


Received September 7, 2018, accepted October 2, 2018, date of publication October 9, 2018, date of current version December 27, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2874803

# Novel 3D Radiomic Features for Computer-Aided Polyp Detection in CT Colonography

YACHENG REN<sup>1</sup>, YIZHI CHEN<sup>1</sup>, LIN LU<sup>2</sup>, AND JUN ZHAO<sup>1,3</sup> , (Member, IEEE)

<sup>1</sup>School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

<sup>2</sup>Department of Radiology, Columbia University Medical Center, New York, NY 10032, USA

<sup>3</sup>SJTU-UH Institute for Medical Imaging Technology, MED-X Research Institute, Shanghai Jiao Tong University, Shanghai 200240, China

Corresponding author: Jun Zhao (junzhao@sjtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 81371634, in part by the National Key Research and Development Program under Grant 2016YFC0104608, in part by the National Basic Research Program of China under Grant 2010CB834302, and in part by the Shanghai Jiao Tong University Medical Engineering Cross Research Funds under Grant YG2017ZD10 and Grant YG2014ZD05.

**ABSTRACT** State-of-the-art computer-aided detection (CAD) systems for colonic polyps in computed tomographic colonography (CTC) tend to yield high detection sensitivities with high false positive (FP) rates. This paper proposes a novel CTC CAD system using 3-D radiomic features that can obtain high detection sensitivity at low FP rate. Our previous shape-based methods are used to generate polyp candidates. Then, a wide variety of radiomic features are developed in the 3-D domain for each initial candidate, including the multiscale Weber local descriptors and the statistical descriptors calculated from various volumetric feature maps on the basis of computed tomography (CT) density, CT gradient, fractal dimension, curvature, and fast radial symmetry transform. These new 3-D radiomic features can characterize polyp candidates in terms of the shape, texture heterogeneity, and salient pattern. FP reduction is finally performed by Random Forests with a developed score rank method to tune the training set. 510 fluid-tagging CT scans from 255 patients with 130 polyps  $\geq 5$  mm were utilized to validate the proposed system in the fivefold cross-validation strategy. The detection result reached 98.5% by-polyp sensitivity at 2.0 FPs per scan for polyps  $\geq 5$  mm. Experimental results indicate that the proposed system yielded a detection performance with high sensitivity and low FP rate. We believe that the proposed system would assist radiologists in increasing the detection accuracy and reducing the interpretation time during colon cancer screening.

**INDEX TERMS** Computer-aided detection, colonic polyp, radiomic feature, Random Forests.

## I. INTRODUCTION

Colorectal cancer is the third most common cause of cancer death and new cancer cases in 2018 for men and women in the U.S. according to the statistics from American Cancer Society [1]. A total of 140250 new cases of colorectal cancer will be diagnosed in the U.S. in 2018, with 50630 deaths resulting from the disease. Adenomatous polyps are important precursors to colorectal cancer, and polyps require 5–15 years to evolve into cancers. Thus, the risk of colorectal cancers can be reduced by early detection and removal of colonic polyps before or during malignant transformation [2], [3].

Computed tomographic colonography (CTC) is a minimally invasive screening technique that detects polyps by using computed tomography (CT) images [3], [4]. To aid radiologists in reducing the detection time and increasing

the detection accuracy during CTC interpretation, computer-aided detection (CAD) has been utilized [5], [6]. To date, various CAD systems have been developed for colonic polyp detection. Vining *et al.* [7] detected colonic polyps using abnormal colon wall thickness. Summers *et al.* [8] calculated the mean, Gaussian, and principal curvatures from the colon wall inner surface to distinguish polyps from false positives (FPs). Yoshida and Näppi [9] adopted shape index (SI) and curvedness (CV) to detect polyp candidates and the quadratic discriminant analysis to reduce FPs. Näppi and Yoshida [10] eliminated FP candidates through feature-guided analysis. Näppi *et al.* [11] detected colorectal masses based on fuzzy merging and wall-thickening analysis. Jerebko *et al.* [12] used an ensemble of support vector machines for polyp candidate classification. Hong *et al.* [13] detected polyps by combining texture and shape analysis

with volume rendering and conformal colon flattening. Konukoglu and Acar [14] utilized heat diffusion patterns for polyp detection. Li *et al.* [15] used the Pareto front to optimize the generation of initial polyp detections. Suzuki *et al.* [16] used a mixture of expert 3D massive-training artificial neural networks to reduce FPs. Kilic *et al.* [17] proposed fuzzy rule-based 3D template matching to detect polyps. Yao *et al.* [18] used a ray-casting algorithm to compute the topographical height map for polyp measurement and FP reduction. van Ravesteijn *et al.* [19] used logistic regression to detect polyps. Zhu *et al.* [20] utilized the Knutsson mapping to improve curvature estimation for FP reduction. Wang *et al.* [21] proposed adaptive SI and CV to detect polyp candidates from decomposed colon structures. Ren *et al.* [3] detected polyp candidates based on SI and multiscale dot enhancement filter and reduced FPs through radiomic features. Tulum *et al.* [5] extracted projection-based features to reduce FP candidates generated by Laplacian of Gaussian filters. Ren *et al.* [6] developed morphological features to further reduce FPs. These CAD systems mainly consist of two stages: detection of polyp candidates and reduction of FP candidates. Considering that any polyp missed in the first stage remains undetected in the final detection results, all possible polyp candidates must be included in the first stage to ensure sufficient overall performance for a CAD system, which likely results in numerous FP candidates for the second stage. Thus, developing informative features or new classification methods to reduce numerous FP candidates is a requirement in the second stage.

Convolutional neural networks (CNNs) seem to be a good solution for FP reduction in the second stage because they can obtain better performance than conventional classifiers in image-based object recognition and classification tasks. CNNs compute image features through a series of convolutional filters, whose kernel elements are determined through the supervised training. That is, CNNs can learn discriminative image features without the need for handcrafted image descriptors. However, the main problem for their applications in CTC is the lack of a large number of annotated CTC training images. Thus, studies about the applications of CNNs in CTC colonic polyp detection are few. Roth *et al.* [22] used 2.5D CNNs and random view aggregation to reduce FPs for polyp detection in CTC. The transformations of translation, rotation, and scaling were used to augment 2.5D views of polyp candidates for the training of CNNs, and the prediction probability of a polyp candidate was computed by averaging multiple prediction scores from the trained CNNs for the candidate. Näppi *et al.* [23] used surface rendering method to generate virtual endoluminal images for the polyp candidates prompted by a CAD system and then applied these virtual endoluminal images to deep transfer learning of a pre-trained CNN for FP reduction. In addition to CNNs, radiomics has provided another good solution for FP reduction in the second stage of a CTC CAD system. As an emerging field, radiomics converts medical images into mineable data through an automated high-throughput extraction of large amounts (200+)

of quantitative features of medical images [3], [24], [25]. It offers a nearly limitless supply of imaging biomarkers that can potentially aid cancer detection and diagnosis, assessment of prognosis, prediction of response to treatment, and monitoring of disease status [26], [27]. Radiomics has been used in lung, breast, prostate, and head-and-neck cancer studies [3], [24], [28], [29], and the potential of radiomic features for comprehensive characterization of tumors in texture and shape has been demonstrated. Thus, we develop a novel CTC CAD system with a comprehensive set of 3D radiomic features for colonic polyp detection in this study. Polyp candidates are detected through our previous candidate detection method [3], [6]. Then, 80 basic radiomic features from other CTC studies and 443 new radiomic features specified for polyps are extracted for each polyp candidate. The resulting 523 radiomic features include multiscale Weber Local Descriptors (MWLDs) and the statistical descriptors calculated from various volumetric feature maps based on CT density, CT gradient, fractal dimension (FD), SI, CV, and fast radial symmetry transform (FRST). Random Forests classifier [30] combined with the tuning and balancing method of the training samples is finally used to classify polyp candidates.

The main contributions in our work are summarized as follows:

- 1) Plentiful and diverse radiomic features are collected and developed to characterize the shape and texture heterogeneity of polyp candidates. A total of 523 quantitative radiomic features (i.e., 80 basic radiomic features and 443 new radiomic features) are extracted to distinguish true positive (TP) detections from FPs. To the best of our knowledge, the FD and FRST feature maps, MWLD, gray-level size zone method (GLSZM) [31], neighborhood gray-tone difference method (NGTDM) [32], gray-level difference method (GLDM) [33], and surrounding region dependence method (SRDM) [34] are the first to be applied for polyp detection in CTC.

- 2) Considering the 3D properties of polyp candidates, we compute or expand all radiomic features in 3D space. The 3D GLSZM, NGTDM, GLDM, and SRDM features and the newly developed 3D MWLD features are calculated from various volumetric feature maps.

- 3) We propose a novel CTC CAD system through radiomics approach. A score rank method is developed to tune the training set, and Random Forests classifier is applied to classify candidates. The proposed CTC CAD system can achieve a high performance of 98.5% by-polyp sensitivity at 2.0 FPs per scan for polyps  $\geq 5$  mm.

The remainder of this paper is organized as follows. We describe our method in Section II and report the experiments and results in Section III. The discussion and conclusion are given in Sections IV and V, respectively.

## II. METHOD

The proposed CTC CAD system includes two stages: polyp candidate detection and FP reduction. Fig. 1 shows the

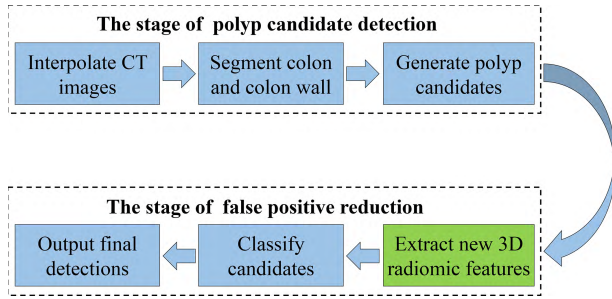


FIGURE 1. Overview of the proposed CTC CAD system.

overview of the proposed CTC CAD system. Our main contribution focuses on the extraction of new 3D radiomic features.

