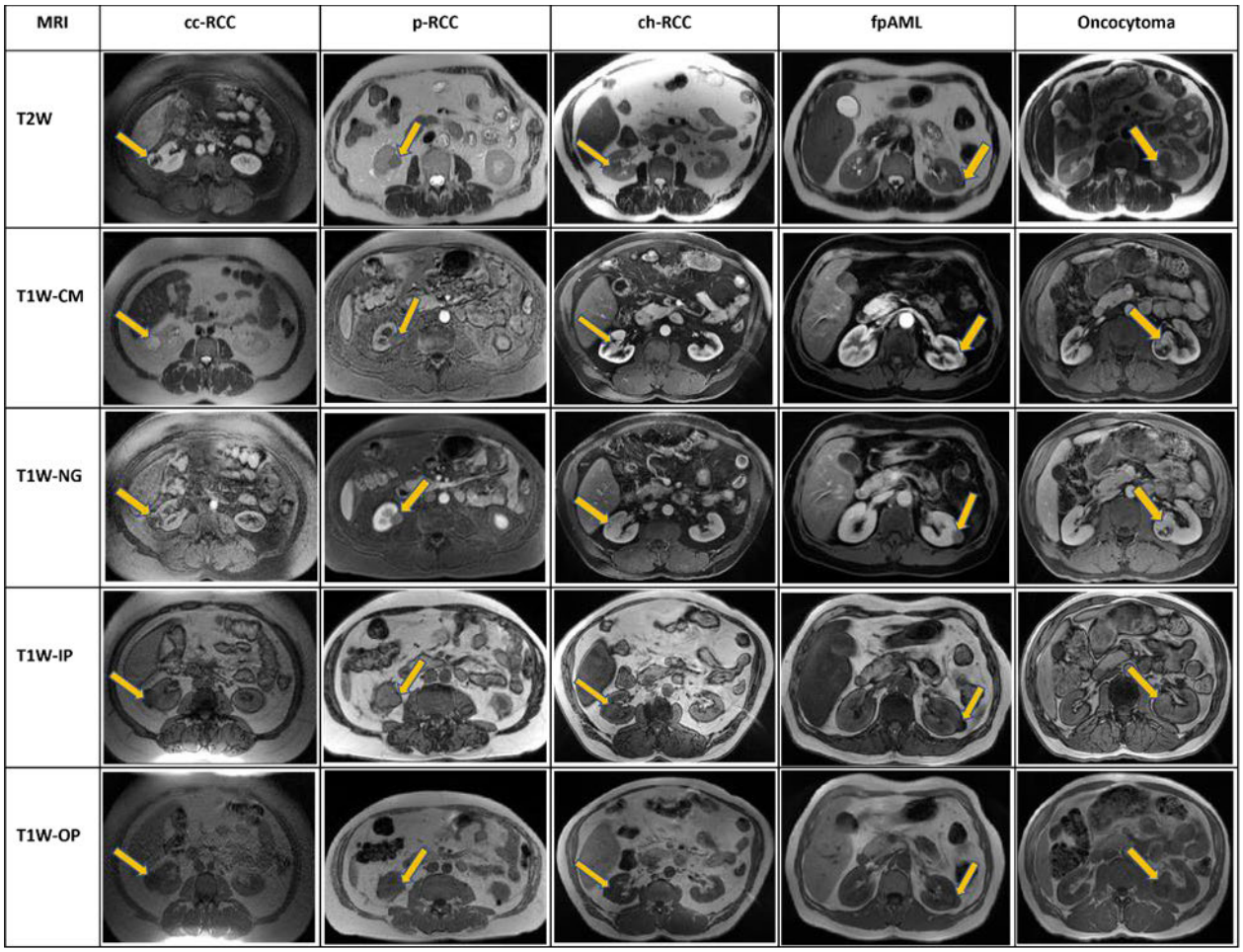

FIGURE 2. Sample axial slices showing malignant (cc-RCC, p-RCC, ch-RCC) and benign (fpAML and oncocytoma) types of SRMs for the same patient acquired with different acquisition protocols namely T2W, T1W-CM, T1W-NG, T1W-IP, and T1W-OP mpMRI sequences. The yellow arrow points to the renal mass in each axial slice.

vertical and horizontal flips, random brightness contrast, and shift-scale rotation using the fast and flexible image augmentation Albumentation library [36].

E. RMs DETECTION ON T2W, T1W-NG, T1W-IP, AND T1W-OP

To develop an appropriate RMs detection model for each mpMRI, we used five-fold cross-validation where each fold contained approximately 17 patient cases and testing it on 22 unseen test data. The training and test datasets in RMs detection were similar to automated kidney segmentation. We trained and tested five different CNN models, including U-Net, U-Net++, Attention U-Net, SegNet, and pre-trained RESNET34 as a backbone in the U-Net model separately on each mpMRI sequence. These models were implemented using CNN layers and trained using an ADAM optimizer with a combination loss function. The combination loss function was used for class imbalance by leveraging the benefits of binary cross entropy (BCE) and Dice Loss [37] The loss function is given by Equation (1).

$$L(y,\hat{y}) = 0.5 * \left(\frac{-1}{N}\sum_i \left[y_i \log\left(\hat{y}_i\right) + (1-y_i)\log\left(1-\hat{y}_i\right)\right]\right) - \left(1 - \frac{\left(2\sum_i y_i\hat{y}_i + 1\right)}{\left(\sum_i y_i + \sum_i \hat{y}_i + 1\right)}\right) \quad (1)$$

where, N is the batch size, y_i is the ground truth pixel value of i, and $\hat{y}_i$ is the predicted pixel value of the pixel. i The first term in Equation (1) is the BCE loss function which compares each of the predicted probabilities to the actual class output which can be either 0 or 1. It penalizes the probabilities based on the distance from the expected value i.e., how close, or far from the actual value. A class imbalance exists when segmenting a small foreground from a large context or background. A combination loss function leverages the dice function that handles the input class imbalance. It controls the trade-off between false positives (FP) and false negatives (FN) by enforcing smooth training using cross-entropy. The

combination loss function is a weighted sum of modified cross-entropy and a dice loss function to encode the learning as shown in Equation (1). One is added to the numerator and denominator of the second term in Equation (1) to ensure that the function has defined edge-case scenarios like $y = \hat{y} = 0$

The models were trained for 50 epochs with model checkpoints to save the best-trained model. The final output layer in each model used a sigmoid activation function to provide a probability map for each pixel. We applied a threshold of 0.5 to classify each pixel as background when the probability was < 0.5 and renal mass or foreground when the probability was > 0.5.

Of the five different trained models, the best model was proposed based on the highest precision, recall, specificity, and average DSC of the RMs segmentation. Deep-learning algorithms were implemented in Python using Keras on top of TensorFlow. All models were trained and evaluated on GPU-accelerated high-performance computers of the Shared Hierarchical Academic Research Computing Network (SHARCNET), a partner organization of the Digital Research Alliance of Canada.

1) STANDARD U-NET MODEL

We used the standard U-Net model as the basic segmentation model owing to its effectiveness in segmenting medical images [38]. Modification regarding the number of filters and network layers was done on the original U-Net model to reduce the network complexity and improve segmentation performance. The U-Net model designed was five-layer deep with 16 filters in the first layer and doubled them at each consecutive layer. We used kernel size of 3×3 with 'RELU' activation with HE_Normal [39] as kernel initializer and 'same' padding at each convolution layer, to minimize overfitting of the model we used a dropout of 0.1 at the initial two layers and then 0.3 afterward. We used the MaxPooling layer after each convolution layer in the encoder path of the U-Net model. The combination loss function with a learning rate of 0.0001 for the ADAM optimizer used in the training. The input to the U-Net model is a 2-channel input with kidney ROI as a first channel and corresponding kidney segmentation as a second channel for better segmentation results avoiding RMs segmentation outside the kidney region.

