

Massive-training support vector regression and Gaussian process for false-positive reduction in computer-aided detection of polyps in CT colonography

Jian-Wu Xu^{a)} and Kenji Suzuki

Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, Chicago, Illinois 60637

(Received 10 November 2010; revised 8 February 2011; accepted for publication 15 February 2011; published 11 March 2011)

Purpose: A massive-training artificial neural network (MTANN) has been developed for the reduction of false positives (FPs) in computer-aided detection (CADE) of polyps in CT colonography (CTC). A major limitation of the MTANN is the long training time. To address this issue, the authors investigated the feasibility of two state-of-the-art regression models, namely, support vector regression (SVR) and Gaussian process regression (GPR) models, in the massive-training framework and developed massive-training SVR (MTSVR) and massive-training GPR (MTGPR) for the reduction of FPs in CADE of polyps.

Methods: The authors applied SVR and GPR as volume-processing techniques in the distinction of polyps from FP detections in a CTC CADE scheme. Unlike artificial neural networks (ANNs), both SVR and GPR are memory-based methods that store a part of or the entire training data for testing. Therefore, their training is generally fast and they are able to improve the efficiency of the massive-training methodology. Rooted in a maximum margin property, SVR offers excellent generalization ability and robustness to outliers. On the other hand, GPR approaches nonlinear regression from a Bayesian perspective, which produces both the optimal estimated function and the covariance associated with the estimation. Therefore, both SVR and GPR, as the state-of-the-art nonlinear regression models, are able to offer a performance comparable or potentially superior to that of ANN, with highly efficient training. Both MTSVR and MTGPR were trained directly with voxel values from CTC images. A 3D scoring method based on a 3D Gaussian weighting function was applied to the outputs of MTSVR and MTGPR for distinction between polyps and nonpolyps. To test the performance of the proposed models, the authors compared them to the original MTANN in the distinction between actual polyps and various types of FPs in terms of training time reduction and FP reduction performance. The authors' CTC database consisted of 240 CTC data sets obtained from 120 patients in the supine and prone positions. The training set consisted of 27 patients, 10 of which had 10 polyps. The authors selected 10 nonpolyps (i.e., FP sources) from the training set. These ten polyps and ten nonpolyps were used for training the proposed models. The testing set consisted of 93 patients, including 19 polyps in 7 patients and 86 negative patients with 474 FPs produced by an original CADE scheme.

Results: With the MTSVR, the training time was reduced by a factor of 190, while a FP reduction performance [by-polyp sensitivity of 94.7% (18/19) with 2.5 (230/93) FPs/patient] comparable to that of the original MTANN [the same sensitivity with 2.6 (244/93) FPs/patient] was achieved. The classification performance in terms of the area under the receiver-operating-characteristic curve value of the MTGPR (0.82) was statistically significantly higher than that of the original MTANN (0.77), with a two-sided p -value of 0.03. The MTGPR yielded a 94.7% (18/19) by-polyp sensitivity at a FP rate of 2.5 (235/93) per patient and reduced the training time by a factor of 1.3.

Conclusions: Both MTSVR and MTGPR improve the efficiency of the training in the massive-training framework while maintaining a comparable performance. © 2011 American Association of Physicists in Medicine. [DOI: [10.1118/1.3562898](https://doi.org/10.1118/1.3562898)]

Key words: computer-aided detection, colonic polyps, support vector regression, Gaussian process regression, false-positive reduction

I. INTRODUCTION

Colorectal cancer is the second leading cause of mortality due to cancer in the United States.¹ Evidence has shown that the risk of colon cancer death could be reduced with early detection and removal of the colonic polyps.² Fiber optic (or

optical) colonoscopy is considered the gold standard diagnostic test as it offers direct biopsy or removal of suspicious colonic polyps.² However, optical colonoscopy is invasive, i.e., it has risks of complications such as perforation, it is expensive, and it has a long examination time and high patient discomfort. Therefore, medical centers are seeking al-

ternative techniques as population screening tools. CT colonography (CTC), also known as virtual colonoscopy, has been proposed as an alternative, less invasive technique for detecting colorectal neoplasms,^{3–6} which has less examination time and less patient discomfort. However, the sensitivity of CTC can be lower for inexperienced readers because there is a long learning curve with CTC reading. This limitation begs for a computer-aided detection (CADe) approach as “a second reader” to assist radiologists in detecting polyps from CTC images.⁷

There has been great interest in the development of automated or semiautomated CADe schemes for the detection of polyps in CTC in the past decade.^{8–12} A CADe scheme for polyp detection is typically composed of candidate detection followed by supervised classification. The task of candidate detection is to achieve high sensitivity in detecting polyps by including as many suspicious lesions as possible. After the polyp candidate detection stage, feature extraction and analysis are performed on the objects detected in CTC. Based on these features, various classifiers have been applied that classify the candidates into polyps and nonpolyps so that FP detections can be reduced while a high level of sensitivity is maintained. Linear and quadratic discriminant analysis were used by Yoshida *et al.*,⁹ as well as by Jerebko *et al.*,¹³ as simple and effective classifiers. Acar *et al.*¹⁴ also applied a linear classifier based on edge-displacement field features. Gokturk *et al.*¹⁰ employed a support vector machine (SVM) to distinguish between polyps and normal tissue. To improve the discriminant ability of SVMs, a committee of SVMs has been proposed to take advantage of combining multiple classifiers.¹⁵ Another popular classifier is the artificial neural network (ANN).¹⁶ Logistic regression has also been employed for reducing false positive (FP) detections where features were ordered according to their relevance.¹⁷ Yao *et al.*¹⁸ employed a topologic height map for FP reduction. Zhu *et al.*¹⁹ developed two-dimensional projection features for distinction between FP and true positive (TP) detections. In summary, all these proposed classifiers operated on extracted geometric, texture, morphologic, and other features from segmented polyp candidates in CTC images. However, the extracted features might be noisy (with errors) due to CTC image reconstruction error, segmentation errors, and other factors. Moreover, it requires not only domain knowledge to design the set of features to be extracted, but also advanced feature selection methods for choosing the most discriminant ones.

Recently, Suzuki *et al.*²⁰ presented a different approach to the reduction of FP detections of polyps, in which an ANN was used as a regression technique instead of a classifier. The inputs to the ANN regression model were voxel values of CTC images rather than computed features from segmented polyp candidates. The ANN was trained with a massive number of subvolumes extracted from 3D CTC volumes together with “teaching” volumes containing the distribution for the “likelihood of being a polyp,” therefore termed *massive-training ANN* (MTANN).^{21,22} As a nonlinear regression technique, MTANN is able to learn to differentiate between the underlying structures of polyps and nonpolyp regions. There-

fore, the trained MTANN is able to enhance polyps and suppress nonpolyps so that the score for a polyp is higher than that of a nonpolyp. The promising performance of the MTANN has been demonstrated in the reduction of FP detections in CTC CADe,^{20,23} computerized detection of lung nodules in low-dose CT,^{21,24–26} and CADe for detecting nodules in chest radiographs.²⁷ However, the computational cost of the training of an MTANN is very high, given the large number of training samples extracted from 3D CTC images. For example, the training of a 3D MTANN with ten polyps and ten FPs took 38 h on a personal computer (Intel, Xeon, 2.7GHz).²⁰ The training time increases much more when a mixture of expert MTANNs is used to reduce a large variety of FPs. It took 244 h to train a mixture of six MTANNs.^{20,23} This drawback hinders the development of a CADe scheme. Recently, a dimension reduction technique based on a Laplacian eigenmap has been used in the MTANN framework to reduce the training time by a factor of 9.5 while maintaining a comparable performance.²⁸

In this study, we investigated the feasibility of two state-of-the-art nonlinear regression techniques, namely, support vector regression (SVR) and Gaussian process regression (GPR), as alternatives to improve the efficiency of training of the massive-training framework for reducing FPs in the computerized detection of polyps in CTC. Unlike ANNs, both SVR and GPR are memory-based methods that store a part of or the entire training data for testing. Therefore, they are generally fast to train and are able to improve the efficiency of the massive-training methodology. Moreover, both SVR and GPR are kernel-based nonlinear regression techniques, where either a kernel or a covariance function is used for implicitly transforming the original image data into a high-dimensional reproducing kernel Hilbert space (RKHS). The transformation is able to capture the inherent nonlinearity underlying the CTC images by enhancing polyps and suppressing nonpolyp objects. Rooted in a maximum margin property, SVR offers excellent generalization ability and robustness to outliers. On the other hand, GPR approaches nonlinear regression from a Bayesian perspective. The Bayesian paradigm provides probabilistic modeling of nonlinear regression. The Bayesian approach to regression specifies *a priori* probability of the parameters to be estimated and it computes the maximum *a posteriori* probability given the observed data samples. Contrary to non-Bayesian schemes where a single parameter is typically chosen by some criterion, the Bayesian probabilistic model produces both the optimal estimated function and the covariance associated with the estimation. Therefore, the Bayesian paradigm offers more information on the estimated parameters than does the non-Bayesian methodology. These two methods have been successfully applied to various regression problems.^{29–32} In this study, we applied SVR and GPR as volume-processing techniques in the distinction of polyps from FP detections in a CTC CADe scheme. Both methods were trained directly with voxel values from CTC images. A 3D scoring method based on a 3D Gaussian weighting function was applied to the outputs of massive-training SVR (MTSVR) and massive-training GPR (MTGPR) for distinction between polyps and

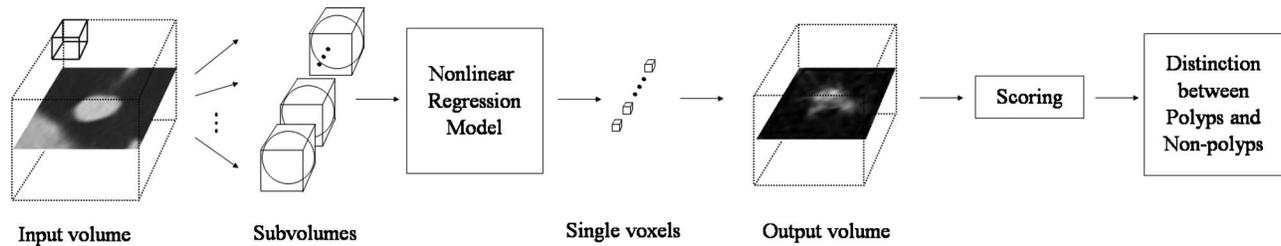


FIG. 1. General framework of massive-training nonlinear regression for distinction between polyps and nonpolyps.

nonpolyps. We tested MTSVR and MTGPR in terms of the training time reduction and FP reduction performance. The novelty of the paper is to provide two alternatives to the core of the massive-training framework, namely, SVR and GPR, other than dimension reduction techniques, to improve the efficiency of the training while maintaining a comparable performance in reducing FPs in the computerized detection of polyps in CTC.

