Computer-Aided Staging of Lymphoma Patients With FDG PET/CT Imaging Based on Textural Information

C. Lartizien, *Member, IEEE*, M. Rogez, E. Niaf, and F. Ricard

***Abstract*—We have designed a computer-aided diagnosis system to discriminate between hypermetabolic cancer lesions and hypermetabolic inflammatory or physiological but noncancerous processes in FDG PET/CT exams of lymphoma patients. Detection performance of the support vector machine (SVM) classifier was assessed based on feature sets including 105 positron emission tomography (PET) and Computed tomography (CT) characteristics derived from the clinical practice and from more sophisticated texture analysis. An original feature selection method based on combining different filter methods was proposed. The evaluation database consisted of 156 lymphomatous and 32 suspicious but nonlymphomatous regions of interest. Different types of training databases including either the PET and CT features or the PET features only, with or without feature selection, were evaluated to assess the added value of multimodality and texture information on classification performance. An optimization study was conducted for each classifier separately to select the best combination of parameters. Promising classification performance was achieved by the SVM classifier combined with the 12 most discriminant PET and CT features with a value of the area under the receiver operating curve of 0.91.**

***Index Terms*—Computed tomography (CT), classification, computer-assisted diagnostic (CAD), image texture analysis, positron emission tomography (PET).**

I. Introduction

THE initial staging and response assessment of cancer remain difficult tasks for radiologists despite the development of diagnostic imaging modalities such as positron emission tomography (PET), multiparametric magnetic resonance (MR) imaging, or contrast computed tomography (CT) scans. Several human observer studies have proven that diagnostic accuracy can be significantly improved by combining different modalities. As an example, PET imaging of the fluorodeoxyglucose (18F-FDG), a glucose analog labeled with a fluor positron emitter combined with a CT scan provides a combination of anatomic and metabolic information that offers advantages over PET alone [1]. However, integrating such a large amount of visual information remains a complex task, especially for junior radiologists with less expertise. This is all the more challenging as malignant and benign tissues may look similar in one type of modality and different in another one. 18F-FDG PET/CT, for instance, has become the recommended exam for the diagnostic and therapeutic follow-up of many cancers, especially lymphoma [2]. This modality is characterized by a very high sensitivity but often lacks specificity due to FDG uptake in noncancerous inflammatory or infectious lesions [3]–[5] or in physiological normal tissues such as brown fat or thymus [6], [7]. This nonspecific uptake can potentially lead to diagnostic discordance, especially during therapeutic follow-up, where it is crucial to determine if residual FDG uptakes reflect remaining cancer lesions or subsequent therapeutic effects or inflammatory localizations [8], [9].

Radiologists may benefit from improved statistical image processing methods to help them facing the difficult diagnostic cases. Computer-aided diagnosis (CAD) systems are well suited for that purpose. While such systems have been widely studied for mammography, CT, and MRI, there have been only a few studies in PET and PET/CT oncology imaging [10]–[12]. We have recently designed an automated diagnostic system that performs guided localization and classification of small and low-contrast lesions in 3-D PET imaging [13]. This CAD system was shown to significantly outperform human detection sensitivities. The number of false detections per image, however, remains high (>5 false detection per organ).

The purpose of this study is to develop a decision system that performs the automatic discrimination of cancerous from benign lesions in FDG PET imaging. We aim at achieving high discrimination performance by 1) combining PET and CT discriminant image features into the decision algorithm to benefit from this complementary modalities and 2) integrating textural information that has been shown promising in discriminating benign from malignant lesions in PET imaging [10]. We focus on prototyping a diagnostic system where the radiologist interviews the CAD on preselected suspicious areas of interest, thus avoiding the challenging task of localization and classification of abnormalities on a per voxel basis. This study builds on a preliminary work that evaluated the diagnostic performance of two classifiers, the support vector machine (SVM) and random decision forest (RDF), each combined with a series of 105 PET and CT characteristics derived from the clinical practice

Manuscript received November 6, 2012; revised April 4, 2013 and July 23, 2013; accepted September 5, 2013. Date of publication September 27, 2013; date of current version May 1, 2014.

C. Lartizien, M. Rogez, and E. Niaf are with the Université de Lyon, CREATIS; CNRS UMR5220; Inserm U1044; INSA-Lyon; Université Lyon 1, 69621 Villeurbanne Cedex, France (e-mail: carole.lartizien@creatis.insa-lyon.fr; emilie.niaf@creatis.insa-lyon.fr; matthieu.rogez@liris.cnrs.fr).

F. Ricard is with the Hospices Civils de Lyon, Département de Médecine Nucléaire, Hôpital Lyon Sud, 69495 Pierre-Benite, France (e-mail: fabien.ricard@chu-lyon.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JBHI.2013.2283658

[standardized uptake value (SUV) peak, for instance] and from more sophisticated texture analysis [14]. The main novelty and strength of this study is that we conducted an extensive analysis to optimize the different elements of the decision system. As an example, we evaluated the gain of multimodality by comparing the diagnostic performance achieved with training databases including either the PET and CT features or the PET features only. Similarly, we estimated the potential added value of texture information over standard clinical metrics such as SUV for this diagnostic task.

One additional original result of this study is an efficient feature selection method combining different "filter" approaches. This method was compared with the genetic algorithm (GA) and the principal component analysis (PCA) which was successfully applied to similar diagnostic challenges in PET neuroimaging [15].

This paper is organized as follows: The clinical PET/CT image database is presented in Section II. Section III shows an overview of the general classification scheme and details the description of the different elements of the decision system. This includes the description of the image features in Section III-A, the description of the original selection method in Section III-B and a short summary of the SVM classifier in Section III-C. Section IV describes the analysis that was conducted to evaluate the detection performance achieved with the different CAD schemes. The results and discussion sections are presented in Sections V and VI, respectively.

II. Database Description

A. Study Population

Twenty-five patients (11 males and 14 females) aged 20 to 79 year-old at diagnosis were included. They had undergone enhanced CT and FDG PET in an integrated exam at baseline for aggressive lymphoma (B-cell lymphoma or Hodgkin disease).