2) U-NET++ MODEL

The improved U-Net model with increased skip connection with deep supervision reduces the semantic gap between the feature maps at the encoder and decoder pathways in the U-Net++ model [40]. We designed the U-Net++ model with five layers of convolutional layers in the encoder path with several filters (8,16,32,64,128) at successive convolutional layers. The activation function was RELU with kernel initializer as 'HE_NORMAL', and 'same' padding. The dropout rate was set to 0.2 after each convolution layer to avoid overfitting issues in training. The ADAM optimizer had a learning rate of 0.001 and a combination loss function. The final layer output was obtained from the deep supervision branch where all segmentation branches were averaged.

3) ATTENTION U-NET MODEL

The boundary of renal masses is problematic when there is less contrast between the RMs and surrounding kidney tissues. To improve the performance of the RMs boundaries an attention U-Net [41] the model was designed. The attention U-Net was six layers deep with an initial filter size of eight and doubled at the following convolutional layers. The convolution layer was designed with 'RELU' activation, 'same' padding, and 'HE_NORMAL' as kernel initializer with a kernel size of 3 X 3. Batch normalization and dropout 0.3 was used to limit the overfitting of the model in training. ADAM optimizer with a learning rate of 0.0001 and a combination loss function.

4) SEGNET MODEL

The SegNet model is an encoder-decoder type architecture identical to the convolutional layers in the VGG16 network [42]. In the SegNet model decoder path pooling indices are transferred from the expansion path instead of the entire feature set, which results in less memory utilization. The designed model had 5 layers in the encoder and decoder paths with two convolutional layers at each stage. The filters were (32,64,128,256,512) in each layer. The kernel size of 3×3 with 'RELU' activation with HE_Normal as kernel initializer and 'same' padding at each convolution layer. MaxPooling was used after every alternate convolutional layer in the encoder path. This model was trained using an ADAM optimizer and combined loss function.

5) PRE-TRAINED RESNET34 BACKBONE IN U-NET MODEL

We implemented ImageNet-based pre-trained RESNET34 as a backbone [43] in the U-Net model to study whether pre-trained weights assist in renal mass segmentation by addressing the vanishing gradient problem. We used the Segmentation Models Python library to develop ImageNet-based pre-trained RESNET34 in the U-Net model.

F. REGISTRATION OF TWO mpMRI SEQUENCE

Generating manual annotations for each type of mpMRI sequence is time-consuming and laborious. Therefore, there are limitations to developing a dedicated model for each mpMRI sequence because of the unavailability of manual annotations. This part of the study was to test whether a registration between the source and target mpMRI images can be used for RMs detection without developing a deep learning-based model from scratch for RMs detection on target mpMRI images. By registering two mpMRI sequences, we can transfer the RMs detection from the source mpMRI to the target mpMRI.

Initially, we identified the best-performing mpMRI sequence for RMs detection as fixed and another mpMRI sequence image as moving images. We registered both

TABLE 1. Comparison of renal mass detection results on T2W MRI sequence.

Model	Precision (%)	Recall/Sensitivity (%)	Specificity (%)	DSC (%)
U-Net	**71.23**	**93.47**	**86.55**	**78.13**
U-Net++	61.25	83.37	82.51	77.35
Attention U-Net	66.51	77.84	85.64	75.57
Pretrained U-Net-RESNET34	52.24	75.55	63.17	61.22
SegNet	61.11	72.27	78.28	73.77

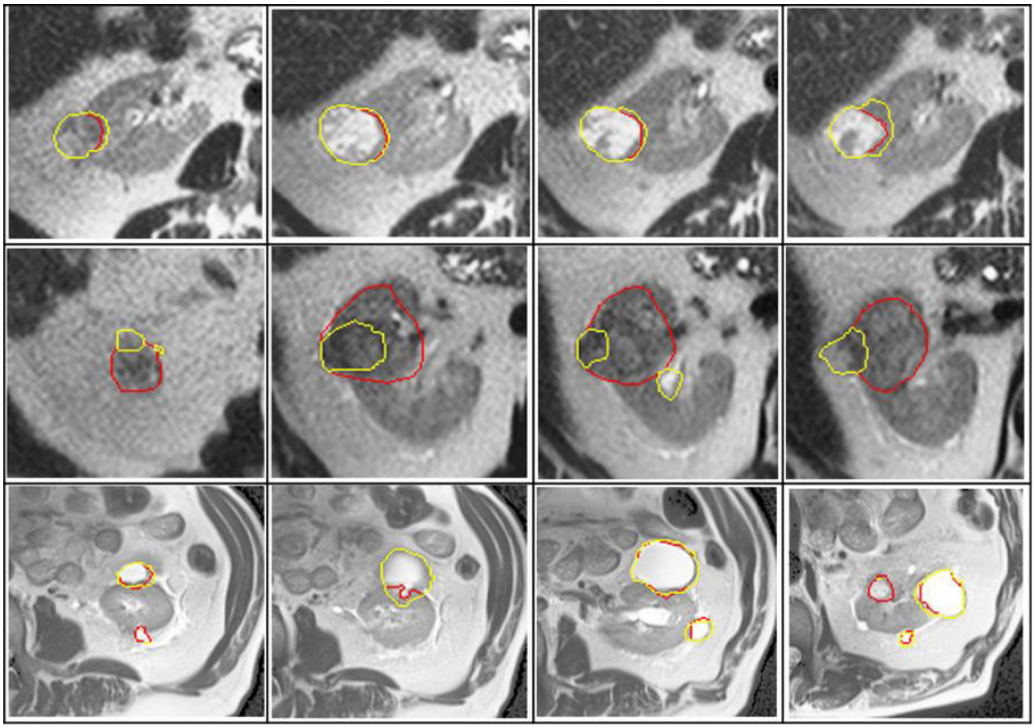

FIGURE 3. Sample renal mass detection on T2W kidney ROI images of different patients are shown in each row where red contours are model predictions and yellow contours are ground truth.

images using the Euler3DTransform of rigid registration by rotation and translation of the 3D images. Once the initial transforms were attained, we used a resampling and gradient descent optimizer with the Mattes Mutual Information similarity metric. Voxel intensity differences were observed between similar types of tissues in the source and target mpMRI sequences. Therefore, mutual information was able to consider matching between voxels of different intensities provided that the shapes and sizes of the voxel areas were similar [44]. The best-performing model predictions of the fixed images were transferred to the registered sequence images. The performance of registered predictions was evaluated using the precision, recall, specificity, and accuracy of RMs detection as compared to the original, non-registered images. The main purpose of the registration was to leverage the best-performing mpMRI sequence predictions for other lower-performing sequences. The registration process was implemented using the SimpleITK registration framework [45], [46].

G. EVALUATION METRICS

We evaluated our models using a 5-fold cross-validation method and tested them on the holdout unseen test data of 22 patient cases. For a fair comparison, these 22 test cases in four different mpMRI sequences namely T2W, T1W-CM, T1W-NG, and T1W-IP were similar in stage 1 of kidney segmentation and stage 2 of the detection of RMs step. To evaluate the automated kidney segmentation, we used DSC as a measure of spatial overlap between models predicted and manually segmented kidneys.

The models for the detection of RMs were evaluated based on precision, recall, specificity, and DSC. A confusion matrix

TABLE 2. Comparison of renal mass detection results on T1W-CM mpMRI sequence.

