A 3-D Riesz-Covariance Texture Model for Prediction of Nodule Recurrence in Lung CT

Pol Cirujeda*, Yashin Dicente Cid, Henning Müller, Daniel Rubin, Todd A. Aguilera, Billy W. Loo, Jr., Maximilian Diehn, Xavier Binefa, and Adrien Depeursinge

***Abstract*— This paper proposes a novel imaging biomarker of lung cancer relapse from 3-D texture analysis of CT images. Three-dimensional morphological nodular tissue properties are described in terms of 3-D Riesz-wavelets. The responses of the latter are aggregated within nodular regions by means of feature covariances, which leverage rich intra- and inter-variations of the feature space dimensions. When compared to the classical use of the average for feature aggregation, feature covariances preserve spatial co-variations between features. The obtained Riesz-covariance descriptors lie on a manifold governed by Riemannian geometry allowing geodesic measurements and differentiations. The latter property is incorporated both into a kernel for support vector machines (SVM) and a manifold-aware sparse regularized classifier. The effectiveness of the presented models is evaluated on a dataset of 110 patients with non-small cell lung carcinoma (NSCLC) and cancer recurrence information. Disease recurrence within a timeframe of 12 months could be predicted with an accuracy of 81.3–82.7%. The anatomical location of recurrence could be discriminated between local, regional and distant failure with an accuracy of 78.3–93.3%. The obtained results open novel research perspectives by revealing the importance of the nodular regions used to build the predictive models.**

***Index Terms*— Cancer recurrence, computer-aided diagnosis, lung computed tomography, radiomics, Riesz-covariance model, texture classification.**

I. Introduction

RECURRENCE modeling in cancerous nodules is an ongoing research topic in the medical imaging domain. Current medical imaging acquisition devices generate big amounts of data and meta-data, such as 3-D scans of patient bodies or particular organs, and associated information such as regions of interest (ROI) delineations and other monitoring information according to its clinical relevance by experts (e.g., patient survival, treatment response). This highlights the need for developing methodologies for managing these big data, particularly focusing on medical assessment such as tissue identification, case retrieval or prediction of disease outcome. The latter belongs to an emerging research field called radiomics [1], which is based on recent evidence that imaging biomarkers have the potential to surrogate invasive biopsy-based molecular assays with the ability to capture intralesional heterogeneity in a non-invasive way [2], [3]. Personalized phenotyping of the manifestation of the disease can be used to diagnose, assess treatment response, and predict prognosis with higher precision [4], [5].

In the particular case of non-small cell lung carcinoma (NSCLC), several studies highlighted the value of imaging biomarkers for predicting patient outcomes [6]. The proportions of solid and ground-glass opacities (GGO) are known to be a predictor of prognosis [7], [8] and invasiveness [9]. The size and shape of solid tumor components was related to cancer aggressiveness [6]–[8], [10]. Ganeshan *et al.* showed multiple correlations between texture features based on 2-D multiscale isotropic Laplacian of Gaussian (LoG) filters and histopathology [11], glucose metabolism and stage [12] as well as patient survival [13], [14]. 2-D LoG filters were also used by Ravanelli *et al.* to predict tumor response to chemotherapy in [15]. Al-Kadi *et al.* were able to predict tumor aggressiveness with high accuracy using texture attributes based on 2-D fractal analysis and corresponding lacunarity. One important limitation of the above-mentioned studies is that texture properties are most often averaged over the gross tumor volume (GTV) and do not explicitly characterize GGO and solid components. The mixed values of texture attributes computed over entire nodular regions are not unequivocal and entail the risk of regrouping tissue with distinct morphological properties and "habitats" [16] (see Figure 1). Mattonen *et al.* [17] separated the analysis of tumor components, although focusing on GGO regions only. In previous work, we specifically learned texture models of solid and GGO nodule components to predict nodule recurrence [18]. The models were based on linear

Manuscript received June 20, 2016; accepted July 12, 2016. Date of publication July 18, 2016; date of current version November 29, 2016. This work was supported by the Swiss National Science Foundation under Grant PZ00P2_154891 and the CIBM. *Asterisk indicates corresponding author.*

*P. Cirujeda is with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona 08018, Spain (e-mail: pol.cirujeda@upf.edu).

Y. D. Cid and H. Müller are with the Institute of Information Systems, University of Applied Sciences Western Switzerland (HES-SO), Sierre 3960, Switzerland and also with the University Hospitals and University of Geneva, Geneva 1205, Switzerland.

D. Rubin is with the Department of Radiology and Medicine (Biomedical Informatics), Stanford University, Stanford, CA 94305 USA.

T. A. Aguilera, B. W. Loo, Jr., and M. Diehn are with the Department of Radiation Oncology and Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA 94305 USA.

X. Binefa is with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona 08018, Spain.

A. Depeursinge is with the Institute of Information Systems, University of Applied Sciences Western Switzerland (HES-SO), Sierre 3960, Switzerland and also with the Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland.

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/TMI.2016.2591921

0278-0062 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

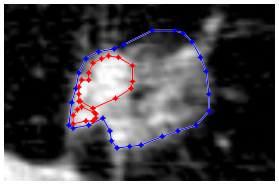


Fig. 1. Example of a lesion with GTV (blue) and solid (red) components annotated in an axial CT slice. The GGO region was excluding the two solid ROIs in this case.

combinations of 2-D steerable Riesz-wavelets learned using support vector machines (SVM).

3-D image analysis approaches are required to best leverage the wealth of information contained in modern radiological images. In particular, 3-D texture features and descriptors can capture complex structural tissue properties and can be used for tissue modeling, segmentation and classification [19]. Imaging biomarkers of NSCLC based on 3-D texture analysis of nodular components are scarce. Parmar *et al.* used 3-D GLCMs and run-length texture analysis along with 2-D coiflet wavelets, 3-D shape and intensity features to predict patient survival, histology and tumor stage [20]. The predictive performance was estimated using a large cohort of 647 patients with NSCLC and showed to be poor for all three outcomes. The latter can potentially be related to the importance of separating the analysis of nodular components (e.g., solid and GGO), which was not the case in this study.

This paper investigates the fusion of visual texture characterization techniques, as provided in previous research lines of the authors, and classification models that can relate the visual observations on computed tomography (CT) scans with the patient's temporal recurrence classes, as annotated by expert clinicians as part of their monitoring. The novelty is to find correlated relationships between observed nodule regions of interest on control cases with known recurrence risk, and to obtain predictive models for newly-observed patients in a quantitative manner. These models are formulated by 3-D Riesz-covariance descriptors for observed texture samples—in the complete 3-D delineated nodule regions—and associated classification models in this non-linear descriptor space, considering different recurrence time windows as different sample classes. The presented method does not use any geometry or structural prior information regarding the nodule boundaries, which is not always easy to determine in an unsupervised manner. It relies on the statistical notion of the covariance of feature observations for modeling texture distributions within regions of interest of arbitrary size. When compared to our previous study based on 2-D Riesz texture models of nodule recurrence (see [18]), the main novelty of the current approach is its natural extension to the 3-D domain via the spatial relation of 3-D Riesz-wavelet features through the covariance-based descriptors. The previous approach used the axial CT with the maximum nodule surface for classification, where the responses of the Riesz filters were aggregated based on the average. The latter discarded spatial co-variations between features, and therefore mixing texture properties of distinct nodular components (see Fig. 1). 3-D Riesz-covariance descriptors do not require a selection of a 2-D slice with the maximum observable area, as all voxel samples within a three-dimensional ROI are considered as observations of a multi-dimensional joint distribution of features, encoding the texture characteristics regardless of its size or region shape. When used as an aggregation function, the average loses the spatial localization of the Riesz filter responses. On the other hand, the feature covariances encode the spatial co-responses of the Riesz filters, providing much richer tissue representations. Overall, the proposed approach provides a flexible framework in which any arbitrary tissue region can be evaluated, regardless of size, shape and connectivity, without the requirement for manual input. The covariance-based descriptors lie on the space of Riemannian geometry, which follows the trend of state-of-the-art work [21], [22], and provides a coherent theoretic framework in order to built classification methods that are sensitive to the topological properties of the described data, e.g. recurrence and locality failure prediction models based on tissue component features.

The rest of the paper is organized as follows: Section II introduces the presented methodology including II-A: the dataset of patients with early stage NSCLC and treated with stereotactic radiation, II-B and II-C: the used texture features and the derived descriptive unit, II-D: an analysis of their Riemannian space geometry for self-contention and II-E: the formulation of two classification algorithms taking into account their spatial constraints — a kernel-based SVM and a manifold-regularized sparse representation. Section III presents the experimental evaluation conducted to test the effectiveness of the presented method for predicting early stage NSCLC recurrence. Finally, Section IV discusses the results and presents our conclusions.

