Computer-Aided Diagnostic System for Early Detection of Acute Renal Transplant Rejection Using Diffusion-Weighted MRI

Mohamed Shehata, *Member, IEEE*, Fahmi Khalifa, *Member, IEEE*, Ahmed Soliman, *Member, IEEE*, Mohammed Ghazal, *Senior Member, IEEE*, Fatma Taher, *Member, IEEE*, Mohamed Abou El-Ghar, Amy C. Dwyer, Georgy Gimel'farb, *Member, IEEE*, Robert S. Keynton, *Member, IEEE*, and Ayman El-Baz, *Senior Member, IEEE*

Abstract*—*Objective: **Early diagnosis of acute renal transplant rejection (ARTR) is critical for accurate treatment. Although the current gold standard, diagnostic technique is renal biopsy, it is not preferred due to its invasiveness, long recovery time (1–2 weeks), and potential for complications, e.g., bleeding and/or infection.** ***Methods:*** **This paper presents a computer-aided diagnostic (CAD) system for early ARTR detection using (3D + b-value) diffusion-weighted (DW) magnetic resonance imaging (MRI) data. The CAD process starts from kidney tissue segmentation with an evolving geometric (level-set-based) deformable model. The evolution is guided by a voxel-wise stochastic speed function, which follows from a joint kidney-background Markov–Gibbs random field model accounting for an adaptive kidney shape prior and on-going kidney-background visual appearances. A B-spline-based three-dimensional data alignment is employed to handle local deviations due to breathing and heart beating. Then, empirical cumulative distribution functions of apparent diffusion coefficients of the segmented DW-MRI at different b-values are collected as discriminatory transplant status features. Finally, a deep-learning-based classifier with stacked nonnegative constrained autoencoders is employed to distinguish between rejected and nonrejected renal transplants.** ***Results:*** **In our initial "leave-one-subject-out" experiment on 100 subjects, 97.0% of the subjects were correctly classified. The subsequent four-fold and ten-fold cross-validations gave the average accuracy of 96.0% and 94.0%, respectively.** ***Conclusion:*** **These results demonstrate the promise of this new CAD system to reliably diagnose renal transplant rejection.** ***Significance:*** **The technology presented here can significantly impact the quality of care of renal transplant patients since it has the potential to replace the gold standard in kidney diagnosis, biopsy.**

***Index Terms*—Renal rejection, CAD system, Deep learning, DW-MRI, ADC.**

I. Introduction

BECAUSE there are up to 17,000 renal transplants per annum in the U.S. and a limited number of donors [1], the assessment of a transplanted kidney is of critical importance to a clinician to ensure renal recovery. The immunological response of the patient to a transplanted kidney is referred to as acute renal transplant rejection (ARTR) and is considered to be the leading cause of renal dysfunction [1] after the transplantation. Early detection of renal dysfunction increases the survival rate of the transplanted kidney [2], [3]. Thus, calling for essential medical biomarkers to assess renal transplants is necessary to distinguish the ARTR from other diagnoses, including acute tubular necrosis (ATN) and immune drug toxicity, especially at an early stage, (i.e., before major changes in creatinine clearance and serum plasma creatinine are detected). Traditional blood tests and urine sampling to evaluate renal transplant dysfunction cannot assess the function of individual kidneys. The glomerular filtration rate (GFR) is often used since it is based on measuring the serum creatinine level and has been approved by the National Kidney Foundation to evaluate the overall kidney function. The GFR is a relatively imprecise and late indicator of renal dysfunction (significant changes in creatinine levels are only detectable after a 60% loss of renal function) [4]. Biopsy, which remains the gold standard for assessing renal transplant dysfunction, uses a relatively small needle to remove a small sample of tissue from the kidney. Biopsies are invasive, expensive, have a high morbidity rate, and may over- or under-estimate the extent of inflammation (an indicator of rejection) in the entire graft [5].

Recently, the medical community has explored more favorable approaches, such as non-invasive imaging, to furnish renal function data on each kidney. Scintigraphy is frequently

Manuscript received December 8, 2017; revised March 19, 2018 and May 10, 2018; accepted June 10, 2018. Date of publication June 25, 2018; date of current version January 18, 2019. This study is made possible by the generous support of the American people through the United States Agency for International Development (USAID). The contents are the responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government. *(Corresponding author: Ayman El-Baz.)*

M. Shehata, F. Khalifa, A. Soliman, R. S. Keynton, and A. El-Baz are with the Department of Bioengineering, University of Louisville, Louisville, KY 40292 USA (e-mail: aselba01@louisville.edu).

M. Ghazal is with the Department of Electrical and Computer Engineering, Abu Dhabi University.

F. Taher is with the College of Technological Innovation, Zayed University.

A. C. Dwyer is with the Kidney Transplantation–Kidney Disease Center, University of Louisville.

G. Gimel'farb is with the Department of Computer Science, University of Auckland.

M. A. El-Ghar is with the Department of Radiology, Urology, and Nephrology Center, University of Mansoura.

Digital Object Identifier 10.1109/TBME.2018.2849987

employed due to its low cost and yielding good functional information, but its low spatial resolution [5] and patient exposure to small doses of radioactivity [6] limits its effectiveness. Computed tomography, provides superior functional and anatomical information, but utilizes nephrotoxic contrast agents (CAs) and also exposes patients to radiation [5]. Recent implementation of magnetic resonance imaging (MRI) have circumvented these shortcomings. For example, dynamic contrast-enhanced (DCE) MRI has been exploited for renal function assessment since it provides both anatomical and functional kidney information [7]–[10]. Unfortunately, DCE employs CAs which have been shown to cause nephrogenic systemic fibrosis; thus, many medical centers are reluctant to use DCE-MRI in patients with renal disease [7]. On the other hand, diffusion-weighted (DW) MRI, an emerging imaging modality for renal function assessment [11], circumvents this major drawback and requires no radiation, while also providing both anatomical and physiological information. DW-MRI measures water molecules in soft tissues (e.g., kidney); when combined with the capillary perfusion and water diffusion, the apparent diffusion coefficients (ADC) can be quantified.

Recently, Liu *et al.* [3] detected early renal allograft dysfunction caused by acute rejection (AR) and ATN using DW-MRI by manually selecting cortical and medullary regions of interest (ROIs). The AR group had lower measured values of ADC, than the control groups, but the ADC values did not differ between the AR and ATN groups. Abou El-Ghar *et al.* [2] studied two groups of renal allograft patients who underwent DW-MRI scans at two b-values: 0 and 800 s/mm^2. Group 1 (49 patients) had stable renal allograft function and Group 2 (21 patients) had acute graft impairment, namely, 10 acute cellular rejection (ACR), 7 ATN, and 4 immunosuppresive toxicity (IT) rejection. The ROI was placed at the middle of the kidney in a single cross-section to calculate a pixel-wise ADC. The ADC values of Group 1 were significantly higher than those of Group 2, and no overlap was detected between the ADCs of Group 1 and the ATN patients of Group 2. However, minimal overlap was observed between the ADCs of Group 1 and the Group 2 patients with the ACR and IT.

Eisenberger *et al.* [12] selected manual ROIs in the upper, middle, and lower poles of the cortex and medulla on several slices to cover large regions of the allograft. Means and standard deviations of the ADC from all b-values were measured. ADCs were significantly reduced in the cortex and medulla for the AR and ATN cases, and their values correlated with the creatinine clearance (CrCl). Kaul *et al.* [13] assessed renal dysfunction with cortical and medullary ADC maps and reported a significant decrease in medullas compared to cortices in normal donor and normal functioning allografts. Both the medulla and cortex ADCs decreased or increased significantly for a rejection or recovery from rejection, respectively.

Palmucci *et al.* [14] evaluated the functionality of 21 transplanted kidneys by comparing the estimated ADCs and true diffusion (TD) with renal function indices. Patients were divided into three groups by their CrCl values with the cortical ADC and TD evaluated in a user-defined ROI of the transplanted kidney for the three groups. A moderate positive correlation was found between the CrCl and both ADC and TD, as well as no difference was observed between the ADC and TD values for the adjacent groups. The subsequent extension [15] of these evaluations to 35 patients revealed a slightly smaller positive correlation than the previously reported study [14]. However, acute rejection responses after transplantation could not be detected. The feasibility of diagnosing ARTR from DW-MRI was evaluated by Xu *et al.* [16] on 26 biopsy-proven rejection (R) and 43 non-rejection (NR) patients. The NR patients showed higher ADCs than those of the R group, and also demonstrated optimum sensitivity and specificity at b_{800} s/mm^2, as evidenced by the ROC curve.

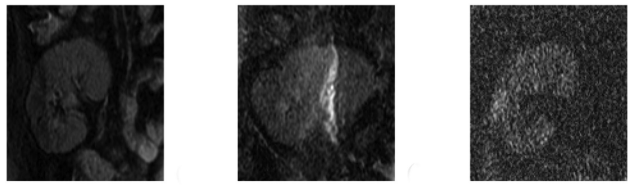

Fig. 1. Coronal cross-sections of raw DW-MRI samples showing (a) similar intensities of kidney and surrounding tissues at b_0, (b) inter-patient anatomical differences at b_0, and (c) image noise at higher b-values (b_{1000}).

Segmenting kidneys from surrounding tissues on abdominal MR images is a key step towards obtaining an accurate diagnosis. In literature, there are different techniques that have been developed to achieve this task. Particularly, the shaped-based segmentation techniques using deformable boundaries have been extensively used due to their flexibly and high accuracy. For example, Leventon *et al.* [17] combined the shape and deformable model by attracting the level-set function to the likely shapes from a training set specified by principal component analysis (PCA). Another study by Litvin *et al.* [18] has employed the shape prior information in curve evolution-based segmentation problems. Based on both intensity and shape prior information, Abdelmunim *et al.* [19], [20] implemented a variational level-set segmentation framework to segment medical shapes (e.g., kidney). Multiple object segmentation was performed using multi-phase level-sets based on shape prior information [21]. The object contour evolution of the level-sets and shape prior estimates were affected by the interactions between neighboring shapes. Gloger *et al.* [22] used Bayesian statistics in addition to shape prior to build a level-set kidney segmentation framework. Wang *et al.* [23] segmented the liver by using a shape-intensity prior level-set after a rough segmentation was obtained by a maximum a posteriori classification of a probability map. Another study by Skalski *et al.* [24] segmented the kidney from CT images using a level-set approach with an ellipsoidal shape constraint. However, achieving an accurate kidney segmentation from DW-MRI is a challenging task because of kidney motions due to inter-patient anatomical differences, low contrast between the kidney and other abdominal structures, especially at higher gradient strengths and duration, or b-values (Fig. 1), low signal-to-noise ratios (SNR) and artifacts that complicate image alignment [25], [26], and geometric distortions due to long acquisition time [27].