### A. POLYP CANDIDATE DETECTION

Polyp candidates are detected through our previous shape-based method [3], [6]. First, the original CT images are linearly interpolated into an isotropic volumetric CT dataset. Second, the colon volume is segmented from the isotropic CT dataset by a partial volume segmentation method [35] that cleanses the tagged materials in CT images. The colon wall is generated by the exclusive or operation between the original colon volume and 3D dilated colon volume [6]. The CT density mapping method [3] is used to correct the pseudo-enhanced CT density in the colon wall. Finally, polyp candidates are generated by thresholding the SI and multi-scale dot enhancement filter [36] that are calculated based on the Gaussian-smoothed geodesic distance field (GDF) [37] within the colon wall.

### B. RADIOMIC FEATURE EXTRACTION

The radiomic features are calculated based on the statistical descriptors. Thus, we introduce the statistical descriptors before extracting radiomic features.

#### 1) STATISTICAL DESCRIPTORS

The statistical descriptors used in this study include intensity features (i.e., first-order statistical descriptors) and texture features (i.e., second-order statistical descriptors). The intensity features are calculated from the image intensity histogram within the volume of interest (VOI), and they can globally describe the distribution of intensity values of individual voxels without concern for spatial relationships. The texture features, which include GLSZM [31], NGTDM [32], GLDM [33], and SRDM [34] features, can quantify intralesion heterogeneity differences in the texture that is defined as the spatial arrangement of voxels of different intensities. That is, GLSZM, NGTDM, GLDM, and SRDM features can quantitatively describe the distribution patterns or spatial relationships between intensity levels of different voxels within a VOI image. To the best of our knowledge, we are the first to utilize GLSZM, NGTDM, GLDM, and SRDM features for CTC polyp detection. Considering the 3D properties of polyp candidates, we compute the 3D GLSZM and NGTDM

features in accordance with [38]. On the basis of 2D GLDM and SRDM features [33], [34], the calculations of GLDM and SRDM features are expanded from 2D to 3D space in this study.

#### a: INTENSITY FEATURES

Twenty intensity features are extracted from the intensity histogram of the VOI image. The first 14 intensity features are from [24]: energy, entropy, kurtosis, maximum, mean, mean absolute deviation, median, minimum, range, root mean square, skewness, standard deviation, uniformity, and variance. The remaining six intensity features are from [39]: the mean and variance of top 25% high intensity values, the mean and variance of top 50% high intensity values, and the mean and variance of top 75% high intensity values.

#### b: 3D GLSZM AND NGTDM FEATURES

The 3D GLSZM and NGTDM features are calculated by accounting for the neighboring properties of voxels with 26-voxel connectivity in the 13 directions of 3D space. However, the 6 voxels at a distance of 1 voxel, the 12 voxels at a distance of  $\sqrt{2}$  voxels, and the 8 voxels at a distance of  $\sqrt{3}$  voxels around the center voxels are treated differently in the calculation of 3D NGTDM features to account for the differences in discretization length [38]. The detailed definitions and computational formulas of 13 3D GLSZM features and 5 3D NGTDM features are given in [38]. In this study, we only list the descriptions of these features.

*Thirteen 3D GLSZM features* [38], [40]: small-zone emphasis, large-zone emphasis, gray-level nonuniformity, zone-size nonuniformity, zone percentage, low-gray-level zone emphasis, high-gray-level zone emphasis, small-zone low-gray-level emphasis, small-zone high-gray-level emphasis, large-zone low-gray-level emphasis, large-zone high-gray-level emphasis, gray-level variance, and zone-size variance.

*Five 3D NGTDM features* [38], [40]: coarseness, contrast, busyness, complexity, and strength.

#### c: 3D GLDM FEATURES

On the basis of 2D GLDM in [33], the 3D GLDM features describe the occurrence of two arbitrary voxels that have a given absolute difference in intensity level and are separated by a specific displacement. Let  $\delta_i$  denote a displacement vector.  $\delta_i = d * D_i$ , where  $d$  refers to a specified intersample spacing,  $i = 1, 2, \dots, 13$ , and  $D_i$  is defined as follows [6]:  $D_1 = (1, 0, 0)$ ;  $D_2 = (0, 1, 0)$ ;  $D_3 = (0, 0, 1)$ ;  $D_4 = (0, 1, 1)$ ;  $D_5 = (1, 0, 1)$ ;  $D_6 = (1, 1, 0)$ ;  $D_7 = (0, 1, -1)$ ;  $D_8 = (1, 0, -1)$ ;  $D_9 = (1, -1, 0)$ ;  $D_{10} = (1, 1, 1)$ ;  $D_{11} = (1, 1, -1)$ ;  $D_{12} = (1, -1, 1)$ ;  $D_{13} = (1, -1, -1)$ .

Four feature measures (i.e., contrast, angular second moment, entropy, and mean) are computed for each displacement vector  $\delta_i$ :

$$I_{\delta_i}(v) = |I(v) - I(v + \delta_i)|, \quad (1)$$

$$p_{\delta_i}(j) = \text{Prob}(I_{\delta_i}(v) = j), \quad (2)$$

$$CON_i = \sum_{j=1}^N j^2 p_{\delta_i}(j), \quad (3)$$

$$ASM_i = \sum_{j=1}^N p_{\delta_i}(j)^2, \quad (4)$$

$$ENT_i = - \sum_{j=1}^N p_{\delta_i}(j) \log p_{\delta_i}(j), \quad (5)$$

$$MEAN_i = \frac{1}{N} \sum_{j=1}^N j p_{\delta_i}(j), \quad (6)$$

where  $I$  is a discrete image,  $v$  is a voxel in  $I$ ,  $I_{\delta_i}$  is the absolute difference image,  $j$  is an intensity level in  $I_{\delta_i}$ ,  $p_{\delta_i}$  is the probability density of  $I_{\delta_i}$ , and  $N$  is the number of intensity levels in  $I_{\delta_i}$ .  $CON_i$ ,  $ASM_i$ ,  $ENT_i$ , and  $MEAN_i$  are contrast, angular second moment, entropy, and mean, respectively. The mean, maximum, minimum, range, and variance of  $CON_i$  are computed as the first five 3D GLDM features. The same operations are also applied to  $ASM_i$ ,  $ENT_i$ , and  $MEAN_i$ . In this manner, 20 3D GLDM features are computed for each value of the intersample spacing  $d$ . Here, the intersample spacing  $d$  is set to 1 and 2, respectively. Thus, 40 3D GLDM features are considered in this study.

#### d: 3D SRDM FEATURES

On the basis of 2D SRDM in [34], the 3D SRDM features are computed from a second-order histogram in two surrounding regions. Let  $I$  denote a quantized image.  $B_1$ ,  $B_2$ , and  $B_3$  denote three cubic windows centered at a voxel  $v$ . The side lengths of  $B_1$ ,  $B_2$ , and  $B_3$  are  $l_1$ ,  $l_2$ , and  $l_3$  ( $l_1 \leq l_2 \leq l_3$ ), which are empirically set to 3, 5, and 7, respectively. The region between  $B_1$  and  $B_2$  is the inner surrounding region  $R_1$ , and the region between  $B_2$  and  $B_3$  is the outer surrounding region  $R_2$  (Fig. 2).  $m$  and  $n$  are the voxel number in regions  $R_1$  and  $R_2$ , respectively. The 3D SRDM features are computed as follows:

$$C_{R_1}(v) = \#\{q | q \in R_1 \text{ and } I(v) - I(q) > \text{thre}\}, \quad (7)$$

$$C_{R_2}(v) = \#\{q | q \in R_2 \text{ and } I(v) - I(q) > \text{thre}\}, \quad (8)$$

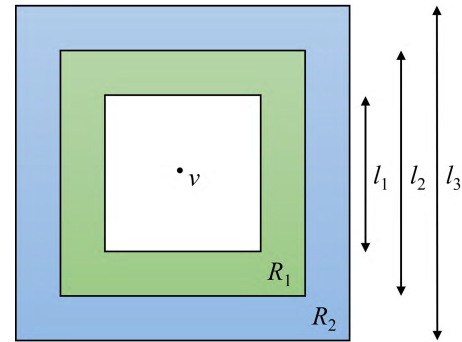
$$A(i, j) = \#\{v | C_{R_1}(v) = i \text{ and } C_{R_2}(v) = j, v \in V\}, \quad (9)$$

$$N = \sum_{i=0}^m \sum_{j=0}^n A(i, j), \quad (10)$$

$$S(i, j) = \begin{cases} 1/A(i, j), & \text{if } A(i, j) > 0, \\ 0, & \text{otherwise,} \end{cases} \quad (11)$$

$$HWS = \frac{1}{N} \sum_{i=0}^m \sum_{j=0}^n j^2 S(i, j), \quad (12)$$

$$VWS = \frac{1}{N} \sum_{i=0}^m \sum_{j=0}^n i^2 S(i, j), \quad (13)$$



**FIGURE 2.** Illustration of surrounding regions for the target voxel  $v$ .  $R_1$  and  $R_2$  indicate the inner and outer surrounding regions, respectively.  $l_1$ ,  $l_2$ , and  $l_3$  are the side lengths of the three cubic windows.

$$DWS = \frac{1}{N} \sum_{k=0}^{m+n} k^2 \left( \sum_{i=0}^m \sum_{j=0}^n S(i, j) \right), \quad (14)$$

$$GWS = \frac{1}{N} \sum_{i=0}^m \sum_{j=0}^n ij S(i, j), \quad (15)$$

where  $q$  is a voxel,  $\text{thre}$  is a given threshold value,  $\#$  denotes the number of elements in the set, and  $V$  is a VOI.  $HWS$ ,  $VWS$ ,  $DWS$ , and  $GWS$  represent four 3D SRDM features (i.e., horizontal-weighted sum, vertical-weighted sum, diagonal-weighted sum, and grid-weighted sum [34], respectively). In this study, the threshold value  $\text{thre}$  is set to 0 and 1, respectively. Therefore, eight 3D SRDM features are computed.

## 2) RADIOMIC FEATURES

A total of 523 radiomic features are extracted from the generated polyp candidates for FP reduction, including 80 basic radiomic features and 443 new radiomic features. The 80 basic radiomic features are the traditional image features expanded from previous CTC CAD studies [9], [10], [41], [42] and comprise 20 intensity features from CT image, CT gradient image, SI image, and CV image, respectively. CT gradient measures the internal structures of the polyp candidate, and CT gradient image is computed by using the  $3 \times 3 \times 3$  Sobel operator. SI measures the topological shape of the local neighborhood of a voxel, whereas CV measures the size of the structure of the effective curvature. Computational details of SI and CV images are given in [9]. The 443 new radiomic features include 66 CT image features, 66 CT gradient features, 87 FD features, 66 SI features, 86 FRST features, and 72 MWLD features. These new radiomic features are presented in the following text. To the best of our knowledge, the MWLD and the feature maps of FD and FRST are the first to be introduced for colonic polyp detection in CTC.