B. FDG PET/CT Protocol

Enhanced FDG PET/CT exams were conducted in daily practice conditions, meaning fasting patients with blood glucose level check prior to the intravenous injection of 5.2 MBq/kg of 18F-FDG. PET/CT acquisition was conducted approximately 60 min after the 18F-FDG injection, using a Gemini PET/CT device (Philips Medical Systems). Patient workup included a cervical–to–upper-thigh transmission contrast-enhanced CT scan with normal but shallow breathing (2 rows of detectors, 120 kV, 100 mAs, rotation: 0.5 s, slice: 3.2 mm, increment: 2.5, pitch: 1.35, injection of 90–120 ml of contrast agent containing 350-mg I/ml Iobitridol, Xenetix 350 at a rate of 3 ml/s). This resulted in CT reconstructed images of $512 \times 512 \times 191$ voxels with dimension $1.17 \times 1.17 \times 2.5$ mm^3. This was followed by a 3-D PET scan (emission data, 3 min per bed position, 6–9 bed positions, bed length 18 cm, overlap 9.6 cm, field of view 57.6 cm). PET data were corrected for attenuation with a CT-based attenuation map and reconstructed with the iterative RAMLA algorithm resulting in reconstructed images of $144 \times 144 \times 234$ voxels with isotropic voxels of 4 mm.

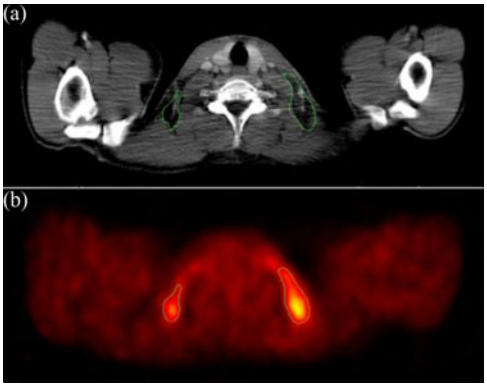

Fig. 1. Example transverse slices of (a) enhanced CT and (b) PET of a patient with benign FDG uptake related to supraclavicular brown fat. This uptake is underlined by the green ROI overlaid over the CT and PET images.

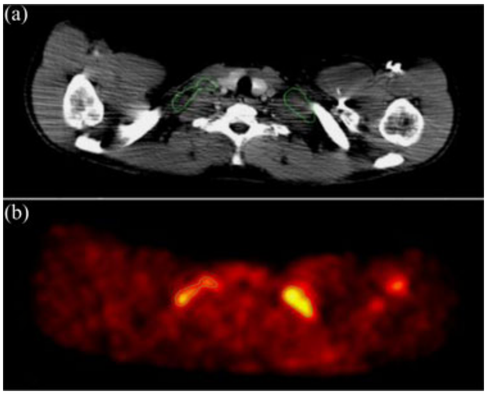

Fig. 2. Example transverse slices of (a) enhanced CT and (b) PET of a patient with malignant FDG uptake related to supraclavicular lymphomatous nodes.

C. Preprocessing and Annotations

The PET images were linearly interpolated to the same voxel size as the CT images and rescaled to SUV values normalized by patient body weight to mimic the common clinical practice. Regions of interest (ROI) fitting the hypermetabolic processes were manually drawn on the PET images by a nuclear physician, then reproduced and adjusted if necessary on the CT images to fit the anatomical contours. When a tissue of interest was visible on several consecutive slices, it was outlined on each slice. The term "ROI" then refers to the stack of those 2-D region delineations. They were categorized as hypermetabolic lymphomatous disease sites (nodal and extranodal sites) or hypermetabolic suspicious but nonlymphomatous sites (brown fat, inflammation, infection, physiologic thymic uptake, etc.) based on experts interpretation, patients follow-up, and biopsy results when available. A total of 156 lymphomatous ROIs (M for malignant) and 32 suspicious nonlymphomatous ROIs (NS for normal suspicious) were analyzed. Fig. 1 shows a transverse slice through the CT [see Fig. 1(a)] and PET [see Fig. 1(b)] images of a patient with two regions of physiological FDG uptake in the brown fat of the cervical supraclavicular area. Similarly, Fig. 2 highlights two regions of lymphomatous FDG uptake in the cervical supraclavicular regions. The mean ROI volume was 27 cm^3 for the suspicious but normal regions, 8.2 cm^3 for the cancerous lymph nodes, and 27 cm^3 for the extranodal cancerous sites (mainly located in the bone, liver, head and neck, and spleen).

III. Description of the CAD System

This paper focuses on a two-class classification problem (cancer lesion-absent and cancer lesion-present). The first step consists in extracting a vector V of some image-based features for each ROI. These features are the input data of a supervised classifier which outputs a malignancy score. An original feature selection protocol combining different filter methods was proposed to select the most discriminant features before the classification step.

A. Feature Extraction

Texture information has been extensively used in the field of pattern recognition. Recent studies focused on their application in PET oncology imaging, mainly in the field of lesion segmentation for radiation treatment planning [10] and for therapy response prediction [16]–[18]. Yu *et al*., for instance showed that some texture features derived from the FDG-PET and CT images, such as PET coarseness or PET busyness, could provide good discrimination between abnormal and normal tissue for head and neck cancer. Tixier *et al*. also investigated the predictive impact of 38 global and local texture features on response to therapy for FDG imaging of the esophageal cancer. They showed that local tumor metabolic nonuniformity characterized by second-order statistics were able to provide nonresponder, partial responder, and complete responder patient identification with higher sensitivity than the standard SUV value. This bibliographic analysis led us to select 105 features extracted from the PET and CT images, including first- and second-order information.

The gray levels within each ROI (PET or CT) were linearly scaled to NGL $=$ 16 levels (NGL stands for "number of gray levels") considering the minimum and maximum ROI voxel values as the lower and upper bounds as suggested in [17]. This allows normalizing voxel intensity among the different patients and reducing noise before the extraction of the texture coefficients.

The mean variation of SUV within each ROI ranged from 0.65 to 15 with a mean value of 3.45, while the mean CT Hounsfield values ranged from 65 to 1935 with a mean value of 296. The 16-levels local resampling was thus sufficient to compromise accuracy and computing time.

1) First-Order Feature: These global characteristics of the gray level frequency distribution included computation of the minimum (min), maximum (max), and mean values over each ROI as well as the corresponding standard deviation (std). The kurtosis and skewness coefficients were also calculated. Kurtosis, for instance, depicts the shape of the gray level distribution relative to a normal distribution, while skewness characterizes the symmetry of the distribution. We also included the PET SUV peak value, which was computed as the mean of the voxel of maximum value in each ROI and its 26 neighbors. This resulted in 7 PET and 6 CT features for each ROI.