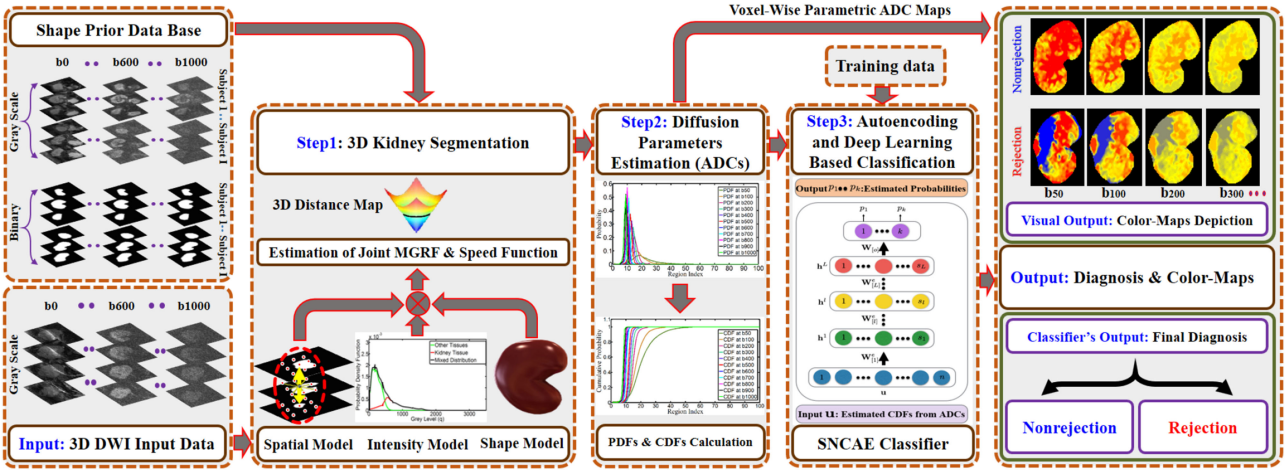

Fig. 2. The presented CAD system for detecting renal rejection from (3D + b-value) DW-MRI.

In total, renal function has been assessed in several clinical studies, but have been limited by the manual delineation of 2D kidney ROIs, which are subjective, depend on the user's knowledge of anatomy, and do not compensate for kidney motions since they do not account for the entire kidney. Several of these studies only investigated the significant statistical differences between pairs of certain b-values, thus only partly capturing kidney function. No study integrated all data analysis steps into a special-purpose CAD system or investigated the fusion of the ADCs at multiple low and high b-values.

To overcome the aforementioned limitations, we develop a fully automated CAD system, shown in Fig. 2, with the ability to delineate the kidney, handle its motion, and fuse the ADC values from segmented DW-MR images acquired at multiple low and high b-values. Since the ultimate goal of this work is to provide an accurate overall non-invasive diagnosis platform for evaluating the transplant status of kidneys, i.e., classify it as either non-rejection or acute rejection, together with providing the clinician with a visual assessment via a color-coded map of the abnormal regions of the transplanted kidney, the presented ARTR-CAD system employs several key components that differentiate this work from previous work [10] by our group. Specifically, the novel components of this CAD system include: (*i*) As an extension for segmenting kidneys from (2D + time) DCE-MRI data [10] to (3D + b-value) DW-MRI data, an adaptive shape prior model and a 4th-order Markov-Gibbs random field (MGRF) model of spatial interactions for 3D images have been integrated into a 3D level-set deformable model to accurately segment a transplanted kidney in the presence of large grey-level inhomogeneities between its different structures, e.g., the cortex and medulla; (*ii*) To accurately diagnose the transplanted kidney, the DW-MR images have been acquired at both low and high b-values, which describe blood perfusion and diffusion, respectively, inside the kidney; (*iii*) To fuse the ADCs at different b-values, stacked auto-encoders (SAE) have been used after establishing voxel-to-voxel correspondences between the segmented DW-MR images of the kidney transplant; and (*iv*) A non-negativity constraint (NC) has been applied to all SAEs, (abbreviated then SNCAEs) for more efficient signal dimensionality reduction and faster learning, i.e., faster convergence to goal parameter estimates for encoding and classification layers when a SNCAE is trained. The SNCAE escapes non-causal negative weighting parameters, thus enhancing the final CAD accuracy.

II. Materials

In total, 110 patients who underwent kidney transplantation provided consent to participate in this study. All scans and biopsies were performed from July 2014 to September 2017, and all kidney transplants were done at Mansoura University (Egypt), with the donated kidneys having been obtained from live donors. However, 10 patients were excluded due to a later on refusal of the study and/or contraindications for the MRI such as metallic prostheses, technical problems, artificial valves, or claustrophobia. The remaining 100 patients, (77 males and 23 females) and range in age from 11 to 54 years (the mean age of 26.3 ± 10.1 years), were included in the study. Patients were divided into two groups: non-rejection (NR) and acute rejection (AR). As a part of routine medical care after transplantation, all patients of both groups were assessed with serum creatinine laboratory values with a normal (basal) level of ≤ 1.3 mg $\cdot$ dl^{-1}. The NR group (30 patients) included patients with healthy graft function. Most of the NR group patients only underwent the DW-MRI scans two weeks after transplantation. The AR group (70 patients) included patients with acute renal rejection, based on renal biopsy histology. All patients of the AR group underwent the DW-MRI scans two weeks after transplantation and just before the renal biopsy. All patients were asked to hold respiration (breath) during the study to reduce respiratory effect. Both the DW-MRI scans and biopsy were included in the final analysis and examined by a nephrologist and a radiologist.

The DW-MR images were acquired before any biopsy procedure by using a 1.5T SIGNA Horizon scanner (General Electric Medical Systems, Milwaukee, WI, USA). Coronal DW-MR images have been obtained by using a body coil and a single-shot spin-echo echo-planar sequence (TR/TE, 8000/61.2 ms; bandwidth, 142 kHz; 1.28×1.28 mm^2 matrix; section thickness of

4 mm; intersection gap of 0 mm; FOV of 36 cm; 7 acquired signals; water signals acquired at different b-values of (b_0, b_{50}, b_{100}, b_{200}, b_{300}, b_{400}, b_{500}, b_{600}, b_{700}, b_{800}, b_{900}, and b_{1000}) s/mm^2). Approximately 50 sections were obtained in 60–120 seconds to cover the whole kidney.

III. Methods

Given an input (3D + b-value) DW-MRI, the CAD system in Fig. 2, performs the following steps: (*i*) kidneys are segmented from surrounding abdominal structures (Section III-A); (*ii*) voxel-wise functional parameters (ADCs) are estimated to form a 3D parametric map for detecting status of the transplanted kidney (Section III-B); and (*iii*) acute rejection (AR) or non-rejection (NR) transplant status is classified in order to use the system as a diagnostic test (Section III-C).

A. 3D Kidney Segmentation

As previously mentioned in the literature, segmenting kidneys from DW-MRI is challenging [25]–[27]. Therefore, our segmentation relies on multiple image features to accurately delineate the kidney and thus facilitates analysis of transplant status. Basic notations and details of the presented segmentation are outlined below.

For describing processing steps, let $\mathbf{p} = (x, y, z)$ denote a voxel at 3D position with discrete Cartesian coordinates (x, y, z) and let $\mathbf{R} = \{(x, y, z) : 0 \leq x \leq X-1; 0 \leq y \leq Y-1; 0 \leq z \leq Z-1\}$ be a finite 3D arithmetic lattice of unit voxels. The lattice has the size of XYZ and supports both grayscale images and their parametric or region (segmentation) maps. A grayscale image, $\mathbf{g} = \{g_{\mathbf{p}} : \mathbf{p} \in \mathbf{R};\ g_{\mathbf{p}} \in \mathbf{Q}\}$, takes voxel-wise values from a finite set, $\mathbf{Q} = \{0, 1, \ldots, Q-1\}$, of Q integer gray levels, i.e., $g : \mathbf{R} \rightarrow \mathbf{Q}$. A region map, $\mathbf{m} = \{m_{\mathbf{p}} : \mathbf{p} \in \mathbf{R};\ m_{\mathbf{p}} \in \mathbf{L}\}$, takes voxel-wise values from a binary set of region labels, $\mathbf{L} = \{0, 1\}$, where 0 and 1 indicate the background and kidney, respectively, i.e., $m : \mathbf{R} \rightarrow \mathbf{L}$.

A 3D geometric (level-set-based) deformable boundary, is employed for the DW-MRI kidney segmentation due to successful results in a wide range of applications, including medical imaging, e.g., for segmenting brain, prostate, liver, kidney etc. [10]. Due to simplicity, flexibility, and ability to handle complex shapes and topological changes independently of surface parametrizations, these deformable boundaries are more popular than the alternative parametric ones. Points of an object-background boundary at each time instant t are specified implicitly as a zero-level set, $B_t = \{\mathbf{p} : \mathbf{p} \in \mathbf{R};\ \Phi(\mathbf{p}, t) = 0\}$, of arguments of a specific higher-dimensional function, $\Phi(\mathbf{p}, t)$, on the lattice $\mathbf{R}$. The function is often a signed distance map

$$\Phi(\mathbf{p}, t) = \begin{cases} d(\mathbf{p}, B_t) & \text{if } \mathbf{p} \text{ is inside the boundary } B_t; \\ 0 & \text{if } \mathbf{p} \text{ is at the boundary } B_t; \\ -d(\mathbf{p}, B_t) & \text{if } \mathbf{p} \text{ is outside the boundary } B_t \end{cases}$$

where $d(\mathbf{p}, B_t) = \min_{\mathbf{b} \in B_t} d(\mathbf{p}, \mathbf{b})$ is the distance from the point $\mathbf{p}$ to the boundary B_t, and $d(\mathbf{p}, \mathbf{b})$ is the Euclidean distance between two lattice points $\mathbf{p}$ and $\mathbf{b}$, as shown in Fig. 3. The function $\Phi(\mathbf{p}, t)$ evolves in discrete time $t = n\tau$ with a fixed step, $\tau > 0$, as [28]:

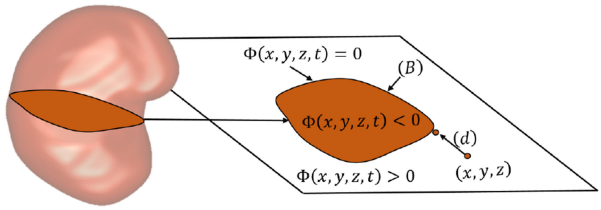

Fig. 3. 3D zero-level set of a function $\Phi(\mathbf{p} = [x, y, z], t)$.

$$\Phi(\mathbf{p}, (n+1)\tau) = \Phi(\mathbf{p}, n\tau) - \tau F_n(\mathbf{p}) |\nabla \Phi(\mathbf{p}, n\tau)| \quad (1)$$

where $n = 0, 1, 2, \ldots$, is the time index; $\nabla \Phi(\mathbf{p}, n\tau)$ is the spatial gradient of $\Phi(\mathbf{p}, n\tau)$:

$$\nabla \Phi(\mathbf{p}, n\tau) = \left[\frac{\partial \Phi(\mathbf{p}, n\tau)}{\partial x}, \frac{\partial \Phi(\mathbf{p}, n\tau)}{\partial y}, \frac{\partial \Phi(\mathbf{p}, n\tau)}{\partial z} \right];$$

$|\mathbf{a}|$ denotes the magnitude of the vector $\mathbf{a}$, and $F_n(\mathbf{p})$ is a speed function guiding the evolution of an initial boundary B_0, defined at the starting instant $t = 0$, i.e., for $n = 0$.

