Computerized Medical Imaging and Graphics 90 (2021) 101911

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

A multiparametric MRI-based CAD system for accurate diagnosis of bladder cancer staging

K. Hammouda [a], F. Khalifa [a], A. Soliman [a], M. Ghazal [b], M. Abou El-Ghar [c], M.A. Badawy [c], H. E. Darwish [d], A. Khelifi [e], A. El-Baz [a,*]

[a] Bioengineering Department, University of Louisville, Louisville, KY, USA
[b] Electrical and Computer Engineering Department, Abu Dhabi University, UAE
[c] Radiology Department, Urology and Nephrology Center, Mansoura University, Egypt
[d] Mathematics Department, Faculty of Science, Mansoura University, Mansoura, Egypt
[e] Computer Science and Information Technology Department, Abu Dhabi University, UAE

ARTICLE INFO

Keywords:
Classification bladder cancer staging
CAD system
Functional features
Texture features
Morphological features

ABSTRACT

Appropriate treatment of bladder cancer (BC) is widely based on accurate and early BC staging. In this paper, a multiparametric computer-aided diagnostic (MP-CAD) system is developed to differentiate between BC staging, especially T1 and T2 stages, using T2-weighted (T2W) magnetic resonance imaging (MRI) and diffusion-weighted (DW) MRI. Our framework starts with the segmentation of the bladder wall (BW) and localization of the whole BC volume (V_t) and its extent inside the wall (V_w). Our segmentation framework is based on a fully connected convolution neural network (CNN) and utilized an adaptive shape model followed by estimating a set of functional, texture, and morphological features. The functional features are derived from the cumulative distribution function (CDF) of the apparent diffusion coefficient. Texture features are radiomic features estimated from T2W-MRI, and morphological features are used to describe the tumors' geometric. Due to the significant texture difference between the wall and bladder lumen cells, V_t is parcelled into a set of nested equidistance surfaces (i.e., iso-surfaces). Finally, features are estimated for individual iso-surfaces, which are then augmented and used to train and test machine learning (ML) classifier based on neural networks. The system has been evaluated using 42 data sets, and a leave-one-subject-out approach is employed. The overall accuracy, sensitivity, specificity, and area under the receiver operating characteristics (ROC) curve (AUC) are 95.24%, 95.24%, 95.24%, and 0.9864, respectively. The advantage of fusion multiparametric iso-features is highlighted by comparing the diagnostic accuracy of individual MRI modality, which is confirmed by the ROC analysis. Moreover, the accuracy of our pipeline is compared against other statistical ML classifiers (i.e., random forest (RF) and support vector machine (SVM)). Our CAD system is also compared with other techniques (e.g., end-to-end convolution neural networks (i.e., ResNet50).

1. Introduction

The most recent released statistics of the American Cancer Society showed that urinary bladder cancer (BC) is the fourth most common cancer among men in the US [1]. Early diagnosis of BC helps clinicians select the best treatment method, which depends on four major essential factors. These factors are the BC staging, T, (T1, T2, T3, and T4); BC grading, G, (G1, G2, and G3), lymph node metastasis, N, (N0, N1, N2, and N3), and distant metastasis, M, (M0 and M1) [2]. Depending on the BC staging, the BC treatment can be widely classified into two main types: non-muscle-invasive bladder cancer (NMIBC, stage $\leq$ T1) and muscle-invasive bladder cancer (MIBC, stage $\geq$ T2). By the time, from 20% to 30% of NMIBC advance to MIBC [3]. So early diagnosis is essential to stop BC from advancing. The MIBC rate of aggressiveness and death rate is very high compared to NMIBC [3].

The cystoscopic examination and histological evaluation of bladder sampled tissue, using transurethral resection (TURB), are considered the gold standard for BC detection. Cystoscopy, however, has its own limitations, especially the difficulty of discriminating between malignant lesions and healthy urothelium. Patients with NMIBC are ultimately

* Corresponding author.
E-mail address: aselba01@louisville.edu (A. El-Baz).

https://doi.org/10.1016/j.compmedimag.2021.101911
Received 19 October 2020; Received in revised form 20 March 2021; Accepted 26 March 2021
Available online 31 March 2021

treated with TURB followed by intravesical chemotherapy [4,5]. The standard treatment of MIBC patients is radical cystectomy (RC) combined with cisplatin-based neoadjuvant chemotherapy (NAC) [3]. However, RC-NAC has poor prognosis, and metastases develop within two years after RC in about 50% of patients [4]. Therefore, treatment decision, prognosis, and follow-up management of patients with BC depend on the accurate differentiation between NMIBC and MIBC [6,3], in which this classification relies mainly on BC staging. Cystoscopic examination with pathological evaluation of the resected tissue is the standard reference for differentiation between BC staging between MIBC and NMIBC [5,7]. From 20% to 30% of BC are incorrectly staged because of variation in performing resection [8,6]. Using multiparametric modalities and multiple examinations can reduce the diagnostic error by fusing information from different modalities and thus enhance BC diagnosis [9,10]. However, these methods are time-consuming, invasive, and costly procedures. It is crucial to develop a noninvasive low-cost approach with high accuracy for BC detection to distinguish between MIBC and NMIBC.

In recent years, artificial intelligence (AI) with deep learning (DL) based on medical imaging modalities, such as magnetic resonance imaging (MRI) and computed tomography (CT), have been exploited as essential diagnostic tools for BC detection, tumor staging, and prediction of the treatment response of tumor recurrence [4,11–13]. Mainly, MRIs have been found to play a crucial role in making early localization of BC and diagnosis of invasiveness. Radiomics is a method that extracts different sets of features from routine medical imaging, using data characterization algorithms [14,15]. These radiomic features enable data to be employed in a decision support system to advance clinical diagnostic, prediction performance, and therapeutic response assessment [16,17]. Various research work on BC diagnosis using MRI modality and radiomic analysis have been conducted using texture features. The latter was extracted from the greyscale pixel values, and high-order derivative maps [18], as well as functional diffusion-weighted MRI (DW-MRI) using apparent diffusion coefficient (ADC) maps [19,20]. In summary, recent studies demonstrated potential to improve patient care with adequate management, which can reduce the cost of unnecessary investigations and treatment and improve the outcome of therapy [11–13].

The critical step for determining the BC staging is the tumor-extension position inside the bladder wall (BW). The tumor's location can be localized inside the BW, intravesical, or extending across perivesical to extravesical (see Fig. 1). However, the tumor's size does not affect the BC staging [2]. The big challenge for differentiation between the NMIBC and MIBC is discrimination between T1 and T2 BC stages. As shown in Fig. 1, both of T3 and T4 BC stages have perivesical or extravesical mass from the BW, so it is not difficult to discriminate them from the T1 BC stage. On the other hand, T1 and T2 stages are visually very close, and the significant difference between them is that the T2 BC stage invades the BW muscle, while the T1 BC stage does not [2,3,5]. Because of the reasons mentioned above, it is essential to build a computer-aided diagnosis (CAD) system to differentiate between T1 and T2 BC stages using MP-MRI and radiomic analysis. To the best of our knowledge, this is the first study to develop a CAD system to discriminate between T1 and T2 stages instead of NMIBC and MIBC.

There is very little research that has developed a CAD system to differentiate between MIBC and NMIBC [21,19] using MRI modality, to the best of our knowledge. Those previous research have many limitations. A pathological tumor is segmented manually. The diagnostic accuracy is related to the MIBC and NMIBC classification; the framework did not work in each stage individually. Besides, T3 and T4 BC stages (i.e., MIBC) have perivesical or extravesical mass from the BW, so it is easy to discriminate them from the T1 BC stage (NMIBC). The big challenge is to differentiate between BC stages, especially T1 and T2 BC stages. Additionally, they used in their analysis the whole tumor without giving the importance for any parts. The extracted features do not reflect the physical meaning of the problem using the entire tumor's radiomic characteristics for each modality.

This manuscript's major objective is to propose an automated CAD system for accurate BC staging classification, especially T1 and T2 BC stages (see Fig. 2). The rest of this paper is organized as follows: Section 3 provides details about the study data. The proposed methodology and employed features are fully explained in Section 3, experimental design and results are outlined in Section 4. Finally, discussion and conclusions are given in Sections 5 and 6, respectively.

2. Bladder MRI data

Forty-two patients who underwent a diagnosis of patients with BC were enrolled in this study after providing consent. The Urology and Nephrology Center, Mansoura University, Egypt, supplied us with the data sets in which the T2-weighted and diffusion MRI were acquired as part of the routine preoperative diagnosis of patients with BC. The patients were enrolled under a protocol approved by the Center's institutional review board (IRB). The T2-weighted and DW-MRI for patients with BC ($n = 42$, $M = 28$, $F = 14$, mean age $= 50 \pm 12.34$ years, age range $= 30$–70 years) were acquired between May 2012 to October 2018, and the diagnosis was confirmed using pathology.