### a: NEW CT IMAGE, CT GRADIENT, FD, AND SI FEATURES

CT image can be perceived as a rugged surface if CT image intensities are regarded as the height above a plane [43].

FD measures the texture complexity of a structure that is not exactly Euclidean (i.e., “in-between dimensions”) giving the potential for a richer description of the examined surface [43]. The FD image is computed from CT image in accordance with [43]. Twenty intensity features and another feature “lacunarity” [43] are calculated from the FD image within each polyp candidate. Lacunarity can measure the “lumpiness” of fractal data and provide the meta-information about calculated FD values in the image. High lacunarity indicates high degree of inhomogeneity in the examined fractal area [43].

Prior to computation of texture features, the full intensity range of the target image is quantized into a smaller number of intensity levels. Here, the intensity values in CT image, CT gradient image, FD image, and SI image are resampled into equally spaced bins by using the bin widths of 25, 25, 0.05, and 0.05, respectively. This quantization process can reduce image noises and normalize intensities across all patient datasets. Then, 66 texture features (i.e., 13 3D GLSZM features, 5 3D NGTDM features, 40 3D GLDM features, and 8 3D SRDM features) are calculated from the quantized CT image, CT gradient image, FD image, and SI image, respectively. That is, 66 new CT image features, 66 new CT gradient features, 87 new FD features, and 66 new SI features are extracted for each generated polyp candidate.

*b: NEW FRST FEATURES*

FRST enhances the points of high local radial symmetry within a scene. On the basis of 2D FRST [44], the 3D FRST is developed to compute the 3D volumetric FRST image for polyp candidates. Let  $O_n$  and  $M_n$  denote the orientation and magnitude projection images at an integer radius  $n$ , respectively.  $v$  is a voxel in the CT image.  $g(v)$  is the gradient vector at  $v$ .  $v_+$  and  $v_-$  are the positively-affected point and negatively-affected point of  $v$ , respectively (Fig. 3). The orientation projection image  $O_n$  and magnitude projection image  $M_n$  are initially zero and can be updated as follows:

$$v_+ = v + \text{round} \left( \frac{g(v)}{\|g(v)\|} n \right), \tag{16}$$

$$v_- = v - \text{round} \left( \frac{g(v)}{\|g(v)\|} n \right), \tag{17}$$

$$O_n(v_+) = O_n(v_+) + 1, \tag{18}$$

$$O_n(v_-) = O_n(v_-) - 1, \tag{19}$$

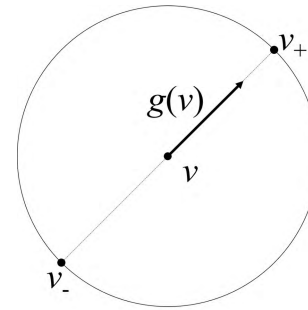
$$M_n(v_+) = M_n(v_+) + \|g(v)\|, \tag{20}$$

$$M_n(v_-) = M_n(v_-) - \|g(v)\|. \tag{21}$$

Given the image noises and the weak gradients inside polyps, the gradient magnitude  $\|g(v)\|$  at  $v$  should be larger than 0.05. Then, the full transform  $S$  is computed as follows:

$$F_n(v) = \frac{M_n(v)}{k_n} \left( \frac{|\min \{O_n(v), k_n\}|}{k_n} \right)^\alpha, \tag{22}$$

$$S = \frac{1}{n_{max}} \sum_{n=1}^{n_{max}} F_n * A_n, \tag{23}$$



**FIGURE 3.** Illustration of 3D FRST.  $v_+$  and  $v_-$  are the positively-affected point and negatively-affected point of voxel  $v$ , respectively. The gradient vector  $g(v)$  at voxel  $v$  points to  $v_+$  and points away from  $v_-$ .

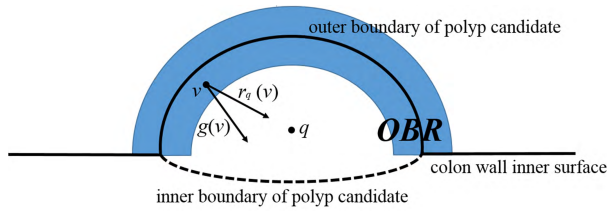
where  $\alpha$  is the radial strictness parameter,  $k_n$  is a scaling factor that normalizes  $O_n$  and  $M_n$  across different radii,  $n_{max}$  is the maximum radius, and  $A_n$  is a three-dimensional Gaussian function with the size  $2n + 1$  and standard deviation  $0.5n$ . Here,  $\alpha$  and  $k_n$  are empirically set to 2 and 26, respectively.  $n_{max}$  is the ceil integer value of the polyp candidate radius. Twenty intensity features are calculated from the 3D FRST image within the polyp candidate region. Then, the intensity values in the 3D FRST image are resampled into equally spaced bins by using a bin width of 0.5. 66 texture features (i.e., 13 3D GLSZM features, 5 3D NGTDM features, 40 3D GLDM features, and 8 3D SRDM features) are calculated from the quantized 3D FRST image. Thus, a total of 86 new FRST features are extracted for each polyp candidate.

*c: NEW MWLD FEATURES*

WLD is inspired by the Weber’s Law, i.e., the human perception of a pattern depends on not only the change of a stimulus (such as sound and lighting) but also the original intensity of the stimulus [45]. WLD is a discriminative descriptor and has been used for texture classification and face detection in 2D digital images [45], [46]. Considering the 3D properties of polyp candidates, we develop the new WLD specifically for polyp candidate classification in this study. The new WLD characterizes each polyp candidate by using a 2D histogram of differential excitation and orientation. Differential excitation is computed as the arctangent value of the ratio between two terms: one is the sum of CT density difference of a current voxel against its neighbors; the other is the CT density of the current voxel. CT density at the outer boundary region of the colonic polyp differs, but that at the other region of the colonic polyp is approximately constant. Thus, we only compute the differential excitation at the outer boundary regions of polyp candidates. Let  $v$  denote a voxel at the outer boundary region  $OBR$  of the polyp candidate  $V$ .  $OBR$  is defined as follows:

$$OBR = \text{dilate}(B), \tag{24}$$

where  $B$  is the outer boundary of the polyp candidate  $V$ , and  $\text{dilate}(\cdot)$  denotes the dilation operation with a radius of  $n$  voxels.  $N_r(v)$  is the set of voxels on the sides of the cubic box that has the center  $v$  and side length  $(2r + 1)$ .  $p$  is a voxel of



**FIGURE 4.** Illustration of orientation component of a polyp candidate. Voxel  $v$  is at the outer boundary region  $OBR$ .  $g(v)$  and  $r_q(v)$  are the unit gradient vector and unit centripetal vector at voxel  $v$ , respectively.  $r_q(v)$  points from voxel  $v$  to the polyp candidate center  $q$ .

$N_r(v)$ , and  $I$  is the CT image.  $\eta(v)$  is the differential excitation at voxel  $v$ , which is computed as follows:

$$\eta(v) = \arctan \left[ \frac{1}{I(v)} \sum_{p \in N_r(v)} (I(p) - I(v)) \right], \quad (25)$$

where  $\eta \in [-\pi/2, \pi/2]$ . Then,  $\eta$  is linearly quantized into  $M$  dominant differential excitations as shown below:

$$\{\xi_i\} = \text{mod} \left( \left\lfloor \frac{\eta + \pi/2}{\pi/M} \right\rfloor, M \right), \quad i = 0, 1, 2, \dots, M - 1. \quad (26)$$

The orientation component  $\theta(v)$  at voxel  $v$  is the angle between the unit gradient vector  $g(v)$  and unit centripetal vector  $r_q(v)$ .  $r_q(v)$  points from  $v$  to the polyp candidate center  $q$  (Fig. 4).  $\theta(v)$  is computed as follows:

$$\theta(v) = \arccos [g(v) \cdot r_q(v)], \quad (27)$$

where  $\theta \in [0, \pi]$ . Then,  $\theta$  is linearly quantized into  $T$  dominant orientations as follows:

$$\{\psi_j\} = \text{mod} \left( \left\lfloor \frac{\theta}{\pi/T} \right\rfloor, T \right), \quad j = 0, 1, 2, \dots, T - 1. \quad (28)$$

The quantized differential excitation  $\xi$  and orientation  $\psi$  are used to compute the 2D histogram  $\{WLD(\xi_i, \psi_j)\}$ ,  $i = 0, 1, 2, \dots, M - 1$ ,  $j = 0, 1, 2, \dots, T - 1$ . The element  $WLD(\xi_i, \psi_j)$  at row  $i$  and column  $j$  of the 2D histogram is the probability of occurrence of the dominant differential excitation  $\xi_i$  on the dominant orientation  $\psi_j$ . Then, all rows of the 2D histogram are concatenated as a 1D feature vector, i.e., a WLD feature vector. The WLD feature vector computes the salient patterns (i.e., differential excitations) and then builds statistics on these salient patterns along with the orientation components of the target voxels. For calculation of a MWLD feature vector, the radius  $r$  is set as different values. For each value of radius  $r$ , we compute a WLD feature vector. All WLD feature vectors are finally concatenated as a MWLD feature vector that can capture multigranularity salient patterns and is robust to rotation and scaling. In this study,  $n$ ,  $M$ , and  $T$  are empirically set to 2, 6, and 4, respectively.  $r$ , in turn, is set to 1, 2, and 3. Thus, 72 ( $6 \times 4 \times 3$ ) new MWLD features are computed for each polyp candidate.