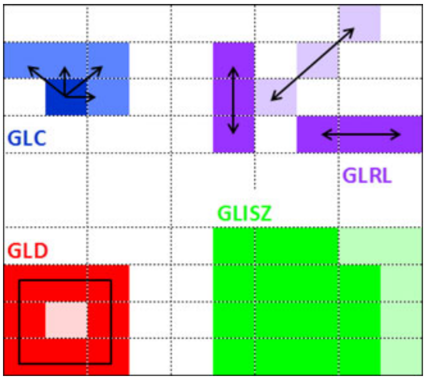

Fig. 3. Schematic description of the four texture matrices.

2) Textural Features: These second-order statistical measures consider the relationship between groups of neighboring voxels in the image. Those used in this study were derived from four matrices illustrated on Fig. 3. All these matrices, originally computed for 2-D images and regular ROI, were adapted in this study to 3-D images and to irregularly shaped ROIs as proposed by Yu *et al*. [10].

a) Gray Level Co-Occurrence Matrix (GLC) [19]–[21]: The GLC element $\mathrm{GLC}_{\theta;,d}(i,j)$ is the joint probability density of the occurrence of gray levels i and j for two voxels separated by the voxel distance d in the angular direction θ. The GLC matrix is a square matrix of dimension NGL image gray levels. One matrix is computed for each value of θ and d. It is the basis for the computation of the Haralick coefficients such as energy, entropy, and correlation [19] as well as coefficients such as "cluster prominence" or "cluster shade" that were more recently proposed by Soh and Clausi [20], [21]. The exhaustive list of the 21 resulting features is given in Table A in the Appendix.

In this study, the GLC matrix was computed in 3-D considering an average value over the 4 angular directions, $\theta = 0°$, $45°$, $90°$ and $135°$ and a voxel distance $d = 1$. We also evaluated the influence of lower frequency variations on the classification performance by considering a distance $d = 3$. As mentioned previously, NGL was set to 16.

b) Gray Level Run-Length Matrix (GLRL) [22], [23]: The GLRL element $\mathrm{GLRL}_{\theta}(i,j)$ is the joint probability density of the occurrence of alignment of j voxels (run length $=$ j) of gray levels i with a spatial relationship defined by the angle θ. As for the GLC matrix, it is an NGLxNGL matrix parameterized by the angle θ. This allows the computation of 11 features described in Table B in the Appendix for each of the PET and CT ROIs. The maximum run length considered in this study was set to 16. As for the GLC matrix, the GLRL matrix was computed in 3-D considering an average value over the 4 angular directions, $\theta = 0°$, $45°$, $90°$ and $135°$.

c) Gray Level Intensity Size Zone Matrix (GLISZ) [24]: The $\mathrm{GLISZ}(i,j)$ element is the probability density of the occurrence of an homogeneous area of j voxels of intensity i. It allows the computation of 11 features following the same definition as that of the GLRL features. In this study, the maximum size of a homogeneous area was set to 16.

d) Gray Level Difference Matrix (GLD) [25]: The GLD(i, j) element is the probability density of the occurrence of a gray levels difference i between a voxel of intensity j and the mean intensity of its neighbors. The GLD matrix enables the computation of three features, namely the coarseness, contrast, and busyness which are thought to correlate well with the human impression. This provided three PET and three CT features for each ROI.

Each of the 188 ROIs defined in Section II-B was thus characterized by a feature vector constituted of 105 elements corresponding to seven PET and six CT first-order features, 21 PET and 21 CT GLCM features, 11 PET and 11 CT GLRL features, 11 PET and 11 CT GLISZ features, and three PET and three CT GLD features.

B. Feature Selection

Reducing the feature space dimension aims to discard uninformative characteristics in order to prevent overfitting, speed up the learning process as well as improve the model's interpretability. The standard approaches produce subsets of discriminative features either by selecting among the original data (data selection) or by deriving new representatives based on the combination of the original set (data reduction). Among the feature selection methods, one can distinguish the "filter" from the "wrapper" approaches. The filter methods propose to rank the features according to some metrics, like statistical tests, and select the best ranked features from this list by setting a threshold on the metric. The main disadvantage of these methods is that they are independent of the classifier performance and strongly rely on the threshold value which in most cases is set arbitrarily. These methods, however, are fast to compute and reproducible. The "wrapper" methods on the other hand, search for the optimal subset tailored to the performance of the classification algorithm. These methods may, however, be computationally intensive and suffer from a weak reproducibility.

We proposed an original method that combines six filter approaches with the purpose of avoiding the arbitrary threshold selection step of the filter approach while keeping fast computation time and reproducible results. This method is as follows:

Six filter methods based on different metrics were first implemented to rank the 105 features independently from the classifier.

This included first the computation of the standard distance metric for each feature defined by

$$\text{dist} = |\overline{x^i_{M}} - \overline{x^i_{\text{NS}}}| / \sqrt{(\sigma^i_{M})^2 + (\sigma^i_{\text{NS}})^2} \tag{1}$$

where $\overline{x^i_M}$ and σ^i_M are the mean and standard deviation of the distribution of feature i and class M computed over the training samples.

The second criterion is the H statistic derived from a Kruskal–Wallis nonparametric one-way analysis of variance by ranks which tests the hypothesis that the two features sets corresponding to the NS and *M* populations came from two distributions with equal median.

The third one is the area under the receiver operating curve (ROC) derived from each of the 105 features separately.

Two indices were also computed from the redundancy and relevance metrics [26], [27] using mutual information (I) as the association parameter. Relevance evaluates the association between feature x^i and the target class l, while redundancy quantifies the correlation between the different features as

$$\text{relevance}\ (x^i) = I(x^i, l) \tag{2}$$

$$\text{redundancy}\ (x^i) = \frac{1}{K} \sum_j I(x^i, x^j) \tag{3}$$

where x^i and x^j are two of the K features.

Combination of these two criteria can be additive or multiplicative thus leading to two metrics referred to as MID and MIQ, respectively, where D and Q stand for difference and quotient

$$\begin{cases} \text{MID}(x^i) = \text{relevance}(x^i) - \text{redundancy}(x^i) \\ \text{MIQ}(x^i) = \text{relevance}(x^i) / \text{redundancy}(x^i). \end{cases} \tag{4}$$

The relevance criterion was also included as the sixth criterion.

The second step of this feature selection method consisted in selecting the most discriminant features from these lists. We proposed to compromise the ranking order achieved by each metric separately by selecting the features that were ranked at least two times among the ten best ranked features by each of the six selection methods. This method, referred to as TOP10 in the following, first characterizes each feature by a metric referred to as "TOP10 frequency" that totalizes the number of times (ranging from 0 to 6) it appears in the top ten best ranked features by each of the six filter method. The final selected feature list is then obtained by ranking the features by decreasing order of this "TOP10 frequency" metric and setting a threshold of 2.