The data were acquired using 3T Ingenia Philips MRI scanners using the following acquisition parameters: FOV $= 240 \times 240 \times 145\,mm^3$, Slice thickness $= 3$ mm, Slice Gap 0.3 mm, TR $= 3000$–5000 ms with shortest TE and Torso Coil, in-plane spatial resolution ranges from $0.5208 \times 0.5208\,mm^2$ to $0.9375 \times 0.9375\,mm^2$. The number of axial slices per subject ranges from 30 to 45. For both DW-MRI acquisition protocols, water signals were acquired at two different *b*-values, b0 and b800, at 100 increments.

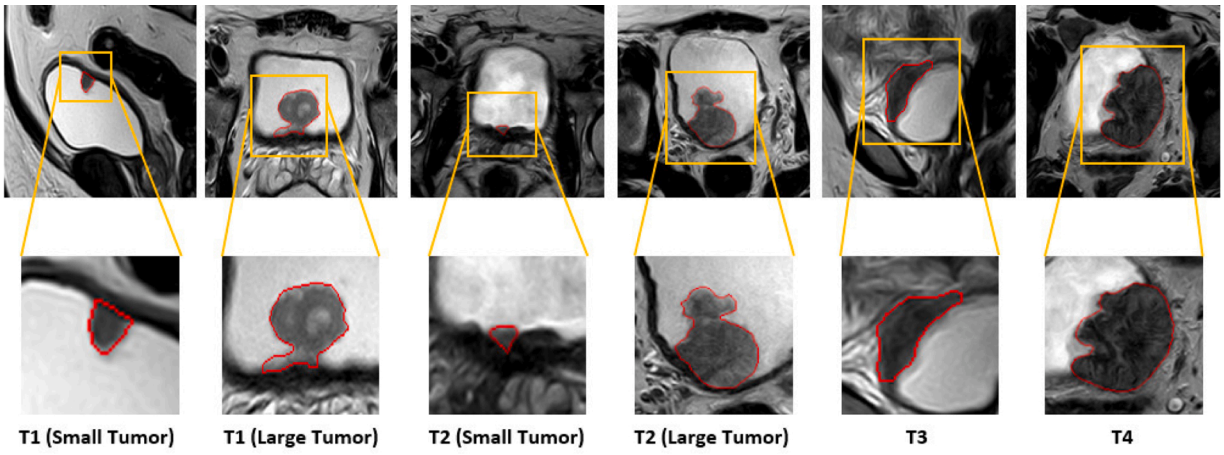

Fig. 1. Examples of different bladder cancer (BC) staging. Enlarged portions show the extent of the tumor into the bladder wall.

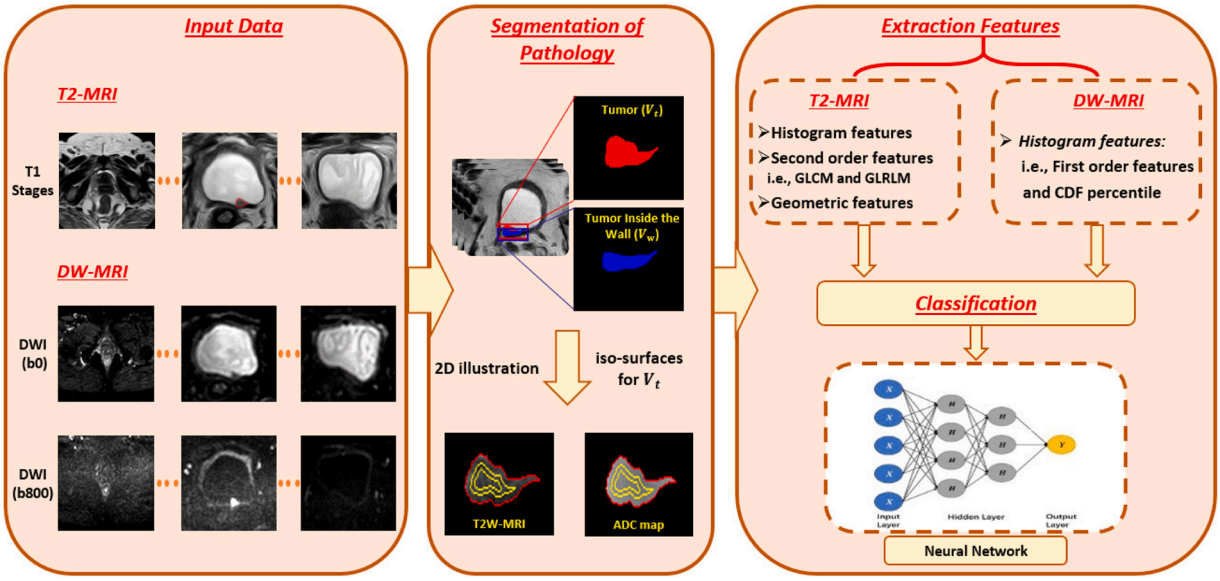

Fig. 2. The proposed multiparametric CAD system for early diagnosis of bladder cancer staging.

3. Methods

In this study, we developed a CAD system (Fig. 2) to discriminate between bladder cancer staging. The overall system applies the following steps to obtain the final diagnosis: Firstly, localization of two VOIs (V_w, V_t) by segmenting the BW with pathology and the BW. After that, extraction of image markers, namely, radiomic and morphological features from the T2-MRI and descriptors of the voxel-wise ADC maps. Finally, diagnosing BC staging utilizes image features to train and test both statistical and cognitive classifiers. Details of the developed CAD system are discussed below.

3.1. Bladder tumor segmentation

Developing an accurate BW segmentation is not a straightforward task and is faced with multiple challenges [22,23] due to the presence of pathology. Generally, pathology affects BW's visual appearance and might appear in more than one region. Also, pathology localization can be intravesical or extending across perivesical to extravesical [2]. Other challenges include motions as geometric distortion (see Fig. 3), image artifacts, low contrast between bladder urinary and surrounding tissues, and anatomical differences between inter-patient (and also between various subjects). To partially overcome the challenges mentioned

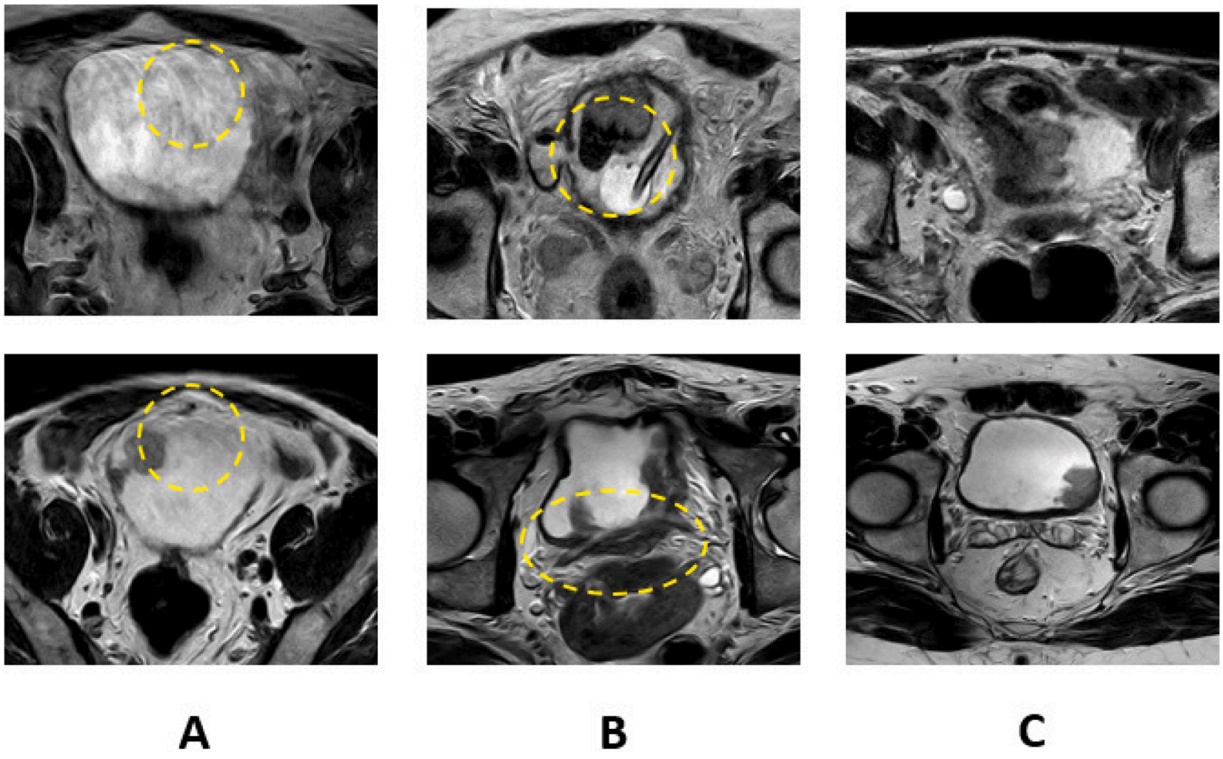

Fig. 3. Challenges for segmenting the bladder wall and pathology using T2-MRI that show with yellow contours (A) geometric distortion/diffused and similar intensities; (B) image artifacts; and (C) anatomical differences between inter-patient and various subjects. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

above for pathological bladder, a deep learning approach was developed using 3D CNN.