Model	Precision (%)	Recall/Sensitivity (%)	Specificity (%)	DSC (%)
U-Net	86.67	82.26	92.34	86.57
U-Net++	**95.34**	**85.41**	**99.64**	**89.79**
Attention U-Net	87.62	80.09	96.27	86.64
Pretrained U-Net-RESNET34	82.49	77.65	92.03	85.65
SegNet	79.67	74.89	84.56	82.21

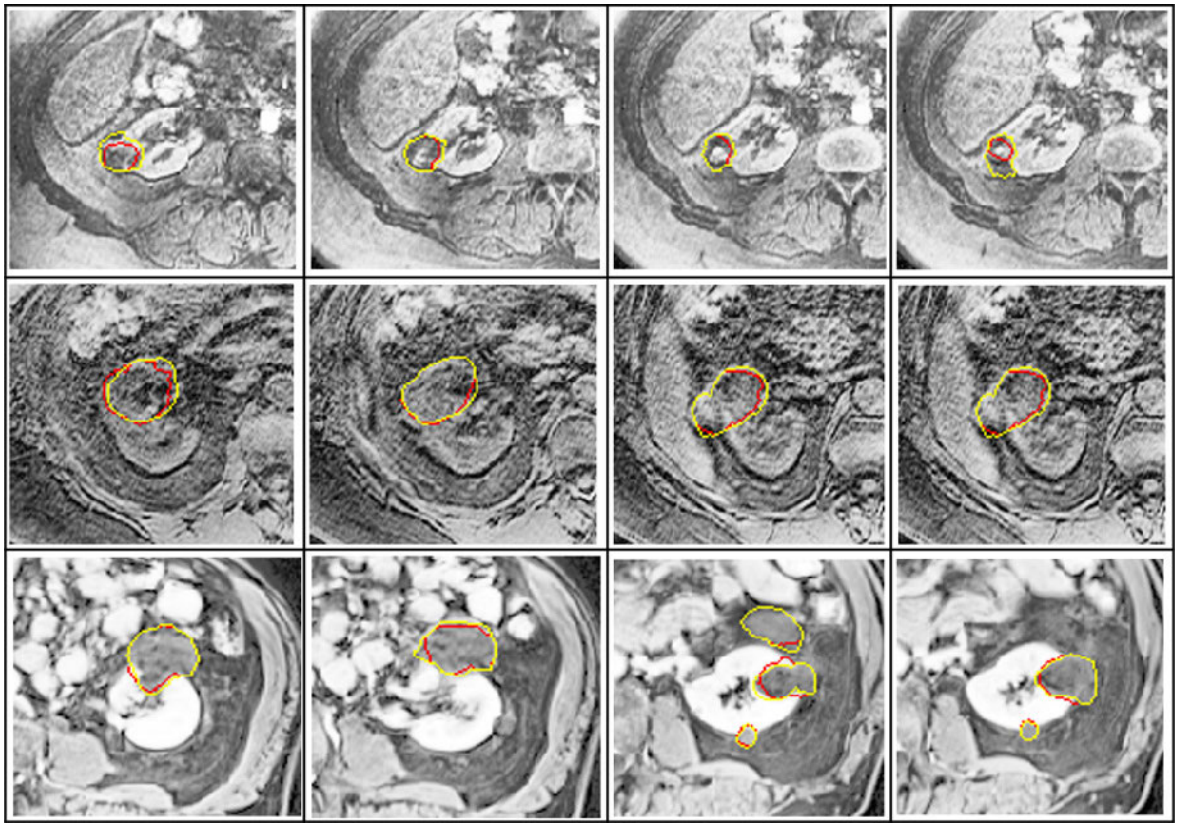

FIGURE 4. Sample renal mass detection on T1W-CM kidney ROI images of different patients are shown in each row where red contours are model predictions and yellow contours are manual segmentation.

was prepared by comparing slices with and without RMs predicted by the algorithm with manual delineation. A true positive (TP) was considered as an overlap between the model-predicted and manually segmented RMs. No detection of RMs in model prediction when manually segmented RMs were present, which was considered a false negative (FN). When the model predicted RMs in a healthy kidney slice without any manually segmented RMs label was considered false positive (FP). No RMs prediction when there were no RMs identified in manual segmentation were considered a true negative (TN). When the TP, FP, FN, and TN values were available for the test data, we calculated the precision, recall, specificity, and accuracy metrics for RMs detection in each model.

III. RESULTS

The attention U-Net model for kidney segmentation reported the kidney segmentation average DSC (mean ± SD) of 89.89% ± 5.0%, 87.64% ± 3.5%, 89.34% ± 5.31%, 85.42% ± 6.0%, and 83.66% ± 7.5% on T2W, T1W-CM,, T1W-NG, T1W-IP, and T1W-OP mpMRI sequences, respectively. The average volume difference (AVD) of kidney segmentation was 9.89% ± 12.59%, 6.14% ± 4.21%, 8.42% ± 10.84%, 9.83% ± 6.41%, and 15.42% ± 13.35% on five mpMRI sequences [28]. The algorithm-generated kidney segmentations served as the ROI for the proposed RMs detection models. Here, we evaluated RMs detection on five different mpMRI sequences using the same test cases acquired on the respective mpMRI sequence protocols. While implementing the detection of RMs on T1W-IP and T1W-OP images, we found that the results were poor compared to the other three mpMRI sequences, namely T2W, T1W-CM, and T1W-NG. This was also evident from the stage 1 results where automated kidney segmentation results were lower in T1W-IP and T1W-OP images. Therefore, we excluded RMs detection on T1W-IP and T1W-OP results in the comparison

TABLE 3. Comparison of renal mass detection results on T1W-NG mpMRI sequence.

Metrics	Precision (%)	Recall/Sensitivity (%)	Specificity (%)	DSC (%)
U-Net	69.72	75.43	87.56	79.62
U-Net++	68.15	74.28	85.20	78.29
Attention U-Net	**71.92**	**78.94**	**93.36**	**82.82**
Pretrained U-Net-RESNET34	67.56	77.62	87.33	78.47
SegNet	66.90	74.53	85.39	77.45

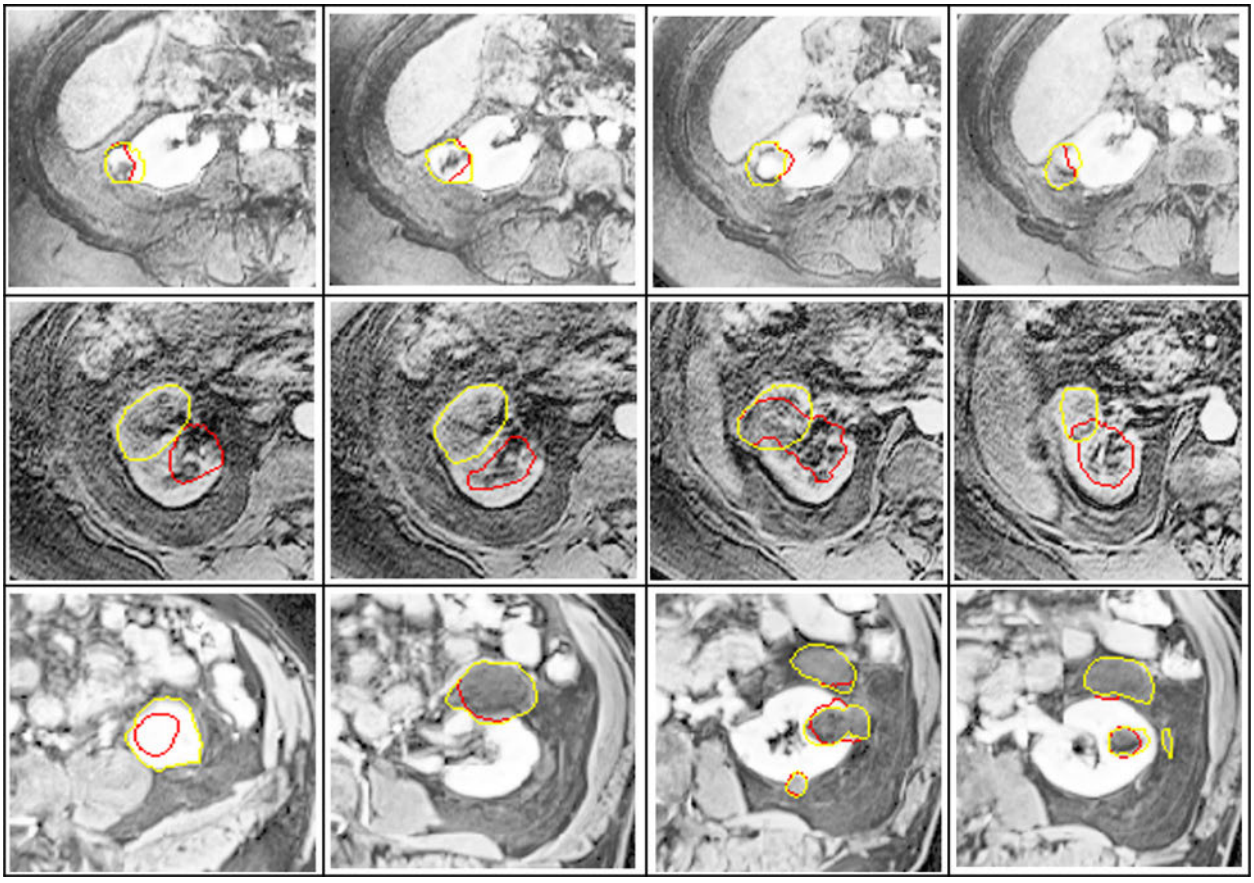

FIGURE 5. Sample renal mass detection on T1W-NG kidney ROI images of different patients shown in each row where red contours are model predictions and yellow contours are ground truth.

study. The RMs detection results were compared for the T2W, T1W-CM, and T1W-NG mpMRI sequences. In the exemplary figures of RMs detection, we showed similar 3 patients' test cases in three patients on T2W, T1W-CM, and T1W-NG images.