### C. CANDIDATE CLASSIFICATION

Random Forests classifier [30] is an ensemble of decision trees. In the training procedure, each tree is constructed by using a bootstrap sample from the entire training set, and each node is split by using the best feature among a subset of features randomly selected without replacement at that node [3]. In the testing procedure, prediction of each testing sample is carried out by a majority vote. Random Forests classifier is robust against overfitting and can train the samples of high-dimensional feature vectors without additional feature selection [6]. Random Forests classifier involves three parameters (i.e., number of trees, number of samples for constructing each tree, and number of features in the random feature subset for splitting each node) and is usually not very sensitive to their values. Therefore, Random Forests classifier is highly suitable for classification task in this study. To avoid large changes in the values of a feature, the values of each individual feature in the training samples are normalized to  $[0, 1]$  by using a linear mapping method (i.e., the min-max normalization method [64]), and the values of the same feature in the testing samples are normalized with the same normalization parameters as the training samples [6]. Considering the classification bias caused by imbalanced training samples, a synthetic minority oversampling technique (SMOTE) [47] is used to balance the training samples before training by oversampling the samples of minority class until the sample number of minority class is equal to that of majority class.

In the stage of polyp candidate detection, all possible candidates are considered. Thus, a true polyp corresponds to one or more TP candidates that possibly include incorrect candidates, such as small fragments. The feature vectors of these incorrect TP candidates negatively influence the classification of Random Forests. Thus, such feature vectors should be removed from the training set before training. In this study, we develop a simple machine learning-based score rank (MLBSR) method to remove the feature vectors of the incorrect TP candidates and redundant FP candidates from the training set: For each round of the fivefold cross-validation used in this study, one fold is used as the testing set, and the remaining four folds are used as the training set [48]. The training set is initially predicted with Random Forests by using a fourfold cross validation. All prediction scores of the training set are collected. Then, all FP candidates corresponding to the training set are sorted in descending order according to their prediction scores. The feature vectors of the top 50% FP candidates are retained in the training set because these difficultly distinguished FP candidates are considered to be informative. For several TP candidates belonging to the same true polyp, only the feature vector of the TP candidate with the highest prediction score is retained in the training set. After the use of MLBSR method, the training set is normalized and then balanced by SMOTE. Random Forests classifier is trained with the training set and evaluated with the testing set [6].

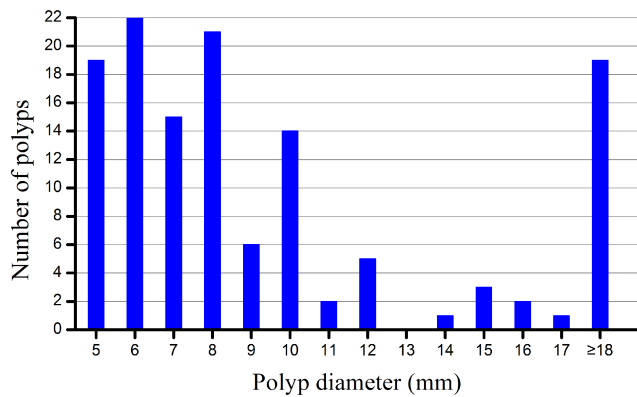


FIGURE 5. Size distribution of polyps.

### III. EXPERIMENTS AND RESULTS

#### A. CTC DATA DESCRIPTION

The oral contrast-enhanced CTC data used in this study included 510 CT scans from 255 patients, which consisted of 195 patients from the Walter Reed Army Medical Center (WRAMC) [49] and 60 patients from The Cancer Imaging Archive (TCIA) [50]. Each patient ingested oral contrast agents and underwent CT scanning on supine and prone positions under the condition of colon insufflation with room air. The WRAMC CTC datasets were acquired through a four-channel or eight-channel CT scanner (GE LightSpeed or LightSpeed Ultra, General Electric Medical Systems) with a table speed of 15 mm/s, 1.25–2.5 mm collimation, 1 mm reconstruction interval, 1.25–2.5 mm slice thickness, 100 mAs, 120 kVp, CT slice size of  $512 \times 512$  pixels, and in-plane pixel spacings ranging from  $0.5293 \text{ mm} \times 0.5293 \text{ mm}$  to  $0.9141 \text{ mm} \times 0.9141 \text{ mm}$  [6]. The TCIA CTC datasets were acquired by using at least a 16-slice CT scanner with 0.5–1.0 mm collimation, 0.98–1.5 pitch, 0.8 mm reconstruction interval, 1–1.25 mm slice thickness, 50 effective mAs, 120 kVp, CT slice size of  $512 \times 512$  pixels, and in-plane pixel spacings ranging from  $0.5859 \text{ mm} \times 0.5859 \text{ mm}$  to  $0.9766 \text{ mm} \times 0.9766 \text{ mm}$  [6], [51]. The OC and CTC confirmed 130 polyps of 5 mm or larger in the 255 patients, of which 102 patients had polyps and 153 had none. These polyps included 15 polyps visible only in either the supine or prone view and 115 polyps visible in both the supine and prone views. A total of 12 flat polyps, 90 sessile polyps, and 28 pedunculated polyps were revealed in the patient datasets, which had 83 polyps ranging from 5 mm to 9 mm and 47 polyps larger than or equal to 10 mm [6]. Fig. 5 shows the polyp size distribution.

#### B. EVALUATION OF THE PROPOSED CAD SYSTEM

An experienced radiologist manually segmented all visible polyps of the CTC datasets in this study. A polyp candidate was counted as a TP if the polyp candidate overlapped a ground truth. Otherwise, the polyp candidate was counted as a FP [3], [6]. A polyp was considered detected if observed on either the supine or prone position [52]. After applying the previous polyp candidate detection method [3], [6], radiomic

features were calculated from each polyp candidate and used for classification. To avoid the overfitting problem in the stage of candidate classification, we used Random Forests classifier to classify polyp candidates and adopted fivefold cross-validation strategy to evaluate the performance of the proposed method. Random Forests classifier constructs each tree using a bootstrap sample from the entire training set and splits each node using the best feature among a subset of features randomly selected without replacement at that node. Thus, Random Forests classifier is robust against overfitting and can train the samples of high-dimensional feature vectors without additional feature selection. In addition, N-fold cross validation completely separates the training set and testing set. It can minimize the evaluation bias and maximize the available datasets, which alleviates overfitting. According to [30] and [53], Random Forests classifier is not very sensitive to the parameter values and can obtain good and stable classification performance if enough trees are constructed. In consideration of the low computational cost and stable classification performance, the parameters of Random Forests classifier in this study were set as follows: the tree number was 100, the sample number for constructing each tree was the sample number of the entire training set, and the number of features in the random feature subset at each node was the square root of the number of all features.

In the fivefold cross validation, all 255 patients were randomly partitioned into five non-intersecting folds. Each fold involved 51 patients. For each cross-validation round, one fold was used as the testing set, and the remaining four folds were used as the training set [3]. After the training set was processed by the MLBSR, min-max normalization, and SMOTE methods, Random Forests classifier was trained with the training set and then used to predict the testing set. Each fold was predicted only once after the fivefold cross validation. All prediction scores of the testing set during cross validation were collected and used for free-response receiver operating characteristics (FROC) analysis. The detection sensitivity and FP rate were quantified as by-polyp sensitivity and the average number of FPs per scan [52], [54], respectively. The normalized area under the FROC curve (nAz) [55] in the range of [0, 5] FPs per scan was used as a performance measure of the FROC curve in the FROC analysis.

#### C. PERFORMANCE ANALYSIS OF THE PROPOSED CAD SYSTEM

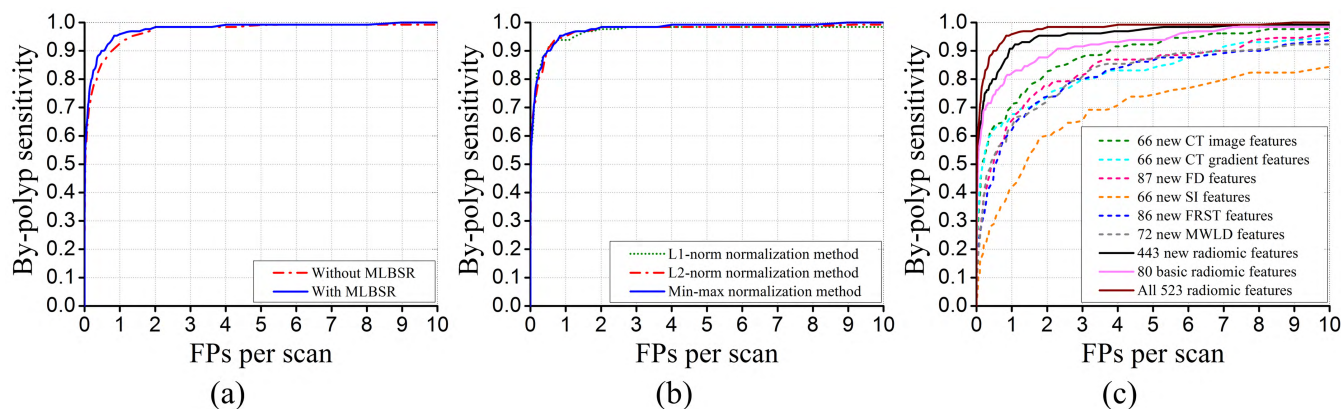
To obtain the enough overall performance, we adjusted the candidate detection method to ensure all polyps detected on either the supine or prone positions in the stage of polyp candidate detection, which produced 23964 FP candidates with two polyp views missed. Then, all 523 radiomic features were calculated from each candidate and sent to Random Forests for classification.

##### 1) ANALYSIS OF THE FP REDUCTION

For Random Forests classifier, the sample number for constructing each tree was set as the sample number of the

**TABLE 1.** The nAz values of Random Forests classifier with different parameters in the testing set during fivefold cross validation. The number of trees is  $n$ , and the number of features in the random feature subset at each node is  $\text{floor}(m * \sqrt{523})$ .