Most of the conventional speed functions quantify visual appearance differences between the object and its background in terms of mean values and variances of image intensities, intensity edges or gradient vector flow, and similar regional signal characteristics. Thus, their guidance may fail if images to be segmented are noisy and/or object-background contrast is low. The noise and inconsistencies due to low-frequency non-uniformity, or inhomogeneity of intensities, are suppressed in part by preprocessing that combines histogram equalization with non-parametric bias correction [29]. To accurately segment the kidneys from noisy and low-contrast DW-MRI, our guiding function accounts for not only regional kidney-background appearance, but also for a kidney shape prior and high-order spatial relations in the goal region map. To provide voxel-wise guidance for the evolving boundary, all appearance and shape descriptors are combined into a joint Markov-Gibbs random field (MGRF) model of a DW-MR image, $\mathbf{g}$, and its binary kidney-background region map, $\mathbf{m}$. The model is specified by a joint probability distribution $P(\mathbf{g}, \mathbf{m}) = P(\mathbf{g}|\mathbf{m})P(\mathbf{m})$, where $P(\mathbf{g}|\mathbf{m})$ and $P(\mathbf{m})$ denote a conditional probability distribution of images, given a map, and an unconditional distribution of region maps, respectively. The latter distribution is factored into two terms: $P(\mathbf{m}) = P_{\text{sp}}(\mathbf{m})P_{\mathbf{V}}(\mathbf{m})$, where $P_{\text{sp}}(\mathbf{m})$ denotes an appearance-based adaptive shape prior, and $P_{\mathbf{V}}(\mathbf{m})$ is a high-order Gibbs probability distribution with potentials $\mathbf{V}$. The potentials evaluate strengths of not only the nearest-neighbor pairwise dependencies, but also of triple- and quad-dependencies, which specify a higher-order spatially homogeneous MGRF model of region maps. These components of the joint image-map model are outlined below.

1) First-Order Kidney/Background Appearance Model: To accurately model DW-MRI appearance, we approximate the empirical marginal probability distribution of intensities with a linear combination of discrete Gaussians (LCDG). The LCDG with two positive dominant components (one each for the kidney

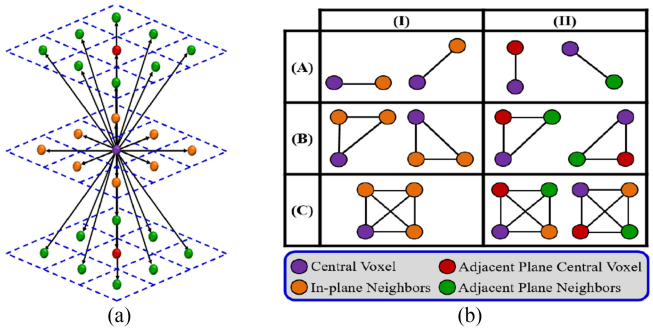

Fig. 4. The nearest 26-neighborhood (a) of a voxel for the 4th-order spatial model and (b) examples of its (A) 2nd-, (B) 3rd-, and (C) 4th-order cliques containing the central voxel (purple) and its neighbors in the same (I) and adjacent (II) planes.

and background) and multiple sign-alternate subordinate components allow for separating the mixed marginal of the DW-MRI voxel-wise intensities into the two distinct LCDGs, each associated with the kidney or background label. More details can be found in [30].

2) Higher-Order MGRF Appearance Model: Compared to other imaging modalities, lower SNRs and frequent artifacts [25], [26], together with geometric distortions due to long acquisition time [27] and larger inhomogeneities of internal structures, such as cortex and medulla, in the DW-MRI hinder the kidney segmentation. To better account for intra-kidney variabilities, spatial dependencies between each voxel and its nearest neighbors in the DW-MR images have been incorporated into our segmentation. Incorporated spatial relationships not only reduce noise impacts, but also reveal homogeneities and thus enhance the overall segmentation accuracy. Unlike more conventional pairwise-spatial homogeneity descriptors (e.g., in [31]–[33]), we use a 4th-order MGRF with analytically estimated potentials to describe those relationships. To find the potential estimates, an initial kidney map, $\mathbf{m}$, is constructed by a simple Bayes classification using joint voxel-wise shape and intensity probabilities. Then, inter-label spatial dependencies in this map, $\mathbf{m}$, are modelled by the 4th-order spatial MGRF with the nearest 26-neighborhood shown in Fig 4(a). This model adds triple and quad clique families to the more conventional 2nd-order Potts MGRF [31]. Let $\mathbf{C}_a$ be a family of s-order cliques of the interaction graph with nodes in the lattice sites $\mathbf{p} \in \mathbf{R}$ and edges connecting interdependent pairs of the sites. Let A clique families describe spatial geometry of interdependencies of region labels in the kidney maps, $\mathbf{m}$. Then the model is specified by the Gibbs probability distribution:

$$P_{\mathbf{V}}(\mathbf{m}) = \frac{1}{Z_{\mathbf{V}}} \exp\left(\sum_{a=1}^{A} \sum_{\mathbf{c}\in\mathbf{C}_a} V_a(m_{\mathbf{p}} : \mathbf{p} \in \mathbf{c})\right)$$

where $\mathbf{V} = [V_a(\boldsymbol{\mu}) : \boldsymbol{\mu}] in \{0,1\}^{\nu_a} \rightarrow (-\infty, \infty) : a = 1, \ldots, A]$ is a collection of potential functions for the families $\mathbf{C}_a$, ν_a is the clique size ($\nu_a \in \{2,3,4\}$) for the family $\mathbf{C}_a$; $\boldsymbol{\mu}$ is a label configuration on the clique, i.e., a pair, triple, or quadruple of binary numbers 0 and 1, and $Z_{\mathbf{V}}$ is the normalizing factor, called the partition function, over the entire population $\mathbb{M} = \{0,1\}^{XYZ}$ of the maps

$$Z_{\mathbf{V}} = \sum_{\mathbf{m}\in\mathbb{M}} \exp\left(\sum_{a=1}^{A} \sum_{\mathbf{c}\in\mathbf{C}_a} V_a(m_{\mathbf{p}} : \mathbf{p} \in \mathbf{c})\right)$$

For equiprobable binary labels, $m_{\mathbf{p}} \in \{0,1\}$, the marginal co-occurrence probabilities over the 2nd-, 3rd-, and 4th-order cliques are $\frac{1}{4}$, $\frac{1}{8}$, and $\frac{1}{16}$, respectively. Provided the cardinalities of the clique families are close to the lattice cardinality for all the families $a = 1, \ldots, A$ and only the equality ("eq") or inequality ("ne") of all the clique-wise labels are taken into account, the corresponding estimates of the 2nd-, 3rd-, and 4th-order potentials are as follows:

$$V_{2:a:\mathrm{eq}} = 4\left(F_{a:\mathrm{eq}}(\mathbf{m}) - \frac{1}{2}\right) = -V_{2:a:\mathrm{ne}}$$

$$V_{3:a:\mathrm{eq}_3} = \frac{16}{3}\left(F_{a:\mathrm{eq}_3}(\mathbf{m}) - \frac{1}{4}\right) = -V_{3:a:\mathrm{eq}_2}$$

$$V_{4:a:\mathrm{eq}} = \begin{bmatrix} V_{4:a:\mathrm{eq}_4} \\ V_{4:a:\mathrm{eq}_3} \\ V_{4:a:\mathrm{eq}_2} \end{bmatrix} = \lambda^* \begin{bmatrix} f_{4:a} \\ f_{3:a} \\ f_{2:a} \end{bmatrix}$$

where $F_{a:\mathrm{eq}}(\mathbf{m})$ denote relative empirical frequencies of the equal binary labels in the cliques of each family $\mathbf{C}_a$ over a given training map $\mathbf{m}$; "eq" and "ne" denote two equal or non-equal labels, respectively, for a 2nd-order clique; "eq$_i$" denote i equal labels for a 3rd- and 4th-order clique; $f_{4:a} = F_{a:\mathrm{eq}_4}(\mathbf{m}) - \frac{1}{8}$; $f_{3:a} = F_{a:\mathrm{eq}_3}(\mathbf{m}) - \frac{1}{2}$; $f_{2:a} = F_{a:\mathrm{eq}_2}(\mathbf{m}) - \frac{3}{8}$; and

$$\lambda^* = \frac{\sum_{a=1}^{A} f_{4:a}^2 + f_{3:a}^2 + f_{2:a}^2}{\sum_{a=1}^{A} \frac{7}{64} f_{4:a}^2 + \frac{1}{4} f_{3:a}^2 + \frac{15}{64} f_{2:a}^2}$$

This approximation is used for computing the higher-order spatial probabilities $\mathrm{Pr}_{\mathbf{V}:\mathbf{p}}(l)$ of each label; $l \in \mathbf{L}$.

3) Appearance-Based Shape Prior: In addition to the distinct visual appearances, the well-known geometric shapes of medical structures can enhance the segmentation accuracy. Relying on this fact, we use an adaptive model of the expected kidney shape to both handle kidney motions, e.g., due to breathing and/or heart beating, and account for the kidney's variability due to inter-patient anatomical differences. In addition, the kidney DW-MR images are very noisy, especially at high b-values.

To build the shape prior, all the b_0-scans (after excluding the test subject) of kidneys formed a training database, and these images have been manually delineated by an MRI expert (a board certified radiologist) to get their binary kidney/background region maps. One of the images was chosen as a database reference. All other images were aligned to the reference by a non-rigid 3D B-spline transformation [34] minimizing the sum of squared voxel-wise intensity differences between the two images. Then the kidney/background labels of the co-aligned region maps were used to learn the shape prior. It is worth mentioning that to select the best reference subject, we performed a normalized cross-correlation between the test subject and every other subject in the shape prior training database. The subject

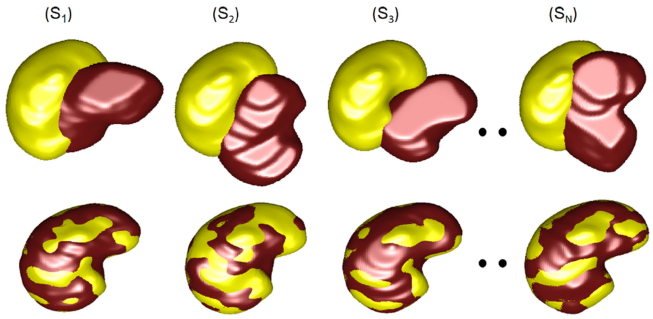

Fig. 5. 3D co-alignment of training DW-MRI to a single reference: the first and second rows present the overlapped 3D kidney volumes before and after the alignment, respectively. Note that the reference subject appears in yellow, while the targets are shown red.

with the maximum correlation was selected to be our reference. The choice of the reference has minimal effect on the shape prior as our presented shape prior depends on both the mapped spatial location in addition to signal appearance. In our presented shape prior, an adaptive search space around the mapped spatial location, which maps each voxel from the test subject to the database, is used in searching for voxels within a predefined tolerance range to the test voxel appearance. This means that any misalignment errors in the registration step will be overcome by the adaptive process. Moreover, the final segmentation takes into account the first-order appearance and the higher order spatial interaction that will overcome any errors from the shape prior segmentation. Fig. 5 illustrates the co-alignment of the training DW-MRI. Adapting the shape prior to each input DW-MR image to be segmented is guided by the visual appearance of the latter image. More details are given in [32], [33].

As DW-MRIs are challenged by motion, which might lead to different kidney masks at different b-values, each b-scan is segmented separately. The shape prior is built by segmenting manually and co-aligning the baseline scans (at b_0 s/mm^2) of all subjects. Then, the shape prior is applied to each b-scan in combination with other probabilistic models for that scan namely, the 1st-order LCDG model of kidney appearance, in terms of voxel-wise intensities, and the 4th-order MGRF model of spatial interactions. This joint probabilistic model provides a stochastic force guiding the evolution of a deformable boundary to segment kidneys at that b-scan. So, the segmentation of all b-value scans uses the same shape prior, but own intensity and spatial interactions models. Algorithm 1 summarizes creating and updating the appearance-guided shape prior for each test DW-MR image to be segmented (test images are first removed from the training set).