C. Supervised Classification With SVM

This study was based on the SVM [28] which consists in deriving a predictive model from a set of training data from two classes based on a feature vector V. This model is then applied to any test image and outputs a decision variable λ_{SVM} that can be thresholded to determine the test image class. SVM classifiers provide low generalization error even with small learning sample datasets. They are based on a solid theoretical background thus leading to reproducible performance and finally, showing a good robustness to noisy or mislabeled data. We have recently shown that they can achieve good detection performance in FDG PET images [13].

The SVM approach is based upon the statistical learning theory [28]. Consider a set of examples (V_n, l_n), where V_n is the vector of K features of sample $n, V_n \in \Re^K, l_n$ is the class of sample $n, l_n \in (-1, +1)$, and $n = (1, 2, \ldots, N)$. The basic idea of the SVM classifier is to find an optimal hyperplane $\omega \cdot V + b$ that maximizes the separation margin $(2/\|\omega\|)$ between the closest data points of each class, referred to as the support vectors. The optimal classifier can be found by solving the quadratic optimization problem

$$\begin{cases} \min \frac{1}{2}\|\omega\|^2 \\ l_n\left((\omega_n \cdot V_n) + b\right) \geq 1, \quad n = (1, 2, \ldots, N). \end{cases} \tag{5}$$

The decision variable $\lambda_{\mathrm{SVM}}(V_n)$ is the signed distance separating V_n from the hyperplane

$$\lambda_{\mathrm{SVM}}(V_n) = \omega \cdot V_n + b. \tag{6}$$

SVM allows solving nonlinear problems by mapping data into a higher dimensional space by means of a kernel function. They also account for nonseparable classes by allowing classification errors which are parameterized by a penalization parameter C.

In this study, all features were normalized prior to classification by applying a standard range transformation for each feature separately so that all values range between 0 and 1. We chose the radial basis function as the kernel function which is parameterized by the standard deviation σ (or $\gamma = 1/2\sigma^2$) of the Gaussian kernel. The values of C and γ were derived based on the optimization study described in Section IV-A. Membership probability outputs were estimated from distances to the margin using Platt's algorithm [29].

IV. Performance of the CAD System

A. Analysis Protocol

1) Evaluation of the TOP10 Selection Method: The TOP10 selection was first applied on the whole series of 188 (32 NS and 156 M) feature vectors, considering separately the 105 PET and CT features and the 53 PET features. This allowed deriving a list of selected features based on the whole sample database.

We then evaluated the reproducibility of the TOP10 method with different random combinations of the data. Six separate subdatabases were indeed generated by randomly drawing 80% of the 188 feature vectors from the original database following a stratified resampling strategy, i.e., guarantying that these databases will contain 17% of NS samples and 83% of M targets as in the original database. The TOP10 selection method was then applied separately for each of these subdatabases, resulting in six different ranked lists of features. We then calculated how many times the TOP10 frequency metric of each specific feature was greater than or equal to 2. This value divided by the number of trials (six in our study) varies from 0 to 1 and estimates the selection frequency of this specific feature. A value of "0" means that this feature was never selected by the TOP10 method in any of the six trials, while a value of "1" indicates that this feature was always selected. This metric referred to as "TOP10 confidence" in the following was computed for each of the selected features.

Performance of the TOP10 method was finally compared to that achieved by one reference wrapper method, the GA [30] which was configured with the following parameters: population size = 5, maximum number of iterations = 30, tournament selection scheme. We also conducted the comparison with one classical data reduction method, the PCA [31] where the number of eigenvectors was set to represent 95% of the total variance.

2) Optimization of the SVM Intrinsic Parameters: An exhaustive grid search was run for the SVM classifier so as to cover the space of γ between 0.001 and 16 using 11 regular intervals on a $\log_2$-scale and the space of C between 0.125 and 4096 using 16 intervals on a $\log_2$-scale, thus resulting in the evaluation of 176 couples of C and γ parameters. A leave-one-out (LOO) cross validation was conducted to estimate classification performance of each couple (C, γ). The LOO strategy avoids to train and test on the same data and is known to produce an unbiased estimation of the general population performance. The basic process of this method consists in using all samples except one to train the SVM and testing then the remaining sample on this predicted model. This process is repeated until each sample is selected as the test one. This allows deriving a predicted score for each sample from which a ROC curve can be computed. The area under the ROC curve (AUC) was used as the quality metric to select the best couple of parameters (C, γ).

An exhaustive grid search analysis was conducted for each combination of the feature vectors (PET/CT features with and without selection, PET features with and without selection, as described in Section IV-B) based on the rapidminer software [32].

3) ROC-Based Performance Evaluation: Performance of the different CAD schemes was first estimated using two different resampling strategies, the LOO cross-validation approach like for the grid search analysis, and the Fukunaga–Hayes (FH) method [33]. Both methods are shown to be well adapted for classifier performance prediction using a limited dataset [34]. The main principle of the FH method is to estimate the classifier performance with different portioning of the database, i.e., different combinations of Ntrain training samples and Ntest test samples, following a hold-out strategy. Fukunaga and Hayes show that the AUC varies linearly with 1/Ntrain and suggest extrapolating the AUC value corresponding to $N \gg$ Ntrain (ideal training based on an infinite number of samples) based on a linear regression between the AUC and 1/Ntrain variables. In this study, the linear regression model was adjusted on 21 points homogeneously sampling Ntrain from 30 to 188 and using 20 replicates per points.

The nonparametric Wilcoxon methodology was then used to fit the experimental ROC curves derived from the LOO analysis and the AUC was selected as the pseudo values in the Dorfman–Berbaum–Metz ANOVA analysis to test for any statistical significance of difference between the different CAD schemes [35]. This statistical analysis was based on the DBM-MRMC software (version 2.33, from the Medical Image Perception Laboratory (http://perception.radiology.uiowa.edu) and the Kurt Rossmann Laboratories for Radiologic Image Research (http://www-radiology.uchicago.edu/krl/).

B. Impact of the Different Parameters on the CAD Performance

We first analyzed the impact of feature selection by comparing the classification performances achieved by the SVM algorithm combined with the PET and CT features and 1) the TOP10 selection method, 2) GA, 3) PCA, and 4) no feature selection.