A. RENAL MASS DETECTION ON T2W IMAGES

A comparison of the RMs detection results on T2W test images using different models is presented and shown in Table 1. The best-performing results are shown in bold. The U-Net model outperformed the other models, including U-Net++, Attention U-Net, Pretrained U-Net RESNET34, and SegNet.

The U-Net model produced precision, sensitivity, and specificity of 71.23%, 93.47%, and 86.55% respectively. The U-Net model generated the highest renal mass segmentation DSC of 78.13%. Among these models, the pre-trained U-Net RESNET34 model performed poorly on T2W test images. Sample RMs detections using the proposed U-Net model on T2W images are shown in Fig. 3.

B. RENAL MASS DETECTION ON T1W-CM IMAGES

A comparison of the results of the RMs detection models for the T1W-CM test cases is presented in Table 2. The best performance metrics are shown in bold. The U-Net++ model with deep supervision (DS) outperformed the other

TABLE 4. Comparison of RMs detection results on T2W, T1W-CM, and T1W-NG mpMRI sequences.

Model	Precision (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	DSC (%)
T2W U-Net	71.23	93.47	86.55	81.50	78.13
T1W-CM U-Net++	**95.34**	**85.41**	**99.64**	**93.47**	**89.79**
T1W-NG Attention U-Net	71.92	78.94	93.36	65.43	82.82

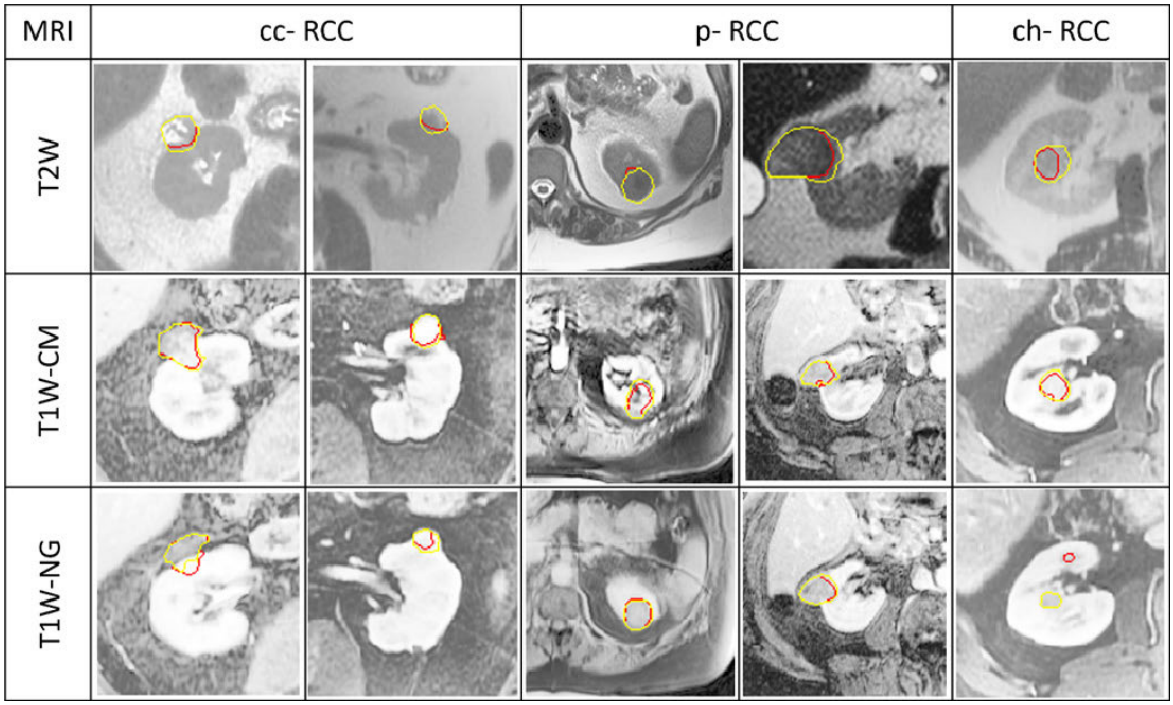

FIGURE 6. Sample clear cell RCC (cc-RCC), papillary RCC (p-RCC), and chromophobe (ch-RCC) detection on 3 MRI sequences where algorithm-generated contours are in red and manual segmentation are in yellow.

state-of-the-art models (i.e., U-Net, Attention U-Net, Pretrained U-Net RESNET34 and SegNet). The U-Net++ with DS produced RMs detection on T1W-CM test cases with precision, recall, and specificity of 95.34%, 85.41%, and 99.64% respectively. The model showed an accuracy of 93.47% and an average DSC of RMs of 89.90%. A sample of localized RMs segmentations using the U-Net++ model on T1W-CM is shown in Fig. 4.

C. RENAL MASS DETECTION ON T1W-NG IMAGES

The comparison results of RMs detection by different models are listed in Table 3, where the values in bold indicate the highest performance metric. The attention U-Net model detected RMs with high performance on T1W-NG test cases as compared to U-Net, U-Net++, Pretrained U-Net RESNET34, and SegNet The attention U-Net model generated precision, recall, and specificity of 71.92%, 78.94%, and 93.36% respectively. The proposed attention U-Net model produced average accuracy and RMs segmentation DSC of 65.43% and 82.82% respectively. The sample RMs detection using the attention U-Net model on T1W-NG is shown in Fig. 5. A summary of RMs detection using U-Net, U-Net++, and attention U-Net on T2W, T1W-CM, and T1W-NG images is presented in Table 4. Due to differences in image acquisition protocol, it was difficult to suggest a single model for the three mpMRI sequences. Compared to T2W and T1W-NG, the results for RMs detection were superior for T1W-CM.

The T2W model produced the highest sensitivity of 93.47% but the lowest specificity of 86.55%, which yielded an accuracy of 81.50%. The T1W-CM model was best best-performing mpMRI sequence for RMs detection with a specificity of 99.64% and an accuracy of 93.47%. The samples of detected malignant RMs of different RCC types on T2W, T1W-CM, and T1W-NG images are shown in Figure 6. Each column in Figure 6 shows different test cases of the patients, and each row shows the corresponding MRI sequence image. The benign fpAML and oncocytoma renal mass detections of these three sequences are shown