Fig. 4 shows the overall segmentation system architecture that contains three stages: preprocessing, 3D CNN, and post-processing. The preprocessing stage normalize all T2-MRI data sets to the same spatial domain using a resampling process-besides extracting the ROI contains all voxels of the bladder with wall and pathology. The preprocessing step makes the data sets suitable for the proposed system and provides faster performance. Secondly, two deep learning architectures (CNN1 and CNN2) based on the DeepMedic network [24] are developed to segment the BW and the pathology. The CNN1 is used to segment the BW with pathology, see Fig 1 in the supplementary material, by feeding the network with the normalized data. Since separate pathology segmentation is almost impossible due to irregular shape and similar visual appearance with the wall, another network (CNN2) is developed, see Fig 2 in the supplementary material. It contains an additional pathway to extract BW only. The second pathway is supplied with 3D data from the adaptive shape prior (ASP) model's output. Finally, a post-processing stage using the conditional random field (CRF) is employed to produce the final output [25] by accounting for small isolated hole due to local minima in training and noise in the input images. More details about CNN1 and CNN2 architecture, layers, etc., can be found in [26].

Using the outputs of CNN1 and CNN2, the bladder's pathology is split into two VOIs (see Fig. 4). The first region, V_t, is the whole part of the pathology and is obtained by subtracting the outputs of CNN1 and CNN2. The second region, V_w, represents the pathology inside the wall. The intersection between the V_t and the output from the CNN2 represents the V_w.

3.2. Features extraction

The primary features utilized in our system are three types: functional and texture, morphological features. The functional features are based on the ADC maps' statistical measures, representing the percentiles of the cumulative distribution function (CDF). Texture features are radiomic features estimated from T2W-MRI. The morphological is used to describe the tumors' geometric.

T1 and T2 BC stages have almost a similar tissue visual appearance. However, the tissue difference between these two stages appears in some parts of the tumor. Thus, some tumor parts have meaningful texture information than others [2,3,5]. For example, the pathology inside the wall is considered an essential element because of the significant difference between cells for wall and bladder lumen [2]. For the reasons above, unlike traditional methods, the tumor (V_t) is divided into a set of nested equidistance surfaces (i.e., iso-surfaces, see Fig. 5 for 2D illustration). All features mentioned above are estimated per iso-surface. Iso-surfaces depend on the distance map generated inside the tumor by finding the minimum Euclidean distance for every inner point to the tumor boundary.

The total number of features derived from T2-MRI are total 24 histogram features, 25 GLCM features, and 16 GLRLM features for each iso-surface. Those features are derived from both segmented MRI data and its gradient image. Besides, we have three morphological features. Thus, the total number of T2-MRI's features are 133 $[(27 + 25 + 16) \times 2 + 3]$. The features extracted from ADC maps for each iso-surface are 24 features. Therefore, the total number of discriminatory features per iso-surface in our pipeline is 157 $(133 + 24)$.

3.2.1. Texture features

Many recent kinds of research for BC have shown the value of the T2W-MRI information about the depth and volume of the BC, the surrounding organ invasion, and the extravesical disease spread [2,27]. The BC staging depends on the pathology's location inside the BW, intravesical, or extending across perivesical to extravesical (see Fig. 1). Thus, T2W-MRI is beneficial to differentiate between the difficult T1 and T2 stages (see Fig. 6).

According to previous research [18,28], the reconstruction of the intensity-based images from noisy signals leads to specific texture loss because of noise smoothness. One way to restore tissue textures beneath the intensity maps' images is by performing high-order derivative operations [18,28]. The first-order derivative map, namely the gradient, is one of the best ways to reflect the heterogeneous patterns of VOIs

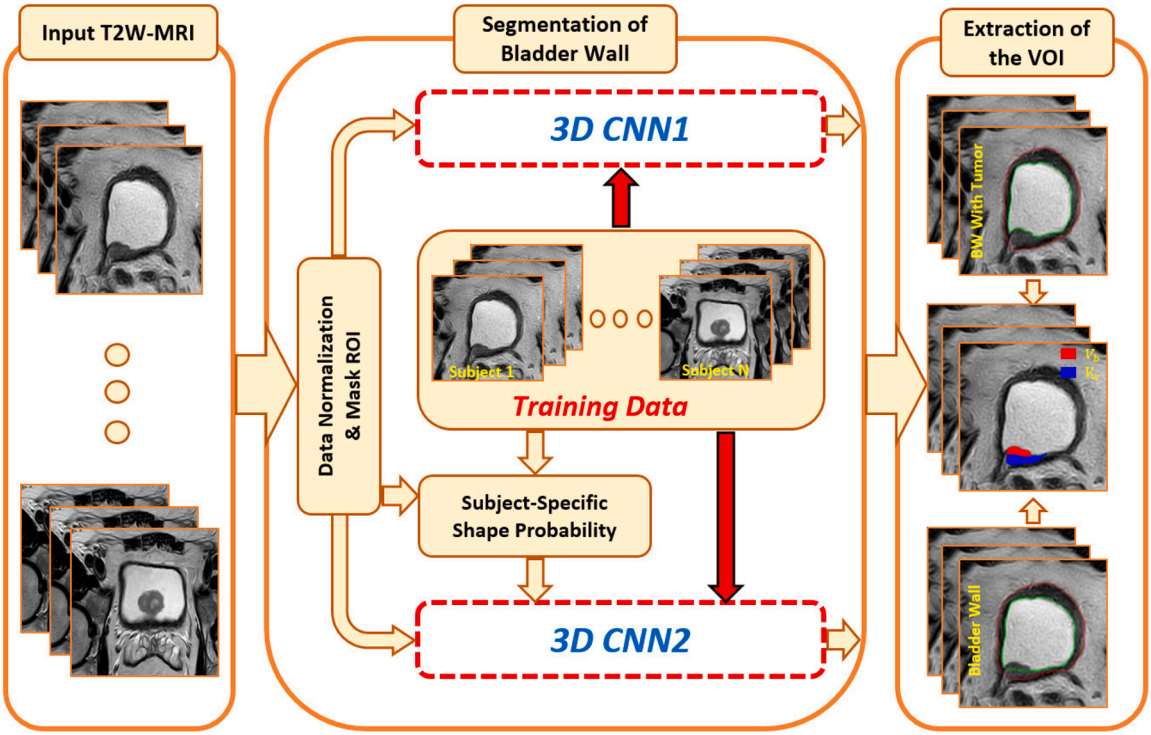

Fig. 4. Pathological bladder segmentation scheme.

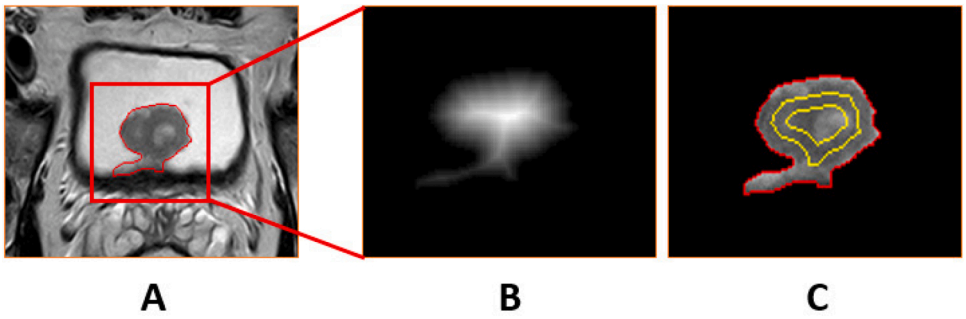

Fig. 5. 2D illustration for the generation of iso-surfaces: (A) an image for bladder from T2-MRI, (B) the distance maps for bladder cancer, and its iso-surfaces (C).

because it amplifies variations of the intensity distribution [28–30]. In this paper, the 3D Sobel kernel is used to calculate the 3D gradient map. Fig. 7 shows the cancerous heterogeneity of an MR slice T2W intensity image and the gradient of the carcinoma.