$m \setminus n$	25	50	75	100	125	150	200	300	400
0.125	0.9385	0.9440	0.9579	0.9460	0.9493	0.9493	0.9555	0.9570	0.9537
0.25	0.9459	0.9538	0.9461	0.9506	0.9555	0.9550	0.9523	0.9568	0.9548
0.5	0.9433	0.9567	0.9581	0.9553	0.9568	0.9546	0.9558	0.9577	0.9581
1.0	0.9505	0.9472	0.9542	0.9591	0.9570	0.9574	0.9568	0.9562	0.9577
2.0	0.9441	0.9492	0.9546	0.9558	0.9569	0.9513	0.9584	0.9565	0.9591
4.0	0.9343	0.9373	0.9512	0.9431	0.9528	0.9512	0.9556	0.9525	0.9526
8.0	0.9343	0.9385	0.9263	0.9499	0.9390	0.9371	0.9414	0.9430	0.9468



**FIGURE 6.** FROC curves of (a) the proposed system with/without the MLBSR method, (b) different normalization methods, and (c) different feature groups for all polyps  $\geq 5$  mm in the fivefold cross validation.

entire training set according to [30]. Table 1 shows the nAz values of Random Forests classifier in the stage of candidate classification with the tree number and the feature number of the feature subset randomly selected at each node equal to different values. The nAz values were relatively stable if Random Forests classifier constructed enough trees. However, the more trees were constructed, the higher computational cost was demanded. As shown in Table 1, several parameter settings make Random Forests classifier yield the maximum nAz value of 0.9591, but the parameter setting used in this study (i.e., the number of trees is equal to 100, and the number of features in the random feature subset at each node is equal to the square root of the number of all features) obtains not only the maximum nAz value but also the lowest computational cost.

The MLBSR and SMOTE methods were used to deal with the imbalanced training set during the training procedure of Random Forests classifier. We conducted the following experiments to evaluate the MLBSR method: (1) only the SMOTE method was used, and (2) both the SMOTE and MLBSR methods were used. Fig. 6(a) shows the FROC curves of the proposed system with/without the MLBSR method. After the use of MLBSR method, the nAz value increased from 0.9462 to 0.9591. Thus, the proposed system

achieved better performance when the MLBSR method was used to fine tune the training set. In this study, the values of each individual feature in the training set were linearly normalized to [0, 1] by the min-max normalization method. We compared the min-max normalization method with the L1-norm and L2-norm normalization methods [56] that divided the values of a feature in the training set by using the L1-norm and L2-norm of the values of this feature, respectively. The FROC curves of the proposed system with the three normalization methods are shown in Fig. 6(b). The min-max normalization method achieved the nAz value of 0.9591, whereas the L1-norm and L2-norm normalization methods yielded the nAz values of 0.9538 and 0.9557, respectively. Thus, the performance of the min-max normalization method was best, and the L2-norm normalization method slightly outperformed the L1-norm normalization method.

All 523 radiomic features used in this study included 80 basic radiomic features (i.e., the traditional image features from previous CTC CAD studies) and 443 newly developed radiomic features. The 443 new radiomic features comprised 6 subgroups of new radiomic features (i.e., 66 new CT image features, 66 new CT gradient features, 87 new FD features, 66 new SI features, 86 new FRST



**TABLE 2. The 40 most important radiomic features with high values of the mean Gini importance scores.**

Basic radiomic features	Intensity_mean, Intensity_median, Intensity_skewness, Intensity_mean_top50% <sup>a</sup> , and Intensity_mean_top75% of CT density; Intensity_median of CT gradient; Intensity_mean_top25% of SI;	
New radiomic features	3D CT image features	GLSZM_large-zone high-gray-level emphasis.
	3D CT gradient features	GLSZM_large-zone emphasis and GLSZM_large-zone low-gray-level emphasis.
	3D SI features	SRDM(0)_horizontal-weighted sum <sup>b</sup> and SRDM(1)_horizontal-weighted sum.
	3D FRST features	Intensity_energy, Intensity_variance_top25%, Intensity_variance_top50%, Intensity_variance_top75%, SRDM(0)_horizontal-weighted sum, SRDM(1)_horizontal-weighted sum, SRDM(1)_diagonal-weighted sum, and SRDM(1)_grid-weighted sum.
	3D FD features	Lacunarity, Intensity_minimum, Intensity_range, Intensity_variance, GLSZM_gray-level nonuniformity, GLSZM_large-zone low-gray-level emphasis, NGTDM_strength, GLDM(1)_mean of contrast <sup>c</sup> , GLDM(1)_maximum of contrast, GLDM(1)_minimum of contrast, GLDM(1)_range of contrast, GLDM(1)_variance of contrast, GLDM(2)_range of contrast, and GLDM(2)_variance of contrast.
	3D MWLD features	WLD_radius(1)_dominant differential excitation(1)_dominant orientation(0) <sup>d</sup> , WLD_radius(1)_dominant differential excitation(5)_dominant orientation(1), WLD_radius(2)_dominant differential excitation(0)_dominant orientation(0), WLD_radius(2)_dominant differential excitation(5)_dominant orientation(1), WLD_radius(3)_dominant differential excitation(0)_dominant orientation(0), and WLD_radius(3)_dominant differential excitation(5)_dominant orientation(1).

<sup>a</sup>Intensity\_mean\_top50% means an intensity feature that is the mean of top 50% high intensity values;

<sup>b</sup>SRDM(0)\_horizontal-weighted sum means a SRDM feature that is the horizontal-weighted sum with the threshold value of 0;

<sup>c</sup>GLDM(1)\_mean of contrast means a GLDM feature that is the mean of contrast with the threshold value of 1;

<sup>d</sup>WLD\_radius(1)\_dominant differential excitation(1)\_dominant orientation(0) means a MWLD feature that is a value of the WLD histogram with the radius, dominant differential excitation, and dominant orientation being 1, 1, and 0, respectively.

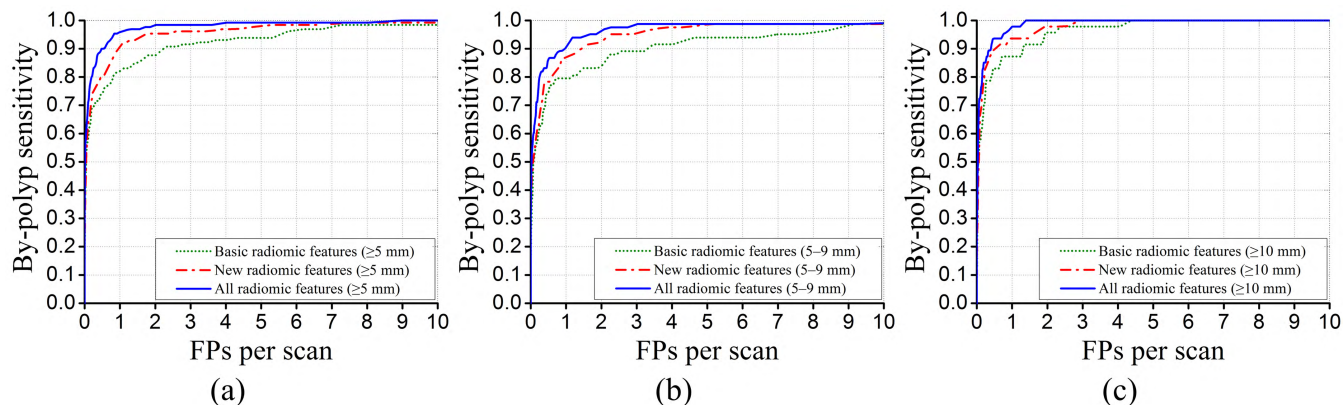
features, and 72 new MWLD features). To analyze the classification performance of the extracted features, different feature groups were fed into Random Forests in the five-fold cross validation for classification of all polyps  $\geq 5$  mm. Fig. 6(c) shows the FROC curves of 9 feature groups (i.e., 6 subgroups of new radiomic features, all 443 new radiomic features, 80 basic radiomic features, and all 523 radiomic features). Among the 6 subgroups of new radiomic features, the 66 new CT image features yielded the maximum nAz value of 0.8033 and performed best, whereas the 66 new SI features obtained the minimum nAz value of 0.5725 and exhibited the poorest performance. The nAz values of the 80 basic radiomic features, 443 new radiomic features, and their combination were 0.8679, 0.9217, and 0.9591, respectively. Therefore, the 80 basic radiomic features outperformed each of the 6 subgroups of new radiomic features. All 443 new radiomic features (i.e., the combination of the 6 subgroups of new radiomic features) performed better than the 80 basic radiomic features, and their combination (i.e., all 523 radiomic features) yielded the best performance for classification of all polyps  $\geq 5$  mm. The classification results indicate that the newly developed radiomic features can characterize colonic polyps and improve the detection performance of the CAD method.

To further demonstrate the effectiveness of the new radiomic features, we computed the mean Gini importance scores [3], [6] of all 523 radiomic features in the fivefold cross validation and then ranked these radiomic features according to their mean Gini importance scores. Table 2 lists the 40 most

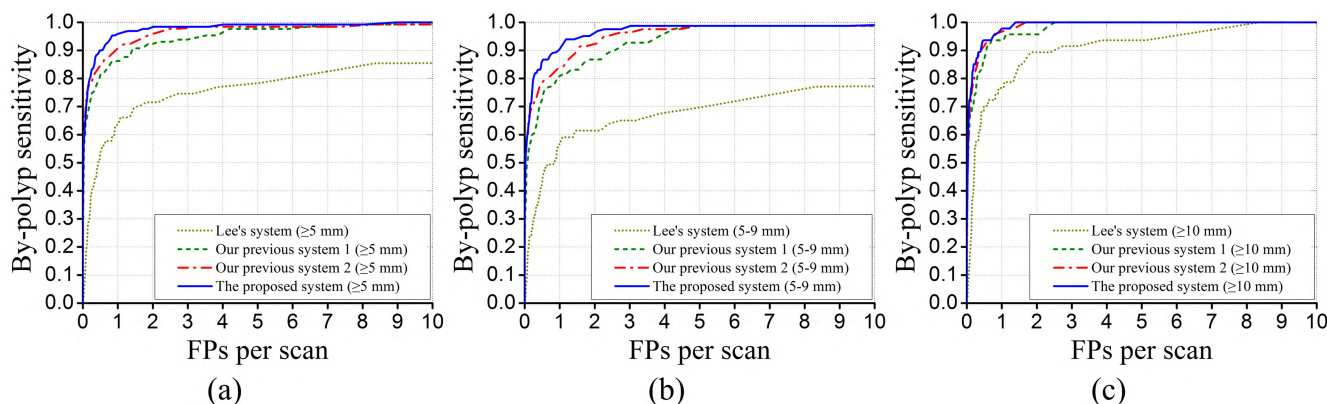
important radiomic features with high values of the mean Gini importance scores, which comprise 7 basic radiomic features (i.e., 5 intensity features of CT image, 1 intensity feature of CT gradient, and 1 intensity feature of SI) and 33 new radiomic features. The 33 most important new features consisted of 1 CT image feature, 2 CT gradient features, 2 SI features, 8 FRST features, 14 FD features, and 6 MWLD features. Thus, the new radiomic features are effective and informative.