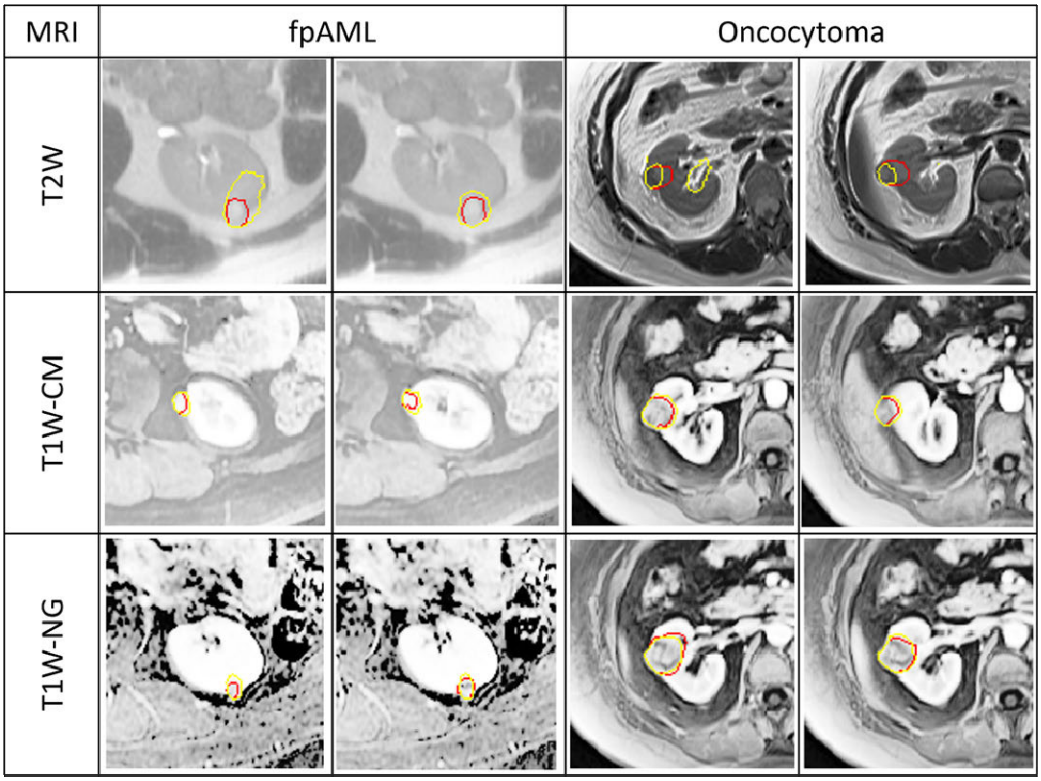

FIGURE 7. Sample benign fat poor AML and oncocytoma detection on 3 MRI sequences where algorithm-generated contours are in red and manual segmentation is in yellow.

in Figure 7. The predictions of the benign renal mass type in two consecutive ROI images are shown in Fig.7 for the three different MRI types. Box plots for the results of RMs detection and segmentation using the proposed models for respective mpMRI sequences on a similar test population are shown in Fig. 8.

D. REGISTRATION OF T1W-CM AND T2W IMAGES

The RMs detection results for the three mpMRI sequences proved that the RMs detection model on T1W-CM performed better than the T2W and T1W-NG sequence models. In practice, T2W and T1W-CM MRI have been studied by clinicians for renal mass characterization. [8] Therefore, we registered T1W-CM and T2W images where T1W-CM images were fixed and T2W images were moving. The results of transferring the RMs detection of T1W-CM on T2W images were compared with the RMs detection on original T2W images as shown in Table 5.

IV. DISCUSSION

The automated detection of RMs is a crucial step in the research of the study of computer-aided kidney cancer classification (e.g., benign vs. malignant) and cancer subtyping [47]. In this study, we compared five different fully automated methods for RMs detection on five mpMRI sequences. We also compared the detection accuracies for each mpMRI sequence. This was the second stage of our cascaded approach to renal mass detection, utilizing the kidney segmentations as ROI detected in the first stage. The best-performing deep learning models were U-Net, U-Net++, and attention U-Net on the T2W, T1W-CM, and T1W-NG sequences, respectively. The segmented RMs average DSC (mean $\pm$ SD) was highest at 89.79 % $\pm$ 6.57 % on the T1W-CM sequence than T2W (78.13 % $\pm$ 11.62%) and T1W-NG (82.82 % $\pm$ 13.83%) sequences. Therefore, we demonstrated that T1W-CM is the most suitable MRI sequence for RMs detection. Our results are significant as they are validated on unseen test cases. The earlier study done by Agarwal et al. [48] for renal mass segmentation on T1W-NG images using an ensemble approach of deep learning models achieved recall and precision of 86.2% and 83.3% on the entire T1W-NG dataset. Their results were not validated on unseen test cases and the ensemble approach of deep learning models requires higher computational power than evaluating the prediction of a single model. The recent systematic review of renal tumor image classification by Amador et al. [49] highlighted that the time-consuming expert knowledge of radiologists requires AI assistance in precise lesion segmentation. Studies evaluating the quantitative assessment of contrast-enhanced imaging have shown that cc-RCC is enhanced to a greater degree than the renal cortex during the corticomedullary phase of enhancement and washout during the nephrographic phase [8]. Compared to T2W and T1W-CM sequences, benign cysts are not

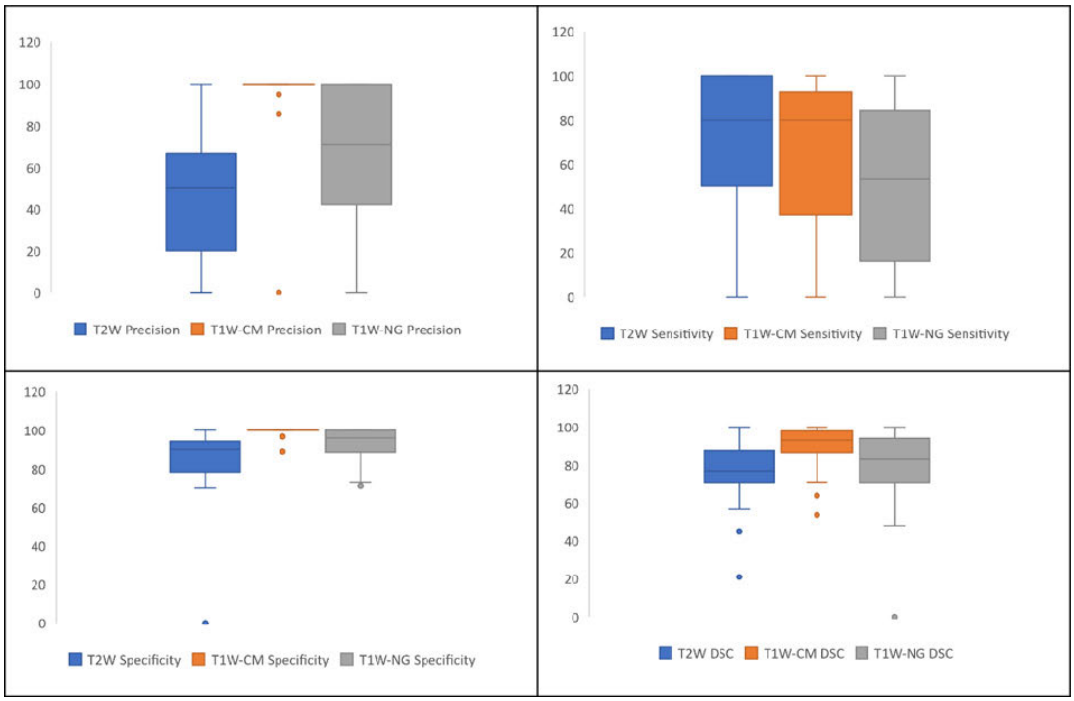

FIGURE 8. Comparison of boxplots of results generated by proposed RMs detection models on T2W, T1W-CM, and T1W-NG mpMRI sequences.

TABLE 5. Comparison of RMs detection original and registered T2W images.

RMs detection Method	Precision (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
Using original T2W images	71.23	**93.47**	**86.55**	**81.50**
Using T1W-CM results transferred on the registered T2W images	**75.55**	82.05	55.00	73.77

prominently visible in T1W-NG [8], [10]. Compared to T1W-CM, these observations may lead to inferior performance of T1W-NG for RMs detection. T1W-NG and T2W resulted in higher false positive detection of RMs because of differences in tumor-to-kidney SI ratio and tumor-to-spleen SI ratio [50], [51]. This difference in the SI of RMs in T2W images may be useful for RMs characterization [52]. As shown in Figure 8, each model showed outliers, outside the overall acceptable model performance. This study showed robustness to vendor-specific variability as the data included MRI scans from clinical MRI systems of multiple vendors. Of the 22 test cases, 10 cases showed multiple RMs, which were different types of solid renal masses and/or cysts. Our models had accurately predicted a variety of RMs in test cases.

Because these mpMRI sequences were acquired using different acquisition protocols (see Appendix), there was a domain shift within them. This domain shift reduces the ability of the models to generalize across images different mpMRI images [35], [53]. Therefore, we used separate models for RMs detections in each mpMRI sequence. The accurately detected RMs on different mpMRI sequences were malignant RCC and its subtypes, and benign solid RMs (fpAML, oncocytomas, angiomyolipoma, and cysts) acquired on heterogeneous MRI systems at different institutions. Our results showed that these RMs detection models worked equally well for both benign and malignant RMs.