After constructing the iso-surfaces on the V_t for T2-MRI, the **radiomic features** are extracted from the segment T2W-MRI image and its gradient [31–33]. Those features consist of three categories: first-order statistics (histogram descriptors), second-order statistics (GLCM and GLRLM descriptors), and morphological features. To estimate the histogram features, the probability distribution function (PDF) is first created for the voxel values in V_t. After that, the CDF is then constructed, and only critical statistical measures (percentiles) are identified from the CDF (see Fig. 8). Table 1 shows the list of those features.

The 2nd-order statistical features are the second set of texture features that are estimated using two GLCM and GLRLM. The 2nd-order statistical features [34] play an essential role in medical image classification based on texture features [34–36]. **GLCM** texture analysis is utilized in our system that is introduced by Haralick et al. [37]. GLCM depends on the estimation of the second-order joint conditional probability density functions, $P(I,J,D,\Theta)$. The probability for each $P(I,J,D,\Theta)$ of moving from grey-level I to J, utilizing the distance D at an angle Θ. For the image with n grey levels, the size of GLCM is $n \times n$. Each matrix is estimated by counting the number of times each pair of grey levels (I,J) occurs at a distance D and in the orientation Θ.

Fig. 9 shows the classical 2D algorithm that utilizes the four orientations: horizontal, vertical, and two diagonal [37,31]. A 3D algorithm

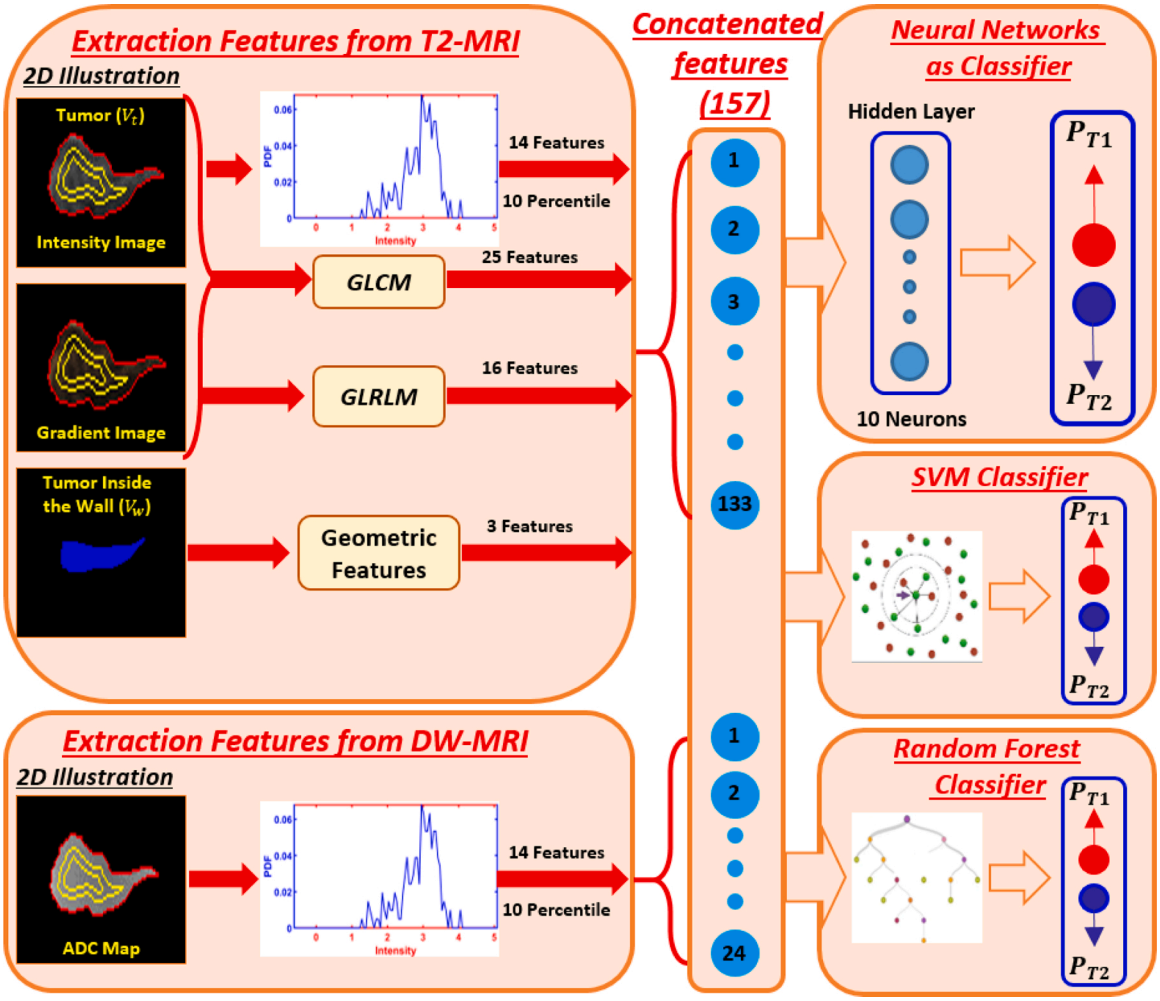

Fig. 6. The architecture of feature extraction and classification of the T1 and T2 BC stages.

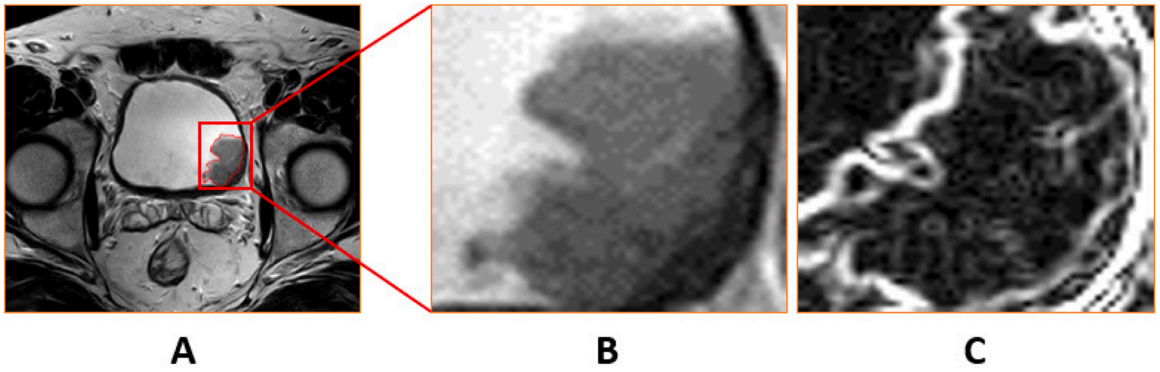

Fig. 7. Illustration of cancerous heterogeneity: (A) T2W intensity image of a bladder carcinoma show in red contour, zoomed intensity (B), gradient (C) of the carcinoma. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

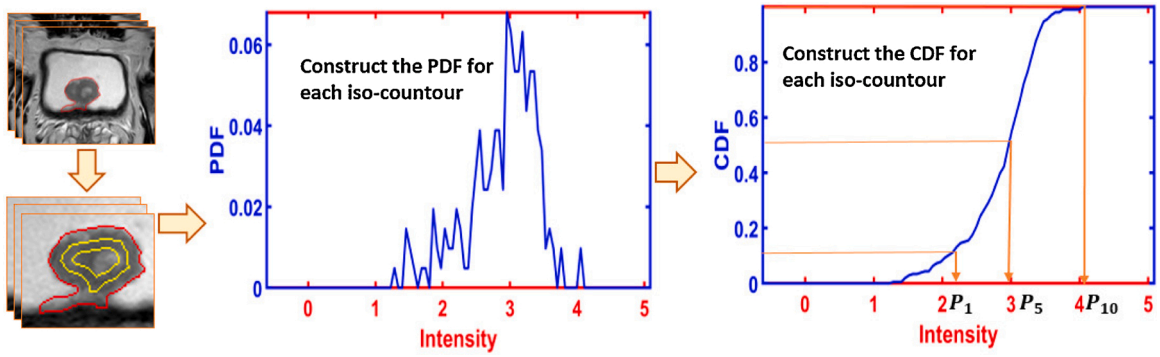

Fig. 8. Demonstration of the extraction of ten percentiles from the cumulative distribution function (CDF) for one iso-surface.

(see Fig. 9) is performed using thirteen directions [37,32,38,39] to get more descriptive information and details about each pixel and its neighbors. In contrast to the 2D scenario, the 3D image elements have 26 neighbors (see Table 2). The analysis for the images in all 26 possible directions provides more precise texture information. To avoid double-counting the gray level co-occurrences, they are accumulated in only thirteen directions. By symmetry, co-occurrences in the opposite directions are found by transposing the resulting matrices. Finally, the isotropic GLCM in all 26 directions is calculated by summing all 13 matrices, and their transposes [32,37–39]. After estimating the GLCM, a

Table 1
The extracted features from V_t and V_w that are histogram features and 2nd-order statistical contains the gray-level co-occurrence matrix (GLCM) and the gray level run length matrix (GLRLM), as well as morphological features.