4) Appearance- and Shape-Guided Deformable Model: Adaptation to both the kidney-background visual appearance, shape prior, and statistical spatial dependencies between kidney labels is one of the main advantages of our segmentation framework. Estimated directly from the input image and a given shape database, these properties guide the evolving deformable boundary by defining, for each voxel $\mathbf{p}$ with intensity $g_{\mathbf{p}} = q$, the speed function [10] of Eq. (1), $F_n(\mathbf{p}) = \kappa\vartheta_{\mathbf{p}}$, where κ is

Algorithm 1: Creating/Updating the Shape Prior.

1: Preprocess the training DW-MR images by bias correction and histogram equalization.
2: Construct the shape database by applying the co-alignment by Glocker *et al.* [34] to the preprocessed DW-MR images.
3: Preprocess the DW-MR image for a test subject and co-align with the shape database.
4: For each voxel, $\mathbf{p} \in \mathbf{R}$, in the test DW-MR image, $\mathbf{g}_{\text{test}}$, calculate its prior shape probabilities as follows:
   A Use the co-aligning deformation field to relate the voxel $\mathbf{p}$ of the test image to the shape database lattice.
   B Initialize a 3D window of size $N_1 \times N_2 \times N_3$, centered at the related voxel in the shape database lattice.
   C Find within the window all the voxels with the corresponding intensity, $g_{\text{test}:\mathbf{p}}$, in all the training images.
   D If necessary, increase the window size and repeat Steps 4B to 4D until a non-empty set of such corresponding training intensities is found.
   E Estimate label probabilities based on relative occurrences of each label in all the training voxels found.

the mean contour curvature and $\vartheta_{\mathbf{p}}$ specifies the magnitude and direction of moving that voxel

$$\vartheta_{\mathbf{p}} = \begin{cases} -\Pr_{\mathbf{p}}(1) & \text{if } \Pr_{\mathbf{p}}(1) > \Pr_{\mathbf{p}}(0);\ \text{i.e., } \Pr_{\mathbf{p}}(1) > 0.5; \\ \Pr_{\mathbf{p}}(0) & \text{otherwise} \end{cases} \tag{2}$$

Here, $\Pr_{\mathbf{p}}(0)$ and $\Pr_{\mathbf{p}}(1)$ are the voxel-wise background and kidney probabilities, respectively

$$\Pr_{\mathbf{p}}(1) = \frac{\Omega_{\text{kd}:\mathbf{p}}}{\Omega_{\text{kd}:\mathbf{p}} + \Omega_{\text{bg}:\mathbf{p}}};\ \Pr_{\mathbf{p}}(0) = \frac{\Omega_{\text{bg}:\mathbf{p}}}{\Omega_{\text{kd}:\mathbf{p}} + \Omega_{\text{bg}:\mathbf{p}}} = 1 - \Pr_{\mathbf{p}}(1)$$

where the variables $\Omega_{\text{kd}:\mathbf{p}}$ and $\Omega_{\text{bg}:\mathbf{p}}$ for the kidney and background, respectively, depend on the voxel-wise probabilities $\Pr(q|l)$; $l \in \mathbf{L}$, for the LCDG submodels of the kidney ($l = 1$) or background ($l = 0$) appearance and on the kidney label probability in the MGRF spatial region map model, $\Pr_{\mathbf{V}:\mathbf{p}}(1)$, and in the adaptive shape prior, $\Pr_{\text{sp}:\mathbf{p}}(1)$, respectively:

$$\Omega_{\text{kd}:\mathbf{p}} = \Pr(q|1)\Pr_{\mathbf{V}:\mathbf{p}}(1)\Pr_{\text{sp}:\mathbf{p}}(1)$$
$$\Omega_{\text{bg}:\mathbf{p}} = \Pr(q|0)(1 - \Pr_{\mathbf{V}:\mathbf{p}}(1)(1 - \Pr_{\text{sp}:\mathbf{p}}(1))$$

Algorithm 2 summarizes the basic steps of the 3D level-set-based kidney segmentation.

B. Estimating and Depicting Diffusion Parameters

In order to accurately evaluate image features, the segmented kidneys are processed to eliminate their motion, being a common challenge in DW-MRIs. The motion in the scans at different

Algorithm 2: DW-MRI Segmentation by Geometric Deformable Boundary.

1: Update the shape prior probability using Step 4 of Algorithm 1.
2: Approximate the marginal of DW-MRI intensities with the LCDG [30] with two dominant components.
3: Form an initial region map, $\mathbf{m}_{\text{ini}}$, using the estimated shape prior and LCDG submodes of kidney and background appearances.
4: Estimate the Gibbs potentials for the 4th-order spatial MGRF map model from $\mathbf{m}_{\text{ini}}$.
5: Find the above speed function [10], $F_n(\mathbf{p})$, using results of Steps 2 to 4.
6: Segment the input image, $\mathbf{g}$, by evolving the level-set function, $\Phi(\mathbf{p}, n\tau)$, of Eq. (1) with the speed function found in Step 5.

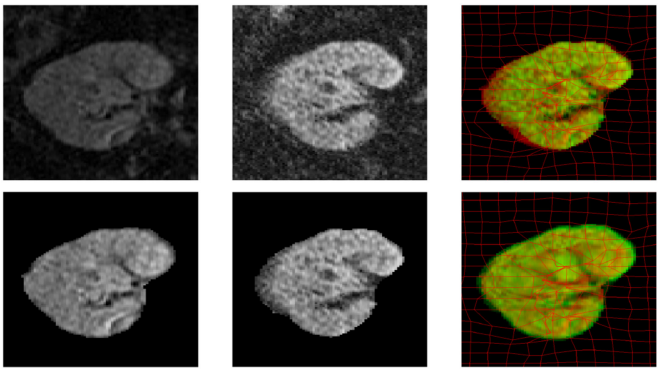

Fig. 6. Typical 2D cross section of the b_0 scan and its segmented kidney (first-column), the b_{300} scan and its segmented kidney (second-column), and the blending overlap before (*upper-row*) and after (*lower-row*) the non-rigid registration for motion handling (third-column).

b-values may result in mis-overlaps of kidneys and false correspondences for the same voxel at different phases, leading to inaccurate evaluations of the discriminative features (i.e., ADCs) to be used to derive the final diagnosis. To overcome this problem, another stage of the non-rigid 3D B-spline registration [34] is applied to both the DW-MRI scans and the segmented mask to map each b-value scan to the baseline b_0-scan. While this additional registration does not eliminate the signal pile-up, it accounts for any potential motion between the same kidney at different b-values and ensures accurate voxel-to-voxel matching for better evaluation of the ADC map. Fig. 6 shows how the non-rigid registration handles any potential misalignment of the subject's segmented kidney.

After segmenting the kidneys and excluding potential motions between the different b-value scans, their discriminatory functional features are evaluated from the images and used to distinguish between the rejected and non-rejected kidney transplants. As a transplant status feature, ARTR-CAD system uses the voxel-wise ADCs defined by Le Bihan [35]

$$\text{ADC} = \frac{1}{b_0 - b} \ln\left(\frac{g_{b:\mathbf{p}}}{g_{0:\mathbf{p}}}\right) = \frac{\ln g_{b:\mathbf{p}} - \ln g_{0:\mathbf{p}}}{b_0 - b} \qquad (3)$$

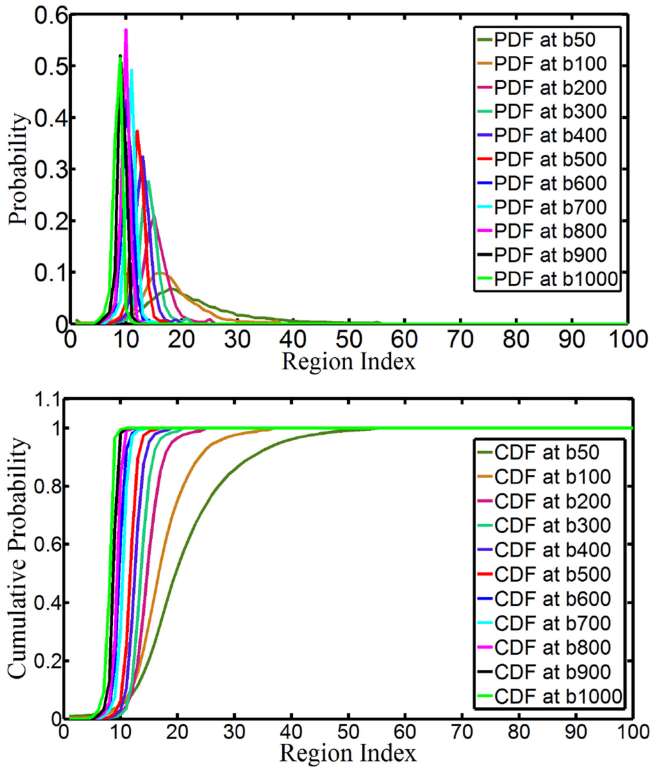

Fig. 7. Empirical ADC distributions and their cumulative distribution functions (CDFs) for one subject at the 11-different b-values.

where the segmented DW-MR images $\mathbf{g}_0$ and $\mathbf{g}_b$ were acquired with the b_0 and a different b-value, respectively.

It is worth noting that conventional classification methods that deals directly with the voxel-wise ADCs of the entire kidney volume as discriminative kidney features face two difficulties: (*i*) varying sizes of input data require either data truncation or zero padding for larger or smaller kidney volumes, respectively, and (*ii*) large data volumes lead to considerable time expenditures for training and classification.

The presented ARTR-CAD system overcomes the above challenges by characterizing the entire 3D ADC maps, collected for each subject at the 11 different b-values, by the CDFs of the ADCs, as shown in Fig. 7. The CDFs are independent of the data size and can be quantified in accordance with the actual signal-to-noise ratio (SNR) of the ADCs. Preliminary experiments have shown that the 1%-accuracy of measuring the ADC within the actual range between the maximum and minimum ADCs for all the b-values and subjects is sufficient to maintain the final classification accuracy. Compared to the empirical probability density functions (PDFs) of the ADCs, the CDFs allow for better differentiation between the PDFs across the whole range of the ADC values. The training CDFs are used for deep learning of a stacked NCAE (SNCAE) classifier detailed in Section III-C. Limiting the input data size to the 11 CDFs helps to overcome the above difficulties of the arbitrary numbers of the original ADCs and notably accelerates the classification. Simultaneously, all the 3D ADCs evaluated for a certain subject

can be displayed as voxel-wise parametric color-coded 3D maps to be visually assessed by radiologists.

C. Autoencoding and Deep Learning-Based Classifier

A rich variety of learnable classifiers with shallow structure [36]–[39] have been used in the CAD systems for organ transplantation prediction using clinical and demographic data of patients, including (but not limited to) artificial neural networks (ANN), support vector machines (SVMs), regression trees, random forests (RFs), decision trees (DTs), k-nearest neighbor (kNN), etc. Compared to classical shallow models, recent advances in deep learning of ANNs [40]–[42] provide more accurate classification ability by: (*i*) automated dimensionality reduction of large scale data [43]; (*ii*) automatic hierarchical feature extraction of more discriminatory features between classes. In this case, the high level (global) features are derived from the low level (local) features for model training [43]–[46]; and (*iii*) flexibility, i.e., a classifier (e.g., a softmax-based or SVM-based) can be built on the extracted features from the deep learning ANN [43], [47].