TABLE I
RANK LIST OF THE MOST DISCRIMINANT PET AND CT FEATURES DERIVED FROM THE TOP10 METHOD*

Features	dist	KW	AUC§	Rel	MID	MIQ	TOP10 frequency (/6)	TOP10 confidence (%)
PET kurtosis	6	7	7 (0.76)	4	103	5	5	66
PET cluster prominence	7	5	5 (0.77)	4	25	7	5	66
CT RL LRLGE	1	3	3 (0.78)	6	60	2	5	83
SUV min	4	4	4 (0.78)	8	49	10	5	100
SUV mean	3	1	1 (0.80)	1	26	80	4	100
SUV peak	8	6	6 (0.77)	9	60	23	4	83
PET sum entropy	2	2	2 (0.78)	3	90	13	4	100
SUV max	10	8	8 (0.76)	12	40	30	3	83
CT mean	39	21	21 (0.72)	2	21	4	2	66
SUV std	16	9	9 (0.76)	11	82	14	2	100
CT ISZ LRLGE	5	17	17 (0.72)	22	10	16	2	83
PET ISZ LRHGE	108	58	58 (0.63)	88	8	1	2	50

*See the appendix for the definition of the acronyms. §The number in brackets corresponds to the AUC value.

TABLE II
RANKED LIST OF THE MOST DISCRIMINANT PET FEATURES DERIVED FROM THE TOP10 METHOD*

Features	dist	KW	AUC§	Rel	MID	MIQ	TOP10 frequency (/6)	TOP10 confidence (%)
SUV min	3	3	3 (0.77)	6	28	6	5	100
SUV std	9	8	8 (0.76)	9	14	8	5	100
Kurtosis	4	6	6 (0.76)	3	19	3	5	100
Cluster prominence	5	4	4 (0.77)	3	21	4	5	100
Sum entropy	1	2	2 (0.78)	2	47	9	5	100
SUV max	8	7	7 (0.76)	10	11	17	4	100
SUV mean	2	1	1 (0.80)	1	20	33	4	100
SUV peak	6	5	5 (0.76)	7	53	13	4	100
ISZ GLN	25	10	10 (0.72)	16	17	7	3	66
RL HGRE	22	22	22 (0.68)	8	43	5	2	50
ISZ SGLGE	20	28	29 (0.65)	18	1	2	2	66
ISZ LRHGE	55	34	34 (0.63)	47	3	1	2	83

*See the appendix for the definition of the acronyms. §The number in brackets corresponds to the AUC value.

We then evaluated the discriminatory performance of the SVM combined with different types of feature vectors and compared them with those achieved by the SVM trained on the series of PET and CT features with and without feature selection. These different feature vectors were composed of the following:

1) The series of 53 PET features. They are referred to as "PET " in the result section.
2) The series of PET features selected from the TOP10 method, referred to as "PET-TOP10."
3) The 13 first-order PET and CT characteristics [see Section III-A1)], referred to as "PET/CT -first order."
4) The seven first-order PET features only, referred to as "PET/CT -first order."

The first two types of vectors were aimed at estimating the added value of PET and CT features over the PET features only while the third and fourth feature sets were aimed at highlighting the influence of texture parameters for this medical diagnostic task.

Finally, as mentioned in Section III-A2a, we also evaluated the influence of the GLC matrix parameters by comparing the discriminatory power achieved with voxel distances $d = 1$ and $d = 3$. This method is referred to as "$\text{GLC}_{d=3}$" in the following.

Note that separate grid search analyses were run for each CAD scheme to determine their optimal intrinsic parameters.

V. RESULTS

A. *Results of the TOP10 Feature Selection Method*

Results of the feature selection described in Section III-B and applied on the whole series of 105 PET and CT features are reported in Table I. Features are ranked separately by decreasing order of each of the six criteria (e.g., the feature giving the highest AUC is ranked first for this criterion, etc.), thus producing six ranking values for each feature (reported in columns 2 to 7 of Table I). The "TOP10 frequency" metric reported in column 8 is then used to sort the features by descending order. Note that the value reported in brackets in column 4 corresponds to the AUC value achieved by each feature separately.

Nine PET and three CT characteristics were selected from the whole series of 105 PET and CT features, including the standard SUV-based metrics evaluated in clinical practice, as well as the PET kurtosis index and five texture parameters including two Haralick coefficients, two characteristics derived from the GLISZ matrix and one from the GLRL matrix. The full names of the texture parameters are given in Tables A and B in the Appendix. Table II reports the characteristics that were selected among the series of 53 PET features. This series of 12 features includes again the clinical SUV metrics and the PET kurtosis index as well as two Haralick coefficients, three indices derived from the GLISZ matrix, and one from the GLRL matrix.

TABLE III
PET/CT AND PET FEATURE SUBSETS DERIVED FROM THE TOP10 SELECTION APPLIED ON THE GLC$_{d=3}$ FEATURE VECTOR*

PET/CT	PET
PET Min	PET Std
PET Kurtosis	PET Min
PET Mean	PET Kurtosis
SUV peak	PET sum entropy
PET sum entropy	PET Max
RL CT LRLGE	PET Mean
PET Max	SUV peak
PET Std	ISZ PET GLN
CT Mean	PET cluster prominence
CT autocorrelation	PET inverse difference
CT Busyness	ISZ PET LRE
ISZ PET SGLGE	ISZ PET SRHGE
ISZ PET SRHGE	
ISZ CT LRLGE	

*See the appendix for the definition of the acronyms.

TABLE IV
OPTIMAL PERFORMANCE OF THE SVM CLASSIFIER COMBINED WITH DIFFERENT SELECTION METHODS APPLIED ON THE PET AND CT FEATURES

	AUC	95% CI*	Nb of features	GS duration (min)
TOP10	0.911	(0.844,0.978)	12	10
PCA	0.892	(0.826,0.959)	21	19
GA	0.903	(0.840,0.967)	64	2100
No-selection	0.878	(0.799,0.955)	105	37

*CI stands for confidence interval.

The last column (column 9) of Tables I and II reports the TOP10 confidence metric derived from the six trials described in Section IV-A1. The majority of the selected features came out in more than 80% of the cases when the TOP10 method was applied on different combinations of the original samples, thus demonstrating the good stability of the TOP10 method.

Table III reports the subsets of selected features from the series of PET/CT (left column) and PET (right column) features based on the GLC coefficients computed with a voxel distance $d = 3$. The clinical PET first-order features (kurtosis, etc.) are well ranked as in Tables I and II thus demonstrating their strong discriminatory power. The selected texture parameters are different from those reported in Tables I and II. The first- and second-order CT parameters are indeed better represented in Table III (five CT features). This might indicate that PET texture parameters capturing lower frequency variations are less discriminant for this specific diagnostic task.