II. Methodology

A. Patient data

A cohort of 110 patients treated with thoracic stereotactic ablative radiotherapy (SABR) from Stanford Cancer Institute was used in the analysis with institutional review board approval. All patients treated with SABR from 2004 to 2013 who had biopsy proven early staging NSCLC were considered but were not included if they received chemotherapy prior to a diagnosis of metastatic disease, had synchronous tumors at diagnosis, or if they received a new diagnosis of primary NSCLC following SABR. Patients were treated using SABR with variable dose fractionation schemes from 25 Gy in one fraction to 60 Gy in five fractions. Treatment planning scans used for analysis were acquired on a General Electric (GE) or Siemens CT simulation scanner used for treatment planning prior to radiation therapy. CT scans had 1.25 to 2.50 mm slice thickness with most scans having in-plane pixel spacing of 0.98 mm with a range of from 0.7 to 1.37 mm. Patients were followed in 3 month intervals at first and monitored for response and recurrence using both clinical exams and cross sectional imaging. The final dataset kept only of patients whose associated CT scan images did not contain fiducial markers or other acquisition artifacts affecting the feature extraction procedure, and with a complete follow-up information along treatment, yielding to the sets defined below.

GTVs were delineated in 3âŁ“D by three treating radiation oncologists from Stanford University who specialize in

thoracic malignancies and stored in the Digital Imaging and Communications in Medicine (DICOM) radiation therapy (RT) format. Each radiation oncologist had 4-18 years of experience at the time of patient treatment. For this study a single radiation oncologist delineated the GGO and solid components of the treated GTV using the lung and mediastinum CT windows. MATLAB was used for the post-processing of CT images and data, including region ground-truth preparation and resampling of volumes in order to have isotropic voxels of $0.8 \times 0.8 \times 0.8$ mm^3 using cubic spline interpolation. The nodules had a reported size ranging from 7 to 1,868 mm^3.

Disease-free survival (DFS) times were collected for 92 patients and divided into three categories:

- patients with "short-term recurrence": cancer relapse occurred within the first 12 months of treatment (23 cases);
- patients with "long-term recurrence": cancer recurrence occurred within the first 24 months of treatment (30 cases);
- patients with "no recurrence" within 24 months of treatment (62 cases).

The patients in the "long-term recurrence" group included all patients of the "short-term recurrence" group.

A second data categorization according to type of recurrence, regardless relapse times, was collected for 32 patients (the 30 previous and two with recurrence after more than 24 months), which are labeled as:

- "local", if the cancer nodule reappeared within the lung parenchyma of the same lobe of the treated lesion (12 cases);
- "regional", metastasis developed in regional hilar or mediastinal lymph nodes (13 cases);
- "distant", if disease recurrence developed at distant metastatic sites including the brain (7 cases).

B. 3-D Riesz-wavelet features

3-D

multiscale Riesz filterbanks are used to characterize the texture of the lung parenchyma in 3-D CT. Each component of the 2$^{\text{nd}}$-order Riesz transform $\boldsymbol{\mathcal{R}}^{(N)}$ of a three-dimensional signal $f(\boldsymbol{x})$ is defined in the Fourier domain as:

$$\widehat{\mathcal{R}^{(n_1,n_2,n_3)}} f(\boldsymbol{\omega}) = (-j)^N \sqrt{\frac{N!}{n_1!n_2!n_3!}} \frac{\omega_1^{n_1}\omega_2^{n_2}\omega_3^{n_3}}{||\boldsymbol{\omega}||^{n_1+n_2+n_3}} \hat{f}(\boldsymbol{\omega}), \quad (1)$$

for all combinations of (n_1, n_2, n_3) with $n_1 + n_2 + n_3 = N$ and $n_{1,2,3} \in \mathbb{N}$. Eq. (1) yields $M = \binom{N+2}{2}$ components $\mathcal{R}^{(n_1,n_2,n_3)}$, which corresponds to all-pass filters pertaining to the 3-D phase only, i.e., the directions defined by the Nth-order partial derivatives of $f(\boldsymbol{x})$ along x, y, z. Band-limited filters are obtained by coupling the Riesz transform with a multi-resolution framework based on non-separable isotropic wavelets (e.g., Simoncelli) [23].

Altogether, the Riesz components $\mathcal{R}^{(n_1,n_2,n_3)}$ form steerable filterbanks, meaning that the local response of each Riesz component to an image rotated by an arbitrary 3-D angle can be derived analytically from a linear combination of the responses of all components of the filterbank.

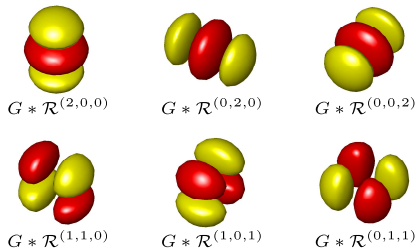


Fig. 2. 2$^{\text{nd}}$-order Riesz kernels $\mathcal{R}^{(n_1,n_2,n_3)}$ convolved with isotropic Gaussian kernels $G(\boldsymbol{x})$.

This work uses the second-order Riesz filterbank (depicted in Figure 2), which qualitatively corresponds to the six unique elements of the Hessian matrix. When compared to lower or higher orders of the Riesz transform, the second-order provides an appropriate balance between the wealth and the dimensionality of the filterbanks. Rotation-covariance is obtained by locally aligning the Riesz components $\mathcal{R}^{(n_1,n_2,n_3)}$ of all scales based on the locally prevailing orientation. The latter is estimated using the uni-directional Riesz components as proposed in [24]. This estimation relies on computing the regularized structure tensor proposed by Chenouard *et al.* [25] using only the three Riesz filters $\mathcal{R}^{(2,0,0)}$, $\mathcal{R}^{(0,2,0)}$ and $\mathcal{R}^{(0,0,2)}$. The sorted collection of eigenvectors of the resulting 3-dimensional matrix composes a rotation matrix $\mathbf{U}_g \in \mathbb{R}^{3\times3}$, where g is the regularization function used when building the structure tensor, in our case a 3-D Gaussian window. Applying Theorem 1 of [26], the steering matrix $\mathbf{S}_{\mathbf{U}_g} \in \mathbb{R}^{M\times M}$ corresponding to $\mathbf{U}_g$ can be obtained and used to steer the Riesz components while preserving their directional inner-relations. This steering matrix is computed from the wavelet coefficients of the first scale using an isotropic Gaussian regularization window with $\sigma = 1$ and is applied to all scales. For every scale s_i with $i \in \{1, 2, 3\}$, locally aligned second order Riesz features are denoted as $\widetilde{\boldsymbol{\mathcal{R}}}^{(2)}_{s_i}$ in the next sections.

C. Riesz-covariance 3-D texture descriptors

Covariance matrices have been used as descriptors in the computer vision and pattern recognition area for feature fusion, big data characterization and multimodal signal processing. Tuzel *et al.* [27] presented an object recognition method for 2-D color images using visual cues such as edges, curvature or color values inside a region of interest of arbitrary size. Fehr *et al.* [28] extended this framework to other domains such as 3-D object recognition in unstructured point clouds. Other approaches, such as [29]-[31] also support the feasibility of covariance-based descriptors for metric learning or spatio-temporal gesture recognition.