Due to the aforementioned advantages, the fully automated ARTR-CAD system utilizes deep learning and auto-encoders with non-negativity constraint (NCAE) as a core ANN architecture for pre-training and classification to distinguish between the non-rejection and acute rejection of kidney transplants. In particular, ARTR-CAD system employs a deep neural network with a stack of auto-encoders (AE) before the output layer that computes a softmax regression, generalizing the common logistic regression to more than two classes. Each AE compresses its input data to capture the most prominent variations and is built separately by greedy unsupervised pre-training [48]. The softmax output layer facilitates the subsequent supervised back-propagation-based fine tuning of the entire classifier by minimizing the total loss (negative log-likelihood) for a given training labeled data. Using the AEs with a non-negativity constraint (NCAE) yields both more reasonable data codes (features) during its unsupervised pre-training and better classification performance after the supervised refinement [49].

Let $\mathbf{W} = \{\mathbf{W}^{\mathrm{e}}_j, \mathbf{W}^{\mathrm{d}}_i : j = 1, \ldots, s;\ i = 1, \ldots, n\}$ be a set of column vectors of weights for encoding (e) and decoding (d) layers of a single AE in Fig. 8. The AE converts an n-dimensional column vector $\mathbf{u} = [u_1, \ldots, u_n]^{\mathsf{T}}$ of input signals, where T denotes vector transposition, into an s-dimensional column vector $\mathbf{h} = [h_1, \ldots, h_s]^{\mathsf{T}}$ of hidden codes (features, or activations), such that $s \ll n$, by a uniform nonlinear transformation of s weighted linear combinations of signals

$$h_j = \sigma\left(\left(\mathbf{W}^{\mathrm{e}}_j\right)^{\mathsf{T}}\mathbf{u}\right) \equiv \sigma\left(\sum_{i=1}^{n} w^{\mathrm{e}}_{j:i} u_i\right)$$

where $\sigma(\ldots)$ is a certain sigmoid, i.e., a differentiable monotone scalar function with values in the range $[0, 1]$. Unsupervised pre-training of the AE minimizes total deviations between each given training input vector $\mathbf{u}_k$; $k = 1, \ldots, K$, and the same-dimensional vector, $\widehat{\mathbf{u}}_{\mathbf{W}:k}$ reconstructed from its code, or activation vector, $\mathbf{h}_k$ (Fig. 8(a)). The total reconstruction error of applying such AE to compress and decompress the K training

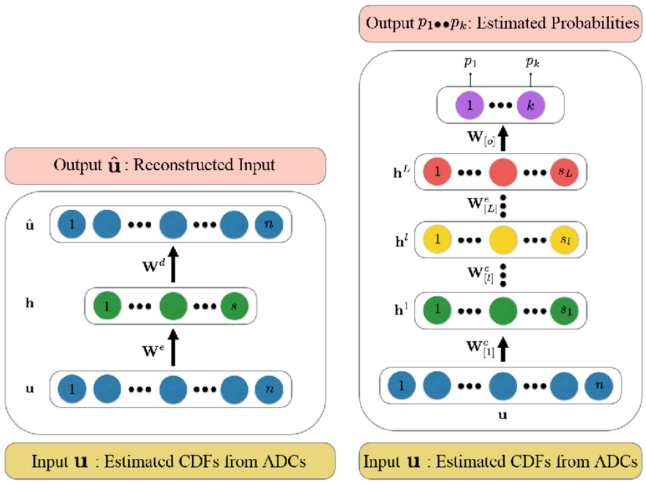

Fig. 8. Illustrative block-diagram of a NCAE (a) and a SNCAE (b) classifiers demonstrating the reconstruction error minimization process to get the updated hidden weight activations $\mathbf{W}$ such that the output $\widehat{\mathbf{u}}$ of the NCAE known as reconstructed input is the same size of the input $\mathbf{u}$ and should be almost the same if we have a good nonlinear transformation through the hidden activators.

input vectors integrates the ℓ_2-norms of the deviations

$$J_{\mathrm{AE}}(\mathbf{W}) = \frac{1}{2K}\sum_{k=1}^{K} \parallel \widehat{\mathbf{u}}_{\mathbf{W}:k} - \mathbf{u}_k \parallel^2 \tag{4}$$

To reduce the number of negative weights and enforce sparsity of the NCAE, the reconstruction error of Eq. (4) is appended, respectively, with quadratic negative weight penalties, $f(w_i) = (\min\{0, w_i\})^2$; $i = 1, \ldots, n$, and Kullback-Leibler (KL) divergence, $J_{\mathrm{KL}}(\gamma||\mathbf{h}_{\mathbf{W}^{\mathrm{e}}})$, of activations, $\mathbf{h}_{\mathbf{W}^{\mathrm{e}}}$, obtained with the encoding weights $\mathbf{W}^{\mathrm{e}}$ for the training data, from a fixed small positive average value, γ, near 0

$$J_{\mathrm{NCAE}}(\mathbf{W}) = J_{\mathrm{AE}}(\mathbf{W}) + \alpha\sum_{j=1}^{s}\sum_{i=1}^{n} f(w_{j:i}) + \beta J_{\mathrm{KL}}(\gamma||\mathbf{h}_{\mathbf{W}^{\mathrm{e}}}) \tag{5}$$

Here, the factors $\alpha \geq 0$ and $\beta \geq 0$ specify relative contributions of the non-negativity and sparsity constraints to the overall loss, $J_{\mathrm{NCAE}}(\mathbf{W})$, and $J_{\mathrm{KL}}(\gamma||\mathbf{h}_{\mathbf{W}^{\mathrm{e}}}) = \sum_{j=1}^{s} h_{\mathbf{W}^{\mathrm{e}}:j} \log(\frac{h_{\mathbf{W}^{\mathrm{e}}:j}}{\gamma}) + (1 - h_{\mathbf{W}^{\mathrm{e}}:j}) \log(\frac{1-h_{\mathbf{W}^{\mathrm{e}}:j}}{1-\gamma})$.

The classifier is built by stacking the NCAE layers with an output softmax layer, as shown in Fig. 8(b). Each NCAE is pre-trained separately in the unsupervised mode by using the activation vector of a lower layer as the input to the upper layer. In our case, initial input data consisted of 11 CDFs, each of size 100, i.e., $n = 1100$. In other words, for quantizing the ADCs, the range between the minimum and maximum ADCs for all the input data sets (i.e., all the sets for 11 b-values and 100 subjects) was divided into 100 steps to keep the chosen 1%-accuracy of initial ADC measurements. The PDFs and then CDFs of the ADCs were built for these quantized values. The bottom NCAE compresses the input vector to s_1 first-level activators, compressed by the next NCAE to s_2 second-level activators, which are reduced in turn by the output softmax layer to s°

values. The number of the NCAE layers and successive data compression ratios for each layer were chosen empirically, on the basis of comparative experiments.

Separate pre-training of the first and second layers by minimizing the loss of Eq. (5) reduces the total reconstruction error, as well as increases sparsity of the extracted activations and numbers of the non-negative weights. The activations of the second NCAE layer, $\mathbf{h}^{[2]} = \sigma(\mathbf{W}^{e}_{[2]}{}^{\mathsf{T}}\mathbf{h}^{[1]})$, are inputs of the softmax classification layer, as sketched in Fig. 8(b) to compute a plausibility of a decision in favor of each particular output class, $c = 1, 2$

$$p(c; \mathbf{W}_{\circ:c}) = \frac{\exp(\mathbf{W}^{\mathsf{T}}_{\circ:c}\mathbf{h}^{[2]})}{\exp(\mathbf{W}^{\mathsf{T}}_{\circ:1}\mathbf{h}^{[2]}) + \exp(\mathbf{W}^{\mathsf{T}}_{\circ:2}\mathbf{h}^{[2]})}; \quad c = 1, 2$$

and $\sum_{c=1}^{2} p(c; \mathbf{W}_{\circ:c}; \mathbf{h}^{[2]}) = 1$. Its separate pre-training minimizes the total negative log-likelihood $J_{\circ}(\mathbf{W}_{\circ})$ of the known training classes, appended with the negative weight penalties

$$J_{\circ}(\mathbf{W}^{o}) = -\frac{1}{K}\sum_{k=1}^{K}\log p(c_k; \mathbf{W}_{\circ:c}) + \alpha\sum_{c=1}^{2}\sum_{j=1}^{s_2} w_{\circ:c:j} \quad (6)$$

Finally, the entire stacked NCAE classifier (SNCAE) is fine-tuned on the labeled training data by the conventional error back-propagation through the network and penalizing only the negative weights of the softmax layer. To find the optimal set of hyper-parameters for the presented SNCAEs, namely, the desired average activation of the hidden units (from 0.05 to 0.9), weight of sparsity penalty term (from 1 to 20), and weight decay parameter (from 10^{-2} to 10^{-7}), a grid search algorithm with the reconstruction error as an optimization metric was employed. At this point, the two-layer SNCAE with the first-layer hidden units ($s_1 = 50$), second-layer hidden units ($s_2 = 5$), output softmax layer ($s^{\circ} = 2$), weight decay parameter ($\alpha = 3 * 10^{-5}$), weight of sparsity penalty term ($\beta = 3$), and desired average activation of the hidden units ($\gamma = 0.1$) gave the best diagnostic accuracy using leave-one-out, four-fold, and 10-fold cross validation scenarios and was accepted for the presented CAD system. Algorithm 3 summarizes classification of kidney transplant status and generation of color ADC maps.

IV. Experimental Results

Segmentation Results: Since the segmentation is an important step in developing any CAD system to assess renal function, we provide a comprehensive analysis of the segmentation results. The performance of the presented segmentation was tested first on the collected DW-MRI data (total of 100 subjects). Fig. 9 shows some segmentation results for different kidney cross-sections (coronal, axial, and sagittal) for two subjects at b_0. The segmentation accuracy was evaluated by three volumetric measures, namely, the Dice similarity coefficient (DSC) [50] in Eq. (7) and percentage kidney volume difference (PKVD) in Eq. (8), volume overlap ratio (VOR) [51] in Eq. (9), and one distance-based metric – the 95-percentile modified Hausdorff distance (MHD) [52] in Eqs. (10,11), which characterize the spatial overlap and distribution of the surface to surface distances between the segmented and ground truth kidneys,

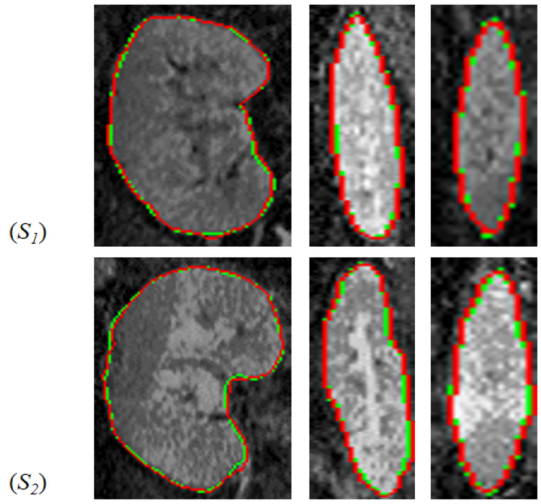

Fig. 9. Our segmentation (red) with respect to the expert's manual ground truth (green): the coronal (left), axial (middle), and sagittal (right) cross-sections for two subjects, S_1 and S_2.

Algorithm 3: Kidney Transplant Status Classification and ADC Color Mapping.