Histogram features	2nd-order statistical		Morphological features (V_w)
	GLCM's features	GLRLM's features	
Mean	Joint average	Short runs emphasis	Volume
Variance	Joint maximum	Long runs emphasis	Surface area
Skewness	Joint entropy	Low grey level run emphasis	Transmural extent
Kurtosis	Joint variance	High grey level run emphasis	
Entropy	Difference average	Short run low grey level emphasis	
Uniformity	Difference variance	Short run high grey level emphasis	
Median	Difference entropy	Long run low grey level emphasis	
Range	Sum average	Long run high grey level emphasis	
Interquartile range	Sum variance	Grey level non-uniformity	
Mean absolute deviation	Sum entropy	Grey level non-uniformity normalised	
Robust mean absolute deviation	Contrast	Run length non-uniformity	
Median absolute deviation	Angular second moment	Run length non-uniformity normalised	
Quartile coefficient of dispersion	Dissimilarity	Run percentage	
ten percentiles	Inverse difference	Grey level variance	
	Inverse difference moment	Run length variance	
	Inverse difference normalised	Run entropy	
	Inverse difference moment normalised		
	Inverse variance		
	Correlation		
	Autocorrelation		
	Cluster tendency		
	Cluster shade		
	Cluster prominence		
	Information correlation 1		
	Information correlation 2		

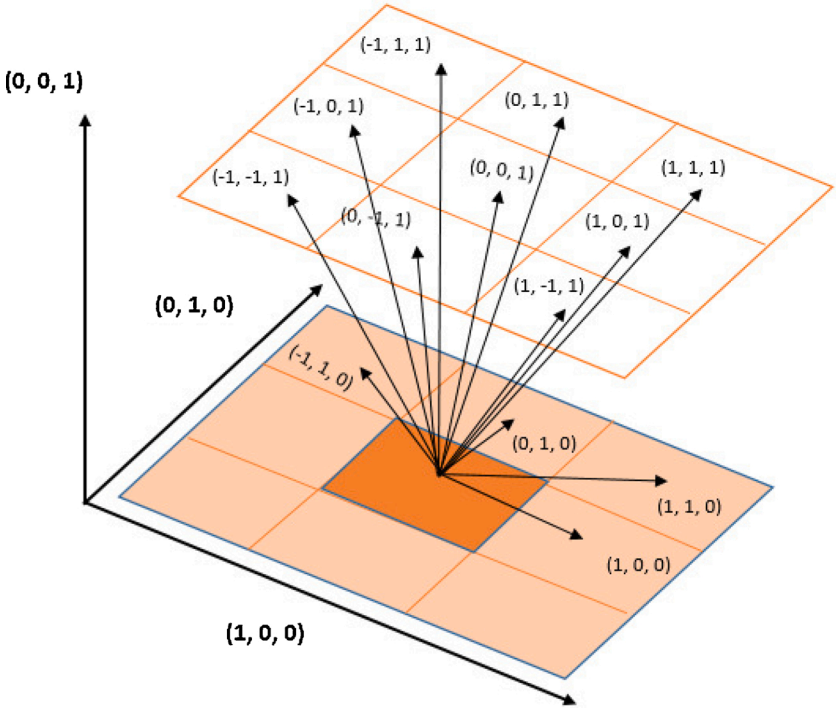

Fig. 9. Adaptation of computational texture algorithms of 3D images that indicates the reference (orange), its neighboring pixels, and all thirteen offset with their orientations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
All thirteen offset with their orientations Θ used for the 3D texture analysis.

XY plane		YZ plane		XZ plane		XYZ space	
Offset	Degree (Θ)	Offset	Degree (Θ)	Offset	Degree (Θ)	Offset	Degree (Θ)
(1,0,0)	0 °	(0,0,1)	0 °	(1,0, 1)	45 °	(1,1,1)	45 °
(1,1,0)	45 °	(0,1,1)	45 °	(− 1,0, 1)	135 °	(− 1,1,1)	135 °
(0,1,0)	90 °	(0, − 1,1)	135 °	–	–	(1, − 1,1)	315 °
(− 1,1,0)	135 °	–	–	–	–	(− 1, − 1,1)	225 °

vector of 25 radiomic features are derived [31,37,40], see Table 1.

On the other hand, **GLRLM** is a method relying on estimating runs of grey levels in the image. A run is a set of consecutive pixels in the image having the same grey level value. Thus, the method for measuring the GLRLM involves counting the number of consecutive pixels with the same grey level in a specific orientation, namely, runs length [38,34]. For 3D images, 13 orientations are shown in Fig. 9 and Table 2 from 3D GLCM. The texture can classify into a coarsest and fine. The coarsest texture appears with many straight pixels with the same grey level, and the fine texture represents a small number of these pixels. Therefore, the texture lengths in all 13 directions can serve as texture description [38, 34]. After calculating the GLRLM, a vector of 16 radiomic features are estimated and used in our framework [31,34,40]), see Table 1.

3.2.2. Functional features

Some recent studies documented and showed that the high-grade tumor and MIBC have lower ADC than the low-grade tumor and NMIBC [7,41]. In our work, ADC is used as a discriminatory feature to diagnose BC staging, especially T1 and T2 stages. In our study, all patients have DW-MRI that is acquired at two scans, the baseline scan ($b = 0$) and the $b = \text{s}\cdot\text{mm}^{-2}$ scan. The DW-MR image marker is the voxel-wise ADCs map that is calculated in accordance with [7,41] as:

$$\text{ADC}(x,y) = \frac{ln\frac{S_0(x,y,z)}{S_1(x,y,z)}}{b_{800} - b_0} \quad (1)$$

where S_0 and S_1 are the signal intensity acquired at the b_0 and b_{800}.

To globally describe the entire ADC map, the CDFs of the estimated ADCs for each subject are constructed. These constructed CDFs have a unified size. Therefore, their use overcomes the challenge related to the variable sizes of different pathology. Like extracting the fourteen histogram features and the CDF percentiles for each iso-surfaces are estimated from T2-MRI (see Table 1). Fig. 8 shows the pipeline for estimating the CDF percentile (ten elements).

After feature extraction, the last step is the classification and evaluation. The features are used to train and test a neural network (NN) as a classifier to develop our goal, the differentiation between T1 and T2 stages. In the experimental results section, we will give more details about the classification.

3.2.3. Morphological features

Another set of features, i.e., morphological features, from T2-MRI are also estimated and utilized in our framework. Morphological features are important for diagnosing the BC staging, as the staging depends on the tumor's transmural, not its size. Among various morphological and shape features, only three features are used to describe the pathology's transmural inside the wall (V_w). For 2D slice, we estimated perimeter and area, while for 3D, surface area and the volume are extracted. Those features are estimated using Minkowski functionals [42]. In addition to those features, pathology transmural extend that measures the radial most extended length of the tumor inside the wall (V_w) is also estimated.

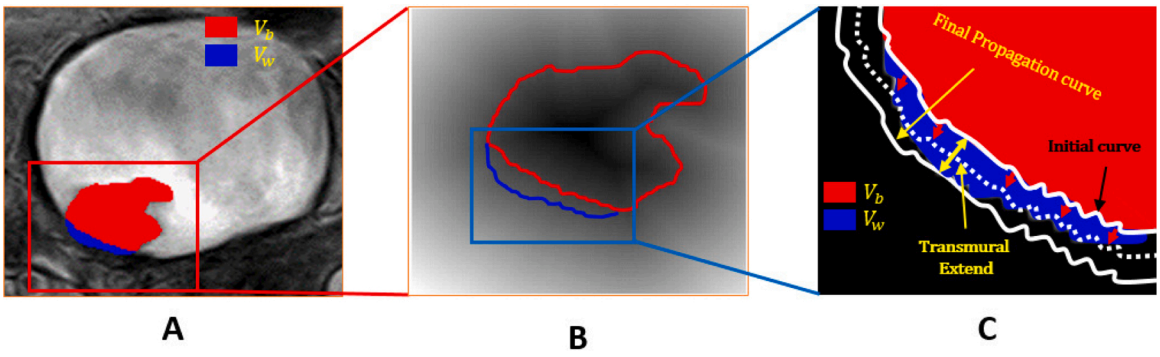

Fig. 10. Generation of transmural extent feature: (A) a T2W intensity image with a tumor inside the bladder (V_b, red) and the wall (V_w, green); (B) Enlarged portion showing the distance maps for the V_b with the edges of both V_b and V_w; illustration of estimating the transmural extent features (vertical distance) (C). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Summary of our CAD system parameter settings. The CNN is for the BC's segmentation, and neural network as a classifier.