## 2) OVERALL PERFORMANCE

The 80 basic radiomic features (i.e., the traditional image features from previous CTC CAD studies), 443 new radiomic features, and their combination (i.e., all 523 radiomic features) were inputted into Random Forests in the fivefold cross validation for detection of different-sized polyps: polyps  $\geq 5$  mm, polyps of 5–9 mm, and polyps  $\geq 10$  mm, respectively. Fig. 7 shows the FROC curves of the 80 basic radiomic features, 443 new radiomic features, and their combination for different polyp sizes. Table 3 tabulates the corresponding detection results. Among the three feature groups, all 523 radiomic features yielded the best classification performance, and the new radiomic features performed better than the basic radiomic features for different polyp sizes. The detection results indicate that the proposed CAD method could achieve a high detection sensitivity with a low FP rate. The FP detections at 98.5% by-polyp sensitivity and 2.0 FPs per scan for polyps  $\geq 5$  mm were counted according to their locations. Among the remaining FPs



**FIGURE 7.** FROC curves of the 80 basic radiomic features, 443 new radiomic features, and their combination (i.e., all 523 radiomic features) for (a) polyps  $\geq 5$  mm, (b) polyps of 5–9 mm, and (c) polyps  $\geq 10$  mm, respectively.



**FIGURE 8.** FROC curves of different CAD systems for (a) polyps  $\geq 5$  mm, (b) polyps of 5–9 mm, and (c) polyps  $\geq 10$  mm, respectively.

**TABLE 3.** Detection results of different feature groups for different polyp sizes.

Feature groups	nAz value		By-polyp sensitivity (FPs/scan)			
	$\geq 5$ mm		5–9 mm		$\geq 10$ mm	
Basic radiomic features	0.8679	87.9% (2.0)	0.8458	89.2% (3.0)	0.9253	87.2% (1.0)
New radiomic features	0.9217	95.4% (2.0)	0.9068	95.6% (3.0)	0.9544	93.6% (1.0)
All radiomic features	0.9591	98.5% (2.0)	0.9408	98.8% (3.0)	0.9750	97.9% (1.0)

after FP reduction, 63.5%, 11.7%, 23.2%, and 1.6% of them were on haustra folds, flat colon walls, ileocecal valves, and rectal tubes, respectively. These FPs were difficultly distinguished from true polyps, as their shapes and internal structures were similar to those of polyps.

**D. DIRECT COMPARISON OF THE PERFORMANCE OF DIFFERENT CAD SYSTEMS**

For a direct comparison between the performance of the proposed CAD system and other CAD systems, Lee’s CAD

system [55], our previous CAD system 1 [3], and our previous CAD system 2 [6] were evaluated in the fivefold cross validation with the same CTC datasets used in this study. Fig. 8 shows the FROC curves of the proposed CAD system, Lee’s CAD system [55], our previous CAD system 1 [3], and our previous CAD system 2 [6] for different polyp sizes. Table 4 shows the corresponding detection performance, which indicates that the proposed CAD system achieved higher sensitivity than the other three CAD systems at the same FP rate for different polyp sizes. Thus, the proposed CAD system outperformed the other three CAD systems for detecting colonic polyps in CTC.

**TABLE 4.** Direct comparison between the performance of the proposed CAD system and the other three CAD systems.

CAD systems	nAz value		By-polyp sensitivity (FPs/scan)			
	$\geq 5$ mm		5–9 mm		$\geq 10$ mm	
Lee's system [55]	0.6831	71.5% (2.0)	0.5953	65.1% (3.0)	0.8380	76.6% (1.0)
Our previous system 1 [3]	0.9064	92.5% (2.0)	0.8671	92.8% (3.0)	0.9548	93.6% (1.0)
Our previous system 2 [6]	0.9398	96.1% (2.0)	0.9099	96.5% (3.0)	0.9710	96.4% (1.0)
The proposed system	0.9591	98.5% (2.0)	0.9408	98.8% (3.0)	0.9750	97.9% (1.0)

**TABLE 5.** Indirect comparison of the performance of different CTC CAD systems.

CAD systems	Method category	# patients for testing	Polyp information	WRAMC database	Evaluation strategy	By-polyp sensitivity	FPs/scan
Summers <i>et al.</i> (2005) [57]	Volume-based conventional method <sup>a</sup>	792	28 polyps $\geq 10$ mm	Yes	Independent test	89.3%	1.1
Nappi <i>et al.</i> (2007) [42]	Volume-based conventional method	32	40 polyps $\geq 6$ mm	Yes	Leave one patient out	95%	3.6
		32	16 polyps $\geq 10$ mm	Yes	Leave one patient out	94%	3.6
Li <i>et al.</i> (2008) [58]	Volume-based conventional method	44	45 polyps 6–9 mm	Yes	Fourfold cross validation	71%	2.7
Summers <i>et al.</i> (2008) [59]	Volume-based conventional method	104	47 polyps $\geq 10$ mm	Yes	Independent test	91.5%	4.8
Wang <i>et al.</i> (2008) [60]	Volume-based conventional method	791	193 polyps 6–9 mm	Yes	Independent test	83%	4.5
Yao <i>et al.</i> (2009) [18]	Volume-based conventional method	792	155 polyps $\geq 6$ mm	Yes	Independent test	95%	1.2
		792	43 polyps $\geq 10$ mm	Yes	Independent test	93%	0.6
Li <i>et al.</i> (2009) [15]	Volume-based conventional method	792	121 polyps $\geq 6$ mm	Yes	Independent test	77.4%	2.9
		792	30 polyps $\geq 10$ mm	Yes	Independent test	91.3%	1.2
Oda <i>et al.</i> (2009) [61]	Volume-based conventional method	104	57 polyps $\geq 6$ mm	Yes	Independent test	91.2%	5.7
Zhu <i>et al.</i> (2010) [54]	Volume-based conventional method	325	125 polyps $\geq 10$ mm	Yes	Independent test	93.1%	1.92
Wang <i>et al.</i> (2010) [62]	Volume-based conventional method	66	96 polyps $\geq 6$ mm	Yes	Leave one patient out	83%	5
Liu <i>et al.</i> (2011) [63]	Volume-based conventional method	791	267 polyps 6–9 mm	Yes	Independent test	62%	5
Roth <i>et al.</i> (2015) [22]	2.5D CNNs	792	173 polyps $\geq 6$ mm	Yes	Independent test	75%	1.5
		792	37 polyps $\geq 10$ mm	Yes	Independent test	95%	1
Proposed method	Volume-based	255	130 polyps $\geq 5$ mm	Yes	Fivefold cross validation	98.5%	2.0
	conventional method	255	83 polyps 5–9 mm	Yes	Fivefold cross validation	98.8%	3.0
		255	47 polyps $\geq 10$ mm	Yes	Fivefold cross validation	97.9%	1.0

<sup>a</sup>Volume-based conventional method means the conventional CTC CAD method that segments 3D volumetric regions as polyp candidates.

## E. INDIRECT COMPARISON OF THE PERFORMANCE OF DIFFERENT CAD SYSTEMS

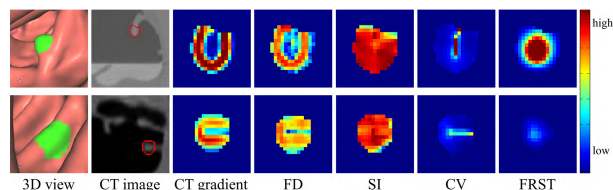
In general, directly comparing the proposed system with other existing systems is difficult given the use of different CTC datasets, different polyp samples, different evaluation strategies, and different evaluation criteria. However, an indirect comparison in similar conditions is still valuable [3]. Table 5 shows the indirect comparison between the performance of the proposed system and other CAD systems with different CTC datasets from the WRAMC database. The proposed system performed comparably to the best among other CTC CAD systems that were validated on the WRAMC datasets.

## IV. DISCUSSION

In this paper, a novel CAD system using radiomics is developed for polyp detection in CTC. The previous shape-based candidate detection method [3], [6] is used to segment polyp

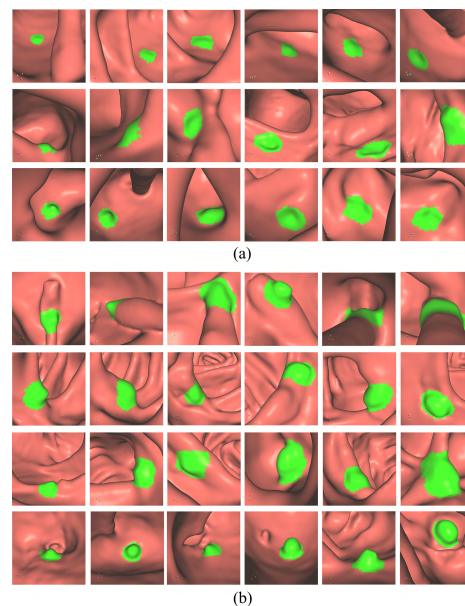
candidates. Then, 523 quantitative radiomic features are computed to characterize each polyp candidate. The generated polyp candidates are finally classified by using Random Forests during the fivefold cross validation. In each round of cross validation, the MLBSR and SMOTE methods are used to process the training set.