As defined in Eq. (1), 2$^{\text{nd}}$-order 3-D Riesz-wavelet features yield to a 6-dimensional filterbank for each scale s_i. Three dyadic scales are chosen to cover the spatial spectrum of nodular regions, leading to a total of 18 Riesz-wavelet features. For a posterior task of region characterization and classification, a compact yet accurate representation of these region-collected

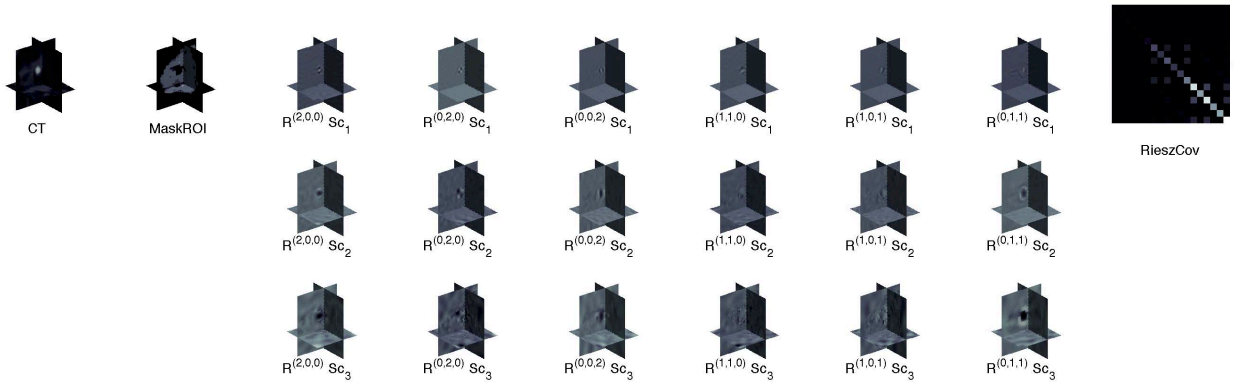


Fig. 3. Feature volumes involved in the descriptor calculation for a given cubic CT volume of $56 \times 56 \times 56$ pixels containing a delineated nodule region. Volumes are plotted by a 3-D representation including three axial slices. The left-most image shows the CT values, followed by the delineated region. Middle subfigures depict the 2^{nd} order 3-D Riesz-wavelet responses for the three wavelet decomposition levels. The 18×18 matrix in the right-most subfigure depicts the resulting 3-D Riesz-covariance descriptor, encoding all covariances between the distributions of the observed texture features. All the voxels within the delineated ROI where considered as samples of an 18-dimensional joint distribution, for the 18 available Riesz-wavelet response features. The resulting multiscale descriptor provides a discriminative signature for the entire 3-D ROI, and encode the tissue characteristics in terms of texture intensities, directionality and scale homogeneity.

features is needed. The feature values obtained for each voxel need to be aggregated over a given region of interest (ROI) to estimate their statistical distributions. The complete space of Riesz-wavelet features constitutes a sparse variety, with a lack of characterization for particular texture entities. Following previous research in which we investigated covariance-based descriptors for shape and texture fusion of 3-D surfaces [21], tissue characterization in 3-D CT imaging [22] or 2-D color image categorization [32], we propose to exploit Riesz-wavelet features in their covariance space. The latter can be locally estimated by computing the covariance matrix of the Riesz features in a given ROI.

The benefits of translating feature magnitudes within a ROI to the space of the covariances of these observations are threefold:

(i) The consideration of voxels within a 3-D region as samples of a multi-dimensional feature distribution implies a loss of structural information, which leads to the robustness to spatial and rotation transformations (due to the steerability of 3-D Riesz-wavelet features) [21], [22].
(ii) The characterization of a set of feature observations by its covariance matrix provides a compact, size and shape independent discriminative signature [32].
(iii) Considering the texture pattern distributions in terms of relationships between Riesz-wavelet features provides a meaningful dictionary of texture possibilities, instead of modelling a sparser space of all the possible feature magnitudes taken separately.

In order to formally define the 3-D Riesz-covariance descriptors, we denote a feature selection function $\Psi(f(\boldsymbol{x}), \boldsymbol{v})$ for a given 3-D CT image $f(\boldsymbol{x})$ and a selected subvolume region $\boldsymbol{v}$ as:

$$\Psi(f(\boldsymbol{x}), \boldsymbol{v}) = \left\{ \widetilde{\boldsymbol{\mathcal{R}}}_{s_i}^{(2)}(p), \forall p \in \boldsymbol{v} \subset \mathbb{R}^3 \right\}, \tag{2}$$

where $\widetilde{\boldsymbol{\mathcal{R}}}_{s_i}^{(2)}(p)$ is the concatenation of the six 3-D Riesz features for each scale s_i.

Then, for a given 3-D ROI $\boldsymbol{v}$ of the CT image, the associated 3-D Riesz-covariance descriptor is:

$$Riesz_{Cov}^{3D}\left(\Psi(f(\boldsymbol{x}), \boldsymbol{v}\right) = \frac{1}{M-1} \sum_{i=1}^{M} \left(\boldsymbol{\psi}_i - \boldsymbol{\mu}\right)\left(\boldsymbol{\psi}_i - \boldsymbol{\mu}\right)^T, \tag{3}$$

where $\boldsymbol{\mu}$ is the vector mean of the set of vectors $\Psi = \{\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_M\}$ within the volumetric neighborhood $\boldsymbol{v}$ made of M samples.

The resulting 18×18 matrix $Riesz_{Cov}^{3D}$ is a symmetric matrix where the diagonal entries represent the variance of each feature channel and the non-diagonal elements represent their pairwise covariance, as seen in Figure 3. The pairing of Riesz-wavelet features with a covariance-based framework is particularly interesting for robust tissue modelling with respect to possible rotations or shape variabilities for further classification tasks. The steerability of the used features ensures that the norm of the wavelet responses remains unaltered for any rotation of the observations. The inherent loss of structural information on the construction of covariance matrices, due to the consideration of feature values as samples of a joint distribution, provides invariance to spatial transformations, shape variabilities, scaling and occlusions.

D. Riemannian geometry of the descriptor space

3-D Riesz-covariance descriptors are embodied by covariance matrices that not only provide a compact and flexible representation but also lie in the Riemannian manifold of symmetric definite positive matrices Sym_d^+ [33]. This has a major impact on their interest as descriptive units, as their spatial variety is geometrically meaningful: 3-D regions sharing similar texture characteristics remain close areas in this descriptor space.

Since symmetric matrices contain only $d(d+1)/2$ independent coefficients, in their upper or lower triangular parts, it is possible to apply a vectorization operation in order to obtain

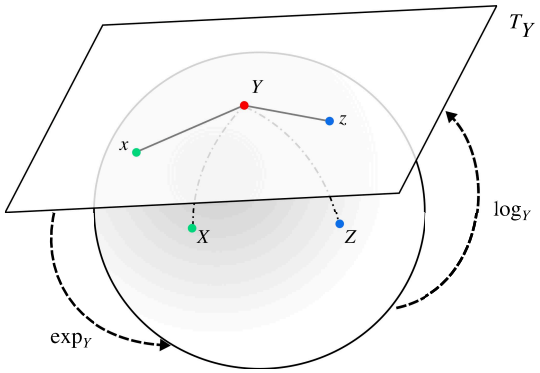


Fig. 4. Mapping of points in a Sym_d^+ manifold to the tangent space T_Y.

a linear orthonormal space for the independent coefficients:

$$\begin{aligned}\hat{\mathcal{X}} &= \text{vec}(X) \\ &= (X_{1,1}, \sqrt{2}X_{1,2}, \ldots, \sqrt{2}X_{1,d}, X_{2,2}, \sqrt{2}X_{2,3}, \ldots, X_{d,d}),\end{aligned} \tag{4}$$

where the $\sqrt{2}$ coefficient in the non-diagonal elements is used for preserving the norm magnitude of the original symmetric matrix and its vectorized form $\|X\|_F = \|\hat{\mathcal{X}}\|_2$ (where $\|X\|_F$ is the Frobenius norm of the matrix X). The obtained vector $\hat{\mathcal{X}}$ lie in the Euclidean space $\mathbb{R}^m$, where $m = d(d+1)/2$.

While this vectorization approximation could be used for the comparison of 3-D Riesz-covariance descriptors, its inherent simplification does not fully exploit the structure of this descriptor space. Instead, the Riemannian manifold can be approximated in close neighborhoods by the Euclidean metric in its tangent space, T_Y, where the symmetric matrix $Y \in Sym_d^+$ is a reference projection point in the manifold [34]. T_Y is formed by a vector space of $d \times d$ symmetric matrices (e.g., corresponding to $Riesz_{Cov}^{3D}$ in our case), and the tangent mapping of a manifold element X to $x \in T_Y$ is made by the point-dependent $\log_Y$ operation (Figure 4):

$$x = \log_Y(X) = Y^{\frac{1}{2}} \log\left(Y^{-\frac{1}{2}} X Y^{-\frac{1}{2}}\right) Y^{\frac{1}{2}}, \tag{5}$$

where the matrix logarithm is computed as

$$\log(X) = U \log(D) U^T, \tag{6}$$

with U and D being the elements of the singular value decomposition (SVD) of $X \in Sym_d^+$.