CNN		Neural network	
Parameter	Value	Parameter	Value
# Conv. Layers	11	# of layers	3
Kernel size	$3 \times 3 \times 3$	# of neurons	157,10,2
Kernel Initialization	He et. al. scheme [43]	# of epochs	100
# Filters	40	Training function name	trainscg
Stride	Urinary Kernel stride	Performance function	Cross Entropy
Patch size	17^3		
# Epochs	35		
Optimizer	RMSProp		
Learning rate	10^{-4}		

The iso-surfaces idea is also applied to estimate the transmural extend feature. The distance maps for the tumor inside the wall are estimated with a negative value. After that, the zero iso-surface is propagated with a constant step until it reaches the tumor's end outside the lumen (see Fig. 10). The maximum deformation distance represents the tumor's transmural extent.

4. Experimental results

The proposed framework was trained and tested on 42 T2W-MRI and DW-MRI subjects with BC, 21 subjects for T1 and T2 BC stages. The MRI scans' discriminative features were trained and tested using different techniques: neural networks, random forest (RF), and support vector machine (SVM). Besides, to manifest the advantage of our framework, the results for a deep feature (end-to-end fashion) using CNN (i.e., ResNet50) and the approach by Xu et al. [21] were presented. Table 3 summarizes our CAD system parameter settings, which contain those of the CNN for the BC's segmentation and neural network classifier. For the SVM classifier, we used the linear kernel function with an automatic kernel scale. Also, we utilized the RF with the decision tree learner type with 41 maximum number of splits, 30 learners.

Due to the limited number of available data sets (42 patients), a leave-one-subject-out (LOSO) method is exploited to learn, test the statistical characteristics of both classes, and conduct validation experiments [44]. Our system's accuracy is assessed by performing a LOSO cross-validation test, a neural network with one hidden layer (ten neurons). The CAD system's results are emphasized and confirmed by presenting two types of analysis, 2D and 3D. Also, the overall diagnostic accuracy is demonstrated for individual modality and their fusion.

Table 4 shows the system accuracy for both 2D and 3D analysis along with the *p*-values, estimated using the statistical Student's t-test. Several metrics have been used for system evaluation: accuracy (ACC), sensitivity (SEN), and specificity (SPE). The receiver operating characteristics (ROC) has also been employed to confirm and support our system's robustness and accuracy. Namely, the area under the curve (AUC) of the ROC is used as an additional evaluation metric. For the 2D analysis, the metrics obtained are 92.86%, 97.05%, 100%, and 0.9705, respectively. On the other hand, for the 3D analysis, the evaluation metrics are 95.24%, 95.24%, 95.24%, and 0.9864, respectively (see Table 4). The statistical metrics are almost the same for 2D and 3D scenarios, with a little higher 3D accuracy and AUC. Figure 11 shows the ROC curves for the proposed neural network (NN) system, 2D and 3D scenarios, that use the individual MRI models and their fusion.

To highlight the advantages of using both T2W-MRI and ADC maps, our pipeline's accuracy is compared against other statistical ML classifiers (RF and SVM), end-to-end convolution neural networks (i.e., ResNet50 [45].), and Xu et al. [21] approach. For the 2D scenario, the overall accuracy, sensitivity, specificity, and AUC for RF are 90.48%, 85.71%, 95.24%, and 0.8968, respectively, for the SVM are 78.57%, 71.43%, 85.71%, and 0.8027, respectively (see Table 1 in the supplementary material). Similarly, for the 3D method, the overall accuracy, sensitivity, specificity, and AUC for RF are 88.10%, 85.71%, 90.48%, and 0.8968, respectively, for the SVM are 83.33%, 76.19%, 90.48%, and 0.8866, respectively (see Table 2 in the supplementary material). Also, to emphasize the advantage of our proposed framework, 2D and 3D scenarios, Fig. 12 shows the corresponding ROC curves for the NN, SVM, RF classifiers using fusion models. The best value for all AUCs is the proposed 3D framework, a neural network, with a value of 0.9864.

Table 4
2D and 3D classification accuracy (ACC), sensitivity (SEN), specificity (SPE) and area under the curve (AUC) for the proposed neural network (NN) method.

	Neural network (2D scenario)				Neural network (3D scenario)				*P*-value
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	
T2W-MRI	88.10 %	85.71 %	90.48 %	0.8912	90.48 %	80.95 %	100 %	0.9524	0.2263
ADC	80.95 %	76.19 %	85.71 %	0.8889	73.81 %	66.67 %	80.95 %	0.8005	0.1562
Morphological features	78.57 %	76.19 %	80.95 %	0.8685	83.33 %	85.71 %	80.95 %	0.8844	0.2356
All features	**92.86** %	**97.05** %	**100** %	**97.05**	**95.24** %	**95.24** %	**95.24** %	**0.9864**	0.1623

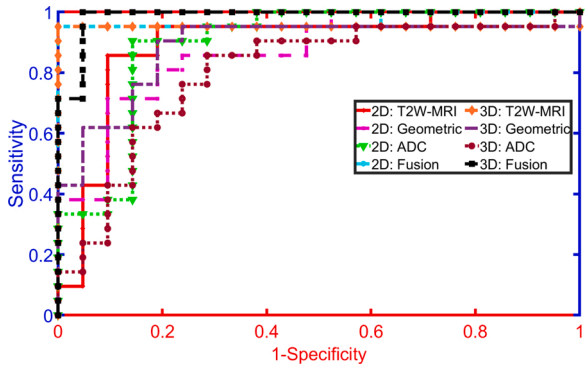

Fig. 11. The receiver operating characteristics (ROC) curves for the proposed 2D and 3D neural network (NN) system using the individual MRI models and their fusion.

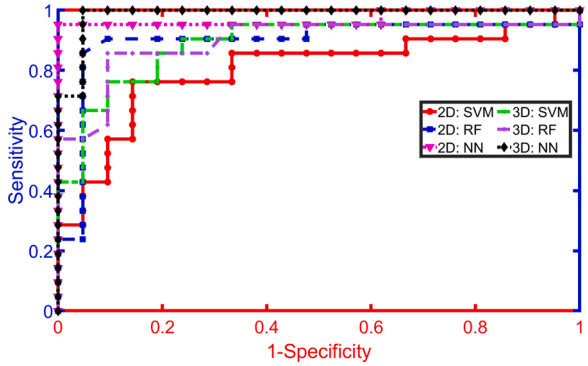

Fig. 12. The receiver operating characteristics (ROC) curves for the neural network (NN), support vector machine (SVM), and random forest (RF) classifiers using fusion models. The ROCs are for both 2D and 3D scenarios.

Table 5
Classification accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the curve (AUC) comparing Xu et al. [21] approach, end-to-end CNN fashion (ResNet50 [45]), and our CAD system (neural network (NN), 3D Scenario) along with the statistical *P*-values.

	ACC	SEN	SPE	AUC	*P*-value
Proposed	95.24 %	95.24 %	95.24 %	0.9864	–
Xu et al. [21]	73.81 %	61.90 %	85.71 %	0.7937	0.0145
ResNet50 [45]	66.67 %	76.19 %	57.14 %	0.7438 %	0.0085

Besides the statistical ML classifiers, we also compared our pipeline against end-to-end convolution neural networks (i.e., ResNet50) and Xu et al. [21] approach, see Table 5. The advantage of the diagnostic capabilities of the proposed framework to differentiate between T1 and T2 BC stages, is first highlighted by comparing it with those obtained with ResNet50 [45]. The ResNet50 is trained and tested using a LOSO method on the T2W-MRI due to the limited number of cases and the overall accuracy, sensitivity, specificity, and AUC are 66.67%, 76.19%, 57.14%, and 0.7438, respectively. Additionally, the results of Xu et al. [21] approach are 73.81%, 61.90%, 85.71%, and 0.7937 for the accuracy, sensitivity, specificity, and AUC, respectively. Also, Fig. 13 shows the ROCs for Xu et al. [21] approach, the ResNet50, and our proposed framework. Besides, the *P*-values in Table 5 document the statistical significant difference between our approach and both Xu et al. [21] and ResNet50.

5. Discussion

Early diagnosis of BC helps physicians to select appropriate treatment interventions and thus increase patients' survival rates. Our goal in this work is the early diagnosis for BC stages, especially T1 and T2 stages. BC staging has many stages, but the most formidable challenge is to differentiate between T1 and T2 BC stages because its visual appearance is almost the same. Also, there is a slight difference in the pathological depth. To the best of our knowledge, this is the first CAD system focused on the closest two BC stages T1 and T2, instead of differentiating between NMIBC and MIBC. We faced many challenges for pathological segmentation, such as geometric distortion image artifacts. The most major limitation in our work is the number of data sets that are only 42 subjects. We used a LOSO technique in our pipeline to avoid this limitation.