II. MATERIALS AND METHODS

In this section, we depict the database used in this study and the performance of a previously reported CADe scheme.⁹ We also present the general massive-training framework of a nonlinear regression technique as a classifier system to differentiate polyps from FP detections. We provide the technical background of SVR and GPR. As nonlinear regression models, both SVR and GPR fit well into the massive-training framework. One main contribution in this paper is to provide SVR and GPR as alternatives to ANN regression in the massive-training framework to improve the computational efficiency in the developmental stage.

II.A. CTC database

We retrospectively collected CTC cases used in this study. The database was acquired at the University of Chicago Medical Center. It consisted of 240 CTC data sets obtained from 120 patients. Each patient followed the standard CTC procedure with precolonoscopy cleansing and colon insufflations with room air or carbon dioxide. Oral contrast was not administered. Both supine and prone positions were scanned with a multidetector-row CT scanner (LightSpeed QX/i, GE Medical Systems, Milwaukee, WI) with collimations between 2.5 and 5.0 mm, reconstruction intervals of 1.0–3.0 mm, and tube currents of 60–120 mA with 120 kVp. The detailed reconstruction intervals were as follows: 1 mm (2 patients), 1.25 mm (2 patients, 7 polyps), 1.5 mm (64 patients, 12 polyps), 2.5 mm (51 patients, 10 polyps), and 3 mm (1 patient). Each reconstructed CT section had a matrix size of 512×512 pixels, with an in-plane pixel size of 0.5–0.7 mm. Optical colonoscopy was also performed for all patients. In this study, we used 5 mm as the lower limit on the clinically important size of polyps. The locations of polyps were confirmed by an expert radiologist based on CTC images and on pathology and colonoscopy reports. A total of 17 patients had 29 colonoscopy-confirmed polyps, 15 of which were 5–9 mm and 14 were 10–25 mm in size. The shapes of

the polyps were pedunculated and sessile (i.e., there were no flat lesions; refer to Ref. 33 for the definition). The whole database was divided into a training set and a testing set. The training set consisted of 27 patients, 10 of which had 10 polyps. We selected ten nonpolyps (i.e., FP sources) from the training set. These ten polyps and ten nonpolyps were used to train the proposed models. The testing set contained 93 patients, including 19 polyps in 7 patients and 86 negative patients. The formulation of CTC images into data patterns as input to the nonlinear regression models is illustrated in Section II B.

An initial CADe scheme for detection of polyps in CTC was applied to the database. The CADe scheme is composed of (1) colon segmentation based on centerline tracing,³⁴ (2) detection of polyp candidates based on shape index and curvedness of the segmented colon,⁹ (3) calculation of 3D pattern features of the polyp candidates,^{7,34,35} and (4) classification of the polyp candidates as polyps or nonpolyps based on quadratic discriminant analysis. The initial CADe scheme yielded a 94.7% (18/19) by-polyp sensitivity with 5.1 (474/93) FPs per patient for the testing set. The major sources of FPs included rectal tubes, stool, haustral folds, colonic walls, and the ileocecal valve.

II.B. General framework of massive-training nonlinear regression

The basic idea of using nonlinear regression techniques to distinguish polyps from nonpolyp objects is to learn the distinctive underlying image structures of different classes. Unlike a classifier that is based on features extracted from segmented objects, a regression model uses individual voxel values as input. It is thus able to differentiate subtle characteristics of different classes in a local scale. For example, the shape index has been proposed as a feature to detect polyps and also retained for classification.⁹ However, the shape index values for a polyp and part of a rectal tube would be very close because both have caplike shapes. Therefore, a classifier failed to separate rectal tubes from polyps based on the shape index.²⁰ However, the underlying image structures are very different. Rectal tubes are hollow in the center, whereas polyps are solid.

The general framework of massive-training nonlinear regression is illustrated in Fig. 1. The input volumes are 3D CTC images. Usually, the pixel size within a CT image is different from the reconstruction interval across CT sections. Moreover, the reconstruction intervals might vary across dif-

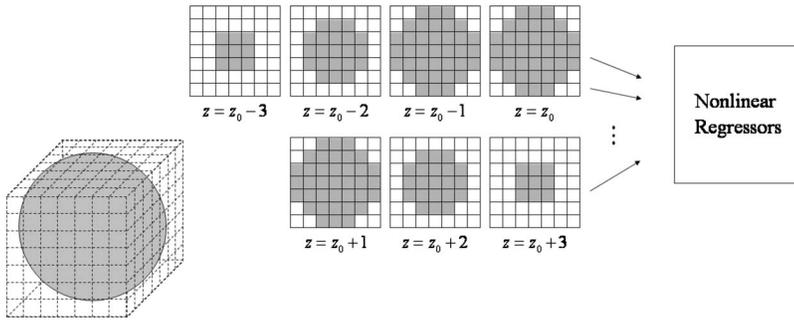


FIG. 2. The spherical-input subvolume and the slice-by-slice representation of the digital quasisphere in a $7 \times 7 \times 7$ voxel cube. Each 7×7 square stands for a certain image slice in the subvolume, where z_0 is the middle slice. The input voxels to the nonlinear regression models are those gray squares in each matrix.

ferent institutions under different protocols. In order to mitigate such variations, we converted the original CTC images into isotropic volume data. The voxel values were normalized to a range between 0 and 1, where 1 corresponded to 1000 Hounsfield unit (HU) and 0 was matched to -1000 HU. Each input volume contained 64 CT slices and each image slice is of 64×64 pixel size. The center of input volume is the detected location of suspicious polyps from the original CADE scheme. If we input each volume directly to the nonlinear regression models, the dimension is prohibitively large ($64 \times 64 \times 64 = 262\,144$). Hence, it requires huge data samples to train a large number of parameters in the regression model, which is not feasible. Therefore, we divide the input volumes into multiple subvolumes. Each subvolume is a cube of $7 \times 7 \times 7$ dimension. We scan the entire input volume with the subvolume voxel by voxel. The other purpose for using the subvolumes as input is to learn different local structures of polyps and nonpolyps so that the regression model is able to differentiate between the two. Because the average shape of polyps is close to a sphere, we can further reduce the number of input voxels for the nonlinear regression model. Figure 2 demonstrates the scheme for extracting a quasisphere from a subvolume cube. The number of voxels in the digital quasisphere is 171, compared to 343 in the original subvolume cube. The gray square in each matrix is the input voxel to the nonlinear regression model. Therefore, the computational cost is reduced dramatically while still preserving the essential image information of polyps.

In this study, we investigated the feasibility of two state-of-the-art nonlinear regression techniques, namely, SVR and GPR, in the massive-training framework. SVR offers robust and nonlinear regression. By employing nonlinear kernel functions, SVR is able to produce a linear combination of functionals in the reproducing kernel Hilbert space induced by the non-negative definite kernel functions. While mapping back to the original image space, the output functional becomes a highly nonlinear function that is able to capture the underlying nonlinear structure of CTC images. On the other hand, GPR is a different approach to nonlinear regression based on Bayesian methodology. GPR is a nonparametric model that places a prior on the nonlinear function directly without parametrization. The prediction based on the learned function is obtained by maximum likelihood of the posterior. We briefly describe SVR and GPR in Secs. II B 1 and II B 2, respectively.

Figure 1 illustrates the nonlinear regression process in the massive-training framework. The inputs to the nonlinear regression model are the voxel values in the quasispherical subvolume V_S . The output of the nonlinear regression model is a continuous scalar value that corresponds to the center voxel in the subvolume, which is defined as

$$O(x, y, z) = \text{NLR}\{I(x-p, y-q, z-r) | (p, q, r) \in V_S\}, \quad (1)$$

where $I(x-p, y-q, z-r)$ is the normalized input voxel to the nonlinear regression model; x , y , and z are the global coordinates; p , q , and r are local coordinates; and $\text{NLR}\{\cdot\}$ is the output of the nonlinear regression model, i.e., SVR or GPR. Therefore, one subvolume V_S corresponds to one output voxel as shown in the left and right hand sides of the nonlinear regression model in Fig. 1. By scanning the entire input CTC volume of $64 \times 64 \times 64$ voxels in size with the input kernel of the regression model voxel by voxel, all output voxels can be obtained. The entire output volume is formed by putting all the output voxels together according to their global coordinates. The output volume is of size $58 \times 58 \times 58$ voxels. The size of the output volume is off by six voxels because scanning is started at three voxels inside the corner of the input volume with the $7 \times 7 \times 7$ input kernel. The output volume is then subject to the scoring method.

The nonlinear regression model enhances polyps and suppresses nonpolyp objects in the output 3D images. In order to distinguish between the two classes, we present a 3D scoring method that translates the output volume into a single scalar value. The score is defined as

$$S = \sum_{(x, y, z) \in V_E} f_G(x, y, z; \sigma_w) \times O(x, y, z). \quad (2)$$

V_E is a volume for evaluation that is large enough to cover a polyp or a nonpolyp object. The criterion to choose a suitable V_E is determined by the standard deviation of a Gaussian weighting function in the scoring method given in Eq. (3). We chose a volume of $21 \times 21 \times 21$ voxels for V_E because it is large enough to cover the Gaussian weighting function. $O(x, y, z)$ is the output voxel value from the trained nonlinear regression model and $f_G(x, y, z; \sigma_w)$ is a 3D Gaussian weighting function with standard deviation σ_w , which is described as

$$f_G(x, y, z; \sigma_w) = \frac{1}{\sqrt{2\pi\sigma_w}} e^{-(x^2+y^2+z^2/2\sigma_w^2)}. \quad (3)$$

The purpose of weighting of the output volume with the 3D Gaussian function is to combine the individual voxel values into a single score. The score is a weighted summation of the output voxel values. Because the 3D Gaussian weighting function is centered at the origin of the volume, the higher the score value is, the more likely the candidate is a polyp. Classification between polyps and nonpolyps is made by thresholding of the scores.

II.B.1. SVR

SVMs are a machine learning technique that maximizes the margin of separation between positive and negative classes. The SVMs achieve this desirable property by implementing the method of structural risk minimization. The principle of structural risk minimization states that the generalization error rate of a SVM on unseen testing data is bounded by the sum of the error rate on training data and an extra term that depends on the *Vapnik–Chervonenkis (VC) dimension*.²⁹ Therefore, a SVM is able to provide a good generalization performance. A SVM was first invented by Vapnik as a powerful tool for pattern recognition^{29,36} and has been successfully applied in handwritten digit recognition,³⁷ face detection,³⁸ text categorization,³⁹ and many others.³⁶ On the other hand, a SVM as a memory-based learning method is very fast to train because part of the training data, called *support vectors*, is stored after the training phase. SVR is a regression version of SVM by incorporating a quantitative response. This is one of the main motivations for use of SVR in the massive-training framework to improve the efficiency.

A SVM has also been adapted for a nonlinear regression problem with a quantitative response.³⁰ Unlike the conventional square loss function that is sensitive to the presence of outliers, the SVR model employs an ε -insensitive loss function that is robust to outliers. The ε -insensitive error measure ignores any errors of size less than ε . It is defined as

$$V_\varepsilon(e) = \begin{cases} 0 & \text{if } |e| < \varepsilon, \\ |e| - \varepsilon, & \text{otherwise.} \end{cases} \quad (4)$$

Therefore, any error falling into the ε -band is not counted toward loss. This is analogous to the SVM classifier, where data samples on the correct side of the decision boundary and far away from it are ignored in the optimization.