Analogously, the exponential mapping of a point $x \in T_Y$ returns its original point representation $X \in Sym_d^+$:

$$X = \exp_Y(x) = Y^{\frac{1}{2}} \exp\left(Y^{-\frac{1}{2}} x Y^{-\frac{1}{2}}\right) Y^{\frac{1}{2}}, \tag{7}$$

where the matrix exponential is computed using the SVD of $x \in T_Y$ as:

$$\exp(x) = U \exp(D) U^T. \tag{8}$$

Pennec *et al.* [33] pointed to the following Riemannian metric that defines the scalar product between two points x_1, x_2 in the tangent space at the reference point Y, T_Y:

$$\langle x_1, x_2 \rangle_Y = \text{trace}\left(x_1 Y^{-1} x_2 Y^{-1}\right). \tag{9}$$

The decision of a projection point in the manifold that is equivalent to the origin of coordinates in the tangent space is not trivial. Since the mean of a set of points is, by definition, the point which minimizes the distances with respect to all points in the set. It can be used as the optimal projection element yielding to the closer tangent space approximation. Due to the convexity of the Sym_d^+ manifold, the mean of a set of covariance matrices $X_{j=1..J}$ on a Riemannian manifold has to be approximated in order to lay on the manifold ensuring:

$$\mu(\{X\}) = \underset{X' \in Sym_d^+}{\text{argmin}} \sum_{j=1}^{J} \delta^2\left(X_j, X'\right), \tag{10}$$

where $\delta^2(X_1, X_2)$ expresses the Riemannian distance between two points $X_1, X_2 \in Sym_d^+$, as defined by [35]:

$$\delta(X_1, X_2) = \sqrt{\text{trace}\left(\log\left(X_1^{-\frac{1}{2}} X_2 X_1^{-\frac{1}{2}}\right)^2\right)}, \tag{11}$$

or more simply $\delta(X_1, X_2) = \sqrt{\sum_{i=1}^{d} \log(\lambda_i)^2}$, where λ_i are the positive eigenvalues of $X_1^{-\frac{1}{2}} X_2 X_1^{-\frac{1}{2}}$.

Several gradient descent procedures are proposed in [36], [37] for the computation of the mean in an iterative manner. Pennec *et al.* [33] presented the following method for computing the geodesic mean of a finite set of points in Sym_d^+, $X_1, ..., X_J$ by an iterative re-projection to the tangent space:

$$\begin{aligned}\hat{\mu}(\{X\})_{t+1} &= \exp_{\mu_t}\left(\frac{1}{J}\sum_{j=1}^{J} \log_{\mu_t}(X_j)\right) \\ &= \mu_t^{\frac{1}{2}} \exp_{\mu_t}\left(\frac{1}{J}\sum_{j=1}^{J} \log\left(\mu_t^{-\frac{1}{2}} X_j \mu_t^{-\frac{1}{2}}\right)\right) \mu_t^{\frac{1}{2}}.\end{aligned} \tag{12}$$

Due to the high dimensionality of manifold points, several dimensionality reduction techniques such as multidimensional scaling, isomap or discriminant diffusion map analysis were presented in the literature [38]-[40]. These techniques provide simpler representations of data samples while preserving their similarity relationships in a lower order space. Their goal is to further perform classical data analysis techniques in the machine learning and classification domains such as linear discriminant analysis, data separability or kernel optimization. In order to evaluate the feasibility of the proposed descriptor space for class differentiation, the previously defined tangent mapping operation was used to project the descriptors to an associated class space of type of recurrence, as determined by clinicians. Using discriminant diffusion map analysis [40] with the defined metric in Eq. 11, we projected the points to a 3-D space preserving their Riemannian distance, as depicted in Figure 5. Such a projection demonstrates the following: *a)* the provided 3-D Riesz-covariance descriptors are able to capture several class entities, *b)* the provided Riemannian metrics and mapping operators are able to provide an adequate kernel for classification, and *c)* this classification separability correlates with clinical knowledge on classes like recurrence

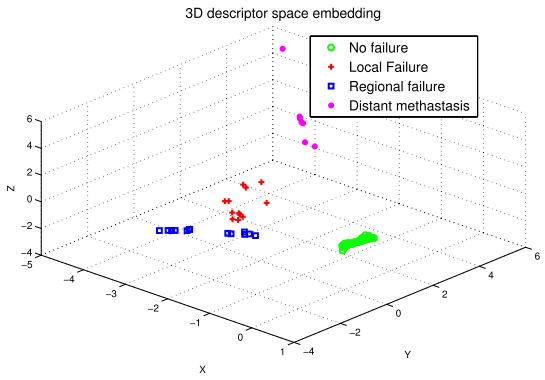


Fig. 5. Descriptor space embedding for the classes of type of recurrence on the follow-up information from ground-truth annotations. Discriminant diffusion map analysis is used together with the Riemannian metrics of the 3-D Riesz-covariance descriptor space for plotting the set of 110 patient nodule delineations in 3-D, together with the labeled information on the dataset. The plot demonstrates the characterization capabilities of the proposed framework, in which texture information preserves a correlation with nodule recurrence information.

locality of the nodules and recurrence time annotations, as is analyzed in this article.

E. Classification methods on the Sym_d^+ manifold

The topological layout of the proposed 3-D Riesz-covariance descriptor yields to focus on geometrically sensitive models that can exploit the Riemannian manifold distribution of the data samples for classification. According to the associated clinical meta-data there are two scenarios of interest for nodule recurrence classification:

(i) a binary separation whether a patient is prone to suffer recurrence within a given time frame or not and
(ii) a multi-class separation according to the possible anatomical localizations of cancer recurrence: no failure, local failure, regional failure, or distant metastasis.

This section includes the presented formulation for both scenarios in which a kernel-based support vector machine (SVM) framework is chosen for binary classification and a manifold-regularized sparse classifier is presented for multi-class datasets.

1) Binary classification via a kernel SVM formulation: Support vector machines constitute a very common classification method in the machine learning literature [41]. In its linear form, this supervised learning methodology attempts to separate a set of S labeled samples $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1}^{S}$ by finding a hyperplane with a normal vector $\boldsymbol{w}$ such that the separation margin, i.e. the distance between this hyperplane and the nearest point of each class, is maximized. The classification decision function is defined as the projection of:

$$h(\boldsymbol{x}) = b + \sum_{i=1}^{S} \alpha_i y_i \langle \boldsymbol{x}_i, \boldsymbol{x} \rangle = b + \langle \boldsymbol{w}, \boldsymbol{x} \rangle, \tag{13}$$

where the term b is a bias magnitude and $\{\alpha_i\}$ are the Lagrangian multipliers that take values only on the support vectors [42]. $\langle \cdot, \cdot \rangle$ denotes the L_2-norm inner product between vectors.

This basic formulation can be extended to the case in which data samples are not linearly separable in their natural distribution by mapping the samples into another transformed space [43]. In the case of 3-D Riesz-Covariance descriptors of nodule regions, the mapping to the tangent space arises in a straightforward way. The introduction of this mapping function ϕ yields to a slight change to Eq. (13) as follows:

$$h(\boldsymbol{x}) = b + \sum_{i=1}^{S} \alpha_i y_i \langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}) \rangle. \tag{14}$$

An associated kernel function $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j) \rangle$ is usually defined instead of an implicit mapping function ϕ. In the case of covariance matrices for signal processing, Yger [44] recently reviewed several kernel formulations that yield useful high-dimensional space mappings.

The natural choice for a kernel-based SVM for the classification of covariance-based descriptors is the tangent mapping function defined in Eq. (5): $\phi_Y(X) = \log_Y(X)$. Using the scalar product in the tangent space and the complementary Riemannian metrics previously defined we can derive the following kernel for symmetric positive-definite matrices:

$$k(X_i, X_j)_Y = \langle \phi(X_i)_Y, \phi(X_j)_Y \rangle, \tag{15}$$

where $Y \in Sym_d^+$ is the projection reference point, set to the mean of the training sample and set as in Eq. (12) in order to maximize the mapping likelihood. Related to this definition with the inner product defined in Eq. (9) we can derive the following kernel:

$$\begin{aligned} k(X_i, X_j)_Y &= \text{trace}\left(\log_Y(X_i) Y^{-1} \log_Y(X_j) Y^{-1}\right) \\ &= \text{trace}\left(\log\left(Y^{-\frac{1}{2}} X_i Y^{-\frac{1}{2}}\right) \log\left(Y^{-\frac{1}{2}} X_j Y^{-\frac{1}{2}}\right)\right). \end{aligned} \tag{16}$$

Finally, in order to enable a computationally efficient implementation of this kernel-based SVM, the following condition is verified thanks to the trace operator and the tangent mapping function defined in Eq. (5):

$$\begin{aligned} k(X_i, X_j)_Y &= \text{trace}\left(Y^{-\frac{1}{2}} \log_Y(X_i) Y^{-\frac{1}{2}}\ Y^{-\frac{1}{2}} \log_Y(X_j) Y^{\frac{1}{2}}\right) \\ &= \langle x_i, x_j \rangle. \end{aligned} \tag{17}$$

Therefore, we can use this kernel trick to perform an equivalent kernel-based classification in the tangent projected space of the 3-D Riesz-covariance descriptors, with respect to the projection reference point at their mean. We refer to [44] for an exhaustive study on covariance-oriented kernels for classification. The same formulation was recently used in electroencephalography signal processing [45], where signals were characterized by their low-level covariance matrices with successful results. Therefore, we propose to test the performance of the same methodology with the developed 3-D Riesz-covariance descriptors.