1: Calculate, using Eq. (3), the ADCs at different b-values for the entire transplanted kidney of each subject.
2: *Classification:*
   A Construct the CDFs of the calculated ADCs over the entire kidney volume at different b-values.
   B Use a SNCAE-based deep ANN classifier trained by unsupervised pre-training and supervised fine tuning together with a leave-one-subject-out (LOSO) approach to discriminate rejection from non-rejection status and get the final diagnosis.
3: *Generation of color ADC maps:* Generate voxel-wise color-coded maps of the ADCs calculated in Step 1 to demonstrate visually perceived differences between the rejection and non-rejection states of kidney transplants at different b-values.

respectively. The ground truth kidney maps were manually outlined by an MRI expert (a board certified radiologist)

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (7)$$

where TP, FP, and FN denote the true positive, false positive, and false negative respectively. Better segmentation is indicated by higher DSC values (ideal segmentation is indicated by a DSC value of one). If there is no overlap, DSC will be assigned a value of zero.

$$PKVD\% = \frac{|GT| - |SR|}{|GT|} * 100 \quad (8)$$

where $|GT|$ and $|SR|$ indicate the total number of voxels in the ground truth and the segmented region, respectively.

$$VOR(GT, SR) = \frac{|GT \cap SR|}{\min(|GT|, |SR|)} \quad (9)$$

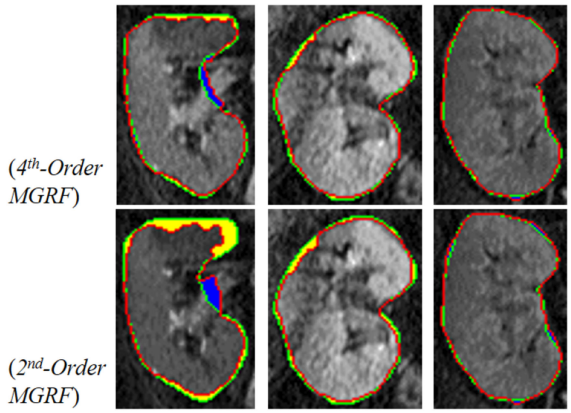

Fig. 10. Our segmentation (red) with respect to the expert's manual ground truth (green) using the 4th-order MGRF (first row) compared to our previous segmentation using the 2nd-order MGRF (second row) [32], [33] for three different subjects (columns), where the first, second, and third columns show large, moderate, and small differences (yellow or false positive (FP) and blue or false negative (FN) regions), respectively.

TABLE I
SEGMENTATION ACCURACY BY THE DSC, MHD(MM), VOR, AND PKVD(%)

	2^{nd}-MGRF [32], [33]			4^{th}-MGRF		
Metric	**Min**	**Max**	**mean±SD**	**Min**	**Max**	**mean±SD**
DSC	0.89	0.96	$0.93_{\pm 0.17}$	0.92	0.97	$0.96_{\pm 0.04}$
MHD (mm)	4.00	8.00	$5.43_{\pm 1.70}$	2.50	6.37	$4.00_{\pm 0.77}$
VOR	0.88	0.98	$0.93_{\pm 0.03}$	0.93	0.99	$0.97_{\pm 0.16}$
PKVD (%)	7.00	19.6	$13.8_{\pm 2.90}$	5.50	16.0	$9.3_{\pm 2.50}$

All metrics are represented by the minimum (Min), maximum (Max), and mean$_{\pm}$standard deviation values.

If set GT is a subset of SR or the converse, then the overlap coefficient is equal to one.

$$HD(A_1, A_1) = \max_{c \in A_1}\{\min_{e \in A_2}\{d(c,e)\}\} \tag{10}$$

where c and e denote points of set A_1 and A_2 respectively, and $d(c,e)$ is the Euclidean distance between c and e. The Modified bidirectional HD (MHD_{Bi}) between SR and its GT is defined as:

$$MHD_{Bi}(GT, SR) = \max\{HD(GT, SR), HD(SR, GT)\} \tag{11}$$

To show the effect of adding the higher-order MGRF model to our segmentation, we compared the current results with our previous segmentation using the 2nd order- MGRF [32], [33], as shown in Fig. 10. In addition, Table I compares both segmentation methods in terms of the DSC, MHD, VOR, and PKVD metrics. As shown in Fig. 10 and documented in Table I, our segmentation has been notably enhanced after adding the higher-order MGRF. This can be explained in part by abilities of the latter model to capture more intricate inhomogeneities of grey levels between different structures, e.g., cortex and medulla.

As shown in Table I, the high accuracy of our segmentation is confirmed by the DSC, MHD, VOR, and PKVD statistics for all the test data sets. Its favorable comparisons with other

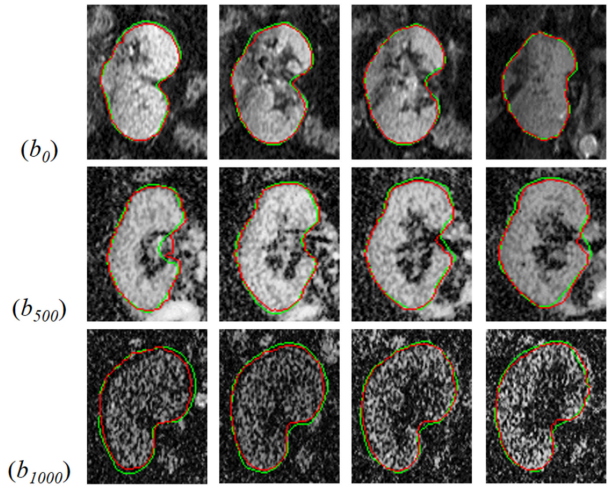

Fig. 11. Additional four coronal cross-sections (columns) for our segmentation of the DW-MRI acquired at different b-values s/mm^2.

TABLE II
DIAGNOSTIC ACCURACY BASED ON THE INPUT CDF FOR INDIVIDUAL b-VALUES, THE FUSED CDFS OF THE REDUCED MODEL INCORPORATING ONLY ADC AT b_{100} AND b_{700} s/mm^2 (F_2), AND THE FUSED CDFS OF THE 11 b-VALUES (F_{11})

Classification Accuracy% ≈												
b_{50}	b_{100}	b_{200}	b_{300}	b_{400}	b_{500}	b_{600}	b_{700}	b_{800}	b_{900}	b_{1000}	F_2	F_{11}
76	85	75	80	71	77	63	62	63	64	72	92	**97**

methods can be found in [32], [33]. In particular, the developed segmentation proved its ability to precisely segment the kidney at higher b_i values. Shown in Fig. 11, coronal cross-sectional segmentation results for three other subjects at b_0 and higher b_i values (b_{500} and b_{1000}) also confirm the high accuracy and robustness to noise of our segmentation.

Diagnostic Results: Following kidney segmentation, ARTR-CAD system classifies the transplant status with the SNCAE-based classifier and CDFs as discriminatory features. A leave-one-subject-out (LOSO) scenario applied to distinguish between the rejection and non-rejection cases from the CDFs of the ADCs for different b-values has correctly classified 97.0% of the cases, namely, 67 out of the 70 rejected and 30 out of the 30 non-rejected kidney transplants. It is worth noting that fusing (by concatenation) all the CDFs calculated at the 11-different b-values improves the final diagnostic accuracy, as shown in Table II, and helps to overcome the challenge of possible artifacts due to chemical shifts, which could occur at one or two b-values, if the artifacts exist.

To investigate the best informative model (i.e., by sacrificing some b-values that may be less important) that can provide a good diagnostic accuracy in differentiating AR from NR renal transplants, variable subset selection was performed using Akaike information criterion (AIC). All the $2^{11} - 1$ models were tested, considering all possible subsets of mean ADC, and the model yielding the lowest AIC was selected as the most informative. The reduced models with the 3-lowest AIC values were the model incorporating ADC at b_{100} and b_{700} s/mm^2 with

TABLE III
DIAGNOSTIC ACCURACY FOR OUR CAD SYSTEM FOR OUR LEVEL-SET SEGMENTATION-BASED 4TH ORDER-MGRF SPATIAL INTERACTIONS, LEVEL-SET SEGMENTATION-BASED ONLY 2ND ORDER-MGRF SPATIAL INTERACTIONS, AND THE LEVEL-SET APPROACH BY (CV)

(Diagnostic Accuracy)% Based on Segmentation		
CV [53]	2^{nd} order-MGRF [32], [33]	4^{th} **order-MGRF**
80.0	95.0	**97.0**

AIC = 58.6, the model incorporating ADC at b_{100} and b_{500} s/mm^2 with AIC = 58.7, and the model incorporating ADC at b_{300}, b_{500}, b_{600}, and b_{700} s/mm^2 with AIC = 59.2, respectively. Therefore, we were motivated to investigate the capability of a new reduced model with the lowest AIC value that integrates only ADC at b_{100} and b_{700} s/mm^2. An accuracy of 92% was obtained, which demonstrates that these two b-values are superior in differentiating the AR from the NR renal transplants. However, we must use all the 11-different b-values available in building the model to get the best (maximum accuracy), as shown in Table II. This can be explained in part by the highly nonlinear nature of the AEs based classifier, which can partition the feature space into regions of nearly arbitrary shape compared to the logistic regression based classifier, which divides feature space along a hyperplane.

The advantages of the presented segmentation technique using up to the 4th order-MGRF spatial interactions on the diagnostic accuracy are highlighted by calculating the ADCs and CDFs for the segmented kidney regions obtained with our previously developed segmentation method using only the 2nd order-MGRF spatial interactions [32], [33] and the Chan and Vese (CV) [53] segmentation method. In each case, the constructed CDFs were used to train and test the SNCAE classifier using the same LOSO scenario. Table III shows that the overall diagnostic accuracies were reduced to 95.0% and 80.0% for the segmentation using the 2nd order-MGRF model and CV segmentation methods, respectively. These results suggest that our better diagnostics can be explained in part by the more accurate and reliable segmentation of the 3D kidneys.

To highlight the advantages of the presented technique over the currently used technique by clinicians at hospitals, we compared the diagnostic accuracy of our method with the current clinical well validated method, which is based on using the clinical biomarkers (i.e., serum plasma creatinine (SPCr)) in which an average basal level for abnormality of 1.3 mg dl^{-1} is used. Deviations from the patient's SPCr basal level were detected using a LOSO approach along with a k-nearest neighbor (kNN) classifier and were used for abnormality detection and a "for cause" biopsy needs to be performed. An additional method that is currently used in the clinical research by clinicians is to place a 2D ROI on the largest mid cross-section of the kidney, using a certain pair of b-values (e.g., b_0 and b_{1000}) s/mm^2. Next, the average ADC value is calculated from this cross-section for each subject. Then, a LOSO approach is used along with a kNN classifier to find the final diagnosis. Comparison of our method accuracy with the aforementioned clinical methods is summarized in Table IV. As documented in Table IV, our approach shows the superiority over the currently used and well-validated clinical methods by clinicians in hospitals and clinical research.

TABLE IV
COMPARATIVE ACCURACY OF ARTR-CAD TO THE STANDARD CLINICAL METHODS USED BY PHYSICIANS

DiagTech	ARTR-CAD	SPCr	manual 2D-ROI
DiagAcc%	**97.0**	73.0	67.0

Note that DiagTech, DiagAcc, and SPCr denote diagnostic technique, diagnostic accuracy, and serum plasma creatine, respectively.