Consider a nonlinear regression model where the dependence of a scalar d on a vector \mathbf{u} is given by

$$d = f(\mathbf{u}) + \nu. \quad (5)$$

The function $f(\cdot)$ and the statistics of noise ν are unknown except that the additive noise ν is statistically independent of the input vector \mathbf{u} . In the massive-training framework, d is the continuous voxel value $T(x, y, z)$ from the corresponding teaching 3D Gaussian function in Eq. (29). The goal of the nonlinear regression model is to estimate the dependence of d on \mathbf{u} , provided there is a set of training data $\{(\mathbf{u}_i, d_i)\}_{i=1}^N$, where (\mathbf{u}_i, d_i) are the sample values of the input vector \mathbf{u} and

model output d , respectively. In SVR, an estimate of d , denoted as g , is expanded in terms of the nonlinear functions in the rich RKHS as follows:

$$g = \sum_{j=0}^L w_j \varphi_j(\mathbf{u}) = \mathbf{w}^T \Psi(\mathbf{u}), \quad (6)$$

where $\Psi(\mathbf{u}) = [\varphi_0(\mathbf{u}), \varphi_1(\mathbf{u}), \dots, \varphi_L(\mathbf{u})]^T$ is the nonlinear functions associated with RKHS, L is the dimension of the feature space which might be infinite, and $\mathbf{w} = [w_0, w_1, \dots, w_L]^T$ is the weights we aim to estimate. SVR achieves this goal by minimizing the following empirical risk:

$$\min_{\mathbf{w}} \sum_{i=1}^N V_\varepsilon(d_i - g_i) + \frac{\lambda}{2} \|\mathbf{w}\|^2, \quad (7)$$

where $V_\varepsilon(\cdot)$ is the ε -insensitive error function defined in Eq. (4), N is the total number of training samples, and λ is the regularization parameter which controls the VC dimension of the model.

Because the cost function in Eq. (4) is not differentiable at the points $\pm\varepsilon$, the optimization problem can be reformulated by introduction of nonnegative *slack variables*.³⁰ If \mathbf{w} is the minimizer of the criterion in Eq. (7), then the solution can be shown to have the form

$$\mathbf{w} = \sum_{i=1}^N (\alpha_i - \alpha'_i) \Psi(\mathbf{u}_i), \quad (8)$$

where α_i and α'_i are positive Lagrange multipliers that maximize the dual objective function

$$\begin{aligned} \max_{\alpha_i, \alpha'_i} & \sum_{i=1}^N d_i (\alpha_i - \alpha'_i) - \varepsilon \sum_{i=1}^N (\alpha_i + \alpha'_i) \\ & - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha'_i) (\alpha_j - \alpha'_j) K(\mathbf{u}_i, \mathbf{u}_j), \end{aligned} \quad (9)$$

subject to the following constraints:

$$\sum_{i=1}^N (\alpha_i - \alpha'_i) = 0,$$

$$0 \leq \alpha_i, \alpha'_i \leq 1/\lambda.$$

$K(\mathbf{u}_i, \mathbf{u}_j)$ is a symmetric non-negative inner-product kernel function defined in the RKHS with the Mercer's theorem²⁹

$$K(\mathbf{u}_i, \mathbf{u}_j) = \Psi(\mathbf{u}_i)^T \Psi(\mathbf{u}_j). \quad (10)$$

The solution depends on the input data samples through the inner-product kernel function. Therefore, even though we might not know the explicit formulation of each nonlinear function $\Psi(\mathbf{u})$, we can still obtain the optimal solution via the inner-product kernel function. In SVR, popular kernel functions include

Linear kernel function: $K(\mathbf{u}_i, \mathbf{u}_j) = \mathbf{u}_i^T \mathbf{u}_j$, (11)

d th degree polynomial function:

$$K(\mathbf{u}_i, \mathbf{u}_j) = (1 + \mathbf{u}_i^T \mathbf{u}_j)^d, \tag{12}$$

Gaussian kernel function:

$$K(\mathbf{u}_i, \mathbf{u}_j) = \exp(-\|\mathbf{u}_i - \mathbf{u}_j\|^2 / 2\sigma^2), \tag{13}$$

Tanh kernel function: $K(\mathbf{u}_i, \mathbf{u}_j) = \tanh(a\mathbf{u}_i^T \mathbf{u}_j + b)$. (14)

Note that only a subset of the solution values $(\alpha_i - \alpha'_i)$ is nonzero and the corresponding data points are called *support vectors*. The two free parameters ε and λ impact the VC dimension of the optimal nonlinear function

$$g(\mathbf{v}) = \sum_{i=1}^N (\alpha_i - \alpha'_i) K(\mathbf{u}_i, \mathbf{v}). \tag{15}$$

ε is the width that controls the tolerance of error measure. If the response d is scaled such that we use $V_\varepsilon(e/\sigma)$ instead, then we might consider using a preset value for ε . The regularization parameter λ can be estimated, for example, by cross-validation. Given any unseen testing data sample \mathbf{v} , the prediction is obtained by plugging \mathbf{v} into Eq. (15)

In our application, we aim at estimating an optimal nonlinear function $g(\mathbf{v})$ so that it is able to characterize the underlying image structures. In the testing stage, we obtained the output of MTSVR $O(x, y, z)$ by setting the variable \mathbf{v} as the input $\{I(x-p, y-q, z-r) | (p, q, r) \in V_s\}$ in Eq. (15).

II.B.2. GPR

GPR is considered as a useful probabilistic regression technique due to its theoretical simplicity and excellent generalization capacity. Whereas SVR aims at direct minimization of the regularized ε -insensitive error function, GPR employs a Bayesian methodology to derive the optimal nonlinear regression model. GPR has been successfully applied to a wide range of areas, such as object categorization³² and others.³¹ It is able to fit any arbitrary-shaped curves due to its nonparametric nature.

A Gaussian process $t(\mathbf{u})$ is a collection of random variables, any finite number of which have a joint Gaussian distribution. It is completely specified by its mean $m(\mathbf{u})$ and covariance function $V(\mathbf{u}, \mathbf{u}')$, which are defined as follows:

$$m(\mathbf{u}) = \mathbf{E}[t(\mathbf{u})], \tag{16}$$

$$V(\mathbf{u}, \mathbf{u}') = \mathbf{E}[(t(\mathbf{u}) - m(\mathbf{u}))(t(\mathbf{u}') - m(\mathbf{u}'))^T]. \tag{17}$$

For notation simplicity, we set the mean as zero. In our case, the random variables represent the value of the function $t(\mathbf{u})$ at location \mathbf{u} . A Gaussian process can be applied in a nonlinear regression model in Eq. (5). To this end, we model the estimated output as

$$t(\mathbf{u}) = \mathbf{w}^T \Phi(\mathbf{u}), \tag{18}$$

where \mathbf{w} is the weight vector that we aim at estimating and $\Phi(\mathbf{u})$ is the basis function. Similar to $\Psi(\mathbf{u})$ used in SVR, $\Phi(\mathbf{u})$ is also associated with the RKHS induced by the covariance function in the Gaussian process.

Bayesian methodology is employed in the Gaussian process to estimate the regression model. We assume *a priori* distribution of weight vector \mathbf{w} such that \mathbf{w} is a multivariate Gaussian random variable with zero mean and covariance function Q_w . We further assume that the additive noise v in the regression model in Eq. (5) follows an independent, identically distributed Gaussian distribution with zero mean and variance σ^2 . Then, the likelihood of the observations \mathbf{d} given the parameter \mathbf{w} can be written as

$$\begin{aligned} p(\mathbf{d}|\mathbf{u}, \mathbf{w}) &= \prod_{i=1}^N p(d_i|\mathbf{u}_i, \mathbf{w}) \\ &= \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(d_i - \mathbf{w}^T \Phi(\mathbf{u}_i))^2}{2\sigma^2}\right) \\ &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{|\mathbf{d} - \mathbf{w}^T \Phi(\mathbf{u})|^2}{2\sigma^2}\right), \end{aligned} \tag{19}$$

where \mathbf{d} is the vector of all individual responses, $\Phi(\mathbf{u})$ is the matrix format of the nonlinear basis function for all the input samples, and $|\mathbf{z}|^2$ denotes the Euclidean length of vector \mathbf{z} .

Given the likelihood (19) and prior information of \mathbf{w} , inference in the Bayesian regression model can be achieved via a posterior distribution based on Bayes' rule, i.e.,

$$\begin{aligned} p(\mathbf{w}|\mathbf{u}, \mathbf{d}) &= \frac{p(\mathbf{d}|\mathbf{u}, \mathbf{w})p(\mathbf{w})}{p(\mathbf{d}|\mathbf{u})} \propto \exp\left(-\frac{|\mathbf{d} - \mathbf{w}^T \Phi(\mathbf{u})|^2}{2\sigma^2}\right) \\ &\quad \times \exp\left(-\frac{\mathbf{w}^T Q_w^{-1} \mathbf{w}}{2}\right) \\ &\propto \exp\left(-\frac{1}{2} \left\{ \mathbf{w}^T \left(\frac{\Phi(\mathbf{u})\Phi(\mathbf{u})^T}{\sigma^2} + Q_w^{-1} \right) \mathbf{w} \right. \right. \\ &\quad \left. \left. - 2\frac{\mathbf{w}^T \Phi(\mathbf{u})\mathbf{d}}{\sigma^2} + \frac{\mathbf{d}^T \mathbf{d}}{\sigma^2} \right\} \right) \\ &\propto \exp\left(-\frac{1}{2}(\mathbf{w} - \bar{\mathbf{w}})^T Q(\mathbf{w} - \bar{\mathbf{w}})\right), \end{aligned} \tag{20}$$

where $Q = \sigma^{-2}\Phi(\mathbf{u})\Phi(\mathbf{u})^T + Q_w^{-1}$ is the posterior covariance function, and $\bar{\mathbf{w}} = \sigma^{-2}Q^{-1}\Phi(\mathbf{u})\mathbf{d}$ is the mean function of the posterior, which is plugged into the final regression model in Eq. (18). Unlike SVR, GPR provides not only the estimate of the weight vector \mathbf{w} , but also its covariance function.