The 523 radiomic features include 3D MWLD features and the statistical features (i.e., intensity features, 3D GLSZM features, 3D NGTDM features, 3D GLDM features, and 3D SRDM features) calculated from the following 3D volumetric feature maps: CT image, CT gradient image, FD image, SI image, CV image, and FRST image. The CT image measures the tissue density of the polyp candidates through the CT density values. The CT gradient image reflects the internal structures of the polyp candidates. The SI and CV images describe the convex shape and shape scale of the polyp candidates, respectively. The FD image describes the texture complexity by measuring the local variations of CT density in



**FIGURE 9.** Examples of 3D feature images of the typical TP and FP candidates. Rows 1 and 2 show a TP and FP candidates, respectively. For visual display of the polyp candidate shape, the 3D views of the polyp candidates are presented in column 1. Column 2 shows the original CT images where the red solid curves delineate the polyp candidate boundaries. Columns 3 to 7 show the feature images of CT gradient, FD, SI, CV, and FRST of the polyp candidates with the display window settings of  $145 \pm 145$ ,  $1.5 \pm 1.5$ ,  $0.5 \pm 0.5$ ,  $1 \pm 1$ , and  $2 \pm 2$ , respectively. The 3D views show that the outer boundaries of the TP and FP candidates are semi-spherical and ridge-likeness, respectively. In the CT images, the polyp candidates show similar CT density with the surrounding normal tissues. The outer boundaries of the polyp candidates are distinct, but their inner boundaries not. The CT gradient and FD images show high intensity values in the outer boundary regions of the polyp candidates. The TP candidate exhibits more voxels with high SI values than the FP candidate. For the polyp candidates, the gentle shape changes of their outer isosurfaces lead to low CV values, whereas the sharp shape changes of their inner isosurfaces cause high CV values. Compared with the semi-spherical TP candidate, the ridge-likeness FP candidate has two flat sides that show lower CV values. The FRST image of the TP candidate exhibits higher intensity values than that of the FP candidate.

the neighborhood of each voxel. The FRST image enhances the local structures of high radial symmetry. Fig. 9 shows the examples of 3D feature images of the typical TP and FP candidates. The histogram-based intensity features are used to globally quantify the distribution characteristics of intensity levels within these feature images. After the feature images of CT density, CT gradient, FD, SI, and FRST are quantized, the 3D GLSZM, NGTDM, GLDM, and SRDM texture features are used to quantitatively represent the texture heterogeneity differences and spatial relationships between intensity levels of different voxels in these quantized feature images. The 3D MWLD features are simple and robust local descriptors developed specifically for polyp candidate classification. With respect to the calculation of 3D WLD features for a polyp candidate, we compute the differential excitation component and orientation component for each voxel in the outer boundary region of the polyp candidate. The differential excitation component describes the local salient pattern, and the orientation component shows the angle between the gradient vector and centripetal vector. The two components of all target voxels are used to generate a 2D WLD histogram. All rows of the 2D WLD histogram are concatenated as a 1D feature vector (i.e., 3D WLD features) that characterizes the frequency information of salient patterns within the outer boundary region of the polyp candidate. The 3D WLD features are dense descriptors because they are computed pixelwise. Furthermore, the 3D MWLD features are obtained by concatenating the WLD feature vectors of different scales. Thus, the 3D MWLD features are multigranularity features and robust to rotation and scaling, which can capture salient patterns in the CT images. To summarize, the 523 totally calculated radiomic features can provide the information of



**FIGURE 10.** Several typical TPs and FPs detected by the proposed CTC CAD system at 98.5% by-polyp sensitivity and 2.0 FPs per scan for polyps  $\geq 5$  mm. (a) From the top down, row 1 presents several flat polyps, and rows 2 and 3 reveal several polyps on the haustral folds and flat colon walls, respectively. (b) From the top down, FPs shown in rows 1, 2, 3, and 4 are located on the rectal tubes, ileocecal valves, haustral folds, and colon walls, respectively.

shape, texture, and salient pattern frequency for each polyp candidate.

We have conducted experiments to evaluate the classification performance of the extracted radiomic features. The 523 radiomic features can be divided into 80 basic radiomic features (i.e., the traditional image features from previous CTC CAD studies) and 443 new radiomic features. All 523 radiomic features were sorted in decreasing order according to their mean Gini importance scores. Among the 40 most important radiomic features listed in Table 2, only 7 features were basic radiomic features, and the other 33 features were new radiomic features. The classification performance of the 80 basic radiomic features, 443 new radiomic features, and the combination of these two feature groups for different polyp sizes are shown in Fig. 7 and Table 3. The 443 new radiomic features outperformed the 80 basic radiomic features, and the combination of the two feature groups achieved the best classification performance. Thus, the new radiomic features are informative and can effectively characterize colonic polyps. However, the proposed radiomic features still have the following limitations: 1) the correlation relationship of the proposed radiomic features need to be further studied since there might be redundant information among these features; 2) some polyps of 5 mm or smaller are not sensitive much to the proposed radiomic features because these polyps might be too small to have the texture. Thus, we intend to address these issues and improve the robustness of the proposed features in future work.

To further validate the system performance, we directly compared the proposed system with Lee's CAD system and our previous two systems on the same CTC datasets (Fig. 8). The direct comparison results tabulated in Table 4 indicate that the proposed system performed better than Lee's CAD system and our previous two systems. Fig. 10 shows several typical TPs and FPs detected by the proposed CTC CAD system at 98.5% by-polyp sensitivity and 2.0 FPs per scan for polyps  $\geq 5$  mm. We also made an indirect comparison between the proposed system and other existing CAD systems in the literatures that used different CTC datasets from the WRAMC database (Table 5). Table 5 shows that the proposed system performed comparably to those state-of-the-art CTC CAD systems. The direct/indirect comparison with other CAD systems reflects the excellent overall performance of the proposed system. However, no methods are robust to all types of colonic polyps since polyps vary a lot not only in shape and size but also the texture or something else. Thus, we will improve the adaptiveness of the proposed CTC CAD system for diverse polyps and expect to evaluate it on a larger CTC database in future research.

## V. CONCLUSION

This study presents a novel CTC CAD system with radiomics for polyp detection. After applying our previous candidate detection method to generate polyp candidates, 80 basic radiomic features and 443 new radiomic features are extracted for each polyp candidate and then sent to Random Forests for classification. The MLBSR method is also developed to adjust the training samples. The proposed system achieved high detection sensitivity at low FP rate and performed better than our previously developed CTC CAD systems. We believe that the proposed CTC CAD system has the potential of assisting radiologists in reducing the workload and human errors during colon cancer screening.

## ACKNOWLEDGMENT

The authors would like to acknowledge Dr. R. Choi at the Virtual Colonoscopy Center, WRAMC, Washington, D.C. for supplying the CTC images.

## REFERENCES

- [1] *Cancer Facts & Figures 2018*, American Cancer Society, Atlanta, GA, USA, 2018.
- [2] M. J. O'Brien et al., "The national polyp study: Patient and polyp characteristics associated with high-grade dysplasia in colorectal adenomas," *Gastroenterology*, vol. 98, no. 2, pp. 371–379, 1990.
- [3] Y. Ren, J. Ma, J. Xiong, L. Lu, and J. Zhao, "High-performance CAD-CTC scheme using shape index, multiscale enhancement filters, and radiomic features," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1924–1934, Aug. 2017.
- [4] S. J. Winawer et al., "Colorectal cancer screening: Clinical guidelines and rationale," *Gastroenterology*, vol. 113, no. 4, pp. 594–642, 1997.
- [5] G. Tulum, B. Bolat, and O. Osman, "A CAD of fully automated colonic polyp detection for contrasted and non-contrasted CT scans," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 4, pp. 627–644, 2017.
- [6] Y. Ren, J. Ma, J. Xiong, Y. Chen, L. Lu, and J. Zhao, "Improved false positive reduction by novel morphological features for computer-aided polyp detection in CT colonography," *IEEE J. Biomed. Health Inform.*, to be published, doi: [10.1109/JBHI.2018.2808199](https://doi.org/10.1109/JBHI.2018.2808199)
- [7] D. J. Vining, Y. Ge, D. K. Ahn, and D. R. Stelts, "Virtual colonoscopy with computer-assisted polyp detection," in *Proc. 1st Int. Workshop Comput.-Aided Diagnosis Comput.-Aided Diagnosis Med. Imag.*, K. Doi, H. MacMahon, M. L. Giger, and K. R. Hoffmann, Eds. Amsterdam, The Netherlands: Elsevier, 1999, pp. 445–452.
- [8] R. M. Summers, C. D. Johnson, L. M. Pusnik, J. D. Malley, A. M. Youssef, and J. E. Reed, "Automated polyp detection at CT colonography: Feasibility assessment in a human population," *Radiology*, vol. 219, no. 1, pp. 51–59, 2001.
- [9] H. Yoshida and J. Nappi, "Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps," *IEEE Trans. Med. Imag.*, vol. 20, no. 12, pp. 1261–1274, Dec. 2001.
- [10] J. Nappi and H. Yoshida, "Feature-guided analysis for reduction of false positives in CAD of polyps for computed tomographic colonography," *Med. Phys.*, vol. 30, no. 7, pp. 1592–1601, 2003.
- [11] J. J. Nappi, H. Frimmel, A. H. Dachman, and H. Yoshida, "Computerized detection of colorectal masses in CT colonography based on fuzzy merging and wall-thickening analysis," *Med. Phys.*, vol. 31, no. 4, pp. 860–872, 2004.
- [12] A. K. Jerebko, J. D. Malley, M. Franaszek, and R. M. Summers, "Support vector machines committee classification method for computer-aided polyp detection in CT colonography," *Acad. Radiol.*, vol. 12, no. 4, pp. 479–486, 2005.
- [13] W. Hong, F. Qiu, and A. Kaufman, "A pipeline for computer aided polyp detection," *IEEE Trans. Vis. Comput. Graphics*, vol. 12, no. 5, pp. 861–868, Sep./Oct. 2006.
- [14] E. Konukoglu and B. Acar, "HDF: Heat diffusion fields for polyp detection in CT colonography," *Signal Process.*, vol. 87, no. 10, pp. 2407–2416, 2007.
- [15] J. Li et al., "Optimizing computer-aided colonic polyp detection for CT colonography by evolving the Pareto front," *Med. Phys.*, vol. 36, no. 1, pp. 201–212, 2009.
- [16] K. Suzuki, H. Yoshida, J. Nappi, S. G. Armato, III, and A. H. Dachman, "Mixture of expert 3D massive-training ANNs for reduction of multiple types of false positives in CAD for detection of polyps in CT colonography," *Med. Phys.*, vol. 35, no. 2, pp. 694–703, 2008.
- [17] N. Kilic, O. N. Ucan, and O. Osman, "Colonic polyp detection in CT colonography with fuzzy rule based 3D template matching," *J. Med. Syst.*, vol. 33, no. 1, pp. 9–18, 2009.
- [18] J. Yao, J. Li, and R. M. Summers, "Employing topographical height map in colonic polyp measurement and false positive reduction," *Pattern Recognit.*, vol. 42, no. 6, pp. 1029–1040, 2009.
- [19] V. F. V. Ravesteijn et al., "computer-aided detection of polyps in CT colonography using logistic regression," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 120–131, Jan. 2010.
- [20] H. Zhu, Y. Fan, H. Lu, and Z. Liang, "Improved curvature estimation for computer-aided detection of colonic polyps in CT colonography," *Acad. Radiol.*, vol. 18, no. 8, pp. 1024–1034, 2011.
- [21] H. Wang et al., "An adaptive paradigm for computer-aided detection of colonic polyps," *Phys. Med. Biol.*, vol. 60, no. 18, pp. 7207–7228, 2015.
- [22] H. R. Roth et al., "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1170–1181, May 2016.
- [23] J. J. Nappi, T. Hironaka, D. Regge, and H. Yoshida, "Deep transfer learning of virtual endoluminal views for the detection of polyps in CT colonography," *Proc. SPIE*, vol. 9785, pp. 97852B-1–97852B-8, Mar. 2016, doi: [10.1117/12.2217260](https://doi.org/10.1117/12.2217260).
- [24] H. J. W. L. Aerts et al., "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Commun.*, vol. 5, Aug. 2014, Art. no. 4006.
- [25] R. Lambin et al., "Radiomics: Extracting more information from medical images using advanced feature analysis," *Eur. J. Cancer*, vol. 48, no. 4, pp. 441–446, 2012.
- [26] R. Gillies, P. Kinahan, and H. Hricak, "Radiomics: Images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2016.
- [27] X. Chen et al., "A radiomics signature in preoperative predicting degree of tumor differentiation in patients with non-small cell lung cancer," *Acad. Radiol.*, to be published, doi: [10.1016/j.acra.2018.02.019](https://doi.org/10.1016/j.acra.2018.02.019).
- [28] F. Khalvati, A. Wong, and M. A. Haider, "Automated prostate cancer detection via comprehensive multi-parametric magnetic resonance imaging texture feature models," *BMC Med. Imag.*, vol. 15, no. 1, pp. 14–27, 2015.