B. Impact of Feature Selection on Classification Performance

Table IV summarizes the classification performance (AUC) achieved by the SVM algorithm combined with the TOP10, GA, and PCA methods as well as the corresponding 95% confidence intervals estimated by the DBM-MRMC software. It also reports the mean computation time to conduct the exhaustive grid search. The TOP10 selection method was shown to process much faster than the GA while achieving slightly better classification performance. Classification and computation performances also compare favorably to those of PCA so we favored the TOP10 method which allows a more straightforward analysis of the salient discriminative features than PCA. Table IV also underlines the improvement achieved when selecting the most discriminant features compared to the AUC value of 0.878 achieved without selection.

TABLE V
OPTIMAL PERFORMANCE OF THE SVM CLASSIFIER WITH DIFFERENT TRAINING DATABASES

	AUC	95% CI*	Nb of features	duration (min)
PET*	0.894	(0.827,0.961)	53	21
PET-TOP10	0.894	(0.821,0.965)	12	10
PET/CT 1st order	0.871	(0.797,0.944)	15	12
PET 1st order	0.825	(0.724,0.926)	8	12
PET/CT GLC$_{d=3}$	0.881	(0.788,0.975)	105	37
PET/CT GLC$_{d=3}$ -TOP10	0.860	(0.761,0.950)	14	11
PET GLC$_{d=3}$	0.880	(0.802,0.957)	53	21
PET GLC$_{d=3}$ -TOP10	0.893	(0.819,0.967)	12	10

*CI stands for confidence interval.

TABLE VI
OPTIMAL PERFORMANCE OF THE SVM CLASSIFIER DERIVED FROM THE FH RESAMPLING METHOD

	PET/CT -TOP10	PET/CT	PET-TOP10	PET
AUC$_{FH}$	0.922±0.0093	0.908±0.012	0.897±0.007	0.919±0.012

Table V shows performance results achieved with the different types of discriminative components. All feature sets achieve high detection performance with AUC values ranging between 0.86 and 0.894, except for the CAD scheme based on the seven first-order PET features whose AUC value is equal to 0.825. The best performances were achieved by the subsets of PET features including the texture parameters with mean AUC values equal to or higher than 0.89 (lines 1 and 2 of Table V). These values did not outperform the AUC value of 0.911 achieved by the SVM classifier combined with the series of 12 selected PET and CT features reported in Table IV. This underlines the positive impact of combining the PET and CT features thus vindicating the multimodality approach for this diagnostic task. The AUC values achieved with the PET or PET and CT features including texture characteristics are also higher than those obtained by accounting for the first-order characteristics only (lines 3 and 4), thus underlining the impact of texture parameters for this discrimination task. The texture parameters computed with a higher voxel distance ($d = 3$ for the GLC matrix) led to overall lower performance than that achieved with $d = 1$.

Table VI shows the AUC values derived from the FH resampling strategies with the following feature vectors: PET and PET/CT with TOP10 feature selection, PET and PET/CT without feature selection. This confirms the ranking order achieved with the LOO strategy (see Tables IV and V) with the best AUC (AUC $= 0.92$) achieved with the series of selected PET and CT features reported in Table I.

The nonparametric global ANOVA analysis was conducted with the DBM-MRMC on the four CAD schemes of Table IV and the eight classification schemes of Table V. This ANOVA analysis indicated that the 12 strategies were not statistically different (with a type I error rate at 0.05), so that we could not perform multiple comparisons based on the AUC differences.

VI. Discussion

We have designed a CAD system to discriminate cancerous from benign but suspicious tissues in FDG PET/CT exams. The best performance (AUC $= 0.911$) was achieved by combining the SVM classifier with a series of texture and SUV-based parameters extracted from the coregistered PET and contrast CT images. This encouraging performance motivates a more extensive clinical validation based on a larger database. We will pay particular attention to increasing the number of suspicious ROI. This clinical evaluation study will be designed to evaluate the gain in accuracy achieved by nuclear physicians of different levels of expertise with the assistance of the CAD scheme.

The highest detection performance was achieved with the series of combined PET and CT features, thus demonstrating the positive impact of combining complementary imaging modalities. Although we could not demonstrate that this performance was statistically higher than that achieved with the series of PET characteristics (AUC $= 0.89$), the observed trend is encouraging. Increasing the size of the sample database may allow achieving sufficient statistical power to conclude that the different CAD schemes do not perform equally.

In their study, Yu *et al.* [10] aimed at discriminating head and neck cancerous (HNC) tissues from high physiological FDG uptakes in normal tissues of this anatomical region. Their study was based on 14 and 15 textural features of PET and CT images, respectively. The best discrimination performance (AUC $= 0.95$) was achieved by combining three textural features, namely the CT coarseness, CT busyness, and PET coarseness and a decision tree-based K-nearest neighbor classifier. The AUC value of 0.95 is higher than the value of 0.91 achieved in this study. The three discriminant features reported previously also do not appear in the best ranked features derived from our analysis. These results cannot be easily compared, first, because the clinical application (lymphoma versus HNC) is different and second, because the discrimination tasks were different. Their study, indeed, focused on FDG uptakes in normal tissues, whereas we were concerned about suspicious FDG uptakes. We also performed another discrimination task [14] where 158 physiological samples extracted from the main organs of interest were added to the 32 suspicious tissues. A full analysis, including the feature selection and classifier optimization steps, was conducted with this new "normal" class and resulted in a best AUC value of 0.97, which is thus closer to the value reported by Yu *et al.*

Other recent studies also addressed the textural characterization of PET imaging but for the specific purpose of tumor therapy response prediction [16]–[18]. They all focused on the PET imaging part and did not explore the potential gain of including CT characteristics as done by Yu and ourselves. In their study, El Naqa *et al.* derived a logistic regression model for cervix cancer and found that the Haralick energy coefficient had a high predictive value [16]. They could not derive similar conclusion for HNC cancer patients where the outcome prediction was mostly driven by SUV and shape descriptive statistics. Tixier *et al.* also investigated the influence of the different types of texture parameters derived from the GLC, GLD, GLISZ, and GLRL matrices in prediction of therapy response in the esophageal cancer [17]. Their study investigated the predictive values of each of 38 features individually, i.e., without any classification or feature combination model. They showed that two features derived from the GLISZ matrix allowed the best patient stratification among the classes of nonresponder, partially responder, and responder patients. Improved performance is likely to be expected by introducing classification strategies, as illustrated in this study. Indeed, the performance achieved by each feature separately (characterized by the AUC value reported in bracket in the fourth column of Tables I and II) was shown to be lower (<0.8) than that achieved by the different classification strategies reported in Tables IV and V.