In order to obtain a prediction for a testing case, we average over all possible parameter values weighted by their posterior probability. This is in contrast to non-Bayesian schemes such as SVR, where a single parameter is typically chosen by maximization of the cost function (7). Therefore, the predictive distribution $t(\mathbf{v})$ for any testing sample \mathbf{v} is

$$\begin{aligned}
p(t(\mathbf{v})|\mathbf{v}, \mathbf{u}, \mathbf{d}) &= \int p(t(\mathbf{v})|\mathbf{v}, \mathbf{w})p(\mathbf{w}|\mathbf{u}, \mathbf{d})d\mathbf{w} \\
&= p(t(\mathbf{v})|\mathbf{v}, \mathbf{w}) \int p(\mathbf{w}|\mathbf{u}, \mathbf{d})d\mathbf{w} \\
&= p(t(\mathbf{v})|\mathbf{v}, \mathbf{w}) \\
&\propto \mathcal{N}\left(\frac{1}{\sigma^2}\Phi(\mathbf{v})^T Q^{-1}\Phi(\mathbf{u})\mathbf{d}, \Phi(\mathbf{v})^T Q^{-1}\Phi(\mathbf{v})\right),
\end{aligned} \tag{21}$$

where $\mathcal{N}(\cdot)$ denotes a Gaussian distribution with mean $(1/\sigma^2)\Phi(\mathbf{v})^T Q^{-1}(\mathbf{u})\mathbf{d}$ and covariance function $\Phi(\mathbf{v})^T Q^{-1}\Phi(\mathbf{v})$. The mean is the optimal estimate for the testing sample \mathbf{v} . The results of Eq. (21) can be alternatively explained by Eq. (18). Because the estimate of weight vector \mathbf{w} follows a Gaussian distribution, $t(\mathbf{v})$ is also a Gaussian random variable with mean $\Phi(\mathbf{v})^T \bar{\mathbf{w}}$ and covariance function $\Phi(\mathbf{v})^T \text{cov}(\mathbf{w})\Phi(\mathbf{v})$. Based on the matrix inversion lemma, we can rewrite the mean as $\Phi(\mathbf{v})^T Q_{\mathbf{w}}(\mathbf{u})(\Phi(\mathbf{u})^T Q_{\mathbf{w}}\Phi(\mathbf{u}) + \sigma^{-2}\mathbf{I})^{-1}\mathbf{d}$. We define $V(\mathbf{u}, \mathbf{u}') = \Phi(\mathbf{u})^T Q_{\mathbf{w}}\Phi(\mathbf{u}')$, which is exactly the covariance function of the Gaussian process in Eq. (17). The covariance function between training samples \mathbf{u} and any testing sample \mathbf{v} is described as $V(\mathbf{v}, \mathbf{u}) = \Phi(\mathbf{v})^T Q_{\mathbf{w}}\Phi(\mathbf{u})$. Therefore, the optimal estimate for any testing sample \mathbf{v} in the nonlinear regression model in GPR is given by

$$t(\mathbf{v}) = V(\mathbf{v}, \mathbf{u})(V(\mathbf{u}, \mathbf{u}') + \sigma^{-2}\mathbf{I})^{-1}\mathbf{d}. \tag{22}$$

The accuracy of the estimate is highly dependent on the covariance function V . In this study, we investigated several popular covariance functions widely used in the Gaussian process

Matern covariance:

$$V(\mathbf{v}, \mathbf{u}) = \left(1 + \frac{\sqrt{3}(\mathbf{v} - \mathbf{u})}{l}\right) \exp\left(-\frac{\sqrt{3}(\mathbf{v} - \mathbf{u})}{l}\right), \tag{23}$$

Neural network covariance:

$$V(\mathbf{v}, \mathbf{u}) = \sin^{-1}\left(\frac{\mathbf{v}^T \mathbf{P} \mathbf{u}}{\sqrt{(1 + \mathbf{v}^T \mathbf{P} \mathbf{v})(1 + \mathbf{u}^T \mathbf{P} \mathbf{u})}}\right), \tag{24}$$

Rational quadratic covariance:

$$V(\mathbf{v}, \mathbf{u}) = \left(1 + \frac{\|\mathbf{v} - \mathbf{u}\|^2}{2\alpha\ell^2}\right)^{-\alpha}, \tag{25}$$

$$\text{Gaussian covariance: } V(\mathbf{v}, \mathbf{u}) = \exp\left(-\frac{\|\mathbf{v} - \mathbf{u}\|^2}{2\ell^2}\right), \tag{26}$$

$$\text{Linear covariance: } V(\mathbf{v}, \mathbf{u}) = \mathbf{v}^T \mathbf{P} \mathbf{u}. \tag{27}$$

The Gaussian covariance function in Eq. (26) is exactly the Gaussian kernel (13) used in SVR. The rational quadratic covariance can be viewed as an infinite sum of Gaussian covariance functions with different length scales. Different

covariance functions characterize different similarity measures between pairs of two data samples.

In the application of the Gaussian process, covariance functions are typically weighted by a smoothness factor, i.e., $\sigma_f V(\mathbf{v}, \mathbf{u}) + \sigma^2 \mathbf{I}$, where σ^2 is the variance of noise and σ_f controls the balance between a covariance function and noise. These two variables, together with free parameters in different covariance functions, are called *hyperparameters*, denoted as θ .

Given a covariance function, it is important to seek an optimal set of hyperparameters to fit the observation data samples. In the Gaussian process, this is achieved by maximization of the marginal log-likelihood function (19), represented by

$$\frac{\partial}{\partial \theta_j} \log p(\mathbf{d}|\mathbf{u}, \theta) = \frac{1}{2} \text{tr}\left(\left(\boldsymbol{\alpha} \boldsymbol{\alpha}^T - V^{-1}\right) \frac{\partial V}{\partial \theta_j}\right), \tag{28}$$

where tr denotes the trace of the matrix, V is the gram matrix with each element as $V(\mathbf{u}_i, \mathbf{u}_j)$, and $\boldsymbol{\alpha}$ is a vector defined as $\boldsymbol{\alpha} = V^{-1}\mathbf{d}$. The inversion of matrix V is computationally intensive with $O(N^3)$. However, once the value is known, we can easily obtain the optimal set of hyperparameters. Note that in a Gaussian process, we only assume that the weight parameter \mathbf{w} , the output $t(\mathbf{u})$, and the additive noise v in the regression model follow Gaussian distributions. No assumptions are imposed on the original input \mathbf{u} in the definition and derivation of a Gaussian process. The Bayesian paradigm is derived based only on these assumptions. In the context of MTGPR, we assume that the individual voxels in the output volume follow a Gaussian distribution, but not the voxels in the input subvolume inside the quasispherical region.

II.B.3. Training of nonlinear regression models

We used ten polyps and ten nonpolyps as training cases, which had been used in our previous studies.²⁰ If we scan the whole $64 \times 64 \times 64$ volume with a $7 \times 7 \times 7$ cube voxel by voxel, there will be 195 112 ($58 \times 58 \times 58$) subvolumes. Because most of the subvolumes overlap and are away from the polyps, we only focus on the subvolumes extracted from a $15 \times 15 \times 15$ cube (i.e., a training volume) centered at the center of the volume object. The $15 \times 15 \times 15$ cube was used in the training phase for the purpose of reduction of the number of training samples, whereas the whole $64 \times 64 \times 64$ volume was used in the testing phase to cover output responses from a regression model sufficiently. To reduce the number of subvolumes further, we selected those subvolumes sampled at every other voxel. Therefore, 512 ($8 \times 8 \times 8$) subvolumes were extracted from each volume object. In total, we extracted 5120 (512×10) subvolumes for each class, with 171 dimensions for each subvolume, according to the methodology described in Sec. II B. Therefore, there are 10 240 training examples in total. Hence, we called the proposed nonlinear regression models as MTSVR and MTGPR. The requirement for selection of a certain number of subvolumes from certain locations in the original volume is to strike a balance between the number of subvolumes and sufficient representation of volumes using subvolumes. The

overlapping subvolumes in the training volume are highly correlated when they are closely spaced; therefore, neighboring subvolumes may be redundant for training. As described in Sec. II B 3, we selected subvolumes at every other voxel inside the $15 \times 15 \times 15$ training volume. This pruning method reduced the number of training subvolumes while maintaining a comparable performance.⁴⁷ The teaching volumes in the massive-training framework contain the distribution for the “likelihood of being a polyp.” The teaching response T in MTSVR and MTGPR for training examples extracted from polyps used a 3D Gaussian weighting function whose peak was located at the center of the polyp. Therefore, we need to extract subvolumes from a whole polyp as well as its background. Recently, Ong *et al.*⁴⁰ proposed to select voxels that belong to the elliptic class of the peak subtype using a one-ring neighborhood (i.e., voxels that satisfy a certain geometric criterion) in detecting polyps in CTC. They demonstrated a better performance with this approach, which limits the feature extraction area in the surface of polyps because their approach focused on geometric (or shape) features. This is different in the massive-training framework where the nonlinear regression model learns the local texture or gray-level information on whole polyps and nonpolyps based on the individual subvolumes. Therefore, we need to extract subvolumes not only from a polyp surface, but also from the inside of the polyp. The teaching response T in MTSVR and MTGPR for training examples extracted from polyps used a 3D Gaussian weighting function whose peak was located at the center of the polyp. On the other hand, we used all zero values as the desired response for nonpolyp training examples. Therefore, both models were able to learn the underlying image structures by enhancing polyps with a 3D Gaussian weighting function and suppressing nonpolyps with zeros. The desired response is described as follows:

$$T(x, y, z) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_T} \exp\left\{-\frac{x^2 + y^2 + z^2}{2\sigma_T^2}\right\} & \text{for a polyp,} \\ 0 & \text{for nonpolyps.} \end{cases} \quad (29)$$

The standard deviation σ_T controls the size of the Gaussian weighting function. The coordinate (x, y, z) is consistent with the one used in Eq. (1). Training of MTSVR and MTGPR involves a large number of subvolume-voxel pairs. The input sample is a vector of length 171 and the teaching response T is a scalar, either a voxel value extracted from the 3D Gaussian weighting function or zero.

MTSVR used quadratic programming to solve the maximization problem in Eq. (9) and to obtain the optimal solution in Eq. (15). MTGPR employed the conjugate gradient for maximizing the marginal log-likelihood function in Eq. (28) and achieved the optimal hyperparameter. Both methods are memory-based because compact kernel/covariance functions are used in both approaches. Therefore, some training examples were stored after the training process. In MTSVR, only a part of training examples was retained, which are called support vectors. On the other hand, all the training

data were stored in MTGPR because the functional evaluation in Eq. (22) required the presence of all training samples. This is very different from the MTANN approach, where all of the training data were discarded after the linear-output back-propagation algorithm.^{41,42} The relevant information was represented in the optimal weights in the linear-output ANN regression model.

We will explore different kernel functions in Eqs. (11)–(14) used in MTSVR. We used grid search to find the best kernel function with optimal parameters. On the other hand, MTGPR is able to compute the hyperparameters through the optimization process for all of the covariance functions in Eqs. (23)–(26). The optimal prediction for any testing samples can have noiseless and noisy formulations in MTGPR. Specifically, the noiseless prediction setup sets $\sigma = 0$ in Eq. (22) and noisy prediction optimizes the parameter. However, we only used noisy prediction in MTGPR in the following experiments in Sec. III, as becomes evident in Sec. IV when we compare these two approaches.