TABLE V
DIAGNOSTIC ACCURACY, SENSITIVITY, AND SPECIFICITY FOR OUR CAD SYSTEM WITH THE SNCAE CLASSIFIER USING DIFFERENT CDF ENCODING STEPS (Δ_i)

	Quality of classification% ≈		
Δ_i:	Accuracy	Sensitivity	Specificity
Δ_1 = 0.01	**97**	**96**	**100**
Δ_2 = 0.02	88	90	83
Δ_3 = 0.04	82	83	80

TABLE VI
DIAGNOSTIC ACCURACY, SENSITIVITY, AND SPECIFICITY FOR OUR CAD SYSTEM WITH THE SNCAE CLASSIFIER USING DIFFERENT STRUCTURES, I.E., DIFFERENT NUMBER OF HIDDEN LAYERS (l) AND HIDDEN NODES AT EACH LAYER (s_l), USING THE SAME INPUT SIZE OF 1100 (11 CDFS EACH OF 100 REGION), $\alpha = 3 * 10^{-5}$, $\beta = 3$, AND $\gamma = 0.1$

	Quality of classification% ≈		
SNCAE Structure:	Accuracy	Sensitivity	Specificity
$s_1 = 5$	85	84	90
$s_1 = 25$	65	67	60
$s_1 = 50$	71	71	70
$s_1 = 50$ and $s_2 = 5$	**97**	**96**	**100**
$s_1 = 50$, $s_2 = 25$, and $s_3 = 5$	83	93	60

In order to evaluate the effect of the CDFs encoding step (Δ) on the overall accuracy, we reconstructed the CDFs using two different Δ_i ($\Delta = 0.02$ and 0.04). Then, we applied our SNCAE classifier on the reconstructed CDFs and the results are shown in Table V. As demonstrated in Table V, the overall accuracy, sensitivity, and specificity, have been greatly reduced. This can be explained in part by the fact that increasing the value of Δ results in losing important data information, thus making the data not well-presented, which in turn affects the classifier performance.

Furthermore, the effect of changing the SNCAE structure on the overall accuracy has been investigated by using different SNCAE layouts (different number of hidden layers (l) and hidden nodes at each layer (s_l)). From the results in Table VI, the network structure with two hidden layers $s_1 = 50$ and $s_2 = 5$, demonstrated the highest accuracy.

In addition to the LOSO approach, we have performed a four-fold cross-validation test where 75% of the data was used for training and the other 25% for testing, and a 10-fold cross-validation test where 90% of the data was used for training and the other 10% for testing, to further validate and justify the performance of the developed SNCAE classifier. As documented

TABLE VII
SENSITIVITY TO A TRAINING SET BASED ON FOUR-FOLD AND 10-FOLD CROSS-VALIDATION SCENARIOS

	four-fold Cross-validating SNCAE			
Testing group (25%)	1	2	3	4
Correct/All	23/25	24/25	24/25	25/25
Average accuracy,%	96.0			
	10-fold Cross-validating SNCAE			
Average accuracy,%	94.0			

TABLE VIII
DIAGNOSTIC RESULTS IN TERMS OF CORRECTLY CLASSIFIED VS. TRUE NON-REJECTION (NR) AND REJECTION (R) CASES, ACCURACY (ACC), SENSITIVITY (SENS), SPECIFICITY (SPEC), AND AREA UNDER THE CURVE (AUC) FOR THE PRESENTED CAD SYSTEM WITH OUR SNCAE CLASSIFIER AND EIGHT CLASSIFIERS FROM THE WEKA COLLECTION [54]

	Classification Performance% ≈					
	NR	AR	Acc	Sens	Spec	AUC
K*	25/30	55/70	80	79	83	0.81
NBT	24/30	62/70	86	89	80	0.85
MCC	28/30	65/70	93	93	93	0.92
Decorate	20/30	62/70	82	89	67	0.82
RT	18/30	60/70	78	86	60	0.75
RF	22/30	65/70	87	93	73	0.85
LDA	25/30	66/70	91	94	83	0.92
SVM_{lin}	24/30	66/70	90	94	80	0.90
SVM_{Qaud}	23/30	64/70	87	91	77	0.87
SVM_{Cub}	23/30	65/70	88	93	77	0.88
SVM_{Gauss}	24/30	66/70	90	94	80	0.90
SNCAE	**30/30**	**67/70**	**97**	**96**	**100**	**0.96**

in Table VII, the diagnostic accuracy of the combined SNCAE classification system is almost independent of the choice of the training and testing data sets. The four-fold and 10-fold cross-validation experiments demonstrated an average accuracy of 96.0% and 94.0%, respectively.

To evaluate capabilities of the developed SNCAE classifier, it has been compared with eight well-known learnable classifiers from the Weka collection [54]: K*, Naive Bayes Tree (NBT), Multi-Class Classifier (MCC), Decorate, Random Tree (RT), Random Forest (RF), linear discriminant analysis (LDA), and support vector machine (SVM) with four different kernels. Table VIII presents their and our diagnostic accuracy in terms of the numbers of correctly classified AR and NR cases with respect to the overall numbers of subjects, sensitivity, specificity, and area under the curve (AUC). Our classifier demonstrated the best total diagnostic accuracy of 97.0% with 100 % specificity (30 out of 30 correctly classified NR transplants), and 96.0% sensitivity (67 out of 70 correctly classified AR transplants). In addition, the SNCAE demonstrated the highest AUC (approaches the top-most unit value). These initial diagnostic results confirm that the presented CAD system holds promises as a reliable non-invasive diagnostic tool. To better compare with our classifier, all other classifiers were tuned several times until best accuracies were obtained.

We assessed statistical significance of the presented SNCAEs classifiers performance using permutation testing [55]. The null distribution, with no relationship between ADC features and the renal allograft status, was simulated by randomly permuting the labels (AR, NR) of the training data set. The classifier was trained on these simulated data using the same procedure as for real data, including hyper-parameter optimization. LOSO cross-validation estimate of accuracy was used as the metric of the SNCAEs classifier performance. Over 1000 simulated data sets, LOSO accuracy never met or exceeded that of the presented SNCAEs classifier trained on real data, resulting in a bootstrap P-value of 0.001.

Together with the automated classification, the presented CAD system also provides voxel-wise parametric ADC maps, which can help in local visual assessment of the transplanted kidney. The color-coded ADC map, which depicts the estimated voxel-wise ADCs of Eq. 3, is more informative than the ADC CDFs collected over the entire kidney and thus hides local peculiarities of spatial behavior of the ADCs. Examples of the ADC maps for two non-rejection and two acute rejection cases for the DW-MRI at the different b-values in Fig. 12 demonstrate some expected relations between the local ADCs, which could be helpful for assessing the transplant and detecting its non-rejection or rejection status.

V. DISCUSSION

Performance of the Presented Technique: After testing the presented ARTR-CAD system on a biopsy-proven cohort of 100 participants in a leave-one-subject-out scenario, our system showed an overall diagnostic accuracy of 97.0% in detecting rejected and non-rejected kidney transplants.

It is worth noting that this high diagnostic accuracy of the presented CAD system is obtained due to its ability to (*i*) capture blood perfusion and diffusion inside the transplanted kidney by collecting and processing data at both low and high b-values; (*ii*) establish voxel-to-voxel matches on the segmented transplanted kidney by 3D non-rigid registration; and (*iii*) use considerably overlapped features for training the SNCAE-based classifier with an enhanced convergence rate.

Currently, the entire processing pipeline is implemented in MATLAB using a Dell Precision WorkStation T7500 (8 Intel Xeon W5590 CPUs @ 3.33 GHz with 48 GB RAM and 4 TB RAID hard drive). For an input DW-MRI data set of size $256 \times 256 \times 50$ voxels of each b-value scan, the segmentation takes around 1.5 min; the non-rigid registration after segmentation takes around 0.33 min; and the feature extraction and final SNCAE-based diagnosis take around 0.17 min. For processing all b-value scans and obtaining final classification, the entire processing takes about 20.3 min on average. We are improving the efficiency of the algorithm to reduce the total running time as recommended by our medical collaborator, by converting our codes to C++ or using the GPU.

Clinical Value of the Contributions: Our preliminary diagnostic results, outlined in this paper, have shown that the presented ARTR-CAD system can differentiate the non-rejection and acute rejection transplants. Moreover, the ARTR-CAD system provides fast diagnostic results (in approximately 20 minutes with this MATLAB version) compared to a week or more turnaround for the final diagnostic biopsy results. These abilities of the ARTR-CAD system will help clinicians to initi-

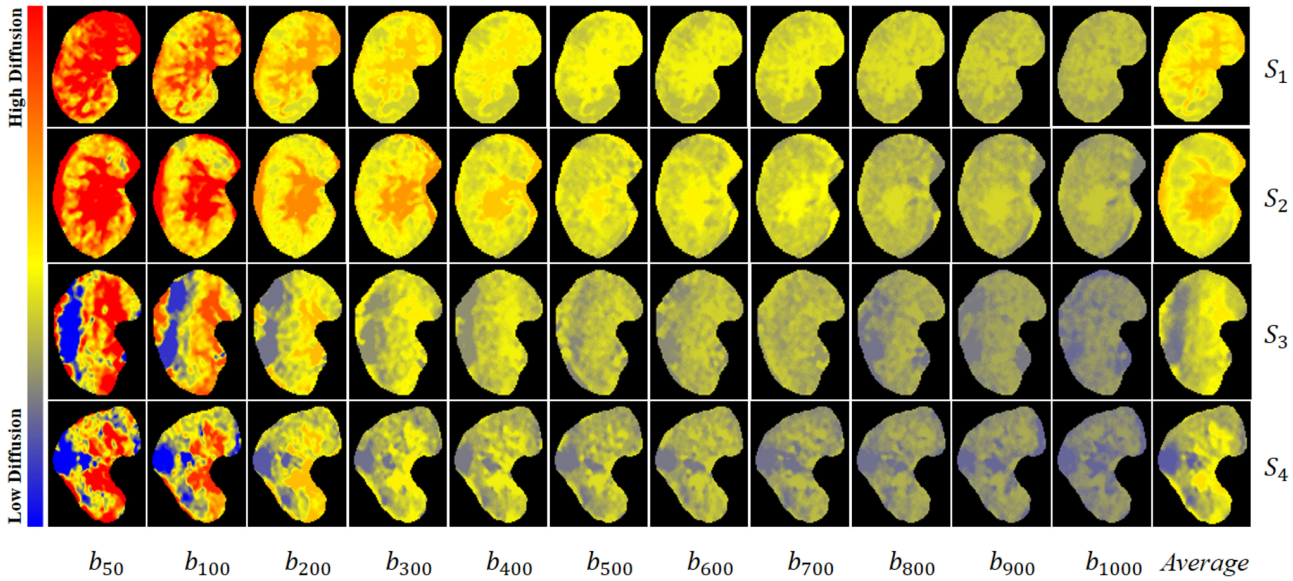

Fig. 12. Pixel-wise parametric maps at different b-values and their average map for two non-rejection (S_1, S_2) and two rejection (S_3, S_4) subjects.

ate timely interventions with appropriate treatments. Thus, the ARTR-CAD system can improve the delivery of healthcare in the USA and worldwide as a new and relatively inexpensive diagnostic tool for early detection of ARTR.