Another contribution of this study is the original feature selection method (TOP10) based on the combination of six filter approaches. This method allowed achieving equivalent detection to that achieved with the GA or PCA. Compromising the performance and computation time tends thus to favor the TOP10 filter approach. Further investigation is however required to better formalize this method. One question is why the TOP10 method applied on the PET and CT features allows achieving better classification performance than without selection with the standard texture parameter (GLC matrix with $d = 1$, see Table IV) whereas it fails when using the same texture parameters except that the GLC coefficients are computed for a voxel distance equal to 3 (see lines 5 and 6 of Table V).

We also evaluated the performance achieved by an RDF based on the good discrimination power achieved by Yu *et al.* [10]. Classification performances were first estimated in [14] based on the PET and CT features and further extended in this study based on the PET features only (data not shown). They were shown to be highly variable depending on the value of the initialization seed and to be outperformed by the SVM performance. Considering the RDF longer computation time, we thus recommend the SVM as the best classification scheme for this application.

Another concern that should be mentioned when computing texture parameters is to ensure that the ROI contains a sufficient number of voxels. As mentioned in the text, the mean volume for the 132 cancerous lymph nodes was 8.2 cm^3 corresponding to 129 PET voxels of 4 mm × 4 mm × 4 mm, which seems sufficient to compute texture parameter. The ten smallest nodes had volumes around 3 and 4 cm^3 corresponding to 50 to 62 PET voxels and around 900–1200 CT voxels after resampling, which is still reasonable considering the voxel distances $d = 1$ and 3 considered in this study.

Finally, we chose to resample the PET data to the CT spatial resolution, which is likely to impact the values of the texture coefficients by adding some correlations between adjacent voxels. This choice to work at the same spatial resolution was mainly motivated by our short-term objective to extend this ROI-based analysis to a voxel-based analysis, thus challenging the task of lesion detection and classification. We chose to work at the best spatial resolution in order to avoid loss of CT texture information. The alternate choices would be to work at the lowest spatial resolution of the PET images or to extract the texture coefficients at the original resolution of each modality which is feasible with our ROI-based approach. We did not evaluate the impact of this resampling step on the CAD performance, which could be the purpose of a further study.

VII. Conclusion

We have designed a CAD system to discriminate suspicious from cancerous tissues in FDG PET/CT exams of lymphoma patients. This study showed that the combination of image texture analysis and automated decision making based on multimodal information offers a promising approach to this clinical challenge. The best performance (AUC = 0.91) was achieved by combining the SVM classifier with a series of first- and second-order textural parameters extracted from the coregistered PET and contrast CT images. Perspectives include: 1) increasing the PET/CT database, especially regarding the samples of suspicious tissues, 2) conducting an extensive clinical study to assess the impact of such a CAD system on the performance of a panel of radiologists with different levels of expertise, and 3) extending this ROI-based analysis to a voxel-based analysis, thus challenging the task of lesion detection and classification.

Appendix

TABLE A
Definition of the Features Derived From the GLC Matrix

Name	Name
Autocorrelation [20]	Sum of squares [19]
Contrast [19]	Sum average [19]
Correlation [19, 21]}	Sum variance [19]
Cluster prominence [20]	Sum entropy [19]
Cluster shade [20]	Difference variance [19]
Dissimilarity [20]	Difference entropy [19]
Energy [19]	Information measure on correlation 1 [19]
Entropy [19]	Information measure on correlation 2 [19]
Inverse difference [21]	Inverse difference normalized [21]
Inverse difference moment [19]	Inverse difference moment normalized [21]
Maximum probability [20]	

TABLE B
Definition and Acronyms of the Features Derived From the GLRL [22], [23] and GLISZ [24] Matrices

Name	Acronyms
short run emphasis	SRE
long run emphasis	LRE
Grey level nonuniformity	GLN
Run length nonuniformity	RLN
Run percentage	RP
Low grey level run emphasis	LGRE
High grey level run emphasis	HGRE
Short run low grey level emphasis	SRLGE
Short run high grey level emphasis	SRHGE
long run low grey level emphasis	LRLGE
long run high grey level emphasis	LRHGE

Acknowledgment

The authors would like to thank A. Susset for her help with the classifier optimization study.

References

[1] V. Kapoor, B. M. McCook, and F. S. Torok, "An introduction to PET-CT imaging," *Radiographics*, vol. 24, pp. 523–43, 2004.

[2] Y. S. Jhanwar and D. J. Straus, "The role of PET in lymphoma," *J. Nucl. Med.*, vol. 47, pp. 1326–1334, 2006.

[3] U. Metser and E. Even-Sapir, "Increased (18)F-fluorodeoxyglucose uptake in benign, nonphysiologic lesions found on whole-body positron emission tomography/computed tomography (PET/CT): accumulated data from four years of experience with PET/CT," *Seminars Nuclear Med.*, vol. 37, pp. 206–22, 2007.

[4] L. Kostakoglu, R. Hardoff, R. Mirtcheva, and S. J. Goldsmith, "PET-CT fusion imaging in differentiating physiologic from pathologic FDG uptake," *Radiographics*, vol. 24, pp. 1411–31, 2004.

[5] G. J. Cook, E. A. Wegner, and I. Fogelman, "Pitfalls and artifacts in 18FDG PET and PET/CT oncologic imaging," *Seminars Nuclear Med.*, vol. 34, pp. 122–33, 2004.

[6] S. Paidisetty and T. M. Blodgett, "Brown fat: Atypical locations and appearances encountered in PET/CT," *Amer. J. Roentgenol.*, vol. 193, pp. 359–66, 2009.

[7] J. Jerushalmi, A. Frenkel, R. Bar-Shalom, J. Khoury, and O. Israel, "Physiologic thymic uptake of 18 F-FDG in children and young adults: a PET/CT evaluation of incidence, patterns, and relationship to treatment," *J. Nucl. Med.*, vol. 50, pp. 849–56, 2009.

[8] P. Castellucci, C. Nanni, M. Farsad, L. Alinari, P. Zinzani, V. Stefoni, G. Battista, D. Valentini, C. Pettinato, M. Marengo, S. Boschi, R. Canini, M. Baccarani, N. Monetti, R. Franchi, L. Rampin, S. Fanti, and D. Rubello, "Potential pitfalls of 18 F-FDG PET in a large series of patients treated for malignant lymphoma: Prevalence and scan interpretation," *Nucl. Med. Commun.*, vol. 26, pp. 689–94, 2005.