2) Multiclass classification using sparse regularized Manifolds: Sparse representation-based methods [46], [47] have shown recent relevance in the machine learning community. These models are specially targeted to complex classification tasks where only a low number of learning

samples might be available, and where a sample to be classified might require strong constraints in order to avoid ambiguous classification decisions. This yields to a sparse model that integrates a regularization term according to the used descriptor space. In the case of 3-D texture analysis, where few patients are available and tissue texture provides a complex class definition, we propose a new formulation adapted to the manifold of 3-D Riesz-covariance descriptors.

As a brief overview in their most general formulation, sparse representation-based classifiers consider a test sample $\boldsymbol{y}$ as a linear combination of elements in a dictionary Φ of training samples from different classes: $\boldsymbol{y} = \Phi\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is the sparse vector indicating the weight coefficients for each element in Φ. The intuition is that the test sample $\boldsymbol{y}$ should be ideally represented, as accurately as possible, by using the smallest number of samples. This sparsity restriction is found via the L_1-norm minimization of $\boldsymbol{\alpha}$:

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \left\{ \|\boldsymbol{\alpha}\|_1 + \|y - \Phi\boldsymbol{\alpha}\|_2^2 \right\}. \tag{18}$$

Then, given $\hat{\boldsymbol{\alpha}}$, the classification label for $\boldsymbol{y}$ is determined by the subset of training samples of a given class i that provides the minimum representation error:

$$class(\boldsymbol{y}) = \underset{i}{\operatorname{argmin}} \left\{ e_i \text{ s.t. } e_i = \|\boldsymbol{y} - \Phi_i\hat{\boldsymbol{\alpha}}_i\|_2 \right\}. \tag{19}$$

This initial approach shares similar fundamentals as in the classical nearest neighbor or nearest subspace classifiers. Eq. (18) intuitively represents an unknown sample as a possible combination of all elements in Φ, but this *collaborativeness* is discarded afterwards as the minimization of the residuals in Eq. (19) determines the closest distance to a single class with the minimal representation error for a unique class decision.

This suggests an important concern: if some subsets of different classes i and j in the training set, Φ_i and Φ_j, are correlated due to similarities in the elements of each class, then the distance between reconstructions $\|e_i\|_2$ and $\|e_j\|_2$ could be very small leading to possible misclassifications. A solution is to avoid the L_1 norm sparsity minimization constraint in Eq. (18) and express the test sample $\boldsymbol{y}$ collaboratively on all the dictionary samples of K classes $\Phi = [\Phi_1, \Phi_2, ..., \Phi_K]$ without forcing any class sparsity prior: then the linear representation solution can be treated as a classical least squares minimization problem:

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \left\{ \|\boldsymbol{y} - X\boldsymbol{\alpha}\|_2^2 \right\}. \tag{20}$$

The main problem is that the solution to this minimization may become unstable and computationally expensive depending on the number of classes or samples (more details can be found in [47]). Therefore, if we want to extend this classification paradigm to the presented 3-D Riesz-covariance descriptor space there is a need for finding a formulation that takes into account its manifold topology and also integrates the sparsity conditions together with the prior knowledge on the geometric descriptor distribution.

This can be achieved by proposing a manifold-aware minimization constraint that relaxes the computational expense

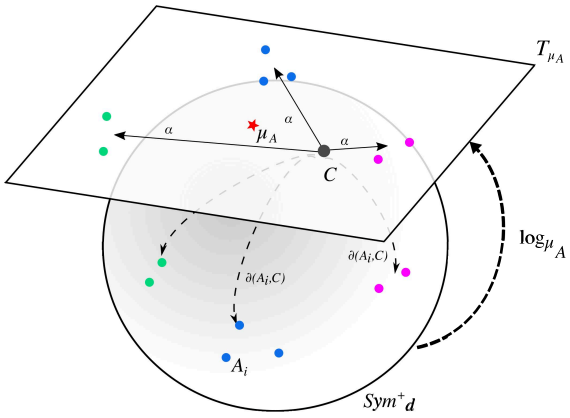


Fig. 6. Schema of the topology of the descriptor space and the sparse classifier integration. The set of $\boldsymbol{\alpha}$ coefficients is computed as an optimization problem on the tangent space with respect to μ_A but the regularization term in the matrix D, which includes the geodesic distances of the samples in the Sym_d^+ manifold penalizes these elements with higher dissimilarity in the descriptor space.

of the method and adds numerical stability. Let A be the whole set of S training samples from K classes, $A = [\{\boldsymbol{a}_1\}, \{\boldsymbol{a}_2\}, ..., \{\boldsymbol{a}_K\}] \in \mathbb{R}^{171\times S}$, where each element in the set $\{a_k\}$ for a class k is a vectorized 3-D Riesz-covariance descriptor sample $\boldsymbol{a}_s = \text{vec}(\log_{\mu_A}(A_s)) \in \mathbb{R}^{171}$, according to Eq. (4). This "dictionary" of training samples is mapped on the tangent space with respect to the mean $\mu_A \in Sym_d^+$ of available learning descriptors. Then, a test sample in the form of a new 3-D Riesz-covariance descriptor $\boldsymbol{c} = \text{vec}(\log_{\mu_A}(C))$ can be expressed as a linear combination of the available set of training samples: $C \approx A\boldsymbol{\alpha}$, being $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, ..., \alpha_s]$ a vector of weights corresponding to each one of the training samples in A. Then, we propose a regularized variation of the minimization expression defined in Eq. (20) as follows:

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \left\{ \|C - A\boldsymbol{\alpha}\|_2^2 + \|D\boldsymbol{\alpha}\|_2^2 \right\}, \tag{21}$$

where D is a diagonal matrix of size $S \times S$ that allows the imposition of prior knowledge on the solution with respect to the training set, using the Riemannian metric defined in Eq. (11). This term contributes to making the least squares solution stable and to introducing sparsity conditions to the vector $\hat{\boldsymbol{\alpha}}$. D is defined as:

$$D = \begin{pmatrix} \delta(A_1, C) & & 0 \\ & \ddots & \\ 0 & & \delta(A_n, C) \end{pmatrix} \tag{22}$$

where A_i and C are the 3-D Riesz-covariance descriptors for training and test samples respectively. See Figure 6 for a schema of the presented classification method. Finally, the solution to the sparse collaborative representation, $\hat{\boldsymbol{\alpha}}$, can be calculated by the following derived expression according to [47]:

$$\hat{\boldsymbol{\alpha}} = \left(A^T A + D^T D\right)^{-1} A^T C, \tag{23}$$

The classification label of the test sample C can be obtained by observing the regularized reconstruction residuals from the

TABLE I
RESULTS FOR THE BINARY CLASSIFICATION OF PATIENT RECURRENCE, USING SHORT- (12 MONTHS) AND LONG-TERM (24 MONTHS) BINARIZATION AND SEVERAL NODULE REGION DESCRIPTORS. PRESENTS THE PERFORMANCE EVALUATION OF THE PRESENTED KERNEL-BASED SVM FORMULATION FOR COVARIANCE-BASED DESCRIPTORS. SHOWS THE RESULTS OF A LINEAR SVM FOR PLAIN VECTORIZED COVARIANCE DESCRIPTORS. FINALLY, ASSESSES THE PERFORMANCE OF A LINEAR SVM USING THE AVERAGE OF THE 3-D RIESZ FILTER RESPONSES WITHIN THE DELINEATED REGION AS TEMPLATES (E.G., CORRESPONDING TO OUR APPROACH IN [18]). THE SHORT-TERM EXPERIMENT INVOLVED 23 RECURRENCES VERSUS 62 REMISSIONS. THE LONG-TERM EXPERIMENT INVOLVED 30 RECURRENCES VERSUS 62 REMISSIONS. ARE EXPRESSED IN TERMS OF CV REPETITION AVERAGES ± STANDARD DEVIATIONS