II.C. Performance evaluation criteria

We used a mean-square error (MSE) to evaluate the training performance of MTSVR and MTGPR as it offers a direct comparison between the teaching response and the output from the models. To assess the performance of the trained models in the testing data, we calculated the area under the receiver-operating-characteristic (ROC) curve (AUC) values,⁴³ a FP reduction rate without removal of TPs, and free-response ROC (FROC) analysis⁴⁴ as performance metrics. The AUC value is calculated based on the maximum-likelihood estimation of the binormal ROC curve.⁴⁵ It offers the overall performance of the regression models after 3D scoring as classifiers to distinguish between polyps and nonpolyps. We conducted statistical tests to determine whether the difference in AUC values from different methods is statistically significant. The FP reduction rate without removal of TPs describes the percentage of FPs that have been eliminated by selecting a threshold without sacrificing any TPs.

III. RESULTS

In this section, we present the performance of the proposed MTSVR and MTGPR in reducing FP detections. We also compare the results to the previous studies on MTANN.²⁰

III.A. Training performance of MTSVR and MTGPR

We manually selected ten representative polyps and ten nonpolyps with different visual appearance such as size, shape, and contrast from the CTC data sets. The ten nonpolyps covered major sources of FPs such as rectal tubes, stool, haustral folds, colonic walls, and the ileocecal valve. The purpose was to make the training data samples represent the database (ideally the whole population). The standard deviation σ_T in the desired response volume (29) was empirically chosen as 4.5 voxels.²⁰

TABLE I. MSE between teaching responses and outputs for training data samples and AUC values and false positive reduction rates without removal of TPs from different regression models with different parameters for testing data samples.

Methods	Model parameters		Training MSE	Testing AUC	FP reduction rate without removal of TPs	
MTSVR	Linear		0.0419	0.7821	0.1391	
	Polynomial	d	2	0.0363	0.7734	0.1623
			4	0.0343	0.7754	0.1656
			6	0.0347	0.7752	0.1634
			8	0.0456	0.7802	0.1367
			10	0.1458	0.7572	0.1574
	Gaussian	σ	0.1	0.0278	0.7930	0.4151
			0.35	0.0178	0.7864	0.5147
			0.7	0.0131	0.7900	0.4519
			1	0.0114	0.7941	0.3885
			5	0.0083	0.8578	0.2378
			10	0.0082	0.8608	0.2085
	tanh	$a=1, b=0$ $a=2, b=0$ $a=3, b=0$ $a=3, b=1$		0.0435	0.7738	0.1428
				0.0422	0.7823	0.1439
			0.0420	0.7834	0.1445	
			0.0420	0.7835	0.1446	
MTGPR	Matern		0.0204	0.8215	0.5042	
	Neural network		0.0304	0.8288	0.4662	
	Rational quadratic		0.0194	0.8245	0.4746	
	Gaussian		0.0235	0.7938	0.4113	
	Gaussian+linear		0.0233	0.8032	0.4430	
MTANN	Number of hidden neurons	10	0.0239	0.7804	0.2965	
		20	0.0182	0.7867	0.3476	
		25	0.0162	0.7707	0.4683	
		30	0.0147	0.8012	0.2822	
		40	0.0117	0.8139	0.2638	
		50	0.0106	0.8080	0.3905	

Table I presents the MSE between the teaching responses and the outputs from MTSVR, MTGPR, and MTANN with different parameters. The four kernel functions in Eqs. (11)–(14) with different parameter values were studied in MTSVR. The Gaussian kernel function offered the best performance in terms of MSE over the other three kernel functions. Linear and polynomial kernel functions produced a much higher MSE. It is interesting to note that the tanh kernel function had a relatively similar performance for different combinations of parameters. We tested four individual covariance functions in Eqs. (23)–(26) in MTGPR. One useful property of the covariance function in the Gaussian process is that the combination of two covariance functions forms another valid covariance function. Therefore, we also applied a popular combination of Gaussian [Eq. (26)] and linear [Eq. (27)] covariance functions in MTGPR. The neural network covariance function produced the largest MSE, whereas other covariance functions achieved a relatively close performance. The combination of Gaussian and linear covariance functions did not provide a performance better than the Gaussian covariance function alone. This was due to the normalization of training samples. We varied the number of neurons in the hidden layer in MTANN. Suzuki *et al.*²⁰ used 25 in their study. Although the MTANN with 25 neu-

rons in the hidden layer did not offer the best MSE for training samples, it achieved the highest FP reduction rate for testing samples.²⁰

In order further to gain qualitative insights into the different performance of the three methods, we applied the best-trained regression models to the training polyps and nonpolyps. Figure 3 presents five representative training polyps and nonpolyps with corresponding output images from MTSVR, MTGPR, and MTANN. We used a Gaussian kernel in Eq. (13) with $\sigma=0.35$ in MTSVR, a Matern covariance function in Eq. (23) in MTGPR, and 25 neurons in the hidden layer in the MTANN. The teaching response images for polyps in the training phase contain the 3D Gaussian function in Eq. (29) shown in the second row of Fig. 3(a). The second row in Fig. 3(b) shows the teaching response images for nonpolyps. All three methods were able to learn the underlying CTC images by enhancing polyps and suppressing nonpolyps. The output volumes for two small polyps (the farthest left and right images) from three methods are stronger and larger than the original ones in the input volumes, which demonstrates the ability of MTSVR, MTGPR, and MTANN to enhance small polyps. However, it should also be noted that the shapes of the output volumes for the farthest left polyp in MTGPR and MTSVR are rounder and larger than the one in MTANN

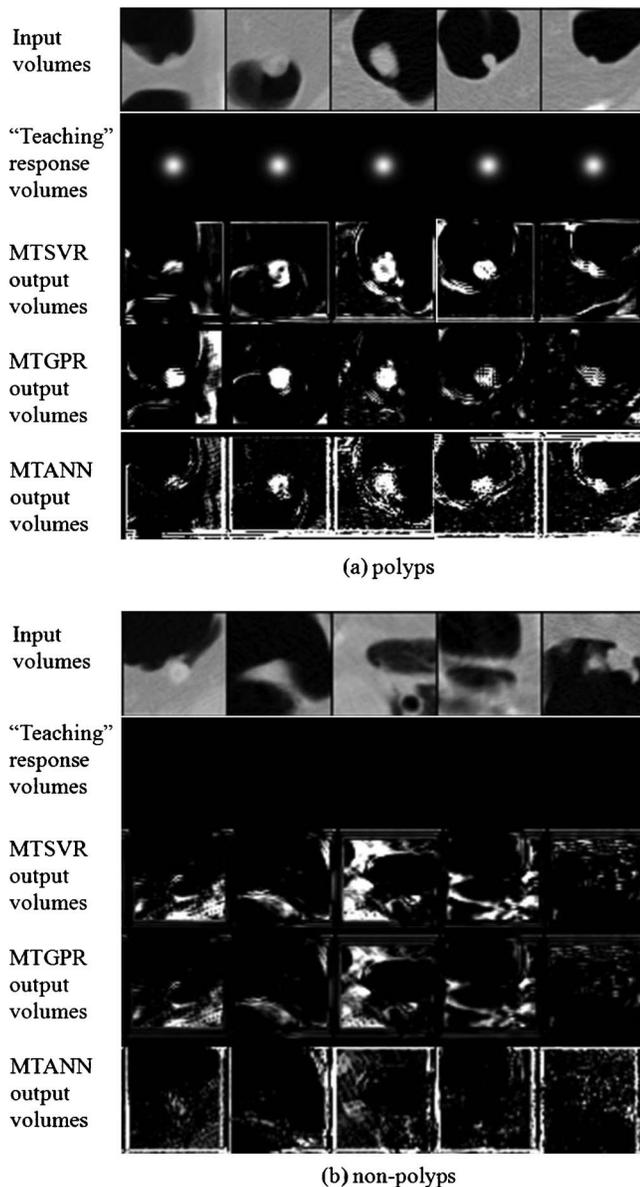


FIG. 3. Illustrations of the central axial slices of representative training (a) polyp volumes and (b) nonpolyp volumes. In the output volumes of MTSVR, MTGPR, and MTANN, polyps are represented by bright voxels in the center, whereas nonpolyps are suppressed and almost dark.

output volume. Figure 3(b) presents the case for training nonpolyps. The center object is dark in the farthest left output volume of MTGPR, whereas it remains in the output volume of MTANN. MTSVR had an intermediate performance.

III.B. Performance comparison on testing CTC data

The main focus of the study is to reduce the number of FP detections in CTC CADE. Although MSE is a good indicator for regression, it might not be strictly correlated with the FP reduction. Therefore, we applied all the trained regression models, not only the best one in terms of MSE, to 474 testing nonpolyps (FPs) and 36 TP volumes (that constitute 18 polyps). Table I presents the AUC values and FP reduction rates

without removal of TPs for MTSVR, MTGPR, and MTANN with different parameters for the testing CTC cases. MTSVR with a Gaussian kernel with $\sigma=10$ produced the best AUC value. This is consistent with the lowest MSE obtained with the same parameter. However, the highest FP reduction rate was achieved with a Gaussian kernel with $\sigma=0.35$, which suggests the difference between two performance metrics. Linear, polynomial, and tanh kernel functions had much lower performance than the Gaussian kernel function. The best performance was obtained with a Matern covariance function in MTGPR, though the rational quadratic covariance function had a slightly better MSE value for the training data samples. The combination of Gaussian and linear covariance functions did not offer performance better than the Gaussian covariance function alone in MTGPR. We also provide the performance for MTANN with different numbers of neurons in the hidden layer in Table I. MTANN with 25 hidden layer neurons achieved the best FP reduction rate. Because the FP reduction rate is the most important criterion in this study, we chose a Gaussian kernel function with $\sigma=0.35$ for MTSVR, a Matern covariance function for MTGPR, and 25 hidden layer neurons for MTANN in the following experiments without further mentioning the specific parameters.

We selected five representative testing polyps and nonpolyps to show different output volumes from MTGPR, MTSVR, and MTANN in Fig. 4. The trained models were able to enhance the testing polyps and suppress the nonpolyps. The ability of all methods to enhance small polyps can be demonstrated again in the middle of the images in Fig. 4(a). In the farthest left and middle images of Fig. 4(b), the testing nonpolyps become dark in the output volumes of MTGPR and MTSVR. However, the nonpolyps can still be seen in the output volumes of MTANN.

Table II presents the two-sided p -values of the differences between two AUC values from three methods. As can be seen from the table, the differences between the AUC values from MTGPR and MTANN and from MTGPR and MTSVR are statistically significant (two-sided p -values <0.05). On the other hand, the AUC values from MTSVR and MTANN is not statistically significant with a two-sided p -value of 0.75. The results demonstrate better performance of MTGPR over MTSVR and MTANN. Figure 5 plots the ROC curves from the three methods.