- [29] N. Maforo, H. Li, W. Weiss, L. Lan, and M. Giger, "SU-D-BRA-02: Radiomics of multi-parametric breast MRI in breast cancer diagnosis: A quantitative investigation of diffusion weighted imaging, dynamic contrast-enhanced, and T2-weighted magnetic resonance imaging," *Med. Phys.*, vol. 42, no. 6, p. 3213, 2015.
- [30] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [31] G. Thibault, J. Angulo, and F. Meyer, "Advanced statistical matrices for texture characterization: Application to cell classification," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 3, pp. 630–637, Mar. 2014.
- [32] M. Amadasun and R. King, "Textural features corresponding to textural properties," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, no. 5, pp. 1264–1274, Sep. 1989.
- [33] J. S. Weszka, C. R. Dyer, and A. Rosenfeld, "A comparative study of texture measures for terrain classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, no. 4, pp. 269–285, Apr. 1976.
- [34] J. K. Kim, J. M. Park, K. S. Song, and H. W. Park, "Detection of clustered microcalcifications on mammograms using surrounding region dependence method and artificial neural network," *J. VLSI Signal Process. Syst.*, vol. 18, no. 3, pp. 251–262, 1998.
- [35] Z. Wang *et al.*, "An improved electronic colon cleansing method for detection of colonic polyps by virtual colonoscopy," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 8, pp. 1635–1646, Aug. 2006.
- [36] Q. Li, S. Sone, and K. Doi, "Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans," *Med. Phys.*, vol. 30, no. 8, pp. 2040–2051, 2003.
- [37] D. G. Kang, D. C. Suh, and J. B. Ra, "Three-dimensional blood vessel quantification via centerline deformation," *IEEE Trans. Med. Imag.*, vol. 28, no. 3, pp. 405–414, Mar. 2009.
- [38] M. Vallières, C. R. Freeman, S. R. Skamene, and I. El Naqa, "A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities," *Phys. Med. Biol.*, vol. 60, no. 14, pp. 5471–5496, 2015.
- [39] Y.-Q. Huang *et al.*, "Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer," *J. Clin. Oncol.*, vol. 34, no. 18, pp. 2157–2164, 2016.
- [40] Z. Li, Y. Wang, J. Yu, Y. Guo, and Q. Zhang, "Age groups related glioblastoma study based on radiomics approach," *Comput. Assist. Surg.*, vol. 22, no. 1, pp. 18–25, 2017.
- [41] J. Näppi and H. Yoshida, "Automated detection of polyps with CT colonography: Evaluation of volumetric features for reduction of false-positive findings," *Acad. Radiol.*, vol. 9, no. 4, pp. 386–397, 2002.
- [42] J. Näppi and H. Yoshida, "Fully automated three-dimensional detection of polyps in fecal-tagging CT colonography," *Acad. Radiol.*, vol. 14, no. 3, pp. 287–300, 2007.
- [43] O. S. Al-Kadi and D. Watson, "Texture analysis of aggressive and nonaggressive lung tumor CE CT images," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 7, pp. 1822–1830, Jul. 2008.
- [44] G. Loy and A. Zelinsky, "Fast radial symmetry for detecting points of interest," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 959–973, Aug. 2003.
- [45] J. Chen *et al.*, "WLD: A robust local image descriptor," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1705–1720, Sep. 2010.
- [46] S. Li, D. Gong, and Y. Yuan, "Face recognition using Weber local descriptors," *Neurocomputing*, vol. 122, pp. 272–283, Dec. 2013.
- [47] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.
- [48] J. Cheng *et al.*, "Retrieval of brain tumors by adaptive spatial pooling and Fisher vector representation," *PLoS ONE*, vol. 11, no. 6, p. e0157112, 2016.
- [49] P. J. Pickhardt *et al.*, "Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults," *New England J. Med.*, vol. 349, no. 23, pp. 2191–2200, 2003.
- [50] K. Clark *et al.*, "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," *J. Digit. Imag.*, vol. 26, no. 6, pp. 1045–1057, 2013.
- [51] C. D. Johnson *et al.*, "Accuracy of CT colonography for detection of large adenomas and cancers," *New England J. Med.*, vol. 359, no. 12, pp. 1207–1217, 2008.
- [52] C. van Wijk, V. F. van Ravesteijn, F. M. Vos, and L. J. van Vliet, "Detection and segmentation of colonic polyps on implicit isosurfaces by second principal curvature flow," *IEEE Trans. Med. Imag.*, vol. 29, no. 3, pp. 688–698, Mar. 2010.
- [53] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
- [54] H. Zhu *et al.*, "Increasing computer-aided detection specificity by projection features for CT colonography," *Med. Phys.*, vol. 37, no. 4, pp. 1468–1481, 2010.
- [55] J.-G. Lee, J. H. Kim, S. H. Kim, H. S. Park, and B. I. Choi, "A straightforward approach to computer-aided polyp detection using a polyp-specific volumetric feature in CT colonography," *Comput. Biol. Med.*, vol. 41, no. 9, pp. 790–801, 2011.
- [56] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Diego, CA, USA, Jun. 2005, pp. 886–893.
- [57] R. M. Summers *et al.*, "Computed tomographic virtual colonoscopy computer-aided polyp detection in a screening population," *Gastroenterology*, vol. 129, no. 6, pp. 1832–1844, 2005.
- [58] J. Li *et al.*, "Wavelet method for CT colonography computer-aided polyp detection," *Med. Phys.*, vol. 35, no. 8, pp. 3527–3538, 2008.
- [59] R. M. Summers *et al.*, "Performance of a previously validated CT colonography computer-aided detection system in a new patient population," *Amer. J. Roentgenol.*, vol. 191, no. 1, pp. 168–174, 2008.
- [60] S. Wang, J. Yao, and R. M. Summers, "Improved classifier for computer-aided polyp detection in CT Colonography by nonlinear dimensionality reduction," *Med. Phys.*, vol. 35, no. 4, pp. 1377–1386, 2008.
- [61] M. Oda *et al.*, "Digital bowel cleansing free colonic polyp detection method for fecal tagging CT colonography," *Acad. Radiol.*, vol. 16, no. 4, pp. 486–494, 2009.
- [62] S. Wang, J. Yao, N. Petrick, and R. M. Summers, "Combining statistical and geometric features for colonic polyp detection in CTC based on multiple kernel learning," *Int. J. Comput. Intell. Appl.*, vol. 9, no. 1, pp. 1–15, 2010.
- [63] J. Liu, S. Kabadi, R. Van Uitert, N. Petrick, R. Deriche, and R. M. Summers, "Improved computer-aided detection of small polyps in CT colonography using interpolation for curvature estimation," *Med. Phys.*, vol. 38, no. 7, pp. 4276–4284, 2011.
- [64] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann, 2011.



**YACHENG REN** received the B.E. degree from the Department of Biomedical Engineering, Huazhong University of Science and Technology, Wuhan, China, in 2013. He is currently pursuing the Ph.D. degree with the Department of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China. His research interests include medical image processing and visualization, computer-aided detection and diagnosis, and machine learning.



**YIZHI CHEN** received the B.E. degree from the Department of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2016, where he is currently pursuing the master's degree with the Department of Biomedical Engineering. His research interests include computer-aided diagnosis and data visualization of medical image.



**LIN LU** received the Ph.D. degree from the Department of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2013. He is currently an Associate Research Scientist with the Department of Radiology, Columbia University Medical Center. He has published over 30 research papers in a number of outstanding journals, such as the *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, *Medical Physics*, the *Journal of Proteome Research*, and the *Journal of Computational Chemistry*. His research interests include image processing, data mining, and machine learning.



**JUN ZHAO** received the Ph.D. degree from the Department of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2006. He was a Visiting Scholar with the University of Iowa, USA, from 2006 to 2007. He is currently a Full Professor and the Associate Dean of the School of Biomedical Engineering, Shanghai Jiao Tong University. He is also active in the international research groups of the IEEE EMBC, SPIE, and ISBI. He has published over 130 research papers in international conferences and journals. His research interests include biomedical imaging, medical image processing, computer-aided detection, and medical applications of synchrotron radiation. He received over 10 grants from the Ministry of Science and Technology of China, the National Science Foundation of China, the Science and Technology Commission, the Education Commission of Shanghai Municipality, GE. He is recognized by a number of awards for academic and educational achievements. He serves as an Associate Editor for *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, and the *International Journal of Biomedical Imaging*.

...