A)	12 MONTHS – SVM KERNEL				24 MONTHS – SVM KERNEL			
Features	accuracy	sensitivity $(TP/TP+FN)$	specificity $(TN/TN+FP)$	F_1-score	accuracy	sensitivity $(TP/TP+FN)$	specificity $(TN/TN+FP)$	F_1-score
GGO	81.33 ± 0.12	87.38 ± 0.05	78.33 ± 0.13	80.75 ± 0.12	54.94 ± 0.12	61.64 ± 0.14	58.51 ± 0.05	53.74 ± 0.07
Solid	82.00 ± 0.15	85.14 ± 0.13	76.67 ± 0.14	78.13 ± 0.14	57.33 ± 0.05	68.98 ± 0.08	50.89 ± 0.02	49.37 ± 0.03
GTV	82.67 ± 0.17	87.62 ± 0.05	78.33 ± 0.13	80.89 ± 0.12	44.69 ± 0.15	63.33 ± 0.22	47.80 ± 0.15	41.77 ± 0.10

B)	12 MONTHS – LINEAR SVM VECT. COVARIANCES				24 MONTHS – LINEAR SVM VECT. COVARIANCES			
Features	accuracy	sensitivity $(TP/TP+FN)$	specificity $(TN/TN+FP)$	F_1-score	accuracy	sensitivity $(TP/TP+FN)$	specificity $(TN/TN+FP)$	F_1-score
GGO	78.67 ± 0.13	83.62 ± 0.07	75.67 ± 0.17	77.42 ± 0.17	49.63 ± 0.15	58.89 ± 0.05	56.92 ± 0.15	52.30 ± 0.16
Solid	80.67 ± 0.09	84.57 ± 0.05	74.32 ± 0.12	75.89 ± 0.12	57.33 ± 0.11	67.11 ± 0.06	58.01 ± 0.03	56.24 ± 0.09
GTV	81.32 ± 0.15	84.38 ± 0.09	75.72 ± 0.18	76.79 ± 0.18	44.87 ± 0.08	57.71 ± 0.11	48.76 ± 0.07	42.86 ± 0.09

C)	12 MONTHS – LINEAR SVM FOR FEATURES AVERAGES				24 MONTHS – LINEAR SVM FOR MEAN OF FEATURES TEMPLATE			
Features	accuracy	sensitivity $(TP/TP+FN)$	specificity $(TN/TN+FP)$	F_1-score	accuracy	sensitivity $(TP/TP+FN)$	specificity $(TN/TN+FP)$	F_1-score
GGO	74.38 ± 0.08	77.65 ± 0.14	68.97 ± 0.07	69.58 ± 0.07	46.67 ± 0.25	50.00 ± 0.23	50.41 ± 0.23	46.41 ± 0.25
Solid	79.90 ± 0.14	85.96 ± 0.09	76.45 ± 0.11	76.20 ± 0.15	53.33 ± 0.20	55.90 ± 0.23	53.60 ± 0.18	52.04 ± 0.19
GTV	75.62 ± 0.17	83.17 ± 0.15	70.24 ± 0.17	68.97 ± 0.19	51.67 ± 0.15	53.62 ± 0.15	52.17 ± 0.16	46.80 ± 0.15

resulting sparse vector $\hat{\boldsymbol{\alpha}}$:

$$class(C) = \underset{i}{\operatorname{argmin}} \left\{ \frac{\|C - A_i \hat{\boldsymbol{\alpha}}_i\|_2}{\|\hat{\boldsymbol{\alpha}}_i\|_2} \, \omega_i \right\}, \qquad (24)$$

where ω_i is a class balancing coefficient that is set to the normalized cardinality of each class sampling in order to minimize the impact of unbalanced data sets.

III. EXPERIMENTAL EVALUATION

A description of the patient cohort and imaging data used in order to evaluate the performance of the presented predictive recurrence model is provided in Section II-A. 3-D Riesz-covariance descriptors for both the GGO and solid components were computed separately, as well as 3-D Riesz-covariance descriptors for the complete nodule region containing the two areas. Following these conditions, the methods were tested according to the characteristics of the presented classification models: kernel SVM for binary classification, or manifold-regularized sparse representation for multi-class modeling. Ten fold cross-validation (CV) schemes were used to estimate the generalization performance. Classification accuracies and confusion matrices were used to measure and analyze the estimated predictive performance of our models.

A. Short- and long-time recurrence model

We implemented the previously defined kernel-based SVM classifier in order to evaluate the feasibility and performance of binary classification on nodule recurrence. A 10-fold CV was used for parameter optimization and estimation of the generalization performance on 70% of the available data. We built the recurrence classification models using the LibSVM software package, taking into special account the library capabilities for parameter learning in order to minimize the impact of unbalanced data [48]. Finally, with the obtained model hyperparameters, 5 Monte-Carlo repetitions of a 10-fold CV partition on the remaining samples were run in order to estimate the average accuracy of binary recurrence prediction as gathered in Table I. Table I gathers the obtained results for the classification of two recurrence annotation classes: (i) recurrence versus non-recurrence on a 12-months follow-up consideration; and recurrence versus non-recurrence on 24-months annotations. These results are also conditioned to the nodule components considered for different 3-D Riesz-covariance descriptor sets: GGO area, solid region, or the GTV combining both components. In the latter, combining both component features also encodes the different region ratio: the second-order moments encoded in the covariance-based descriptors provide signatures according to the Riesz-wavelet responses for each region. Three model classification formulations are tested: the Riemannian kernel-based SVM classifier as presented in this paper fully exploits the geometric topology of the presented 3-D Riesz covariance-based descriptors, using the mean of the classified descriptor set on the Sym_d^+ descriptor space as the kernel projection point. Second, a linear SVM for plain vectorized 3-D Riesz-covariance descriptors (according to Eq. (4), without the mapping to any tangent space. Finally, a linear SVM using the mean of the features obtained within each delineated region: although using the average of a feature sampling set as an aggregation function can filter out salient feature values we provide this experiment as a straightforward step from our work presented in [18], where the average of 2-D features on the CT slice of maximum diameter was used. We believe the better performance

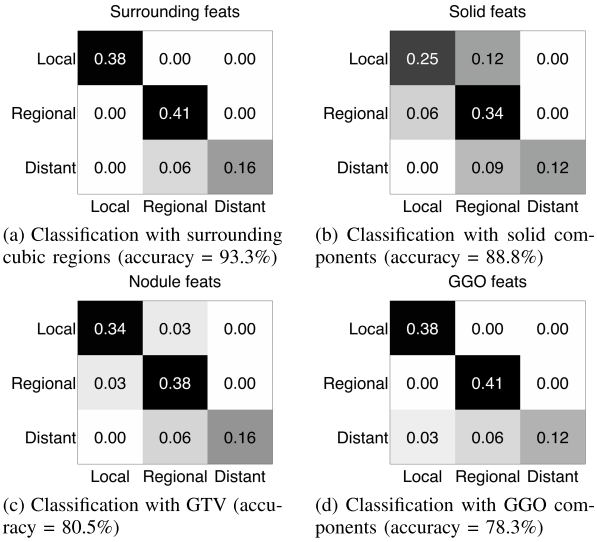
TABLE II
COMPARISON WITH OTHER STUDIES PREDICTING TUMOR RECURRENCE, AGGRESSIVENESS, STAGE, PATIENT SURVIVAL AND HAZARD

Study	number of patients	predicted outcome	separation of tumor components	texture analysis	reported performance
Al–Kadi *et al.* 2008 [49]	15	aggressive vs. less aggress. adeno.	no	2–D	0.83 (accuracy)
Ganeshan *et al.* 2010 [12]	17	below vs. above stage II	no	2–D	0.7 (kappa)
Ganeshan *et al.* 2012 [13]	54	survival (low vs. high)	no	2–D	0.6 (AUC)
Mattonen *et al.* 2014 [17]	46	tumor recurrence (yes/no)	yes	2–D	0.8 (AUC)
Depeursinge *et al.* 2015 [18]	91	tumor recurrence (yes/no)	yes	2–D	0.79 (AUC)
Depeursinge *et al.* 2015 [18]	91	tumor recurrence (patient hazard)	yes	2–D	0.81 (CI)
Parmar *et al.* 2015 [20]	647	survival (patient hazard)	no	2–D & 3–D	0.61 (CI)
Parmar *et al.* 2015 [20]	647	histology	no	2–D & 3–D	0.64 (AUC)
Parmar *et al.* 2015 [20]	647	tumor stage	no	2–D & 3–D	0.64 (AUC)
This work	85	tumor recurrence (yes/no)	yes	3–D	0.83 (accuracy)
This work	32	failure re–localization (local/regional/distant)	no	3–D	0.93 (accuracy)

measurements found in Table I. A support the motivation of the present work: using covariances of features in the 3-D domain as discriminative signatures provides a more sensitive data representation by means of its second-order statistics. Furthermore, lying on a particular geometrical variety allows the development of more accurate classification techniques.