The chemical shift imaging: T1-weighted in phase (T1W-IP) and opposed-phase (T1W-OP) MRI sequence performed poorly on automated RMs detection with an average DSC of less than 50%; therefore, we didn't report these results. In T1W-IP images, the decreased SI was >25% in ccRCCs but other RCCs rarely showed minimal SI drop [8]. The drop in SI was non-specific, which may explain the poor performance of T1W-IP images for RMs detection.

We compared these model performances in an ablation study using binary cross entropy loss function. The RMs detection models were unable to produce an average DSC of 60% in all categories of mpMRI sequences. The main reason was that RMs were small areas to detect on a large kidney background. The combined loss

TABLE 6. Multi-parametric MRI technique used for renal masses including imaging at 1.5 and 3 Tesla[a].

Pulse Sequence	Dual echo T1W GRE			T2W TSE/FSE	Volume Interpolated T1W 3D GRE[b]		Diffusion weighted imaging [c]
	2D GRE	3D GRE		Single shot TSE/FSE	3T	1.5T	Single shot echo-planar imaging
		3T	1.5T				
Physiology	Breath Hold	Breath Hold	Breath hold	Respiratory Triggered Breath hold	Breath Hold		Breath Hold
Duration	21 sec.	16 sec.	20 sec.	3-4 min. 22 sec.	20 sec.		21 sec
Fat Suppression	N/A	N/A	N/A	N/A	Chemical or Spectral Inversion Recovery		Spectral Inversion Recovery
TE (IP/OP)[c] ; TR (msec)	(4.6/2.3);160-180	(2.5/1.3);5.5 and (2.2/1.1);4.0	(4.6/2.3);7.6	83-88;1030	1.7-2.5; 4.0-4.5	1.4;4.3	60.8-74;2075-4600
Flip angle (degrees)	70	10-12	10	180	10-12	10-12	90
Bandwidth (Hz)	260	700	313	450	325-460	488	250-1446
Number of excitations	1	0.7-1	1	Half-Fourier	1	1	2
Acceleration factor	2	2	1	1 2	2	2	2
Matrix Size	256/320 x 134/152	294 x 224	192 x 320	170 x 256	256 x 320	132 x 320	130-38;96-75
Field of view (cm)	25 x 35	25 x 35	25 x 35	25 x 35	25 x 35	25 x 35	40-380;28-75
Slice thickness (mm)	5-6	3-4	3-5	5	2.5-4	2.5-4	6

[a]Imaging performed on clinical 1.5 Tesla (Symphony, Avanto or Aera, Siemens Healthcare; Optima, GE Healthcare or Achieva, Philips Healthcare) or 3 Tesla (TRIO, Siemens Healthcare; Discovery 750W, General Electric Healthcare) systems.
[b]VIBE (Siemens Healthcare), LAVA (General Electric Healthcare), THRIVE (Philips Healthcare).
[c]Diffusion weighted imaging performed with two b values (0 and 600 mm2/sec) with ADC map automatically derived.
[d]IP=in phase, OP=opposed phase

function has a dice score involved in it which helps to improve training. This was verified as the model's performance was improved. RMs can be multiple instances in the same image, therefore we investigated the YOLOv4 model [54] for RMs detection. It was implemented on T2W mpMRI sequence images. The mean average precision (mAP) compared the ground-truth bounding box to the detected box returned a score of 40%. The lower mAP score resulted from false RMs detection in the YOLOv4 model.

Image registration for RMs detection on target MRI sequences showed that RMs detection on one MRI sequence (source) could be transferred or overlaid on another MRI sequence (target) without the need to develop a dedicated CNN model on the target MRI sequence. In this study T2W images were registered on the T1W-CM images. Owing to

registration differences, the transfer of RMs detection of T1W-CM images on target T2W images did not achieve results similar to those of T1W-CM. We observed that five T2W test images were acquired with higher resolution than the corresponding T1W-CM images. Although down-sampling of the T2W images was performed in the registration process, there was a difference between T1W-CM and registered T2W images. Therefore, some RMs detections were false negative detections, resulting in reduced sensitivity. This helped to improve the precision of the RMs detection in registered T2W images by eliminating the false positive predictions in RMs detection in the deep learning model trained with original T2W images. This study demonstrated that registration could help in RMs detection in MRI sequences for which a trained deep-learning model is unavailable.

The limitation of this study was that there were three different models for three different MRI sequences for renal mass detection. Therefore, in future studies, it will be worth exploring domain adaptation techniques using Generative Adversarial Networks (GAN) to propose a single generalized deep learning model on different MRI sequences. Designing and developing a separate renal mass detection model for each multi-parametric MRI was computationally exhaustive and time-demanding. Therefore, in future research work instead of developing a dedicated deep learning model for each MRI sequence, one can implement image registration techniques to transfer renal mass detection results on target MRI sequences for which a renal mass detection model is unavailable.

V. CONCLUSION

The best-performing deep learning models were U-Net, U-Net++, and attention U-Net on the T2W, T1W-CM, and T1W-NG sequences, respectively. Of the five different mpMRI sequences, the T1W-CM sequence was the most suitable for RMs detection compared to the T2W and T1W-NG images. We also demonstrated that segmentations can be transferred from the source to the target domain via image registration when manual segmentations are not available for the target mpMRI sequence.

APPENDIX

MRI RENAL MASS PROTOCOL

The multi-parametric MRI technique was used for renal masses, including imaging at 1.5 and 3 Tesla[a].

CONFLICT OF INTEREST

The author of this article has no conflict of interest to disclose.

REFERENCES

[1] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, and F. Bray, "Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012," *Int. J. Cancer*, vol. 136, no. 5, pp. E359–E386, Mar. 2015, doi: 10.1002/ijc.29210.

[2] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2022," *CA, A Cancer J. Clinicians*, vol. 72, no. 1, pp. 7–33, Jan. 2022, doi: 10.3322/caac.21708.

[3] D. R. Brenner, A. Poirier, R. R. Woods, L. F. Ellison, J.-M. Billette, A. A. Demers, S. X. Zhang, C. Yao, C. Finley, N. Fitzgerald, N. Saint-Jacques, L. Shack, D. Turner, and E. Holmes, "Projected estimates of cancer in Canada in 2022," *Can. Med. Assoc. J.*, vol. 194, no. 17, pp. E601–E607, May 2022, doi: 10.1503/cmaj.212097.

[4] A. Volpe, T. Panzarella, R. A. Rendon, M. A. Haider, F. I. Kondylis, and M. A. S. Jewett, "The natural history of incidentally detected small renal masses," *Cancer*, vol. 100, no. 4, pp. 738–745, Feb. 2004, doi: 10.1002/cncr.20025.

[5] A. M. May, A. Guduru, J. Fernelius, S. J. Raza, F. Davaro, S. A. Siddiqui, and Z. A. Hamilton, "Current trends in partial nephrectomy after guideline release: Health disparity for small renal mass," *Kidney Cancer*, vol. 3, no. 3, pp. 183–188, Nov. 2019, doi: 10.3233/kca-190066.

[6] M. C. S. Wong, W. B. Goggins, B. H. K. Yip, F. D. H. Fung, C. Leung, Y. Fang, S. Y. S. Wong, and C. F. Ng, "Incidence and mortality of kidney cancer: Temporal patterns and global trends in 39 countries," *Sci. Rep.*, vol. 7, no. 1, pp. 1–10, Nov. 2017, doi: 10.1038/s41598-017-15922-4.

[7] N. Schieda, S. Krishna, I. Pedrosa, S. D. Kaffenberger, M. S. Davenport, and S. G. Silverman, "Active surveillance of renal masses: The role of radiology," *Radiology*, vol. 302, no. 1, pp. 11–24, Jan. 2022, doi: 10.1148/radiol.2021204227.