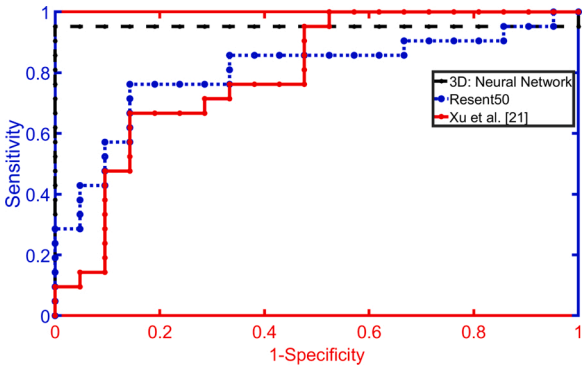

Fig. 13. The receiver operating characteristics (ROC) curves for the neural network (NN) and ResNet50 use the T2-MRI model.

The development of the proposed framework with high accuracy is our target that was achieved by using MP-MRIs (T2-MRI and DW-MRI) in which hand-crafted features are derived for each modality. The morphological and texture features are extracted from the T2-MRI modality, and the voxel-wise ADC map is derived from the DW-MRI modality. For more precision, the morphological features are extracted from only the V_w that reflect the physical meaning of the differentiation between T1 and T2 BC stages.

The accuracy of the diagnostic results for the proposed framework, 3D method, and NN as a classifier emphasizes that the morphological feature had higher accuracy than the DW-MRI modality. Moreover, the T2-MRI modality had the highest accuracy metrics among all categories. The ROC curves support the accuracy metrics in which the order for the classes with high accuracy are T2-MRI, morphological features, and DW-MRI. But it is essential to mention that feature fusion increases the system's overall accuracy, which is the main reason for using MP-MRI. Accuracy diagnostic will eventually lead to better treatment and improve the outcome of the treatment. Our CAD system's average processing time for a test subject is 62.97 seconds for all three processing stages (CNN segmentation, feature extraction, and classification. Namely, 45.59 seconds for the segmentation step, 9.52 seconds for features extraction, and 7.86 seconds for neural network classification. Our CAD software is primarily implemented in Python and Matlab, and the experiments were conducted on a Dell Precision workstation with an Intel (R) Xeon (R) W-2155 CPU running at 3.3 GHz and 128 GiB RAM.

Our work reported the diagnostic results for both 2D and 3D scenarios. The AUCs' values for both of them are 0.9705 and 0.9864, respectively, and also the accuracy values are 92.86% and 95.24%, respectively. Although the 3D analysis has shown a marginal increase over the 2D, the *P*-values demonstrates no statistically significant difference between 2D and 3D scenarios for individual modality and their fusion. The results, however, reflect the importance of 3D processing as it provides a comprehensive evaluation with complete descriptive information and details about BC. So, this ensures that our 3D approach is less sensitive for selecting a specific cross-section. Our proposed framework results were also confirmed by comparing different statistical ML classifiers, namely NN, RF, and SVM. For 2D and 3D scenarios, the RF classifier's diagnostic results had higher values than the SVM, and

both had smaller amounts than our proposed framework, neural network. Also, the AUCs of the Roc curves prove the value of the accuracy metrics.

Similarly, the diagnostic results for the deep features, ResNet50 and Xu et al. [21] approach were meager compared with the proposed neural network's framework. Besides, the ResNet50's accuracy, T2W-MRI modality, metrics are approximately the same as RF and SVM classifiers, with RF being slightly better. Therefore, these comparisons demonstrate the benefits of our CAD system. In particular, our proposed results compared with the ResNet50 provide us the importance of pathological segmentation, selecting the hand-crafted features, selecting a suitable classifier, and using MP-MRIs (T2-MRI and DW-MRI) as well as using morphological features for the V_w.

One of our limitations is the limited size of the data set, thus increasing our data will allow us to train and test our CAD system with other approaches (e.g., K-fold cross-validation or an independent training set) than LOSO, which is known to produce biased accuracy estimates. Besides, we used end-to-end convolution neural networks (i.e., ResNet50) on the T2W-MRI modality. The diagnostic results are almost the same as RF and SVM classifiers, but it is low compared with the neural network. Therefore, constructing the end-to-end convolution neural networks (CNN) with multi pathways for using MP-MRIs instead of only one model, e.g., the first pathway for T2W-MRI and the second one for DW-MRI that will enhance our diagnostic results.

6. Conclusion

In our work, a multiparametric, T2-weighted-magnetic resonance imaging (MRI) and diffusion-weighted MRI, computer-aided diagnostic (MP-CAD) system was developed to discriminate between T1 and T2 BC stages. The MRI scans' hand-crafted discriminative features were trained and tested for two various techniques: neural networks and statistical ML classifiers, as well as the ResNet50, which was used to validate our proposal results. The diagnostic results approved and confirmed that the proposed framework, neural network, is the best among all techniques.

Declaration of interests

None.

Authors' contribution

K. Hammouda, F. Khalifa, A. Soliman, M. Ghazal, and A. El-Baz: conceptualization, developing the proposed methodology for the analysis, and formal analysis. As well as software, validation, and visualization. K. Hammouda and F. Khalifa: prepared initial draft. M. Ghazal, M. Abou El-Ghar, and A. El-Baz: funding acquisition. M. Abou El-Ghar, and M.A. Badawy: collected local MRI data and participated in formal analysis. K. Hammouda, F. Khalifa, H.E. Darwish, and A. El-Baz: review and edit the revised version. A. El-Baz: project administration.

Declaration of Competing Interest

The authors report no declarations of interest.

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.compmedimag.2021.101911.

References

[1] American Cancer Society, 2021. Cancer Facts and Figures.

[2] Hammouda, K., Khalifa, F., Ghazal, M., Elmogy, M., Darwish, H.E., Elmaghraby, A., El-Baz, A., et al., 2018. MRI markers for early assessment of bladder cancer: a review. In: 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT). IEEE, pp. 185–191.

[3] Witjes, J.A., Lebret, T., Compérat, E.M., Cowan, N.C., De Santis, M., Bruins, H.M., Hernandez, V., Espinós, E.L., Dunn, J., Rouanne, M., et al., 2017. Updated 2016 eau guidelines on muscle-invasive and metastatic bladder cancer. Eur. Urol. 71 (3), 462–475.

[4] Suarez-Ibarrola, R., Hein, S., Reis, G., Gratzke, C., Miernik, A., 2019. Current and future applications of machine and deep learning in urology: a review of the literature on urolithiasis, renal cell carcinoma, and bladder and prostate cancer. World J. Urol. 1–19.

[5] Babjuk, M., B“ohle, A., Burger, M., Capoun, O., Cohen, D., Compérat, E.M., Hernández, V., Kaasinen, E., Palou, J., Rouprêt, M., et al., 2017. EAU guidelines on non-muscle-invasive urothelial carcinoma of the bladder: update 2016. Eur. Urol. 71 (3), 447–461.

[6] Wu, S., Zheng, J., Li, Y., Yu, H., Shi, S., Xie, W., Liu, H., Su, Y., Huang, J., Lin, T., 2017. A radiomics nomogram for the preoperative prediction of lymph node metastasis in bladder cancer. Clin. Cancer Res. 23 (22), 6904–6911.

[7] Takeuchi, M., Sasaki, S., Ito, M., Okada, S., Takahashi, S., Kawai, T., Suzuki, K., Oshima, H., Hara, M., Shibamoto, Y., 2009. Urinary bladder cancer: diffusion-weighted MR imaging-accuracy for diagnosing T stage and estimating histologic grade. Radiology 251 (1), 112–121.

[8] Turker, P., Bostrom, P.J., Wroclawski, M.L., van Rhijn, B., Kortekangas, H., Kuk, C., Mirtti, T., Fleshner, N.E., Jewett, M.A., Finelli, A., et al., 2012. Upstaging of urothelial cancer at the time of radical cystectomy: factors associated with upstaging and its effect on outcome. BJU Int. 110 (6), 804–811.

[9] Jakse, G., Algaba, F., Malmstr”om, P.-U., Oosterlinck, W., 2004. A second-look TUR in T1 transitional cell carcinoma: why? Eur. Urol. 45 (5), 539–546.

[10] Kobayashi, S., Koga, F., Kajino, K., Yoshita, S., Ishii, C., Tanaka, H., Saito, K., Masuda, H., Fujii, Y., Yamada, T., et al., 2014. Apparent diffusion coefficient value reflects invasive and proliferative potential of bladder cancer. J. Magn. Reson. Imaging 39 (1), 172–178.