[9] F. M. Paes, D. G. Kalkanis, P. A. Sideras, and A. N. Serafini, "FDG PET/CT of extranodal involvement in non-Hodgkin lymphoma and Hodgkin disease," *Radiographics*, vol. 30, pp. 269–91, 2010.

[10] H. Yu, C. Caldwell, C. Mah, I. Poon, J. Balogh, R. MacKenzie, N. Khaouam, and R. Tirona, "Automated radiation targeting in head-and-neck cancer using region-based texture analysis of PET and CT images," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 75, pp. 618–625, 2009.

[11] Y. Nie, L. Q. Q. Li, F. Li, Y. Pu, D. Appelbaum, and K. O. Doi, "Integrating PET and CT information to improve diagnostic accuracy for lung nodules: a semiautomatic computer aided method," *J. Nucl. Med.*, vol. 47, pp. 1075–1080, 2006.

[12] G. V. Saradhi, G. Gopalakrishnan, A. S. Roy, R. Mullick, R. Manjeshwar, K. Thielemans, and U. Patil, "A framework for automated tumor detection in thoracic FDG PET images using texture-based features," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2009, pp. 97–100.

[13] C. Lartizien, S. Marache-Francisco, and R. Prost, "Automatic detection of lung and liver lesions in 3D positron emission tomography images: A pilot study," *IEEE Trans. Nucl. Sci.*, vol. 59, no. 1, pp. 102–12, Feb. 2012.

[14] C. Lartizien, M. Rogez, A. Susset, F. Giammarile, E. Niaf, and F. Ricard, "Computer aided staging of lymphoma patients with FDG PET/CT imaging based on textural information," presented at IEEE Int. Symp. Biomed. Imag., Barcelona, Spain, 2012.

[15] I. A. Illán, J. M. Górriz, J. Ramírez, D. Salas-Gonzalez, M. M. López, F. Segovia, R. Chaves, M. Gómez-Rio, C. G. Puntonet, and A. D. N. Initiative, "18 F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis," *Inform. Sci.*, vol. 181, pp. 903–16, 2011.

[16] I. E. Naqa, P. W. Grigsby, A. Apte, E. Kidd, E. Donnelly, D. Khullar, S. Chaudhari, D. Yang, M. Schmitt, R. Laforest, W. L. Thorstad, and J. O. Deasy, "Exploring feature-based approaches in PET imaging for predicting cancer treatment outcomes," *Pattern Recog.*, vol. 42, pp. 1162–1171, 2009.

[17] F. Tixier, C. C. Le Rest, M. Hatt, N. Albarghach, O. Pradier, J.-P. Metges, L. Corcos, and D. Visvikis, "Intratumor heterogeneity characterized by textural features on baseline 18 F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer," *J. Nucl. Med.*, vol. 52, pp. 369–378, 2011.

[18] J. George, P. Claes, K. Vuncks, S. Tejpar, C. M. Deroose, J. Nuyts, D. Loecks, and P. Suetens, "A textural feature based tumor therapy response prediction model for longitudinal evaluation with PET imaging," presented at IEEE Int. Symp. Biomed. Imag., Valencia, Spain, 2012.

[19] R. Haralick, K. Shanmuggan, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man Cybern.*, vol. SMC-3, no. 6, pp. 610–21, Nov. 1973.

[20] L.-K. Soh, "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, no. 2, pp. 779–795, Mar. 1999.

[21] D. A. Clausi, “An analysis of co-occurrence texture statistics as a function of grey level quantization,” *Can. J. Remote Sensing*, vol. 28, pp. 45–62, 2002.

[22] H. H. Loh, J. G. Leu, and R. C. Luo, “The analysis of natural textures using run length features,” *IEEE Trans. Ind. Electron.*, vol. 35, no. 2, pp. 323–328, May 1988.

[23] X. Tang, “Texture information in run-length matrices,” *IEEE Trans. Imag. Process.*, vol. 7, no. 11, pp. 1602–09, Nov. 1998.

[24] G. Thibault, B. Fertil, C. Navarro, S. Pereira, P. Cau, N. Levy, J. Sequeira, and J.-L. Mari, “Texture indexes and gray level size zone matrix. Application to cell nuclei classification,” in *Proc. Pattern Recog. Inform. Process.*, Minsk, Belarus, 2009, pp. 140–145.

[25] M. Amadasum and R. King, “Textural features corresponding to textural properties,” *IEEE Trans. Syst., Man Cybern.*, vol. 19, no. 5, pp. 1264–74, Sep./Oct. 1989.

[26] C. Ding and H. Peng, “Minimum redundancy feature selection from microarray gene expression data,” *J. Bioinformatics Comput. Biol.*, vol. 3, pp. 185–205, 2005.

[27] H. Peng, F. Long, and C. Ding, “Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy,” *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 27, no. 8, pp. 1226–38, Aug. 2005.

[28] V. Vapnik, *Statistical learning theory*. New York, NY, USA: Wiley, 1995.

[29] J. Platt, “Probabilistic outputs for support vector machines,” in *Advances in Large Margin Classifiers*. Cambridge, MA, USA: MIT Press, 1999.

[30] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA, USA: Addison Wesley, 1989.

[31] I. T. Jolliffe, *Principal Component Analysis*. New York, NY, USA: Springer, 2002.

[32] I. Mierswa, M. Wurst, R. Klinkenberg, M. Scholz, and T. Euler, “YALE: Rapid prototyping for complex data mining tasks,” in *Proc. 12th ACM SIGKDD Int. Conf. Knowledge Discovery Data Mining*, 2006, pp. 935–940.

[33] K. Fukunaga and R. R. Hayes, “Effects of sample size on classifier design,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 8, pp. 873–85, Aug. 1989.

[34] B. Sahiner, H. P. Chan, and L. Hadjiiski, “Classifier performance prediction for computer-aided diagnosis using a limited dataset,” *Med. Phys.*, vol. 35, pp. 1559–70, 2008.

[35] D. D. Dorfman, K. S. Berbaum, and C. E. Metz, “ROC characteristic rating analysis: Generalization to the population of readers and patients with the jackknife method,” *Invest. Radiol.*, vol. 27, pp. 723–31, 1992.

Authors’ photographs and biographies not available at the time of publication.