[8] N. K. Ramamurthy, B. Moosavi, M. D. F. McInnes, T. A. Flood, and N. Schieda, "Multiparametric MRI of solid renal masses: Pearls and pitfalls," *Clin. Radiol.*, vol. 70, no. 3, pp. 304–316, Mar. 2015, doi: 10.1016/j.crad.2014.10.006.

[9] J. J. Nikken and G. P. Krestin, "MRI of the kidney—State of the art," *Eur. Radiol.*, vol. 17, no. 11, pp. 2780–2793, Nov. 2007, doi: 10.1007/s00330-007-0701-3.

[10] J. M. Willatt, H. K. Hussain, S. Chong, M. Kappil, S. F. Azar, P. S. Liu, J. A. Ruma, and K. M. Elsayes, "MR imaging in the characterization of small renal masses," *Abdominal Imag.*, vol. 39, no. 4, pp. 761–769, Aug. 2014, doi: 10.1007/s00261-014-0109-x.

[11] K. Al Nasibi, J. S. Pickovsky, F. Eldehimi, T. A. Flood, L. T. Lavallee, A. K. Tsampalieros, and N. Schieda, "Development of a multiparametric renal CT algorithm for diagnosis of clear cell renal cell carcinoma among small (≤ 4 cm) solid renal masses," *Amer. J. Roentgenology*, vol. 219, no. 5, pp. 814–823, Nov. 2022, doi: 10.2214/ajr.22.27971.

[12] N. Schieda, M. S. Davenport, S. G. Silverman, B. Bagga, D. Barkmeier, Z. Blank, N. E. Curci, A. M. Doshi, R. T. Downey, E. Edney, E. Granader, I. Gujrathi, R. M. Hibbert, N. Hindman, C. Walsh, T. Ramsay, A. B. Shinagare, and I. Pedrosa, "Multicenter evaluation of multiparametric MRI clear cell likelihood scores in solid indeterminate small renal masses," *Radiology*, vol. 303, no. 3, pp. 590–599, Jun. 2022, doi: 10.1148/radiol.211680.

[13] E. Roussel, U. Capitanio, A. Kutikov, E. Oosterwijk, I. Pedrosa, S. P. Rowe, and M. A. Gorin, "Novel imaging methods for renal mass characterization: A collaborative review," *Eur. Urology*, vol. 81, no. 5, pp. 476–488, May 2022, doi: 10.1016/j.eururo.2022.01.040.

[14] M. G. Lubner, "Radiomics and artificial intelligence for renal mass characterization," *Radiologic Clinics North Amer.*, vol. 58, no. 5, pp. 995–1008, Sep. 2020, doi: 10.1016/j.rcl.2020.06.001.

[15] 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.

[16] F. Zabihollahy, N. Schieda, S. Krishna, and E. Ukwatta, "Automated classification of solid renal masses on contrast-enhanced computed tomography images using convolutional neural network with decision fusion," *Eur. Radiol.*, vol. 30, no. 9, pp. 5183–5190, Sep. 2020, doi: 10.1007/s00330-020-06787-9.

[17] A. Anush, G. Rohini, S. Nicola, E. M. WalaaEldin, and U. Eranga, "Deep-learning-based ensemble method for fully automated detection of renal masses on magnetic resonance images," *J. Med. Imag.*, vol. 10, no. 2, p. 24501, Mar. 2023, doi: 10.1117/1.jmi.10.2.024501.

[18] X.-Y. Chen, Y. Zhang, Y.-X. Chen, Z.-Q. Huang, X.-Y. Xia, Y.-X. Yan, M.-P. Xu, W. Chen, X.-L. Wang, and Q.-L. Chen, "MRI-based grading of clear cell renal cell carcinoma using a machine learning classifier," *Frontiers Oncol.*, vol. 11, pp. 1–8, Oct. 2021, doi: 10.3389/fonc.2021.708655.

[19] D. K. Dwivedi, Y. Xi, P. Kapur, A. J. Madhuranthakam, M. A. Lewis, D. Udayakumar, R. Rasmussen, Q. Yuan, A. Bagrodia, V. Margulis, M. Fulkerson, J. Brugarolas, J. A. Cadeddu, and I. Pedrosa, "Magnetic resonance imaging radiomics analyses for prediction of high-grade histology and necrosis in clear cell renal cell carcinoma: Preliminary experience," *Clin. Genitourinary Cancer*, vol. 19, no. 1, pp. 12–21, Feb. 2021, doi: 10.1016/j.clgc.2020.05.011.

[20] D. Said, S. J. Hectors, E. Wilck, A. Rosen, D. Stocker, O. Bane, A. T. Beksaç, S. Lewis, K. Badani, and B. Taouli, "Characterization of solid renal neoplasms using MRI-based quantitative radiomics features," *Abdominal Radiol.*, vol. 45, no. 9, pp. 2840–2850, Sep. 2020, doi: 10.1007/s00261-020-02540-4.

[21] N. E. Canvasser, F. U. Kay, Y. Xi, D. F. Pinho, D. Costa, A. D. de Leon, G. Khatri, J. R. Leyendecker, T. Yokoo, A. Lay, N. Kavoussi, E. Koseoglu, J. A. Cadeddu, and I. Pedrosa, "Diagnostic accuracy of multiparametric magnetic resonance imaging to identify clear cell renal cell carcinoma in cT1a renal masses," *J. Urol.*, vol. 198, no. 4, pp. 780–786, Oct. 2017, doi: 10.1016/j.juro.2017.04.089.

[22] I. L. Xi, Y. Zhao, R. Wang, M. Chang, S. Purkayastha, K. Chang, R. Y. Huang, A. C. Silva, M. Vallières, 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, doi: 10.1158/1078-0432.ccr-19-0374.

[23] W. Wang, K. Cao, S. Jin, X. Zhu, J. Ding, and W. Peng, "Differentiation of renal cell carcinoma subtypes through MRI-based radiomics analysis," *Eur. Radiol.*, vol. 30, no. 10, pp. 5738–5747, Oct. 2020, doi: 10.1007/s00330-020-06896-5.

[24] I. Aslam, F. Aamir, M. Kassai, L. A. Crowe, P.-A. Poletti, S. D. Seigneux, S. Moll, L. Berchtold, and J.-P. Vallée, "Validation of automatically measured T1 map cortico-medullary difference (ΔT1) for eGFR and fibrosis assessment in allograft kidneys," *PLoS One*, vol. 18, no. 2, Feb. 2023, Art. no. e0277277, doi: 10.1371/journal.pone.0277277.

[25] A. J. Daniel, C. E. Buchanan, T. Allcock, D. Scerri, E. F. Cox, B. L. Prestwich, and S. T. Francis, "Automated renal segmentation in healthy and chronic kidney disease subjects using a convolutional neural network," *Magn. Reson. Med.*, vol. 86, no. 2, pp. 1125–1136, Aug. 2021, doi: 10.1002/mrm.28768.

[26] T. L. Kline, M. E. Edwards, J. Fetzer, A. V. Gregory, D. Anaam, A. J. Metzger, and B. J. Erickson, "Automatic semantic segmentation of kidney cysts in MR images of patients affected by autosomal-dominant polycystic kidney disease," *Abdominal Radiol.*, vol. 46, no. 3, pp. 1053–1061, Mar. 2021, doi: 10.1007/s00261-020-02748-4.

[27] Y. Kim, Y. Ge, C. Tao, J. Zhu, A. B. Chapman, V. E. Torres, A. S. L. Yu, M. Mrug, W. M. Bennett, M. F. Flessner, D. P. Landsittel, and K. T. Bae, "Automated segmentation of kidneys from MR images in patients with autosomal dominant polycystic kidney disease," *Clin. J. Amer. Soc. Nephrology*, vol. 11, no. 4, pp. 576–584, 2016, doi: 10.2215/cjn.08300815.

[28] R. Gaikar, F. Zabihollahy, M. W. Elfaal, A. Azad, N. Schieda, and E. Ukwatta, "Transfer learning-based approach for automated kidney segmentation on multiparametric MRI sequences," *J. Med. Imag.*, vol. 9, no. 3, p. 36001, Jun. 2022, doi: 10.1117/1.jmi.9.3.036001.