The difference between ROC curves from MTGPR and MTSVR can be further investigated by examining the score distributions. We present the histograms of scores from MTSVR and MTGPR in Figs. 6 and 7. The dynamic range of the scores from MTGPR is larger than the one from MTSVR. The distribution of nonpolyps in MTSVR has a bell shape centering in the middle. On the other hand, the distribution of nonpolyps in MTGPR is more concentrated in the center with longer tails. The distributions of polyps in both methods are much closer than those of nonpolyps. The scores for polyps in MTGPR tend to have a larger dynamic range. Although both methods were able to eliminate half of the FP detections, the underlying score distributions are very different.

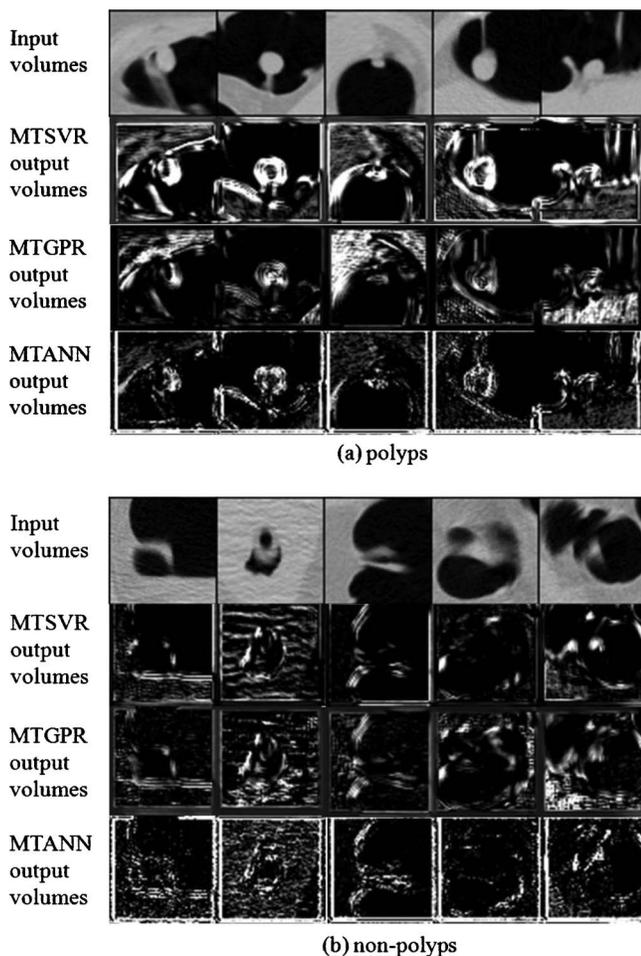


FIG. 4. Illustrations of the central axial slices of representative testing (a) polyp volumes and (b) nonpolyp volumes. Testing polyps are enhanced in the center of the output volumes, whereas testing nonpolyps are suppressed.

We evaluated the performance of the proposed MTGPR and MTSVR for FP reduction by using FROC analysis. Figure 8 shows the FROC curves for comparison. The FROC curves indicate that MTGPR was able to reduce 50% (239/474) of the FP detections without removing any of the 36 TP volumes, i.e., a 94.7% (18/19) by-polyp sensitivity was obtained at a FP rate of 2.52 (235/93) per patient, whereas MTSVR achieved a 94.7% (18/19) by-polyp sensitivity with a FP rate of 2.47 (230/93) per patient. The FP rates for MTSVR and MTGPR are very close. However, MTGPR achieved a slightly better AUC value than did MTSVR. This is due to the outward shape of the FROC curve for MTGPR in the drop-off region. MTANN was able to eliminate 48.5% (230/474) of the FPs, which results in a 94.7% (18/19) by-

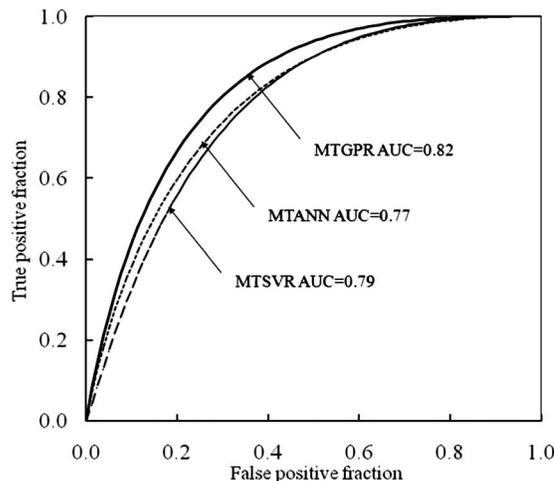


FIG. 5. The ROC curves for MTSVR, MTGPR, and MTANN with AUC values.

polyp sensitivity with 2.62 (244/93) FPs per patient. MTGPR and MTSVR offered slightly better performance than MTANN in terms of the FP reduction rate.

III.C. Computational efficiency comparison

One of the main contributions of the paper is to improve the computational efficiency of the massive-training framework in the development phase of a CAde scheme by using SVR and GPR while maintaining a comparable performance. In the development stage of a new CAde scheme, a low training computational cost is crucial because one would change some parameters of the massive-training model, alter training cases, or optimize the parameters of an initial detection scheme.

Let N be the total number of training samples, m the number of support vectors in MTSVR, d_l the dimension of training input samples, in the case where $m/N \ll 1$, the number of operations for MTSVR is $O(m^3 + m^2N + mNd_l)$.³⁶ In our application, the ratio m/N is usually around 0.2. In MTGPR, the inversion of the Gram matrix is required in order to carry out the prediction in Eq. (18). The computational complexity of inverting the Gram matrix is $O(N^3)$. Great efforts have been devoted to reducing the computation cost of GPR for large data sets.³¹ Because there is redundancy in the training data due to the approach we used to extract subvolumes in MTGPR, we employed *projected process* approximation methodology to reduce the computational cost to $O(s^2N)$, where s is the number of latent function values.³¹ We used a random selection method to choose half positive and half

TABLE II. Statistical comparisons among the performance (AUC values) of MTSVR, MTGPR, and MTANN in the distinction between polyps and nonpolyps. The AUC values with standard deviation and two-sided p values are shown.

	MTGPR (AUC=0.82 ± 0.03)	MTANN (AUC=0.77 ± 0.03)
MTSVR (AUC=0.79 ± 0.03)	0.04	0.75
MTGPR	...	0.03

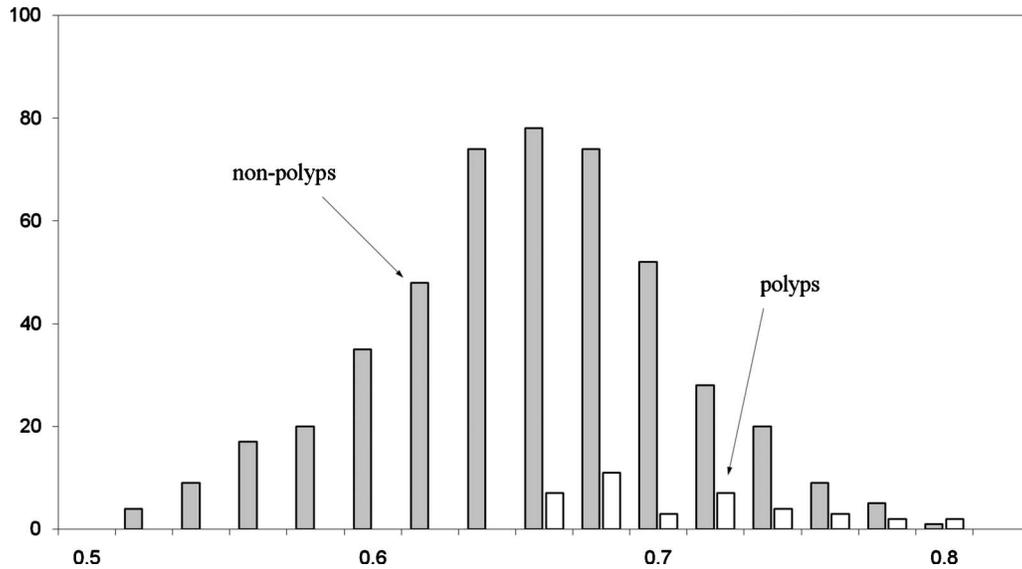


FIG. 6. Histogram of the scores from MTSVR with a Gaussian kernel function with $\sigma=0.35$ for 36 true positive volumes (from 18 polyps) and 474 false positives (nonpolyps) produced by the original CADe scheme for the detection of polyps in CTC.

negative training samples as the number s .³¹ Therefore, both MTSVR and MTGPR scale linearly with the number of the training samples N . On the other hand, the computational complexity of the MTANN depends on the dimension of training input samples d_I , the number of hidden neurons N_H , the number of training samples N , and the number of iterations to train the ANN by using the back-propagation algorithm.⁴⁶ The training of an MTANN was performed 500 000 times.²⁰ Therefore, the computational costs of MTSVR and MTGPR would be lower than that of the MTANN in most situations.

We compared the computational costs of MTSVR, MTGPR, and MTANN on a workstation (Intel, Xeon, 2.7GHz,

1GB RAM). The training of the original MTANN took 38 h, whereas that of the MTSVR and that of the MTGPR took 12 min and 25 h, respectively. The training time was reduced by factors of 190 and 1.3 with the MTSVR and the MTGPR, respectively. Compared to the MTANN coupled with a Laplacian eigenmap for dimension reduction in Ref. 28, the MTSVR offered a comparable performance in terms of AUC values and the FP reduction rate without removal of TPs while reducing the training time even more (i.e., by a factor of 20). Although the MTGPR did not reduce the training time as much as the MTSVR did, it obtained a statistically significant increase in AUC values over the MTANN.

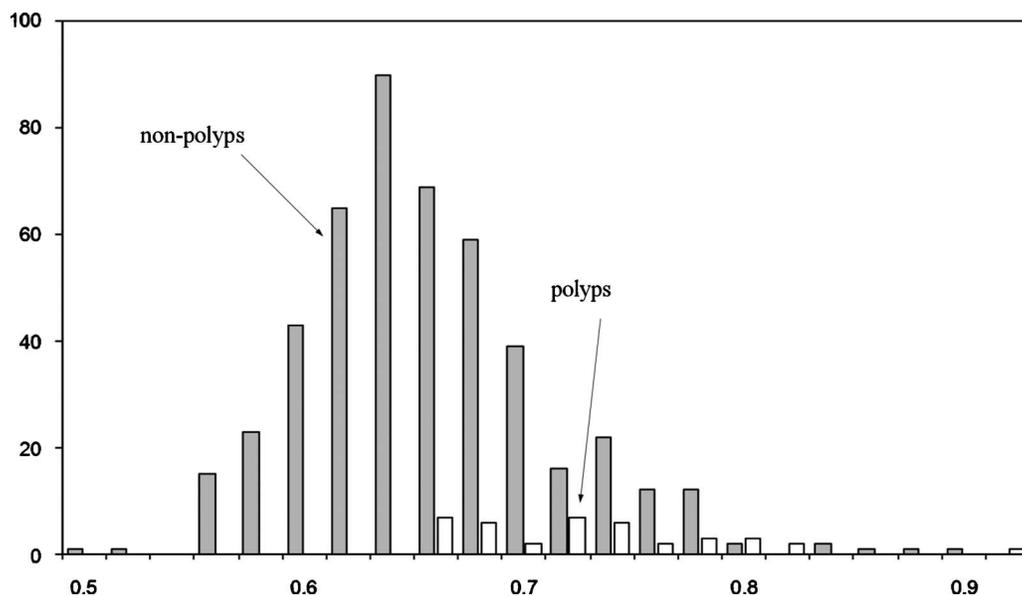


FIG. 7. Histogram of the scores from MTGPR with a Matern covariance function for 36 true positive volumes (from 18 polyps) and 474 false positives (nonpolyps) produced by the original CADe scheme for the detection of polyps in CTC.