B. Locality recurrence model

Besides temporal recurrence, there is also a particular clinical interest in exploring the relationship of a particular nodule texture and its potential correlation with a failure re-localization. For this reason, a second experiment was conducted to train the manifold-regularized sparse classifier with the provided local annotation from the monitoring of these patients suffering nodule recurrence. The computed descriptors for this classification task were also conditioned by the nodule components used (i.e., GGO, solid and GTV). In addition, we built descriptors associated with a surrounding volume defined as the closest isotropic cube containing the entire nodule region. This cube contains both internal and external regions of the nodule. The motivation to consider such a volume is that for locality classification tasks, some Riesz-wavelet responses belonging to the non-affected tissue around the nodule might contain information on cancer invasiveness and external tumor margin. We implemented the formulation proposed in the previous section with a 5-fold CV scheme. The latter was used in order to generalize as much as possible the method due to the reasonably low number of patients (32 in total) in this dictionary-based classification approach. Results are presented in terms of classification accuracy and associated confusion matrices in Figure 7. Finally, Table II compares the presented model for short- and long-time recurrence prediction, as well as locality recurrence classification, with a review of other state-of-the-art studies. Please note the "reported performance" values provide heterogeneous criteria according to each referenced approach, which are based on different datasets and validation strategies.



(a) Classification with surrounding cubic regions (accuracy = 93.3%)

(b) Classification with solid components (accuracy = 88.8%)

(c) Classification with GTV (accuracy = 80.5%)

(d) Classification with GGO components (accuracy = 78.3%)

Fig. 7. Confusion matrices of the tests with several feature regions for the manifold-regularized sparse classification method, for locality recurrence modeling. Each matrix cell specifies the percentage of true positive elements classified for local (12 cases), regional (13 cases), and distant (7 cases). Different feature regions were tested in order to model and correlate the texture at several nodule levels with the locality recurrence on the learning set of control patients.

IV. DISCUSSION AND CONCLUSIONS

This paper describes a novel framework for leveraging rich 3-D texture information contained in modern radiology imaging protocols such as CT. We focused on building 3-D imaging biomarkers of NSCLC nodule recurrence. The proposed descriptors are based on feature covariance and therefore leverage rich intra- and inter-variations of the feature space, i.e., the local responses of 3-D Riesz-wavelets. These descriptors lie on a manifold governed by Riemannian geometry, which allows geodesic measurements and differentiations between tissue samples. The incorporation of the latter into recent machine learning techniques (i.e., SVMs and sparse regularized manifolds), led to families of manifold-aware classifiers that could be used directly to map the imaging biomarkers to disease outcomes. One major methodological contribution of this work is to use a texture aggregation function that preserves spatial co-variations between features. This represents an important solution to overcome the limitations of classical aggregation functions (e.g., the average), where the interactions between local responses of texture operator pairs are mixed between distinct nodular components.

In a first step, the methods were used to predict NSCLC nodule recurrence from pre-treatment CT scans. The experimental results presented in Section III-A and Table I raise

several interesting considerations for a clinical use of the proposed methods. Choosing a binarization time window of 12 months enables to consider patients who have suffered immediate recurrence. This allows to model the failure and non-failure classes with a tighter correlation with respect to the clinical information we gathered from patient monitoring, and we can consider that there is a stable relationship within texture information and follow-up labels. Accuracies close to 83% place the proposed methodology amongst acceptable performances regarding other methods (see Table II). With a 24-month time span, we observed an expected behavior, which means significantly lower classification accuracy. With this temporal consideration the patient categorization varies, considering numbers of samples and including patients with possibly different biological processes than only the nodule observations. Finally, regarding the kernel-based SVM implementation with respect to a simple descriptor vectorization space as commented in Eq. (4), we observed slightly better performance in the case that takes into account the Riemannian geometry of the descriptor space (see Table I A versus B). This is consistent with other approaches in the literature using similar descriptor paradigms and validates the choice of the 3-D Riesz-covariance descriptors for particular medical imaging applications that can benefit from their 3-D, spatial and rotation invariance. The importance of preserving spatial co-variations between features is highlighted by large predictive performance improvements between covariance- versus average-based feature aggregation in Table I A, B versus C. The results obtained with the average (see Table I C) are consistent with our previous study where solid components of the tumor yielded best results [18]. However, thanks to the covariance-based aggregation, the use of larger ROIs including heterogeneous tissue architectures (e.g., GTV and GGO) improves the performance of the predictive models (see Table I A and B), which was not the case when using feature averages over the GTV or GGO (see Table I C). Considering all these results, we conclude that the predictive models of temporal recurrence from a binomial point of view were encouraging and compare with the state-of-the-art (see Table II). This is not at all a final study and it leaves many continuation lines, particularly in a more formal definition of recurrence time spans. This yields to consider the exploration of continuous time models, such as using the Cox-LASSO [50] or survival SVMs [51] based on the proposed 3-D Riesz-covariance descriptors.

In a second step, the proposed methods were used to predict the type of NSCLC nodule recurrence based on the manifold-regularized sparse classifier. Unlike the previous experiment, the performance of the predictive models was found to be strongly dependent on the type of ROI used (see Figure 7), which indicates that the nodule regions do not all carry information about recurrence locality. In terms of nodular components, it can be observed that the solid parts contain more information than GGO, yielding accuracies of 88.8% and 78.3%, respectively. However, the main observation was that the morphological properties of the lung tissue external to the nodule (i.e., surrounding cubic regions) contain precious information about the type of recurrence. The latter has been little investigated so far with the exception of the study of Dilger *et al.* [52]. Our findings are in accordance with theirs showing that texture properties computed from regions surrounding lung nodules have value in the prediction of nodule malignancy. These observations open new research directions on how to leverage morphological tissue properties to evaluate cancer invasiveness and require further validation and investigations. Future experiments separating the three regions, i.e., surrounding tissue, solid and GGO will be carried out to better define the responsibilities of the latter in cancer relapse. Since our approach focuses on texture properties of nodular tissue, we plan to combine it with other features such as tumor shape, intensity and size, modeling orthogonal information. We recognize several limitations of the current work, including a small number of patients in the three categories of recurrence type, and no sensitivity analysis to the hand-drawn delineation of nodular components.

REFERENCES

[1] S. Napel and M. Giger, "Special section guest editorial: Radiomics and imaging genomics: Quantitative imaging for precision medicine," *J. Med. Imag.*, vol. 2, no. 4, p. 041001, 2015.

[2] R. Lambin *et al.*, "Radiomics: Extracting more information from medical images using advanced feature analysis," *Eur. J. Cancer*, vol. 48, no. 4, pp. 441–446, 2012.

[3] M. Gerlinger *et al.*, "Intratumor heterogeneity and branched evolution revealed by multiregion sequencing," *N. Eng. J. Med.*, vol. 366, no. 10, pp. 883–892, 2012.

[4] A. M. Rutman and M. D. Kuo, "Radiogenomics: Creating a link between molecular diagnostics and diagnostic imaging," *Eur. J. Radiol.*, vol. 70, no. 2, pp. 232–241, May 2009.

[5] R. Mirnezami, J. Nicholson, and A. Darzi, "Preparing for precision medicine," *N. Eng. J. Med.*, vol. 366, no. 6, pp. 489–491, 2012.

[6] W. D. Travis *et al.*, "International association for the study of lung cancer/American Thoracic Society/European Respiratory Society international multidisciplinary classification of lung adenocarcinoma," *J. Thoracic Oncol.*, vol. 6, no. 2, pp. 244–285, 2011.

[7] K. Kodama *et al.*, "Prognostic value of ground-glass opacity found in small lung adenocarcinoma on high-resolution CT scanning," *Lung Cancer*, vol. 33, no. 1, pp. 17–25, 2001.