[29] J. Guo, A. Odu, and I. Pedrosa, "Deep learning kidney segmentation with very limited training data using a cascaded convolution neural network," *PLoS One*, vol. 17, no. 5, May 2022, Art. no. e0267753, doi: 10.1371/journal.pone.0267753.

[30] B. R. Herts, S. G. Silverman, N. M. Hindman, R. G. Uzzo, R. P. Hartman, G. M. Israel, D. A. Baumgarten, L. L. Berland, and P. V. Pandharipande, "Management of the incidental renal mass on CT: A white paper of the ACR incidental findings committee," *J. Amer. College Radiol.*, vol. 15, no. 2, pp. 264–273, Feb. 2018, doi: 10.1016/j.jacr.2017.04.028.

[31] M. Czarniecki, R. Gautam, P. L. Choyke, and B. Turkbey, "Imaging findings of hereditary renal tumors, a review of what the radiologist should know," *Eur. J. Radiol.*, vol. 101, pp. 8–16, Apr. 2018, doi: 10.1016/j.ejrad.2018.01.026.

[32] H. J. Meyer, A. Pfeil, D. Schramm, A. G. Bach, and A. Surov, "Renal incidental findings on computed tomography: Frequency and distribution in a large non selected cohort," *Medicine*, vol. 96, no. 26, p. 7039, 2017.

[33] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *NeuroImage*, vol. 31, no. 3, pp. 1116–1128, Jul. 2006, doi: 10.1016/j.neuroimage.2006.01.015.

[34] R. Chelghoum, A. Ikhlef, A. Hameurlaine, and S. Jacquir, "Transfer learning using convolutional neural network architectures for brain tumor classification from MRI images," in *Proc. IFIP Int. Conf. Artif. Intell. Appl. Innov.*, vol. 1, 2020, pp. 189–200.

[35] M. Ghafoorian, A. Mehrtash, T. Kapur, N. Karssemeijer, E. Marchiori, M. Pesteie, C. R. G. Guttmann, F.-E. de Leeuw, C. M. Tempany, B. van Ginneken, A. Fedorov, P. Abolmaesumi, B. Platel, and W. M. Wells, "Transfer learning for domain adaptation in MRI: Application in brain lesion segmentation," in *Proc. 20th Int. Conf.*, 2017, pp. 516–524.

[36] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and flexible image augmentations," *Information*, vol. 11, no. 2, p. 125, Feb. 2020, doi: 10.3390/info11020125.

[37] S. Jadon, "A survey of loss functions for semantic segmentation," in *Proc. IEEE Conf. Comput. Intell. Bioinf. Comput. Biol. (CIBCB)*, Oct. 2020, pp. 1–7, doi: 10.1109/CIBCB48159.2020.9277638.

[38] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf.*, 2015, pp. 234–241.

[39] C. Xu and H. Wang, "Research on a convolution kernel initialization method for speeding up the convergence of CNN," *Appl. Sci.*, vol. 12, no. 2, pp. 1–24, 2022, doi: 10.3390/app12020633.

[40] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical*, vol. 11045. Cham, Switzerland: Springer, 2018, pp. 3–11.

[41] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glock, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.

[42] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.

[43] 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.

[44] R. Vs and K. Revathy, "Using mutual information and cross correlation as metrics for registration of images," *J Theor. Appl. Inf.*, vol. 4, pp. 1–24, Jan. 2008.

[45] B. C. Lowekamp, D. T. Chen, L. Ibáñez, and D. Blezek, "The design of SimpleITK," *Frontiers Neuroinform.*, vol. 7, pp. 1–14, May 2013, doi: 10.3389/fninf.2013.00045.

[46] Z. Yaniv, B. C. Lowekamp, H. J. Johnson, and R. Beare, "SimpleITK image-analysis notebooks: A collaborative environment for education and reproducible research," *J. Digit. Imag.*, vol. 31, no. 3, pp. 290–303, Jun. 2018.

[47] R. Gaikar, "Development of machine learning algorithms for kidney cancer diagnosis from multi-parametric MRI and histopathology images," Ph.D. thesis, School Eng., Univ. Guelph, Guelph, ON, Canada, 2023. [Online]. Available: https://hdl.handle.net/10214/28046

[48] A. Agarwal, N. Schieda, M. Elfaal, and E. Ukwatta, "Deep-learning-based ensemble method for fully automated detection of renal masses on magnetic resonance images," *J. Med. Imag.*, vol. 10, no. 2, pp. 1–24, Mar. 2023.

[49] S. Amador, F. Beuschlein, V. Chauhan, J. Favier, D. Gil, P. Greenwood, R. R. de Krijger, M. Kroiss, S. Ortuño-Miquel, A. Patocs, A. Stell, and A. Walch, "Deep learning approaches applied to image classification of renal tumors: A systematic review," *Arch. Comput. Methods Eng.*, vol. 31, no. 2, pp. 615–622, Mar. 2024, doi: 10.1007/s11831-023-09995-w.

[50] M. S. Chung, H. J. Choi, M.-H. Kim, and K.-S. Cho, "Comparison of T2-weighted MRI with and without fat suppression for differentiating renal angiomyolipomas without visible fat from other renal tumors," *Amer. J. Roentgenology*, vol. 202, no. 4, pp. 765–771, Apr. 2014, doi: 10.2214/ajr.13.11058.

[51] H. J. Choi, J. K. Kim, H. Ahn, C.-S. Kim, M.-H. Kim, and K.-S. Cho, "Value of T2-weighted MR imaging in differentiating low-fat renal angiomyolipomas from other renal tumors," *Acta Radiologica*, vol. 52, no. 3, pp. 349–353, Apr. 2011, doi: 10.1258/ar.2010.090491.

[52] Y. Zheng, S. Wang, Y. Chen, and H.-Q. Du, "Deep learning with a convolutional neural network model to differentiate renal parenchymal tumors: A preliminary study," *Abdominal Radiol.*, vol. 46, no. 7, pp. 3260–3268, Jul. 2021, doi: 10.1007/s00261-021-02981-5.

[53] W. Yan et al., "The domain shift problem of medical image segmentation and vendor-adaptation by UNet-GAN," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, 2017, pp. 597–609.
[54] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

ROHINI GAIKAR received the B.E. degree in electronics engineering and the M.E. degree in electronics and telecommunication engineering from Savitribai Phule Pune University, India, in 2003 and 2010, respectively. She is currently pursuing the Ph.D. degree in biomedical engineering with the University of Guelph, ON, Canada. From 2011 to 2018, she was an Assistant Professor with the D. Y. Patil College of Engineering, Akurdi, Pune, India. Her research interest includes the development of medical image processing algorithms using artificial intelligence-based algorithms. She is also working on the development of virtual biopsy algorithms for kidney cancer subtyping. She received the MITACS Internship and an Ontario Graduate Scholarship (OGS) for her research work.

AZAR AZAD received the Ph.D. degree in clinical biochemistry from the Diagnostic Medical Laboratory Sciences, Tehran University of Medical Sciences, in 1991. Her postdoctoral research was focused on molecular pathway analysis in angiogenesis and apoptosis at the University of Toronto, in 1997. She was with the University of Toronto as the Co-Director of the Banting and Best Diabetes Core Laboratory and as the Scientific Staff with the Mount Sinai Hospital, until 2017. She is currently an Entrepreneur and the Founder of A.I. VALI Inc.

NICOLA SCHIEDA is currently an Associate Professor and the Director of abdominal and pelvic MRI and prostate imaging with the University of Ottawa. He is actively involved in clinical research, with a focus on genitourinary imaging, adrenal, kidney, and prostate cancers. He is a member of the Radiological Society of North America, American Roentgen Ray Society, and the Society of Abdominal Radiology; and the Chair of American College of Radiology Genitourinary Continuing Professional Improvement Program.

ERANGA UKWATTA (Senior Member, IEEE) received the Ph.D. degree from Western University, Canada, in 2013. He was a Multicenter Postdoctoral Fellow with Johns Hopkins University and the University of Toronto. He is currently an Assistant Professor with the University of Guelph, Canada, and an Adjunct Professor of systems and computer engineering with Carleton University, Canada. He has more than 90 journal articles and conference proceedings papers. His research interests include medical image segmentation and registration, deep learning, and computational modeling.