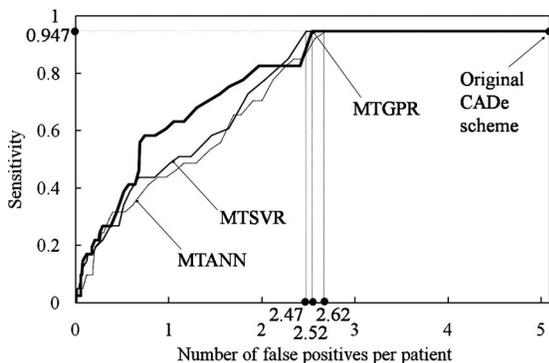


FIG. 8. FROC curves for MTSVR, MTGPR, and MTANN for the testing CTC cases. The performance of the original CADe scheme is shown on the far right with a 94.7% sensitivity at 5.09 FPs per patient. MTSVR achieves the best specificity with 2.47 FPs per patient, followed by MTGPR with 2.52 FPs per patient and MTANN with 2.62 FPs per patient.

IV. DISCUSSION

We used Eq. (16) to predict the output for any testing data samples in MTGPR. This is called *noisy prediction* because of the σ^2 term in the formulation. This is realistic in modeling the output since we do not have direct access to function values themselves, but only noisy version thereof Eq. (5). From an optimization perspective, σ^2 serves as regularization in Eq. (16). Another modeling in MTGPR is called *noiseless prediction*, where σ^2 is omitted in Eq. (16). This approach focuses on incorporating the knowledge about the nonlinear function that training data samples provide without using any *a priori* information about the function. In order to compare these two approaches in MTGPR, we also trained noiseless prediction with different covariance functions. Table III presents the performance comparison between two approaches. It is evident from the table that the noisy prediction method outperforms the noiseless prediction one. Therefore, noisy prediction is more realistic and achieves better performance.

The standard deviation σ_w in Eq. (3) controls the shape of the 3D Gaussian weighting function which, in turn, determines the output scores for objects (i.e., polyp candidates). We investigated the effect of σ_w parameter change on the performance of the MTSVR and MTGPR. We changed the standard deviation σ_w in the range between 0.5 and 15. The FP reduction rates without removal of TPs on the testing cases are shown in Fig. 9. It is interesting to note that the curves for the MTSVR and MTGPR exhibit a similar trend. Because the performance was the highest at a standard de-

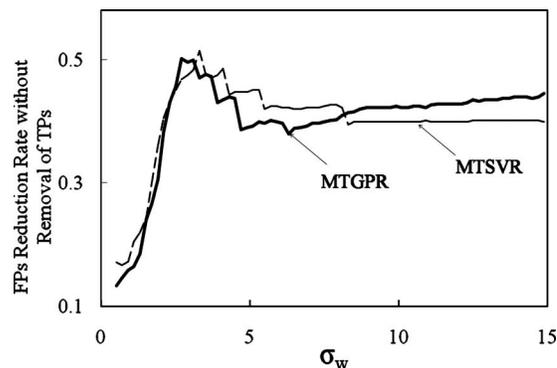


FIG. 9. FP reduction rates without removal of TPs versus parameter σ_w in the 3D Gaussian weighting function for MTGPR and MTSVR, respectively.

viation of 2.7 for the MTGPR and 3.3 for the MTSVR, we used these two values in our experiments. These results were consistent with that in the distinction between benign and malignant nodules in CT images by means of MTANNs.⁴⁷

Both MTSVR and MTGPR are memory-based methods because a compact kernel and covariance functions are used in the formulations. Therefore, part or all of the training data were required during testing phase. This is inherent in any kernel-based learning methods.^{29,30} On the other hand, MTANN discarded all the training data after training phase where information from training data was extracted.

In MTSVR, the Gaussian kernel function offered the best performance both for the training set and for the testing set. This suggests that there exists a nonlinear structure underlying the CTC images, while linear and polynomial kernel functions failed to capture the nonlinearity. On the other hand, different covariance functions in MTGPR produced quite close performance. This suggests that CTC images are quite robust to different similarity measures offered by different covariance functions in MTGPR. It is interesting to notice that MTSVR with a Gaussian kernel function outperforms MTGPR with a Gaussian covariance function in terms of FP reduction rate. This is because two methods have very different mathematical formulations. The kernel width of σ in MTSVR was selected using the grid search strategy, while it was optimized through maximization of the marginal probability density function in MTGPR.

One limitation of our study is the limited number of CTC cases with polyps used in the experiments. A larger data set of CTC might provide more realistic and reliable evaluation

TABLE III. Performance comparisons of noisy and noiseless predictions in MTGPR.

	Noisy prediction		Noiseless prediction	
	Testing AUC	FP reduction rate without removal of TPs	Testing AUC	FP reduction rate without removal of TPs
Matern	0.8215	0.5021	0.8182	0.4662
Neural network	0.8288	0.4662	0.8233	0.4345
Rational quadratic	0.8245	0.4746	0.8192	0.4683
Gaussian	0.7938	0.4113	0.7893	0.2532

of the proposed methods. However, it should be noted that both SVMs and Gaussian process were designed to have good generalization ability.²⁹ In addition, the proposed methods were tested by use of a widely “accepted” leave-one-lesion-out cross-validation test which is generally unbiased.^{48–50} Especially for a small number of cases, a leave-one-out cross-validation test provides a pessimistically biased estimate; in other words, the performance estimate obtained with this test is lower than the “true” performance.^{48,51} Also, a study in Ref. 52 showed that a leave-one-out cross-validation test provides a performance estimate with good generalization ability. Thus, we expect that the performance estimates for both MTSVR and MTGPR reported in this paper would be comparable to (or potentially better than) the performance obtained when applied to a larger data set.

Evaluation with a larger data set is desirable because it would provide more *reliable* performance estimates (i.e., precise or with a lower variance), but not necessarily offer more accurate evaluation estimates (i.e., close to true performance or with a smaller bias). The authors in Ref. 52 stated, “Please note that the bias is the primary concern of inaccuracy in the estimated performance levels that we should try to minimize, even at a cost of slight increase in the variance of estimated performance levels.” Our primary interest is closeness to the true performance and generalization of performance estimators. When we discuss performance evaluation, the accuracy of the evaluation is most important; thus, researchers have been seeking an unbiased estimator.^{48–54} As stated earlier, the performance estimate by a leave-one-out cross-validation test provides a pessimistically biased estimate^{48–50} with good generalization.^{48,51} Therefore, we expect that the performance estimates reported in this paper would be comparable to (or potentially better than) the performance obtained when applied to a larger data set. On the other hand, if our primary interest is precision of evaluation (as opposed to accuracy), evaluation with a large data set is much more important. This is true for other studies where the cohort population is seriously important, such as population-based studies in medicine. In other words, the larger your sample size, the more sure you can be that the answers truly reflect the population. However, our study interest is different from that. Therefore, we would like to leave evaluation with a large data set as our future work.

The dimension of the input vector to the model is closely related to the training time of the proposed MTSVR and MTGPR. In our previous studies, we presented two dimension reduction techniques, namely, Laplacian eigenmap²⁸ and principal component analysis (PCA),⁵⁵ in the MTANN framework to improve the efficiency in the training. Both methods were able to reduce the MTANN training time while maintaining a comparable performance in terms of the AUC value and FP reduction rate. The application of dimension reduction techniques, such as PCA and Laplacian eigenmaps, in the MTSVR and MTGPR would be important and interesting. Based on our previous studies,^{28,55} we expect that similar results would hold for MTSVR and MTGPR. We will leave this topic as our future research work.

V. CONCLUSION

We developed a 3D MTSVR and a 3D MTGPR to reduce FPs in a CADe scheme for the detection of polyps in CTC. Both MTSVR and MTGPR were able to eliminate around half of the FP detections from the original CADe scheme without sacrificing the sensitivity. We compared the performance against previously presented MTANN (Ref. 20) that was based on artificial neural networks. MTGPR achieved statistically significant increase of the AUC value over MTANN and MTSVR. Both MTSVR and MTGPR obtained a slightly better FP rate than did MTANN at the same sensitivity level. With MTSVR, the training time was reduced by a factor of 190 compared to that of MTANN. Therefore, both MTSVR and MTGPR could be useful for improving the specificity of a CADe scheme for the detection of polyps in CTC.

ACKNOWLEDGMENTS

The authors are grateful to Ms. E. F. Lanzl for improving the manuscript. This work was supported by Grant No. R01CA120549 from the National Cancer Institute/National Institutes of Health and partially by NIH Grant Nos. S10 RR021039 and P30 CA14599. CAD technologies developed at the University of Chicago have been licensed to companies including R2 Technology (Hologic), Riverain Medical, Deus Technology, Median Technologies, Mitsubishi Space Software, General Electric, and Toshiba. It is the policy of the University of Chicago that investigators disclose publicly actual or potential significant financial interests that may appear to be affected by research activities.

^{a)} Author to whom correspondence should be addressed. Electronic mail: jwxu@uchicago.edu; Telephone: (773) 834-5108; Fax: (773) 702-0371.

¹ A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, and M. J. Thun, “Cancer statistics, 2009,” *Ca-Cancer J. Clin.* **59**, 225–249 (2009).

² S. J. Winawer, R. H. Fletcher, L. Miller, F. Godlee, M. H. Stolar, C. D. Mulrow, S. H. Woolf, S. N. Glick, T. G. Ganiats, J. H. Bond, L. Rosen, J. G. Zapka, S. J. Olsen, F. M. Giardiello, J. E. Sisk, R. Van Antwerp, C. Brown-Davis, D. A. Marciniak and R. J. Mayer, “Colorectal cancer screening: Clinical guidelines and rationale,” *Gastroenterology* **112**, 594–642 (1997).

³ C. Gene Coin, F. C. Wollett, J. T. Coin, M. Rowland, R. K. DeRamos, and R. Dandrea, “Computerized radiology of the colon: A potential screening technique,” *Comput. Radiol.* **7**, 215–221 (1983).

⁴ A. S. Chaoui, M. A. Blake, M. A. Barish, and H. M. Fenlon, “Virtual colonoscopy and colorectal cancer screening,” *Abdom. Imaging* **25**, 361–367 (2000).