Although the ARTR-CAD system has shown the ability to differentiate rejected from non-rejected renal transplants with a high accuracy of 97.0%, this study is limited by the small and limited diversity of the sample size (total of 100 subjects from one race). This can be justified in part by the difficulties faced in acquiring data, especially when it is related to patient data. This brings us to another limitation related to the training and testing methods using the "leave-one-subject-out" scenario. Thus, we followed the suggestions made by the biostatistics people [56] on how to train and test the developed model (ARTR-CAD) system in the case of a limited number of subjects (small database), which is mainly based on different cross-validation scenarios. Therefore, we are currently running another study at different U.S. sites (e.g., the University of Louisville and the University of Michigan) to be able to test and confirm the accuracy of the presented technique on a larger and more diverse cohort of patients with different races and/or cultural backgrounds to reflect a wider population of kidney transplant patients. One additional limitation, in most situations, there is a lack in determining the best number of hidden layers and/or units without training "empirically" several networks and estimating both the generalization error of each network and the final accuracy based on several cross-validation techniques. If you have too few hidden units, you will get high training error and high generalization error due to underfitting and high statistical bias. If you have too many hidden units, you may get low training error but still have high generalization error due to overfitting and high variance.

VI. Conclusions and Future Work

To conclude, the ARTR-CAD system for early detection of renal transplant rejection from the (3D + b-value) DW-MRI data combines existing and new techniques for non-rigid image alignment, kidney segmentation with a deformable boundary, estimation of spatial diffusion parameters (ADCs), and classification of the transplanted kidney status using CDFs of the ADCs as integral status descriptions and a trainable SNCAE classifier. Despite this paper being focused on a global kidney status diagnosis, i.e., rejection vs. non-rejection, capabilities of our CAD system can be enhanced in various ways. In particular, clinical reports and clinical biomarkers, such as creatinine clearance (CrCl) and serum plasma creatinine (SPCr) could be integrated with image markers towards detecting different types and causes of renal rejection, e.g., antibody-mediated rejection and T-cell or cellular rejection, which is very critical for determining the proper treatment. Moreover, we are exploring the possibility of combining these clinical biomarkers with the image biomarkers on a larger data set to find out whether the diagnostic results can be improved in such a way. In the future, we plan to investigate diagnostic abilities of other imaging modalities, such as diffusion tensor imaging (DTI) and blood oxygen level dependent (BOLD) MRI.

Acknowledgment

This study is made possible by the generous support of the American people through the United States Agency for International Development (USAID). The contents are the responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government.

References

[1] F. Khalifa *et al.*, "Models and methods for analyzing DCE-MRI: A review," *Med. Phys.*, vol. 41, no. 12, pp. 1–32, 2014.
[2] M. Abou-El-Ghar *et al.*, "Role of diffusion-weighted MRI in diagnosis of acute renal allograft dysfunction: A prospective preliminary study," *Brit. J. Radiol.*, vol. 85, no. 1014, pp. e206–e211, 2014.
[3] G. Liu *et al.*, "Detection of renal allograft rejection using blood oxygen level-dependent and diffusion weighted magnetic resonance imaging: A retrospective study," *BMC Nephrol.*, vol. 15, no. 1, p. 158, 2014.

[4] G. Myers *et al.*, "Recommendations for improving serum creatinine measurement: A report from the laboratory working group of the national kidney disease education program," *Clin. Chem.*, vol. 52, no. 1, pp. 5–18, 2006.

[5] R. Katzberg *et al.*, "Functional, dynamic, and anatomic MR urography: Feasibility and preliminary findings," *Acad. Radiol.*, vol. 8, no. 11, pp. 1083–1099, 2001.

[6] V. Lee *et al.*, "Dynamic three-dimensional MR renography for the measurement of single kidney function: Initial experience 1," *Radiology*, vol. 227, no. 1, pp. 289–294, 2003.

[7] E. Hodneland *et al.*, "In vivo estimation of glomerular filtration in the kidney using DCE-MRI," in *Proc. Int. Symp. Image Signal Process. Anal.*, 2011, pp. 755–761.

[8] F. G. Zöllner *et al.*, "Assessment of 3D DCE-MRI of the kidneys using non-rigid image registration and segmentation of voxel time courses," *Comput. Med. Imag. Graph.*, vol. 33, no. 3, pp. 171–181, 2009.

[9] E. Hollis *et al.*, "Towards non-invasive diagnostic techniques for early detection of acute renal transplant rejection: A review," *Egypt. J. Radiol. Nuclear Med.*, vol. 48, pp. 257–269, 2016.

[10] F. Khalifa *et al.*, "Dynamic contrast-enhanced MRI-based early detection of acute renal transplant rejection," *IEEE Trans. Med. Imag.*, vol. 32, no. 10, pp. 1910–1927, Oct. 2013.

[11] J. Zhang *et al.*, "New magnetic resonance imaging methods in nephrology," *Kidney Int.*, vol. 85, no. 4, pp. 768–778, 2014.

[12] U. Eisenberger *et al.*, "Evaluation of renal allograft function early after transplantation with diffusion-weighted MR imaging," *Eur. Radiol.*, vol. 20, no. 6, pp. 1374–1383, 2010.

[13] A. Kaul *et al.*, "Assessment of allograft function using diffusion-weighted magnetic resonance imaging in kidney transplant patients," *Saudi J. Kidney Dis. Transplantation*, vol. 25, no. 6, 2014, Art. no. 1143.

[14] S. Palmucci *et al.*, "Magnetic resonance with diffusion-weighted imaging in the evaluation of transplanted kidneys: Preliminary findings," *Transplantation Proc.*, vol. 43, pp. 960–966, 2011.

[15] S. Palmucci *et al.*, "Magnetic resonance with diffusion-weighted imaging in the evaluation of transplanted kidneys: Updating results in 35 patients," *Transplantation Proc.*, vol. 44, pp. 1884–1888, 2012.

[16] J. Xu *et al.*, "Value of diffusion-weighted MR imaging in diagnosis of acute rejection after renal transplantation," *J. Zhejiang Univ. Med. Sci.*, vol. 39, no. 2, pp. 163–167, 2010.

[17] M. E. Leventon *et al.*, "Statistical shape influence in geodesic active contours," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2000, vol. 1, pp. 316–323.

[18] A. Litvin and W. C. Karl, "Coupled shape distribution-based segmentation of multiple objects," *Int. Conf. Inf. Process. Med. Imag.*, vol. 19, pp. 345–356, 2005.

[19] H. E. A. El Munim and A. A. Farag, "Curve/surface representation and evolution using vector level sets with application to the shape-based segmentation problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 945–958, Jun. 2007.

[20] H. Abdelmunim *et al.*, "A kidney segmentation approach from DCE-MRI using level sets," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–6.

[21] P. Yan *et al.*, "Modeling interaction for segmentation of neighboring structures," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 2, pp. 252–262, Mar. 2009.

[22] O. Gloger *et al.*, "Prior shape level set segmentation on multistep generated probability maps of MR datasets for fully automatic kidney parenchyma volumetry," *IEEE Trans. Med. Imag.*, vol. 31, no. 2, pp. 312–325, Feb. 2012.

[23] J. Wang *et al.*, "Shape–intensity prior level set combining probabilistic atlas and probability map constrains for automatic liver segmentation from abdominal CT images," *Int. J. Comput. Assisted Radiol. Surg.*, vol. 11, no. 5, pp. 817–826, 2016.

[24] A. Skalski *et al.*, "Kidney segmentation in ct data using hybrid level-set method with ellipsoidal shape constraints," *Metrol. Meas. Syst.*, vol. 24, no. 1, pp. 101–112, 2017.

[25] V. Vishnevskiy *et al.*, "Simultaneous denoising and registration for accurate cardiac diffusion tensor reconstruction from MRI," *Med. Image Comput. Comput.-AssistedIntervention*, vol. 9349, pp. 215–222, 2015.

[26] H. Jensen *et al.*, "Locally orderless registration for diffusion weighted images," *Med. Image Comput. Comput.-AssistedIntervention*, vol. 9350, pp. 305–312, 2015.

[27] P.-T. Yap *et al.*, "Brain tissue segmentation based on diffusion MRI using l0 sparse-group representation classification," *Med. Image Comput. Comput.-Assist ed Intervention*, pp. 132–139, 2015.

[28] S. Osher and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*. New York, NY, USA: Springer, 2006.

[29] N. Tustison *et al.*, "N4ITK: Improved N3 bias correction," *IEEE Trans. Med. Imag.*, vol. 29, no. 6, pp. 1310–1320, Jun. 2010.

[30] A. El-Baz *et al.*, "Precise segmentation of 3-D magnetic resonance angiography," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 7, pp. 2019–2029, Jul. 2012.

[31] A. Farag *et al.*, "Precise segmentation of multi-modal images," *IEEE Trans. Image Process.*, vol. 15, no. 4, pp. 952–968, Apr. 2006.

[32] M. Shehata *et al.*, "A novel framework for automatic segmentation of kidney from DW-MRI," in *Proc. IEEE Int. Symp. Biomed. Imag., From Nano Macro*, 2015, pp. 951–954.

[33] M. Shehata *et al.*, "A level set-based framework for 3D kidney segmentation from diffusion MR images," in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 4441–4445.

[34] B. Glocker *et al.*, "Non-rigid registration using discrete MRFs: Application to thoracic CT images," in *Proc. MICCAI Workshop Eval. Methods Pulmonary Image Registration*, pp. 147–154, 2010.

[35] D. Le Bihan and E. Breton, "Imagerie de diffusion in-vivo par résonance magnétique nucléaire," *Comptes-Rendus de l'Académie des Sci.*, vol. 93, no. 5, pp. 27–34, 1985.

[36] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 2013.

[37] L. Breiman, *Classification and Regression Trees*. Boca Raton, FL, USA: CRC Press, 1984.

[38] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[39] P. Cunningham and S. J. Delany, "k-nearest neighbour classifiers," *Multiple Classifier Syst.*, vol. 34, pp. 1–17, 2007.

[40] G. B. Goh *et al.*, "Deep learning for computational chemistry," *J. Comput. Chem.*, vol. 38, no. 16, pp. 1291–1307, 2017.

[41] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.

[42] D. Shen *et al.*, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, 2017.

[43] G. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[44] H. A. Song and S.-Y. Lee, "Hierarchical representation using NMF," in *Proc. Neural Inf. Proc.*, 2013, pp. 466–473.

[45] D. Erhan *et al.*, "Why does unsupervised pre-training help deep learning?," *J. Mach. Learn. Res.*, vol. 11, pp. 625–660, 2010.

[46] Y. Bengio *et al.*, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[47] Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.

[48] Y. Bengio *et al.*, "Greedy layer-wise training of deep networks," in *Proc. 19th Int. Conf. Adv. Neural. Inf. Process. Syst.*, 2007, vol. 19, pp. 153–160.

[49] E. Hosseini-Asl *et al.*, "Deep learning of part-based representation of data using sparse autoencoders with nonnegativity constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 12, pp. 2486–98, Dec. 2016.

[50] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[51] K. Kuzmin *et al.*, "Synergy landscapes: A multilayer network for collaboration in biological research," in *Proc. Int. Conf. School Netw. Sci.*, 2016, pp. 205–212 2016.

[52] G. Gerig *et al.*, "Valmet: A new validation tool for assessing and improving 3D object segmentation," in *Proc. Med. Image Comput. Comput.-Assisted Intervention*, pp. 516–523, 2001.

[53] T. F. Chan and L. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.

[54] M. Hall *et al.*, "The WEKA data mining software: An update," *ACM SIGKDD Explorations Newslett.*, vol. 11, no. 1, pp. 10–18, 2009.

[55] M. Ojala and G. C. Garriga, "Permutation tests for studying classifier performance," *J. Mach. Learn. Res.*, vol. 11, no. Jun, pp. 1833–1863, 2010.

[56] G. James *et al.*, *An Introduction to Statistical Learning*, vol. 112, New York, NY, USA: Springer, 2013.