[8] K. Suzuki, H. Asamura, M. Kusumoto, H. Kondo, and R. Tsuchiya, "Early peripheral lung cancer: Prognostic significance of ground glass opacity on thin-section computed tomographic scan," *Ann. Thoracic Surg.*, vol. 74, no. 5, pp. 1635–1639, 2002.

[9] H. Matsuguma *et al.*, "Comparison of three measurements on computed tomography for the prediction of less invasiveness in patients with clinical stage I non-small cell lung cancer," *Ann. Thoracic Surg.*, vol. 95, no. 6, pp. 1878–1884, 2013.

[10] M. Yanagawa *et al.*, "Prognostic importance of volumetric measurements in stage I lung adenocarcinoma," *Radiology*, vol. 272, no. 2, pp. 557–567, Aug. 2014.

[11] B. Ganeshan *et al.*, "Non-small cell lung cancer: Histopathologic correlates for texture parameters at CT," *Radiology*, vol. 266, no. 1, pp. 326–336, 2013.

[12] B. Ganeshan, S. Abaleke, R. C. D. Young, C. R. Chatwin, and K. A. Miles, "Texture analysis of non-small cell lung cancer on unenhanced computed tomography: Initial evidence for a relationship with tumour glucose metabolism and stage," *Cancer Imag.*, vol. 10, pp. 137–143, Jul. 2010.

[13] B. Ganeshan, E. Panayiotou, K. Burnand, S. Dizdarevic, and K. A. Miles, "Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: A potential marker of survival," *Eur. Radiol.*, vol. 22, no. 4, pp. 796–802, 2012.

[14] T. Win *et al.*, "Tumor heterogeneity and permeability as measured on the CT component of PET/CT predict survival in patients with non-small cell lung cancer," *Clin. Cancer Res.*, vol. 19, pp. 3591–3599, May 2013.

[15] M. Ravanelli *et al.*, "Texture analysis of advanced non-small cell lung cancer (NSCLC) on contrast-enhanced computed tomography: Prediction of the response to the first-line chemotherapy," *Eur. Radiol.*, vol. 23, no. 12, pp. 3450–3455, 2013.

[16] R. A. Gatenby, O. Grove, and R. J. Gillies, "Quantitative imaging in cancer evolution and ecology," *Radiology*, vol. 269, no. 1, pp. 8–14, 2013.

[17] S. A. Mattonen, D. A. Palma, C. J. A. Haasbeek, S. Senan, and A. D. Ward, "Early prediction of tumor recurrence based on CT texture changes after stereotactic ablative radiotherapy (SABR) for lung cancer," *Med. Phys.*, vol. 41, no. 3, p. 033502, 2014.

[18] A. Depeursinge, M. Yanagawa, A. N. Leung, and D. L. Rubin, "Predicting adenocarcinoma recurrence using computational texture models of nodule components in lung CT," *Med. Phys.*, vol. 42, no. 4, pp. 2054–2063, May 2015.

[19] A. Depeursinge, A. Foncubierta-Rodríguez, D. Van De Ville, and H. Müller, "Three-dimensional solid texture analysis and retrieval in biomedical imaging: Review and opportunities," *Med. Image Anal.*, vol. 18, no. 1, pp. 176–196, 2014.

[20] C. Parmar *et al.*, "Radiomic feature clusters and prognostic signatures specific for lung and head & neck cancer," *Sci. Rep.*, vol. 5, p. 11044, Jun. 2015.

[21] P. Cirujeda, Y. Dicente Cid, X. Mateo, and X. Binefa, "A 3D scene registration method via covariance descriptors and an evolutionary stable strategy game theory solver," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 306–329, 2015.

[22] P. Cirujeda *et al.*, "3D Riesz-wavelet based Covariance descriptors for texture classification of lung nodule tissue in CT," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 7909–7912.

[23] M. Unser, D. Sage, and D. Van De Ville, "Multiresolution monogenic signal analysis using the Riesz–Laplace wavelet transform," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2402–2418, Nov. 2009.

[24] Y. D. Cid, H. Müller, A. Platon, P.-A. Poletti, and A. Depeursinge, "Locally-oriented wavelet transforms for 3-D solid texture classification," *IEEE Trans. Image Process.*, to be published.

[25] N. Chenouard and M. Unser, "3D steerable wavelets and monogenic analysis for bioimaging," *in Proc. IEEE Int. Symp. Biomed. Imag. Nano Macro*, Mar. 2011, pp. 2132–2135.

[26] M. Unser, N. Chenouard, and D. Van De Ville, "Steerable pyramids and tight wavelet frames in $L_2(\mathbb{R}^d)$," *IEEE Trans. Image Process.*, vol. 20, no. 10, pp. 2705–2721, Oct. 2011.

[27] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on Riemannian manifolds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1713–1727, Oct. 2008.

[28] D. Fehr *et al.*, "Compact covariance descriptors in 3D point clouds for object recognition," in *Proc. ICRA*, 2012, pp. 1793–1798.

[29] R. Sivalingam, V. Morellas, D. Boley, and N. Papanikolopoulos, "Metric learning for semi-supervised clustering of region covariance descriptors," in *Proc. ACM/IEEE Int. Conf. Distrib. Smart Cameras*, 2009, pp. 1–8.

[30] A. Sanin, C. Sanderson, M. T. Harandi, and B. C. Lovell, "Spatio-temporal covariance descriptors for action and gesture recognition," in *Proc. IEEE Workshop Appl. Comput. Vis.*, Jan. 2013, pp. 103–110.

[31] W. Huang, Z. Lin, J. Yang, and J. Wang, "Text localization in natural images using stroke feature transform and text covariance descriptors," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1241–1248.

[32] P. Cirujeda and X. Binefa, "Medical image classification via 2D color feature based covariance descriptors," in *Proc. Working Notes CLEF*, 2015.

[33] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian framework for tensor computing," *Int. J. Comput. Vis.*, vol. 66, no. 1, pp. 41–66, 2006.

[34] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Log-Euclidean metrics for fast and simple calculus on diffusion tensors," *Magn. Reson. Med.*, vol. 56, no. 2, pp. 411–421, 2006.

[35] W. Förstner and B. Moonen, "A metric for covariance matrices," in *Geodesy-The Challenge of the 3rd Millennium*. Springer, New York, 2003, pp. 299–309.

[36] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Commun. Pure Appl. Math.*, vol. 30, pp. 509–541, Sep. 1977.

[37] M. Moakher, "A differential geometric approach to the geometric mean of symmetric positive-definite matrices," *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 3, pp. 735–747, 2005.

[38] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.

[39] Q. Tian, N. Sebe, M. S. Lew, E. Loupias, and T. S. Huang, "Content-based image retrieval us ing wavelet-based salient points," *Proc. SPIE*, vol. 4315, pp. 425–436, Dec. 2000.

[40] Y. Huang, X. F. Zha, J. Lee, and C. Liu, "Discriminant diffusion maps analysis: A robust manifold learner for dimensionality reduction and its applications in machine condition monitoring and fault diagnosis," *Mech. Syst. Signal Process.*, vol. 34, no. 1, pp. 277–297, 2013.

[41] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[42] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, Sep. 2004.

[43] B. Schölkopf and A. J. Smola, *Learning With Kernels—Support Vector Machines, Regularization, Optimization & Beyond*. Cambridge, MA, USA: MIT Press, 2001.

[44] F. Yger, "A review of kernels on covariance matrices for BCI applications," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Sep. 2013, pp. 1–6.

[45] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Classification of covariance matrices using a Riemannian-based kernel for BCI applications," *Neurocomputing*, vol. 112, pp. 172–178, Jul. 2013.

[46] J. Wright *et al.*, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

[47] D. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 471–478.

[48] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, May 2011.

[49] O. S. Al-Kadi and D. Watson, "Texture analysis of aggressive and nonaggressive lung tumor CE CT images," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 7, pp. 1822–1830, Jul. 2008.

[50] N. Simon, J. H. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for Cox's proportional hazards model via coordinate descent," *J. Statist. Softw.*, vol. 39, no. 5, pp. 1–13, Mar. 2011.

[51] L. Evers and C.-M. Messow, "Sparse kernel methods for high-dimensional survival data," *Bioinformatics*, vol. 24, no. 14, pp. 1632–1638, May 2008.

[52] S. K. N. Dilger *et al.*, "Improved pulmonary nodule classification utilizing quantitative lung parenchyma features," *J. Med. Imag.*, vol. 2, no. 4, p. 041004, 2015.