⁵ C. D. Johnson and A. H. Dachman, “CT colonography: The next colon screening examination?,” *Radiology* **216**, 331–341 (2000).

⁶ D. J. Vining, “Virtual colonoscopy,” *Gastrointest Endosc Clin. N. Am.* **7**, 285–291 (1997).

⁷ H. Yoshida and A. H. Dachman, “CAD techniques, challenges, and controversies in computed tomographic colonography,” *Abdom. Imaging* **30**, 26–41 (2005).

⁸ D. J. Vining, Y. Ge, D. K. Ahn, and D. R. Stelts, “Virtual colonoscopy with computer-assisted polyp detection,” in *Computer-Aided Diagnosis in Medical Imaging*, edited by K. Doi (Elsevier Science, Amsterdam, 1999), pp. 445–452.

⁹ H. Yoshida and J. Näppi, “Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps,” *IEEE Trans. Med. Imaging* **20**, 1261–1274 (2001).

¹⁰ S. B. Gokturk, C. Tomasi, B. Acar, C. F. Beaulieu, D. S. Paik, R. B. Jeffrey, Jr., J. Yee, and S. Napel, “A statistical 3-D pattern processing method for computer-aided detection of polyps in CT colonography,”

- IEEE *Trans. Med. Imaging* **20**, 1251–1260 (2001).
- ¹¹R. M. Summers, W. S. Selbie, J. D. Malley, L. M. Pusanik, A. J. Dwyer, N. A. Courcoutsakis, D. J. Shaw, D. E. Kleiner, M. C. Sneller, C. A. Langford, S. M. Holland, and J. H. Shelhamer, “Polypoid lesions of airways: Early experience with computer-assisted detection by using virtual bronchoscopy and surface curvature,” *Radiology* **208**, 331–337 (1998).
 - ¹²R. M. Summers, C. D. Johnson, L. M. Pusanik, J. D. Malley, A. M. Youssef, and J. E. Reed, “Automated polyp detection at CT colonography: Feasibility assessment in a human population,” *Radiology* **219**, 51–59 (2001).
 - ¹³A. Jerebko, S. Lakare, P. Cathier, S. Periaswamy, and L. Bogoni, “Symmetric curvature patterns for colonic polyp detection,” in Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Vol. 9, pp. 169–176, 2006.
 - ¹⁴B. Acar, C. F. Beaulieu, S. B. Gokturk, C. Tomasi, D. S. Paik, R. B. Jeffrey, Jr., J. Yee, and S. Napel, “Edge displacement field-based classification for improved detection of polyps in CT colonography,” *IEEE Trans. Med. Imaging* **21**, 1461–1467 (2002).
 - ¹⁵A. 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.* **12**, 479–486 (2005).
 - ¹⁶A. K. Jerebko, R. M. Summers, J. D. Malley, M. Franaszek, and C. D. Johnson, “Computer-assisted detection of colonic polyps with CT colonography using neural networks and binary classification trees,” *Med. Phys.* **30**, 52–60 (2003).
 - ¹⁷V. F. van Ravesteijn, C. van Wijk, F. M. Vos, R. Truyen, J. F. Peters, J. Stoker, and L. J. van Vliet, “Computer-aided detection of polyps in CT colonography using logistic regression,” *IEEE Trans. Med. Imaging* **29**, 120–131 (2010).
 - ¹⁸J. Yao, J. Li, and R. M. Summers, “Employing topographical height map in colonic polyp measurement and false positive reduction,” *Pattern Recogn.* **42**, 1029–1040 (2009).
 - ¹⁹H. Zhu, Z. Liang, P. J. Pickhardt, M. A. Barish, J. You, Y. Fan, H. Lu, E. J. Posniak, R. J. Richards, and H. L. Cohen, “Increasing computer-aided detection specificity by projection features for CT colonography,” *Med. Phys.* **37**, 1468–1481 (2010).
 - ²⁰K. Suzuki, H. Yoshida, J. Nappi, and A. H. Dachman, “Massive-training artificial neural network (MTANN) for reduction of false positives in computer-aided detection of polyps: Suppression of rectal tubes,” *Med. Phys.* **33**, 3814–3824 (2006).
 - ²¹K. Suzuki, S. G. Armato, F. Li, S. Sone, and K. Doi, “Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography,” *Med. Phys.* **30**, 1602–1617 (2003).
 - ²²K. Suzuki, H. Abe, H. MacMahon, and K. Doi, “Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (MTANN),” *IEEE Trans. Med. Imaging* **25**, 406–416 (2006).
 - ²³K. Suzuki, H. Yoshida, J. Nappi, S. G. Armato, and I. 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.* **35**, 694–703 (2008).
 - ²⁴H. Arimura, S. Katsuragawa, K. Suzuki, F. Li, J. Shiraishi, S. Sone, and K. Doi, “Computerized scheme for automated detection of lung nodules in low-dose computed tomography images for lung cancer screening,” *Acad. Radiol.* **11**, 617–629 (2004).
 - ²⁵F. Li, H. Arimura, K. Suzuki, J. Shiraishi, Q. Li, H. Abe, R. Engelmann, S. Sone, H. MacMahon, and K. Doi, “Computer-aided detection of peripheral lung cancers missed at CT: ROC analyses without and with localization,” *Radiology* **237**, 684–690 (2005).
 - ²⁶K. Suzuki, “A supervised ‘lesion-enhancement’ filter by use of a massive-training artificial neural network (MTANN) in computer-aided diagnosis (CAD),” *Phys. Med. Biol.* **54**, S31–S45 (2009).
 - ²⁷K. Suzuki, J. Shiraishi, H. Abe, H. MacMahon, and K. Doi, “False-positive reduction in computer-aided diagnostic scheme for detecting nodules in chest radiographs by means of massive training artificial neural network,” *Acad. Radiol.* **12**, 191–201 (2005).
 - ²⁸K. Suzuki, J. Zhang, and J. Xu, “Massive-training artificial neural network coupled with Laplacian-eigenfunction-based dimensionality reduction for computer-aided detection of polyps in CT colonography,” *IEEE Trans. Med. Imaging* **29**, 1907–1917 (2010).
 - ²⁹V. N. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed. (Springer, New York, 1998).
 - ³⁰A. J. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Stat. Comput.* **14**, 199–222 (2004).
 - ³¹C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, MA, 2005).
 - ³²A. Kapoor, K. Grauman, R. Urtasun, and T. Darrell, “Gaussian processes for object categorization,” *Int. J. Comput. Vis.* **88**, 169–188 (2010).
 - ³³A. Lostumbo, K. Suzuki, and A. H. Dachman, “Flat lesions in CT colonography,” *Abdom. Imaging* **35**, 578–583 (2010).
 - ³⁴J. Nappi, A. H. Dachman, P. MacEneaney, and H. Yoshida, “Automated knowledge-guided segmentation of colonic walls for computerized detection of polyps in CT colonography,” *J. Comput. Assist. Tomogr.* **26**, 493–504 (2002).
 - ³⁵J. Nappi and H. Yoshida, “Feature-guided analysis for reduction of false positives in CAD of polyps for computed tomographic colonography,” *Med. Phys.* **30**, 1592–1601 (2003).
 - ³⁶C. J. C. Burges, “A tutorial on support vector machines for pattern recognition,” *Data Min. Knowl. Discov.* **2**, 121–167 (1998).
 - ³⁷C. J. C. Burges and B. Schölkopf, “Improving the accuracy and speed of support vector machines,” *Advances in Neural Information Processing Systems 9* (MIT Press, Cambridge, MA, 1997), pp. 375–381.
 - ³⁸E. Osuna, R. Freund, and F. Girosi, “Training support vector machines: An application to face detection,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 130–136, 1997.
 - ³⁹T. Joachims, *Learning to Classify Text Using Support Vector Machines—Methods, Theory, and Algorithms* (Springer, New York, 2002).
 - ⁴⁰J. L. Ong and A.-K. Seghouane, “From point to local neighbourhood: Polyp detection in CT colonography using geodesic ring neighbourhoods,” *IEEE Trans. Image Process.* (in press).
 - ⁴¹K. Suzuki, I. Horiba, N. Sugie, and M. Nanki, “Extraction of left ventricular contours from left ventriculograms by means of a neural edge detector,” *IEEE Trans. Med. Imaging* **23**, 330–339 (2004).
 - ⁴²K. Suzuki, I. Horiba, and N. Sugie, “Neural edge enhancer for supervised edge enhancement from noisy images,” *IEEE Trans. Pattern Anal. Mach. Intell.* **25**, 1582–1596 (2003).
 - ⁴³C. E. Metz, “ROC methodology in radiologic imaging,” *Invest. Radiol.* **21**, 720–733 (1986).
 - ⁴⁴J. P. Egan, G. Z. Greenberg, and A. I. Schulman, “Operating characteristics, signal detectability, and the method of free response,” *J. Acoust. Soc. Am.* **33**, 993–1007 (1961).
 - ⁴⁵C. E. Metz, B. A. Herman, and J. H. Shen, “Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data,” *Stat. Med.* **17**, 1033–1053 (1998).
 - ⁴⁶K. Suzuki, I. Horiba, and N. Sugie, “Efficient approximation of neural filters for removing quantum noise from images,” *IEEE Trans. Signal Process.* **50**, 1787–1799 (2002).
 - ⁴⁷K. Suzuki, F. Li, S. Sone, and K. Doi, “Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network,” *IEEE Trans. Med. Imaging* **24**, 1138–1150 (2005).
 - ⁴⁸K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. (Academic, San Diego, 1990).
 - ⁴⁹B. Sahiner, H. P. Chan, and L. Hadjiiski, “Classifier performance estimation under the constraint of a finite sample size: Resampling schemes applied to neural network classifiers,” *Neural Networks* **21**, 476–483 (2008).
 - ⁵⁰B. Sahiner, H. P. Chan, and L. Hadjiiski, “Classifier performance prediction for computer-aided diagnosis using a limited dataset,” *Med. Phys.* **35**, 1559–1570 (2008).
 - ⁵¹K. Fukunaga and R. R. Hayes, “Estimation of classifier performance,” *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 1087–1101 (1989).
 - ⁵²Q. Li and K. Doi, “Reduction of bias and variance for evaluation of computer-aided diagnostic schemes,” *Med. Phys.* **33**, 868–875 (2006).
 - ⁵³Q. Li and K. Doi, “Comparison of typical evaluation methods for computer-aided diagnostic schemes: Monte Carlo simulation study,” *Med. Phys.* **34**, 871–876 (2007).
 - ⁵⁴Q. Li, “Reliable evaluation of performance level for computer-aided diagnostic scheme,” *Acad. Radiol.* **14**, 985–991 (2007).
 - ⁵⁵K. Suzuki, J. Xu, and I. Sheu, in *Principal-component massive-training machine-learning regression for false-positive reduction in computer-aided detection of polyps in CT colonography*, in Machine Learning in Medical Imaging (MLMI), Vol. 6357 (Springer-Verlag, Berlin, Beijing, China, 2010), pp. 182–189.