[11] Yang, R., Du, Y., Weng, X., Chen, Z., Wang, S., Liu, X., 2020. Automatic recognition of bladder tumours using deep learning technology and its clinical application. Int. J. Med. Robot. Comput. Assist. Surg. e2194.

[12] Lucas, M., Jansen, I., van Leeuwen, T.G., Oddens, J.R., de Bruin, D.M., Marquering, H.A. Deep learning-based recurrence prediction in patients with non-muscle-invasive bladder cancer, Eur. Urol. Focus.

[13] Shkolyar, E., Jia, X., Chang, T.C., Trivedi, D., Mach, K.E., Meng, M.Q.-H., Xing, L., Liao, J.C., 2019. Augmented bladder tumor detection using deep learning. Eur. Urol. 76 (6), 714–718.

[14] Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., Van Stiphout, R.G., Granton, P., Zegers, C.M., Gillies, R., Boellard, R., Dekker, A., et al., 2012. Radiomics: extracting more information from medical images using advanced feature analysis. Eur. J. Cancer 48 (4), 441–446.

[15] Gillies, R.J., Kinahan, P.E., Hricak, H., 2016. Radiomics: images are more than pictures, they are data. Radiology 278 (2), 563–577.

[16] Huang, Y.-Q., Liang, C.-H., He, L., Tian, J., Liang, C.-S., Chen, X., Ma, Z.-L., Liu, Z.-Y., et al., 2016. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. J. Clin. Oncol. 34 (18), 2157–2164.

[17] Lambin, P., Leijenaar, R.T., Deist, T.M., Peerlings, J., De Jong, E.E., Van Timmeren, J., Sanduleanu, S., Larue, R.T., Even, A.J., Jochems, A., et al., 2017. Radiomics: the bridge between medical imaging and personalized medicine. Nat. Rev. Clin. Oncol. 14 (12), 749.

[18] Xu, X., Zhang, X., Tian, Q., Zhang, G., Liu, Y., Cui, G., Meng, J., Wu, Y., Liu, T., Yang, Z., et al., 2017. Three-dimensional texture features from intensity and high-order derivative maps for the discrimination between bladder tumors and wall tissues via MRI. Int. J. Comput. Assist. Radiol. Surg. 12 (4), 645–656.

[19] Xu, X., Zhang, X., Tian, Q., Wang, H., Cui, L.-B., Li, S., Tang, X., Li, B., Dolz, J., Ayed, I.b., et al., 2019. Quantitative identification of nonmuscle-invasive and muscle-invasive bladder carcinomas: a multiparametric MRI radiomics analysis. J. Magn. Reson. Imaging 49 (5), 1489–1498.

[20] Xu, X., Wang, H., Du, P., Zhang, F., Li, S., Zhang, Z., Yuan, J., Liang, Z., Zhang, X., Guo, Y., et al., 2019. A predictive nomogram for individualized recurrence stratification of bladder cancer using multiparametric MRI and clinical risk factors. J. Magn. Reson. Imaging 50 (6), 1893–1904.

[21] Xu, X., Liu, Y., Zhang, X., Tian, Q., Wu, Y., Zhang, G., Meng, J., Yang, Z., Lu, H., 2017. Preoperative prediction of muscular invasiveness of bladder cancer with radiomic features on conventional MRI and its high-order derivative maps. Abdom. Radiol. 42 (7), 1896–1905.

[22] Hammouda, K., Khalifa, F., Soliman, A., Ghazal, M., Abou El-Ghar, M., Haddad, A., Elmogy, M., Darwish, H., Keynton, R., El-Baz, A., 2019. A deep learning-based approach for accurate segmentation of bladder wall using MR images. In: 2019 IEEE International Conference on Imaging Systems and Techniques (IST). IEEE, pp. 1–6.

[23] Dolz, J., Xu, X., Rony, J., Yuan, J., Liu, Y., Granger, E., Desrosiers, C., Zhang, X., Ben Ayed, I., Lu, H., 2018. Multiregion segmentation of bladder cancer structures in MRI with progressive dilated convolutional networks. Med. Phys. 45 (12), 5482–5493.

[24] Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B., 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. 36, 61–78.

[25] Kr“ahenb”uhl, P., Koltun, V., 2011. Efficient inference in fully connected CRFs with Gaussian edge potentials. Advances in Neural Information Processing Systems 109–117.

[26] Hammouda, K., Khalifa, F., Soliman, A., Abdeltawab, H., Ghazal, M., Abou El-Ghar, M., Haddad, A., Darwish, H.E., Keynton, R., El-Baz, A., 2020. A 3D CNN with a learnable adaptive shape prior for accurate segmentation of bladder wall using MR images. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, pp. 935–938.
[27] de Haas, R.J., Steyvers, M.J., Fütterer, J.J., 2014. Multiparametric MRI of the bladder: ready for clinical routine? Am. J. Roentgenol. 202 (6), 1187–1195.
[28] Hu, Y., Liang, Z., Song, B., Han, H., Pickhardt, P.J., Zhu, W., Duan, C., Zhang, H., Barish, M.A., Lascarides, C.E., 2016. Texture feature extraction and analysis for polyp differentiation via computed tomography colonography. IEEE Trans. Med. Imaging 35 (6), 1522–1531.
[29] Han, F., Wang, H., Zhang, G., Han, H., Song, B., Li, L., Moore, W., Lu, H., Zhao, H., Liang, Z., 2015. Texture feature analysis for computer-aided diagnosis on pulmonary nodules. J. Digit. Imaging 28 (1), 99–115.
[30] Song, B., Zhang, G., Lu, H., Wang, H., Zhu, W., Pickhardt, P.J., Liang, Z., 2014. Volumetric texture features from higher-order images for diagnosis of colon lesions via CT colonography. Int. J. Comput. Assist. Radiol. Surg. 9 (6), 1021–1031.
[31] Zwanenburg, A., Vallières, M., Abdalah, M.A., Aerts, H.J., Andrearczyk, V., Apte, A., Ashrafinia, S., Bakas, S., Beukinga, R.J., Boellaard, R., et al., 2020. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology 295 (2), 328–338.
[32] Kolossváry, M., Kellermayer, M., Merkely, B., Maurovich-Horvat, P., 2018. Cardiac computed tomography radiomics. J. Thorac. Imaging 33 (1), 26–34.
[33] Kolossváry, M., Karády, J., Szilveszter, B., Kitslaar, P., Hoffmann, U., Merkely, B., Maurovich-Horvat, P., 2017. Radiomic features are superior to conventional quantitative computed tomographic metrics to identify coronary plaques with napkin-ring sign. Circ. Cardiovasc. Imaging 10 (12), e006843.
[34] Galloway, M.M., 1975. Texture analysis using gray level run lengths. Comput. Graph. Image Process. 4 (2), 172–179.
[35] Mohanty, A.K., Beberta, S., Lenka, S.K., 2011. Classifying benign and malignant mass using GLCM and GLRLM based texture features from mammogram. Int. J. Eng. Res. Appl. 1 (3), 687–693.
[36] Singh, K., 2016. A comparison of gray-level run length matrix and gray-level co-occurrence matrix towards cereal grain classification. Int. J. Comput. Eng. Technol. (IJCET) 7 (6), 9–17.
[37] Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybern. 3 (6), 610–621.
[38] Ávila, M., Caballero, D., Antequera, T., Durán, M.L., Caro, A., Pérez-Palacios, T., 2018. Applying 3D texture algorithms on MRI to evaluate quality traits of loin. J. Food Eng. 222, 258–266.
[39] Moya, L., Zakeri, H., Yamazaki, F., Liu, W., Mas, E., Koshimura, S., 2019. 3d gray level co-occurrence matrix and its application to identifying collapsed buildings. ISPRS J. Photogram. Remote Sens. 149, 14–28.
[40] Ashrafinia, S., et al., 2019. Quantitative Nuclear Medicine Imaging Using Advanced Image Reconstruction and Radiomics. Johns Hopkins University (Ph.D. Thesis).
[41] Avcu, S., Koseoglu, M., Ceylan, K., Dbulutand, M., Unal, O., 2011. The value of diffusion-weighted MRI in the diagnosis of malignant and benign urinary bladder lesions. Br. J. Radiol. 84 (1006), 875–882.
[42] Legland, D., Kiêu, K., Devaux, M.-F., 2007. Computation of minkowski measures on 2d and 3d binary images. Image Anal. Stereol. 26 (2), 83–92.
[43] He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. Proceedings of the IEEE International Conference on Computer Vision 1026–1034.
[44] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. ACM SIGKDD Explor. Newsl. 11 (1), 10–18.
[45] He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 770–778.